Routing Protocol - BGP

BGP

Enterprise Network

BGP

ISP

AS 3000

AS 2000

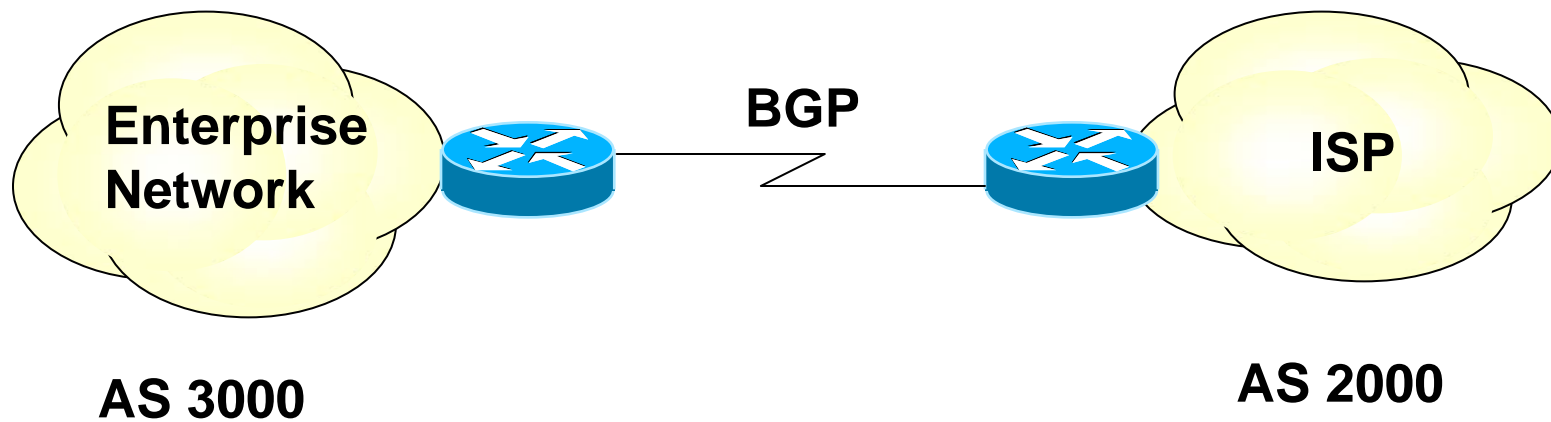❑ **BGP is using between Autonomous Systems**

BGP(cont.)

- RFC 1771(BGPv4)
  - Support CIDR
- Transfer the AS information to reach destination
- Using TCP(Port 179)
- Incremental Update
- keepalive

Why BGP

- Suitable to large network

- Policy based routing

- Path for IN/OUT traffic

- Filtering

Autonomous System(AS)

- The group of routers and networks under same routing administration
- The need of AS
  - Maintain independence of routing policy
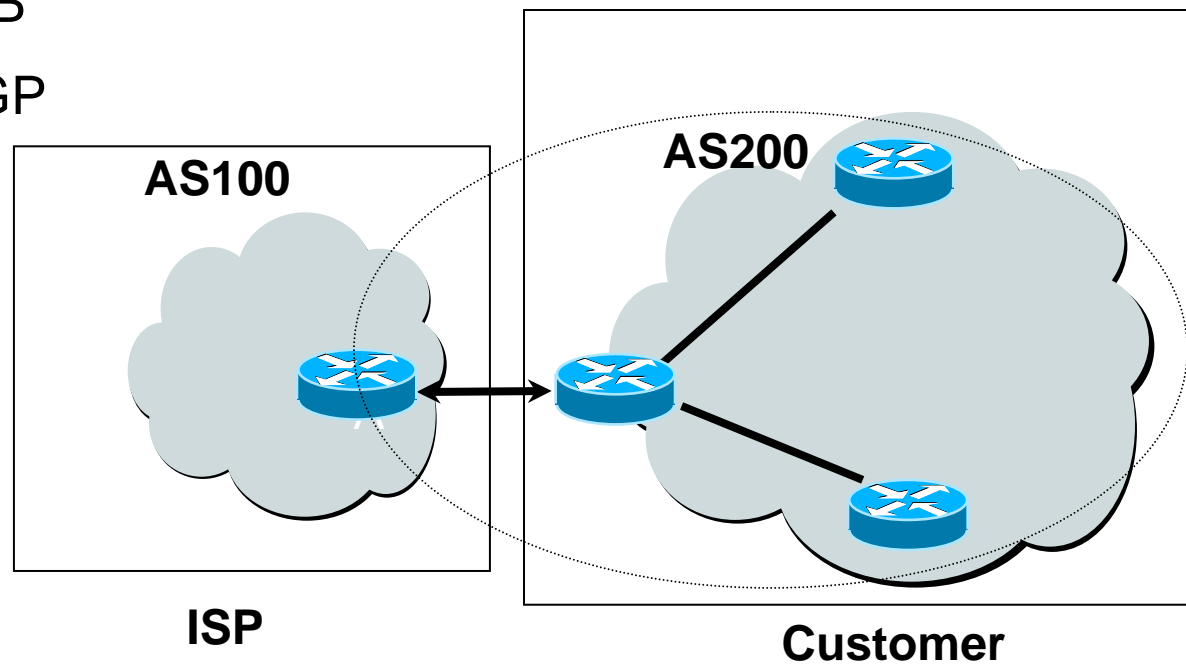 - Localize the errors or faults of specific network

Autonomous System(AS)

