

IPv4/IPv6 Tutorial

5th August 2015

SANOG 26

Srinath Beldona

srinath_beldona@yahoo.com

Agenda

- Routing Basics
- Review of IPv6 Addressing (Optional)
- Introduction to IPv4/IPv6 OSPF routing
- Or
- Introduction to IPv4/IPv6 ISIS Routing
- Introduction to Multi-Protocol BGP (IPv4/IPv6)
- Conclusion

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::

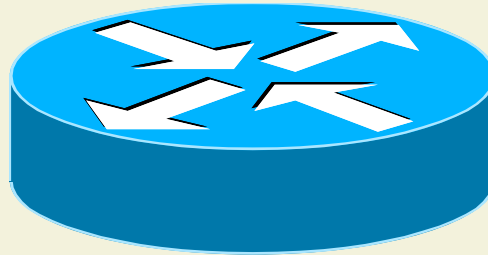
IPv4

- 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

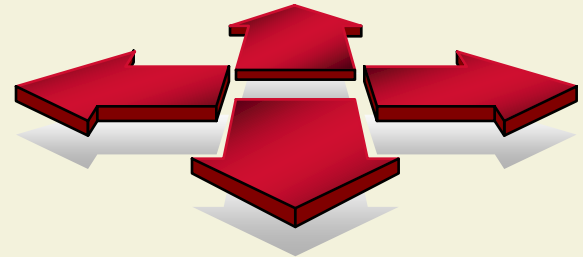
forward packet, forward packet, forward packet,
forward packet...

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

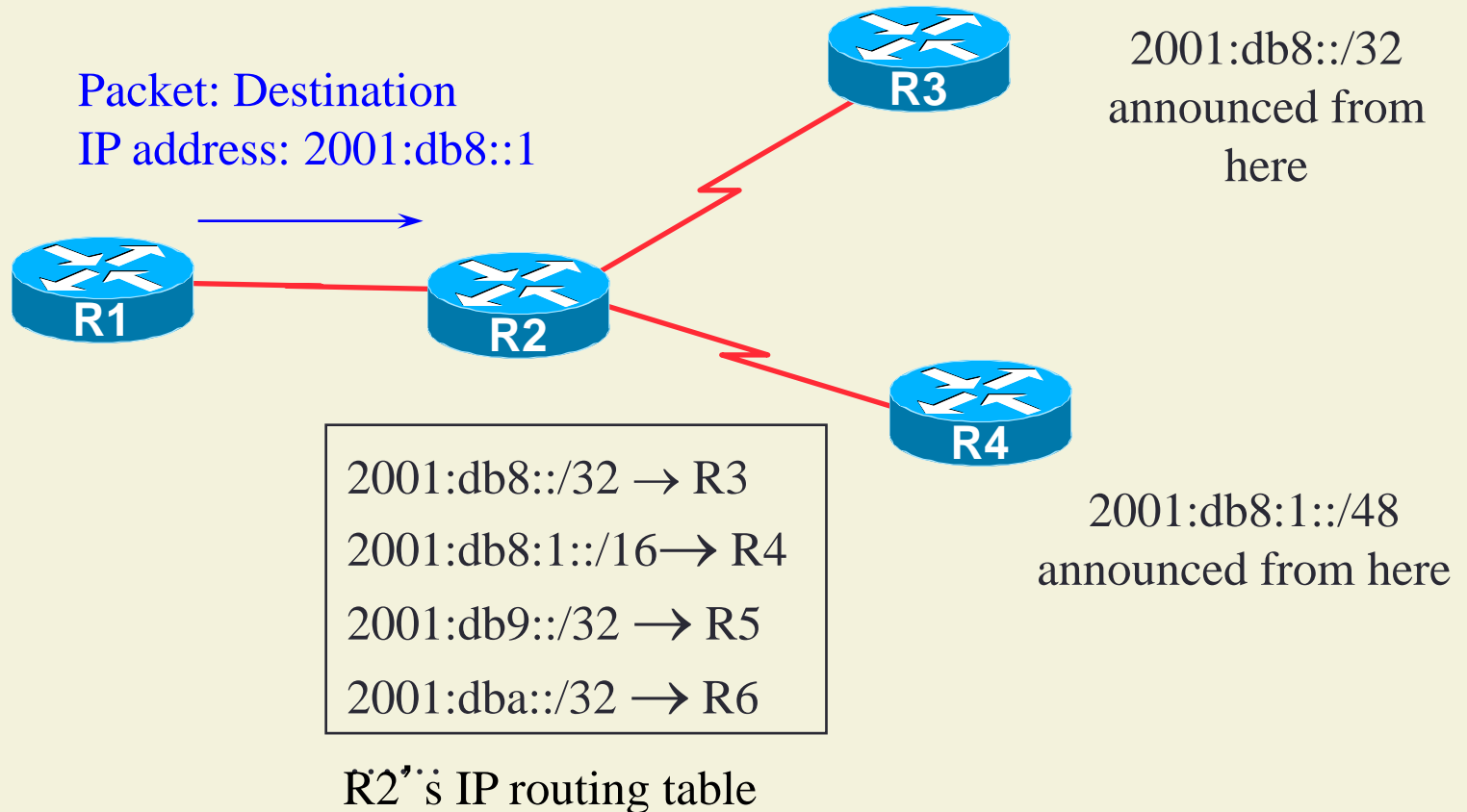
- Path derived from information received from a routing protocol
- Several alternative paths may exist
 - best path stored in **forwarding** table
- Decisions are updated periodically or as topology changes (event driven)
- Decisions are based on:
 - topology, policies and metrics (hop count, filtering,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

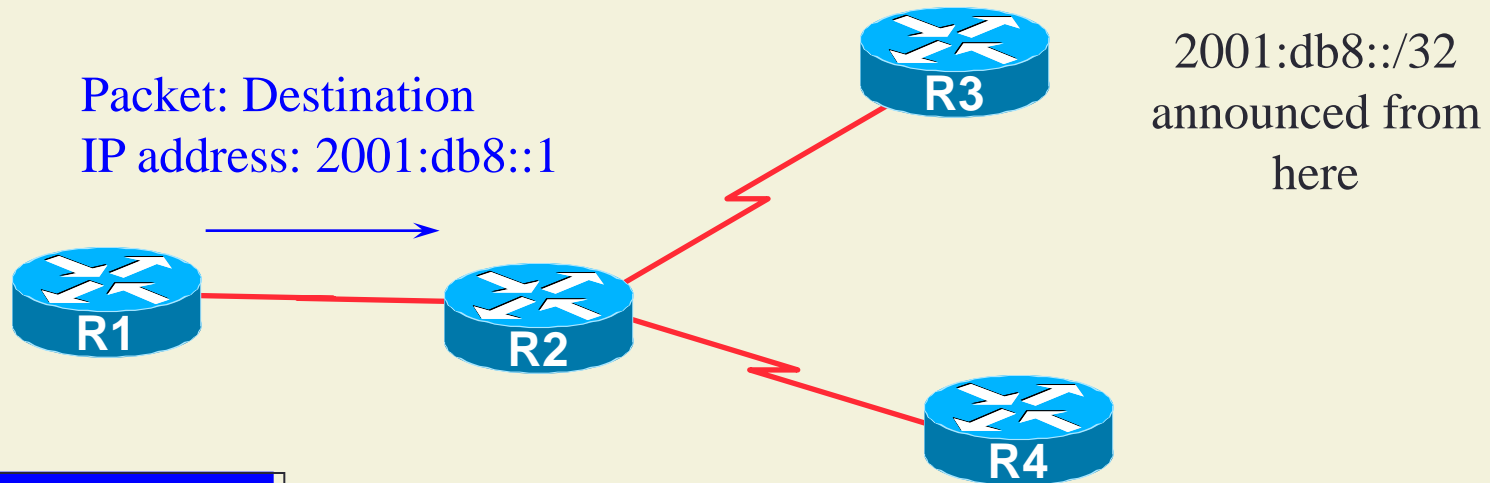
IP route lookup

- 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
.....

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

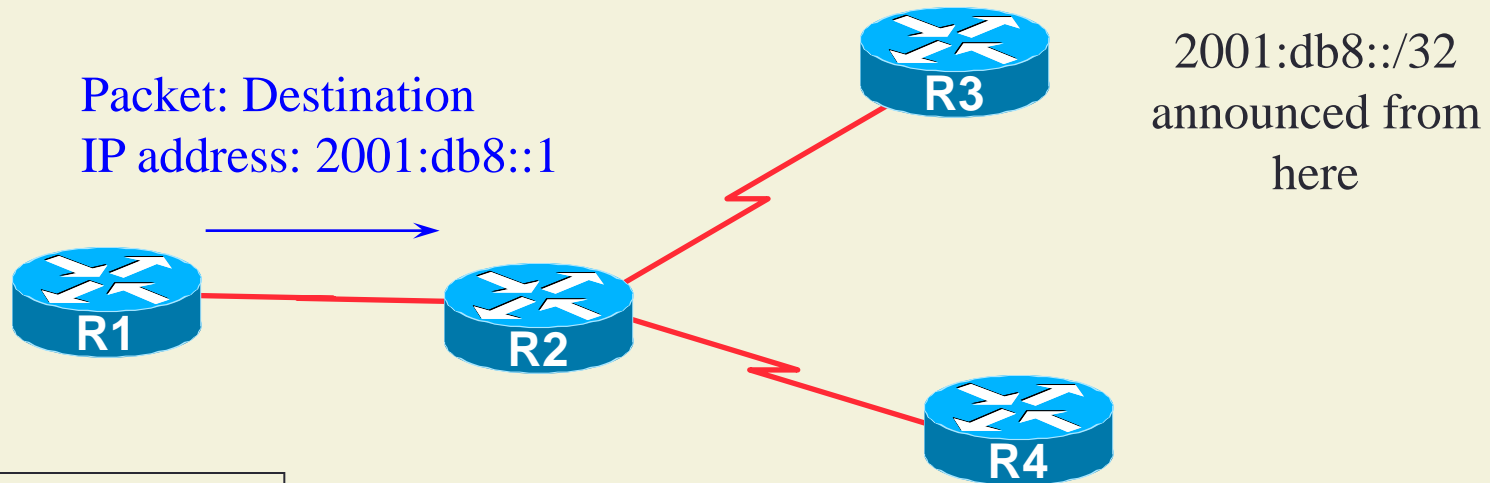
Match!

2001:db8:1::/48
announced from here

R2' s IP routing table

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
.....	

2001:db8::1 && ffff:ffff:ffff:: 2001:db8:1::/16
announced from here

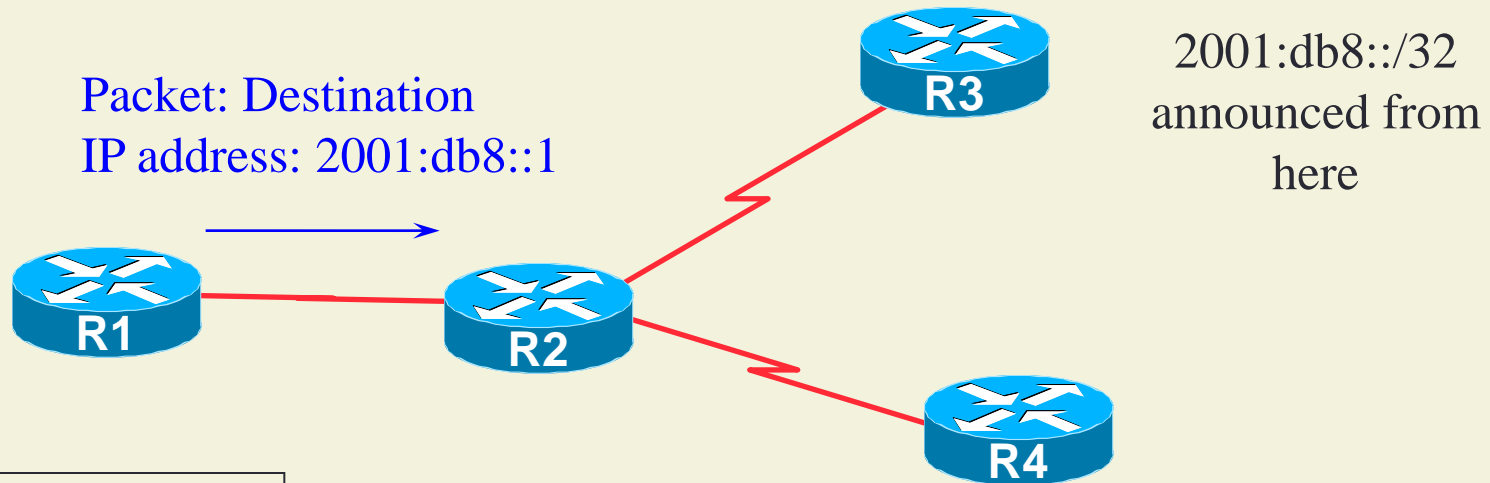
vs.

2001:db8:1:: && ffff:ffff:ffff::
Match as well!

R2' s IP routing table

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::

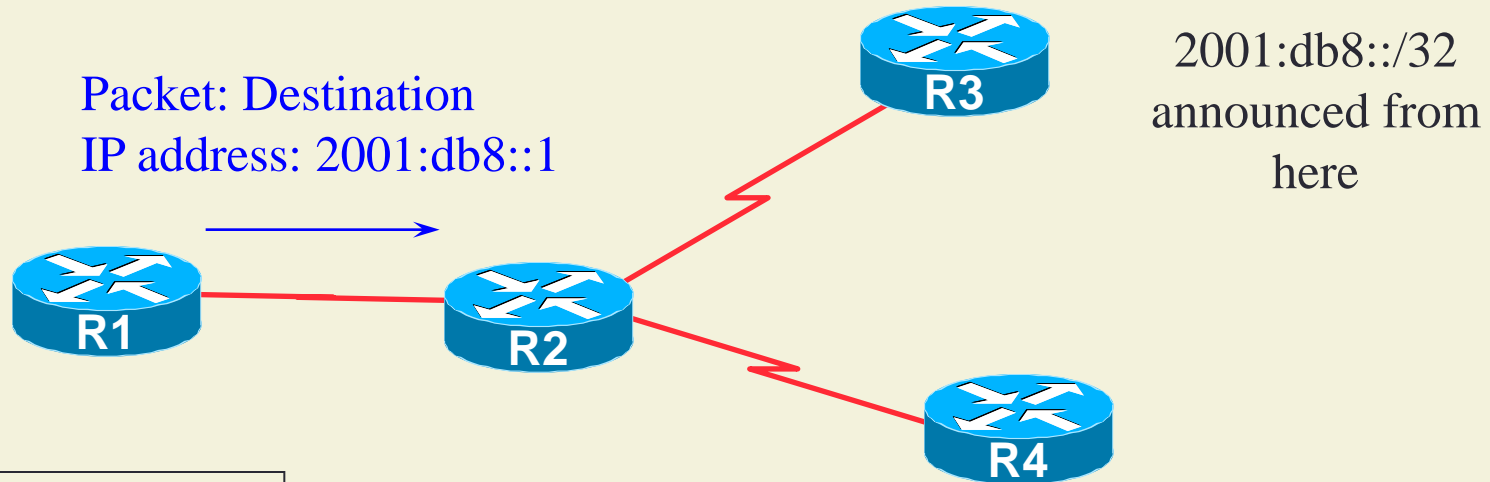
vs.

2001:db9:: && ffff:ffff::

Does not match!

IP route lookup: Longest match routing

- Based on destination IP address



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

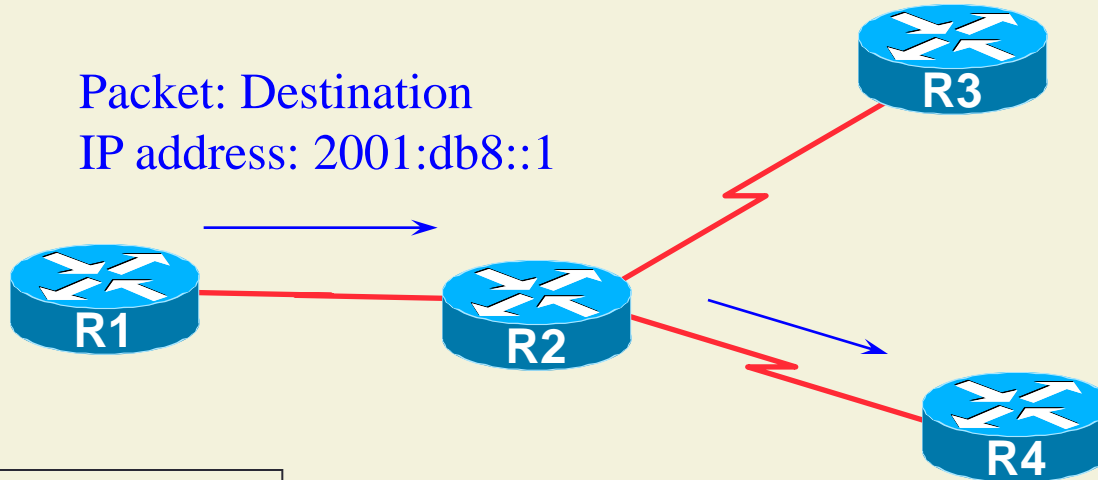
R2' s IP routing table

2001:db8::1 && ffff:ffff::
vs. 2001:dba:: && ffff:ffff::
Does not match!

IP route lookup: Longest match routing

- Based on destination IP address

Packet: Destination
IP address: 2001:db8::1



2001:db8::/32 → R3

2001:db8:1::/48 → R4

2001:db9::/32 → R5

2001:dba::/32 → R6

.....

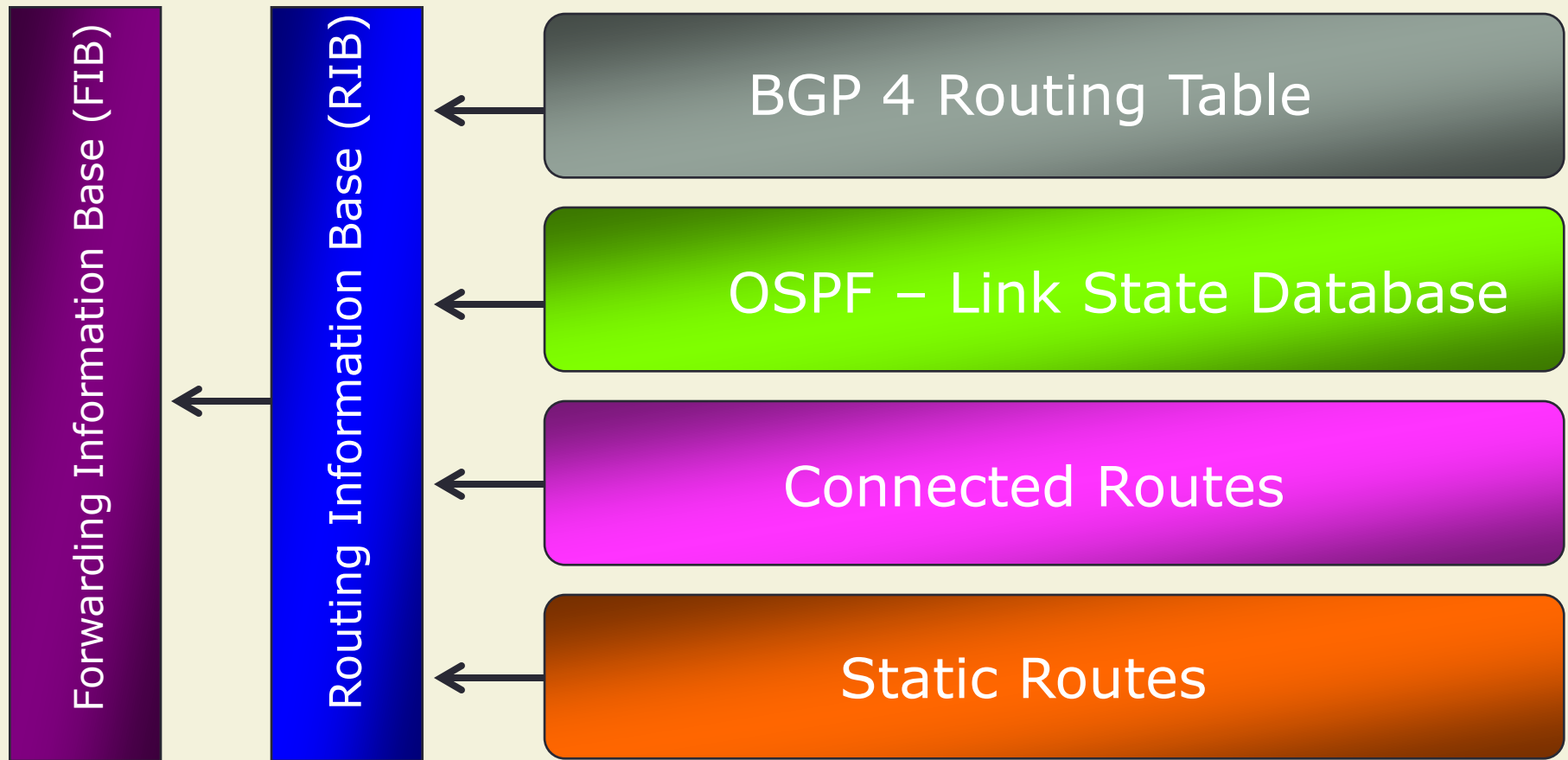
← Longest match, 48 bit netmask

R2' s IP routing table

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

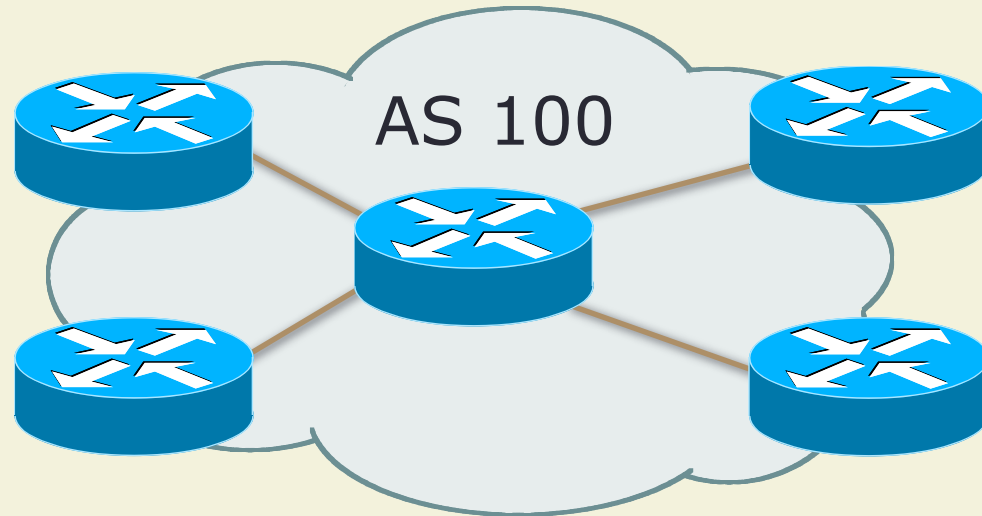
Egress Traffic

- 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)

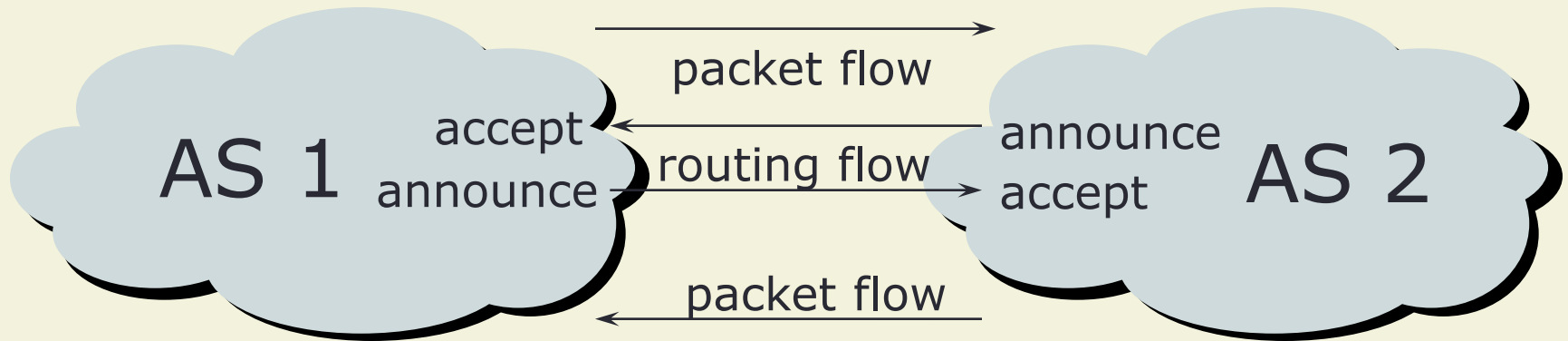


- Collection of networks with same routing policy
- Single routing protocol
- Usually under single ownership, trust and administrative control

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

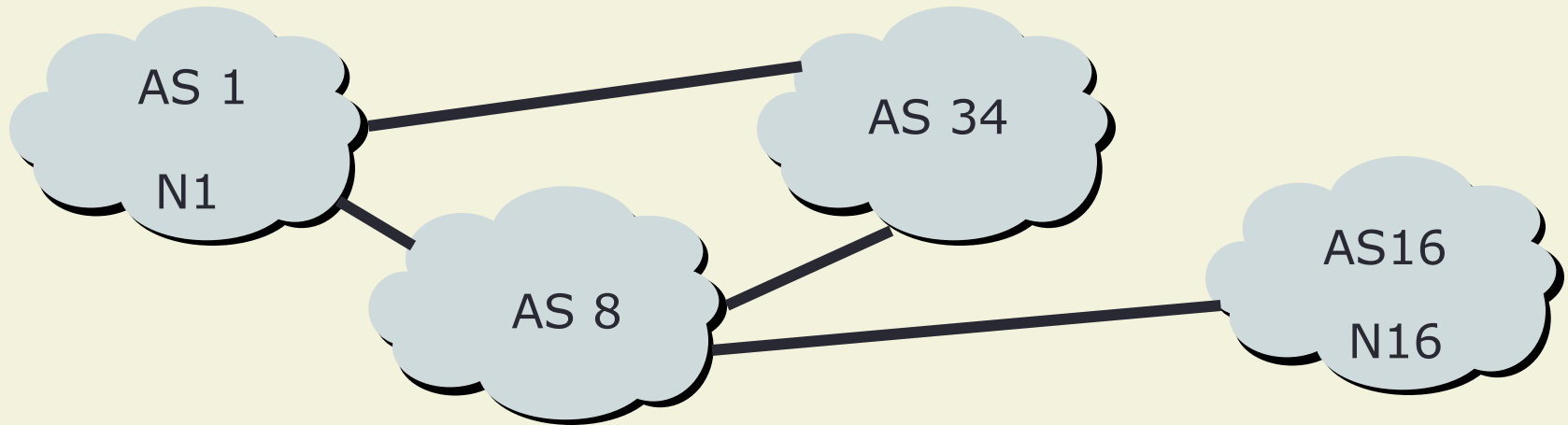
AS2 must announce to AS1

AS1 must accept from AS2

Routing flow and Traffic flow

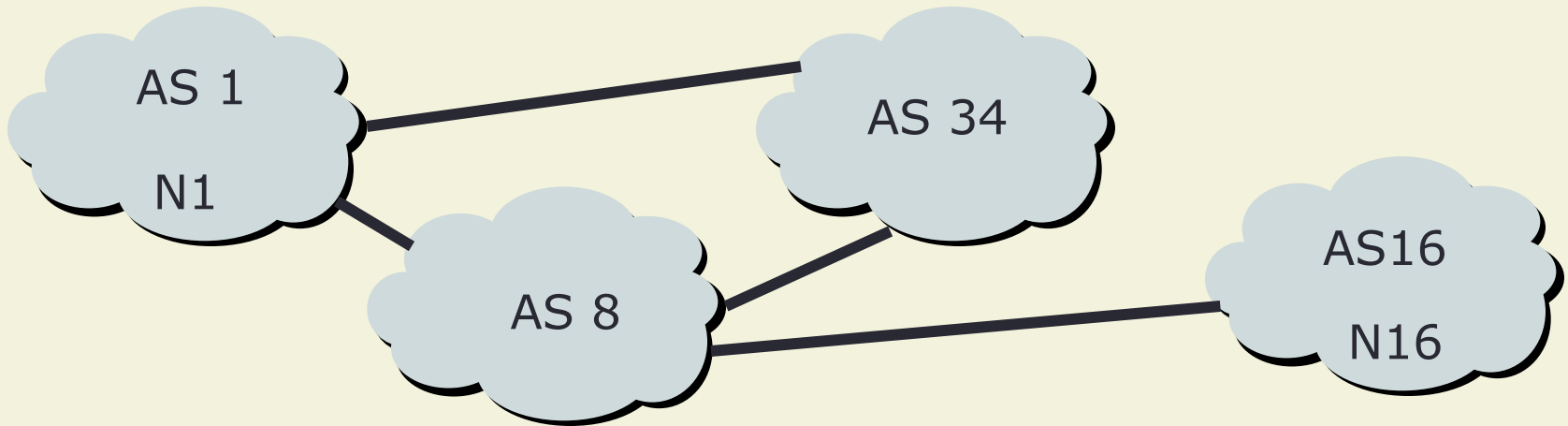
- Traffic flow is always in the opposite direction of the flow of Routing information
 - Filtering outgoing routing information inhibits traffic flow inbound
 - Filtering inbound routing information inhibits traffic flow outbound

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

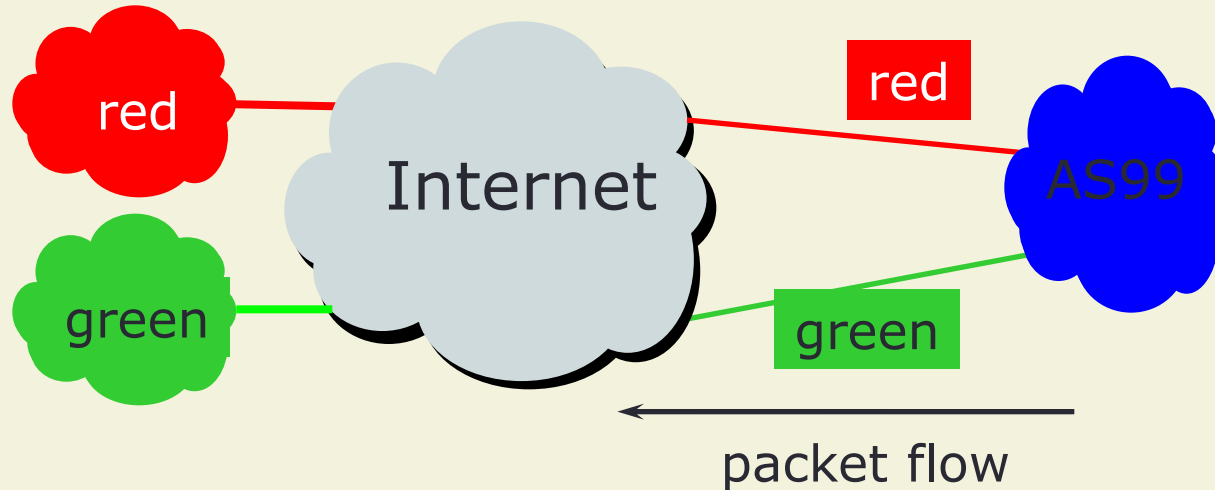


- As multiple paths between sites are implemented it is easy to see how policies can become quite complex.

Routing Policy

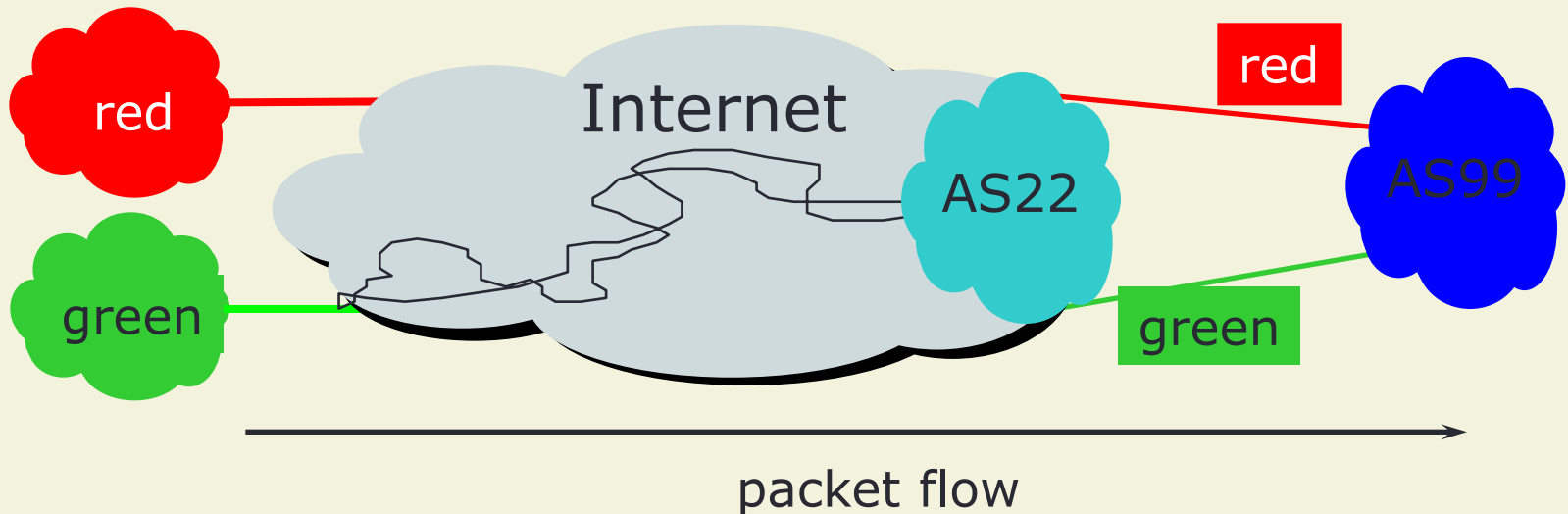
- Used to control traffic flow in and out of an ISP network
- ISP makes decisions on what routing information to accept and discard from its neighbours
 - Individual routes
 - Routes originated by specific ASes
 - Routes traversing specific ASes
 - Routes belonging to other groupings
 - Groupings which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

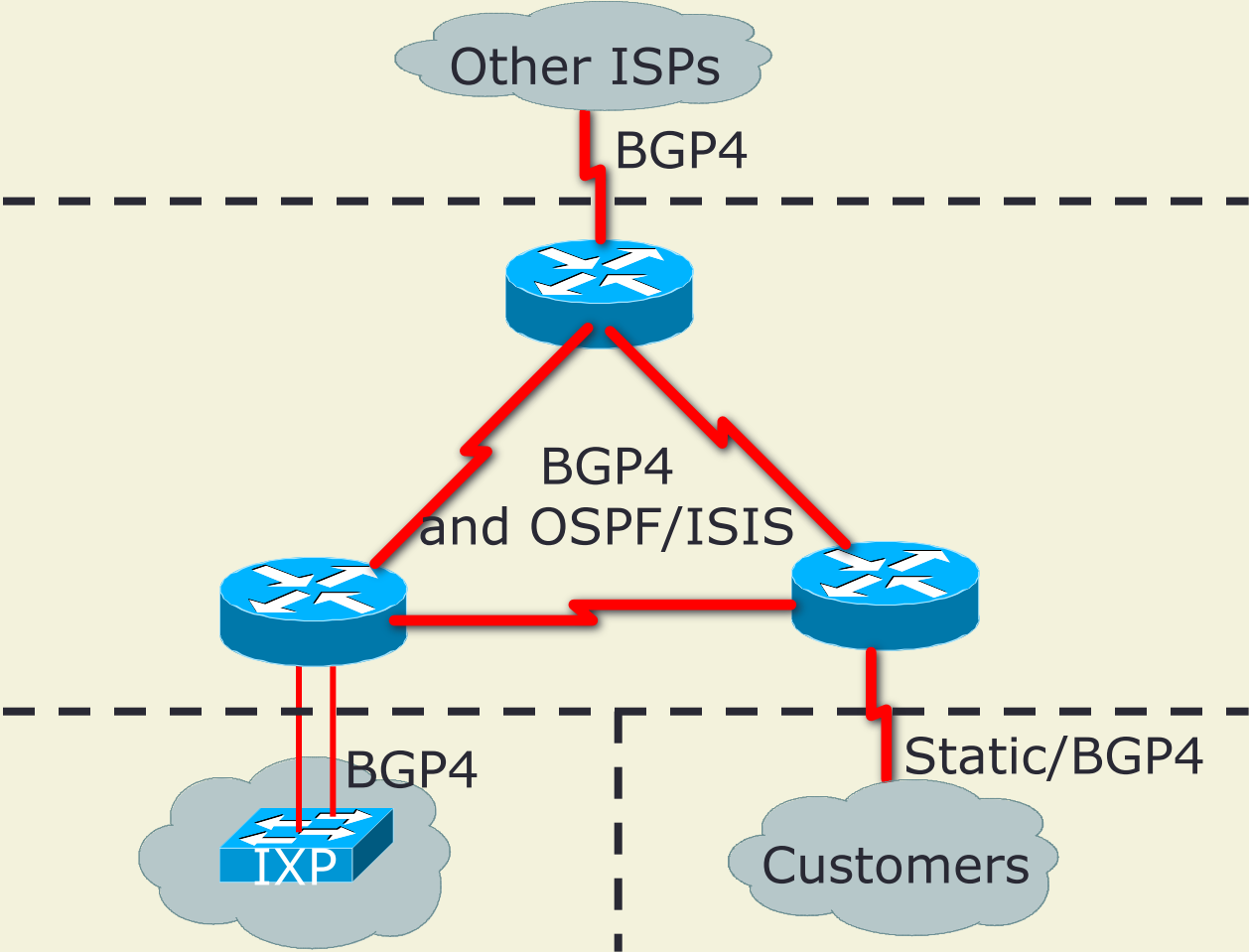
Interior versus Exterior Routing Protocols

- 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

The IPv6 Protocol & IPv6 Standards

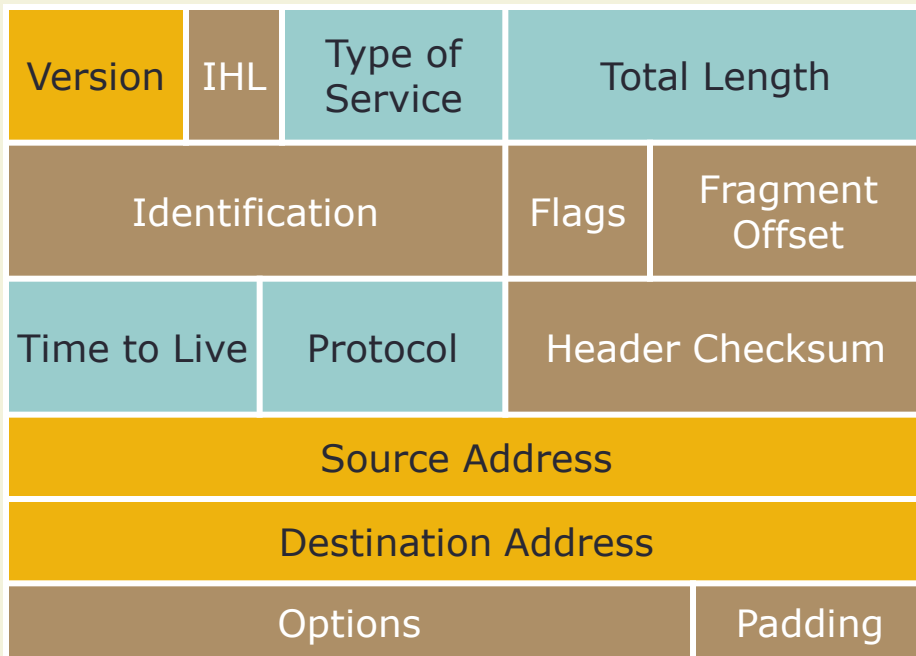
ISP Workshops

So what has really changed?

