CS 540
Computer Networks II

Sandy Wang
chwang_98@yahoo.com

6. ROUTING PROTOCOLS – BGP

Topics

1. Overview
2. LAN Switching
3. IPv4
4. IPv6
5. Tunnels
6. Routing Protocols -- RIP, RIPng
7. Routing Protocols -- OSPF
8. IS-IS
9. Midterm Exam
10. BGP
11. MPLS
12. Transport Layer -- TCP/UDP
13. Congestion Control & Quality of Service (QoS)
14. Access Control List (ACL)
15. Final Exam

Reference Books

- **Cisco CCNA Routing and Switching ICND2 200-101 Official Cert Guide, Academic Edition** by Wendel Odom -- July 10, 2013**.** ISBN-13: 978-1587144882
- **The TCP/IP Guide: A Comprehensive, Illustrated Internet Protocols Reference** by Charles M. Kozierok – October 1, 2005. ISBN-13: 978-1593270476
- **Data and Computer Communications (10th Edition) (William Stallings Books on Computer and Data Communications)** by Williams Stallings – September 23, 2013. ISBN-13: 978-0133506488

http://class.svuca.edu/~sandy/class/CS540/

Interior and Exterior Gateway Protocols

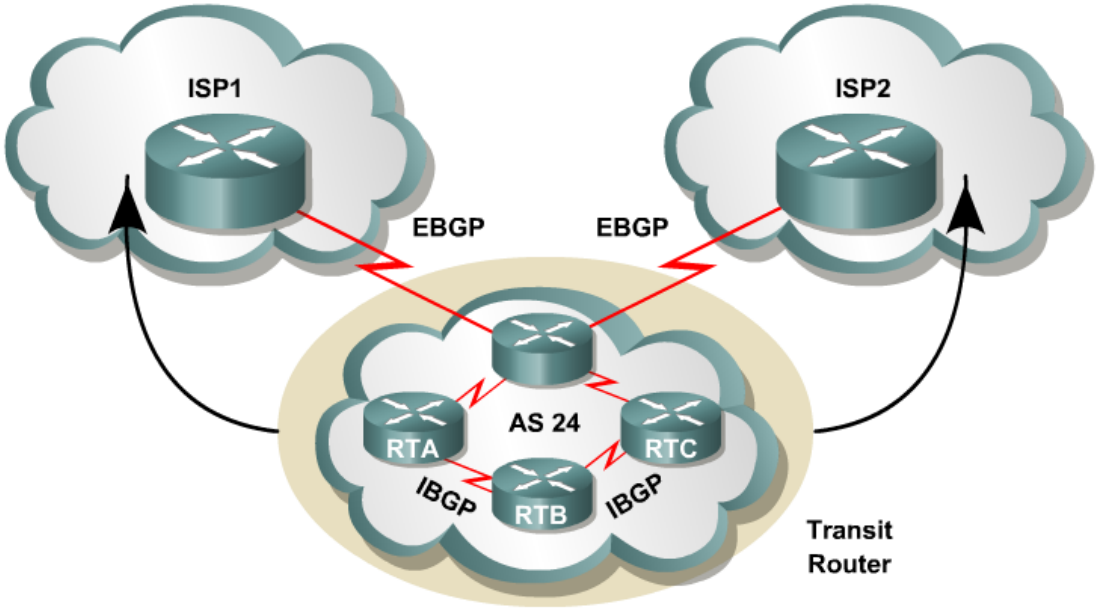
	Interior Gateway Protocols				Exterior Gateway Protocols
	Distance Vector Routing Protocols		**Link State Routing Protocols**		**Path Vector**
Classful	**RIPv1** (1982/1988)	**IGRP** (1985)			**EGP** (1982)
Classless	**RIPv2** (1994)	**EIGRP** (1992)	**OSPFv2** (1991)	**IS-IS** (1990)	**BGPv4** (1995)
IPv6	**RIPng** (1997)	**EIGRP for IPv6** (not yet released)	**OSPFv3** (1999)	**IS-IS for IPv6** (2000)	**BGPv4 for IPv6** (1999)

- **Note**: IGRP and EIGRP are Cisco proprietary protocols. They are meant as an alternative between the limited RIP routing protocol and the more complicated and resource intensive OSPF and IS-IS routing protocols. IGRP was discontinued with IOS 12.2 in 2005.
- The dates shown are when the RFC or other document was finalized. The protocol may have been implemented earlier than this date.

BGP -- Border Gateway Protocol

- An inter-Autonomous Routing Protocol
- RFC 1105, 1989 -- A Border Gateway Protocol (BGP)
- RFC 1163, 1990 -- A Border Gateway Protocol (BGP)
- RFC 1267, 1991 -- A Border Gateway Protocol 3 (BGP-3)
- RFC 1771, 1995 – A Border Gateway Protocol 4 (BGP-4)
- Latest RFC 4271 -- A Border Gateway Protocol 4 (BGP-4)
- Use TCP as the transport protocol (Use port 179)
  - Eliminates the need to implement explicit update fragmentation, retransmission, acknowledgement, and sequencing.

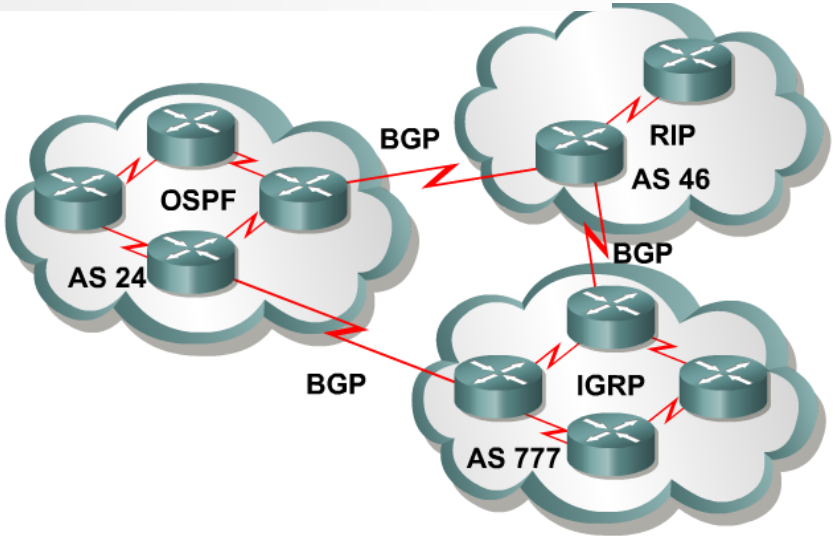
BGP Basics



- BGP is a path vector routing protocol.
- BGP is a distance vector routing protocol, in that it relies on downstream neighbors to pass along routes from their routing table.
- BGP uses a list of AS numbers through which a packet must pass to reach a destination.

Overview of autonomous systems

EGPs, such as BGP, are used to interconnect autonomous systems.

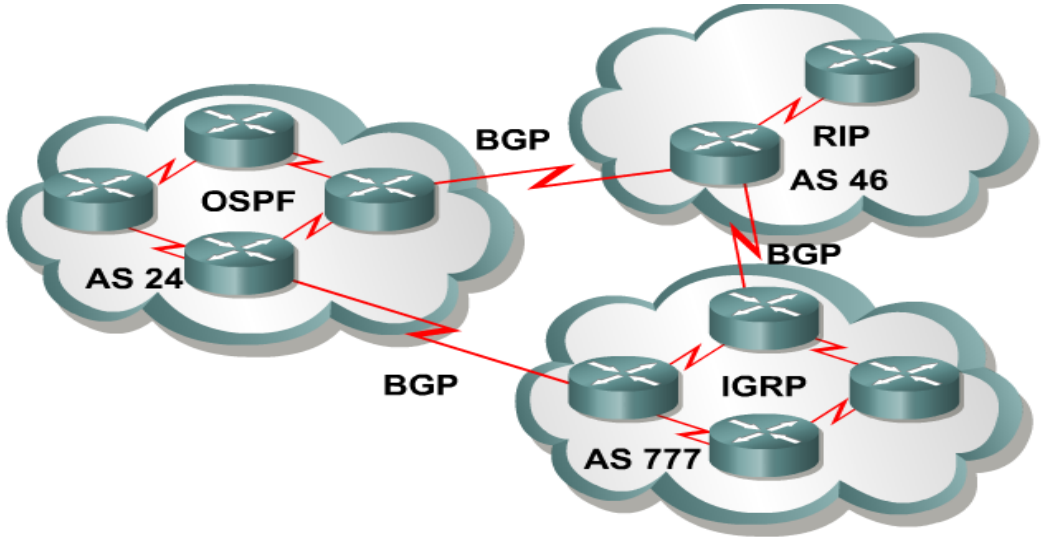


- An AS is a group of routers that share similar routing policies and operate within a single administrative domain.
- An AS can be a collection of routers running a single IGP, or it can be a collection of routers running different protocols all belonging to one organization.
- In either case, the outside world views the entire Autonomous System as a single entity.

Autonomous System Number

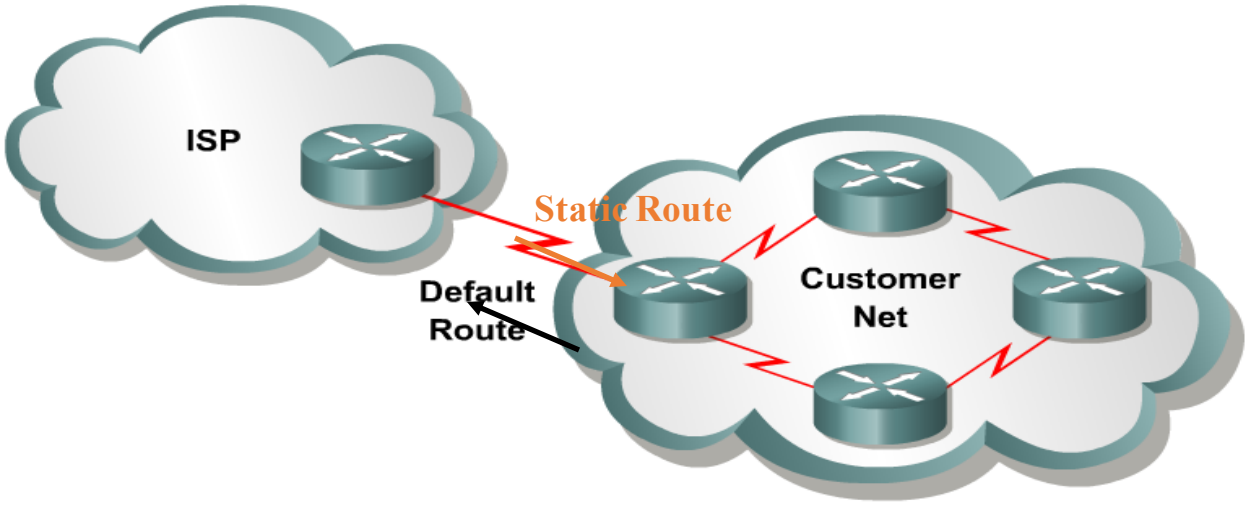
- A unique ASN is allocated to each AS for use in BGP routing. AS numbers are important because the ASN uniquely identifies each network on the Internet.
- Internet Assigned Numbers Authority (IANA)
- IANA allocates AS numbers to Regional Internet Registries (RIRs). Each AS has an identifying number that is assigned by an Internet registry or a service provider.
- This number is between **1 and 65,535**.
- AS numbers within the range of **64,512 through 65,535** are reserved for **private** use.
- Because of the finite number of available AS numbers, an organization must present justification of its need before it will be assigned an AS number.
- Extend to 32-bit in RFC 6793 -- BGP Support for Four-Octet Autonomous System (AS) Number Space

Overview of autonomous systems



- Today, the Internet Assigned Numbers Authority (IANA) is enforcing a policy whereby organizations that connect to a single provider and share the provider's routing policies use an AS number from the private pool, 64,512 to 65,535.

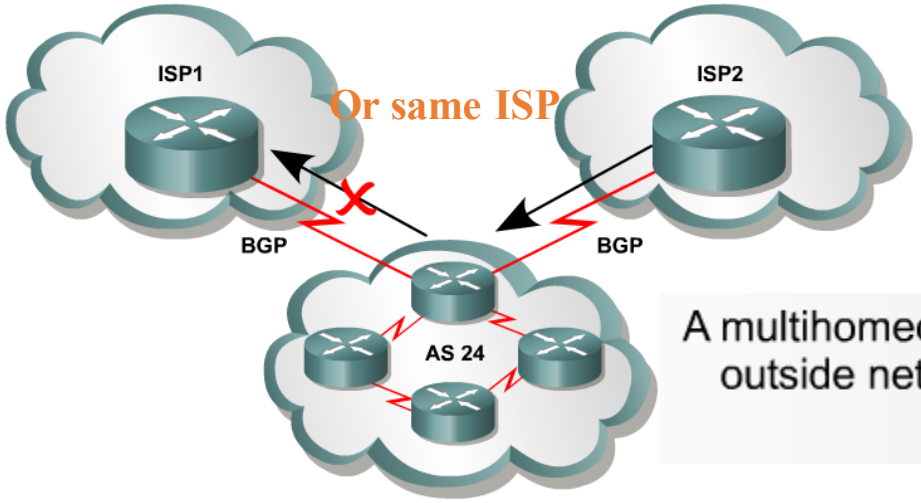
Single-homed autonomous systems



A single-homed AS can be configured with a default route to reach outside networks.

- If an AS has only **one exit point** to outside networks, it is considered a **single-homed system**.
- Single-homed autonomous systems are often referred to as **stub** networks or stubs.
- Stubs can rely on a **default route** to handle all traffic destined for non-local networks.
- BGP is **not** normally needed in this situation.

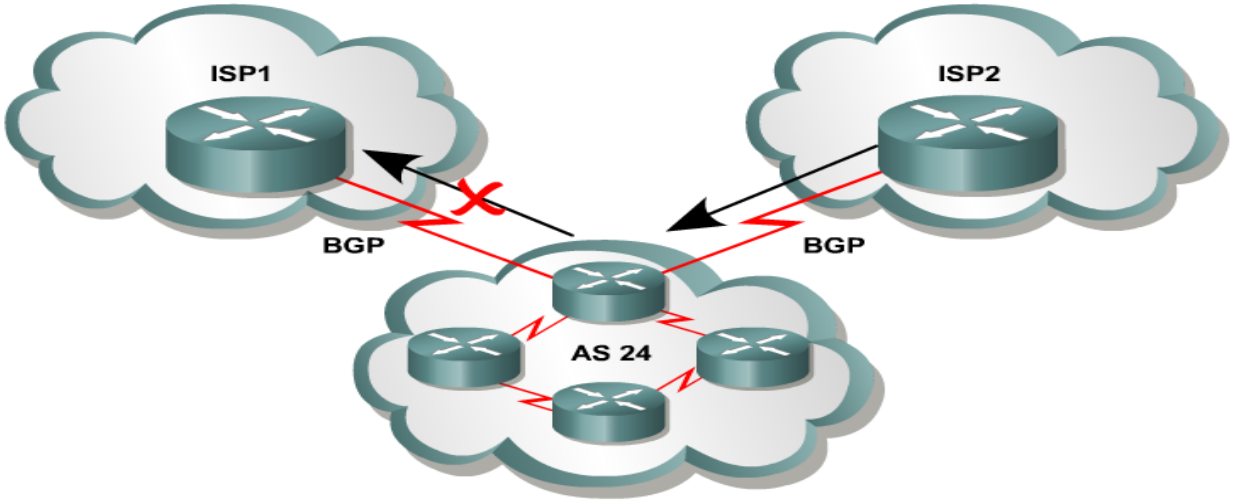
Multihomed nontransit autonomous systems



A multihomed nontransit AS features more than one exit point to outside networks, but does not allow traffic to pass from one outside connection to another.

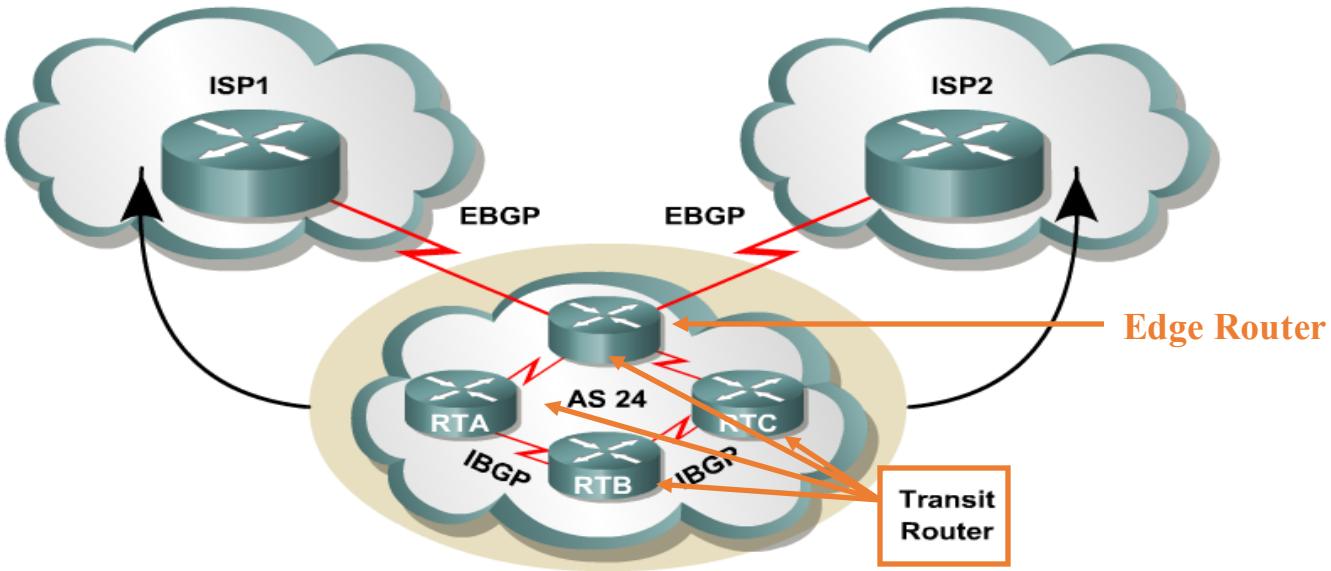
- Multihomed nontransit autonomous systems do not really need to run BGP4 with their providers.
- It is usually **recommended** and **often required** by ISPs.
- As it will be seen later in this module, BGP4 offers numerous advantages, including increased control of route propagation and filtering.