- Interior Gateway Protocol
  - ◆ Used inside AS
  - ◆ RIP, OSPF, IS-IS, IGRP, etc
- Exterior Gateway Protocol
  - ◆ Used between ASs
  - ◆ EGP, BGP
- The scope of AS number
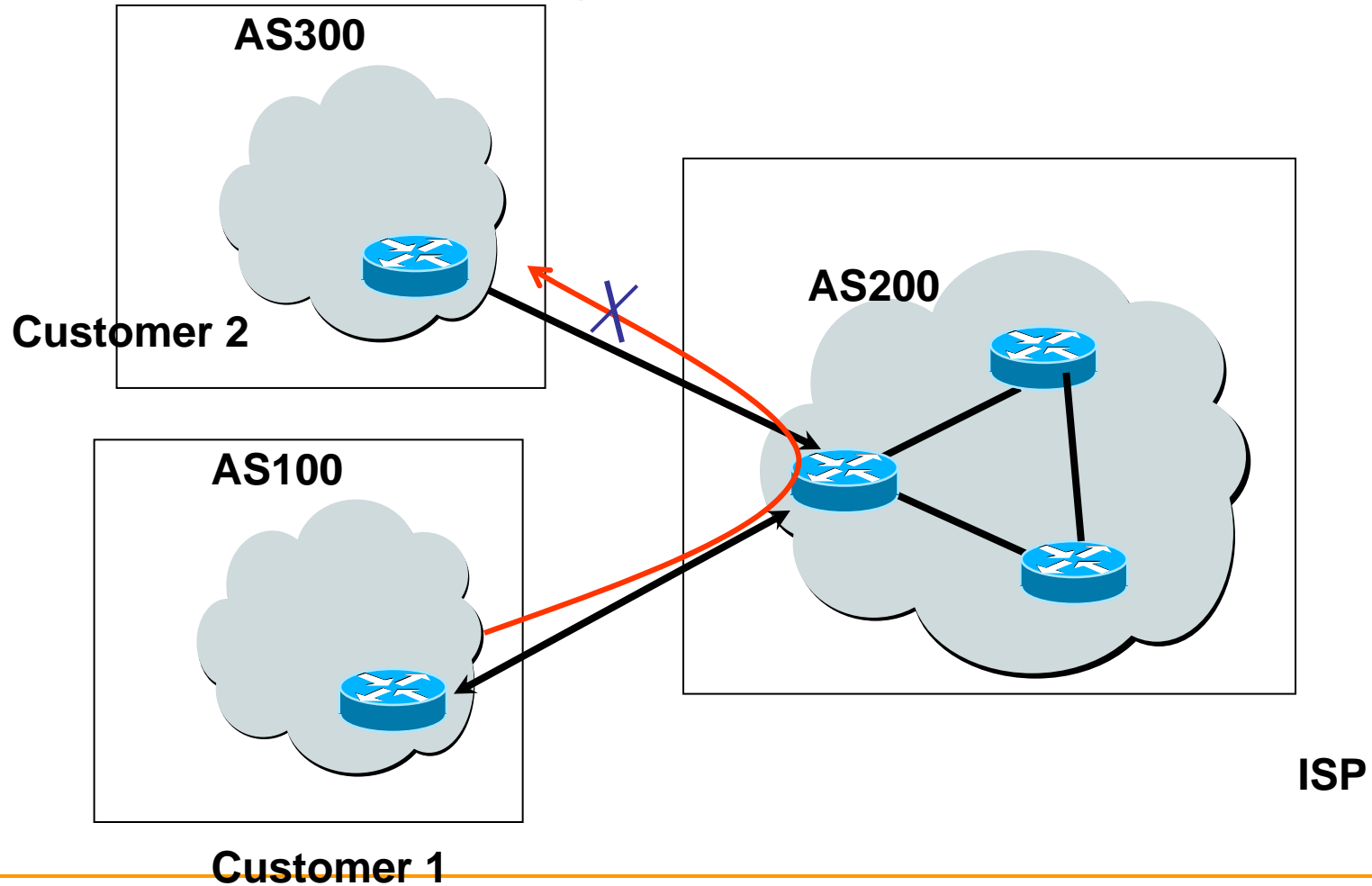 - ◆ 1-65535(64512- 65535 for private)