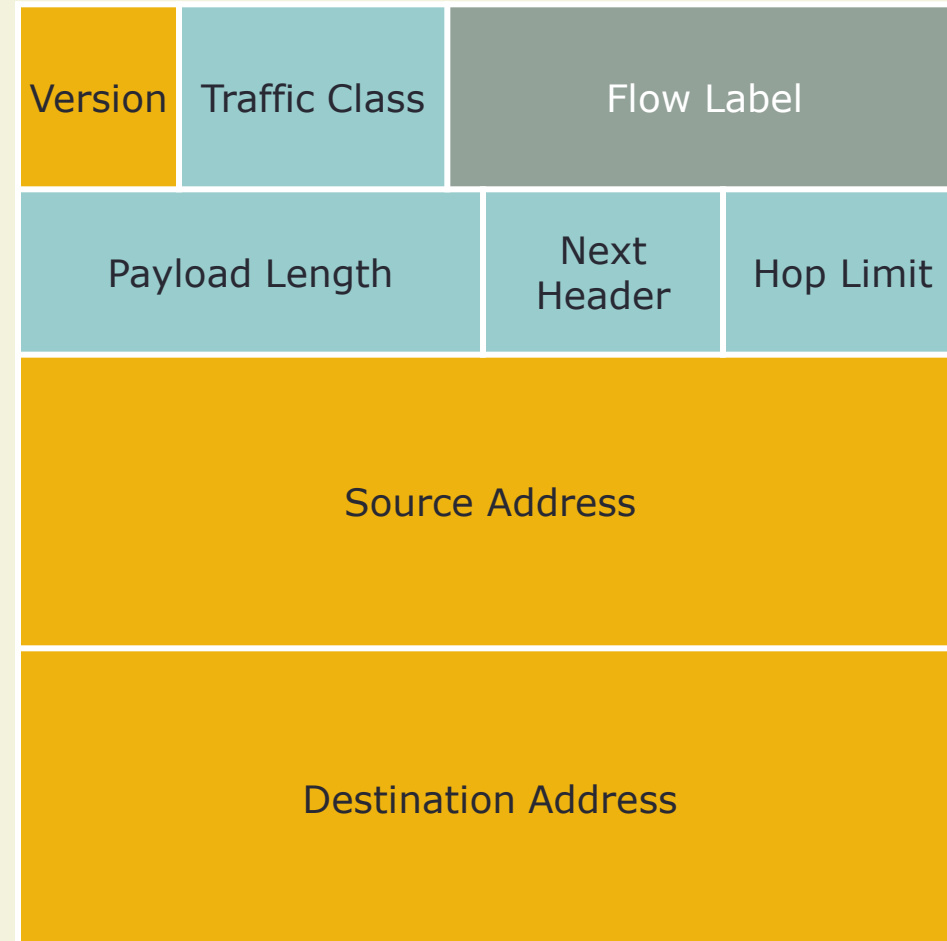
- Expanded address space
 - Address length quadrupled to 16 bytes
- Header Format Simplification
 - Fixed length, optional headers are daisy-chained
 - IPv6 header is twice as long (40 bytes) as IPv4 header without options (20 bytes)
- No checksum at the IP network layer
- No hop-by-hop fragmentation
 - Path MTU discovery
- 64 bits aligned
- Authentication and Privacy Capabilities
 - IPsec is mandated
- No more broadcast

IPv4 and IPv6 Header Comparison

IPv4 Header



IPv6 Header



- Legend**
- Field's name kept from IPv4 to IPv6
 - Fields not kept in IPv6
 - Name and position changed in IPv6
 - New field in IPv6

IPv6 Header

- Version = 4-bit value set to 6
- Traffic Class = 8-bit value
 - Replaces IPv4 TOS field
- Flow Label = 20-bit value
- Payload Length = 16-bit value
 - The size of the rest of the IPv6 packet following the header – replaces IPv4 Total Length
- Next Header = 8-bit value
 - Replaces IPv4 Protocol, and indicates type of next header
- Hop Limit = 8-bit value
 - Decreased by one every IPv6 hop (IPv4 TTL counter)
- Source address = 128-bit value
- Destination address = 128-bit value

Header Format Simplification

- Fixed length
 - Optional headers are daisy-chained
- 64 bits aligned
- IPv6 header is twice as long (40 bytes) as IPv4 header without options (20 bytes)
- IPv4 contains 10 basic header fields
- IPv6 contains 6 basic header fields
 - No checksum at the IP network layer
 - No hop-by-hop fragmentation

Header Format – Extension Headers



- All optional fields go into extension headers
- These are daisy chained behind the main header
 - The last 'extension' header is usually the ICMP, TCP or UDP header
- Makes it simple to add new features in IPv6 protocol without major re-engineering of devices
- Number of extension headers is not fixed / limited

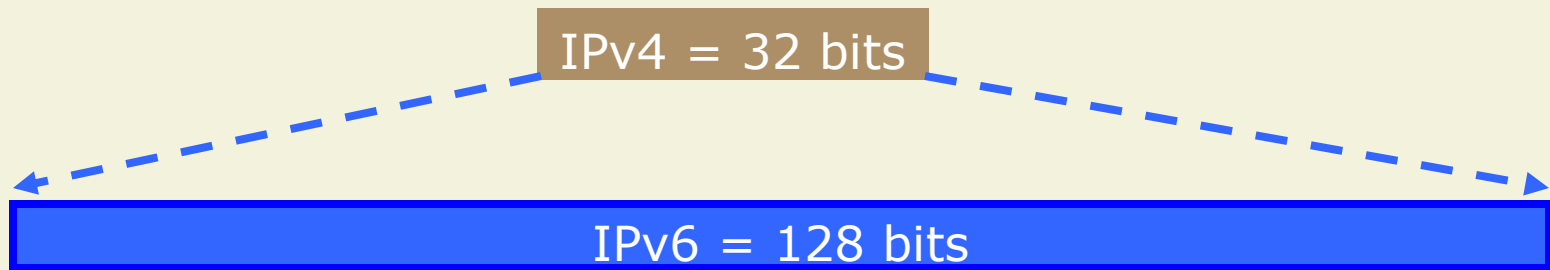
Header Format – Common Headers

- Common values of Next Header field:
 - 0 Hop-by-hop option (extension)
 - 2 ICMP (payload)
 - 6 TCP (payload)
 - 17 UDP (payload)
 - 43 Source routing (extension)
 - 44 Fragmentation (extension)
 - 50 Encrypted security payload (extension, IPSec)
 - 51 Authentication (extension, IPSec)
 - 59 Null (No next header)
 - 60 Destination option (extension)

Header Format – Ordering of Headers

- Order is important because:
 - Hop-by-hop header has to be processed by every intermediate node
 - Routing header needs to be processed by intermediate routers
 - At the destination fragmentation has to be processed before other headers
- This makes header processing easier to implement in hardware

Larger Address Space




- IPv4
 - 32 bits
 - = 4,294,967,296 possible addressable devices
- IPv6
 - 128 bits: 4 times the size in bits
 - = 3.4×10^{38} possible addressable devices
 - = 340,282,366,920,938,463,463,374,607,431,768,211,456
 - $\sim 5 \times 10^{28}$ addresses per person on the planet

How was the IPv6 Address Size Chosen?

- Some wanted fixed-length, 64-bit addresses
 - Easily good for 10^{12} sites, 10^{15} nodes, at .0001 allocation efficiency
 - (3 orders of magnitude more than IPv6 requirement)
 - Minimizes growth of per-packet header overhead
 - Efficient for software processing
- Some wanted variable-length, up to 160 bits
 - Compatible with OSI NSAP addressing plans
 - Big enough for auto-configuration using IEEE 802 addresses
 - Could start with addresses shorter than 64 bits & grow later
- Settled on fixed-length, 128-bit addresses

IPv6 Address Representation (1)

- 16 bit fields in case insensitive colon hexadecimal representation
 - 2031:0000:130F:0000:0000:09C0:876A:130B
- Leading zeros in a field are optional:
 - 2031:0:130F:0:0:9C0:876A:130B
- Successive fields of 0 represented as ::, but only once in an address:
 - 2031:0:130F::9C0:876A:130B is ok
 - 2031::130F::9C0:876A:130B is **NOT** ok
- 0:0:0:0:0:0:0:1 → ::1 (loopback address)
- 0:0:0:0:0:0:0:0 → :: (unspecified address)

IPv6 Address Representation (2)

- :: representation
 - RFC5952 recommends that the rightmost set of :0: be replaced with :: for consistency
 - 2001:db8:0:2f::5 rather than 2001:db8::2f:0:0:0:5
- IPv4-compatible (not used any more)
 - 0:0:0:0:0:0:192.168.30.1
 - = ::192.168.30.1
 - = ::C0A8:1E01
- In a URL, it is enclosed in brackets (RFC3986)
 - [http://\[2001:db8:4f3a::206:ae14\]:8080/index.html](http://[2001:db8:4f3a::206:ae14]:8080/index.html)
 - Cumbersome for users, mostly for diagnostic purposes
 - Use fully qualified domain names (FQDN)
 - ⇒ The DNS has to work!!

IPv6 Address Representation (3)

- Prefix Representation
 - Representation of prefix is just like IPv4 CIDR
 - In this representation you attach the prefix length
 - Like IPv4 address:
 - 198.10.0.0/16
 - IPv6 address is represented in the same way:
 - 2001:db8:12::/40

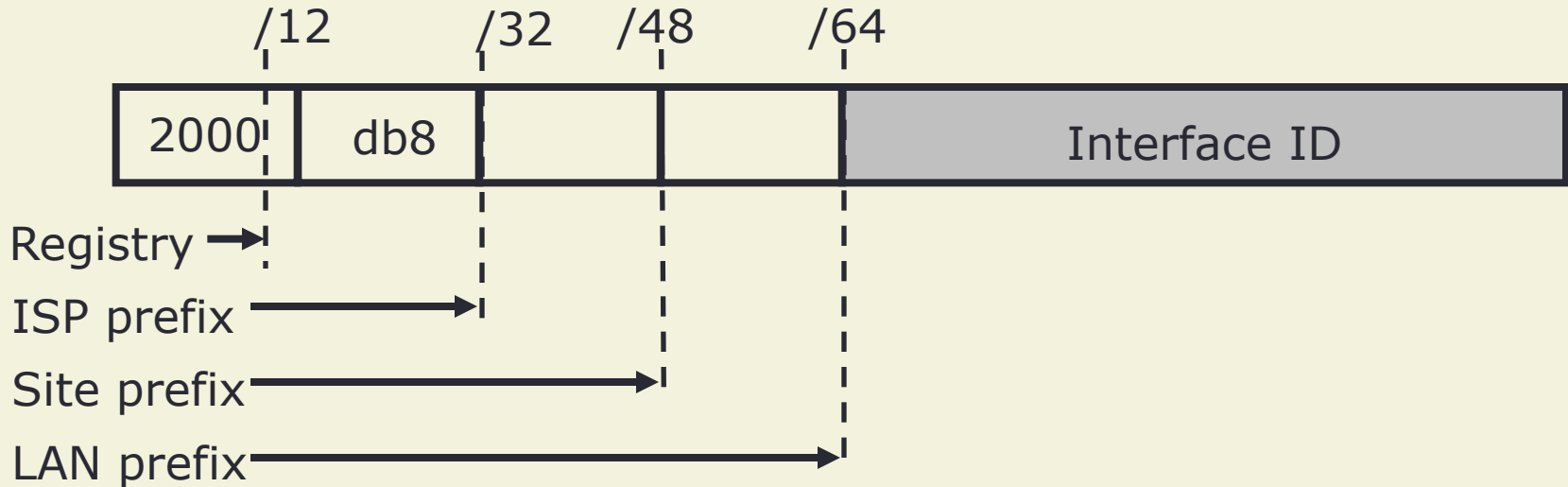
IPv6 Addressing

- IPv6 Addressing rules are covered by multiple RFCs
 - Architecture defined by RFC 4291
- Address Types are :
 - Unicast : One to One (Global, Unique Local, Link local)
 - Anycast : One to Nearest (Allocated from Unicast)
 - Multicast : One to Many
- A single interface may be assigned multiple IPv6 addresses of any type (unicast, anycast, multicast)
 - No Broadcast Address → Use Multicast

IPv6 Addressing

Type	Binary	Hex
Unspecified	000...0	::/128
Loopback	000...1	:::1/128
Global Unicast Address	0010	2000::/3
Link Local Unicast Address	1111 1110 10	FE80::/10
Unique Local Unicast Address	1111 1100 1111 1101	FC00::/7
Multicast Address	1111 1111	FF00::/8

IPv6 Address Allocation



- The allocation process is:
 - The IANA is allocating out of 2000::/3 for initial IPv6 unicast use
 - Each registry gets a /12 prefix from the IANA
 - Registry allocates a /32 prefix (or larger) to an IPv6 ISP
 - Policy is that an ISP allocates a /48 prefix to each end customer

IPv6 Addressing Scope

- 64 bits reserved for the interface ID
 - Possibility of 2^{64} hosts on one network LAN
 - In theory 18,446,744,073,709,551,616 hosts
 - Arrangement to accommodate MAC addresses within the IPv6 address
- 16 bits reserved for the end site
 - Possibility of 2^{16} networks at each end-site
 - 65536 subnets equivalent to a /12 in IPv4 (assuming a /28 or 16 hosts per IPv4 subnet)

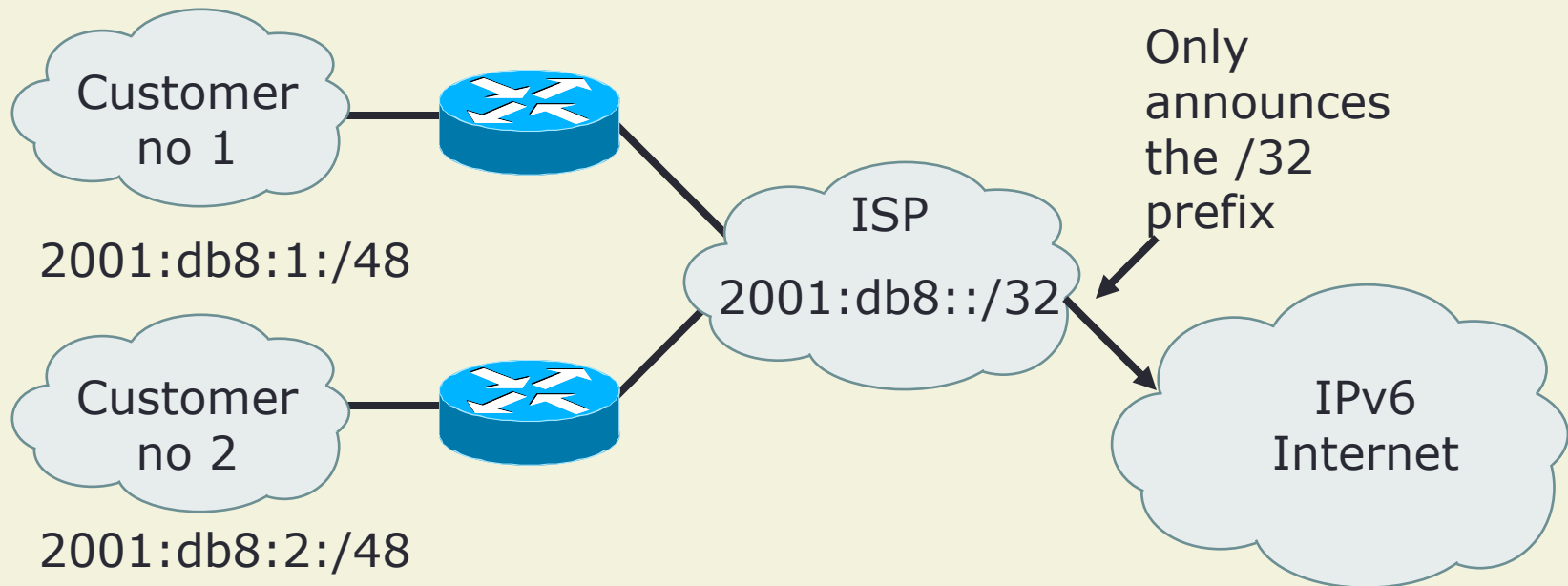
IPv6 Addressing Scope

- 16 bits reserved for each service provider
 - Possibility of 2^{16} end-sites per service provider
 - 65536 possible customers: equivalent to each service provider receiving a /8 in IPv4 (assuming a /24 address block per customer)
- 29 bits reserved for all service providers
 - Possibility of 2^{29} service providers
 - i.e. 536,870,912 discrete service provider networks
 - Although some service providers already are justifying more than a /32

How to get an IPv6 Address?

- IPv6 address space is allocated by the 5 RIRs:
 - AfriNIC, APNIC, ARIN, LACNIC, RIPE NCC
 - ISPs get address space from the RIRs
 - Enterprises get their IPv6 address space from their ISP
- 6to4 tunnels 2002::/16
 - Last resort only and now mostly useless
- (6Bone)
 - Was the IPv6 experimental network since the mid 90s
 - Now retired, end of service was 6th June 2006 (RFC3701)

Aggregation hopes



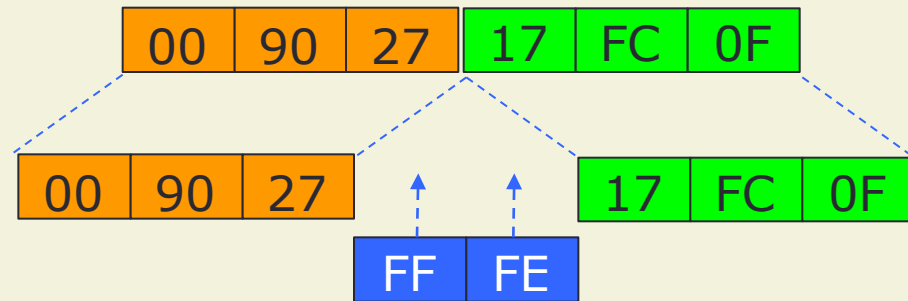
- Larger address space enables aggregation of prefixes announced in the global routing table
- Idea was to allow efficient and scalable routing
- **But current Internet multihoming solution breaks this model**

Interface IDs

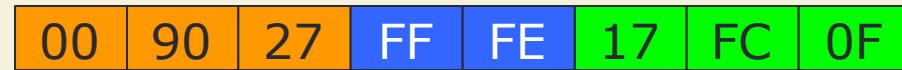
- Lowest order 64-bit field of unicast address may be assigned in several different ways:
 - Auto-configured from a 64-bit EUI-64, or expanded from a 48-bit MAC address (e.g., Ethernet address)
 - Auto-generated pseudo-random number (to address privacy concerns)
 - Assigned via DHCP
 - Manually configured

EUI-64

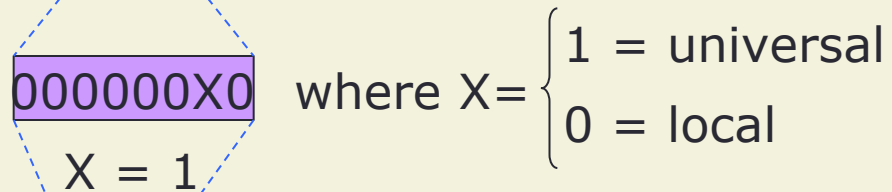
Ethernet MAC address
(48 bits)



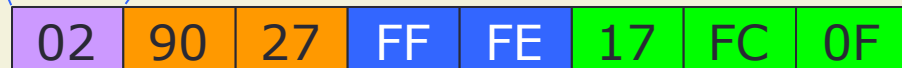
64 bits version



Scope of the EUI-64 id



EUI-64 address



- EUI-64 address is formed by inserting FFFE between the **company-id** and the **manufacturer extension**, and setting the “u” bit to indicate scope
 - Global scope: for IEEE 48-bit MAC
 - Local scope: when no IEEE 48-bit MAC is available (eg serials, tunnels)

IPv6 Addressing Examples

LAN: 2001:db8:213:1::/64

Ethernet0

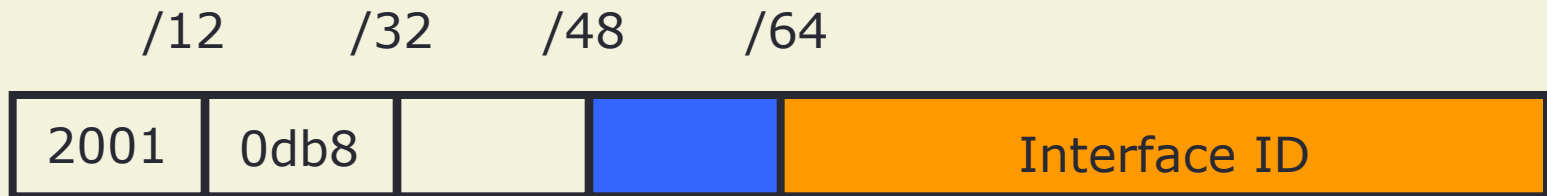


```
interface Ethernet0
  ipv6 address 2001:db8:213:1::/64 eui-64
```

MAC address: 0060.3e47.1530

```
router# show ipv6 interface Ethernet0
Ethernet0 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::260:3EFF:FE47:1530
Global unicast address(es):
  2001:db8:213:1:260:3EFF:FE47:1530, subnet is 2001:db8:213:1::/64
Joined group address(es):
  FF02::1:FF47:1530
  FF02::1
  FF02::2
MTU is 1500 bytes
```

IPv6 Address Privacy (RFC 4941)



- Temporary addresses for IPv6 host client application, e.g. Web browser
- Intended to inhibit device/user tracking but is also a potential issue
 - More difficult to scan all IP addresses on a subnet
 - But port scan is identical when an address is known
- Random 64 bit interface ID, run DAD before using it
- Rate of change based on local policy
- Implemented on Microsoft Windows XP/Vista/7 and Apple MacOS 10.7 onwards
 - Can be activated on FreeBSD/Linux with a system call

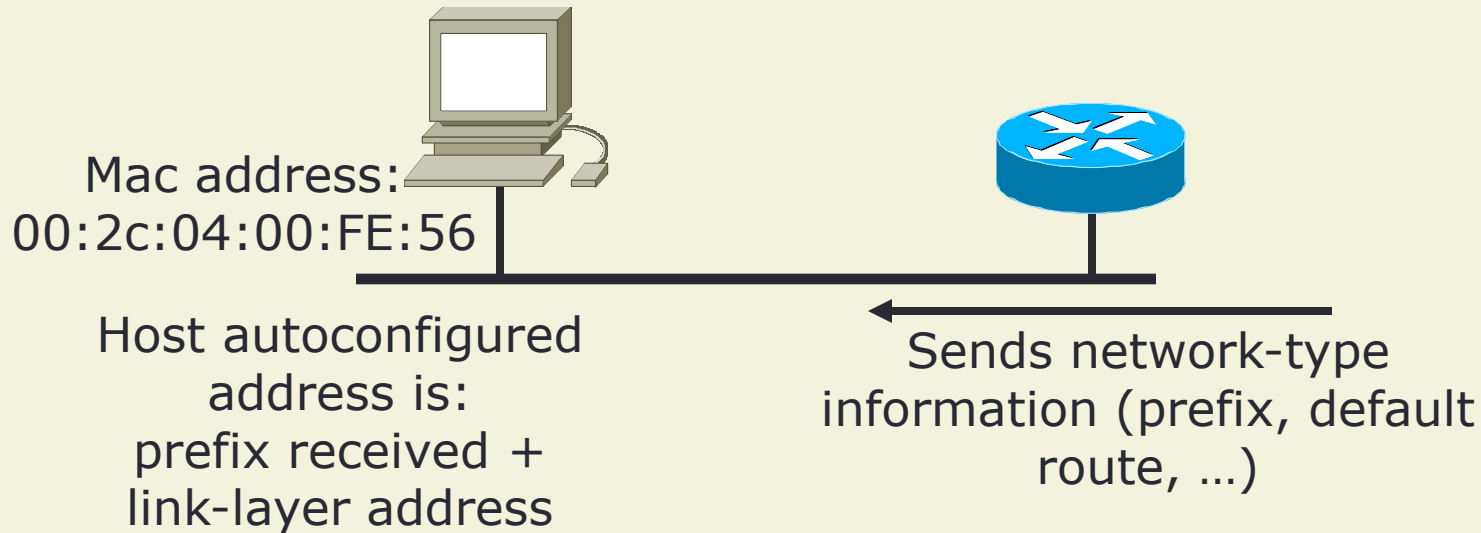
Host IPv6 Addressing Options

- Stateless (RFC4862)
 - SLAAC – Stateless Address AutoConfiguration
 - Booting node sends a “router solicitation” to request “router advertisement” to get information to configure its interface
 - Booting node configures its own Link-Local address
- Stateful
 - DHCPv6 – required by most enterprises
 - Manual – like IPv4 pre-DHCP
 - Useful for servers and router infrastructure
 - Doesn't scale for typical end user devices

IPv6 Renumbering

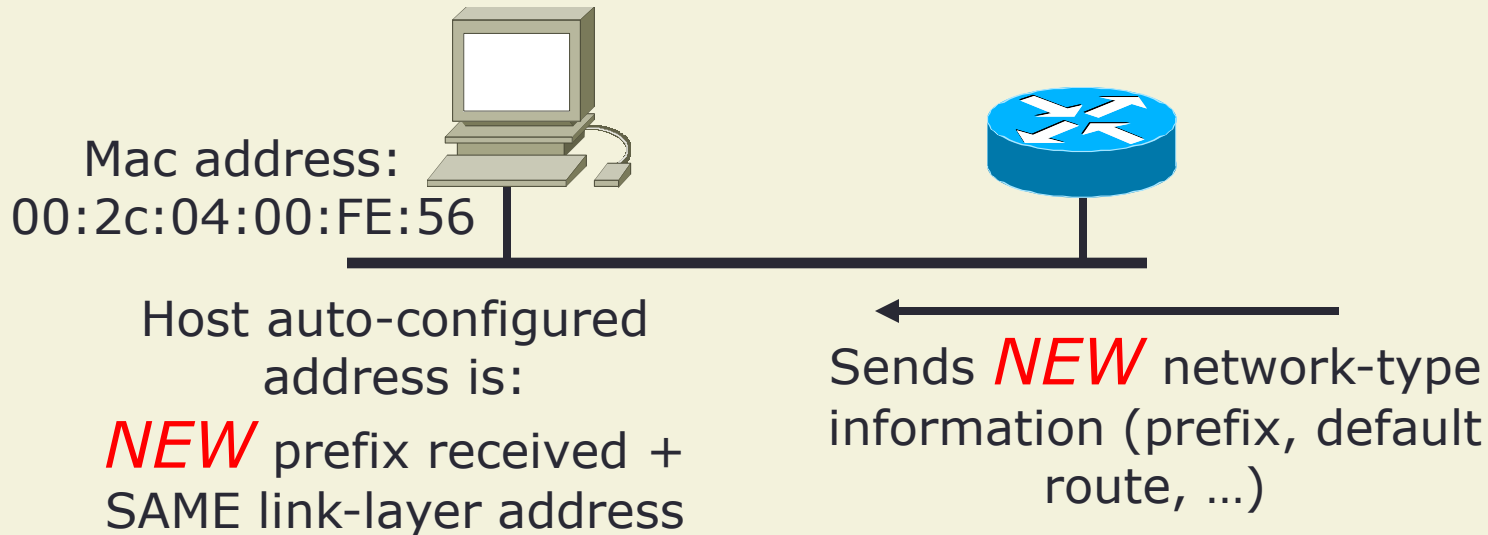
- Renumbering Hosts
 - Stateless:
 - Hosts renumbering is done by modifying the RA to announce the old prefix with a short lifetime and the new prefix
 - Stateful:
 - DHCPv6 uses same process as DHCPv4
- Renumbering Routers
 - Router renumbering protocol was developed (RFC 2894) to allow domain-interior routers to learn of prefix introduction / withdrawal
 - **No known implementation!**

Auto-configuration



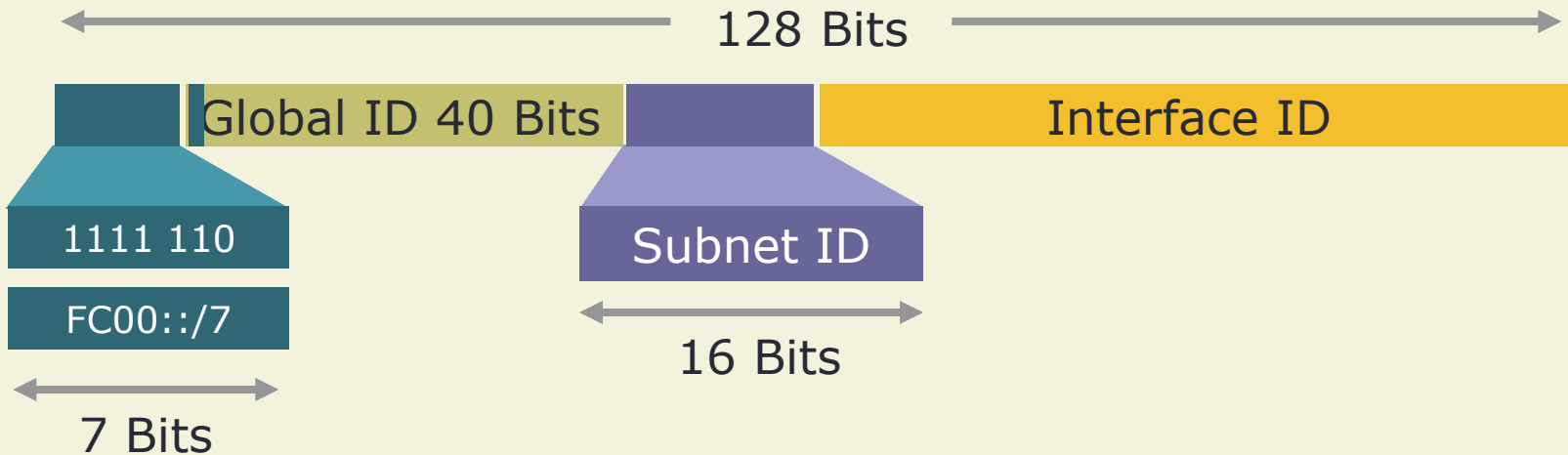
- PC sends router solicitation (RS) message
- Router responds with router advertisement (RA)
 - This includes prefix and default route
 - RFC6106 adds DNS server option
- PC configures its IPv6 address by concatenating prefix received with its EUI-64 address

Renumbering



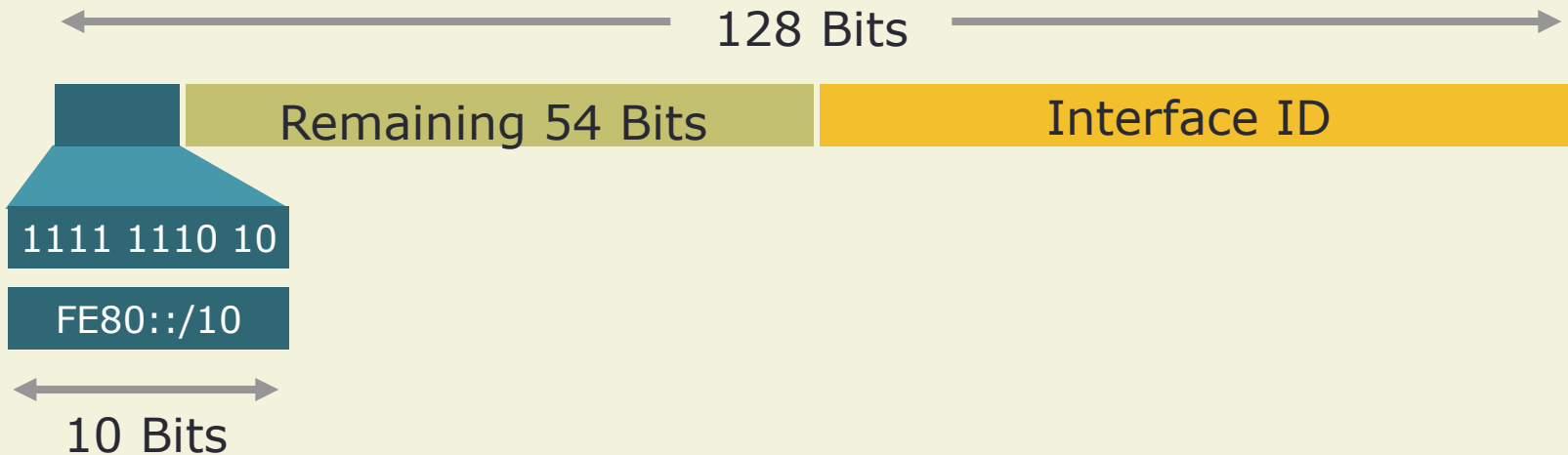
- Router sends router advertisement (RA)
 - This includes the new prefix and default route (and remaining lifetime of the old address)
- PC configures a new IPv6 address by concatenating prefix received with its EUI-64 address
 - Attaches lifetime to old address

Unique-Local



- Unique-Local Addresses Used For:
 - Local communications & inter-site VPNs
 - Local devices such as printers, telephones, etc
 - Site Network Management systems connectivity
- Not routable on the Internet
- Reinvention of the deprecated site-local?