Multihomed nontransit autonomous systems



- ***Incoming route advertisements influence your outgoing traffic, and outgoing advertisements influence your incoming traffic.***
- If the provider advertises routes into your AS via BGP, your internal routers have more accurate information about external destinations.
  - BGP also provides tools for setting routing policies for external destinations.
- If your internal routes are advertised to the provider via BGP, you have influence over which routes are advertised at which exit point.
  - BGP also provides tools for your influencing (to some degree) the choices the provider makes when sending traffic into your AS.

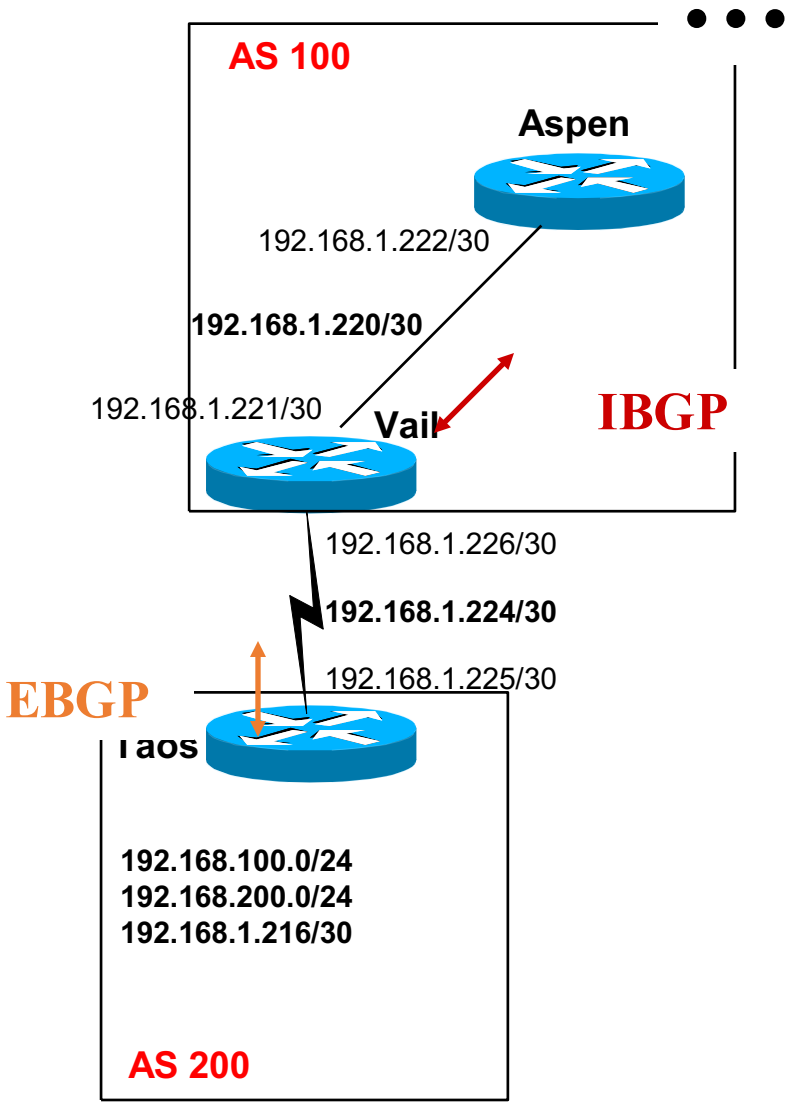
Multi-homed Transit Autonomous Systems

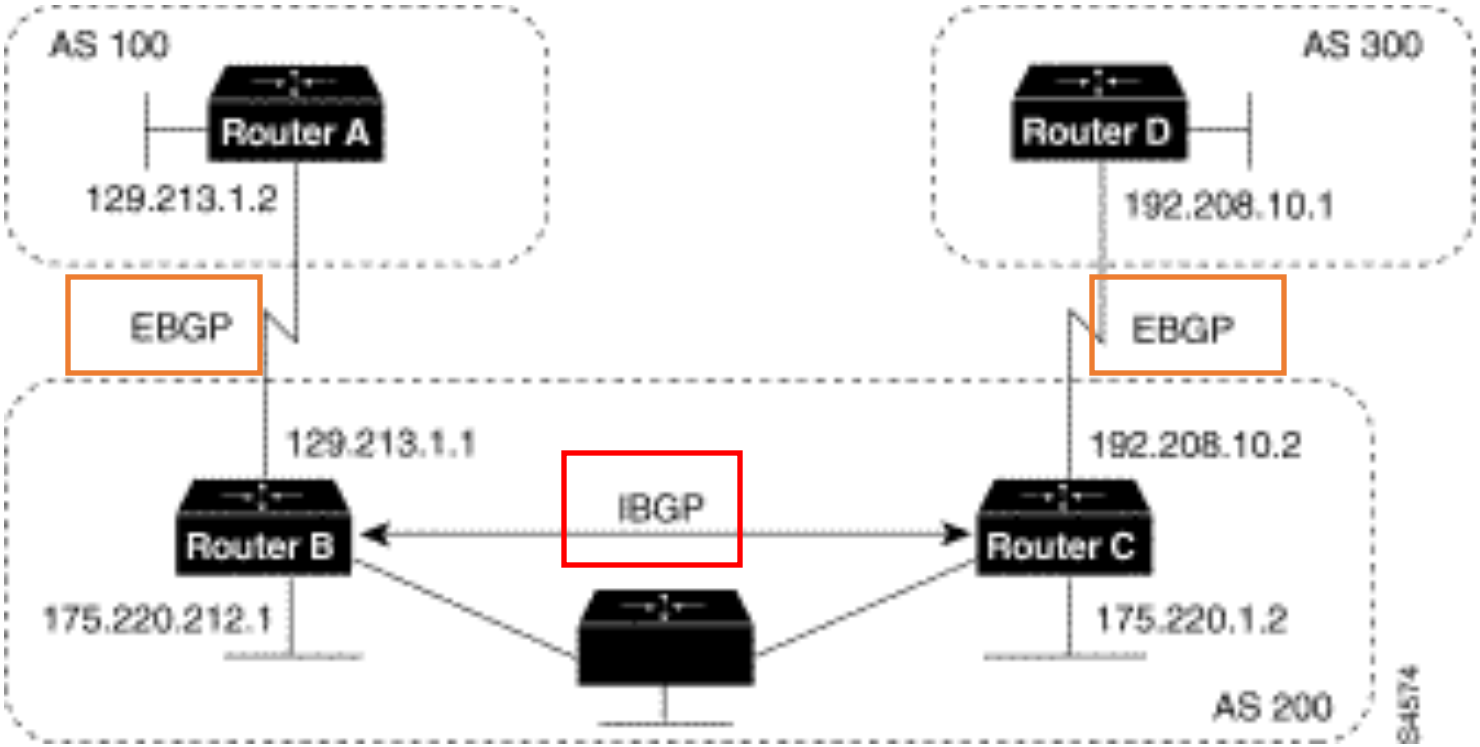


- When BGP is running inside an AS, it is referred to as **Internal BGP (IBGP).**
- When BGP runs between autonomous systems, it is called **External BGP (EBGP).**
- If the role of a BGP router is to route IBGP traffic, it is called a **transit router.**
- Routers that sit on the boundary of an AS and that use EBGP to exchange information with the ISP are called **border or edge routers**.

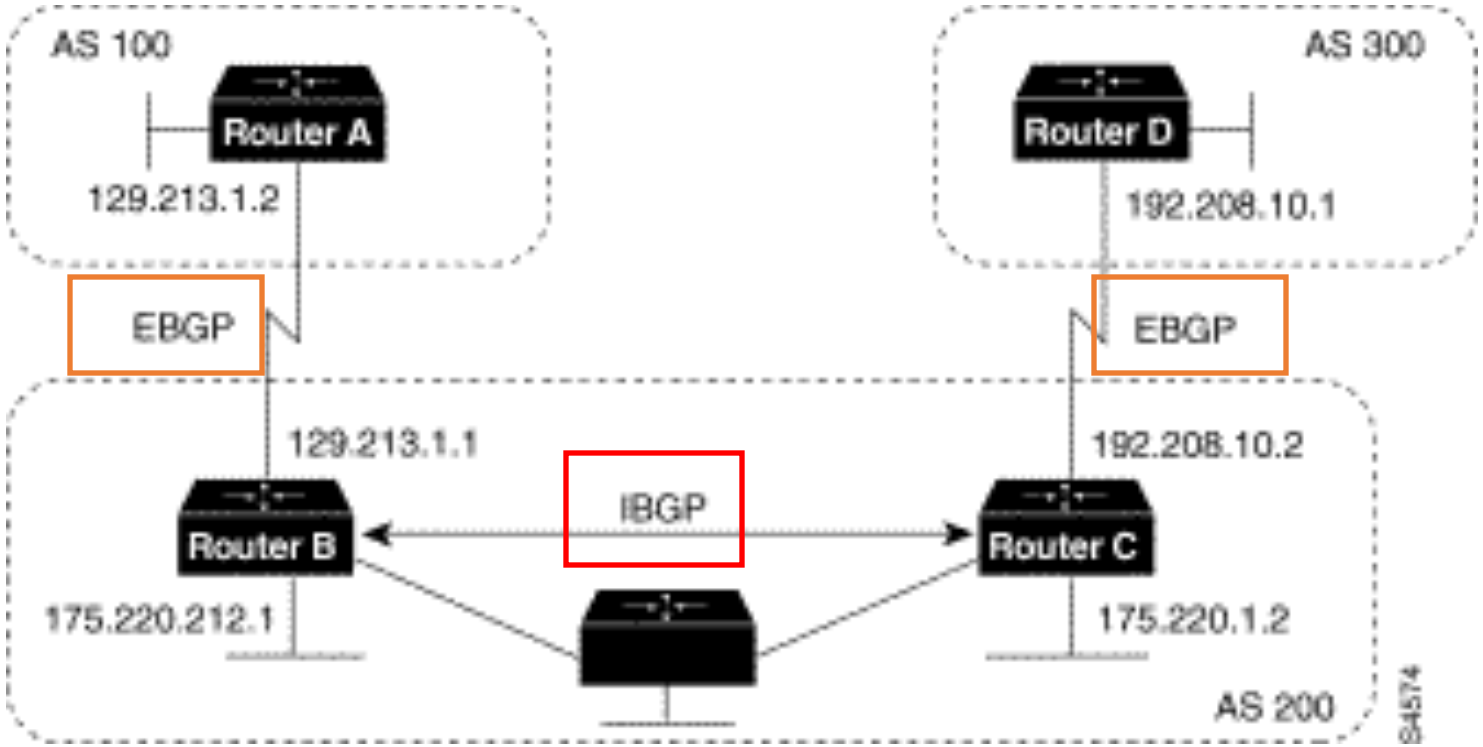
IBGP vs EBGP

- When BGP is running inside an AS, it is referred to as **Internal BGP (IBGP)**.
  - If a BGP router's role is to route IBGP traffic, it is called a transit router.
- When BGP runs between autonomous systems, it is called **External BGP (EBGP)**.
  - Routers that sit on the boundary of an AS and use EBGP to exchange information with the ISP are called border routers.



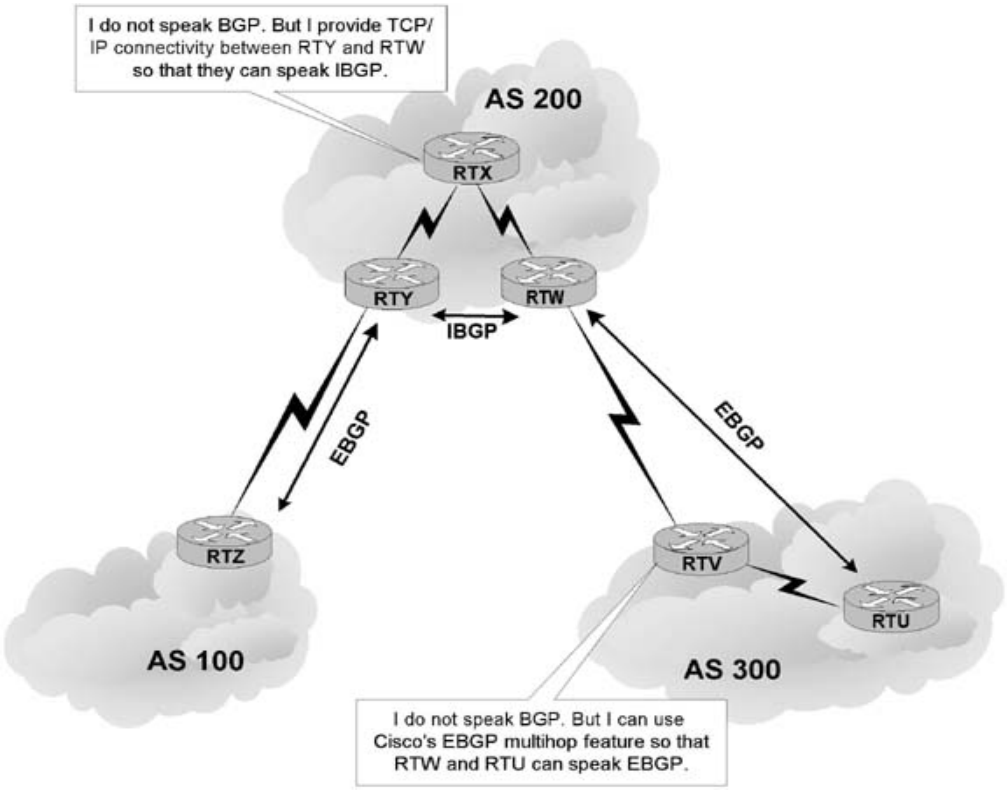


- Routers A and B are running **EBGP (BGP)**, and Routers B and C are running **IBGP**.
- Note that the **EBGP (BGP)** peers are directly connected and that the **IBGP** peers are not. (They can be.)
- As long as there is an **IGP** running that allows the two neighbors to reach one another, IBGP peers do not have to be directly connected.



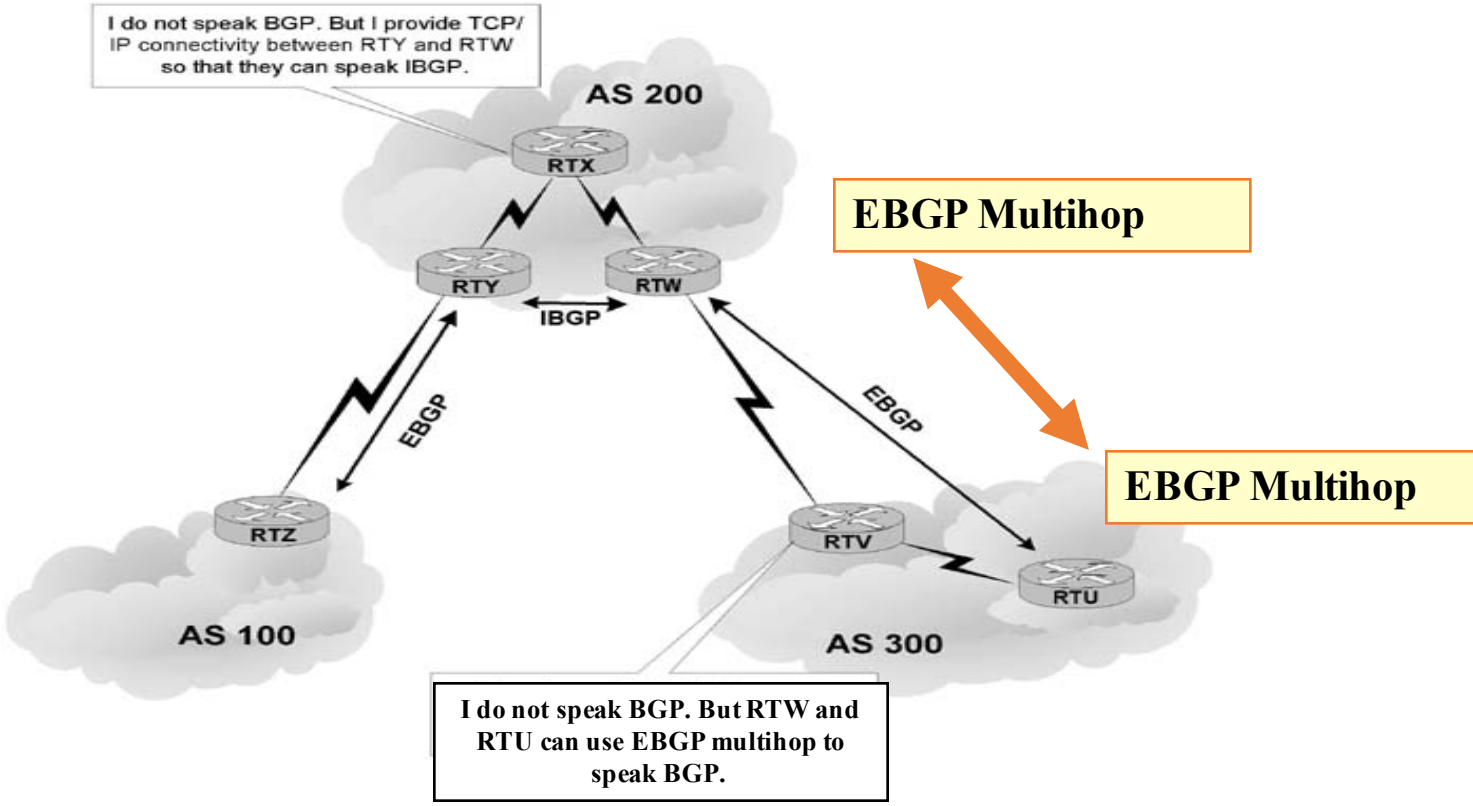
- All **BGP** speakers within an AS must establish a peer relationship with each other, that is, the **BGP** speakers within an AS must be fully meshed logically. (later)
- BGP4 provides two techniques that alleviate the requirement for a logical full mesh: confederations and route reflectors. (later)
- AS 200 is a ***transit AS*** for AS 100 and AS 300---that is, AS 200 is used to transfer packets between AS 100 and AS 300.