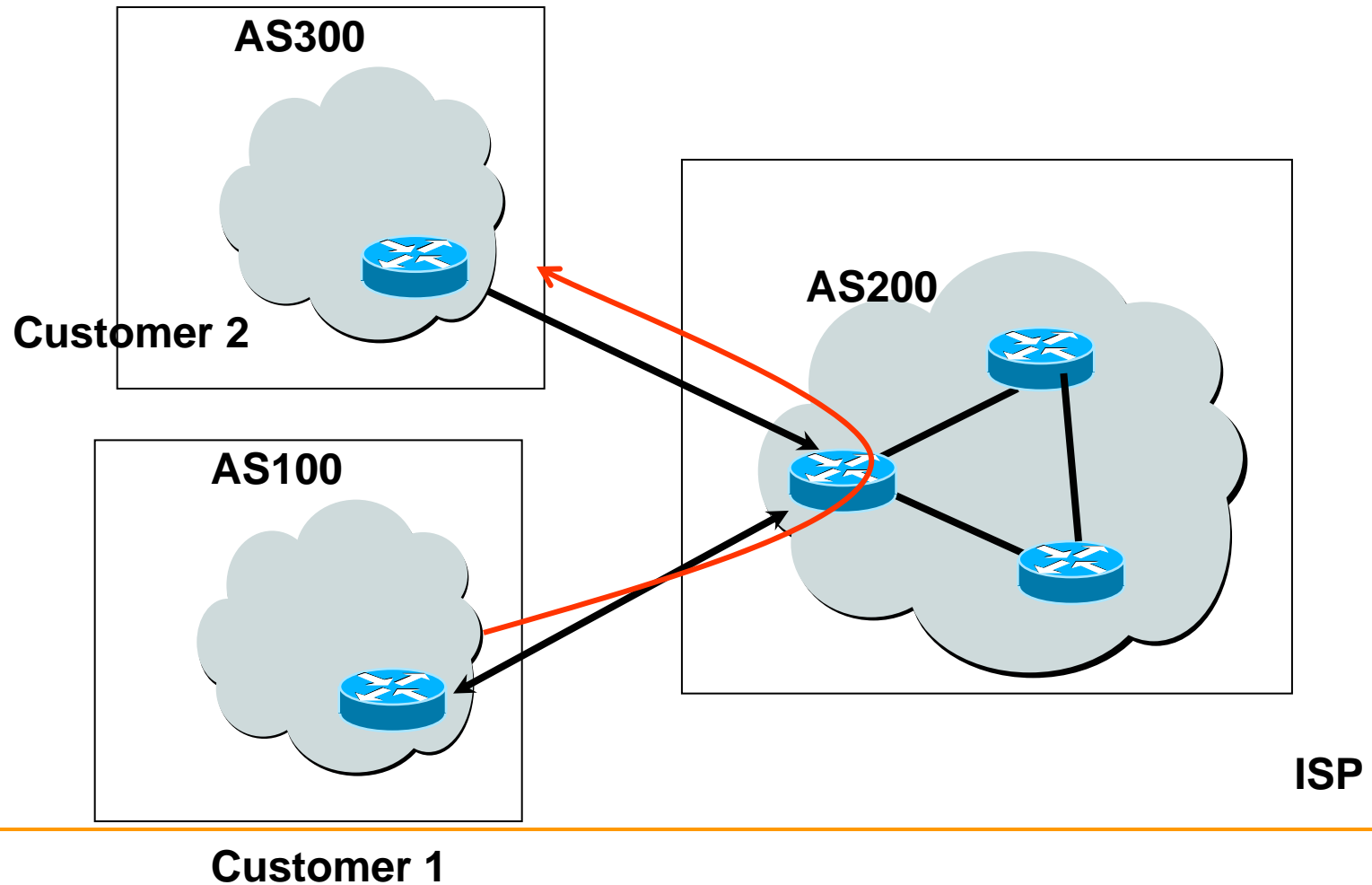# AS connection

- Single-Homed AS
  - Static
  - IGP
  - BGP

AS connection

- MultiHomed Nontransit AS

AS connection

- MultiHomed Transit AS

AS300

Customer 2

AS100

AS200

ISP
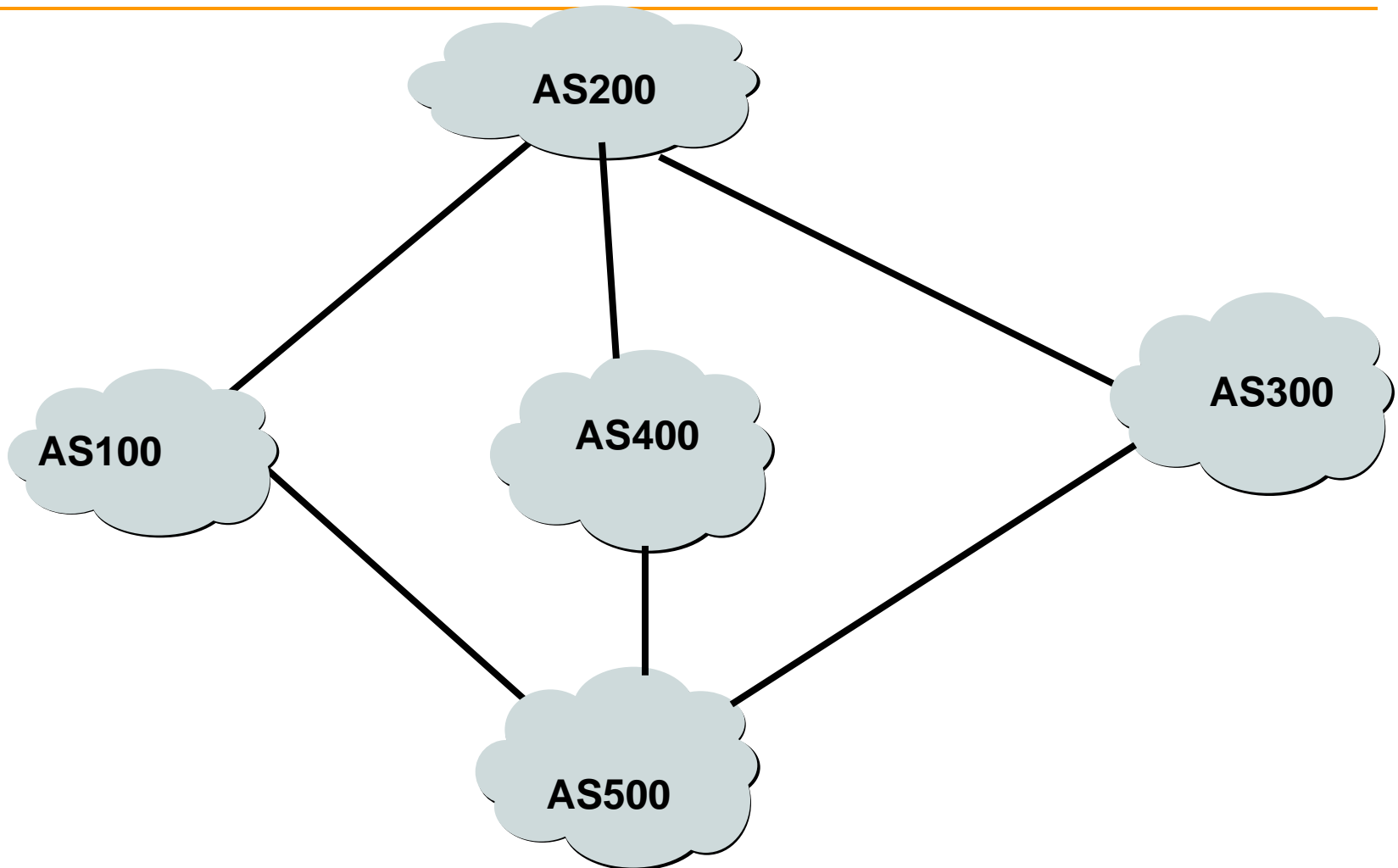
Customer 1

Policy-based routing

Path-vector routing

- IGP announce networks and describe the cost to reach those networks.
- BGP announces pathways and the networks that are reachable at the end of the pathway. BGP describes the pathway by using attributes which are similar to metrics.
- The adminsitrator define routing policies.