Link-Local



- Link-Local Addresses Used For:
 - Communication between two IPv6 device (like ARP but at Layer 3)
 - Next-Hop calculation in Routing Protocols
- Automatically assigned by Router as soon as IPv6 is enabled
 - Mandatory Address
- Only Link Specific scope
- Remaining 54 bits could be Zero or any manual configured value

Multicast use

- Broadcasts in IPv4
 - Interrupts all devices on the LAN even if the intent of the request was for a subset
 - Can completely swamp the network (“broadcast storm”)
- Broadcasts in IPv6
 - Are not used and replaced by multicast
- Multicast
 - Enables the efficient use of the network
 - Multicast address range is much larger

IPv6 Multicast Address

- IP multicast address has a prefix FF00::/8
- The second octet defines the lifetime and scope of the multicast address.

8-bit	4-bit	4-bit	112-bit
1111 1111	Lifetime	Scope	Group-ID

Lifetime	
0	If Permanent
1	If Temporary

Scope	
1	Node
2	Link
5	Site
8	Organisation
E	Global

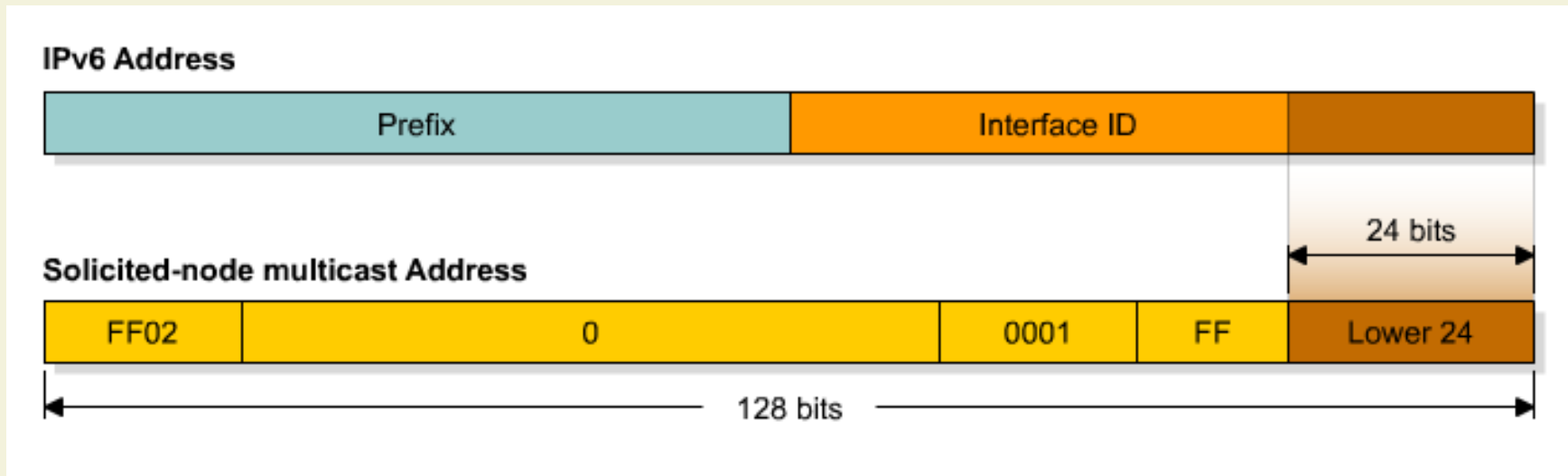
IPv6 Multicast Address Examples

- RIPng
 - The multicast address AllRIPRouters is **FF02::9**
 - Note that 02 means that this is a permanent address and has link scope
- OSPFv3
 - The multicast address AllSPFRouters is **FF02::5**
 - The multicast address AllDRouters is **FF02::6**
- EIGRP
 - The multicast address AllEIGRPRouters is **FF02::A**

Solicited-Node Multicast

- Solicited-Node Multicast is used for Duplicate Address Detection
 - Part of the Neighbour Discovery process
 - Replaces ARP
 - Duplicate IPv6 Addresses are rare, but still have to be tested for
- For each unicast and anycast address configured there is a corresponding solicited-node multicast address
 - This address is only significant for the local link

Solicited-Node Multicast Address



- Solicited-node multicast address consists of FF02:0:0:0:0:1:FF::/104 prefix joined with the lower 24 bits from the unicast or anycast IPv6 address

Solicited-Node Multicast

```
R1#sh ipv6 int e0
Ethernet0 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::200:CFF:FE3A:8B18
  No global unicast address is configured
  Joined group address(es):
    FF02::1
    FF02::2
    FF02::1:FF3A:8B18
  MTU is 1500 bytes
  ICMP error messages limited to one every 100 milliseconds
  ICMP redirects are enabled
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND advertised reachable time is 0 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 200 seconds
  ND router advertisements live for 1800 seconds
  Hosts use stateless autoconfig for addresses.
R1#
```

Solicited-Node Multicast Address

IPv6 Anycast

- An IPv6 anycast address is an identifier for a set of interfaces (typically belonging to different nodes)
 - A packet sent to an anycast address is delivered to one of the interfaces identified by that address (the “nearest” one, according to the routing protocol’s measure of distance).
 - [RFC4291 describes IPv6 Anycast in more detail](#)
- In reality there is no known implementation of IPv6 Anycast as per the RFC
 - Most operators have chosen to use IPv4 style anycast instead

Anycast on the Internet

- A global unicast address is assigned to all nodes which need to respond to a service being offered
 - This address is routed as part of its parent address block
- The responding node is the one which is closest to the requesting node according to the routing protocol
 - Each anycast node looks identical to the other
- Applicable within an ASN, or globally across the Internet
- Typical (IPv4) examples today include:
 - Root DNS and ccTLD/gTLD nameservers
 - SMTP relays and DNS resolvers within ISP autonomous systems

MTU Issues

- Minimum link MTU for IPv6 is 1280 octets (versus 68 octets for IPv4)
 - ⇒ on links with MTU < 1280, link-specific fragmentation and reassembly must be used
- Implementations are expected to perform path MTU discovery to send packets bigger than 1280
- Minimal implementation can omit PMTU discovery as long as all packets kept ≤ 1280 octets
- A Hop-by-Hop Option supports transmission of “jumbograms” with up to 2^{32} octets of payload

IPv6 Neighbour Discovery

- Protocol defines mechanisms for the following problems:
 - Router discovery
 - Prefix discovery
 - Parameter discovery
 - Address autoconfiguration
 - Address resolution
 - Next-hop determination
 - Neighbour unreachability detection
 - Duplicate address detection
 - Redirects

IPv6 Neighbour Discovery

- Defined in RFC 4861
- Protocol built on top of ICMPv6 (RFC 4443)
 - Combination of IPv4 protocols (ARP, ICMP, IGMP,...)
- Fully dynamic, interactive between Hosts & Routers
- Defines 5 ICMPv6 packet types:
 - Router Solicitation
 - Router Advertisement
 - Neighbour Solicitation
 - Neighbour Advertisement
 - Redirect

IPv6 and DNS

- Hostname to IP address:

IPv4	www.abc.test.	A	192.168.30.1
------	---------------	---	--------------

IPv6	www.abc.test	AAAA	2001:db8:c18:1::2
------	--------------	------	-------------------

IPv6 and DNS

- IP address to Hostname:

IPv4 1.30.168.192.in-addr.arpa. PTR www.abc.test.

IPv6 2.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.1.0.0.0.8.
1.c.0.8.b.d.0.1.0.0.2.ip6.arpa PTR
www.abc.test.

IPv6 Technology Scope

<i>IP Service</i>	<i>IPv4 Solution</i>	<i>IPv6 Solution</i>
Addressing Range	32-bit, Network Address Translation	128-bit, Multiple Scopes
Autoconfiguration	DHCP	Serverless, Reconfiguration,
Security	IPSec	IPSec Mandated,
Mobility	Mobile IP	works End-to-End with Direct Routing
Quality-of-Service	Differentiated Service,	Differentiated Service,
IP Multicast	IGMP, Integrated Multicast Service	MLD, Integrated Multicast Service
	BGP	BGP, Scope Identifier

What does IPv6 do for:

- Security
 - Nothing IPv4 doesn't do – IPSec runs in both
 - But IPv6 mandates IPSec
- QoS
 - Nothing IPv4 doesn't do –
 - Differentiated and Integrated Services run in both
 - So far, Flow label has no real use

IPv6 Security

- IPsec standards apply to both IPv4 and IPv6
- All implementations required to support authentication and encryption headers (“IPsec”)
- Authentication separate from encryption for use in situations where encryption is prohibited or prohibitively expensive
- Key distribution protocols are not yet defined (independent of IP v4/v6)
- Support for manual key configuration required

IP Quality of Service Reminder

- Two basic approaches developed by IETF:
 - “Integrated Service” (int-serv)
 - Fine-grain (per-flow), quantitative promises (e.g., x bits per second), uses RSVP signalling
 - “Differentiated Service” (diff-serv)
 - Coarse-grain (per-class), qualitative promises (e.g., higher priority), no explicit signalling
 - Signalled diff-serv (RFC 2998)
 - Uses RSVP for signalling with course-grained qualitative aggregate markings
 - Allows for policy control without requiring per-router state overhead

IPv6 Support for Int-Serv

- 20-bit Flow Label field to identify specific flows needing special QoS
 - Each source chooses its own Flow Label values; routers use Source Addr + Flow Label to identify distinct flows
 - Flow Label value of 0 used when no special QoS requested (the common case today)
- Originally standardised as RFC 3697

IPv6 Flow Label

- Flow label has not been used since IPv6 standardised
 - Suggestions for use in recent years were incompatible with original specification (discussed in RFC6436)
- Specification updated in RFC6437
 - RFC6438 describes the use of the Flow Label for equal cost multi-path and link aggregation in Tunnels

IPv6 Support for Diff-Serv

- 8-bit Traffic Class field to identify specific classes of packets needing special QoS
 - Same as new definition of IPv4 Type-of-Service byte
 - May be initialized by source or by router enroute; may be rewritten by routers enroute
 - Traffic Class value of 0 used when no special QoS requested (the common case today)

IPv6 Standards

- Core IPv6 specifications are IETF Draft Standards
→ well-tested & stable
 - IPv6 base spec, ICMPv6, Neighbor Discovery, PMTU Discovery,...
- Other important specs are further behind on the standards track, but in good shape
 - Mobile IPv6, header compression,...
 - For up-to-date status: www.ipv6tf.org
- 3GPP UMTS Rel. 5 cellular wireless standards (2002) mandate IPv6; also being considered by 3GPP2

IPv6 Status – Standardisation

- Several key components on standards track...
 - Specification (RFC2460)
 - ICMPv6 (RFC4443)
 - RIP (RFC2080)
 - IGMPv6 (RFC2710)
 - Router Alert (RFC2711)
 - Autoconfiguration (RFC4862)
 - DHCPv6 (RFC3315 & 4361)
 - IPv6 Mobility (RFC3775)
 - GRE Tunnelling (RFC2473)
 - DAD for IPv6 (RFC4429)
 - ISIS for IPv6 (RFC5308)
 - Neighbour Discovery (RFC4861)
 - IPv6 Addresses (RFC4291 & 3587)
 - BGP (RFC2545)
 - OSPF (RFC5340)
 - Jumbograms (RFC2675)
 - Radius (RFC3162)
 - Flow Label (RFC6436/7/8)
 - Mobile IPv6 MIB (RFC4295)
 - Unique Local IPv6 Addresses (RFC4193)
 - Teredo (RFC4380)
 - VRRP (RFC5798)
- IPv6 available over:
 - PPP (RFC5072)
 - FDDI (RFC2467)
 - NBMA (RFC2491)
 - Frame Relay (RFC2590)
 - IEEE1394 (RFC3146)
 - Facebook (RFC5514)
 - Ethernet (RFC2464)
 - Token Ring (RFC2470)
 - ATM (RFC2492)
 - ARCnet (RFC2497)
 - FibreChannel (RFC4338)

Recent IPv6 Hot Topics

- IPv4 depletion debate
 - IANA IPv4 pool ran out on 3rd February 2011
 - <http://www.potaroo.net/tools/ipv4/>
- IPv6 Transition “assistance”
 - CGN, 6rd, NAT64,IVI, DS-Lite, 6to4, A+P...
- Mobile IPv6
- Multihoming
 - SHIM6 “dead”, Multihoming in IPv6 same as in IPv4
- IPv6 Security
 - Security industry & experts taking much closer look

Conclusion

- Protocol is “ready to go”
- The core components have already seen several years field experience

The IPv6 Protocol & IPv6 Standards

ISP Workshops

IPv6 Addressing

ISP Workshops

Where to get IPv6 addresses

- Your upstream ISP
- Africa
 - AfriNIC – <http://www.afrinic.net>
- Asia and the Pacific
 - APNIC – <http://www.apnic.net>
- North America
 - ARIN – <http://www.arin.net>
- Latin America and the Caribbean
 - LACNIC – <http://www.lacnic.net>
- Europe and Middle East
 - RIPE NCC – <http://www.ripe.net/info/ncc>

Internet Registry Regions



Getting IPv6 address space (1)

- From your Regional Internet Registry
 - Become a member of your Regional Internet Registry and get your own allocation
 - Membership usually open to all network operators
 - General allocation policies are outlined in RFC2050
 - RIR specific policy details for IPv6 allocations are listed on the individual RIR website
 - Open to all organisations who are operating a network
 - Receive a /32 (or larger if you will have more than 65k /48 assignments)

Getting IPv6 address space (2)

- From your upstream ISP
 - Receive a /48 from upstream ISP's IPv6 address block
 - Receive more than one /48 if you have more than 65k subnets
- If you need to multihome:
 - Apply for a /48 assignment from your RIR
 - Multihoming with provider's /48 will be operationally challenging
 - Provider policies, filters, etc

Using 6to4 for IPv6 address space

- Some entities still use 6to4
 - Not recommended due to operational problems
 - Read <http://datatracker.ietf.org/doc/draft-ietf-v6ops-6to4-to-historic> for some of the reasoning why
- FYI: 6to4 operation:
 - Take a single public IPv4 /32 address
 - 2002:<ipv4 /32 address>::/48 becomes your IPv6 address block, giving 65k subnets
 - Requires a 6to4 gateway
 - 6to4 is a means of connecting IPv6 islands across the IPv4 Internet

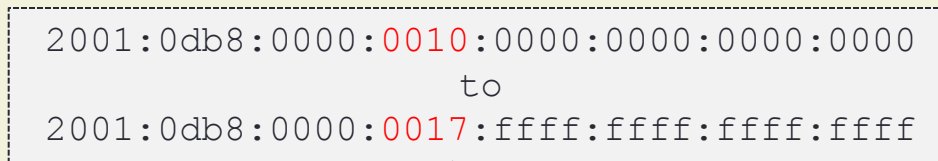
Nibble Boundaries

- IPv6 offers network operators more flexibility with addressing plans
 - Network addressing can now be done on nibble boundaries
 - For ease of operation
 - Rather than making maximum use of a very scarce resource
 - With the resulting operational complexity
- A nibble boundary means subnetting address space based on the address numbering
 - Each number in IPv6 represents 4 bits = 1 nibble
 - Which means that IPv6 addressing can be done on 4-bit boundaries

Nibble Boundaries – example

- Consider the address block 2001:db8:0:10::/61
 - The range of addresses in this block are:

```
2001:0db8:0000:0010:0000:0000:0000:0000
                    to
2001:0db8:0000:0017:ffff:ffff:ffff:ffff
```



- Note that this subnet only runs from 0010 to 0017.
- The adjacent block is 2001:db8:0:18::/61

```
2001:0db8:0000:0018:0000:0000:0000:0000
                    to
2001:0db8:0000:001f:ffff:ffff:ffff:ffff
```

- The address blocks don't use the entire nibble range

Nibble Boundaries – example

- Now consider the address block 2001:db8:0:10::/60
 - The range of addresses in this block are:

```
2001:0db8:0000:0010:0000:0000:0000:0000
to
2001:0db8:0000:001f:ffff:ffff:ffff:ffff
```

- Note that this subnet uses the entire nibble range, 0 to f
- Which makes the numbering plan for IPv6 simpler
 - This range can have a particular meaning within the ISP block (for example, infrastructure addressing for a particular PoP)

Addressing Plans – Infrastructure

- All Network Operators should obtain a /32 from their RIR
- Address block for router loop-back interfaces
 - Number all loopbacks out of **one** /64
 - /128 per loopback
- Address block for infrastructure (backbone)
 - /48 allows 65k subnets
 - /48 per region (for the largest multi-national networks)
 - /48 for whole backbone (for the majority of networks)
 - Infrastructure/backbone usually does NOT require regional/geographical addressing
 - Summarise between sites if it makes sense

Addressing Plans – Infrastructure

- What about LANs?
 - /64 per LAN
- What about Point-to-Point links?
 - Protocol design expectation is that /64 is used
 - /127 now recommended/standardised
 - <http://www.rfc-editor.org/rfc/rfc6164.txt>
 - (reserve /64 for the link, but address it as a /127)
 - Other options:
 - /126s are being used (mimics IPv4 /30)
 - /112s are being used
 - Leaves final 16 bits free for node IDs
 - Some discussion about /80s, /96s and /120s too

Addressing Plans – Infrastructure

- NOC:
 - ISP NOC is “trusted” network and usually considered part of infrastructure /48
 - Contains management and monitoring systems
 - Hosts the network operations staff
 - take the last /60 (allows enough subnets)
- Critical Services:
 - Network Operator’s critical services are part of the “trusted” network and should be considered part of the infrastructure /48
 - For example, Anycast DNS, SMTP, POP3/IMAP, etc
 - Take the second /64
 - (some operators use the first /64 instead)

Addressing Plans – ISP to Customer

- Option One:
 - Use ipv6 unnumbered
 - Which means no global unicast ipv6 address on the point-to-point link
 - Router adopts the specified interface's IPv6 address
 - Router doesn't actually need a global unicast IPv6 address to forward packets

```
interface loopback 0
  ipv6 address 2001:db8::1/128
interface serial 1/0
  ipv6 address unnumbered loopback 0
```

Addressing Plans – ISP to Customer

- Option Two:
 - Use the second /48 for point-to-point links
 - Divide this /48 up between PoPs
 - Example:
 - For 10 PoPs, dividing into 16, gives /52 per PoP
 - Each /52 gives 4096 point-to-point links
 - Adjust to suit!
 - Useful if ISP monitors point-to-point link state for customers
 - Link addresses are **untrusted**, so do not want them in the first /48 used for the backbone &c
 - Aggregate per router or per PoP and carry in iBGP (not ISIS/OSPF)

Addressing Plans – Customer

- Customers get **one** /48
 - Unless they have more than 65k subnets in which case they get a second /48 (and so on)
- In typical deployments today:
 - Several ISPs are giving small customers a /56 and single LAN end-sites a /64, e.g.:
 - /64 if end-site will only ever be a LAN
 - /56 for small end-sites (e.g. home/office/small business)
 - /48 for large end-sites
 - This is another very active discussion area
 - Observations:
 - Don't assume that a mobile endsite needs only a /64
 - Some operators are distributing /60s to their smallest customers!!

Addressing Plans – Customer

- Consumer Broadband Example:
 - DHCPv6 pool is a /48
 - DHCPv6 hands out /60 per customer
 - Which allows for 4096 customers per pool
- Business Broadband Example:
 - DHCPv6 pool is a /48
 - DHCPv6 hands out /56 per customer
 - Which allows for 256 customers per pool
 - If BRAS has more than 256 business customers, increase pool to a /47
 - This allows for 512 customers at /56 per customer
 - Increasing pool to /46 allows for 1024 customers
 - BRAS announces entire pool as one block by iBGP

Addressing Plans – Customer

- Business “leased line”:
 - /48 per customer
 - One stop shop, no need for customer to revisit ISP for more addresses until all 65k subnets are used up
- Hosted services:
 - One physical server per vLAN
 - One /64 per vLAN
 - How many vLANs per PoP?
 - /48 reserved for entire hosted servers across backbone
 - Internal sites will be subnets and carried by iBGP

Addressing Plans – Customer

- Geographical delegations to Customers:
 - Network Operator subdivides /32 address block into geographical chunks
 - E.g. into /36s
 - Region 1: 2001:db8:1xxx::/36
 - Region 2: 2001:db8:2xxx::/36
 - Region 3: 2001:db8:3xxx::/36
 - etc
 - Which gives 4096 /48s per region
 - For Operational and Administrative ease
 - Benefits for traffic engineering if Network Operator multihomes in each region

Addressing Plans – Customer

- Sequential delegations to Customers:
 - After carving off address space for network infrastructure, Network Operator simply assigns address space sequentially
 - Eg:
 - Infrastructure: 2001:db8:0::/48
 - Customer P2P: 2001:db8:1::/48
 - Customer 1: 2001:db8:2::/48
 - Customer 2: 2001:db8:3::/48
 - etc
 - Useful when there is no regional subdivision of network and no regional multihoming needs

Addressing Plans – Routing Considerations

- Carry Broadband pools in iBGP across the backbone
 - Not in OSPF/ISIS
- Multiple Broadband pools on one BRAS should be aggregated if possible
 - Reduce load on iBGP
- Aggregating leased line customer address blocks per router or per PoP is undesirable:
 - Interferes with ISP's traffic engineering needs
 - Interferes with ISP's service quality and service guarantees

Addressing Plans – Traffic Engineering

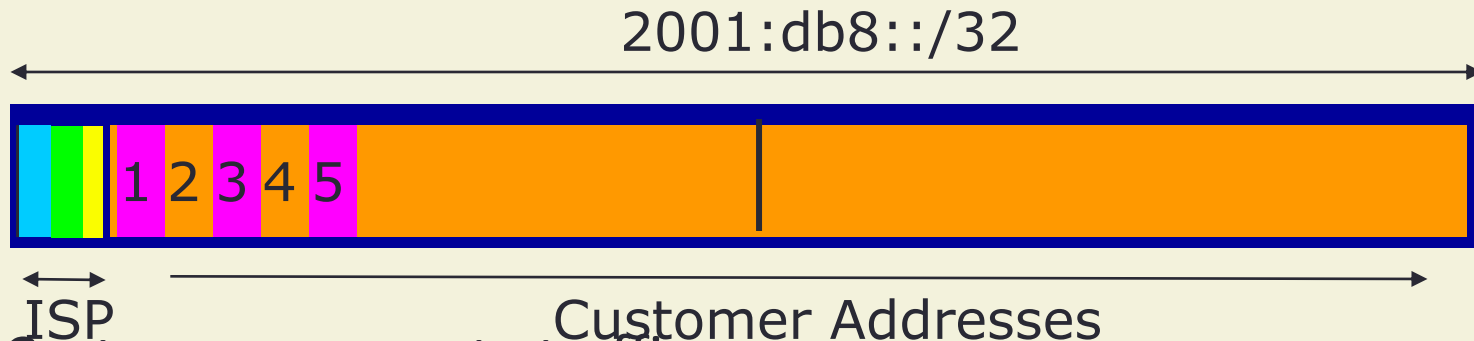
- Smaller providers will be single homed
 - The customer portion of the ISP's IPv6 address block will usually be assigned sequentially
- Larger providers will be multihomed
 - Two, three or more external links from different providers
 - Traffic engineering becomes important
 - Sequential assignments of customer addresses will negatively impact load balancing

Addressing Plans – Traffic Engineering

- ISP Router loopbacks and backbone point-to-point links make up a small part of total address space
 - And they don't attract traffic, unlike customer address space
- Links from ISP Aggregation edge to customer router needs one /64
 - Small requirements compared with total address space
 - Some ISPs use IPv6 unnumbered
- Planning customer assignments is a very important part of multihoming
 - Traffic engineering involves subdividing aggregate into pieces until load balancing works

Unplanned IP addressing

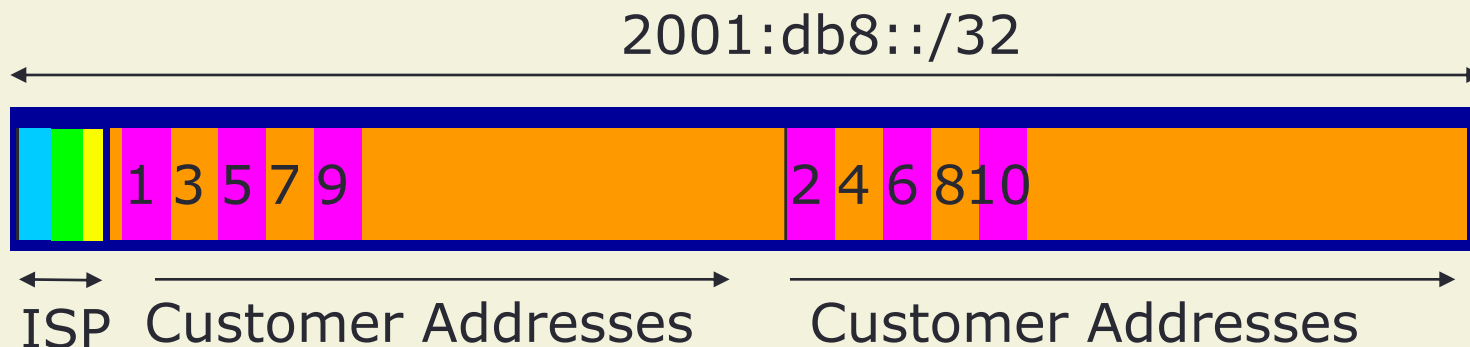
- ISP fills up customer IP addressing from one end of the range:



- Customers generate traffic
 - Dividing the range into two pieces will result in one /33 with all the customers and the ISP infrastructure the addresses, and one /33 with nothing
 - No loadbalancing as all traffic will come in the first /33
 - Means further subdivision of the first /33 = harder work

Planned IP addressing

- If ISP fills up customer addressing from both ends of the range:



- Scheme then is:
 - First customer from first /33, second customer from second /33, third from first /33, etc
- This works also for residential versus commercial customers:
 - Residential from first /33
 - Commercial from second /33

Planned IP Addressing

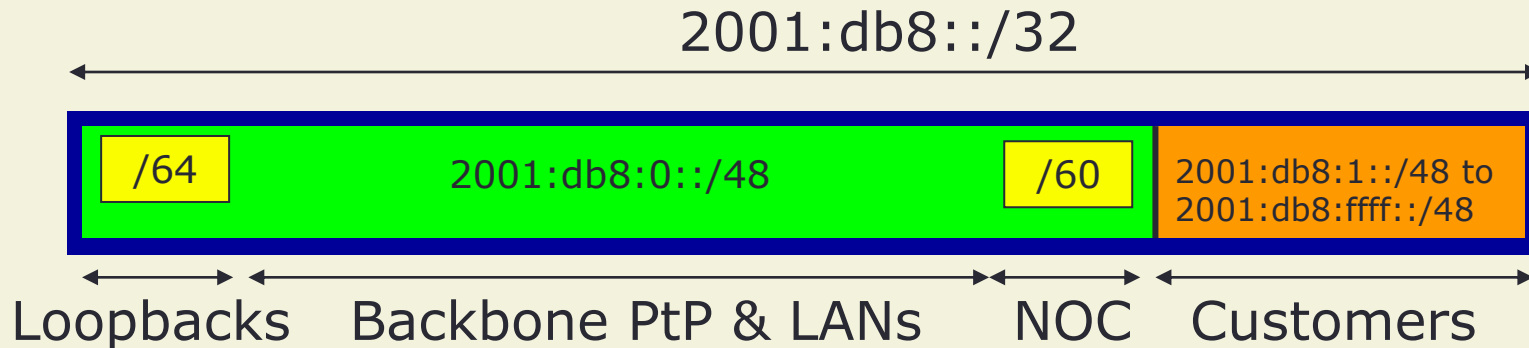
- This works fine for multihoming between two upstream links (same or different providers)
- Can also subdivide address space to suit more than two upstreams
 - Follow a similar scheme for populating each portion of the address space
- Consider regional (geographical) distribution of customer delegated address space
- Don't forget to always announce an aggregate out of each link

Addressing Plans – Advice

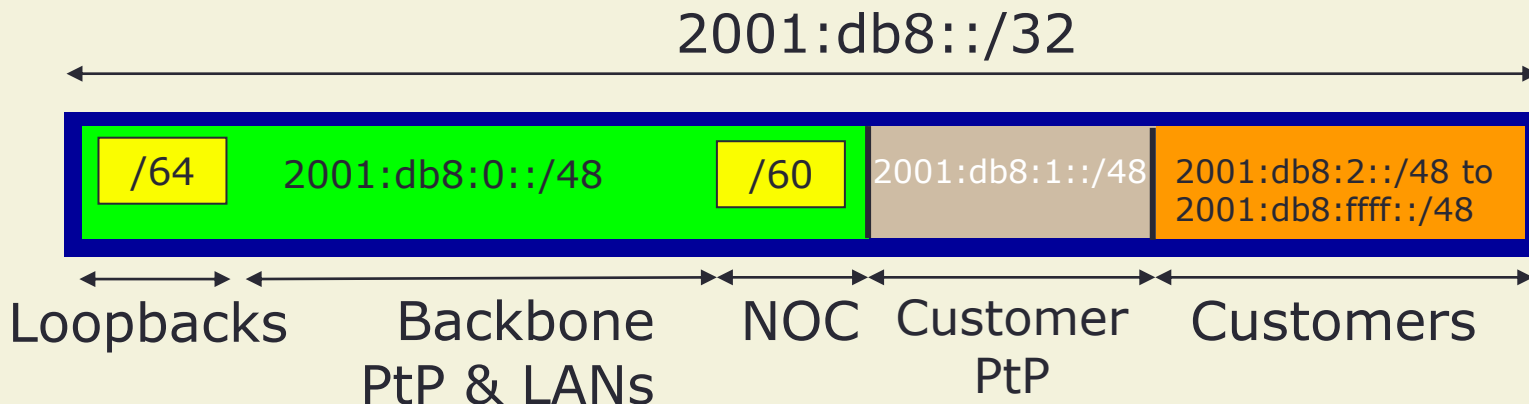
- Customer address assignments should not be reserved or assigned on a per PoP basis
 - Follow same principle as for IPv4
 - Subnet aggregate to cater for multihoming needs
 - Consider regional delegation
 - ISP iBGP carries customer nets
 - Aggregation within the iBGP not required and usually not desirable
 - Aggregation in eBGP is very necessary
- Backbone infrastructure assignments:
 - Number out of a **single** /48
 - Operational simplicity and security
 - Aggregate to minimise size of the IGP

Addressing Plans – Scheme

- Looking at Infrastructure:



Alternative:



Addressing Plans Planning

- Registries will usually allocate the next block to be contiguous with the first allocation
 - (RIRs use a sparse allocation strategy – industry goal is aggregation)
 - Minimum allocation is /32
 - Very likely that subsequent allocation will make this up to a /31 or larger (/28)
 - So plan accordingly

Addressing Plans (contd)

- Document infrastructure allocation
 - Eases operation, debugging and management
- Document customer allocation
 - Customers get /48 each
 - Prefix contained in iBGP
 - Eases operation, debugging and management
 - Submit network object to RIR Database

Addressing Tools

- Examples of IP address planning tools:
 - NetDot netdot.uoregon.edu (recommended!!)
 - HaCi sourceforge.net/projects/haci
 - IPAT nethead.de/index.php/ipat
 - freeipdb home.globalcrossing.net/~freeipdb/
- Examples of IPv6 subnet calculators:
 - ipv6gen code.google.com/p/ipv6gen/
 - sipcalc www.routemeister.net/projects/sipcalc/