EBGP vs IBGP



- **EBGP** peers must be **directly connected**, but there are certain exceptions to this requirement.
- In contrast, **IBGP** peers merely **require TCP/IP connectivity** within the same AS.
  - As long as RTY can communicate with RTW using TCP, both routers can establish an IBGP session.
  - If needed, an IGP such as OSPF can provide IBGP peers with routes to each other.

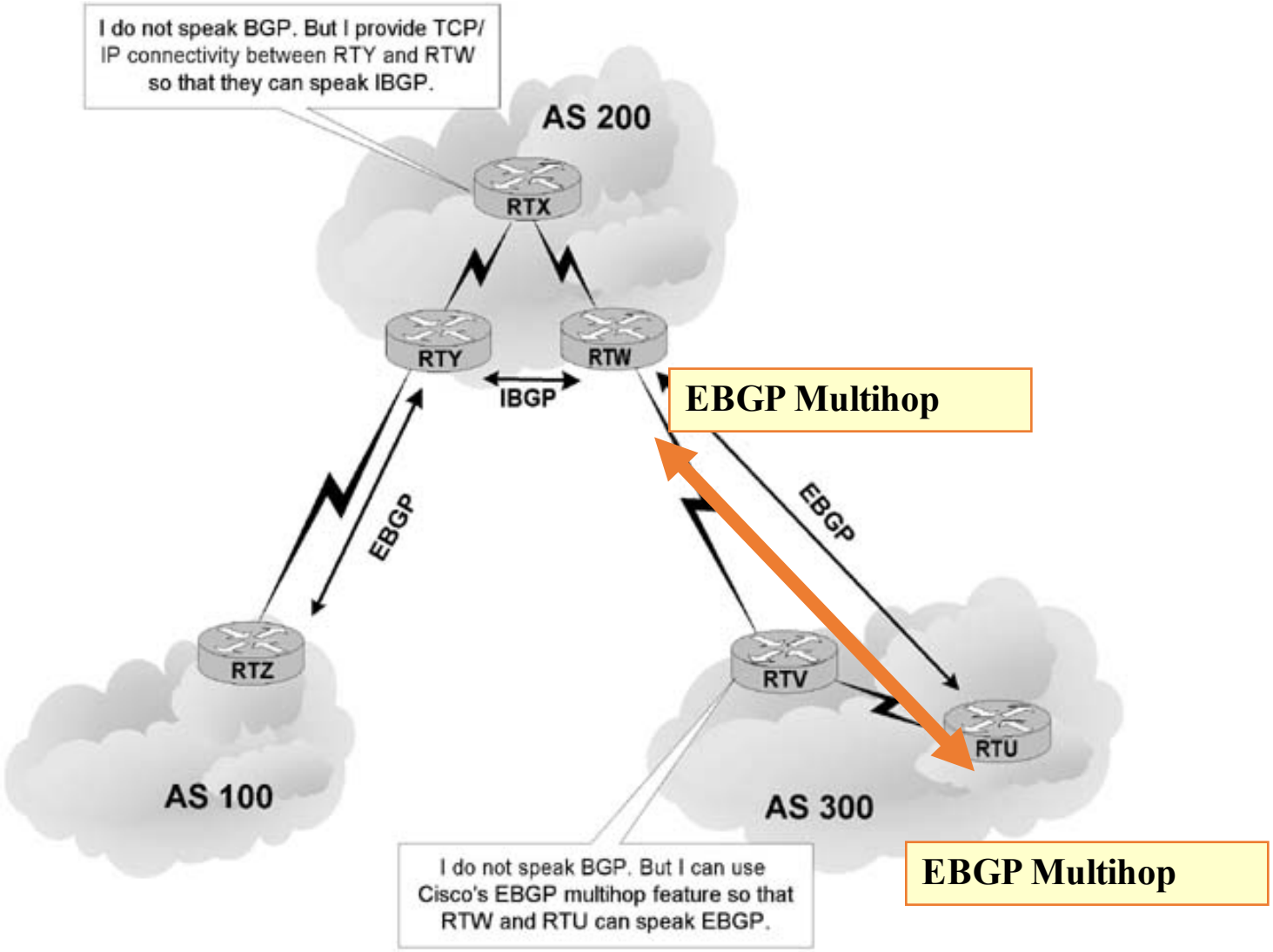
EBGP



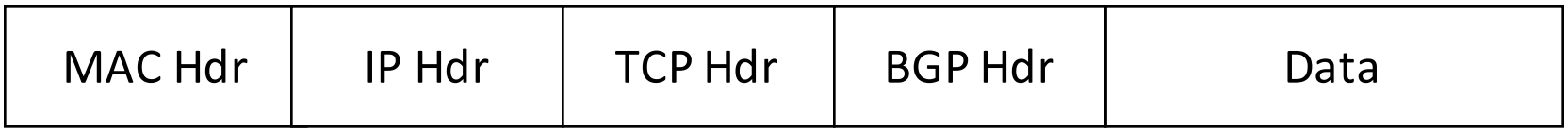
- **EBGP** neighbors must be directly connected in order to establish an **EBGP** session.
- However, **EBGP multihop** is a Cisco IOS option allows RTW and RTU to be logically connected in an **EBGP** session, despite the fact that RTV does not support BGP.
- The **EBGP** multihop option is configured on each peer with the following command:

```
Router(config-router)#neighbor IP-address ebgp-multihop
  [hops]
```

EBGP


I do not speak BGP. But I provide TCP/ IP connectivity between RTY and RTW so that they can speak IBGP.
AS 200
RTX
RTY
RTW
IBGP
EBGP Multihop
EBGP
EBGP
RTZ
RTV
RTU
AS 100
AS 300
I do not speak BGP. But I can use Cisco's EBGP multihop feature so that RTW and RTU can speak EBGP.
EBGP Multihop

BGP Packet Format



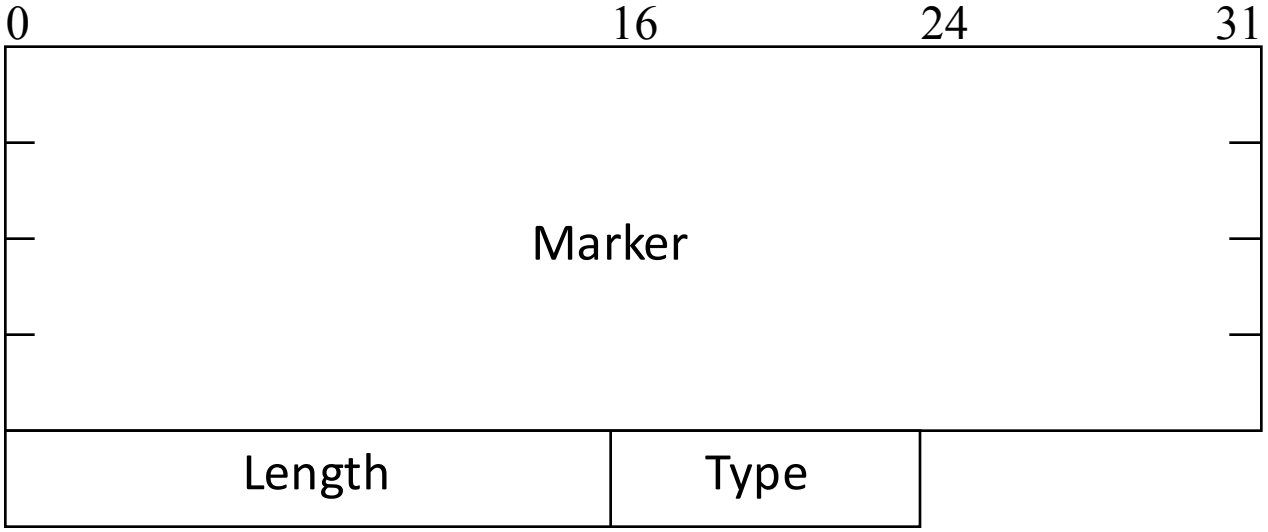
- TCP Port 179

BGP Message Types

- Before establishing a BGP peer connection the two neighbors must perform the standard TCP three-way handshake and open a TCP connection to port 179.
- After the TCP session is established, BGP peers exchanges several messages to open and confirm connection parameters and to send BGP routing information.
- All BGP messages are unicast to the one neighbor over the TCP connection.
- There are four BGP message types:
  - **Type 1: OPEN**
  - **Type 2: KEEPALIVE**
  - **Type 3: UPDATE**
  - **Type 4: NOTIFICATION**

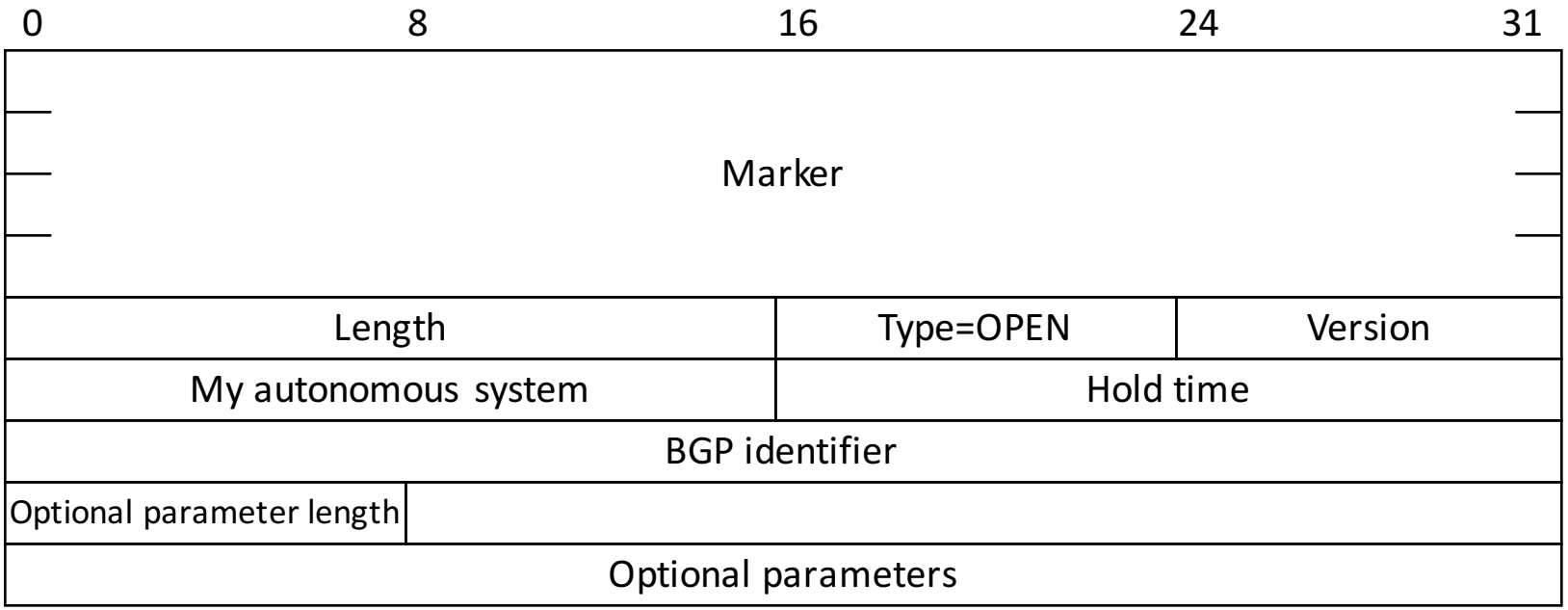
BGP messages

- BGP header format



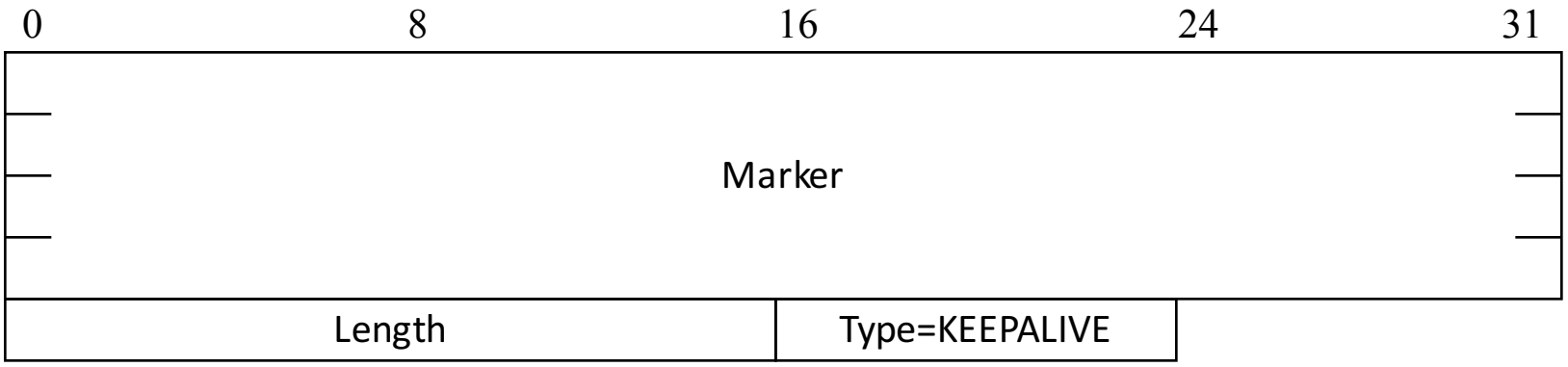
  - Marker: 16-octet that is used to detect loss of synchronization and to authenticate messages when authentication is supported.
  - Length: indicates the total length of the message in octets, including the BGP header.
  - Type: indicates the type of the message.

OPEN message



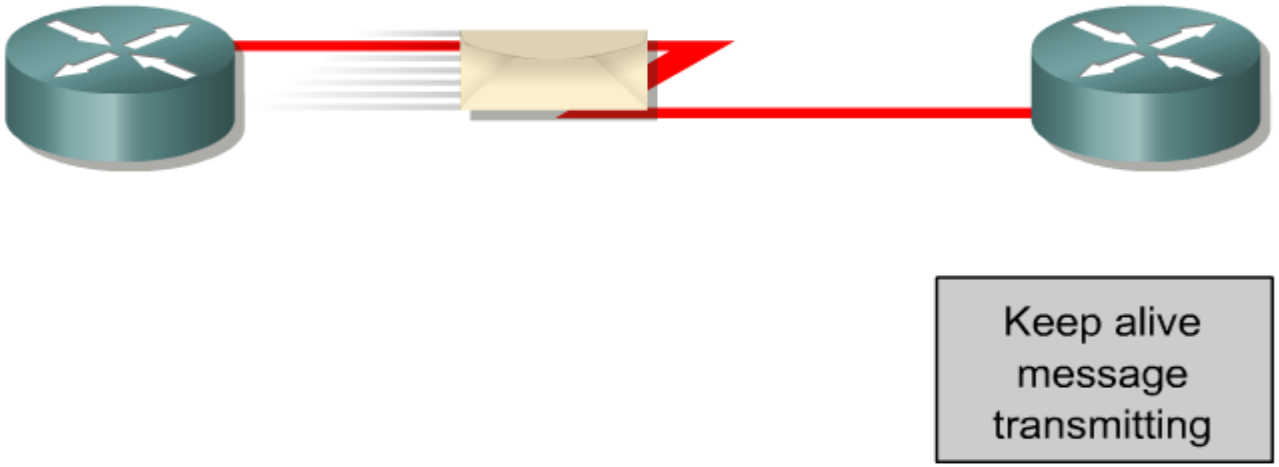
- Version: the protocol version number of the message.
- My autonomous system: The AS number of the sending router.
- Hold time: the number of seconds between the transmission of successive KEEPALIVE messages.
- BPG identifier: the sending BGP router.
- Optional parameter: a list of optional parameters, encoded in TLV structure.

KEEPALIVE message



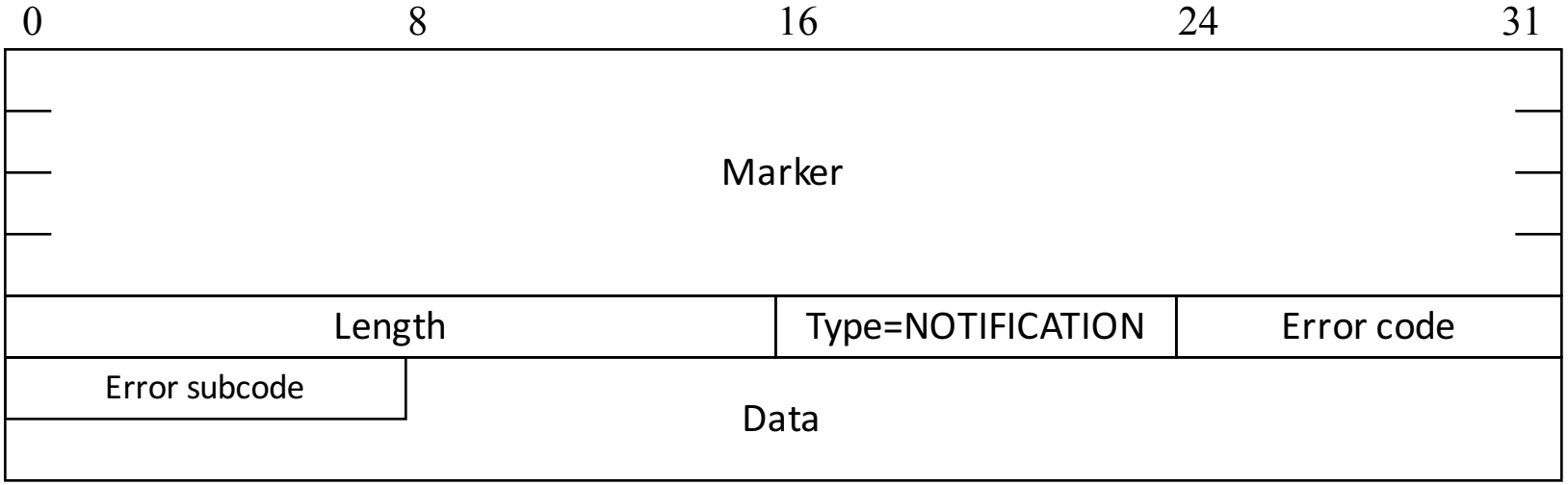
- If the hold time is zero, then KEEPALIVE messages will not be sent.