BGP Database

- Neighbor table
  - List of BGP neighbors
- BGP forwarding table
  - List of all networks learned from each neighbor
- IP routing table
  - List of best path to destination networks

BGP Message

- Open
    - Version number
    - AS number
    - Holdtime
    - Router ID
- Keepalive
- Update
- Notification
    - When error is detected

eBGP Peering

- EBGP (External BGP)
  - EBGP has peering relationship between routers in different ASs
  - EBGP is normally running on the same subnet
  - Must be connected directly
  - Exceptions : use multi-hop

AS100   **AS200**   **AS300**

EBGP neighbors

EBGP neighbors
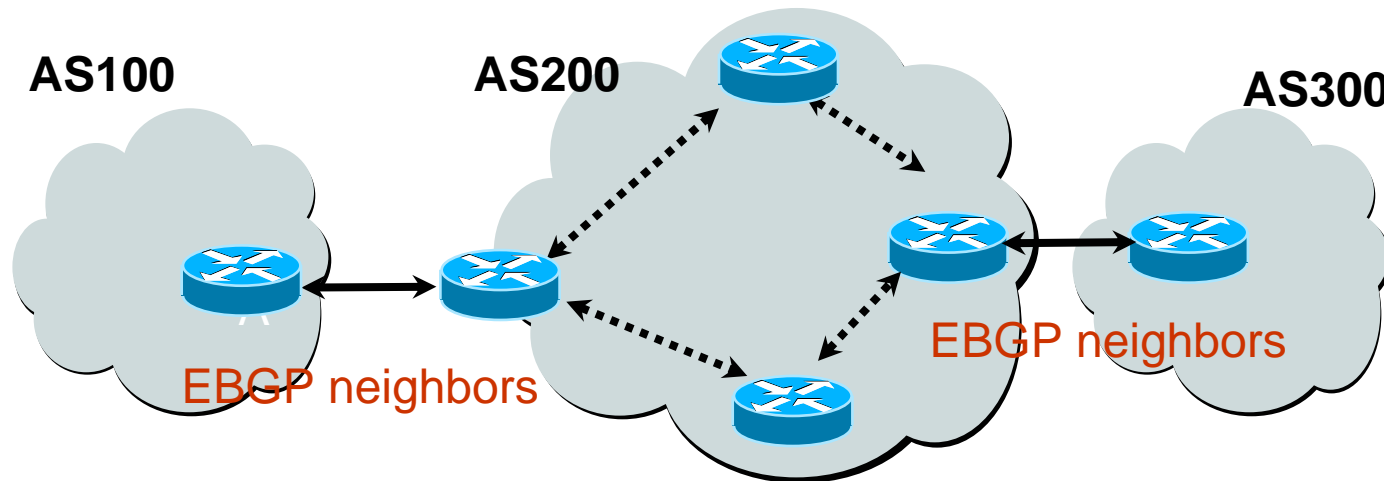
iBGP Peering I

- IBGP (Internal BGP)
  - IBGP has peering inside same AS
  - Don't have to be connected directly