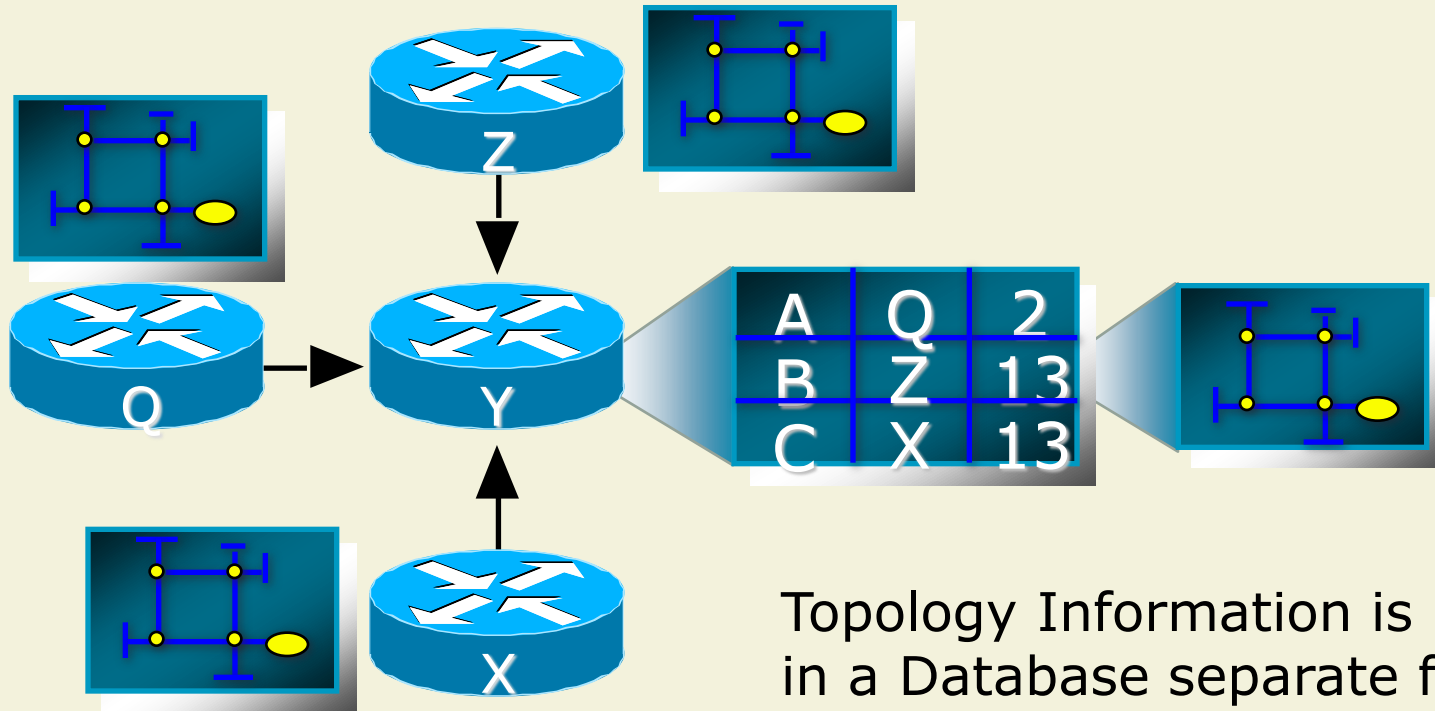
Introduction to OSPF

ISP Workshops

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

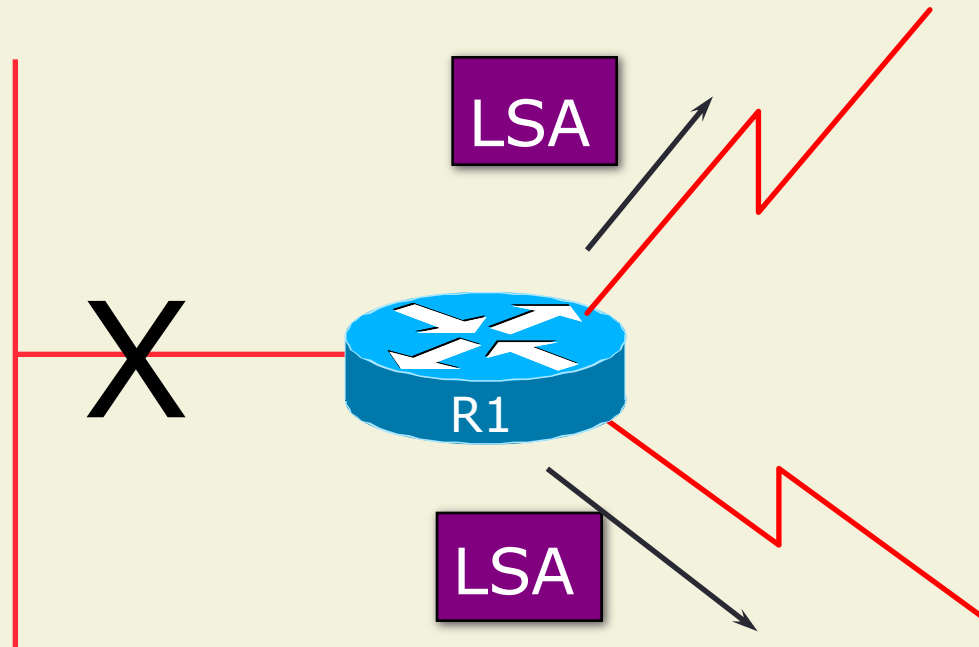


Topology Information is kept in a Database separate from the Routing Table

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

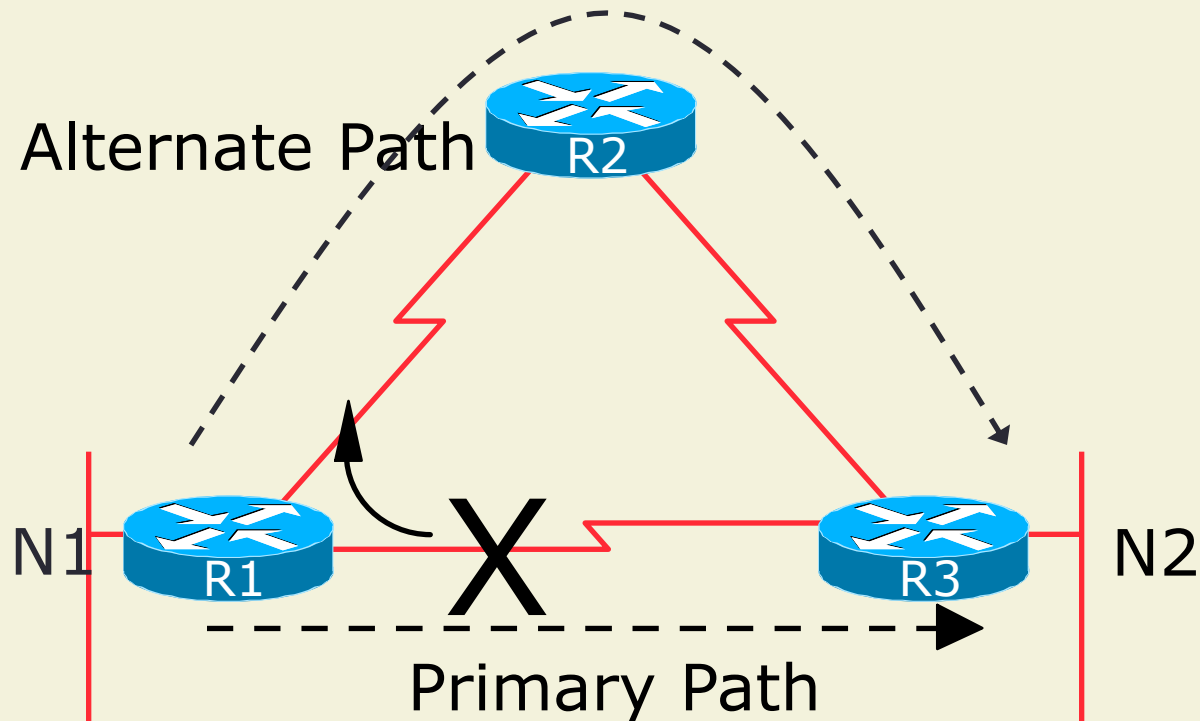
Low Bandwidth Utilisation



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

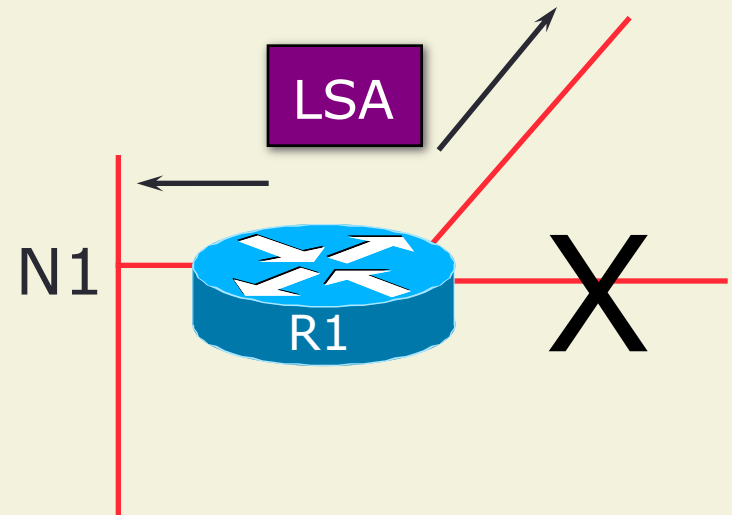
Fast Convergence

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



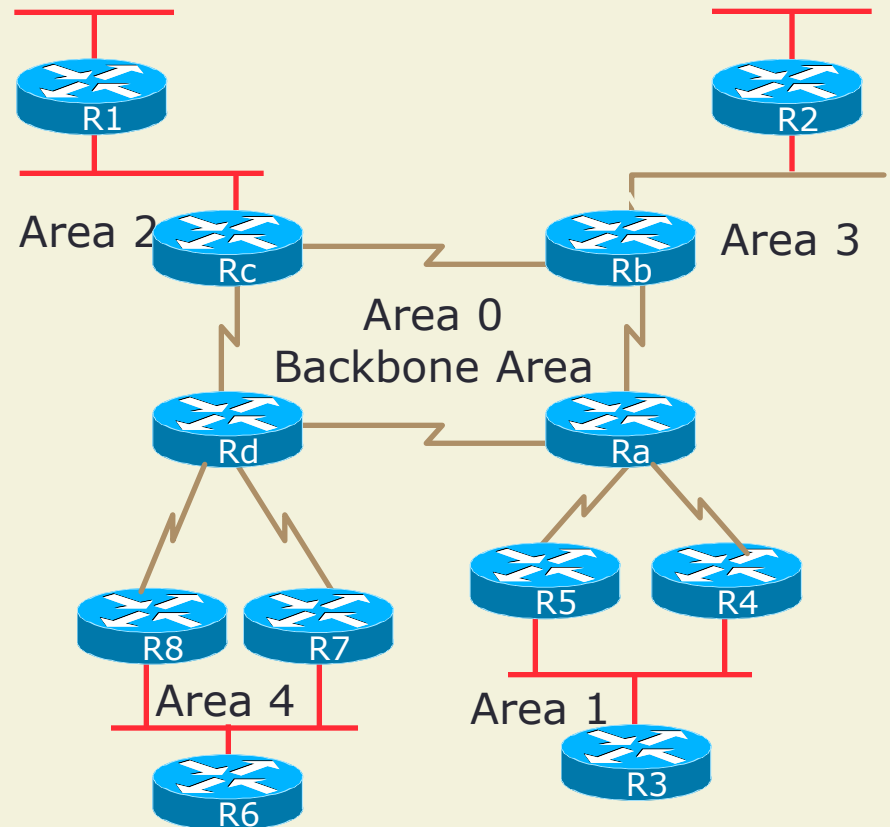
Fast Convergence

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



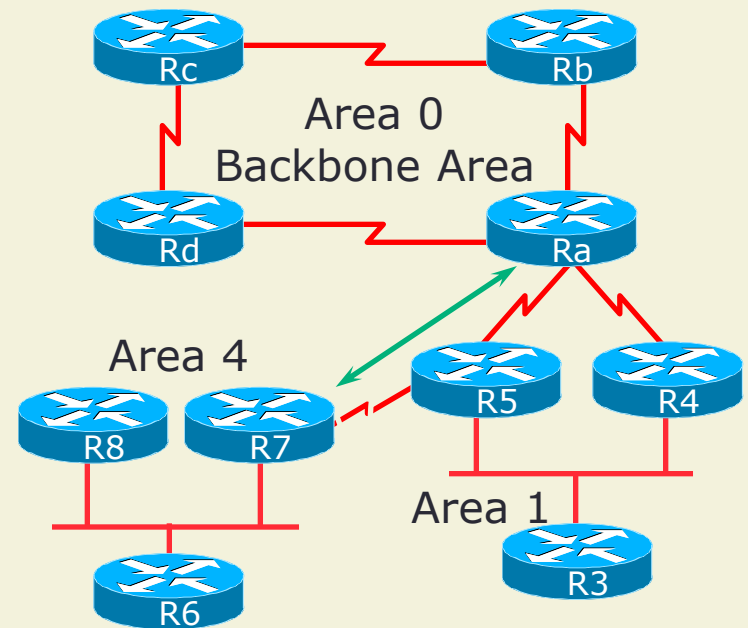
OSPF Areas

- 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

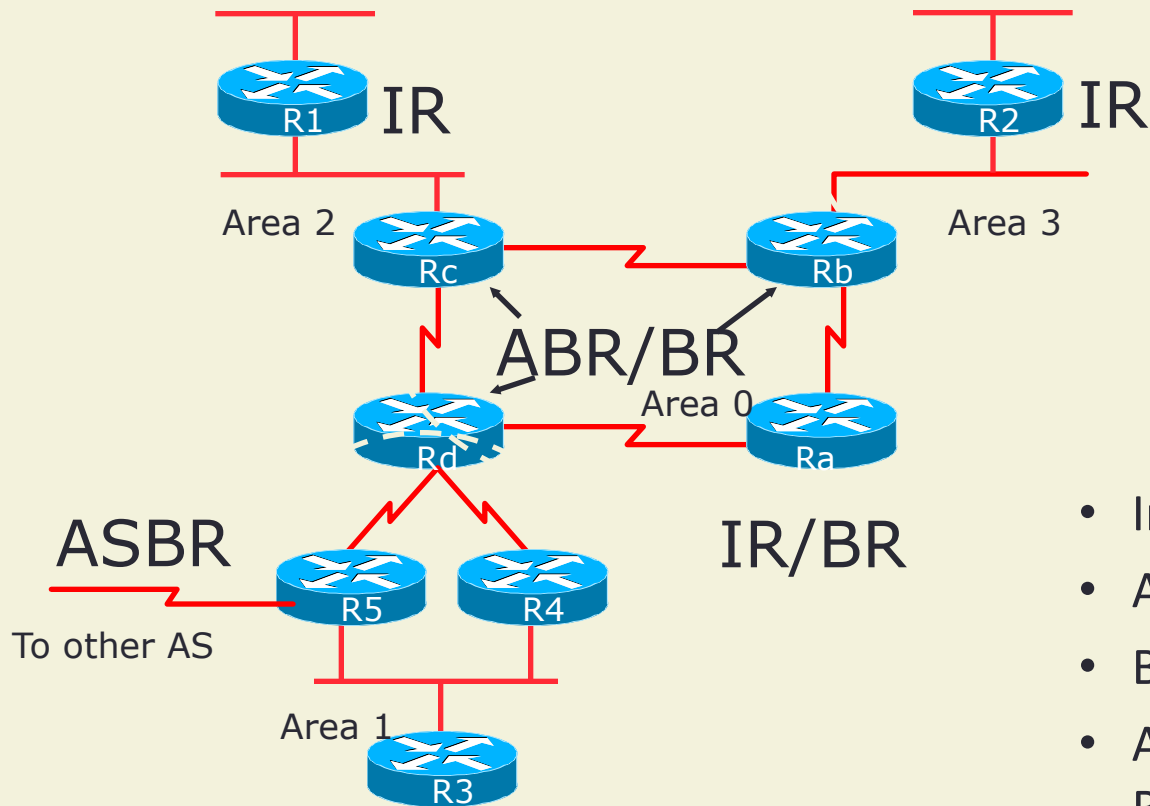


Virtual Links between OSPF Areas

- 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

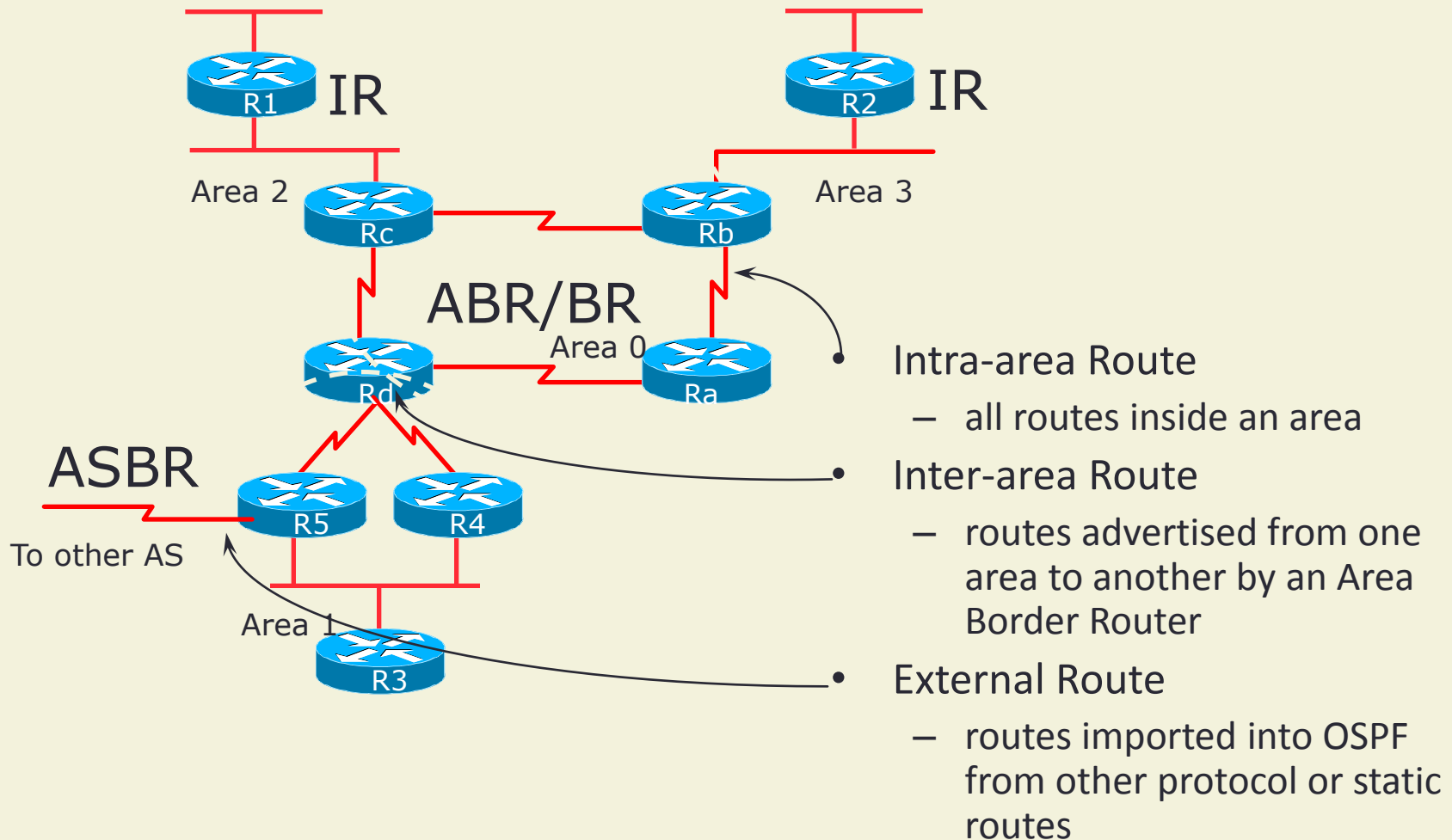


Classification of Routers



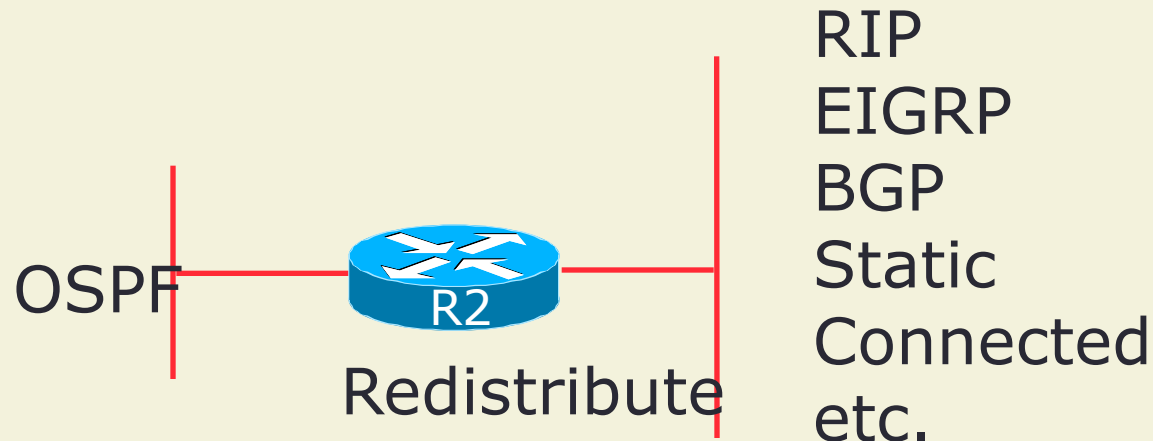
- Internal Router (IR)
- Area Border Router (ABR)
- Backbone Router (BR)
- Autonomous System Border Router (ASBR)

OSPF Route Types



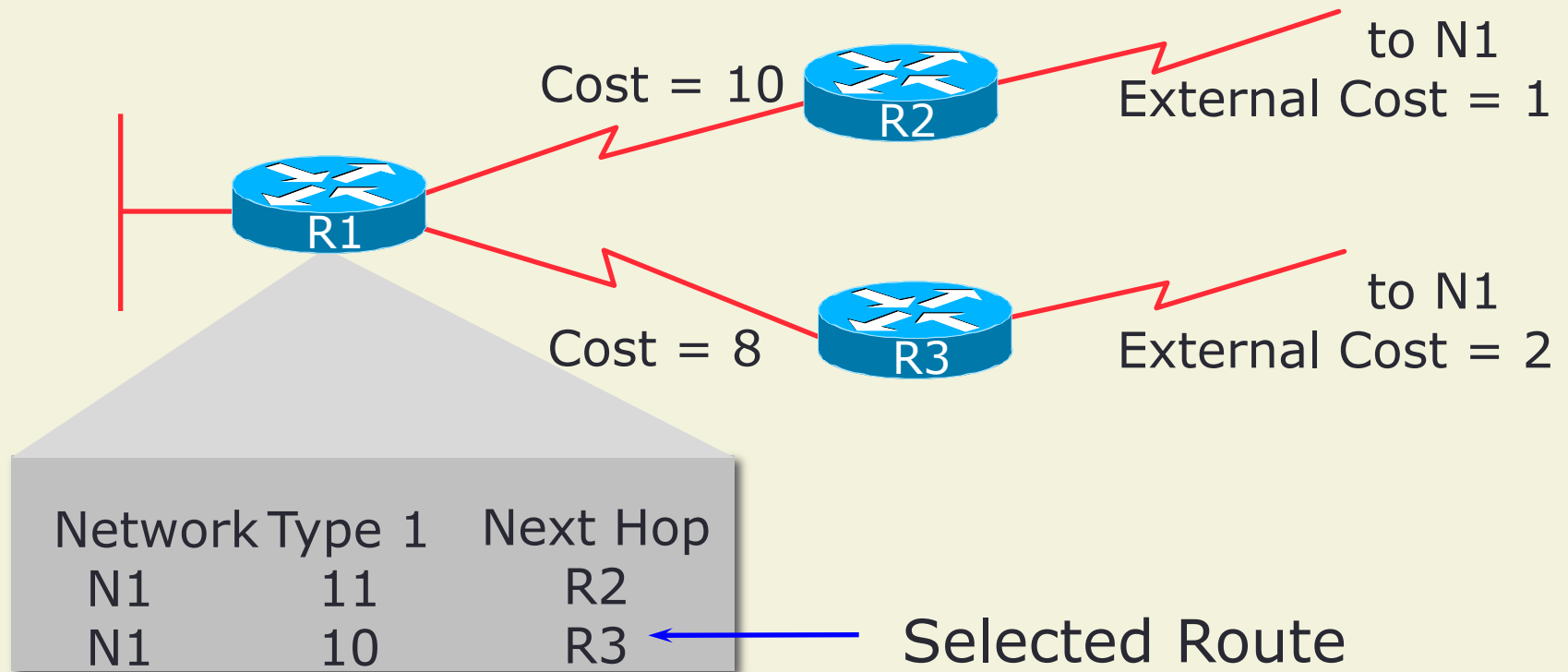
External Routes

- 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)



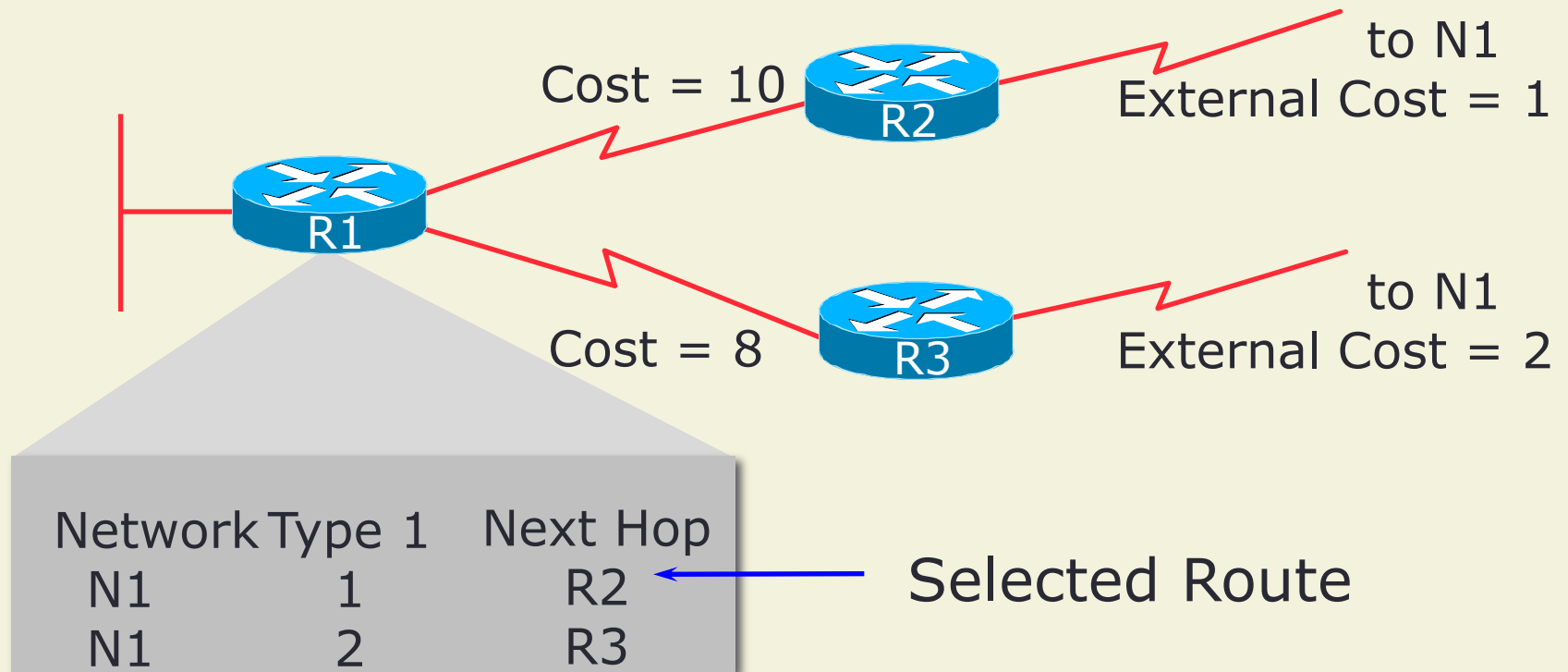
External Routes

- 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

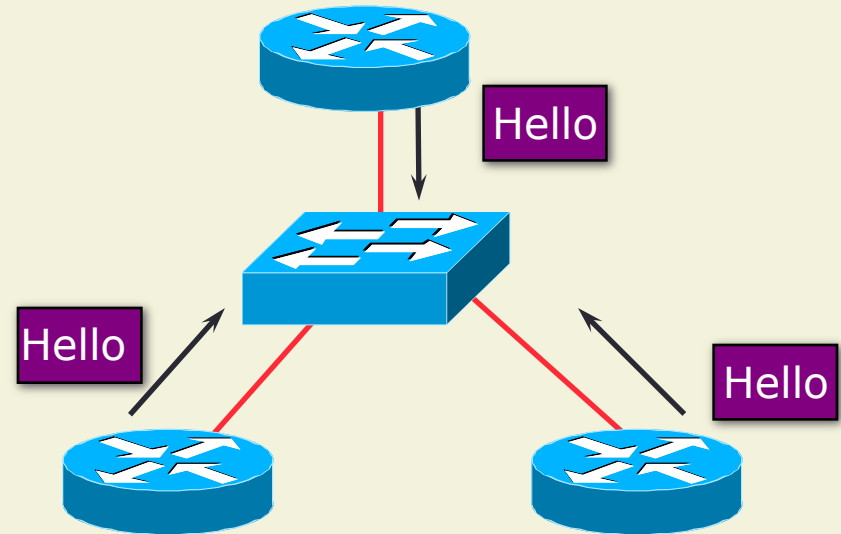


Topology/Link State Database

- 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

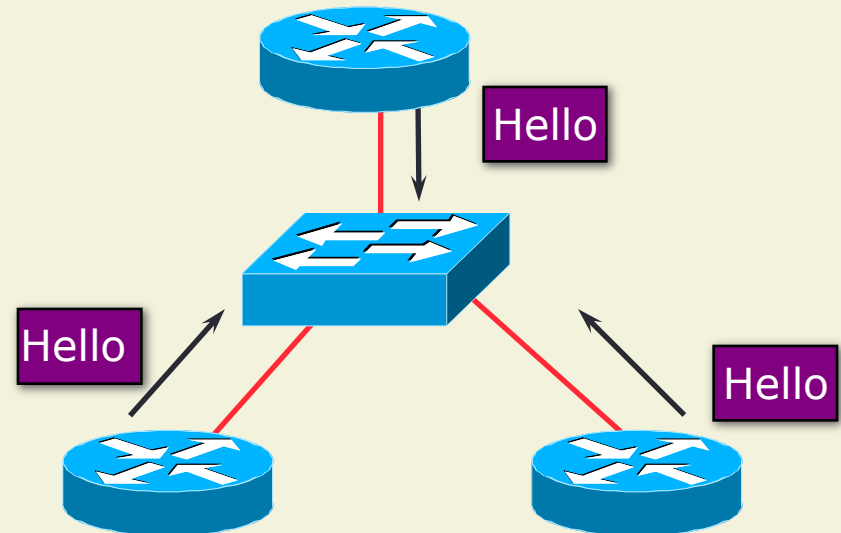
The Hello Protocol

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



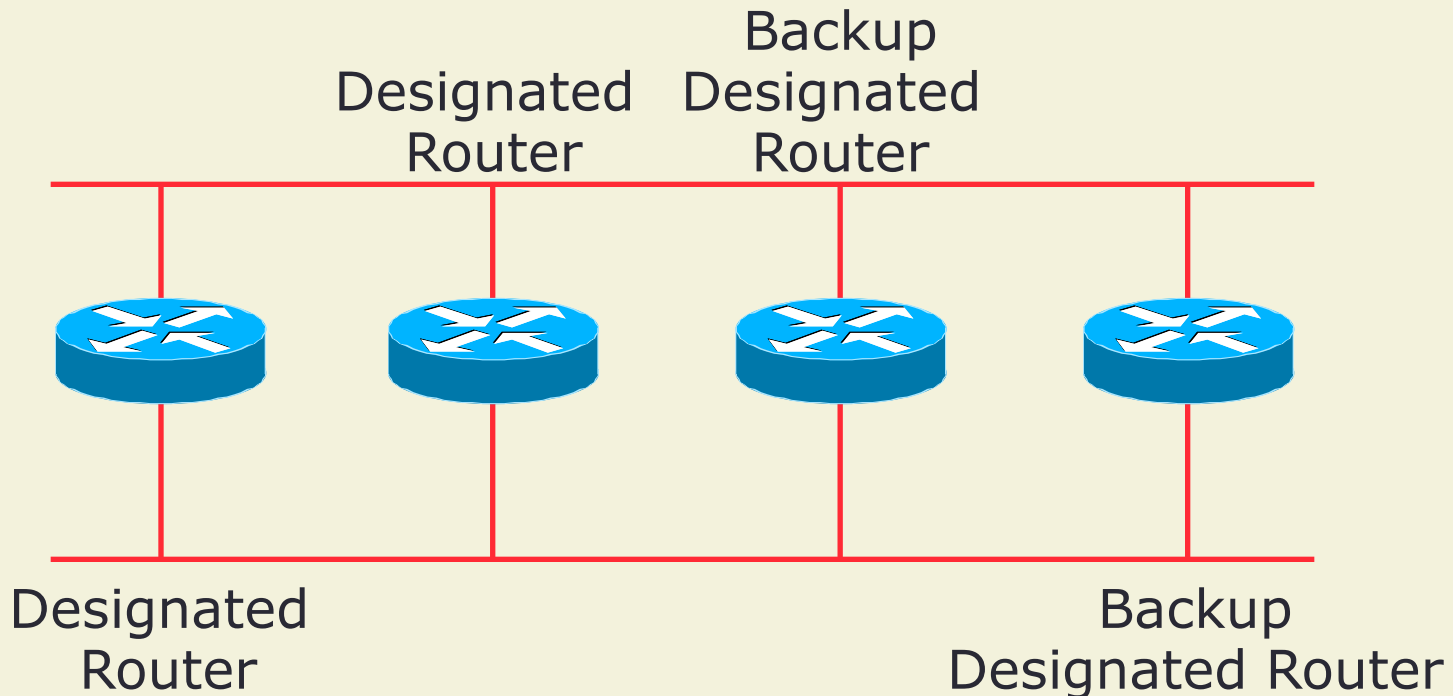
The Hello Packet

- 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)



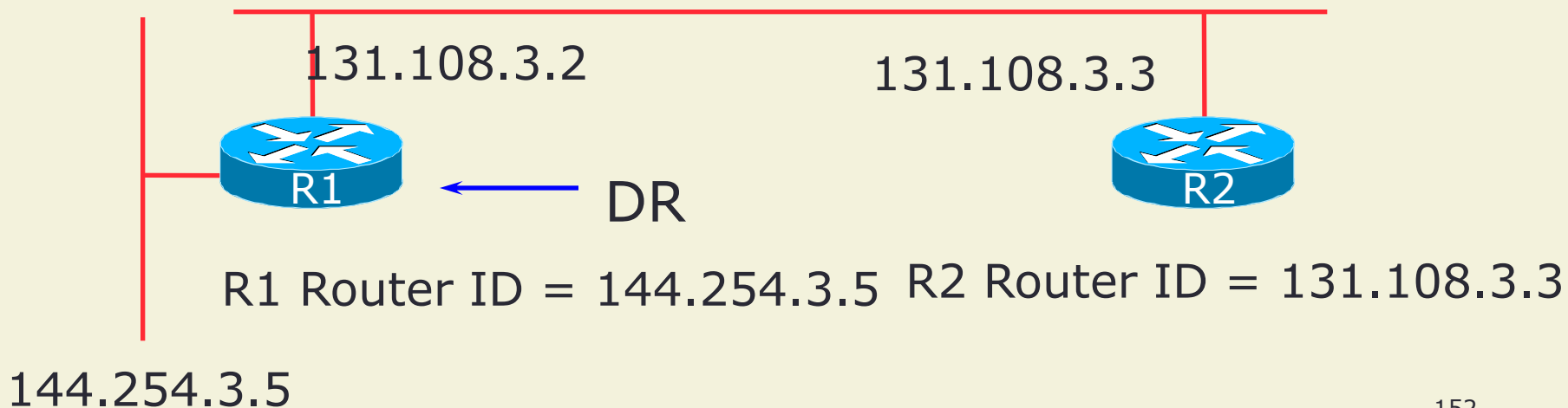
Designated Router

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



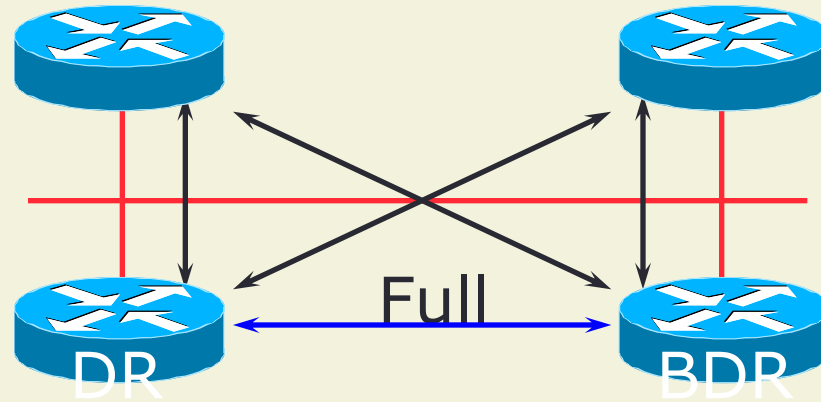
Designated Router by Priority

- 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



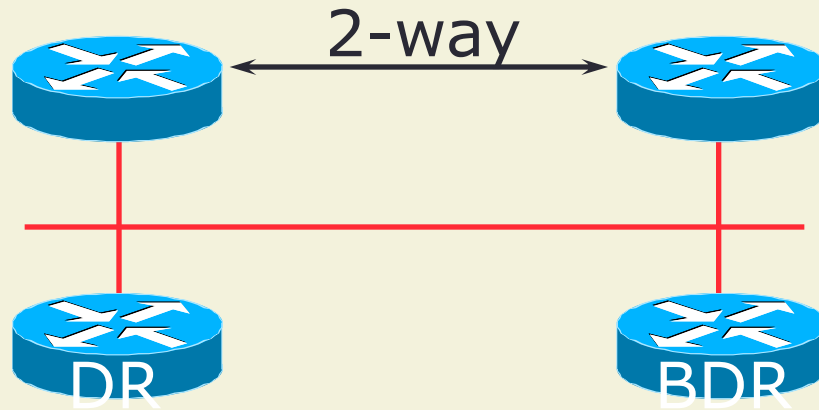
Neighbouring States

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



Neighbouring States

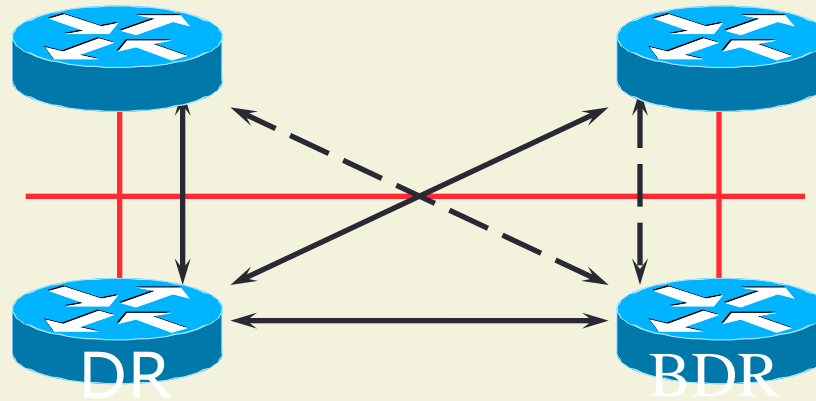
- 2-way
 - Router sees itself in other Hello packets
 - DR selected from neighbours in state 2-way or greater



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

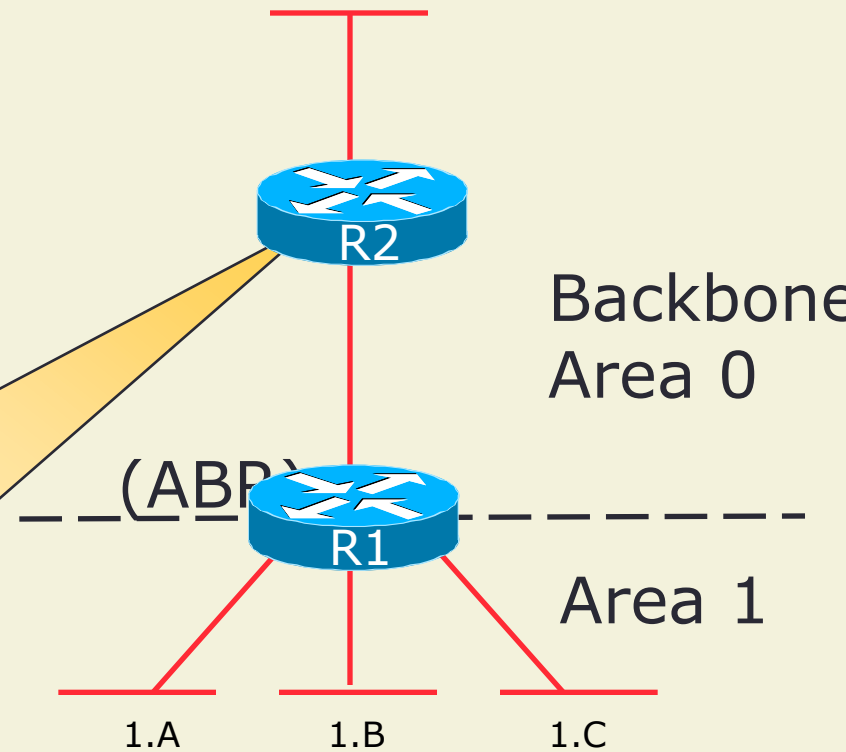
- Prefix or all subnets
- Prefix or all networks
- ‘Area range’ command