Type 2: BGP Keepalive Message



- This message type is sent periodically between peers to maintain connections and verify paths held by the router sending the keepalive.
- If a router accepts the parameters specified in its neighbor's Open message, it responds with a Keepalive.
- Subsequent Keepalives are **sent every 60 seconds** by Cisco default or equal to one-third the agreed-upon hold time (180 seconds).
- If the periodic timer is set to a value of zero (0), no keepalives are sent.

NOTIFICATION message

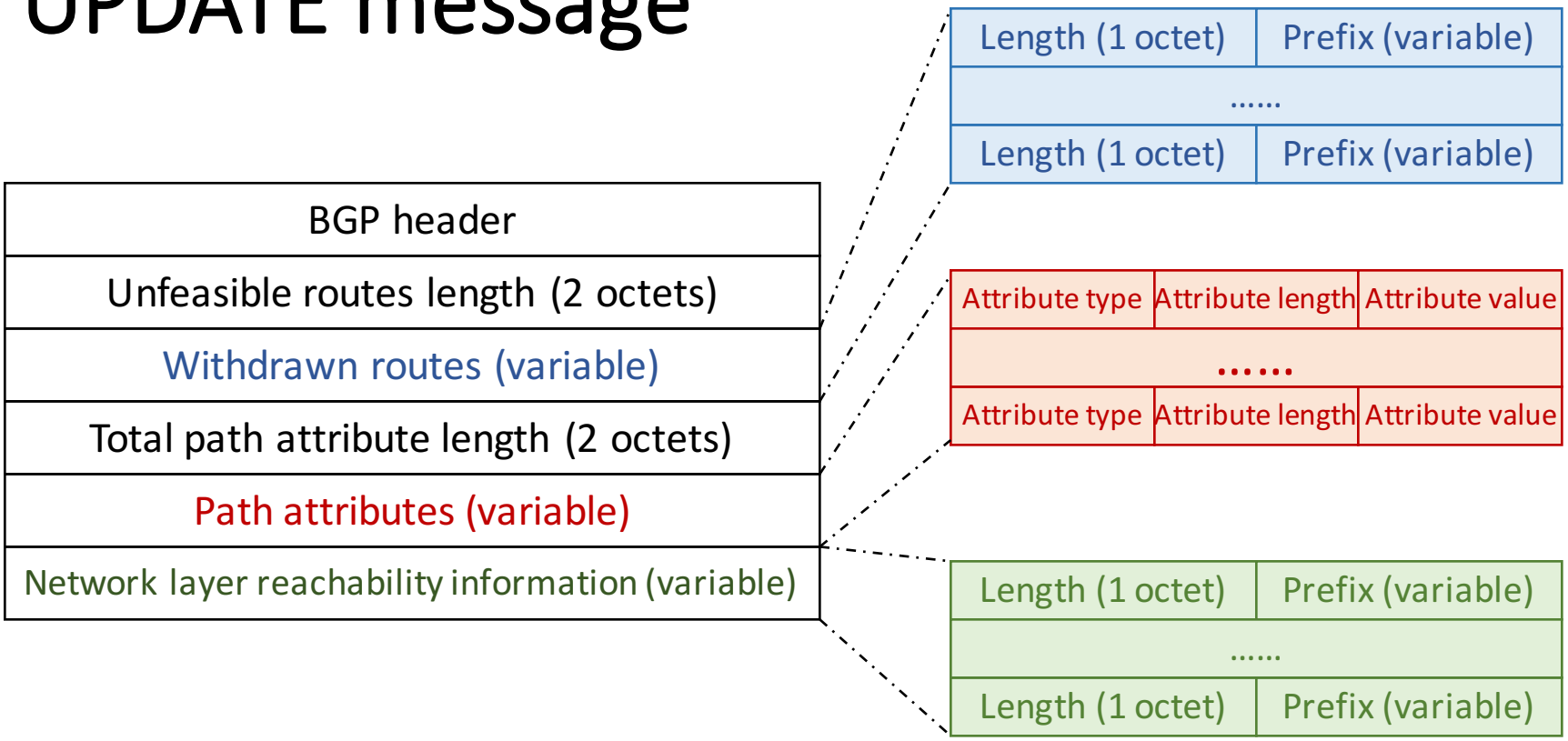


- Error code: the type of error condition.
- Error subcode: specific information about the nature of the error.
- Data: the reason for the notification.

BGP Notification Message Error Code

Type Code	Name	Sub-Code and Description
1	Message Header Error	1: Connection not synchronized 2: Bad message length 3: Bad message type
2	Open Message Error	1: Unsupported Version number 2: Bad peer AS 3: Bad BGP Identifier 4: Unsupported optional parameter 5: Authentication failure 6: Unacceptable hold time
3	Update Message Error	1: Malformed attribute list 2: Unrecognized well-known attribute 3: Mission well-known attribute 4: Attribute flags error 5: Attribute length error 6: Invalid ORIGIN attribute 7: AS Routing loop 8: Invalid NEXT_HOP attribute 9: Optional attribute error 10: Invalid network field 11: Malformed AS_PATH
4	Hold Time Expired	
5	Finite State Machine Error	
6	Cease	

UPDATE message



- Unfeasible routes length: the total length of the withdrawn routes field in octets.
- Withdrawn routes: a list of IP address prefixes for the routes that need to be withdrawn from BGP routing tables.
- Total path attribute length: the total length of the Path Attributes field in octets.
- Path attributes: a variable length sequence of path attributes.
- NLRI: a list of IP prefixes.

Type 3: BGP Update Message

Network-Layer Reachability Information (NLRI)

- This is one or more (Length, Prefix) tuples that advertise IP address prefixes and their lengths.
- 192.168.160.0/19
  - Prefix = 192.168.160.0
  - Prefix Length = 19

Path Attributes

- This is described later, providing the information that allows BGP to choose a shortest path, detect routing loops, and determine routing policy.

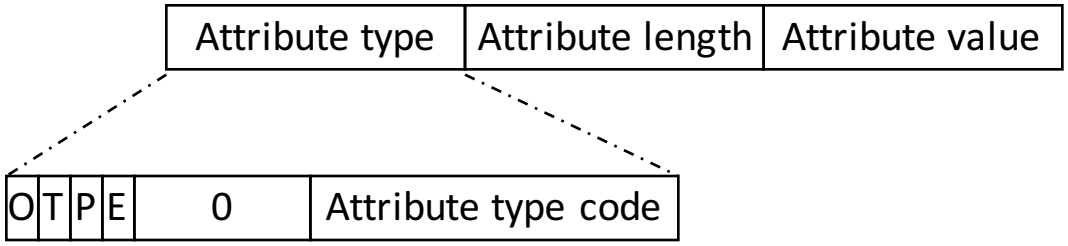
Withdrawn Routes

- These are (Length, Prefix) tuples describing destination that have become unreachable and are being withdrawn from service.

BGP Attribute Categories

- **Well-Known Mandatory** : Must be recognized and included in all BGP Update messages
- **Well-Known Discretionary**: Must be recognized but may or may not be included in a specific Update message
- **Optional Transitive**: Not required to be supported. Accept the attribute and pass on to peers
- **Optional Non-transitive**: Not required to be supported. Quietly ignore and not advertise to other peers.

Update message (cont.)



- **Attribute flag (1 octet):**
  - O bit: attribute is optional (O=1), or required (O=0).
  - T bit: an optional attribute is transitive (T=1), or non-transitive (T=1). Set to 1 for Well-known attributes
  - P bit: the information in the optional transitive attribute is partial (P=1), or complete (P=0). Set to 0 for well-known attributes and for optional non-transitive attributes
  - E bit: the attribute length is two octets (E=1), or one octet (E=0).

BGP Attributes

Type Code	Name	Value Code	Description	
1	Origin	0	IGP	Well-Known Mandatory
		1	EGP	
		2	Incomplete	
2	AS_PATH	1	AS_SET	Well-Known Mandatory
		2	AS_SEQUENCE	
		3	AS_CONFED_SET	
		4	AS_CONFED_SEQUENCE	
3	NEXT_HOP	0	Next-Hop IP Address	Well-Known Mandatory
4	MULTI_EXIT_DISC	0	4-octet MET	Optional Non-transitive
5	LOCAL_PREF	0	4-octet LOCAL_PREF	Well-Known Discretionary
6	ATOMIC_AGGREGATE	0	None	Well-Known Discretionary
7	AGGREGATOR	0	AS Number and IP Address of aggregator	Well-Known Discretionary
8	COMMUNITY	0	4-octet community identifier	Optional Transitive
9	ORIGINATOR_ID	0	4-octet router ID of originator	Optional Non-transitive
10	CLUSTER_LIST	0	Variable-length list of cluster IDs	Optional Non-transitive

Update message (cont.)

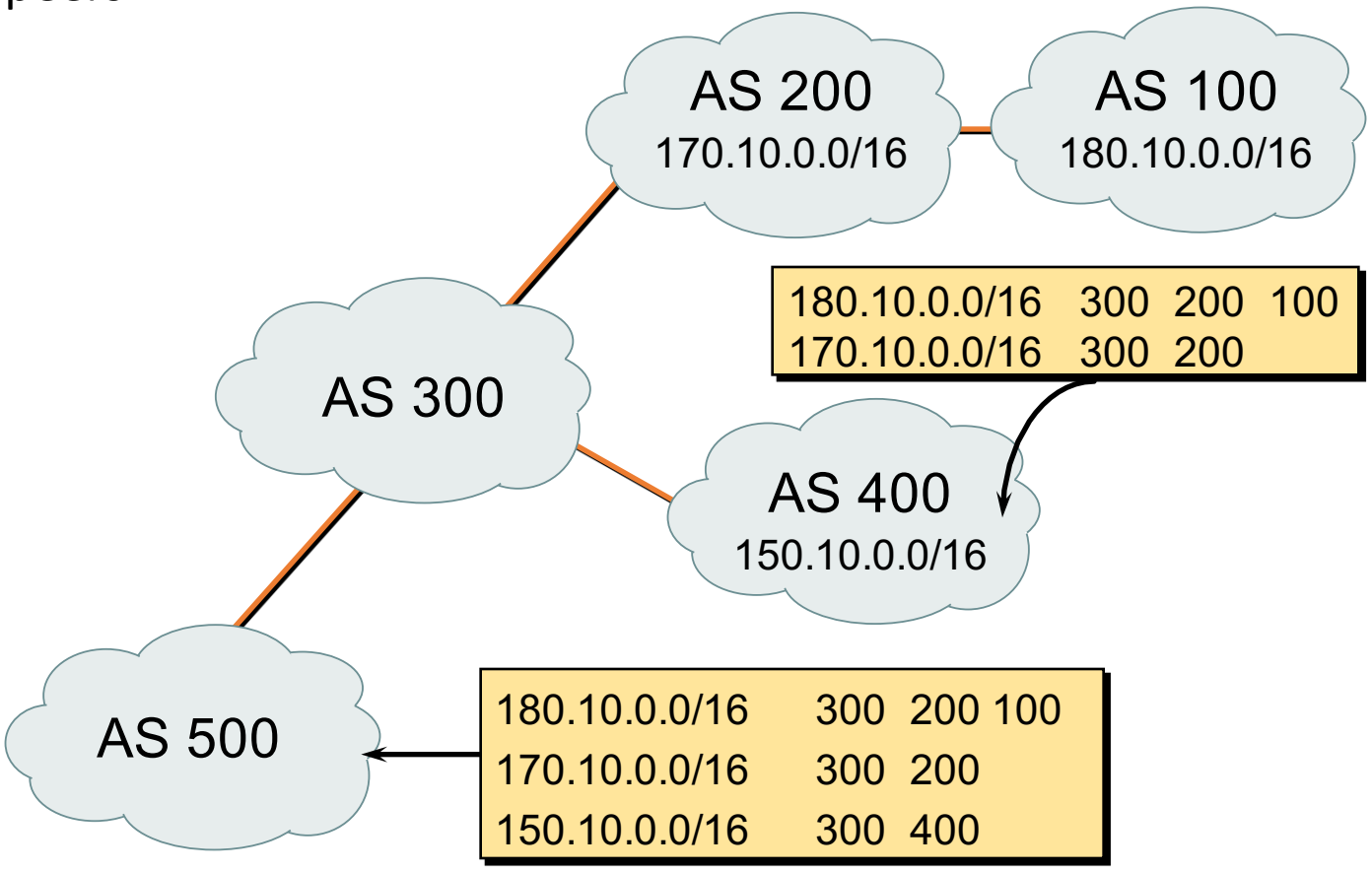
- **Attribute type code:**
  - ORIGIN (type code 1): defines the origin of the NLRI.
  - AS_PATH (type code 2): lists the sequence of Ass that the route have traversed to reach the destination.
  - NEXT_HOP (type code 3): defines the IP address of the border router that should be used as the next hop to the destination listed in the NLRI.
  - MULTI_NEXT_DISC (type code 4): discriminates among multiple entry/exit points to a neighboring AS and gives a hint to the neighboring AS about the preferred path.
  - LOCAL_PREF(type code 5): informs other BGP routers within the same AS of its degree of preference for an advertised route.
  - ATOMIC_AGGREGATE (type code 6): informs other BGP routers that it selected a less specific route without selecting a more specific one that is included in it.
  - AGGREGATOR (type code 7): specifies the last AS number that formed the aggregate route followed by the IP address of the BGP router that formed the aggregate route.

The ORIGIN Attribute

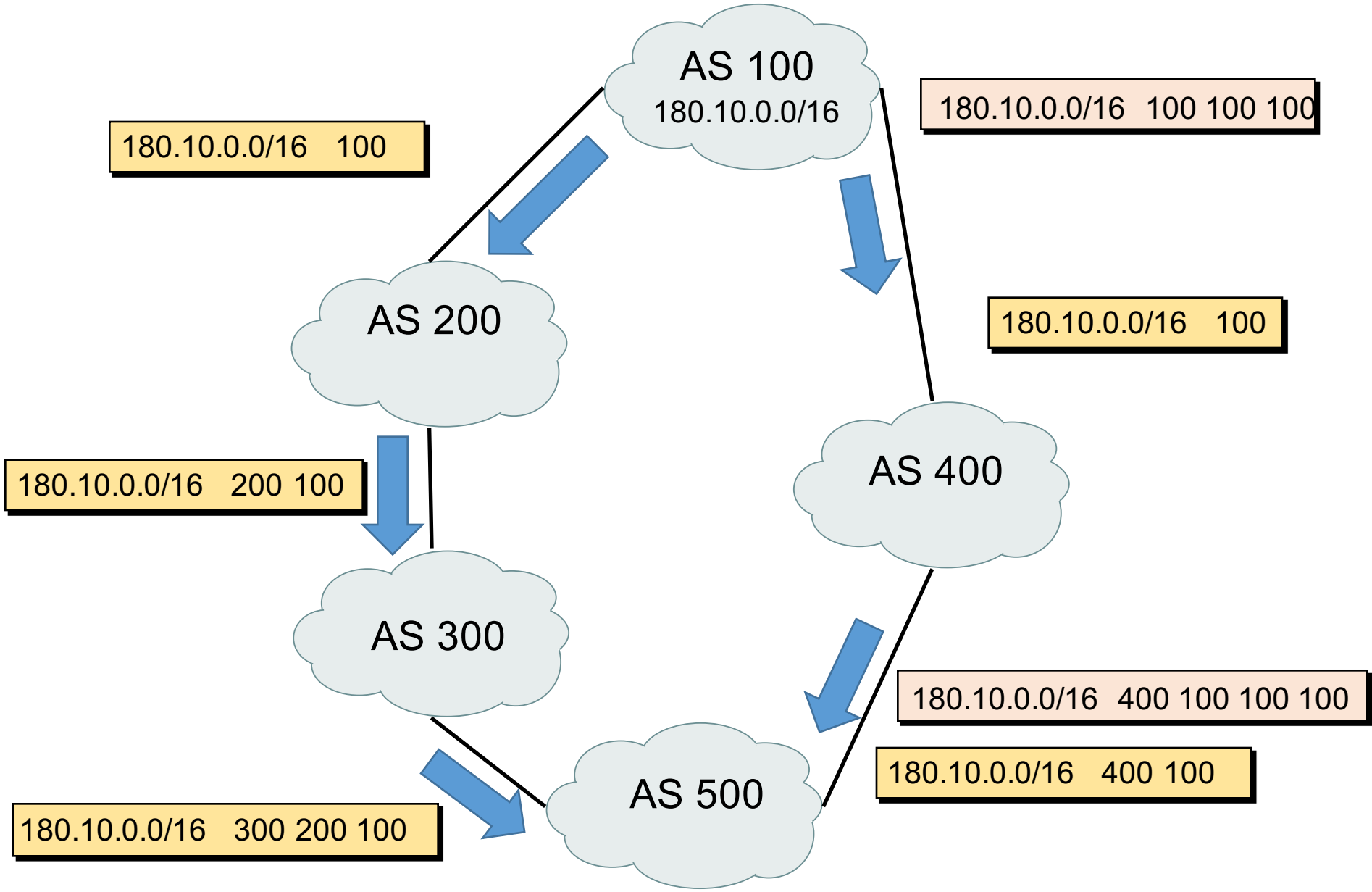
- **IGP** – The router was learned from a protocol internal to the originating AS.
- **EGP** – The route was learned from the Exterior Gateway Protocol
- **Incomplete** – The router was learned by some other means. Routes learns through redistribution has the incomplete origin attribute, because there is no way to determine the original source of the route.