AS400

AS200

AS100

AS300

IBGP

IBGP

IBGP neighbors
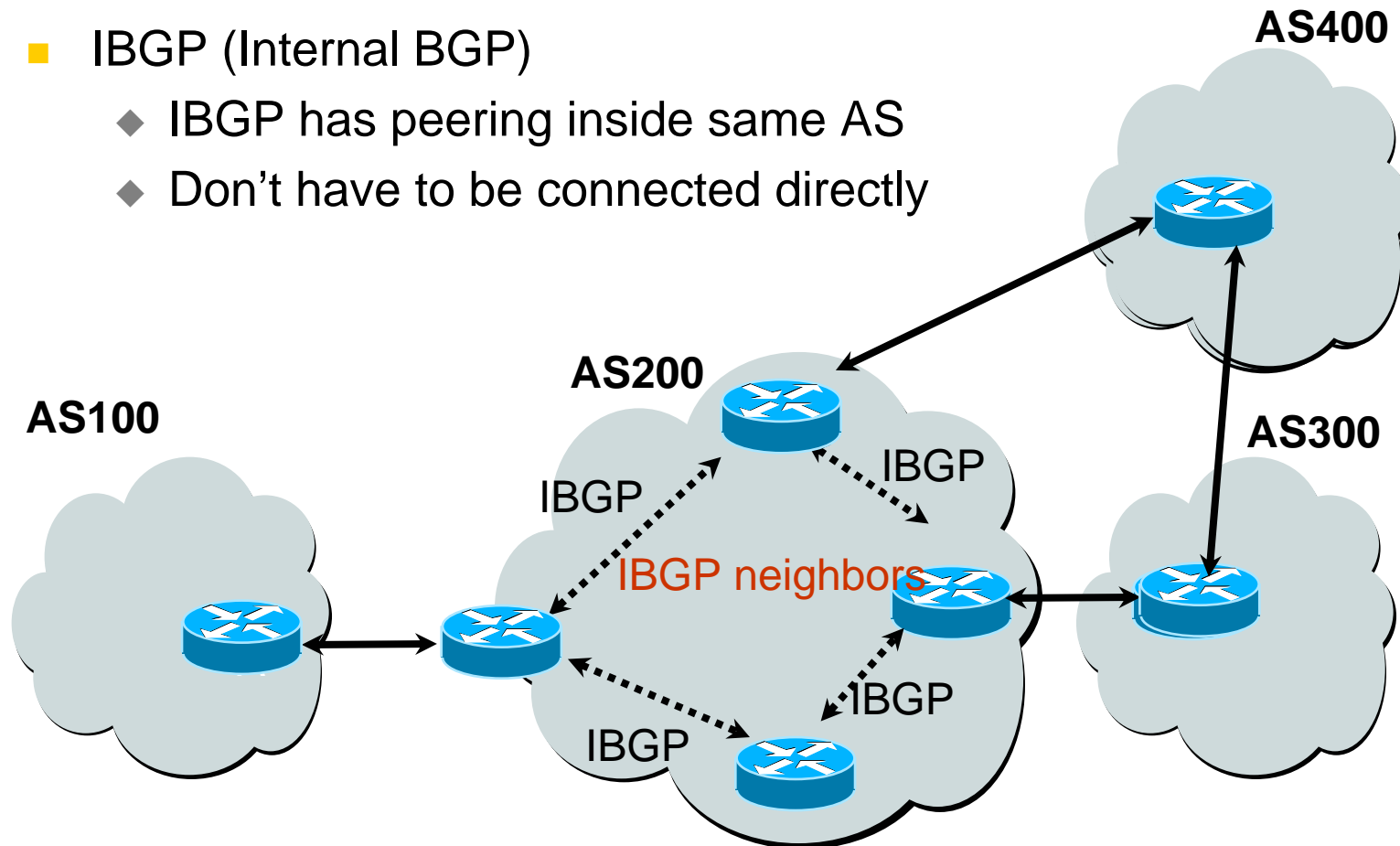
IBGP

IBGP

IBGP split horizon rule
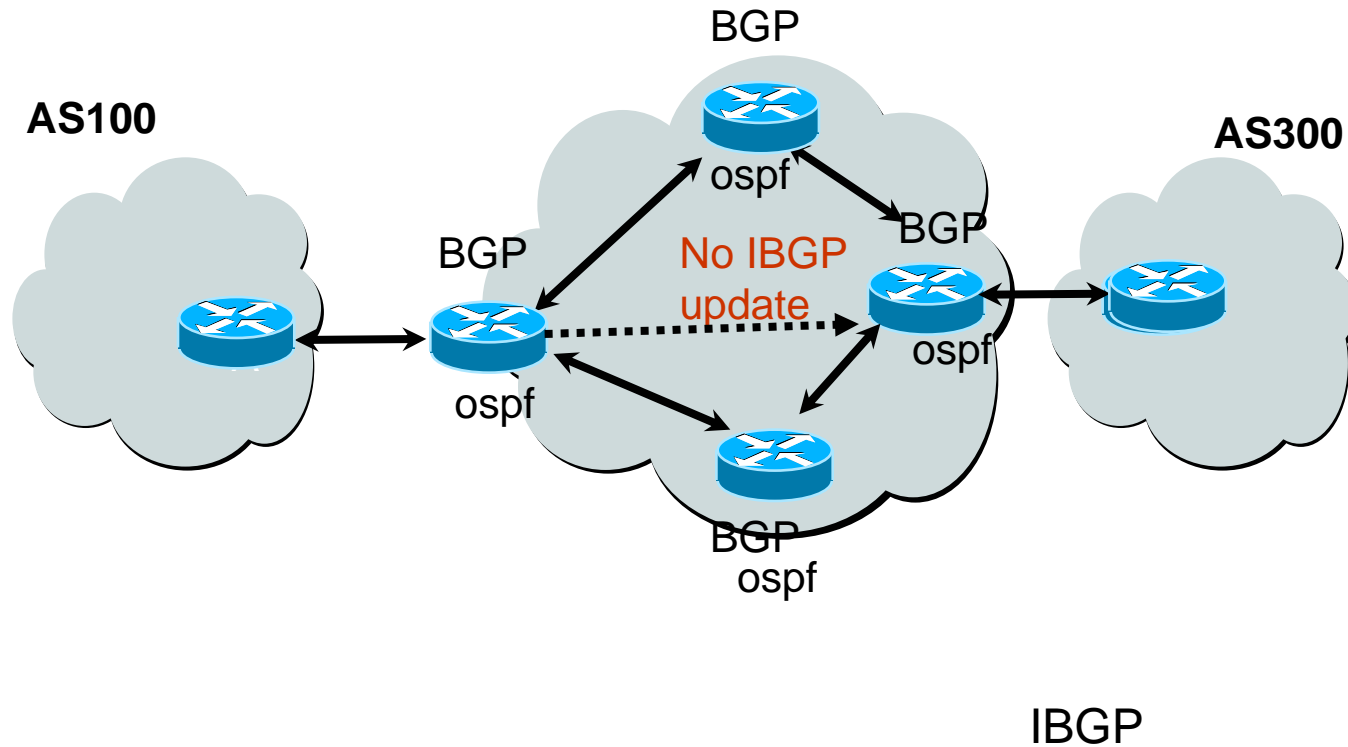
- By default, routes learned via IBGP are never propagated to other IBGP peers