With summarisation

Network	Next Hop
1	R1

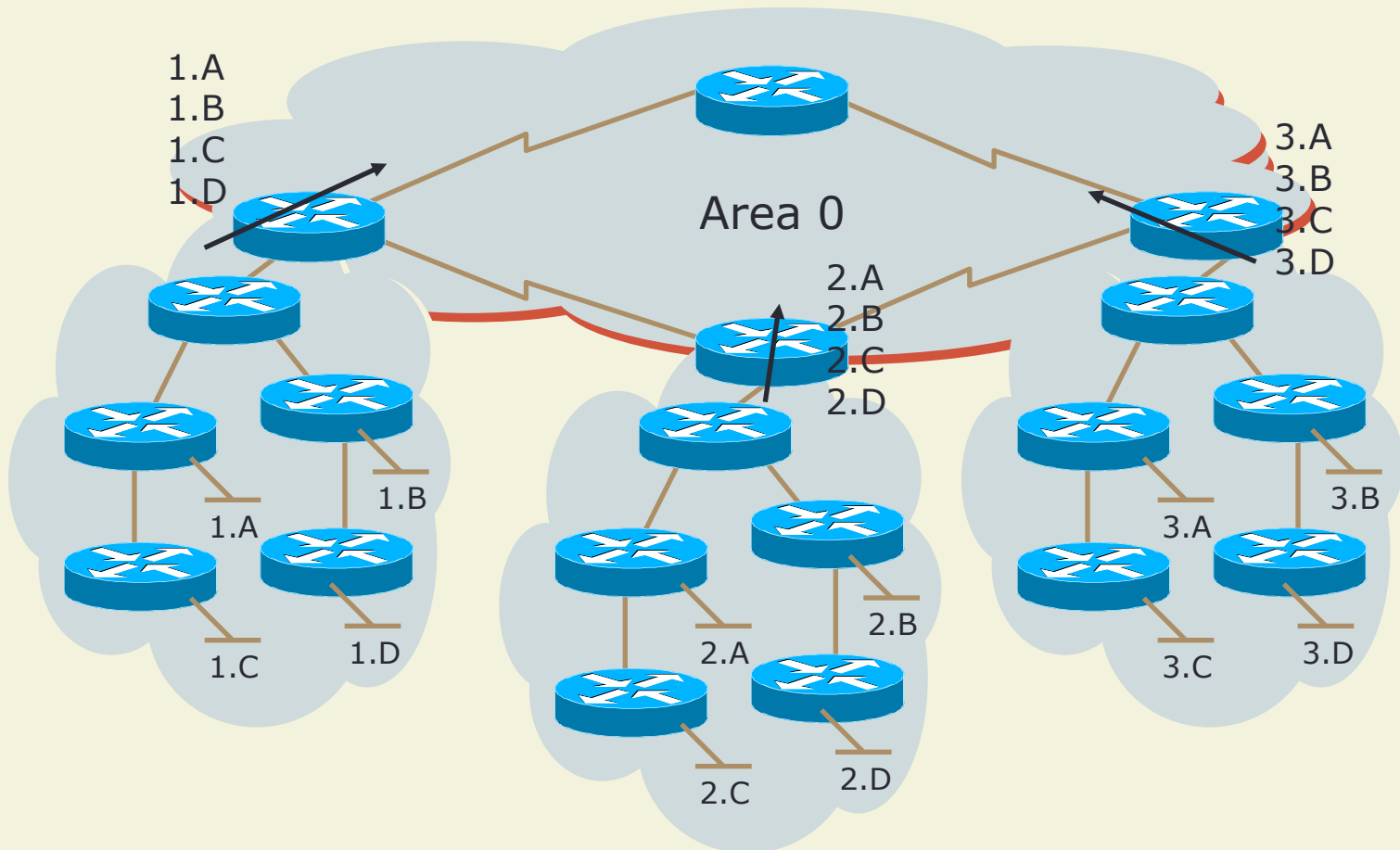
Without summarisation

Network	Next Hop
1.A	R1
1.B	R1
1.C	R1



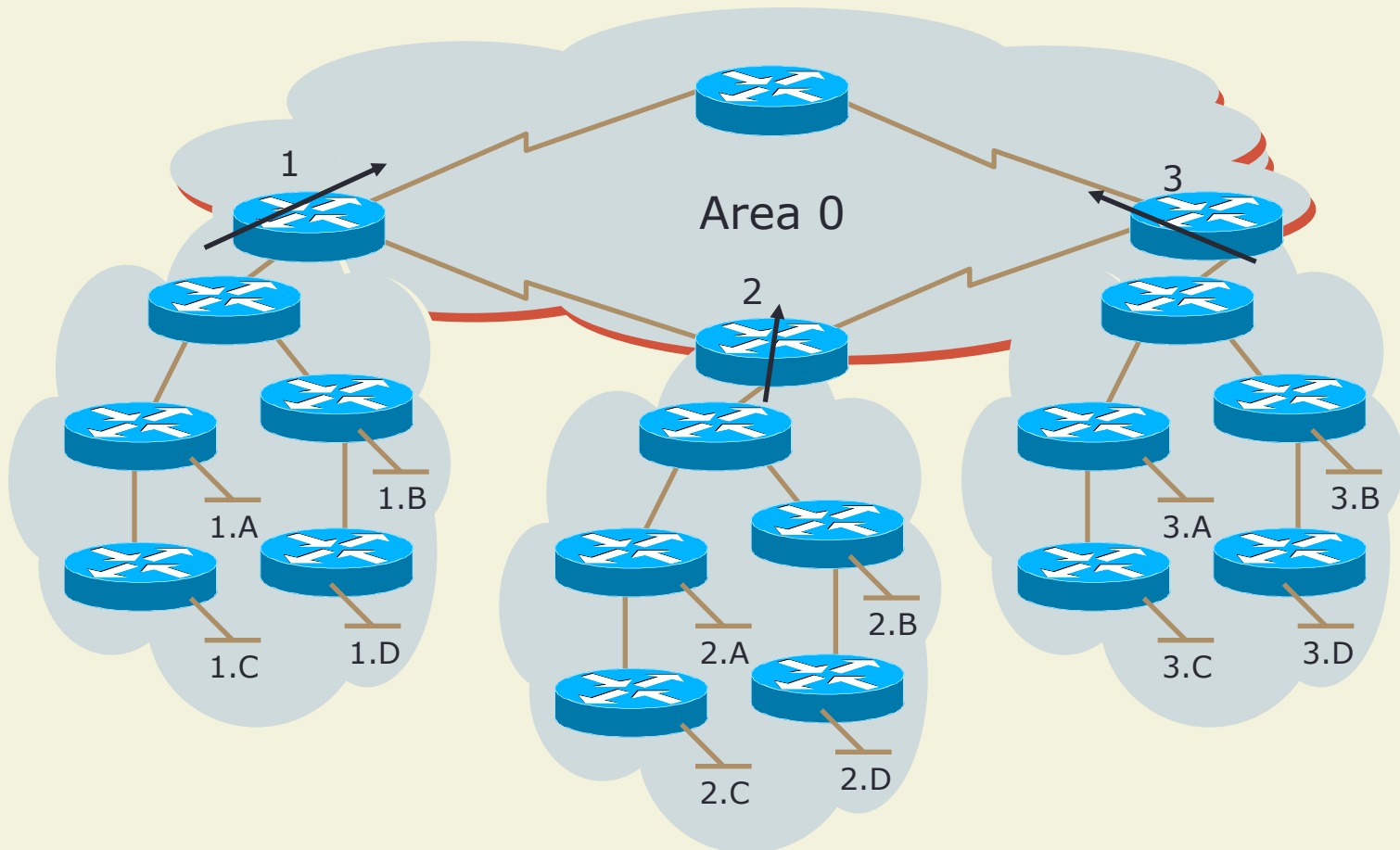
No Summarisation

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



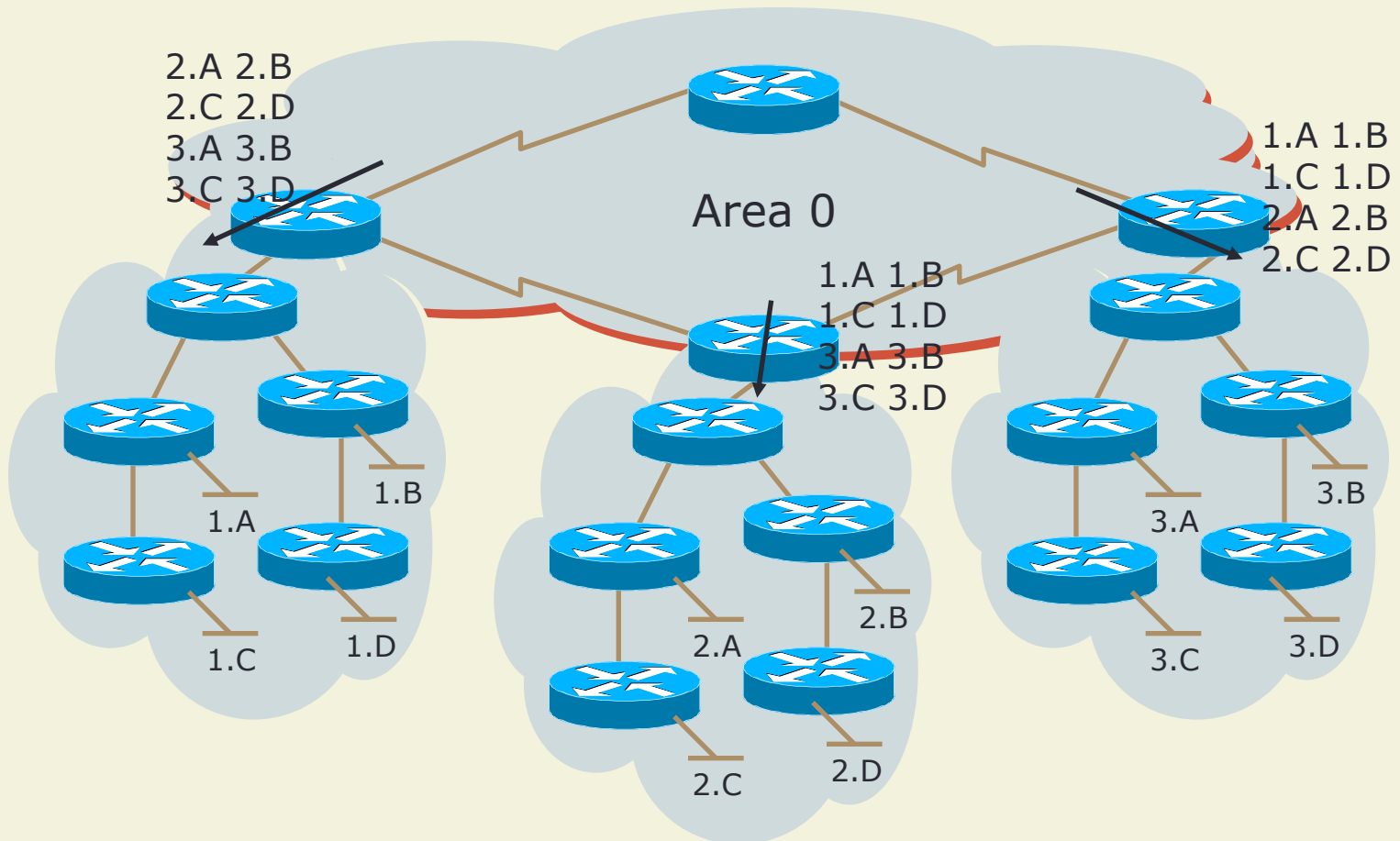
With Summarisation

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



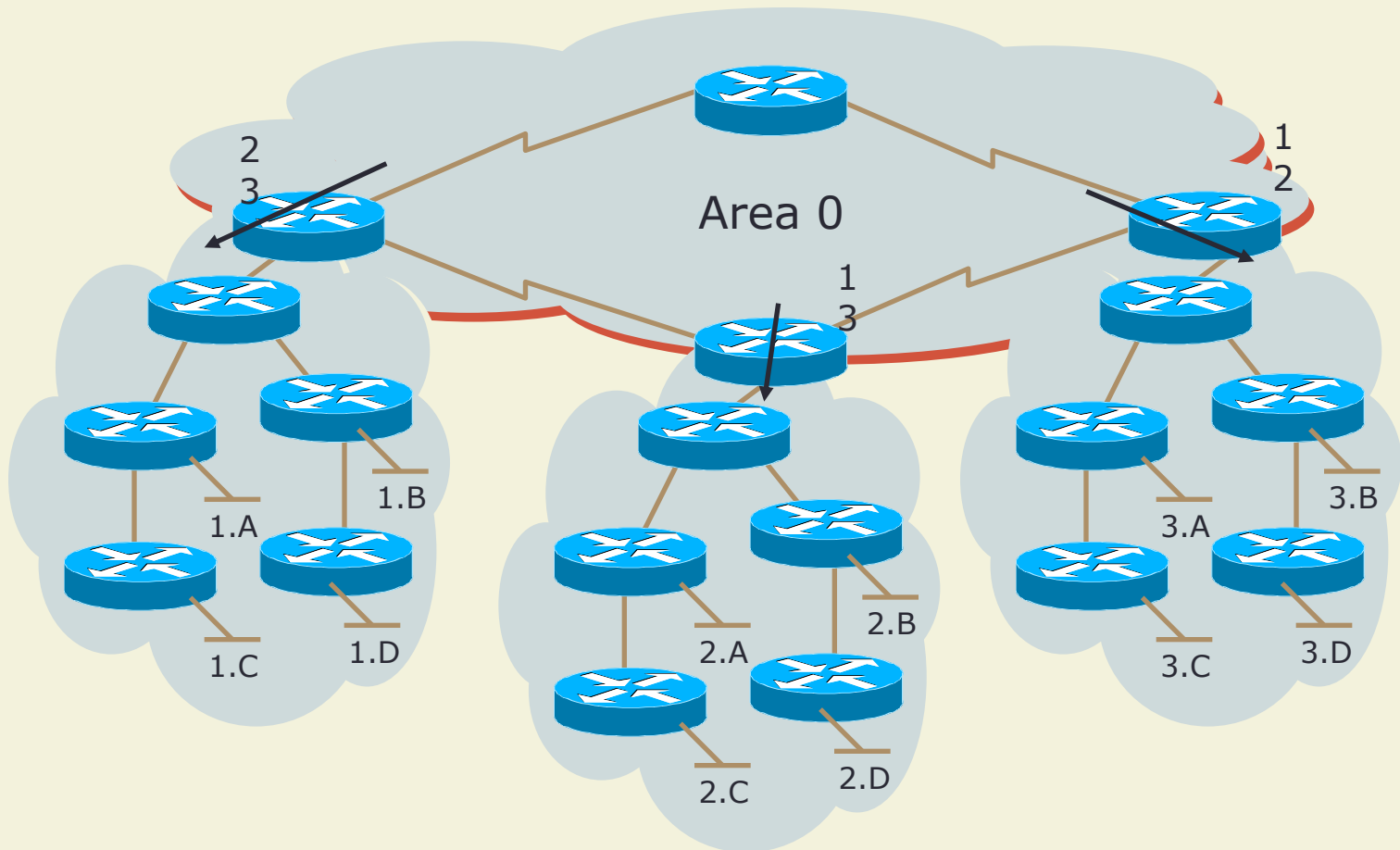
No Summarisation

- 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

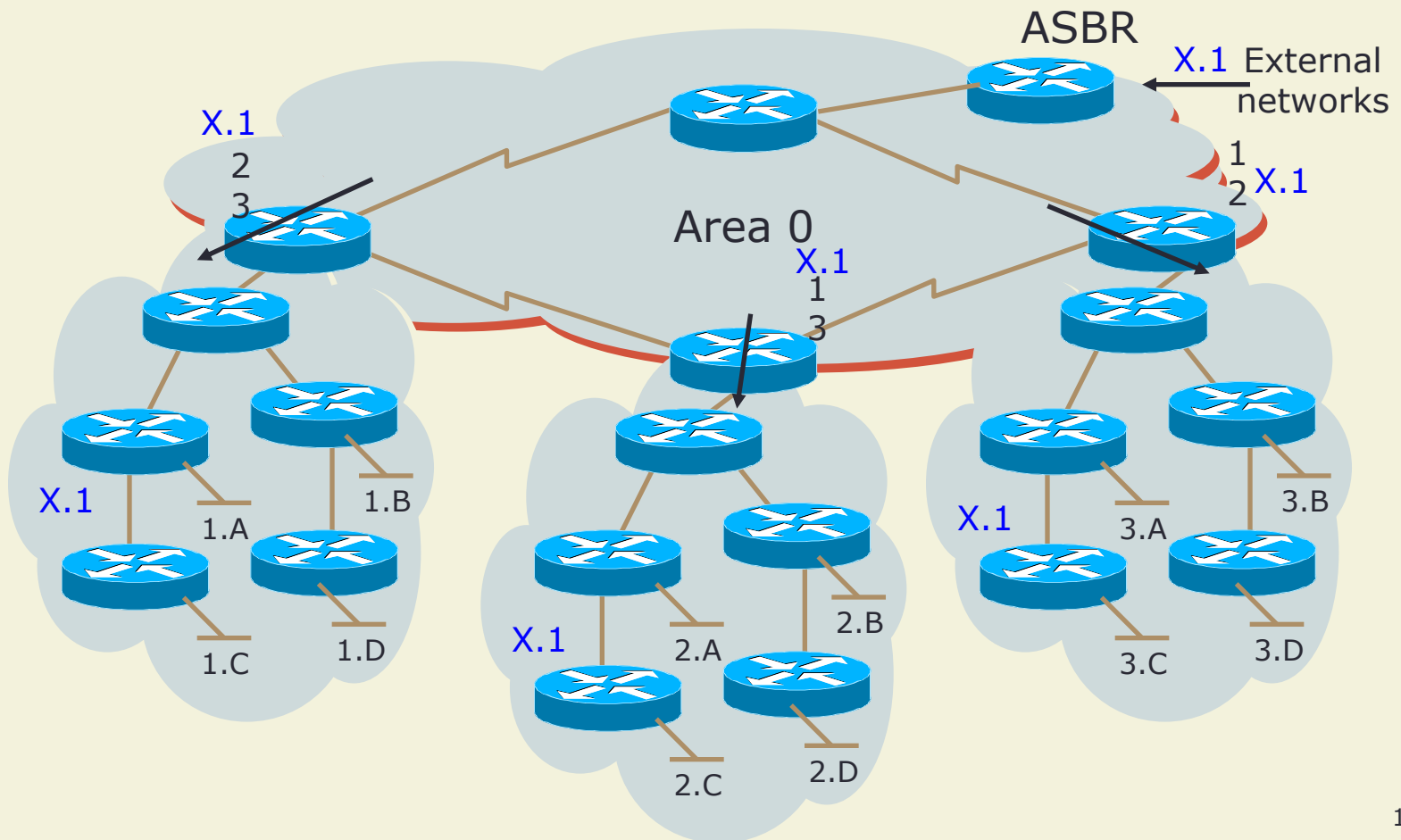


Types of Areas

- 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

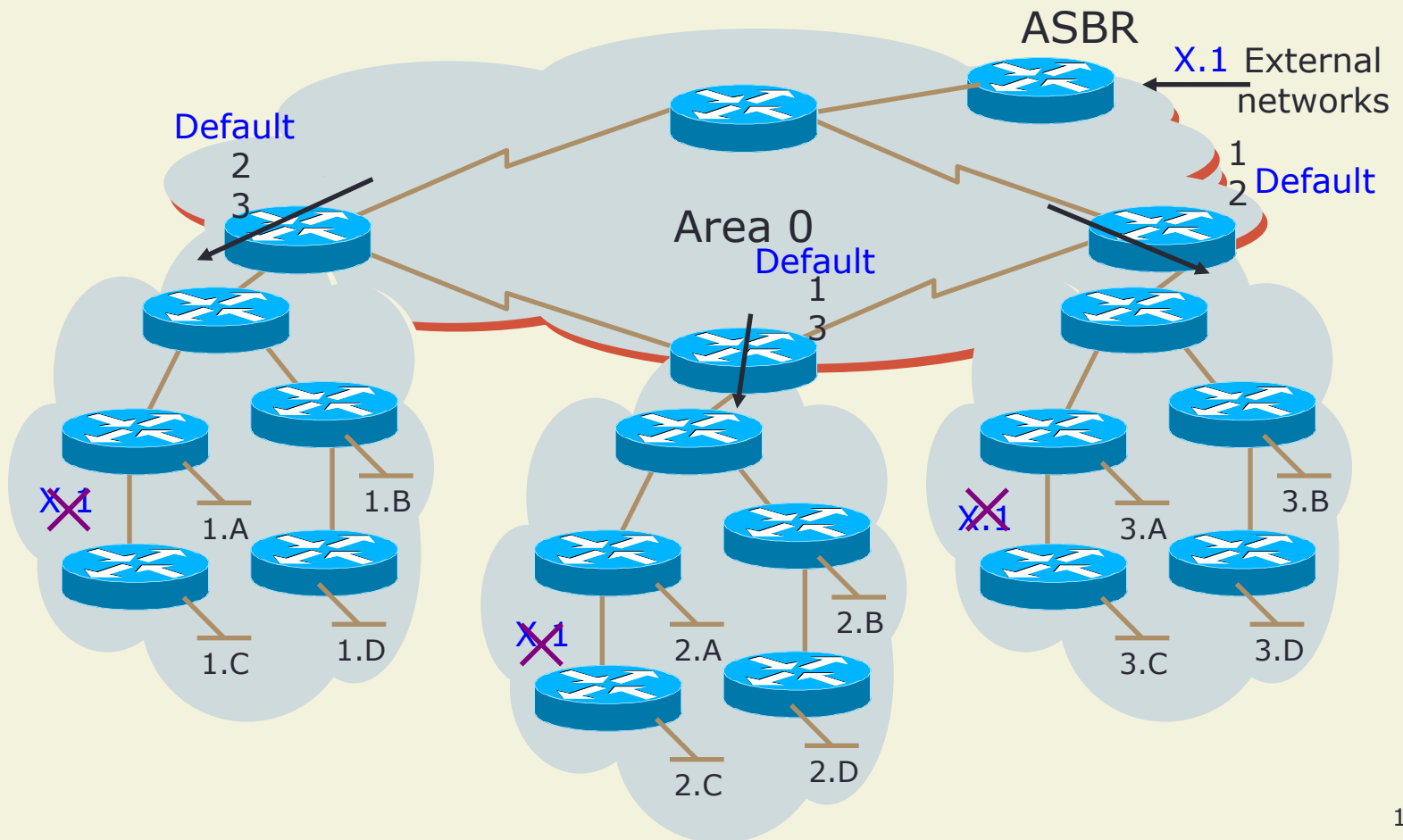
Regular Area (Not a Stub)

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



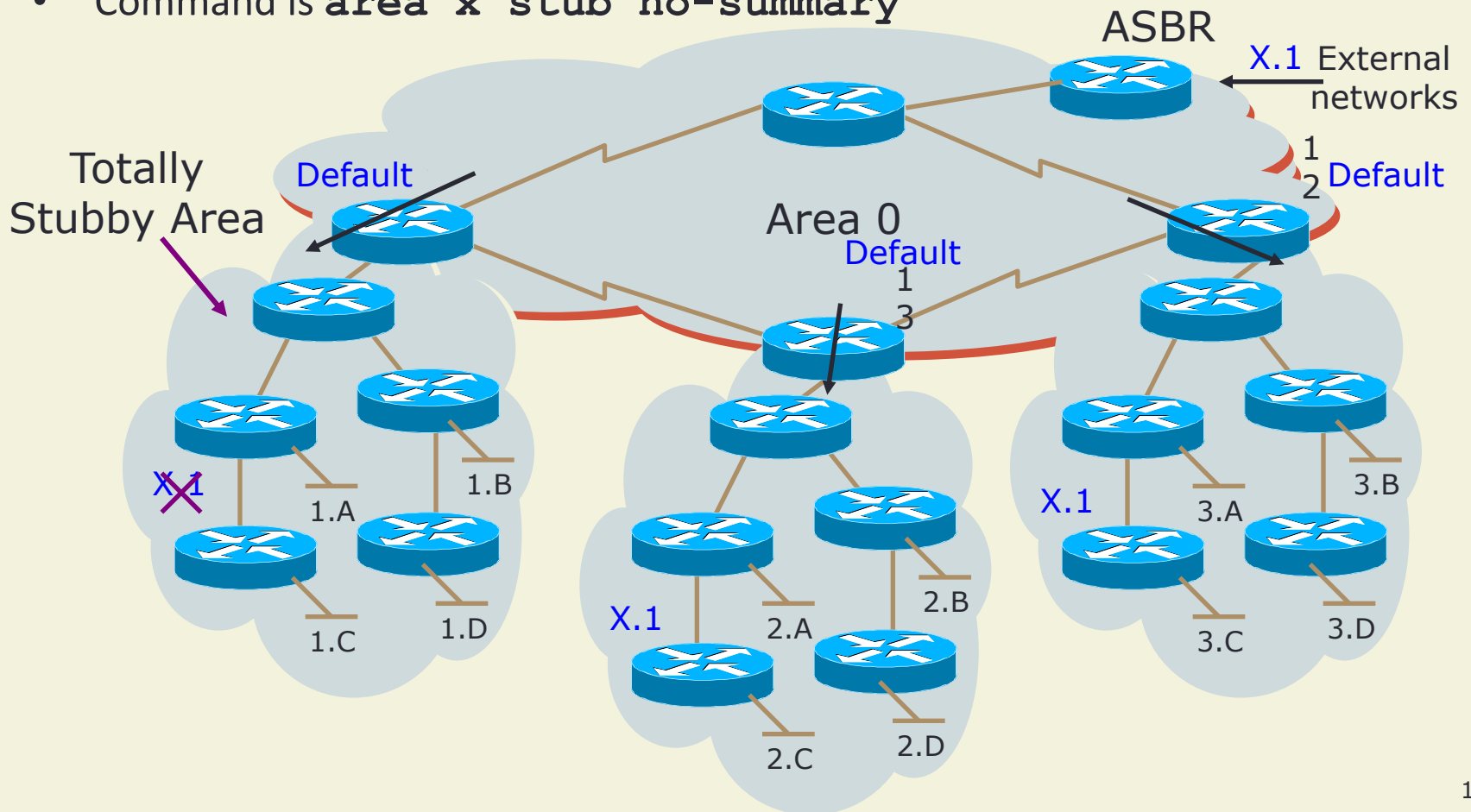
Normal Stub Area

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



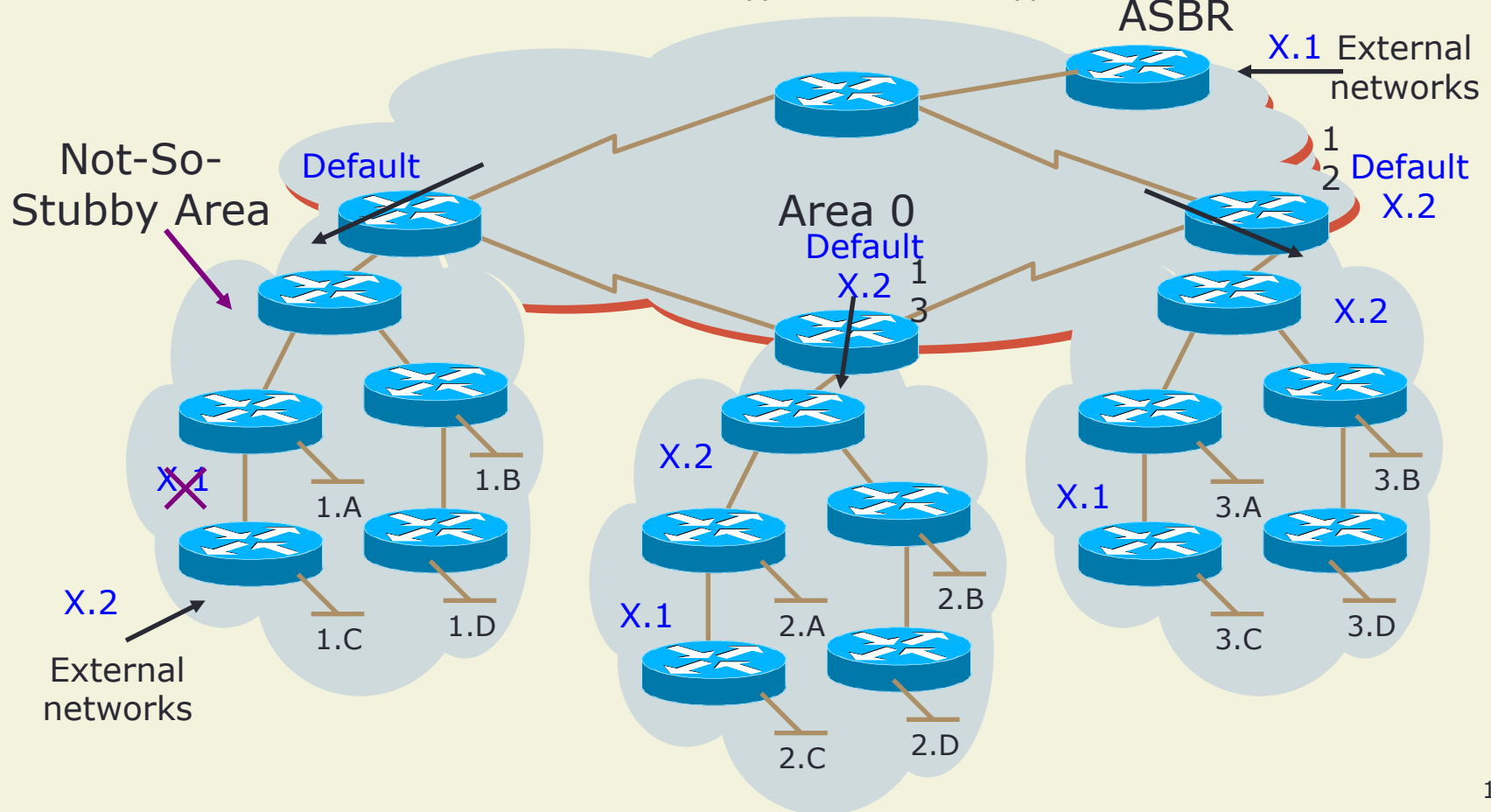
Totally Stubby Area

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



Not-So-Stubby Area

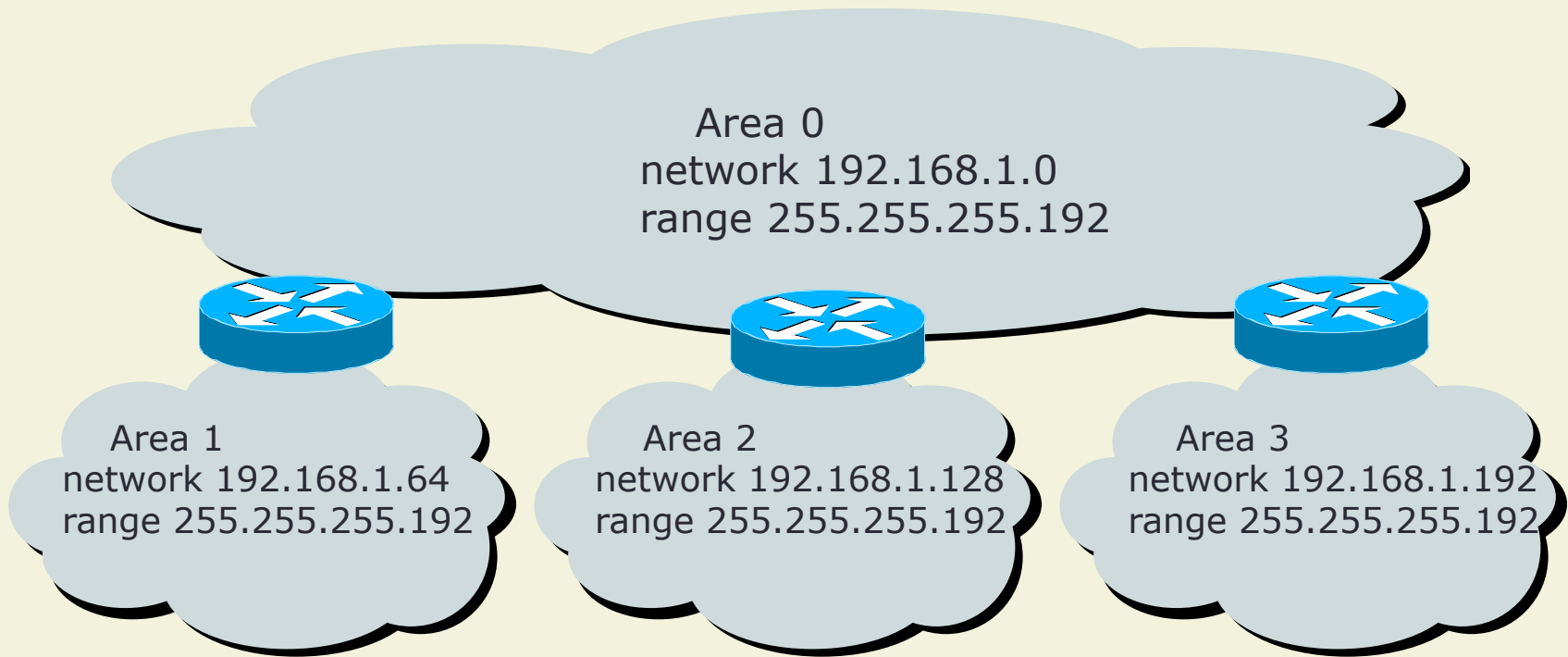
- Capable of importing routes in a limited fashion
- Type-7 LSA's carry external information within an NSSA
- NSSA Border routers translate selected type-7 LSAs into type-5 external network LSAs



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

OSPF for IPv6

ISP Workshops

Recap: OSPFv2

- April 1998 was the most recent revision (RFC 2328)
- OSPF uses a 2-level hierarchical model
- SPF calculation is performed independently for each area
- Typically faster convergence than DVRPs
- Relatively low, steady state bandwidth requirements

OSPFv3 overview

- OSPF for IPv6
- Based on OSPFv2, with enhancements
- Distributes IPv6 unicast prefixes
- Runs directly over IPv6
- Ships-in-the-night with OSPFv2
- OSPFv3 does **not** carry IPv4 prefixes
 - RFC5838 proposes an extension which adds address family support

OSPFv3 / OSPFv2 Similarities

- Basic packet types
 - Hello, DBD, LSR, LSU, LSA
- Mechanisms for neighbor discovery and adjacency formation
- Interface types
 - P2P, P2MP, Broadcast, NBMA, Virtual
- LSA flooding and aging
- Nearly identical LSA types

V2, V3 Differences

OSPFv3 runs on a Link instead of per IP Subnet

- A link by definition is a medium over which two nodes can communicate at link layer
- In IPv6 multiple IP subnet can be assigned to a link and two nodes in different subnet can communicate at link layer therefore OSPFv3 is running per link instead of per IP subnet
- An Interface connect to a link and multiple interface can be connected to a link

V2, V3 Differences (Cont.)

Support of Multiple Instance per Link

- New field (instance) in OSPF packet header allow running multiple instance per link
- Instance ID should match before packet being accepted
- Useful for traffic separation, multiple areas per link and address families (RFC5838)

V2, V3 Differences (Cont.)

Address Semantic Change in LSA

- Router and Network LSA carry only topology information
- Router LSA can be split across multiple LSAs; Link State ID in LSA header is a fragment ID
- Intra area prefix are carried in a new LSA payload called intra-area-prefix-LSAs
- Prefix are carried in payload of inter-area and external LSA

V2, V3 Differences (Cont.)

Generalisation of Flooding Scope

- In OSPFv3 there are three flooding scope for LSAs (link-local scope, area scope, AS scope) and they are coded in LS type explicitly
- In OSPFv2 initially only area and AS wide flooding was defined; later opaque LSAs introduced link local scope as well

V2, V3 Differences (Cont.)