The AS-Path Attribute

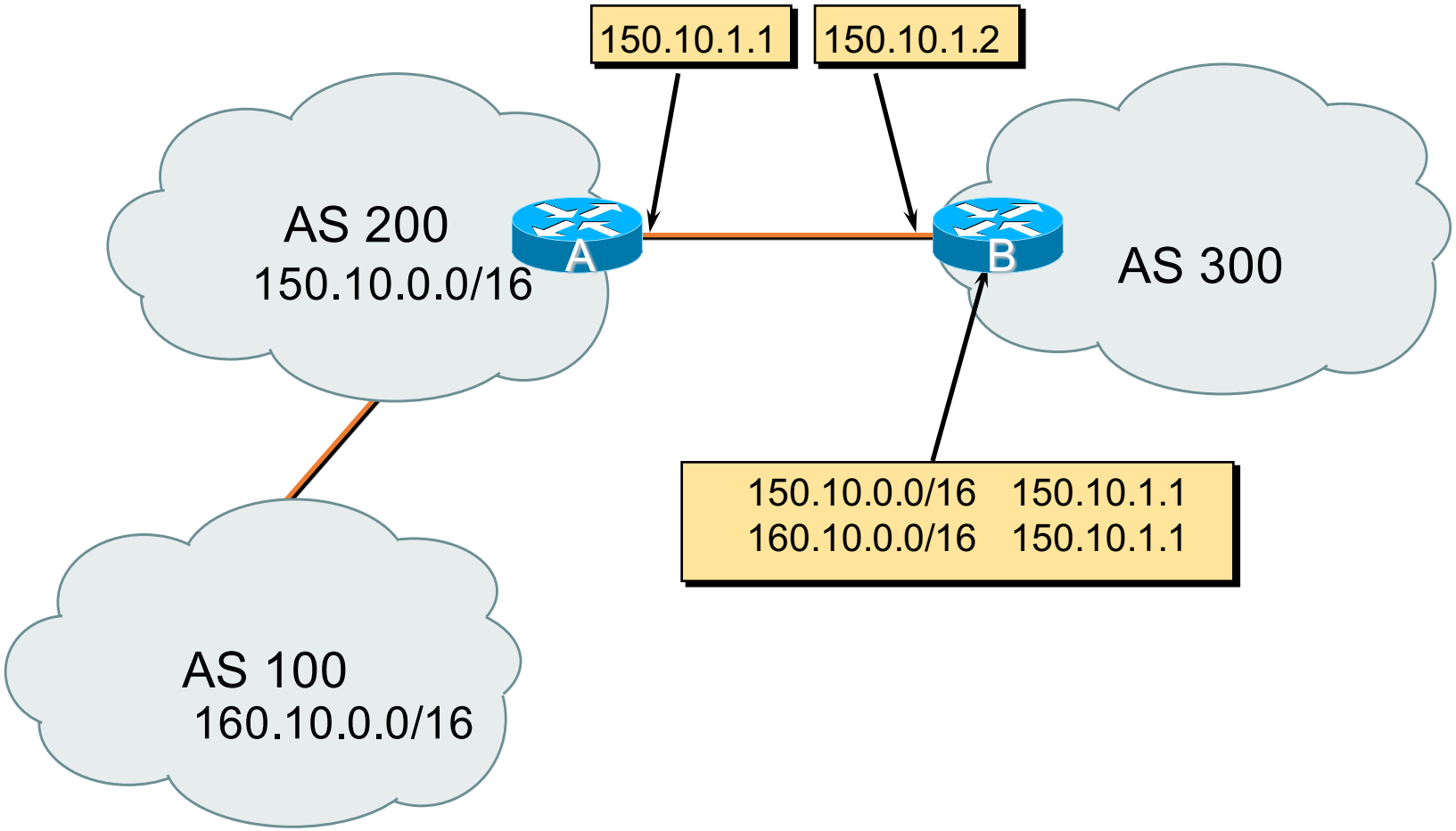
- Sequence of ASs a route has traversed – Prepended only when the route is advertised between EBGP peers
- Loop detection
- Apply policy



The AS-Path Attribute



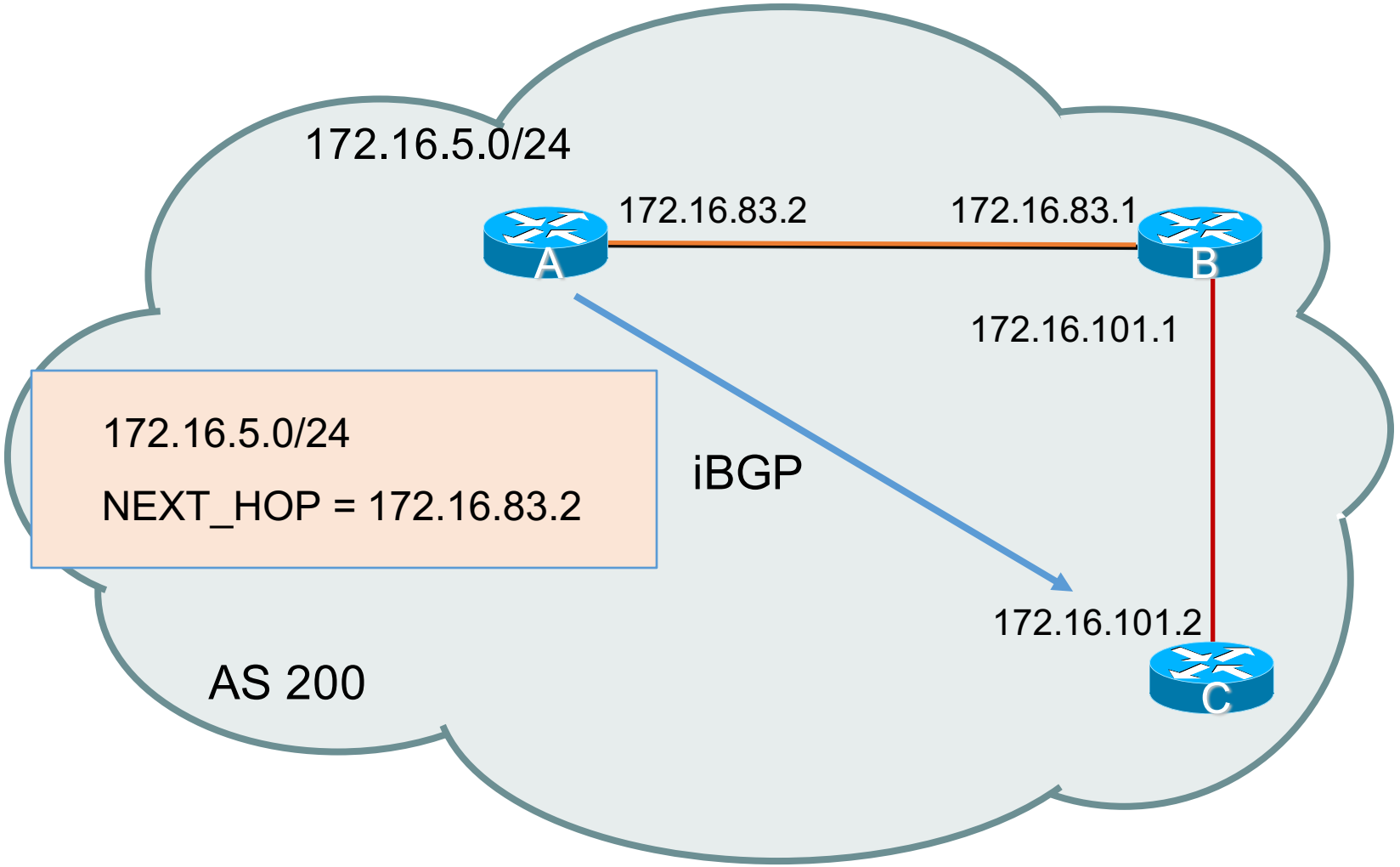
The Next Hop Attribute



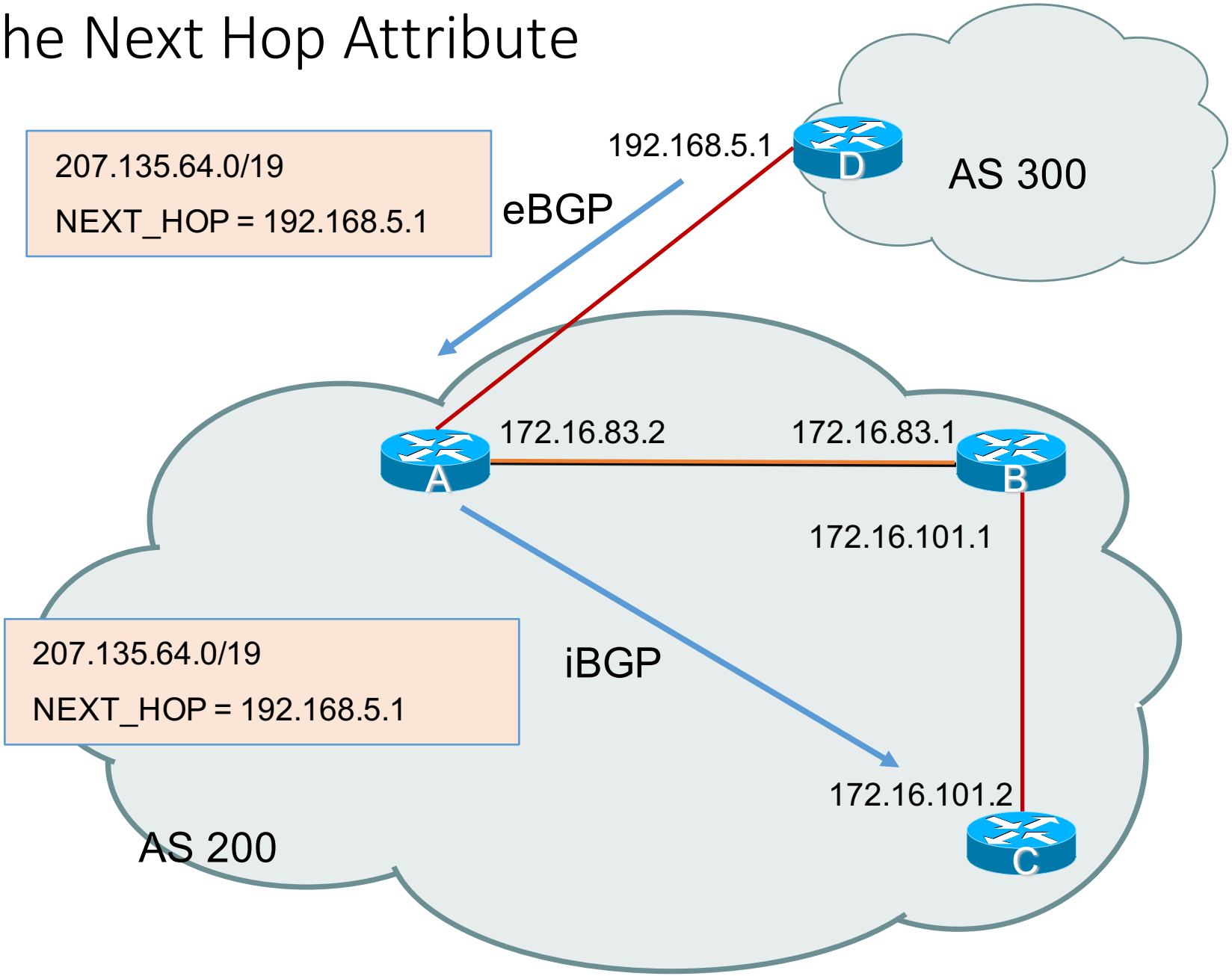
Next Hop (continued)

1. Advertising to different AS – the NEXT_HOP is the IP address of the advertising router's interface
2. Advertising to the same AS and the route is in the same AS – NEXT_HOP is the IP address of the originating router
3. Advertising to the same AS and the route is in a different AS – NEXT_HOP is the IP address of the external peer from which the route was learned

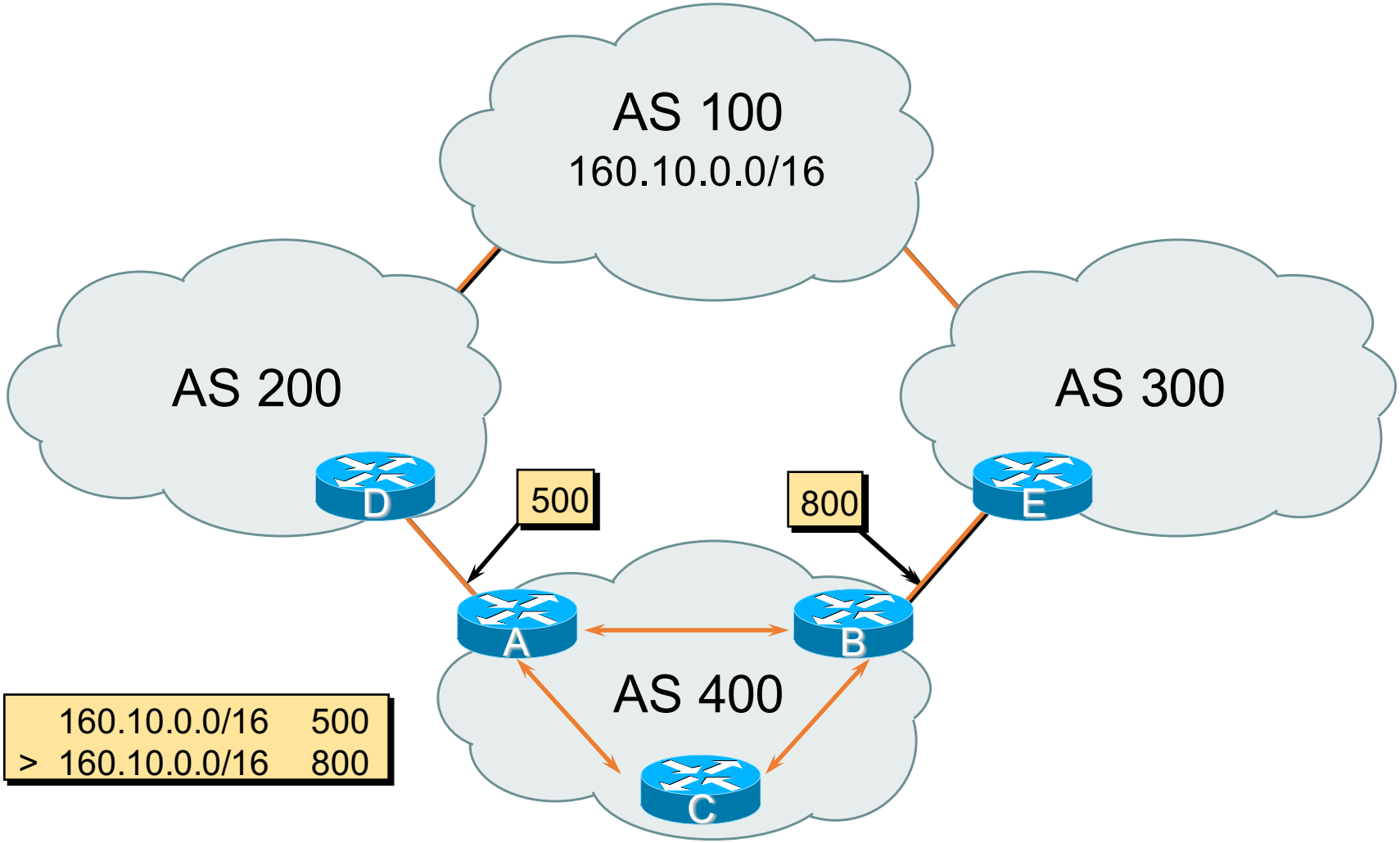
The Next Hop Attribute



The Next Hop Attribute



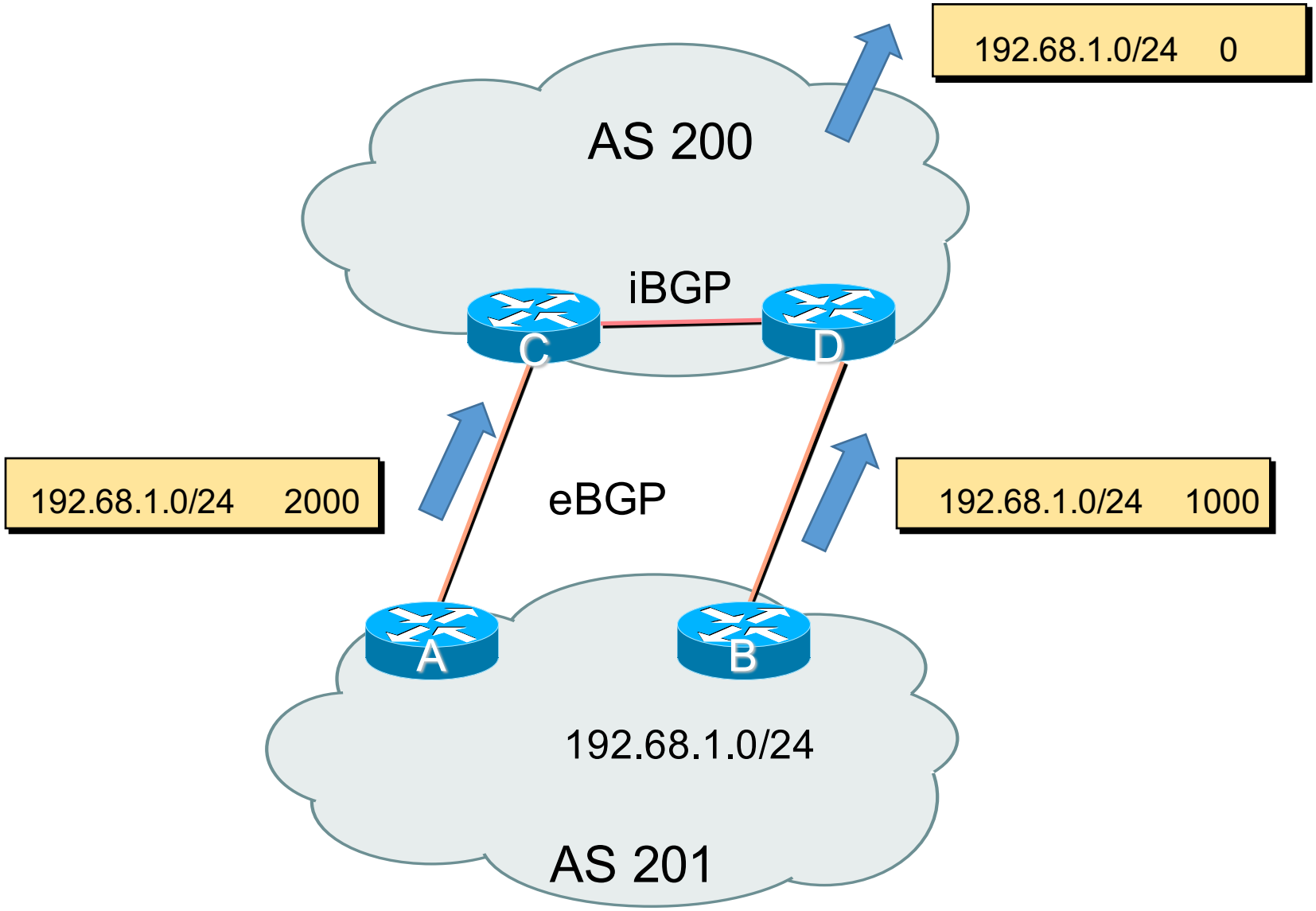
The Local Preference Attribute



The Local Preference Attribute

- Local to an AS – non-transitive
  - Only between iBGP peers
  - Local preference set to 100 when heard from neighbouring AS
- Used to influence BGP path selection
  - Determines best path for outbound traffic
- Path with highest local preference wins