BGP & IGP synchrnization
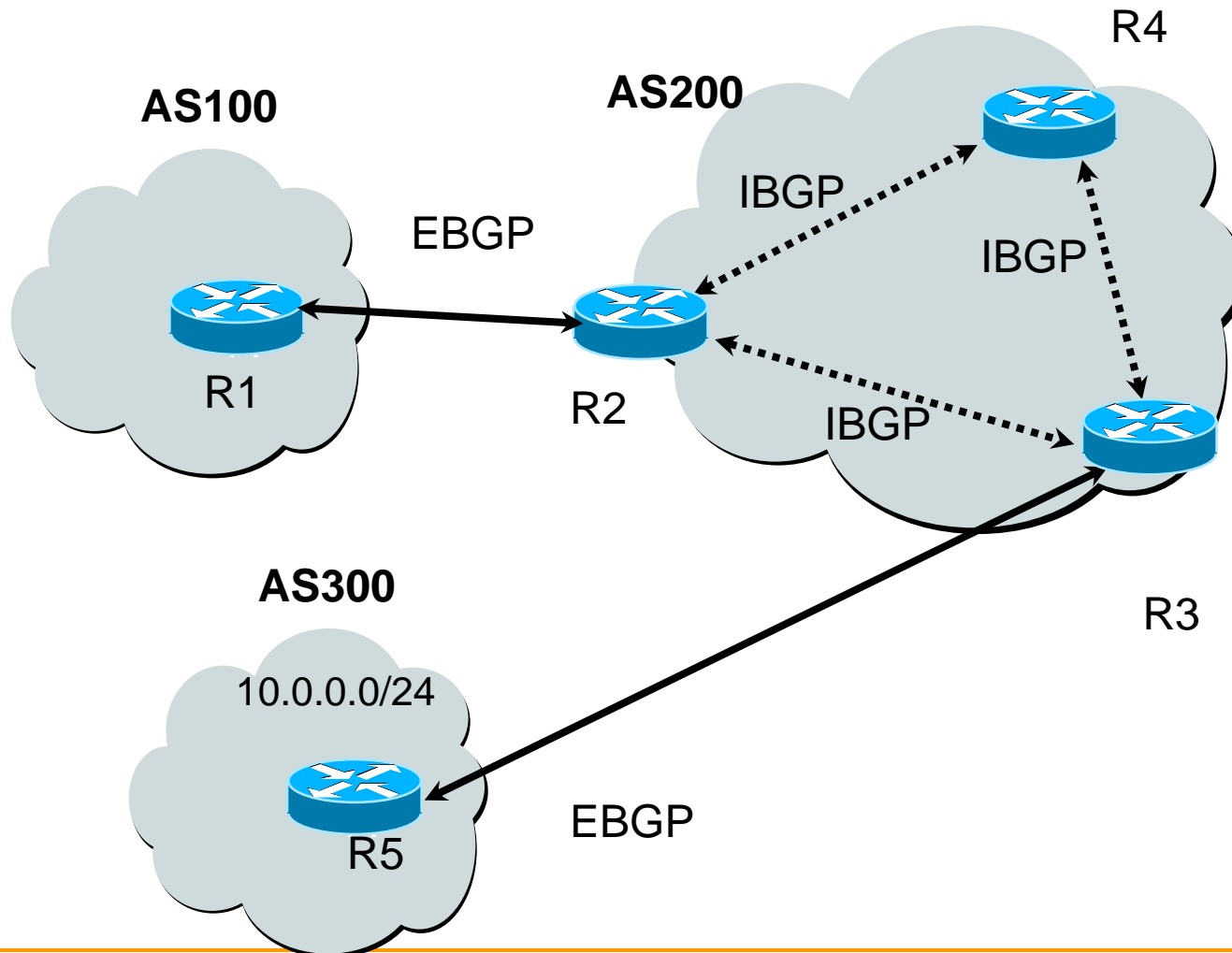
Path Attributes

- Make routing policy with adjusting Path Attribute
- Types
  - Well-known mandatory
    - Origin, as-path, next-hop
  - Well-known discretionary : may not in update message
    - Local-pref
  - Optional transitive
    - Aggregator, community
  - Optional non-transitive
    - Multi-exit-disc, originator_id, cluster_list

Path Attributes(cont.)

- ORIGIN(1)
- AS_PATH(2)
- NEXT_HOP(3)
- MULTI_EXIT_DISC(4)
- LOCAL_PREF(5)
- ATOMIC_AGGREGATE(6)
- AGGREGATOR(7)
- COMMUNITY(8)
- ORIGINATOR_ID(9)
- Cluster List(10)
- WEIGHT(CISCO ONLY)

origin

- The source(origin) of a speicifi routing update.
  - IGP
  - BGP
  - Incomplete
- Priority
  - IGBP > EBGP > Incomplete

AS_Path

- The lists of AS which go through to get reach to destination

- Used for Loop detection

- Apply routing policy to each AS path

AS_Path (cont.)

AS100

130.13.0.0/16

AS200

140.14.0.0/16

Network	as-path
130.13.0.0/16	100
140.14.0.0/16	
150.15.0.0/16	300
160.16.0.0/16	300 400
170.17.0.0/16	300 500

Network	as-path
130.13.0.0/16	200 100
140.14.0.0/16	200
150.15.0.0/16	
160.16.0.0/16	400
170.17.0.0/16	500

AS300

150.15.0.0/16

Network	as-path
130.13.0.0/16	300 200 100
140.14.0.0/16	300 200
150.15.0.0/16	300
160.16.0.0/16	300 400
170.17.0.0/16	