Explicit Handling of Unknown LSA

- The handling of unknown LSA is coded via U-bit in LS type
- When U bit is set, the LSA is flooded with the corresponding flooding scope, as if it was understood
- When U bit is clear, the LSA is flooded with link local scope
- In v2 unknown LSA were discarded

V2, V3 Differences (Cont.)

Authentication is Removed from OSPF

- Authentication in OSPFv3 has been removed
- OSPFv3 relies now on the IPv6 authentication header since OSPFv3 run over IPv6
- Autype and Authentication field in the OSPF packet header therefore have been suppressed

V2, V3 Differences (Cont.)

OSPF Packet format has been changed

- The mask field has been removed from Hello packet
- IPv6 prefix are only present in payload of Link State update packet

V2, V3 Differences (Cont.)

Two New LSAs Have Been Introduced

- Link-LSA has a link local flooding scope and has three purposes:
 - The router link local address
 - List all IPv6 prefixes attached to the link
 - Assert a collection of option bit for the Router-LSA
- Intra-area-prefix-LSA
 - Used to advertise router's IPv6 address within the area

Inter-Area Prefix LSA

- Describes the destination outside the area but still in the AS
- Summary is created for one area, which is flooded out in all other areas
- Originated by an ABR
- Only intra-area routes are advertised into the backbone
- Link State ID simply serves to distinguish inter-area-prefix-LSAs originated by the same router
- Link-local addresses must never be advertised in inter-area-prefix-LSAs

LSA Types

	LSA Function Code	LSA Type
Router-LSA	1	0x2001
Network-LSA	2	0x2002
Inter-Area-Prefix-LSA	3	0x2003
Inter-Area-Router-LSA	4	0x2004
AS-External-LSA	5	0x4005
Group-membership-LSA	6	0x2006
Type-7-LSA	7	0x2007
Link-LSA	8	0x2008
Intra-Area-Prefix-LSA ^{NEW}	9	0x2009

Configuring OSPFv3 in Cisco IOS

- Similar to OSPFv2
 - Prefixing existing Interface and Exec mode commands with “**ipv6**”
- Interfaces configured directly
 - Replaces **network** command
 - (Also available in OSPFv2 from IOS 12.4)
- “Native” IPv6 router mode
 - Not a sub-mode of **router ospf**

Configuring OSPFv3

- Setting up the OSPFv3 process:
`[no] ipv6 router ospf <process ID>`
- Applying the OSPFv3 process to an interface:
`interface <router-int-name>`
`[no] ipv6 ospf <process ID> area <area ID>`
- Configuring summarisation:
`ipv6 router ospf <process ID>`
`[no] area <area ID> range <prefix>/<length>`

OSPFv3 exec mode commands

- Exec mode commands:

```
show ipv6 ospf [<process ID>]
```

```
clear ipv6 ospf [<process ID>]
```

- Showing new LSA:

```
show ipv6 ospf [<process ID>] database link
```

```
show ipv6 ospf [<process ID>] database prefix
```

OSPFv3 Authentication

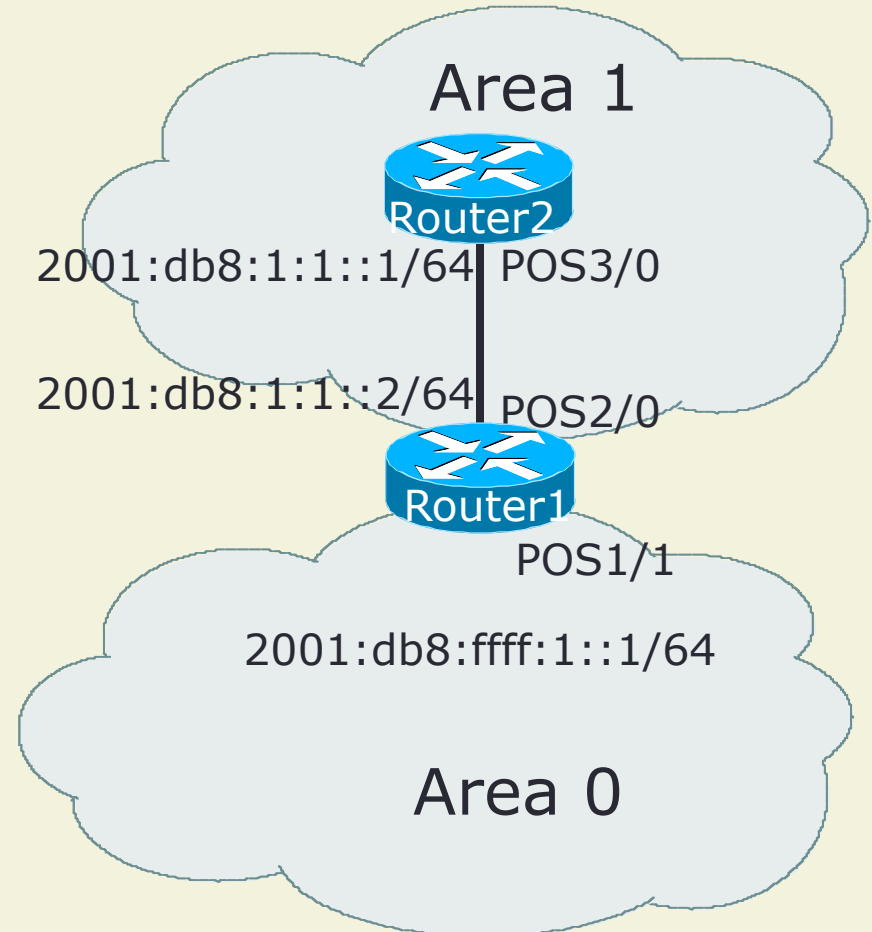
- Configuring authentication per area:
 - SPI value has to be unique per area:
`ipv6 router ospf <process ID>`
`area 0 authentication ipsec spi 256 md5 <passwd>`
- Disabling authentication on a specific link when area authentication is activated:
`interface fastethernet 0/0`
`ipv6 ospf authentication null`
- Configuring authentication per interface:
 - SPI value has to be unique per link:
`interface fastethernet 0/0`
`ipv6 ospf authentication ipsec spi 256 md5 <passwd>`

OSPFv3 Debug Commands

- Adjacency is not appearing
 - `[no] debug ipv6 ospf adj`
 - `[no] debug ipv6 ospf hello`
- SPF is running constantly
 - `[no] debug ipv6 ospf spf`
 - `[no] debug ipv6 ospf flooding`
 - `[no] debug ipv6 ospf events`
 - `[no] debug ipv6 ospf lsa-generation`
 - `[no] debug ipv6 ospf database-timer`
- General purpose
 - `[no] debug ipv6 ospf packets`
 - `[no] debug ipv6 ospf retransmission`
 - `[no] debug ipv6 ospf tree`

OSPFv3 Configuration Example

```
Router1#  
interface POS1/1  
  ipv6 address 2001:db8:ffff:1::1/64  
  ipv6 ospf 100 area 0  
!  
interface POS2/0  
  ipv6 address 2001:db8:1:1::2/64  
  ipv6 ospf 100 area 1  
!  
  ipv6 router ospf 100  
    log-adjacency-changes  
!  
  
Router2#  
interface POS3/0  
  ipv6 address 2001:db8:1:1::1/64  
  ipv6 ospf 100 area 1  
!  
  ipv6 router ospf 100  
    log-adjacency-changes
```



OSPFv3 Interface Status

```
Router2#sh ipv6 ospf int pos 3/0
POS3/0 is up, line protocol is up
  Link Local Address FE80::290:86FF:FE5D:A000, Interface ID 7
  Area 1, Process ID 100, Instance ID 0, Router ID 10.1.1.4
  Network Type POINT_TO_POINT, Cost: 1
  Transmit Delay is 1 sec, State POINT_TO_POINT,
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:02
  Index 1/1/1, flood queue length 0
  Next 0x0(0)/0x0(0)/0x0(0)
  Last flood scan length is 3, maximum is 3
  Last flood scan time is 0 msec, maximum is 0 msec
  Neighbor Count is 1, Adjacent neighbor count is 1
    Adjacent with neighbor 10.1.1.3
  Suppress hello for 0 neighbor(s)
```

OSPFv3 Neighbour Status

```
Router2#sh ipv6 ospf neighbor detail
```

```
Neighbor 10.1.1.3
```

```
  In the area 1 via interface POS3/0
```

```
Neighbor: interface-id 8, link-local address FE80::2D0:FFFF:FE60:DFFF
```

```
Neighbor priority is 1, State is FULL, 12 state changes
```

```
Options is 0x630C34B9
```

```
Dead timer due in 00:00:33
```

```
Neighbor is up for 00:49:32
```

```
Index 1/1/1, retransmission queue length 0, number of retransmission 1
```

```
First 0x0(0)/0x0(0)/0x0(0) Next 0x0(0)/0x0(0)/0x0(0)
```

```
Last retransmission scan length is 2, maximum is 2
```

```
Last retransmission scan time is 0 msec, maximum is 0 msec
```

OSPFv3 entries in Routing Table

```
Router2#sh ipv6 route
IPv6 Routing Table - 5 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
OI 2001:db8:FFFF:1::/64 [110/2]
    via FE80::2D0:FFFF:FE60:DFFF, POS3/0
C 2001:db8:1:1::/64 [0/0]
  via ::, POS3/0
L 2001:db8:1:1::1/128 [0/0]
  via ::, POS3/0
L FE80::/10 [0/0]
  via ::, Null0
L FF00::/8 [0/0]
  via ::, Null0
```

OSPFv3 link troubleshooting

- Next router address in OSPFv3 is a link-local address

```
OI 2001:db8:FFFF:1::/64 [110/2]
```

```
via FE80::2D0:FFFF:FE60:DFFF, POS3/0
```

- How to troubleshoot??

- SSH to neighbouring router needs extended SSH command, for example:

```
ssh FE80::2D0:FFFF:FE60:DFFF /source-int POS3/0
```

- Source interface has to be specified – a router with multiple interfaces has no idea which interface the remote link local address is attached to

Cisco IOS OSPFv3 Database Display

```
Router2# show ipv6 ospf database
```

```
OSPF Router with ID (3.3.3.3) (Process ID 1)
```

Router Link States (Area 0)

Link ID	ADV Router	Age	Seq#	Checksum	Link count
0	1.1.1.1	2009	0x8000000A	0x2DB1	1
0	3.3.3.3	501	0x80000007	0xF3E6	1

Net Link States (Area 0)

Link ID	ADV Router	Age	Seq#	Checksum
7	1.1.1.1	480	0x80000006	0x3BAD

Inter Area Prefix Link States (Area 0)

ADV Router	Age	Seq#	Prefix
1.1.1.1	1761	0x80000005	2001:db8:2:2::/64
1.1.1.1	982	0x80000005	2001:db8:2:4::2/128

Link (Type-8) Link States (Area 0)

Link ID	ADV Router	Age	Seq#	Checksum	Interface
11	3.3.3.3	245	0x80000006	0xF3DC	Lo0
7	1.1.1.1	236	0x80000008	0x68F	Fa2/0
7	3.3.3.3	501	0x80000008	0xE7BC	Fa2/0

Intra Area Prefix Link States (Area 0)

Link ID	ADV Router	Age	Seq#	Checksum	Ref lstype
0	1.1.1.1	480	0x80000008	0xD670	0x2001
107	1.1.1.1	236	0x80000008	0xC05F	0x2002
0	3.3.3.3	245	0x80000006	0x3FF7	0x2001

Cisco IOS OSPFv3 Detailed LSA Display

```
show ipv6 ospf 1 database inter-area prefix
```

```
LS age: 1714
LS Type: Inter Area Prefix Links
Link State ID: 0
Advertising Router: 1.1.1.1
LS Seq Number: 80000006
Checksum: 0x25A0
Length: 36
Metric: 1
Prefix Address: 2001:db8:2:2::
Prefix Length: 64, Options: None
```

```
show ipv6 ospf 1 database link
```

```
LS age: 283
Options: (IPv6 Router, Transit Router, E-Bit, No Type 7-to-5, DC)
LS Type: Link-LSA (Interface: Loopback0)
Link State ID: 11 (Interface ID)
Advertising Router: 3.3.3.3
LS Seq Number: 80000007
Checksum: 0xF1DD
Length: 60
Router Priority: 1
Link Local Address: FE80::205:5FFF:FEAC:1808
Number of Prefixes: 2
Prefix Address: 2001:db8:1:3::
Prefix Length: 64, Options: None
Prefix Address: 2001:db8:1:3::
Prefix Length: 64, Options: None
```

Conclusion

- Based on existing OSPFv2 implementation
- Similar CLI and functionality

Introduction to ISIS

ISP Workshops

IS-IS Standards History

- ISO 10589 specifies OSI IS-IS routing protocol for CLNS traffic
 - A Link State protocol with a 2 level hierarchical architecture
 - Type/Length/Value (TLV) options to enhance the protocol
- RFC 1195 added IP support
 - Integrated IS-IS
 - I/IS-IS runs on top of the Data Link Layer

IS-IS Standards History

- RFC5308 adds IPv6 address family support to IS-IS
- RFC5120 defines Multi-Topology concept for IS-IS
 - Permits IPv4 and IPv6 topologies which are not identical
 - (Required for an incremental roll-out of IPv6 on existing IPv4 infrastructure)

ISIS Levels

- ISIS has a 2 layer hierarchy
 - Level-2 (the backbone)
 - Level-1 (the areas)
- A router can be
 - Level-1 (L1) router
 - Level-2 (L2) router
 - Level-1-2 (L1L2) router

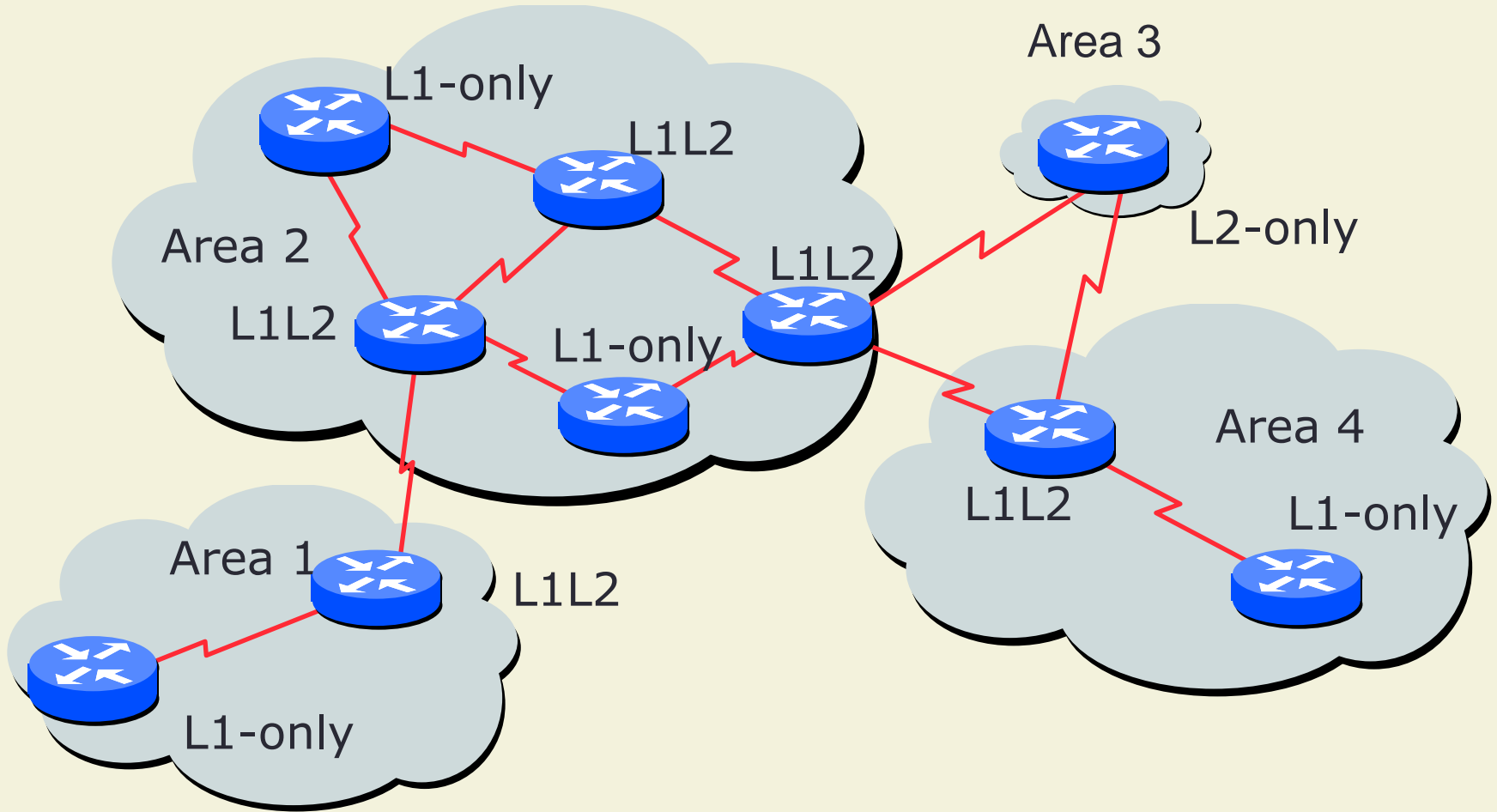
ISIS Levels

- Level-1 router
 - Has neighbours only on the same area
 - Has a level-1 LSDB with all routing information for the area
- Level-2 router
 - May have neighbours in the same or other areas
 - Has a Level-2 LSDB with all routing information about inter-area
- Level-1-2 router
 - May have neighbours on any area.
 - Has two separate LSDBs: level-1 LSDB & level-2 LSDB

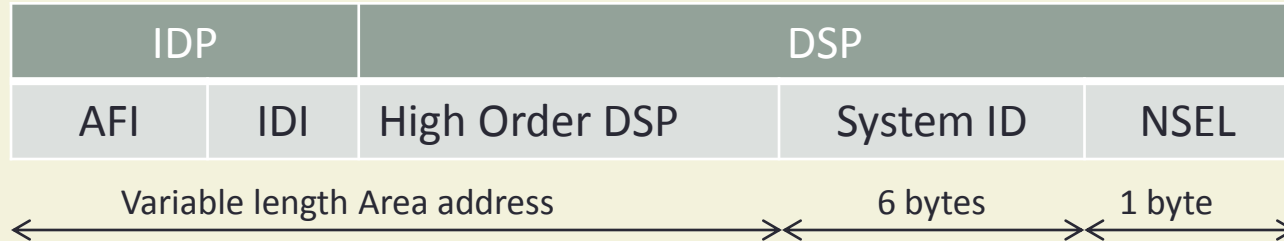
Backbone & Areas

- ISIS does not have a backbone area as such (like OSPF)
- Instead the backbone is the contiguous collection of Level-2 capable routers
- ISIS area borders are on links, not routers
- Each router is identified with a unique Network Entity Title (NET)
 - NET is a Network Service Access Point (NSAP) where the n-selector is 0
 - (Compare with each router having a unique Router-ID with IP routing protocols)

Example: L1, L2, and L1L2 Routers



NSAP and Addressing



- NSAP: Network Service Access Point
 - Total length between 8 and 20 bytes
 - Area Address: variable length field (up to 13 bytes)
 - System ID: defines an ES or IS in an area.
 - NSEL: N-selector; identifies a network service user (transport entity or the IS network entity itself)
- NET: the address of the network entity itself

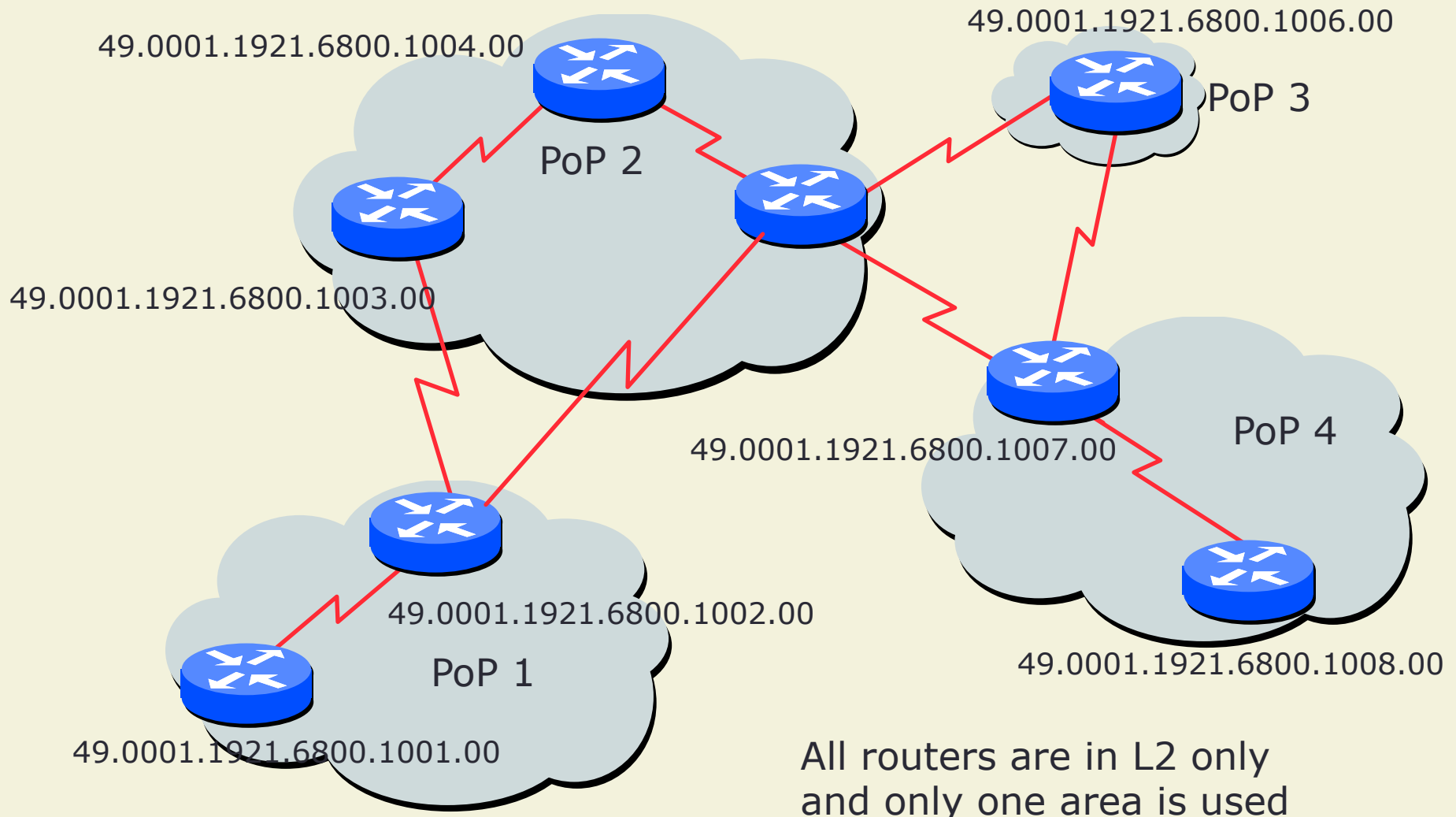
Addressing Common Practices

- ISPs typically choose NSAP addresses thus:
 - First 8 bits – pick a number (usually 49)
 - Next 16 bits – area
 - Next 48 bits – router loopback address
 - Final 8 bits – zero
- Example:
 - NSAP: 49.0001.1921.6800.1001.00
 - Router: 192.168.1.1 (loopback) in Area 1

Addressing & Design Practices

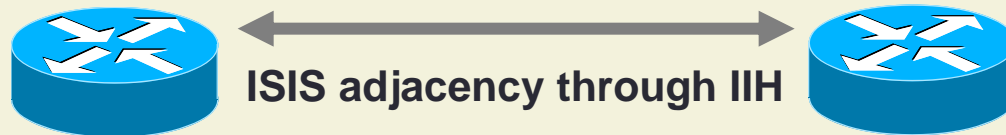
- ISPs usually only use one area
 - Multiple areas only come into consideration once the network is several hundred routers big
- NET begins with 49
 - “Private” address range
- All routers are in L2 only
 - Note that Cisco IOS default is L1L2
 - Set L2 under ISIS generic configuration (can also be done per interface)

Typical ISP Design



Adjacencies

- Hello Protocol Data Units (PDUs) are exchanged between routers to form adjacencies



- Area addresses are exchanged in IIH PDUs
 - Intermediate-System to Intermediate System Hello PDUs
 - (PDU is ISIS equivalent of a packet)

Link State PDU (LSP)

- Each router creates an LSP and floods it to neighbours
- A level-1 router will create level-1 LSP(s)
- A level-2 router will create level-2 LSP(s)
- A level-1-2 router will create
 - level-1 LSP(s) and
 - level-2 LSP(s)

The ISIS LSP

- LSPs have a Fixed Header and TLV coded contents
- The LSP header contains
 - LSP-id (Sequence number)
 - Remaining Lifetime (Checksum)
 - Type of LSP (level-1, level-2)
 - Attached bit (Overload bit)
- The LSP contents are coded as TLV (Type, Length, Value)
 - Area addresses
 - IS neighbours
 - Authentication Information

Link State Database Content

- Each router maintains a separate LSDB for level-1 and level-2 LSPs
- The LSDB contains:
 - LSP headers and contents
 - SRM bits: set per interface when router has to flood this LSP
 - SSN bits: set per interface when router has to send a PSNP for this LSP

Flooding of LSPs

- New LSPs are flooded to all neighbors
- All routers get all LSPs
- Each LSP has a sequence number
- There are 2 kinds of flooding:
 - Flooding on a p2p link
 - Flooding on LAN

Flooding on a p2p link

- Once the adjacency is established both routers send CSNP packet
- Missing LSPs are sent by both routers if not present in the received CSNP
- Missing LSPs may be requested through PSNP

Flooding on a LAN

- Each LAN has a Designated Router (DIS)
- The DIS has two tasks
 - Conducting the flooding over the LAN
 - Creating and updating a special LSP describing the LAN topology (Pseudonode LSP)
- DIS election is based on priority
 - Best practice is to select two routers and give them higher priority – then in case of failure one provides deterministic backup for the other
 - Tie break is by the highest MAC address

Flooding on a LAN

- DIS conducts the flooding over the LAN
- DIS multicasts CSNP every 10 seconds
- All routers on the LAN check the CSNP against their own LSDB (and may ask specific re-transmissions with PSNPs)

Complete Sequence Number PDU

- Describes all LSPs in your LSDB (in range)
- If the LSDB is large, multiple CSNPs are sent
- Used on 2 occasions:
 - Periodic multicast by DIS (every 10 seconds) to synchronise the LSDB over LAN subnets
 - On p2p links when link comes up

Partial Sequence Number PDUs

- PSNPs Exchanged on p2p links (ACKs)
- Two functions
 - Acknowledge receipt of an LSP
 - Request transmission of latest LSP
- PSNPs describe LSPs by its header
 - LSP identifier
 - Sequence number
 - Remaining lifetime
 - LSP checksum

Network Design Issues

- As in all IP network designs, the key issue is the addressing lay-out
- ISIS supports a large number of routers in a single area
- When network is so large requiring the use of areas, use summary-addresses
- >400 routers in the backbone is quite doable

Network Design Issues

- Link cost
 - Default on all interfaces is 10
 - (Compare with OSPF which sets cost according to link bandwidth)
 - Manually configured according to routing strategy
- Summary address cost
 - Equal to the best more specific cost
 - Plus cost to reach neighbour of best specific
- Backbone has to be contiguous
 - Ensure continuity by redundancy
- Area partitioning
 - Design so that backbone can **NOT** be partitioned

Scaling Issues

- Areas vs. single area
 - Use areas where
 - sub-optimal routing is not an issue
 - areas with one single exit point
- Start with L2-only everywhere
 - Future implementation of level-1 areas will be easier
 - Backbone continuity is ensured from start

ISIS for IPv6

ISP Workshops

Topics Covered

- IS-IS standardisation
- IS-IS for IPv6
- Multi-Topology IS-IS

ISIS Standards History

- ISO 10589 specifies the OSI IS-IS routing protocol for CLNS traffic
- RFC 1195 added IPv4 support
 - Also known as Integrated IS-IS (I/IS-IS)
 - I/IS-IS runs on top of the Data Link Layer
- RFC5308 adds IPv6 address family support
- RFC5120 defines Multi-Topology concept
 - Permits IPv4 and IPv6 topologies which are not identical
 - Permits roll out of IPv6 without impacting IPv4 operations

Integrated IS-IS for IPv6 Overview

- 2 Type/Length/Values (TLV) added to support IPv6 routing
- IPv6 Reachability TLV (0xEC)
 - Describes network reachability such as IPv6 routing prefix, metric information and some option bits
- IPv6 Interface Address TLV (0xE8)
 - Contains a 128 bit address
 - For Hello PDUs, must contain the link-local address (FE80::/10)
 - For LSP, must only contain the non link-local address

Integrated IS-IS for IPv6 Overview

- A new Network Layer Protocol Identifier (NLPID) is defined
 - Allowing IS-IS routers with IPv6 support to advertise IPv6 prefix payload using 0x8E value
 - IPv4 and OSI uses different values

ISIS for IPv6

IS-IS for IPv6

- A single SPF runs per level for OSI, IPv4 and IPv6
 - All routers in an area must run the same set of protocols [IPv4-only, IPv6-only, IPv4-IPv6]
 - L2 routers don't have to be configured similarly but no routing hole must exist

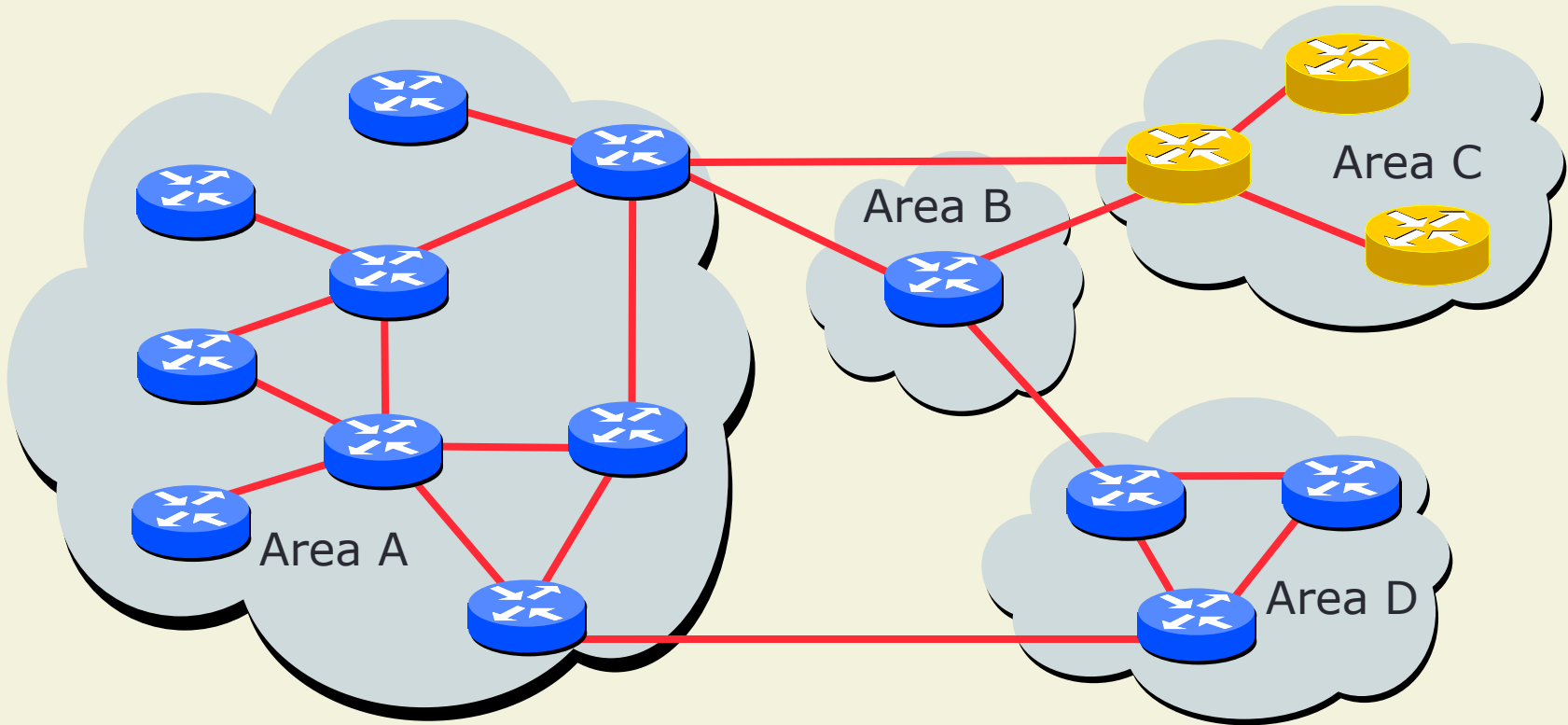
Simple SPF rules


- If IS-IS is used for both IPv4 and IPv6 in an area, both protocols must support the same topology within this area:
 - “no adjacency-check” between L2 routers over-rides this, **but must be used with caution**
- All interfaces configured with IS-ISv6 must support IPv6
- All interfaces configured with IS-IS for both protocols must support both of them
 - IPv6 configured tunnel won't work, GRE should be used in this configuration
- Otherwise, consider Multi-Topology IS-IS (separate SPF)


Single SPF IS-IS for IPv6 restrictions

- IS-IS for IPv6 uses the same SPF for both IPv4 and IPv6.
- Therefore:
 - Not suitable for an existing IPv4 IS-IS network where operator wants to turn on scattered IPv6 support
 - If using IS-IS for both IPv4 and IPv6 then the IPv4 and IPv6 topologies MUST match exactly. Cannot run IS-IS IPv6 on some interfaces, IS-IS IPv4 on others.
 - Will only form adjacencies with similarly-configured routers.
 - For example, an IS-IS IPv6-only router will not form an adjacency with an IS-IS IPv4/IPv6 router. (Exception is over L2-only interface)
 - Cannot join two IPv6 areas via an IPv4-only area. L2 adjacencies will form OK but IPv6 traffic will black-hole in the IPv4 area.

IS-IS Hierarchy & IPv6 example



 IPv4-IPv6 enabled router

 IPv4-only enabled router

Configuring IS-IS for IPv6

- CLI is familiar:
- IPv6 address family mode enables features specific to IPv6:

```
router isis as64512
  net 49.0001.0001.0001.00
  set-overload-bit on-startup wait-for-bgp
!
  address-family ipv6
    set-overload-bit on-startup wait-for-bgp
!
```

- **Configure IS-IS for IPv6 on interfaces**
 - Interface must be IPv6 enabled, eg. IPv6 address set

IS-IS for IPv6

Specific Attributes (1)

- Entering address-family sub-mode
`[no] address-family ipv6`
- IPv6 address-family sub-mode.
`[no] adjacency-check`
 - Enables or disables adjacency IPv6 protocol-support checks. If checking is enabled (default condition when IS-IS IPv6 is configured) then the router will not form an adjacency with a neighbor not supporting IS-IS IPv6.
`[no] distance <1-254>`
 - Sets the administrative distance of IS-IS IPv6. Note that the administrative distance is applied to routes in the IPv6 routing table only.

IS-IS for IPv6

Specific Attributes (2)

[no] maximum-paths <1-4>

- Sets the maximum number of paths allowed for a route learnt via IS-IS IPv6. Note that this applies to the IPv6 routing table only.

[no] default-information originate [route-map <name>]

- Configures origination of the IPv6 default route (::) by IS-IS. Used in the same manner as the existing IPv4 default-information command.

[no] summary-prefix <prefix> [level-1|level-2|level-1-2]

- Configures IPv6 summary prefixes. Command is used in same manner as the existing IPv4 summary-prefix command.