The Multi-Exit Discriminator (MED) Attribute



The Multi-Exit Discriminator Attribute

- Inter-AS Metric: Optional non-transitive
  - Carried in eBGP updates to inform another AS of the preferred ingress point
  - Lowest MED value is preferred

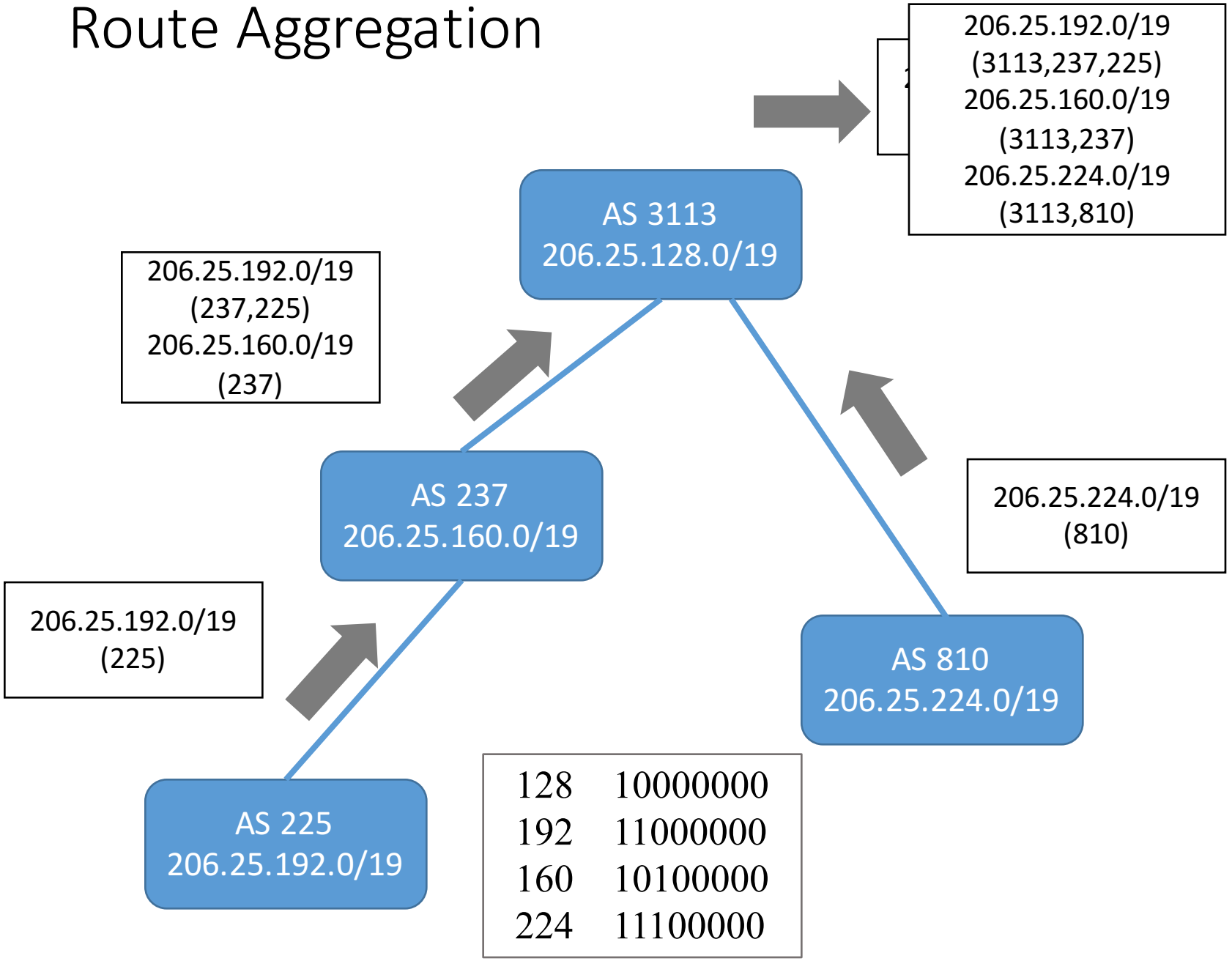
- Used only to influence traffic between two directly connected AS
  - Not passed beyond the receiving AS
  - Metric reset to 0 on announcement to next AS

- Comparable if paths are from same AS

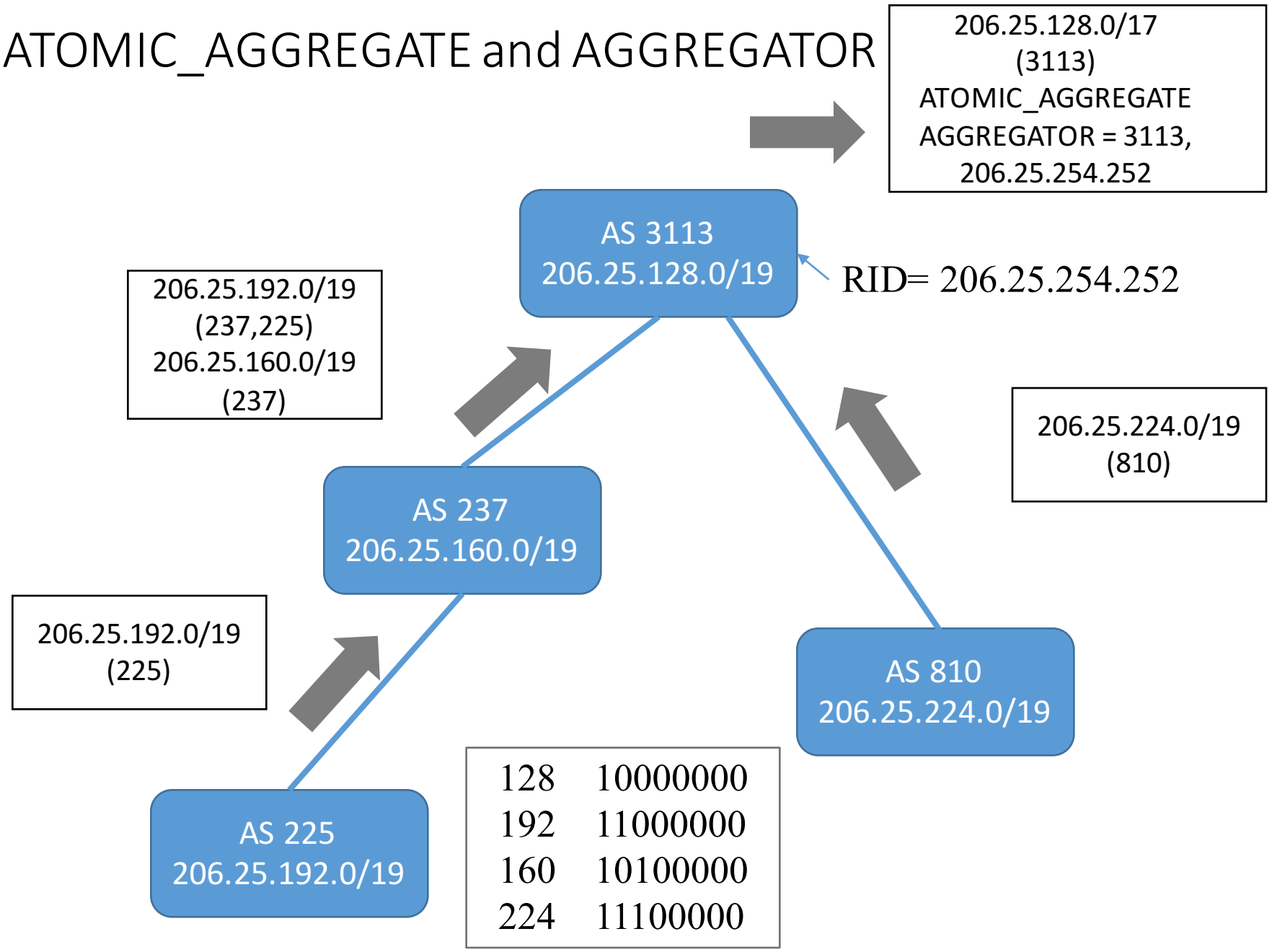
The ATOMIC_AGGREGATE and AGGREGATOR Attributes

- The **ATOMIC_AGGREGATE** is a well-know discretionary attribute. Used to alert downstream routers that a loss of path information has occurred.
- Receiving router must attach the ATOMIC_AGGREGATE attribute
- The **AGGREGATOR** attribute is optional transitive. Provide information about where the aggregation was performed by including the AS number and the Router ID that originated the aggregate route.

Route Aggregation



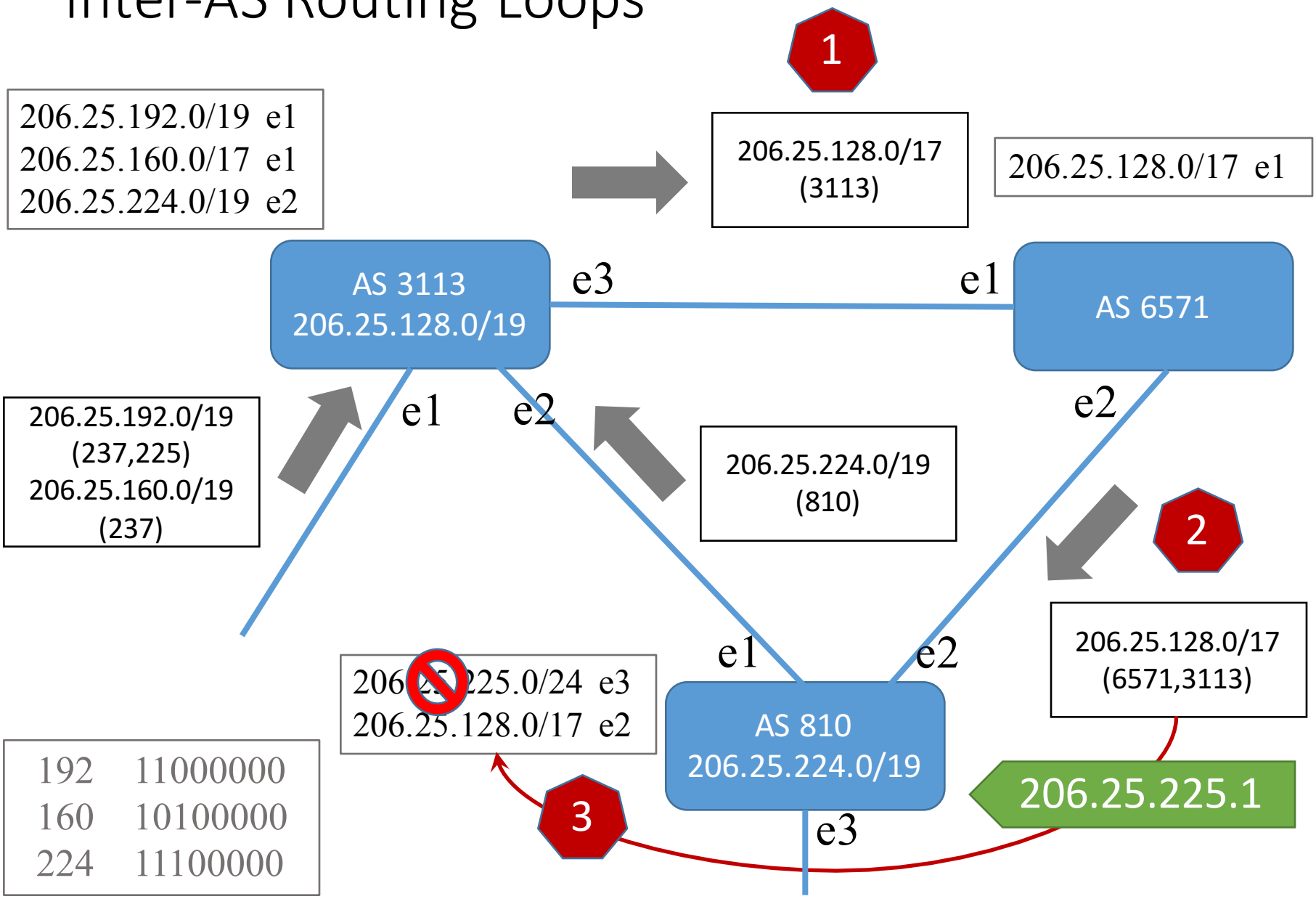
ATOMIC_AGGREGATE and AGGREGATOR



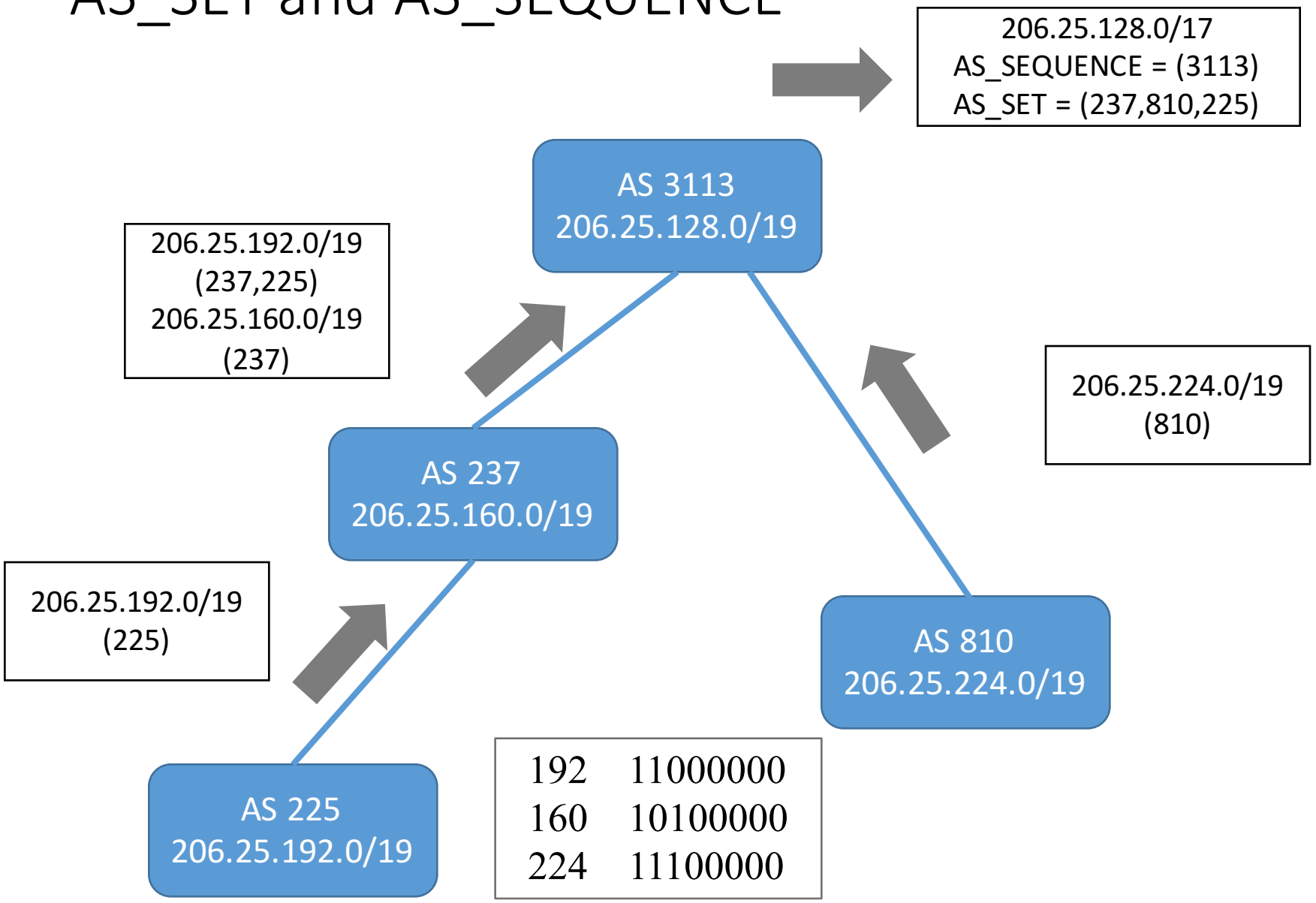
AS_SET and AS_SEQUENCE

- AS_SEQUENCE: Ordered list of AS numbers
- AS_SET: Unordered list of the AS numbers
  - When a BGP router aggregates routes learned from other AS, it can include all those AS numbers in the AS_PATH as an AS_SET