AS400

160.16.0.0/16

AS500

170.17.0.0/16

AS_Path Filtering

AS200

130.13.1.6/30

130.13.1.5/30

150.15.X.X
AS300

AS100

RTA

❑ **How to block 150.15.x.x via AS200 originating from AS300 on RTA?**

AS_Path Filtering(cont.)

```
RTA#
router bgp 100
  neighbor 130.13.1.6 remote-as 200
  neighbor 130.13.1.6 filter-list 10 in
ip as-path access-list 10 deny 300
ip as-path access-list 10 permit .*
```
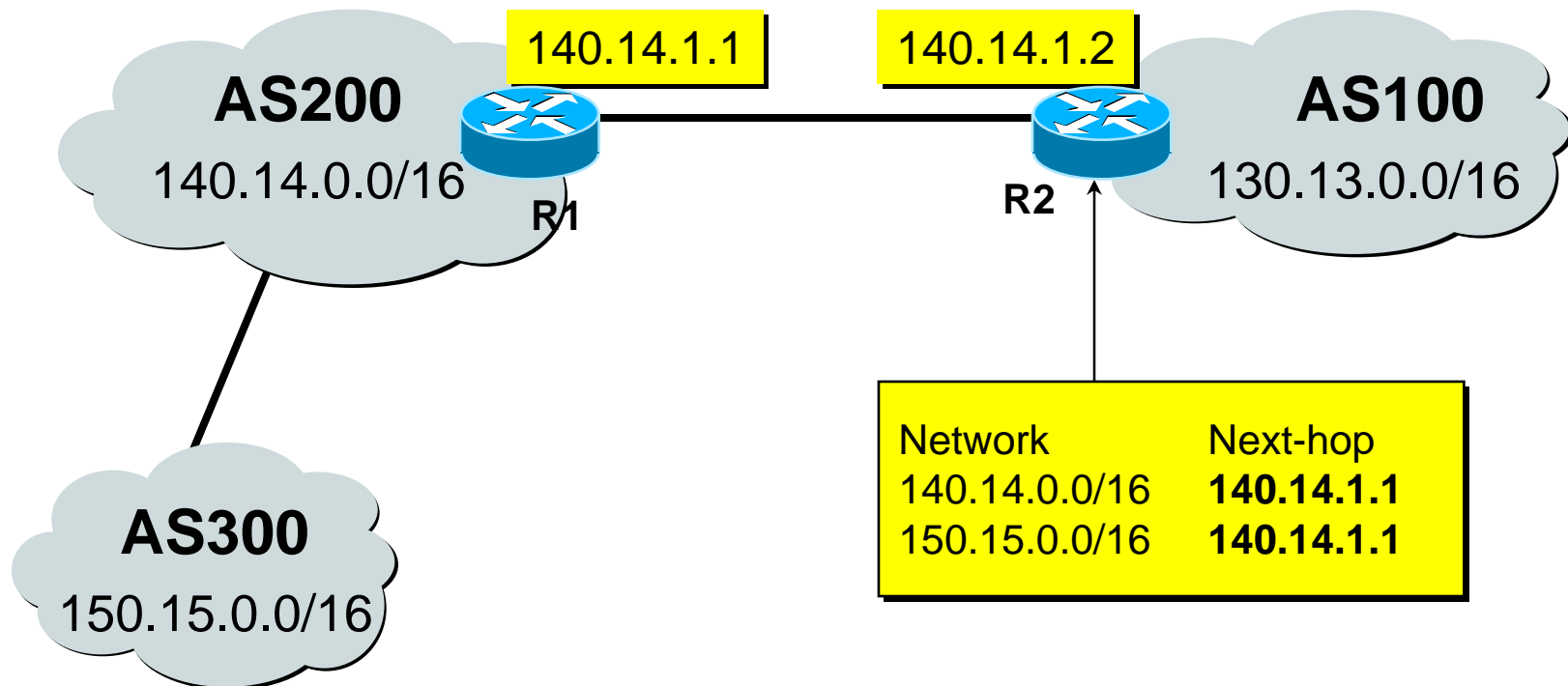
Next hop behavior

- BGP is an AS-by-AS routing protocol, not router-by-router routing protocol
- In BGP, the next hop means the IP address to reach the next AS

Next_Hop

EBGP Next-Hop

Next_Hop (cont.)