[no] set-overload-bit on-startup wait-for-bgp

- Set overload bit so that the router does not enter transit path until iBGP is running

IS-IS for IPv6

Specific Attributes (3)

```
[no] redistribute <protocol> [metric <value>]  
    [metric-type {internal|external}] [level-  
    1|level-1-2|level-2] [route-map <name>]
```

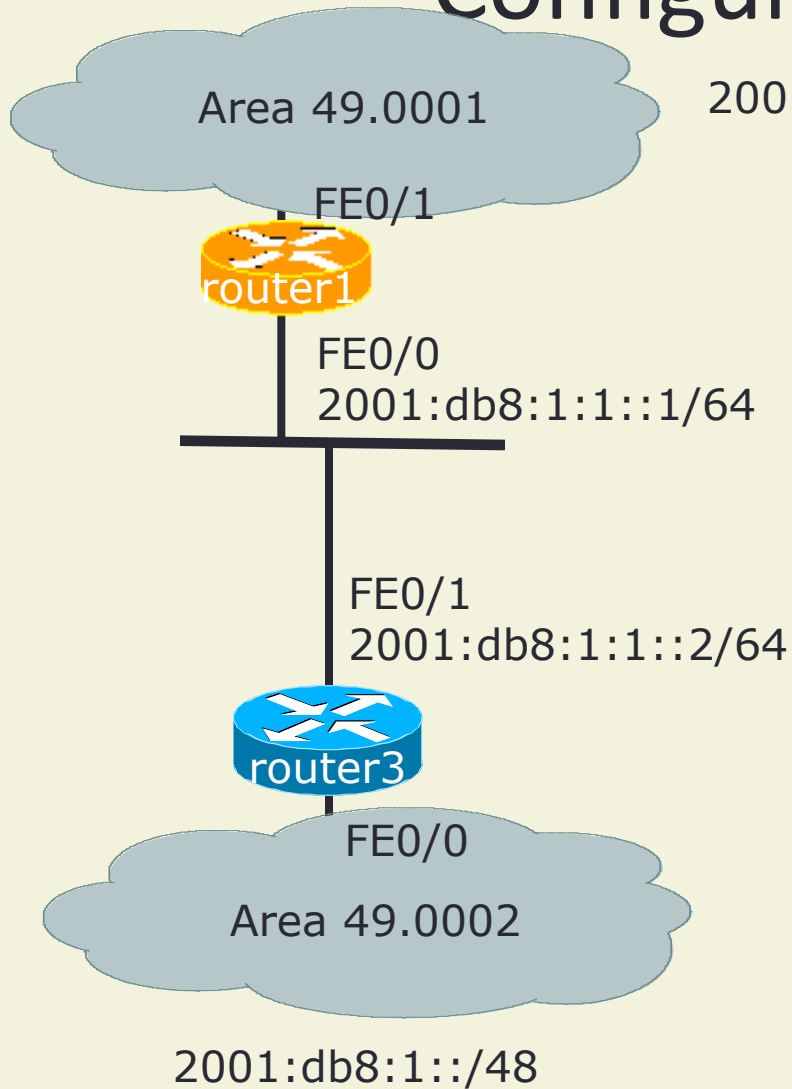
- Configures redistribution of routes learnt from other IPv6 sources into IS-IS. Command is used in same manner as existing IPv4 redistribute command.

```
[no] redistribute isis {level-1|level-2} into  
    {level-1|level-2} distribute-list <prefix-list-  
    name>
```

- Configures IS-IS inter-area redistribution of IPv6 routes. Command is used in same manner as existing IPv4 redistribute isis command.
- Leaving address-family sub-mode
`exit-address-family`
- Showing the I/IS-ISv6 configuration
`show ipv6 protocols [summary]`

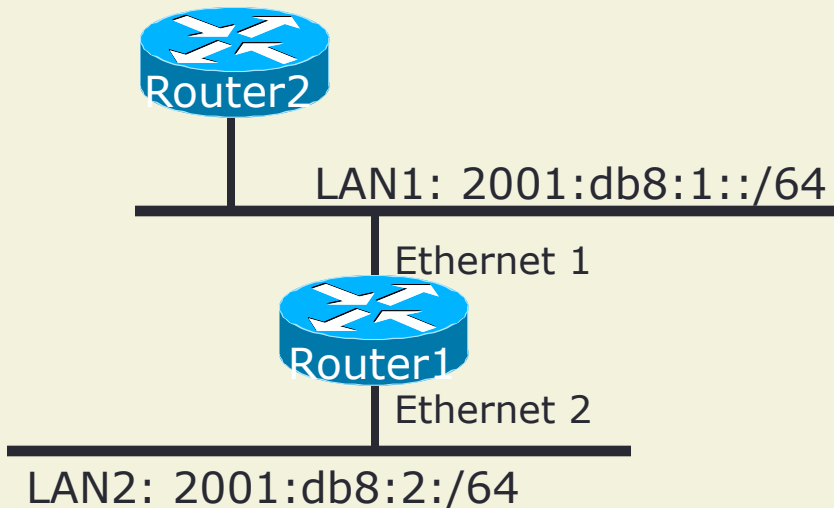
ISIS for IPv6

Configuration Example



```
Router1#  
interface fastethernet0/0  
  ipv6 address 2001:db8:1:1::1/64  
  ipv6 router isis  
  isis circuit-type level-2-only  
  
router isis  
  net 49.0001.1921.6801.0001.00  
  address-family ipv6  
  redistribute static  
  exit-address-family
```

IS-IS dual stack configuration



Dual IPv4/IPv6 configuration.
Redistributing both IPv6 static routes
and IPv4 static routes.

```
Router1#  
interface ethernet 1  
 ip address 10.1.1.1 255.255.255.0  
 ipv6 address 2001:db8:1::1/64  
 ip router isis  
 ipv6 router isis  
  
interface ethernet 2  
 ip address 10.2.1.1 255.255.255.0  
 ipv6 address 2001:db8:2::1/64  
 ip router isis  
 ipv6 router isis  
  
router isis  
 net 42.0001.0000.0000.072c.00  
 redistribute static  
!  
 address-family ipv6  
  redistribute static  
 exit-address-family
```

ISIS Display (1)

```
router1#sh ipv6 route isis
IPv6 Routing Table - default - 46 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
       B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       D - EIGRP, EX - EIGRP external, ND - Neighbor Discovery, l - LISP
       O - OSPF Intra, OI - OSPF Inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
I2 2001:DB8::2/128 [115/2]
    via FE80::C801:3CFF:FE4F:8, FastEthernet0/0
I2 2001:DB8::3/128 [115/20]
    via FE80::C802:3CFF:FE4F:0, Serial1/0
I2 2001:DB8::4/128 [115/22]
    via FE80::C801:3CFF:FE4F:8, FastEthernet0/0
I2 2001:DB8::5/128 [115/40]
    via FE80::C802:3CFF:FE4F:0, Serial1/0
I2 2001:DB8::6/128 [115/42]
    via FE80::C801:3CFF:FE4F:8, FastEthernet0/0
    via FE80::C802:3CFF:FE4F:0, Serial1/0
```

ISIS Display (2)

```
router1#sh isis ipv6 rib
IS-IS IPv6 process workshop, local RIB
* 2001:DB8::2/128
    via FE80::C801:3CFF:FE4F:8/FastEthernet0/0, type L2 metric 2 LSP [7/8]
* 2001:DB8::3/128
    via FE80::C802:3CFF:FE4F:0/Serial1/0, type L2 metric 20 LSP [2/8]
* 2001:DB8::4/128
    via FE80::C801:3CFF:FE4F:8/FastEthernet0/0, type L2 metric 22 LSP [8/8]
* 2001:DB8::5/128
    via FE80::C802:3CFF:FE4F:0/Serial1/0, type L2 metric 40 LSP [4/8]
* 2001:DB8::6/128
    via FE80::C801:3CFF:FE4F:8/FastEthernet0/0, type L2 metric 42 LSP [5/8]
    via FE80::C802:3CFF:FE4F:0/Serial1/0, type L2 metric 42 LSP [5/8]
* 2001:DB8::7/128
    via FE80::C802:3CFF:FE4F:0/Serial1/0, type L2 metric 60 LSP [A/8]
* 2001:DB8::8/128
    via FE80::C801:3CFF:FE4F:8/FastEthernet0/0, type L2 metric 62 LSP [6/8]
    via FE80::C802:3CFF:FE4F:0/Serial1/0, type L2 metric 62 LSP [6/8]
...
```

ISIS Display (3)

```
Router2#sh clns is-neighbors detail
```

```
Tag Workshop:
```

System Id	Interface	State	Type	Priority	Circuit Id	Format
router1	Fa0/0	Up	L2	64	Router2.01	Phase V
Area Address(es): 49.0001						
IP Address(es): 10.0.15.1*						
IPv6 Address(es): FE80::C800:3CFF:FE4F:8						
Uptime: 00:07:31						
NSF capable						
Interface name: FastEthernet0/0						
Router4	Se1/0	Up	L2	0	00	Phase V
Area Address(es): 49.0001						
IP Address(es): 10.0.15.18*						
IPv6 Address(es): FE80::C803:3CFF:FE4F:0						
Uptime: 00:07:32						
NSF capable						
Interface name: Serial1/0						
Router14	Fa0/1	Up	L2	64	Router14.02	Phase V
Area Address(es): 49.0001						
IP Address(es): 10.0.15.26*						
IPv6 Address(es): FE80::C80D:3CFF:FE50:6						
Uptime: 00:08:40						
NSF capable						
Interface name: FastEthernet0/1						

Multi-topology ISIS

Multi-Topology IS-IS extensions

- Multi-Topology is used by ISPs who are deploying IPv6 on an existing IPv4 infrastructure:
 - Running single topology ISIS means that enabling ISIS IPv6 on a point to point link must be done simultaneously at both ends
 - Otherwise the adjacency will go down, leading to possible breakage in the network
 - Adding new routers on a broadcast media in a single topology ISIS is very tricky
 - ISIS for IPv6 must be enabled on all devices on the broadcast media at the same time
 - Otherwise breakage in the network could occur due to adjacencies going down

Multi-Topology IS-IS extensions

- IS-IS for IPv6 assumes that the IPv6 topology is the same as the IPv4 topology
 - Single SPF running, multiple address families
 - Some networks may be like this, but some others may not be
- Multi-Topology IS-IS solves this problem
 - New TLV attributes introduced
 - New Multi-Topology ID #2 for IPv6 Routing Topology
 - Two topologies maintained:
 - ISO/IPv4 Routing Topology
 - IPv6 Routing Topology

Multi-Topology IS-IS Restrictions

- This feature is not compatible with the previous single SPF model
 - New TLV are used to transmit and advertise IPv6 capabilities
 - All routers that run IS-IS for IPv6 need to enable multi-topology within the network
 - A transition mode is provided for existing IS-IS IPv6 network to migrate to Multi-Topology IS-IS IPv6

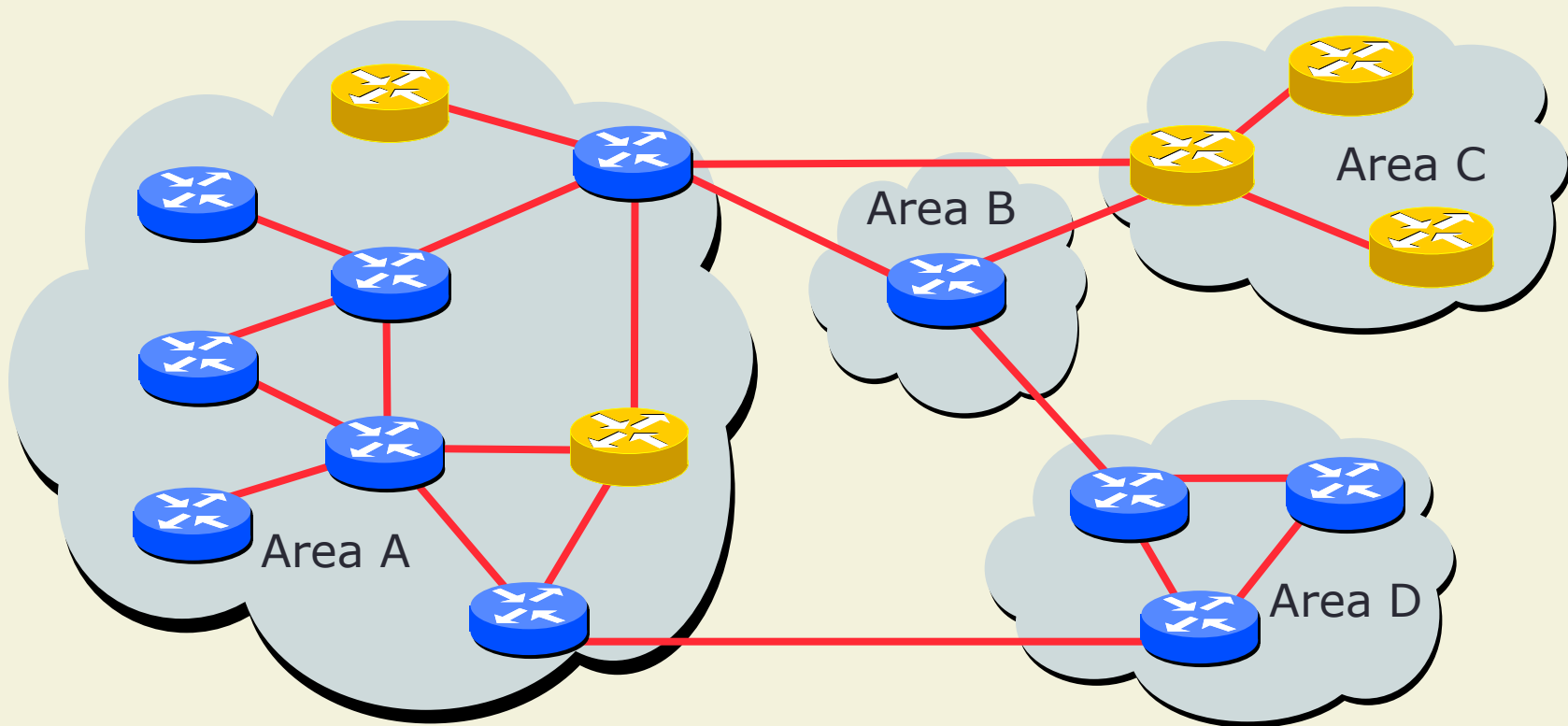
Multi-Topology IS-IS Restrictions

- IPv4 or IPv6 or IPv4/IPv6 may be configured on the interface for either level-1, level-2 or level-1-2
- If IPv4 and IPv6 are configured on the same interface, they must be running the same IS-IS level
 - IPv4 cannot be configured to run on ISIS level-1 only on an interface while IPv6 is configured to run ISIS level-2 only on the same interface.

Multi-Topology IS-IS Restrictions

- All routers on a LAN or point-to-point link must have at least one common supported topology (IPv4 or IPv6) when operating in Multi-Topology IS-IS mode
 - N.B. a router that is not operating in Multi-Topology IS-IS IPv6 mode cannot form adjacency with Multi-Topology IS-IS IPv6 router, even though IPv6 is the common supported topology. However, if IPv4 is the common supported topology between those two routers, adjacency should be formed.
- Wide metrics are required to be enabled globally within the Autonomous System
 - (Default for most ISPs these days anyway)

Multi-Topology IS-IS example



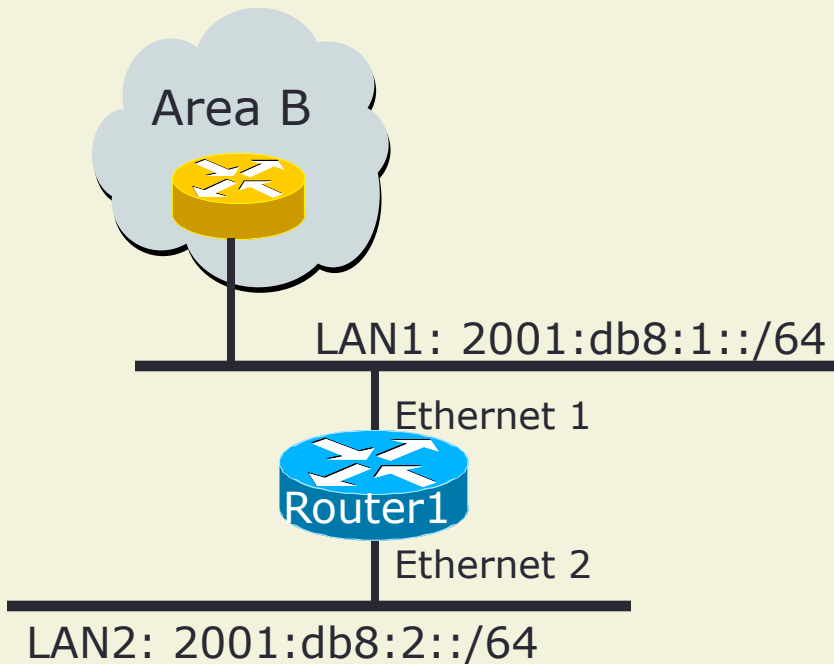
IPv4-IPv6 enabled router



IPv4-only enabled router

The Multi-Topology software will create two topologies inside Area:
IPv4 and IPv6.
IPv4-only routers will be excluded from the IPv6 topology

Multi-Topology ISIS Configuration example



- The optional keyword **transition** may be used for transitioning existing IS-IS IPv6 single SPF mode to MT IS-IS
- Wide metric is mandated for Multi-Topology to work

```
Router1#  
interface Ethernet 1  
  ip address 10.1.1.1 255.255.255.0  
  ipv6 address 2001:db8:1::1/64  
  ip router isis  
  ipv6 router isis  
  isis ipv6 metric 20  
  
interface Ethernet 2  
  ip address 10.2.1.1 255.255.255.0  
  ipv6 address 2001:db8:2::1/64  
  ip router isis  
  ipv6 router isis  
  isis ipv6 metric 20  
  
router isis isp  
  net 49.0000.0100.0000.0500.00  
  metric-style wide  
  !  
  address-family ipv6  
    multi-topology  
  exit-address-family
```

Narrow to Wide Metrics Transition

- When migrating from narrow to wide metrics, care is required
 - Narrow and wide metrics are NOT compatible with each other
 - Migration is a two stage process, using the “transition” keyword
- Networks using narrow metrics should first configure across all routers:

```
router isis isp  
metric-style transition
```

- Once the whole network is changed to transition support, the metric style can be changed to wide:

```
router isis isp  
metric-style wide
```

Multi-Topology IS-IS Display

```
Router2# show clns neighbors detail
```

```
Tag workshop:
```

System Id	Interface	SNPA	State	Holdtime	Type	Protocol
Router2	Fa0/0	ca01.3c4f.0008	Up	7	L2	M-ISIS

```
Area Address(es): 49.0001
```

```
IP Address(es): 10.0.15.2*
```

```
IPv6 Address(es): FE80::C801:3CFF:FE4F:8
```

```
Uptime: 00:01:46
```

```
NSF capable
```

```
Topology: IPv4, IPv6
```

```
Interface name: FastEthernet0/0
```

```
Router2# show isis database detail
```

```
Tag workshop:
```

```
IS-IS Level-2 Link State Database:
```

LSPID	LSP Seq Num	LSP Checksum	LSP Holdtime	ATT/P/OL
router1.00-00	* 0x00000006	0xD3D1	1112	0/0/0

```
Area Address: 49.0001
```

```
Topology: IPv4 (0x0)
```

```
IPv6 (0x2)
```

```
NLPID: 0xCC 0x8E
```

```
Hostname: router1
```

```
IP Address: 10.0.15.241
```

```
IPv6 Address: 2001:DB8::1
```

```
Metric: 2 IS-Extended Router2.01
```

```
Metric: 20 IS-Extended Router3.00
```

```
Metric: 2 IS-Extended Router13.02
```

```
Metric: 2 IS (MT-IPv6) Router2.01
```

```
Metric: 20 IS (MT-IPv6) Router3.00
```

```
Metric: 2 IS (MT-IPv6) Router13.02
```


Multi-Topology IS-IS Support

- In Cisco IOS:
 - Supported in 12.2SRE, 12.2SXH, 12.4T, and 15.0 onwards
 - The commands for MT are in 12.3 and 12.4 but do not work
 - The only workaround is to use single topology or change to the knowing working releases
- In Cisco IOS-XE:
 - Supported in 3.3 or later
- In Cisco IOS-XR:
 - Supported in 3.9 or later
 - Note: MT is enabled by default
- In Juniper JunOS:
 - Supported in 9.0 or later

Introduction to BGP

ISP Workshops

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

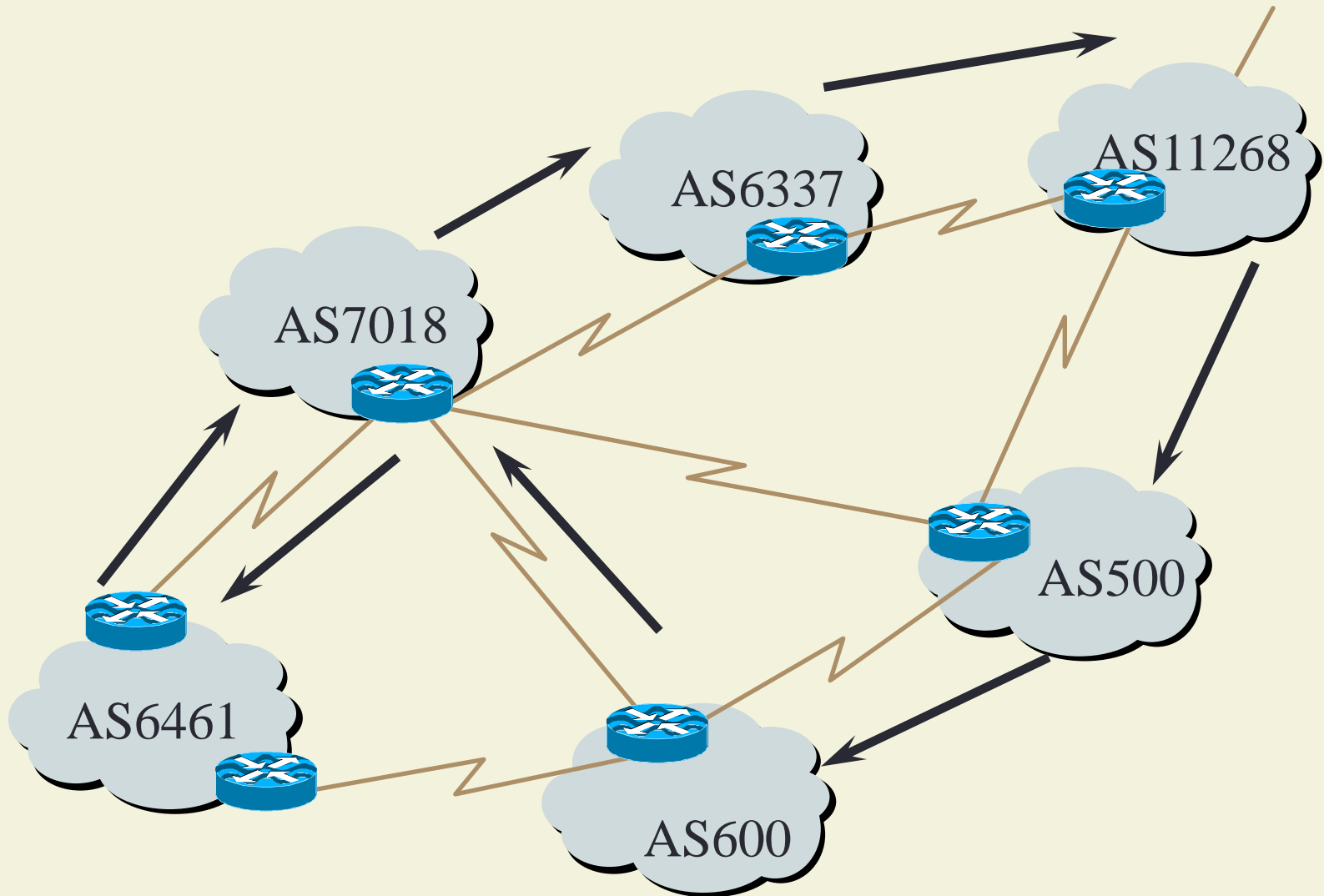
- BGP is classified as a *path vector* routing protocol (see RFC 1322)
 - A path vector protocol defines a route as a pairing between a destination and the attributes of the path to that destination.

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



AS Path

Path Vector Protocol



Definitions

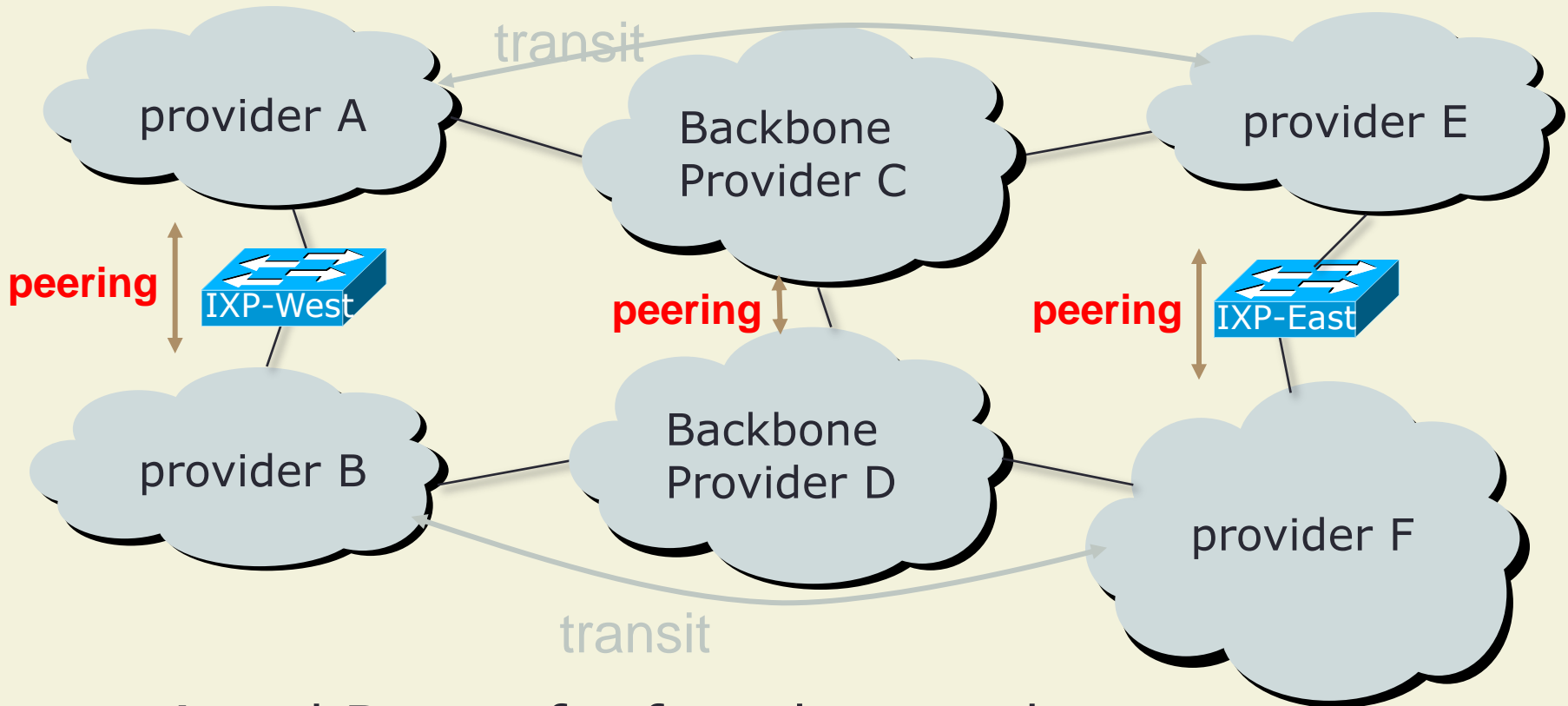
- **Transit** – carrying traffic across a network, usually for a fee
- **Peering** – exchanging routing information and traffic
- **Default** – where to send traffic when there is no explicit match in the routing table

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

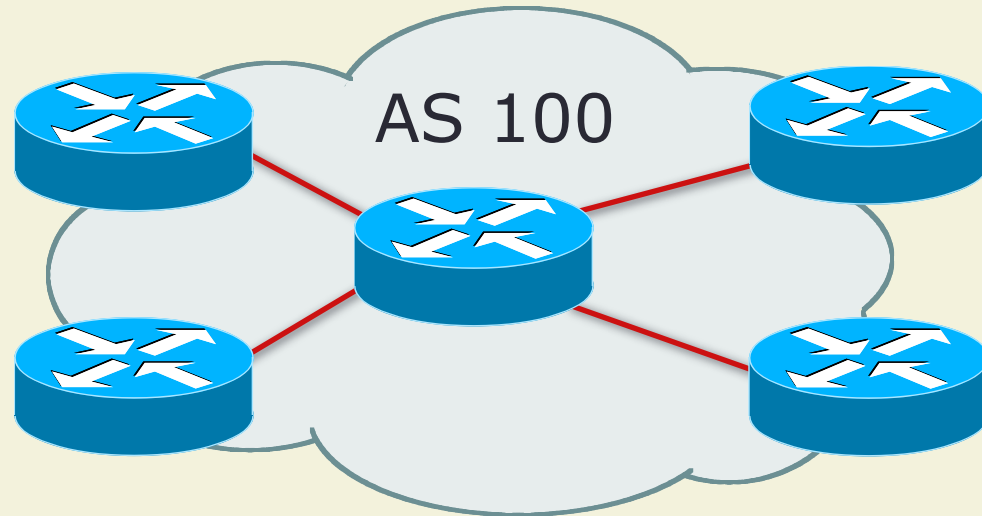
NB: is not related to where an ISP is in the hierarchy

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
```

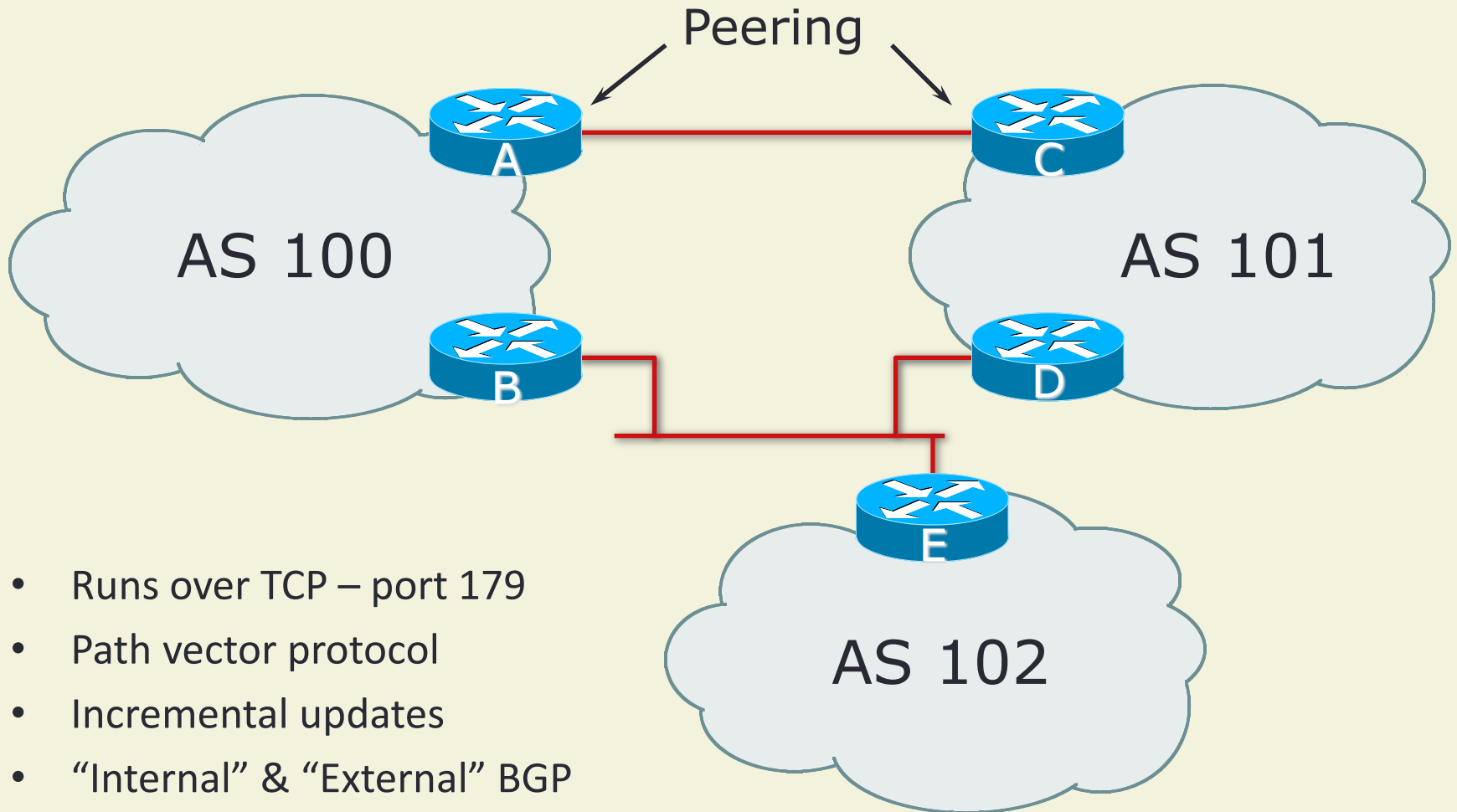
- IOS will display ASNs in plain notation by default

- Dot notation is optional:

```
router bgp 2.4
```

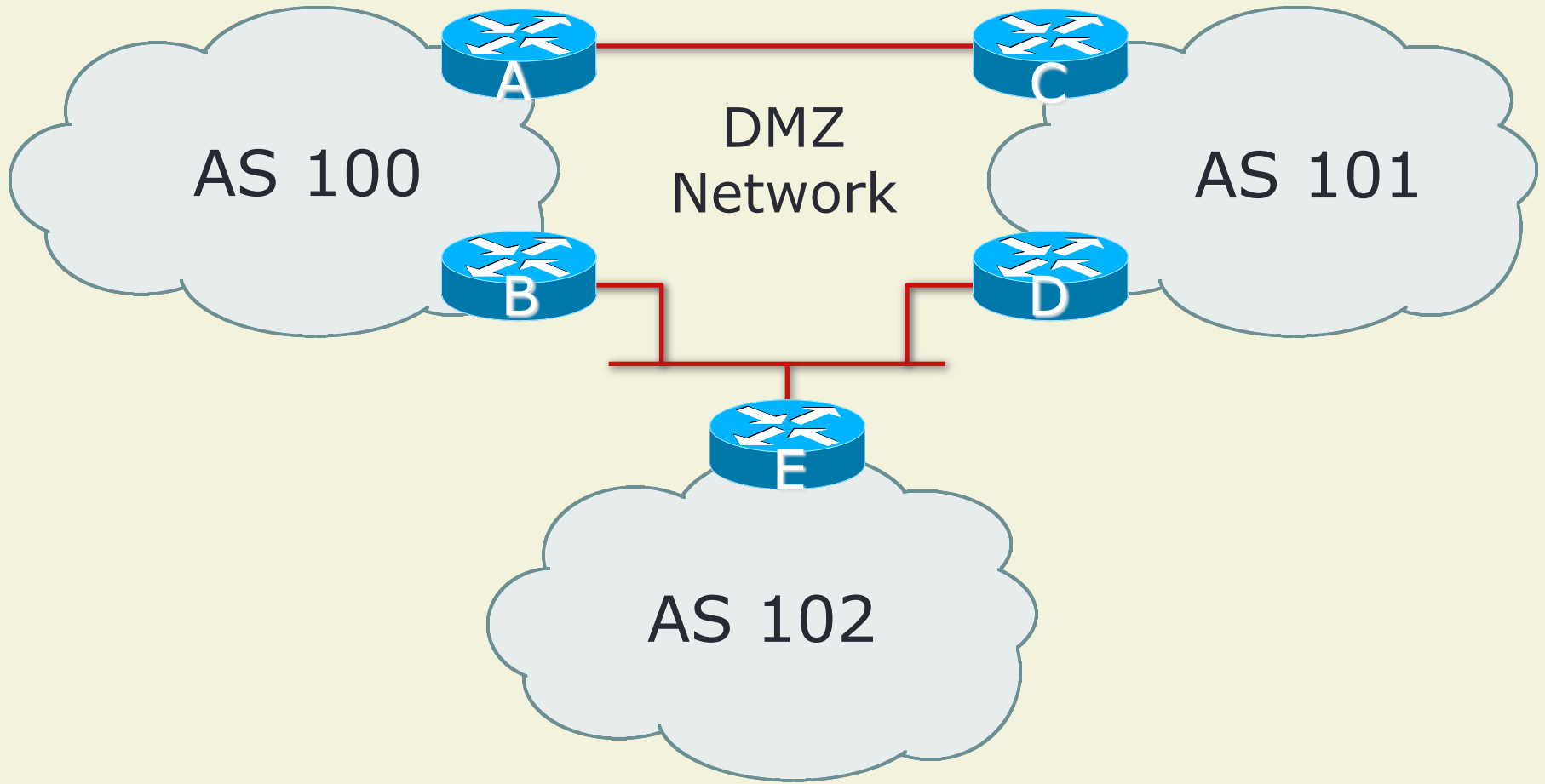
```
bgp asnotation dot
```

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

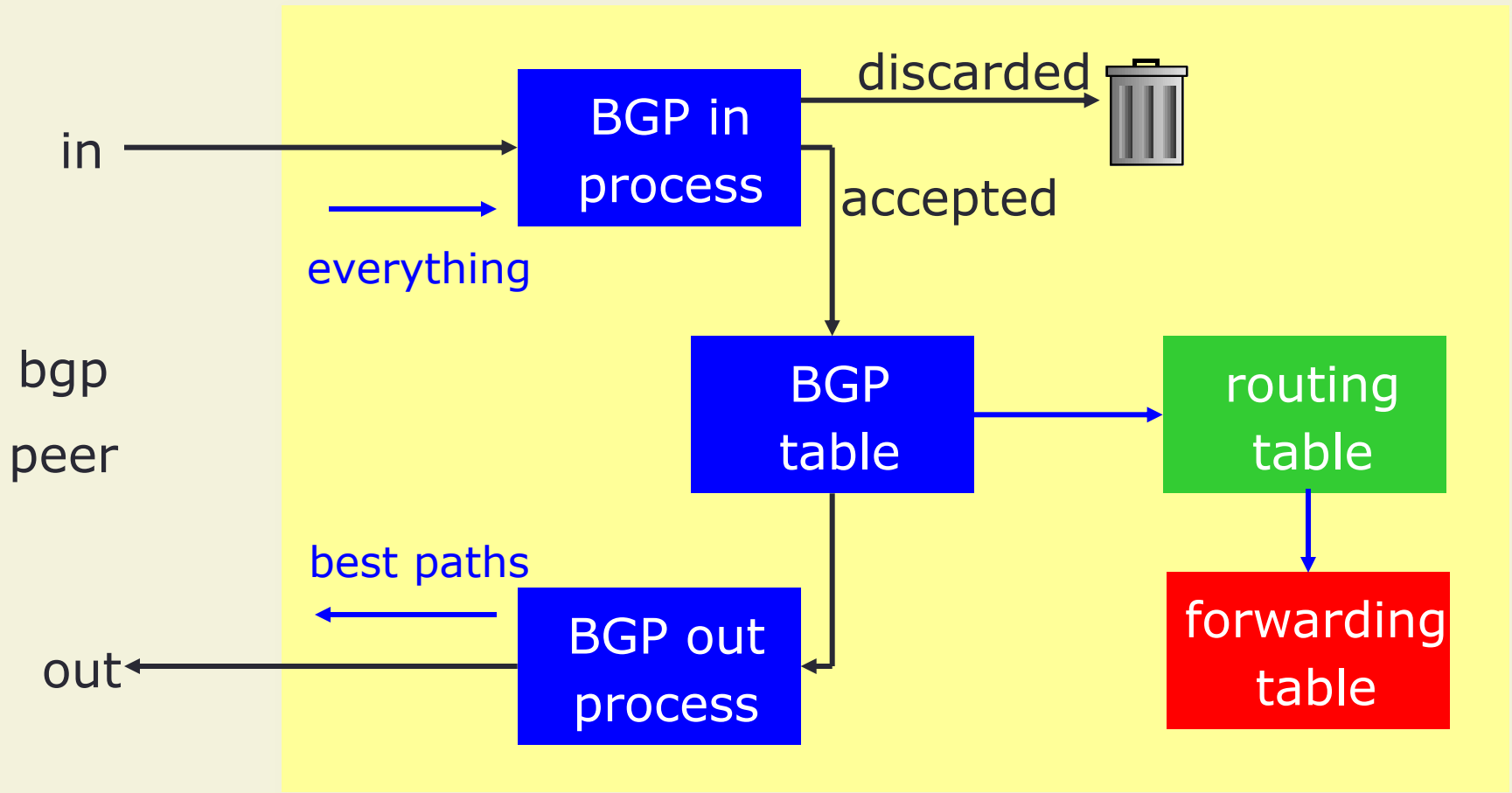
BGP General Operation

- 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

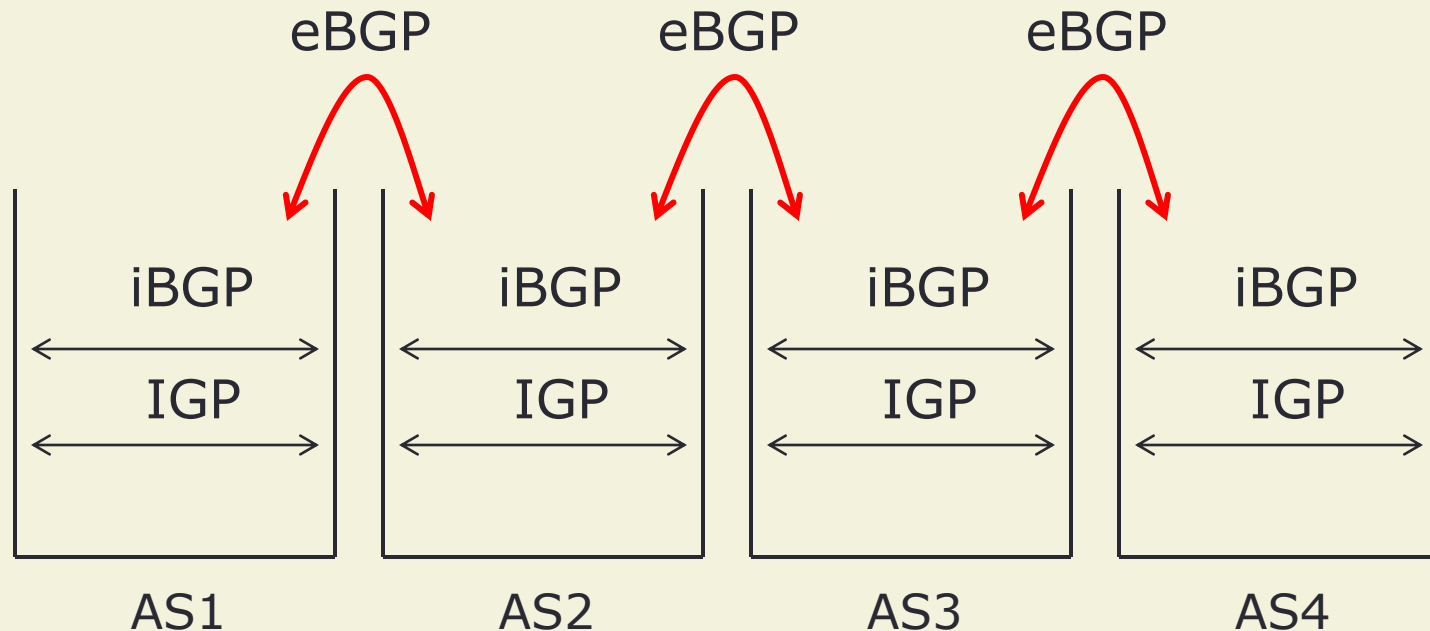


eBGP & iBGP

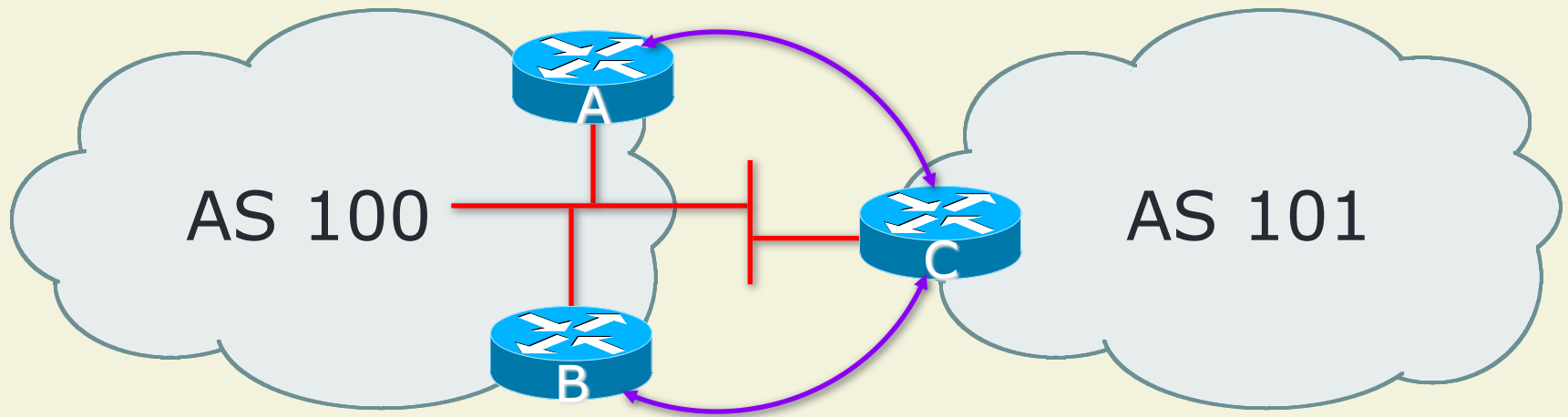
- BGP used internally (iBGP) and externally (eBGP)
- iBGP used to carry
 - Some/all Internet prefixes across ISP backbone
 - ISP' s customer prefixes
- eBGP used to
 - Exchange prefixes with other ASes
 - Implement routing policy

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
!
```

ip address on
ethernet interface

Local ASN

Remote ASN

ip address of Router
C ethernet interface

Inbound and
outbound filters

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
!
```

ip address on
ethernet interface

Local ASN

Remote ASN

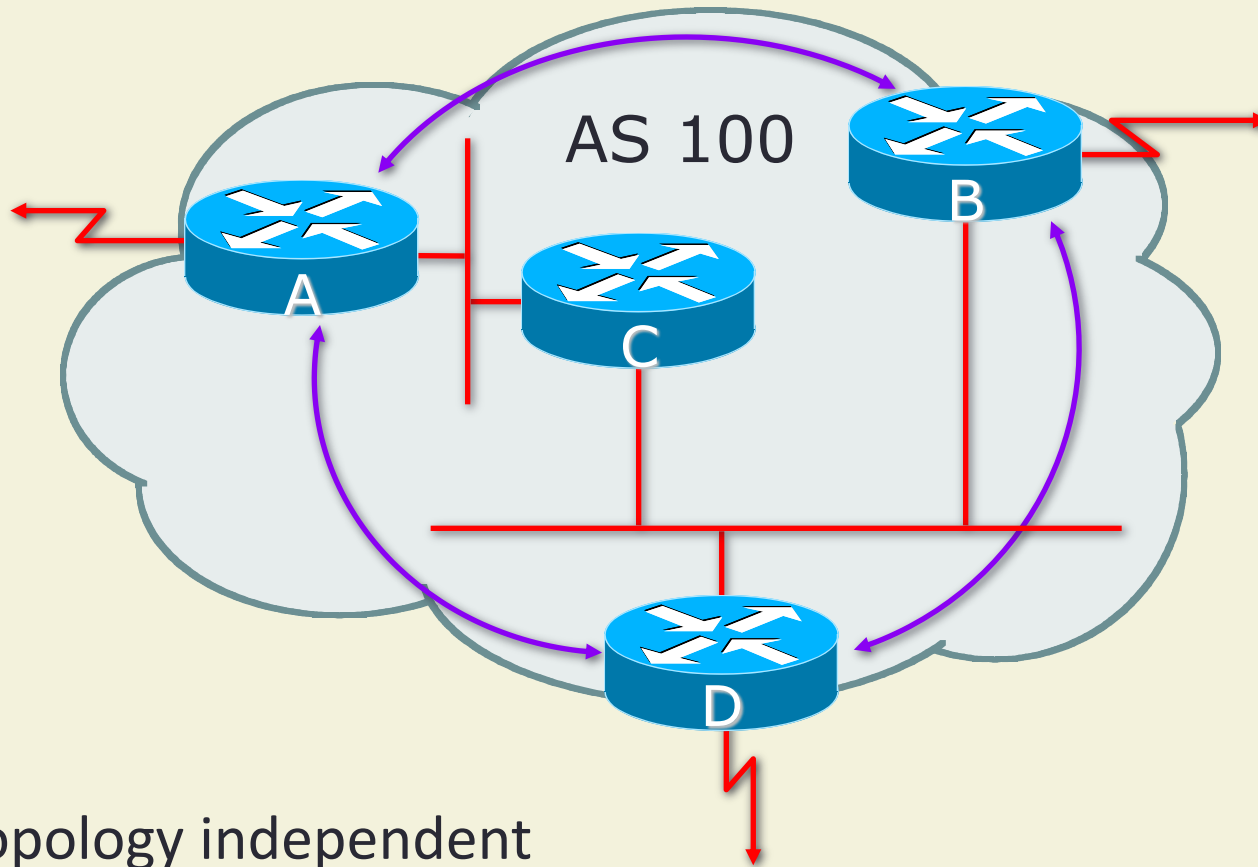
ip address of Router
A ethernet interface

Inbound and
outbound filters

Internal BGP (iBGP)

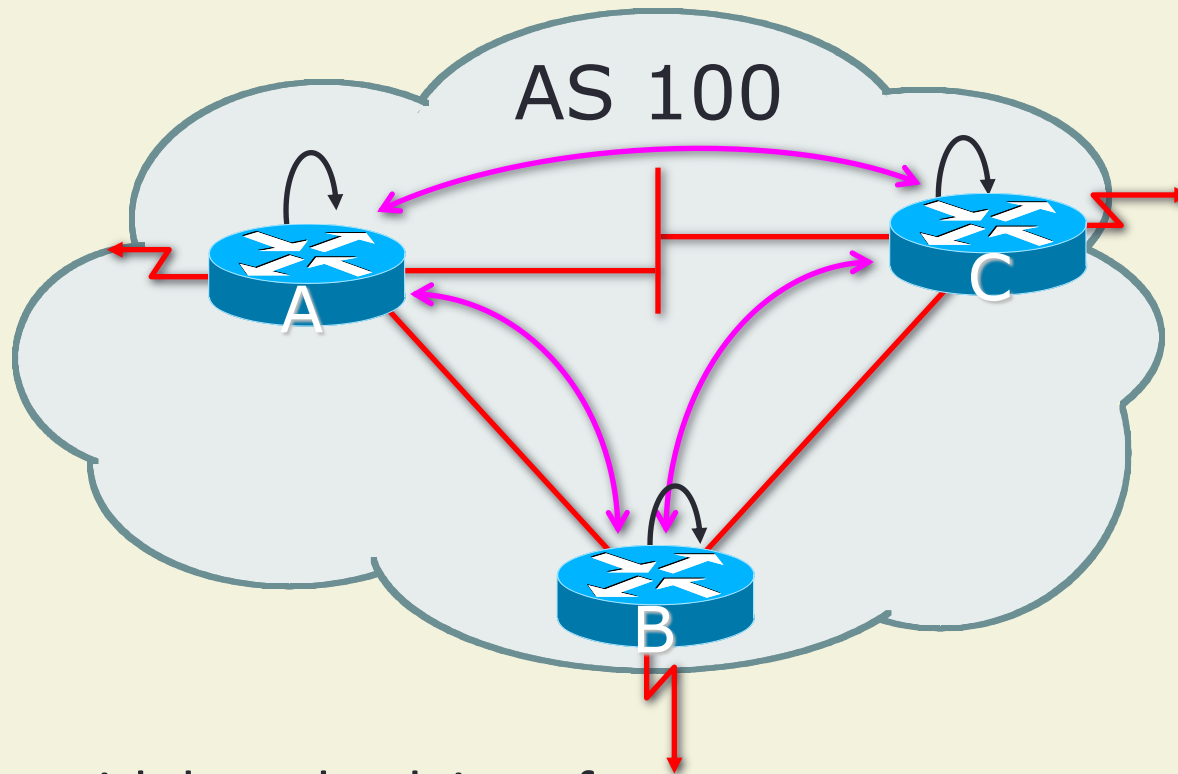
- BGP peer within the same AS
- Not required to be directly connected
 - IGP takes care of inter-BGP speaker connectivity
- iBGP speakers must be fully meshed:
 - They originate connected networks
 - They pass on prefixes learned from outside the ASN
 - They do not pass on prefixes learned from other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

ip address on loopback interface

Local ASN

Local ASN

ip address of Router B loopback interface

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
!
```

ip address on loopback interface

Local ASN

Local ASN

ip address of Router A loopback interface

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 serial10
```

- 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

BGP Enhancements for IPv6

ISP Workshops

Adding IPv6 to BGP...

- RFC4760
 - Defines Multi-protocol Extensions for BGP4
 - Enables BGP to carry routing information of protocols other than IPv4
 - e.g. MPLS, IPv6, Multicast etc
 - Exchange of multiprotocol NLRI must be negotiated at session startup
- RFC2545
 - Use of BGP Multiprotocol Extensions for IPv6 Inter-Domain Routing

RFC4760

- New optional and non-transitive BGP attributes:
 - MP_REACH_NLRI (Attribute code: 14)
 - Carry the set of reachable destinations together with the next-hop information to be used for forwarding to these destinations (RFC2858)
 - MP_UNREACH_NLRI (Attribute code: 15)
 - Carry the set of unreachable destinations
- Attribute contains one or more Triples:
 - AFI Address Family Information
 - Next-Hop Information
(must be of the same address family)
 - NLRI Network Layer Reachability Information

RFC2545

- IPv6 specific extensions
 - Scoped addresses: Next-hop contains a global IPv6 address and/or potentially a link-local address
 - NEXT_HOP and NLRI are expressed as IPv6 addresses and prefix
 - Address Family Information (AFI) = 2 (IPv6)
 - Sub-AFI = 1 (NLRI is used for unicast)
 - Sub-AFI = 2 (NLRI is used for multicast RPF check)
 - Sub-AFI = 3 (NLRI is used for both unicast and multicast RPF check)
 - Sub-AFI = 4 (label)

BGP Considerations

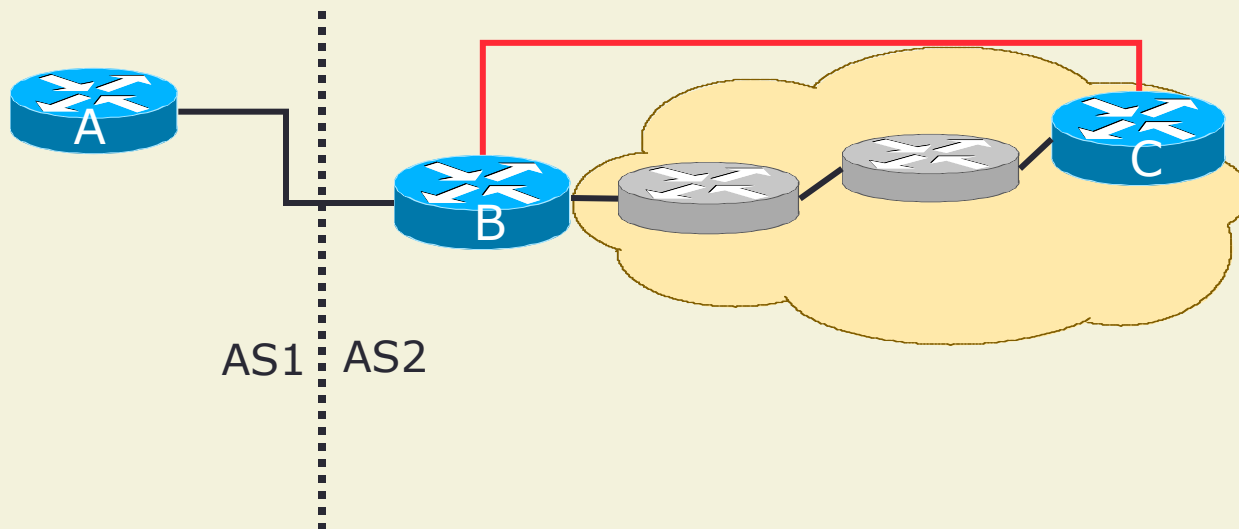
- Rules for constructing the NEXTHOP attribute:
 - When two peers share a common subnet, the NEXTHOP information is formed by a global address and a link local address
 - Redirects in IPv6 are restricted to the usage of link local addresses

Routing Information

- Independent operation
 - One RIB per protocol
 - e.g. IPv6 has its own BGP table
 - Distinct policies per protocol
- Peering sessions can be shared when the topology is congruent

BGP next-hop attribute

- Next-hop contains a global IPv6 address (or potentially a link local address)
- Link local address is set as a next-hop only if the BGP peer shares the subnet with both routers (advertising and advertised)



More BGP considerations

- TCP Interaction
 - BGP runs on top of TCP
 - This connection could be set up either over IPv4 or IPv6
- Router ID
 - When no IPv4 is configured, an explicit bgp router-id needs to be configured
 - BGP identifier is a 32 bit integer currently generated from the router identifier – which is generated from an IPv4 address on the router
 - This is needed as a BGP identifier, is used as a tie breaker, and is sent within the OPEN message

BGP Configuration

- Two options for configuring BGP peering
- Using link local addressing
 - ISP uses FE80:: addressing for BGP neighbours
 - **NOT RECOMMENDED**
 - There are plenty of IPv6 addresses
 - Unnecessary configuration complexity
- Using global unicast addresses
 - As with IPv4
 - **RECOMMENDED**

BGP Configuration

- Cisco IOS assumes that all BGP neighbours will be IPv4 unicast neighbours
 - We need to remove this assumption

```
router bgp 100
  no bgp default ipv4-unicast
```

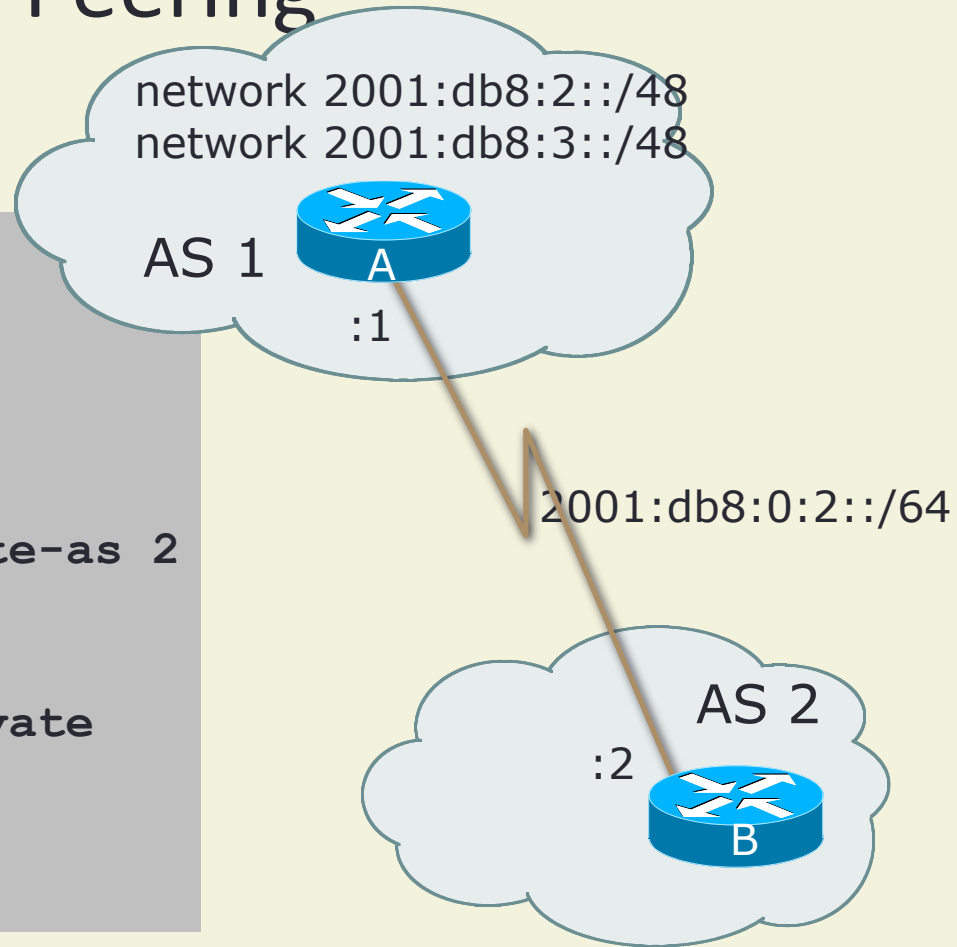
- Failing to do this will result in all neighbours being defined as IPv4 unicast neighbours
 - Non-IPv4 neighbours will have no specific unicast IPv4 configuration
 - Cluttered configuration, confusing troubleshooting and diagnosis

BGP Configurations

Regular Peering

Router A

```
router bgp 1
  no bgp default ipv4-unicast
  bgp router-id 1.1.1.1
  neighbor 2001:db8:0:2::2 remote-as 2
  !
  address-family ipv6
  neighbor 2001:db8:0:2::2 activate
  network 2001:db8:2::/48
  network 2001:db8:3::/48
  !
```



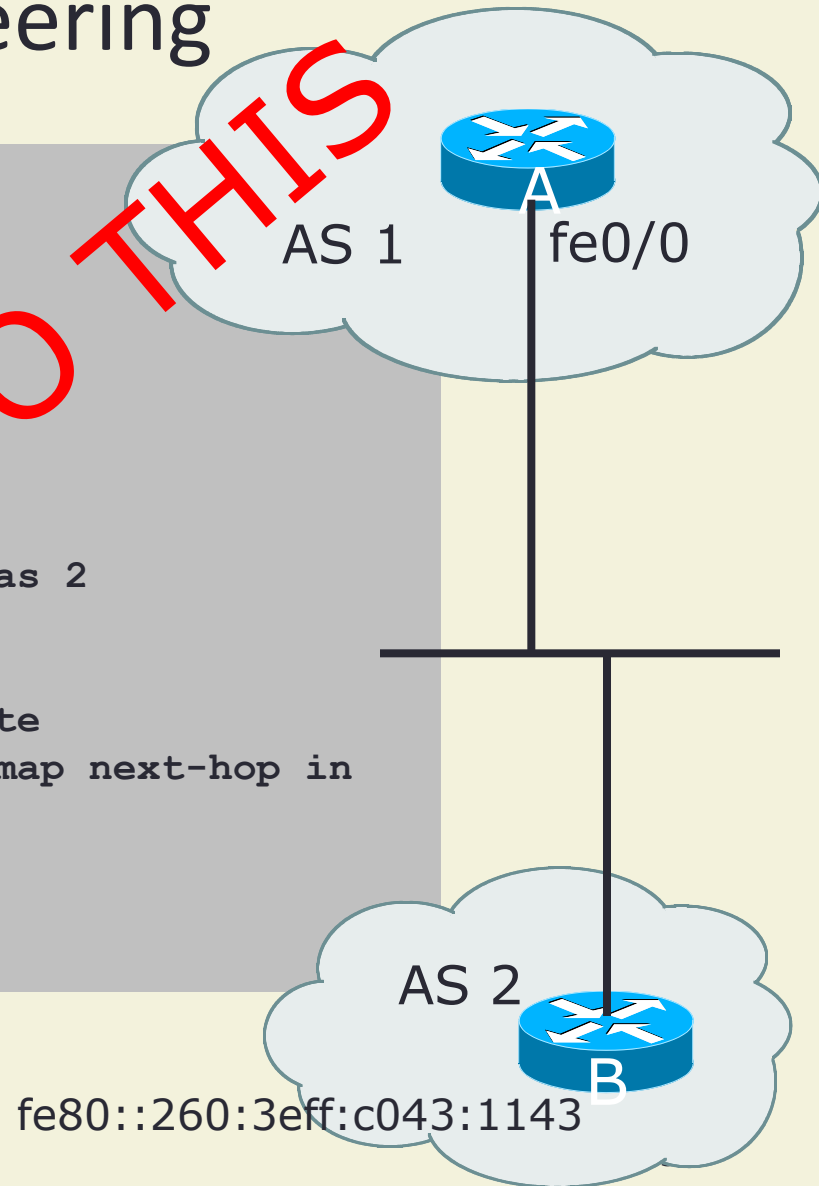
`activate` means that the BGP peering is activated for this particular address family

BGP Configurations

Link Local Peering

Router A

```
interface fastethernet 0/0
  ipv6 address 2001:db8:ffc0:1::1/64
  !
router bgp 1
  no bgp default ipv4-unicast
  bgp router-id 1.1.1.1
  neighbor fe80::260:3eff:c043:1143 remote-as 2
  !
  address-family ipv6
    neighbor fe80::260:3eff:c043:1143 activate
    neighbor fe80::260:3eff:c043:1143 route-map next-hop in
  !
route-map next-hop permit 5
  set ipv6 next-hop 2001:db8:ffc0:1::1
  !
```



BGP Configuration

IPv4 and IPv6

```
router bgp 10
  no bgp default ipv4-unicast
  neighbor 2001:db8:1:1019::1 remote-as 20
  neighbor 172.16.1.2 remote-as 30
!
  address-family ipv4
  neighbor 172.16.1.2 activate
  neighbor 172.16.1.2 prefix-list ipv4-ebgp in
  neighbor 172.16.1.2 prefix-list v4out out
  network 172.16.0.0
  exit-address-family
!
  address-family ipv6
  neighbor 2001:db8:1:1019::1 activate
  neighbor 2001:db8:1:1019::1 prefix-list ipv6-ebgp in
  neighbor 2001:db8:1:1019::1 prefix-list v6out out
  network 2001:db8::/32
  exit-address-family
!
! Continued -->
```

BGP Configuration

IPv4 and IPv6

```
ip prefix-list ipv4-ebgp permit 0.0.0.0/0 le 32
!  
ip prefix-list v4out permit 172.16.0.0/16
!  
ipv6 prefix-list ipv6-ebgp permit ::/0 le 128
!  
ipv6 prefix-list v6out permit 2001:db8::/32
!
```

- Compare IPv4 prefix filters with IPv6 prefix filters

```
ip prefix-list <name> permit|deny <ipv4 address>
```

```
ipv6 prefix-list <name> permit|deny <ipv6 address>
```

BGP Configuration

IPv4 and IPv6

- When configuring the router, recommendation is:
 - Put all IPv6 configuration directly into IPv6 address family
 - Put all IPv4 configuration directly into IPv4 address family
- Router will sort generic from specific address family configuration when the configuration is saved to NVRAM or displayed on the console
- Example follows...
 - Notice how activate is added by the router to indicate that the peering is activated for the particular address family

BGP Address Families Applied Configuration

```
router bgp 10
  no bgp default ipv4-unicast
  !
  address family ipv4
    neighbor 172.16.1.2 remote-as 30
    neighbor 172.16.1.2 prefix-list ipv4-ebgp in
    neighbor 172.16.1.2 prefix-list v4out out
    network 172.16.0.0
  !
  address-family ipv6
    neighbor 2001:db8:1:1019::1 remote-as 20
    neighbor 2001:db8:1:1019::1 prefix-list ipv6-ebgp in
    neighbor 2001:db8:1:1019::1 prefix-list v6out out
    network 2001:db8::/32
  !
  ip prefix-list ipv4-ebgp permit 0.0.0.0/0 le 32
  ip prefix-list v4out permit 172.16.0.0/16
  ipv6 prefix-list ipv6-ebgp permit ::/0 le 128
  ipv6 prefix-list v6out permit 2001:db8::/32
```

Generic Configuration

Specific Configuration

BGP Address Families

End result

```
router bgp 10
  no bgp default ipv4-unicast
  neighbor 2001:db8:1:1019::1 remote-as 20
  neighbor 172.16.1.2 remote-as 30
!
  address-family ipv4
  neighbor 172.16.1.2 activate
  neighbor 172.16.1.2 prefix-list ipv4-ebgp in
  neighbor 172.16.1.2 prefix-list v4out out
  network 172.16.0.0
  exit-address-family
!
  address-family ipv6
  neighbor 2001:db8:1:1019::1 activate
  neighbor 2001:db8:1:1019::1 prefix-list ipv6-ebgp in
  neighbor 2001:db8:1:1019::1 prefix-list v6out out
  network 2001:db8::/32
  exit-address-family
!
ip prefix-list ipv4-ebgp permit 0.0.0.0/0 le 32
ip prefix-list v4out permit 172.16.0.0/16
ipv6 prefix-list ipv6-ebgp permit ::/0 le 128
ipv6 prefix-list v6out permit 2001:db8::/32
```

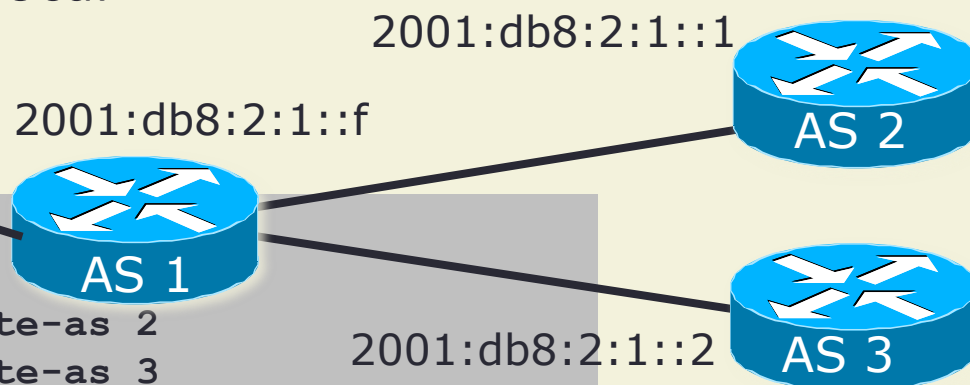
Generic Configuration

Specific Configuration

BGP Configuration

Manipulating Attributes

- Prefer routes from AS 2 (local preference)



```
router bgp 1
  no bgp default ipv4-unicast
  neighbor 2001:db8:2:1::1 remote-as 2
  neighbor 2001:db8:2:1::2 remote-as 3
!
  address-family ipv6
  neighbor 2001:db8:2:1::1 activate
  neighbor 2001:db8:2:1::1 prefix-list in-filter in
  neighbor 2001:db8:2:1::1 route-map fromAS2 in
  neighbor 2001:db8:2:1::2 activate
  neighbor 2001:db8:2:1::2 prefix-list in-filter in
  network 2001:db8::/32
  exit-address-family
!
route-map fromAS2 permit 10
  set local-preference 120
```

BGP Configuration

Carrying IPv4 inside IPv6 peering

- IPv4 prefixes can be carried inside an IPv6 peering
 - Note that the next-hop for received prefixes needs to be “fixed”
- Example

```
router bgp 1
  neighbor 2001:db8:0:2::2 remote-as 2
  !
  address-family ipv4
    neighbor 2001:db8:0:2::2 activate
    neighbor 2001:db8:0:2::2 route-map ipv4 in
  !
  route-map ipv4 permit 10
    set ip next-hop 131.108.1.1
```

BGP Status Commands

- IPv6 BGP show commands take ipv6 as argument

```
show bgp ipv6 unicast <parameter>
```

```
Router5>sh bgp ipv6 uni 2001:DB9:4::/48
BGP routing table entry for 2001:DB9:4::/48, version 20
Paths: (1 available, best #1, table default)
  Advertised to update-groups:
    1
  Local
    2001:DB9::4 (metric 64) from 2001:DB9::4 (10.20.15.227)
      Origin IGP, metric 0, localpref 100, valid, internal, best
```

- IPv4 BGP show commands can also use this format:

```
show bgp ipv4 unicast <parameter>
```

BGP Status Commands

- Display summary information regarding the state of the BGP neighbours
`show bgp ipv6 unicast summary`

```
Router1>sh bgp ipv6 uni sum
BGP router identifier 10.10.15.224, local AS number 10
BGP table version is 28, main routing table version 28
18 network entries using 2880 bytes of memory
38 path entries using 3040 bytes of memory
9/6 BGP path/bestpath attribute entries using 1152 bytes of memory
4 BGP AS-PATH entries using 96 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 7168 total bytes of memory
BGP activity 37/1 prefixes, 95/19 paths, scan interval 60 secs

Neighbor          V    AS MsgRcvd MsgSent   TblVer  InQ  OutQ Up/Down  State/PfxRcd
2001:DB8::2       4    10    185    182     28   0    0 02:36:11      16
2001:DB8::3       4    10    180    181     28   0    0 02:36:08      11
2001:DB8:0:4::1  4    40    153    152     28   0    0 02:05:39       9
```

↑
Neighbour Information

↑
BGP Messages Activity

Conclusion

- BGP extended to support multiple protocols
 - IPv6 is but one more address family
- Operators experienced with IPv4 BGP should have no trouble adapting
 - Configuration concepts and CLI is familiar format

Thank You