Inter-AS Routing Loops



AS_SET and AS_SEQUENCE

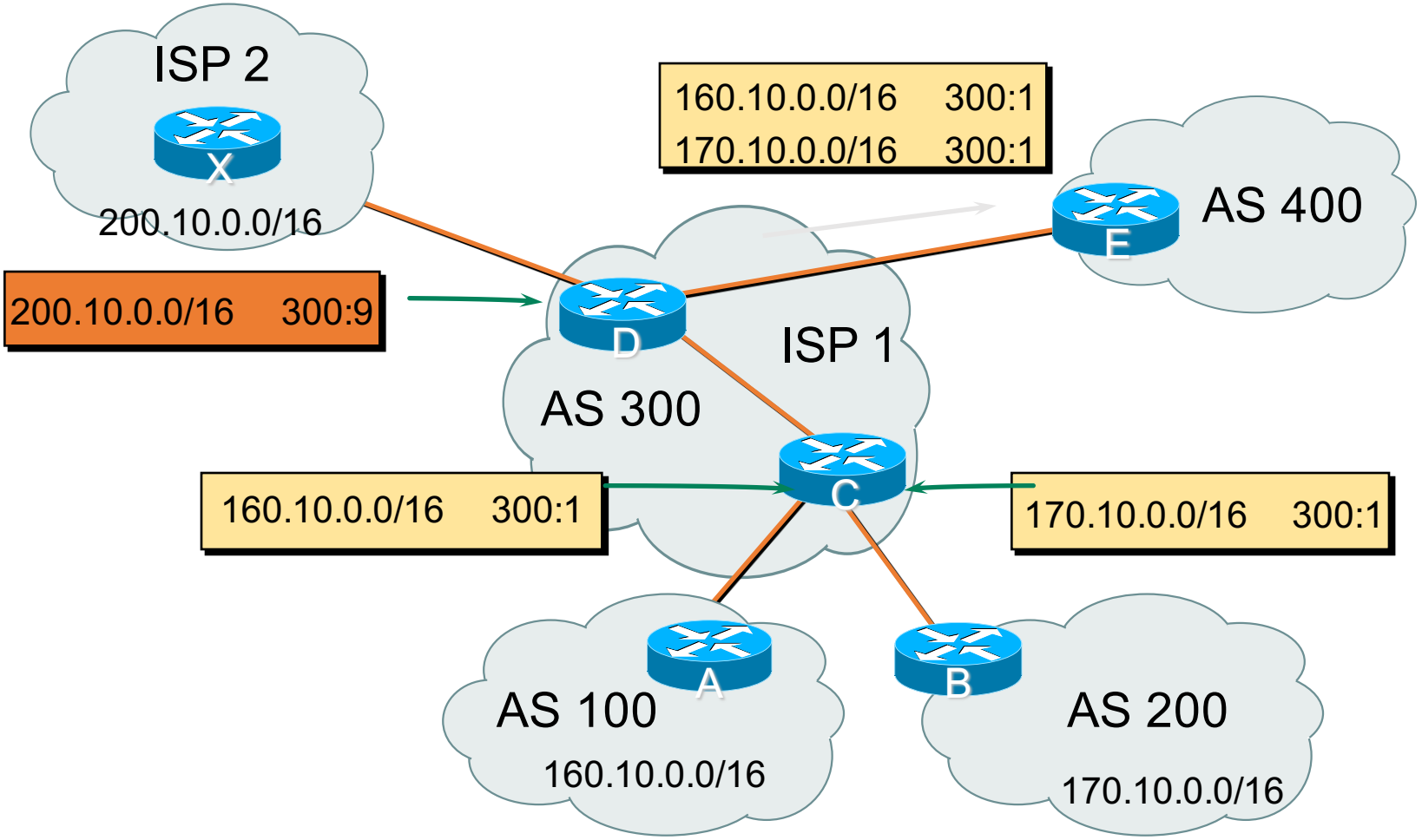


The COMMUNITY Attribute

put the LABEL on the route. (tag)
says the route belong to a community
the receiving party use the label for filter

- Optional transitive attribute to simplify policy enforcement
- Originally Cisco-specific. Standardized in RFC 1997.
- Identify a destination as a member of a community
- Represented as two 16bit integers AA:NN (AA is the AS Number)
- Well-known Communities:
  - INTERNET : no value. Can be advertised freely
  - NO_EXPORT (0xFFFFFF01) : cannot be advertised outside the confederation
  - NO_ADVERTISE (0xFFFFFF02) : cannot be advertised at all
  - LOCAL_AS (0xFFFFFF03) : Cannot be advertised to EBGP peers

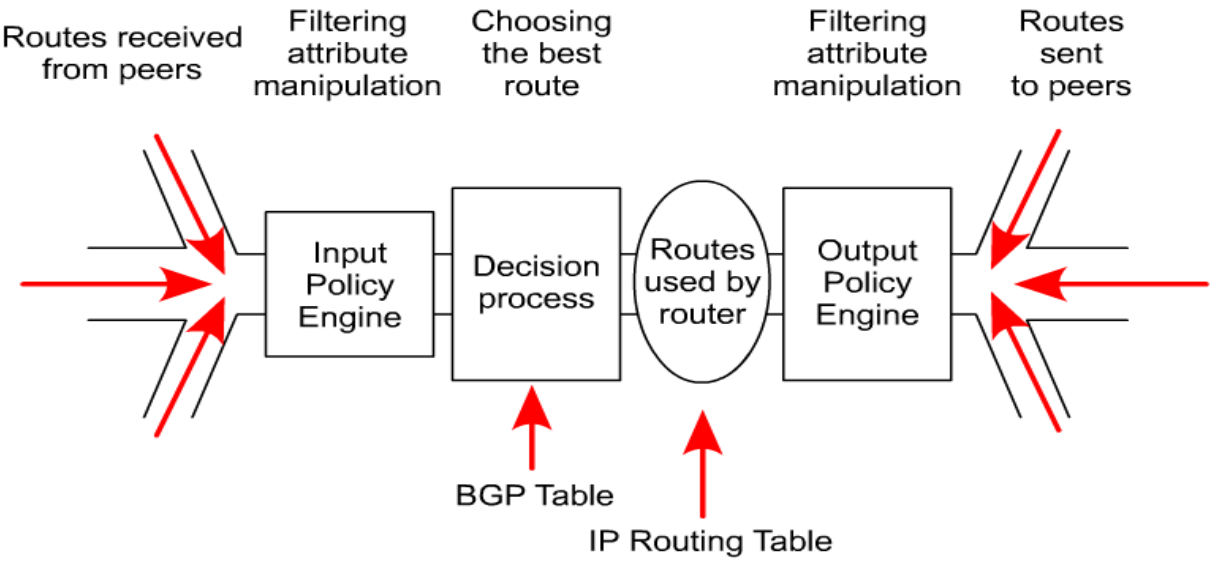
Community


ISP 2
X
200.10.0.0/16
200.10.0.0/16 300:9
160.10.0.0/16 300:1
170.10.0.0/16 300:1
AS 400
E
D
ISP 1
AS 300
C
160.10.0.0/16 300:1
170.10.0.0/16 300:1
A
B
AS 100
160.10.0.0/16
AS 200
170.10.0.0/16

The ORIGINATOR_ID and CLUSTER_LIST Attributes

- Optional non-transitive attributes used by route reflectors to prevent routing loops
- ORIGINATOR_ID: Router ID of the originator of the route in the local AS
- CLUSTER_LIST: A sequence of route reflector cluster IDs through which the route has traversed.

BGP Routing



- BGP is so flexible because it is a fairly simple protocol.
- Routes are exchanged between BGP peers via UPDATE messages.
- BGP routers receive the UPDATE messages, run some policies or filters over the updates, and then pass on the routes to other BGP peers.

BGP Route Selection Algorithm

Summary of the BGP Path Selection Process

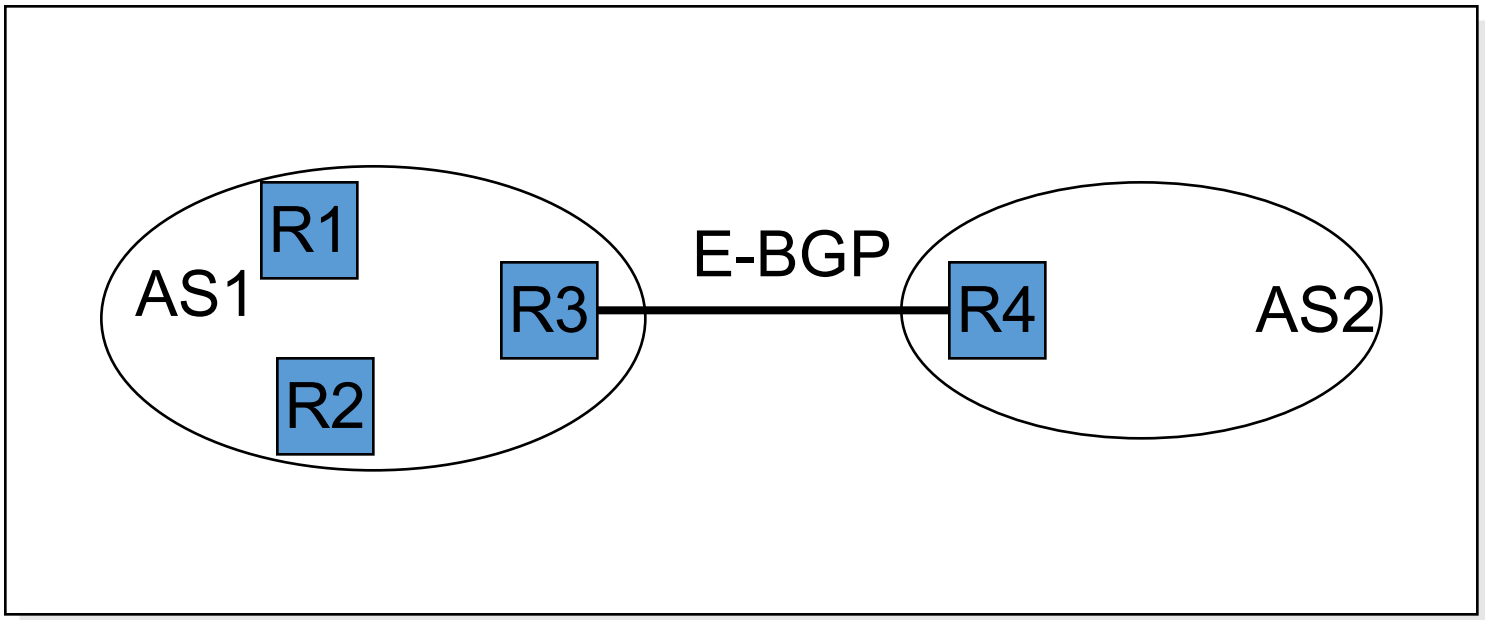
- BGP selects only one path as the best path.
- When the path is selected, BGP puts the selected path in its routing table and propagates the path to its neighbors.
- BGP uses the following criteria, in the order presented, to select a path for a destination:

must be available

1. If the path specifies a next hop that is inaccessible, drop the update
2. Prefer the path with the **largest weight**.
3. If the weights are the same, prefer the path with the **largest local preference**.
4. If the local preferences are the same, prefer the **path that was originated by BGP** running on this router.
5. If no route was originated, prefer the route that has the **shortest AS_path**.
6. If all paths have the same AS_path length, prefer the path with the **lowest origin** type (where IGP is lower than EGP, and EGP is lower than Incomplete).
7. If the origin codes are the same, prefer the path with the **lowest MED attribute**.
8. If the paths have the same MED, prefer the **external path** over the internal path.
9. If the paths are still the same, prefer the path through the **closest IGP neighbor**.
10. Prefer the path with the lowest IP address, as specified by the BGP **router ID**.

Internal vs. External BGP

- BGP can be used by R3 and R4 to learn routes
- How do R1 and R2 learn routes?



Internal BGP (I-BGP)

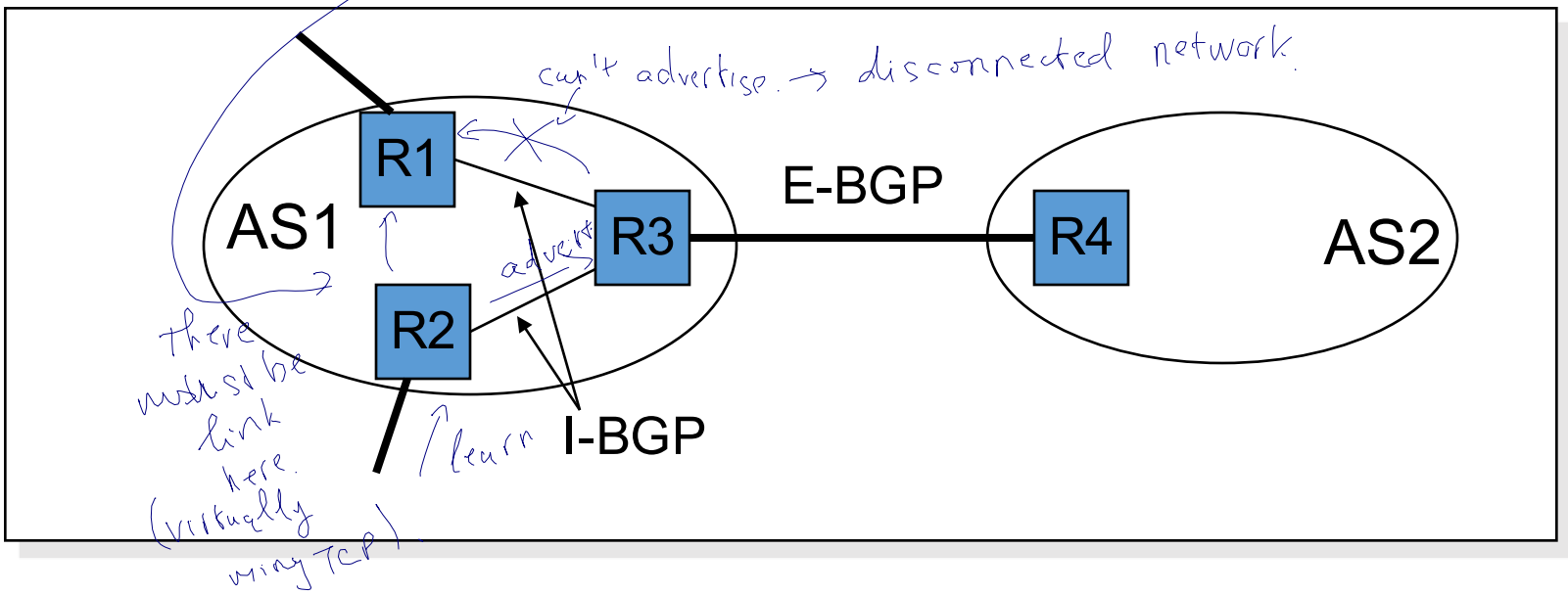
- Same messages as E-BGP
- Different rules about re-advertising prefixes:
  - Prefix learned from E-BGP can be advertised to I-BGP neighbor and vice-versa, but
  - Prefix learned from one I-BGP neighbor cannot be advertised to another I-BGP neighbor
  - Reason: no AS PATH within the same AS and thus danger of looping.

Internal BGP (I-BGP)

- R3 can tell R1 and R2 prefixes from R4
- R3 can tell R4 prefixes from R1 and R2
- R3 cannot tell R2 prefixes from R1

R2 can only find these prefixes through a *direct connection* to R1
Result: I-BGP routers must be fully connected (via TCP)!

- contrast with E-BGP sessions that map to physical links



Reasons For Fully Meshed iBGP

- To prevent BGP routing loops within an AS
- To ensure that all routers along the path of a BGP route know how to forward packets to the destination

Route Flap Dampening

- Route flap
  - Going up and down of path or change in attribute
    - BGP WITHDRAW followed by UPDATE = 1 flap
    - eBGP neighbour going down/up is NOT a flap
  - Ripples through the entire Internet
  - Wastes CPU
- Route Dampening is to To limit the propagation of flapping routes. Advertise stable routes only
- Described in RFC2439

Route Dampening

- Route Dampening Parameters:
  - **History state:** After a single route flap, the route is assigned a penalty, and the dampening state of the route is set to History. Each time the route flaps, the penalty increases.
  - **Suppress limit:** If the penalty exceeds the suppress limit, the route is Dampened. When the route is in Damp state, the router is not considered for best path selection and is not advertised to BGP peers.
  - **Half life:** The penalty of a route is decreased based on the half-life period. The default half-life period is 15 minutes. The penalty on the route is reduced every 5 seconds.
  - **Reuse limit:** When the penalty falls below the reuse limit, the route is unsuppressed. When the penalty falls below half of reuse limit, the history of the route is cleared.
  - **Maximum suppress limit:** It is the maximum amount of time a route can be suppressed for.
  - Cisco Default: Penalty = 1000 per flap. Suppress limit = 2000, half-life period = 15 minutes, reuse-limit = 750 and maximum suppress-limit = 60 minutes.

Operation

- Add penalty (1000) for each flap
  - Change in attribute gets penalty of 500
- Exponentially decay penalty

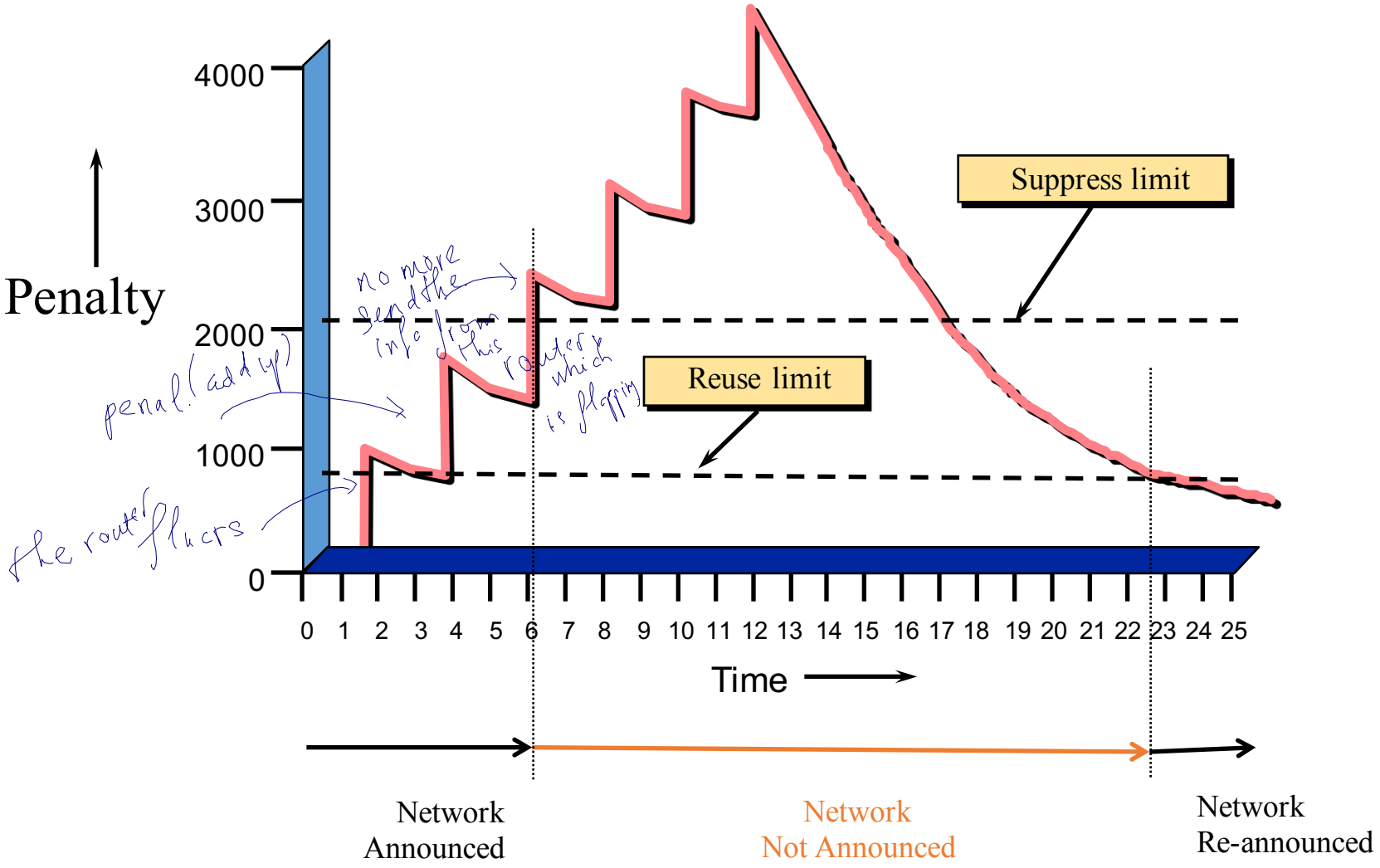
  half life determines decay rate
- Penalty above suppress-limit

  do not advertise route to BGP peers
- Penalty decayed below reuse-limit

  re-advertise route to BGP peers

  penalty reset to zero when it is half of reuse-limit

Operation

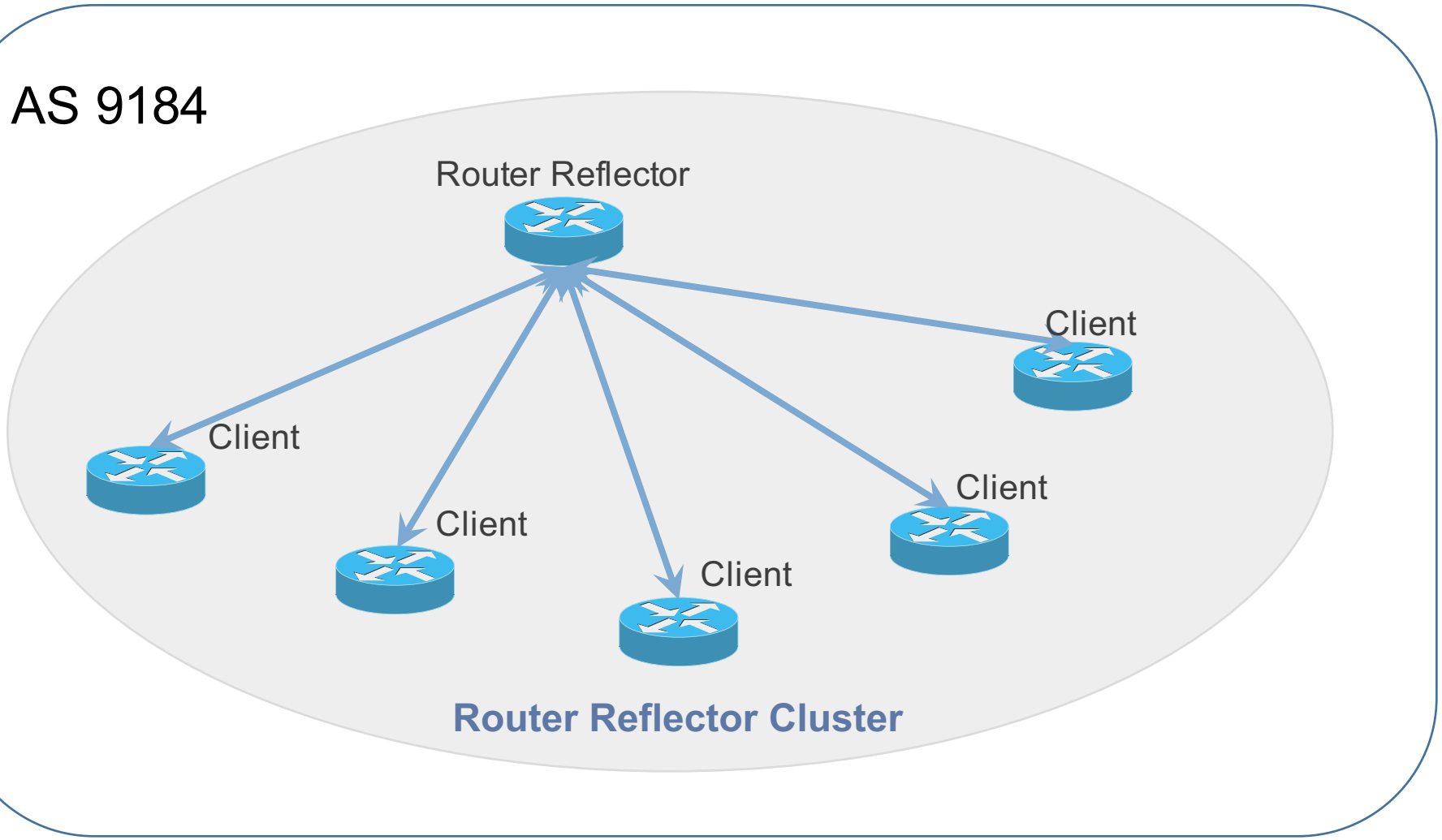

4000
3000
2000
1000
0
Penalty
Suppress limit
Reuse limit
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Time
Network Announced
Network Not Announced
Network Re-announced

Route Reflector

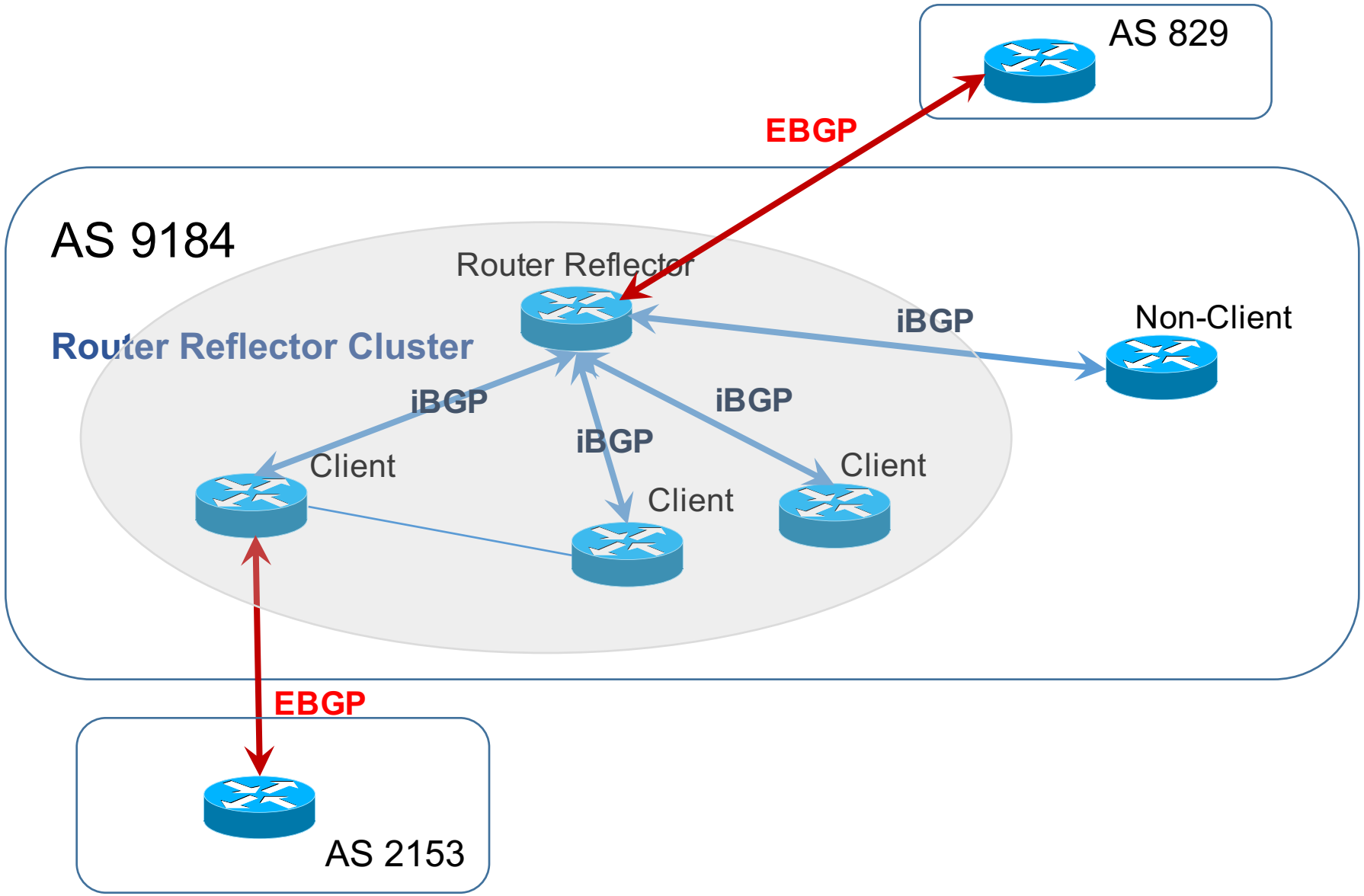
- All iBGP peers must be fully meshed which requires n(n-1)/2 iBGP connections in the AS.
- A router is configured as a route reflector (RR)
- Other iBGP routers (clients) peer with the RR only
- Rules to Advertise
  - Routes learned from a non-client iBGP peer – Advertise to clients only
  - Routes learned from a client – Advertise to all non-clients and clients, except the originating client
  - Routes learned from an eBGP peer – Advertise to all clients and non-clients

Router Reflector

reduce the comms between each router, send to reflector & reflector send to every one else

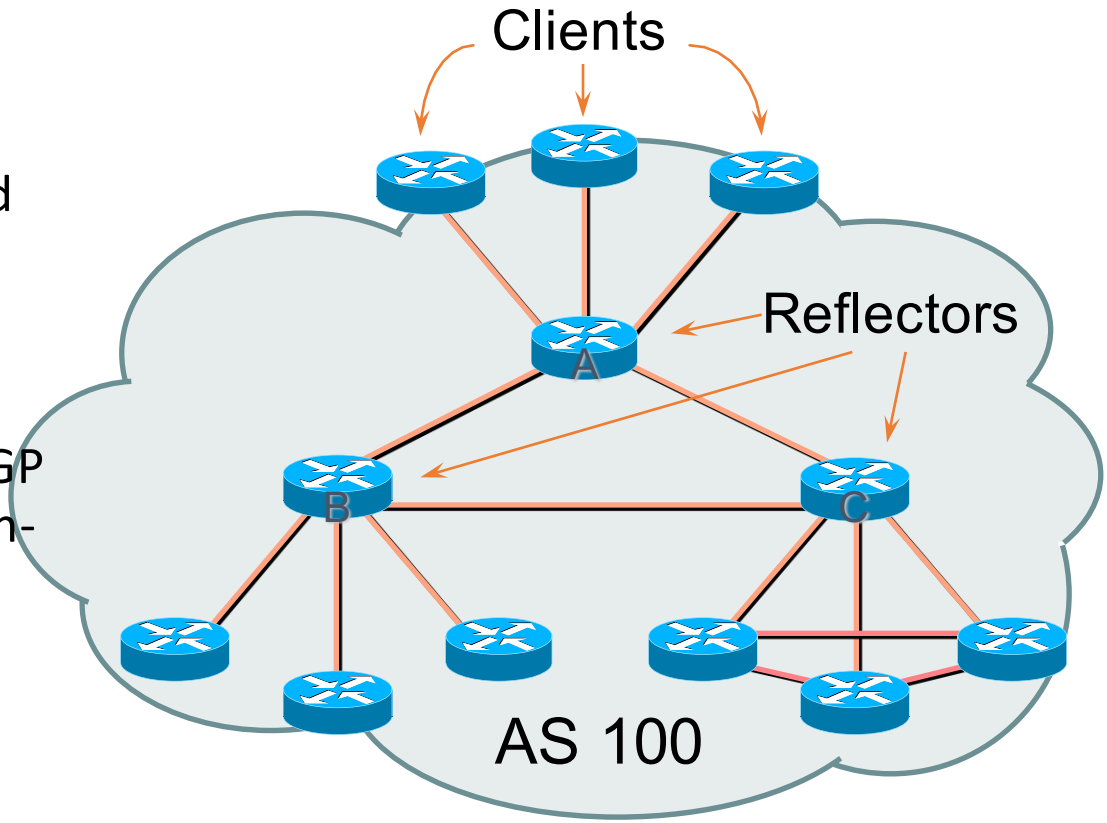


Router Reflection Cluster Peering Relationships



Route Reflector

- Reflector receives path from clients and non-clients
- Selects best path
- If best path is from client, reflect to other clients and non-clients
- If best path is from non-client, reflect to clients only
- If a route is received from an EBGP peer, reflect to all clients and non-clients
- Described in RFC2796



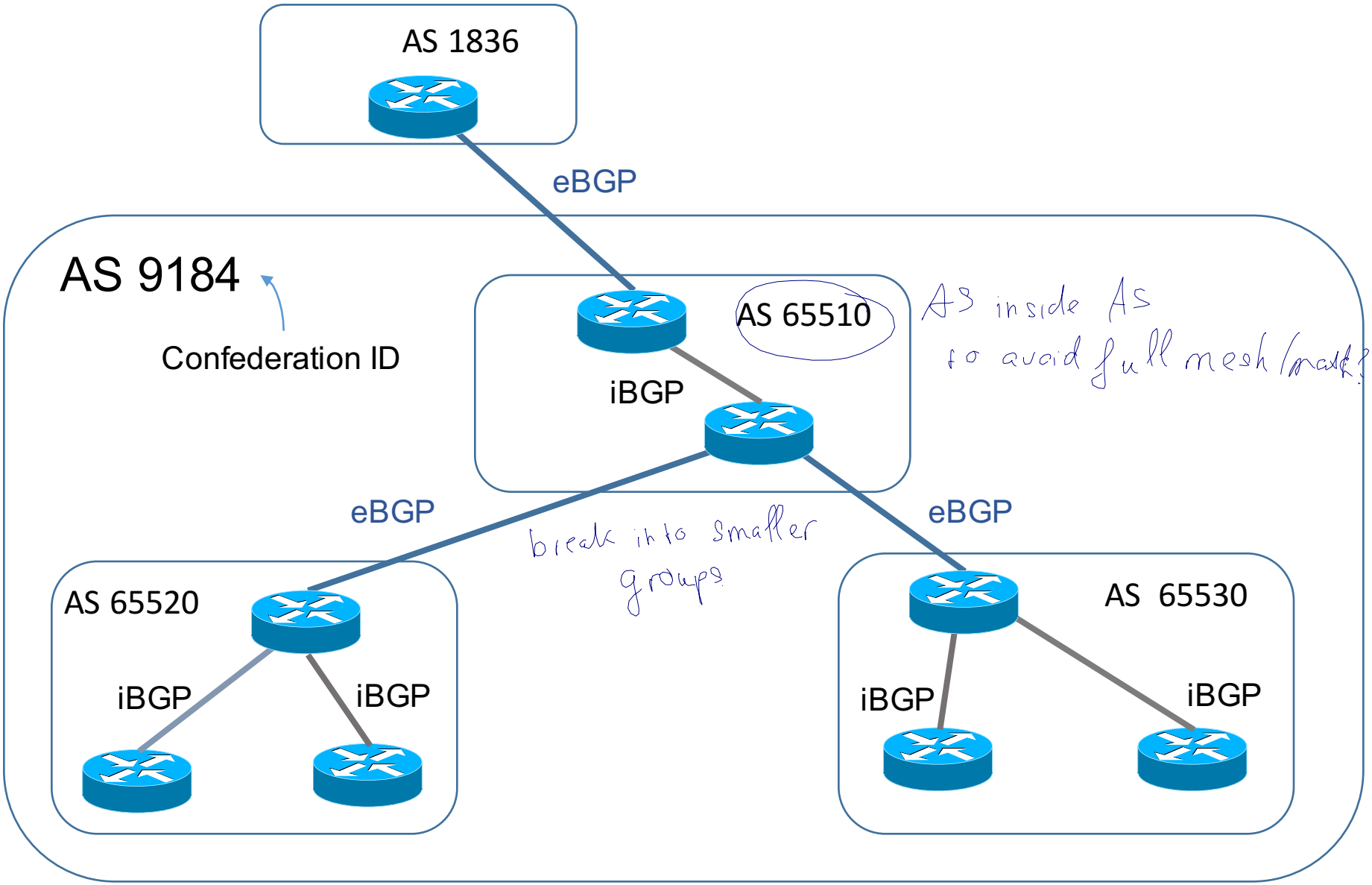
Route Reflectors: Loop Avoidance

- Originator_ID attribute
  - Carries the RID of the originator of the route in the local AS (created by the RR)
    - An RR does not advertise a route back to the originator
    - If the originator receives an update with its own RID, the update is ignored
- Cluster_list attribute
  - The local cluster-id is added when the update is sent by the RR
  - Cluster-id is router-id (address of loopback)
    - When an RR receives an update, it checks the CLUSTER_LIST. Ignores the update if it sees its own cluster ID in the list

Confederations

- Member Autonomous Systems: An AS subdivided into a group of sub-autonomous systems. Usually use AS number in the reserved range 64512 to 65535.
- iBGP to peers in the same member AS and eBGP to peers in other member AS
- Confederation ID: the AS number of the entire confederation which is represented to peers outside of the confederation as the AS number of the entire confederation.
- AS_CONFED_SEQUENCE: ordered list of AS numbers within the confederation
- AS_CONFED_SET: Unordered list of AS numbers within the confederation
- When update sent to a peer external to the confederation, the AS_CONFED_SEQUENCE and AS_CONFED_SET is stripped, and the confederation ID is prepended. External peers see the confederation as a single AS.
- Choosing a route: EBGP external to the confederation > EBGP to member AS > IBGP

Confederation



Inter-AS Routing Loops