IBGP Next-Hop

AS200

140.14.0.0/16

140.14.1.1

AS300

150.15.0.0/16

140.14.1.2

AS100

130.13.0.0/16

R1

R2

130.13.1.1

130.13.1.2

R3

Network	Next-hop
140.14.0.0/16	**140.14.1.1**
150.15.0.0/16	**140.14.1.1**

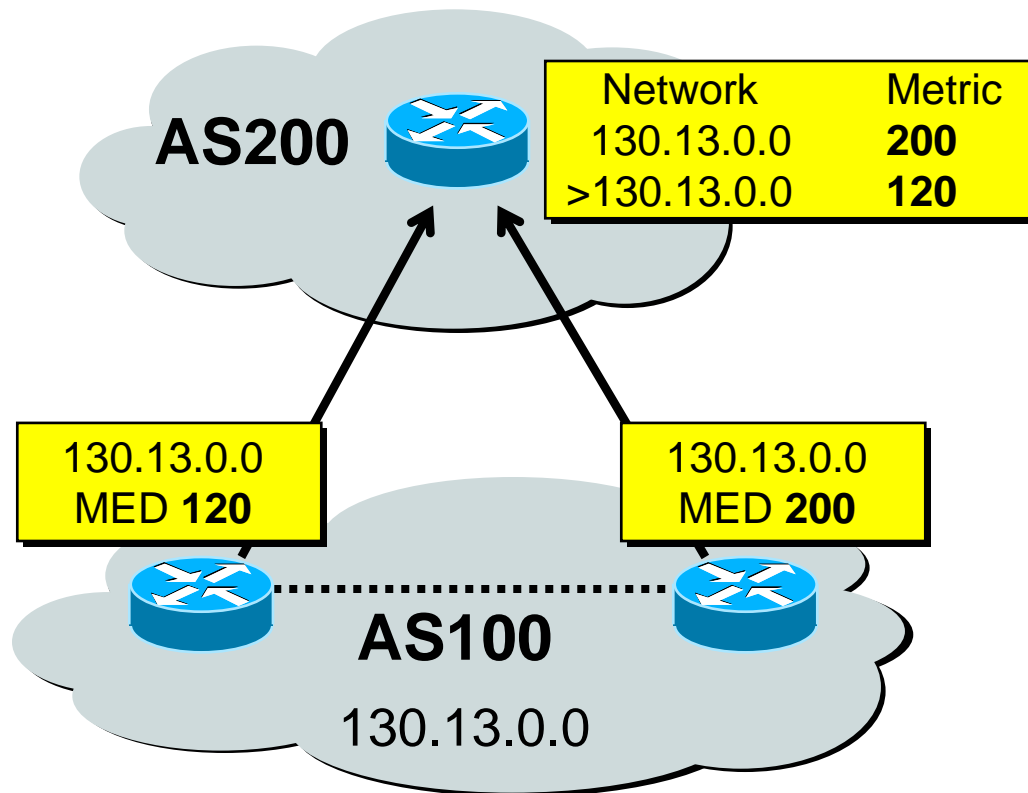| Network	Next-hop
140.14.0.0/16	**140.14.1.1**
150.15.0.0/16	**140.14.1.1**
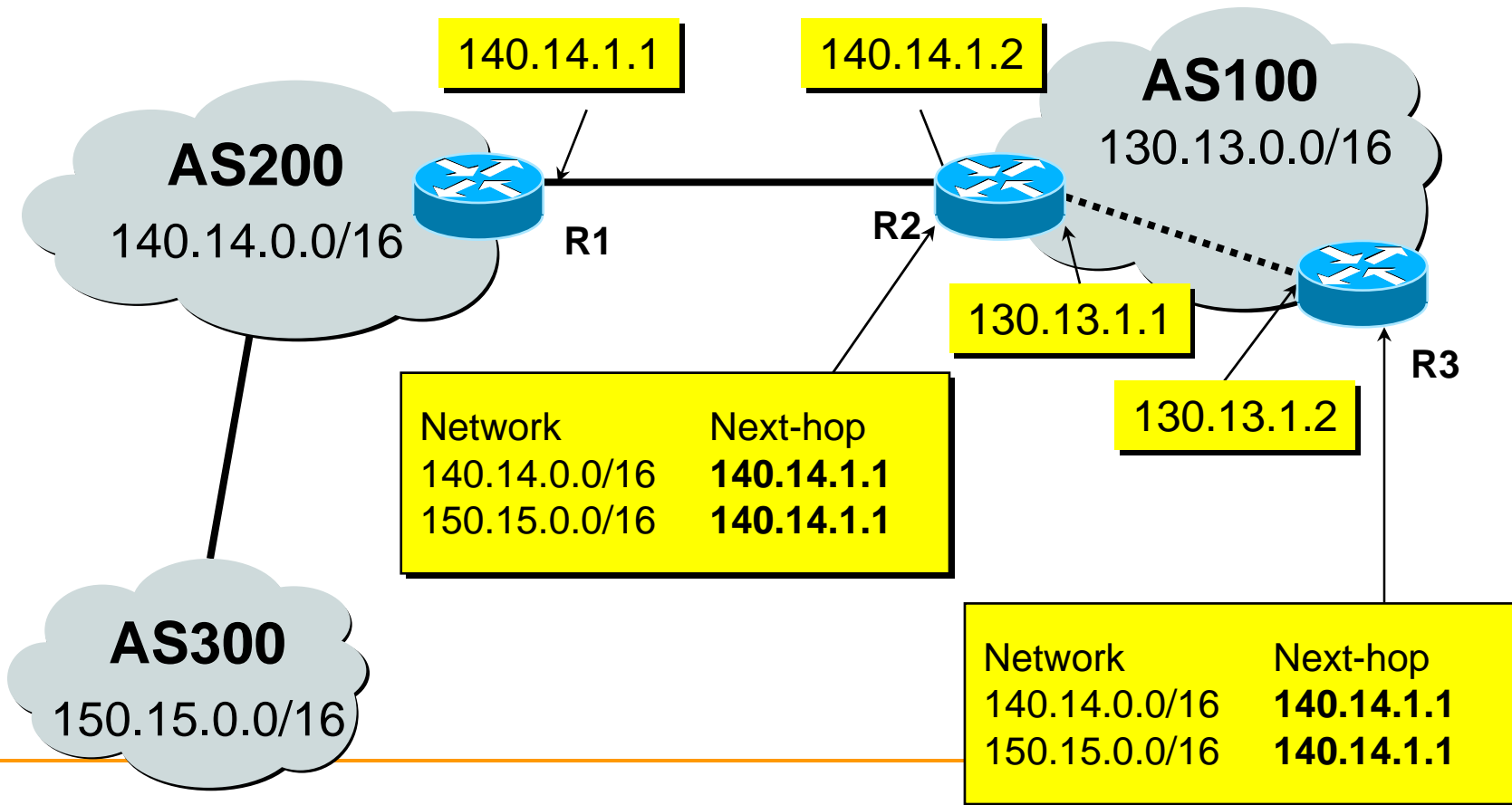
BGP Command

□ **Next-hop-self**

Router(config-router)#

Neighbor {ip-address | peer-group-name} next-hop-self

- ◆ Forces all updates for this neighbor to be advertised with this router as the next hop
- ◆ The ip address used for the next-hop-self will be the same as the source IP address of the BGP packet.

Multi_Exit_Discriminator

Network	Metric
130.13.0.0	**200**
>130.13.0.0	**120**

AS200

130.13.0.0
MED **120**

130.13.0.0
MED **200**

AS100

130.13.0.0

Local_Preference

AS400
160.16.0.0/16

AS200

AS300

local-pref
500

local-pref
700

AS100

network	local-pref
160.16.0.0/16	**500**
>160.16.0.0/16	**700**

Weight

Network	Weight
160.16.0.0	**20000**
>160.16.0.0	**45000**
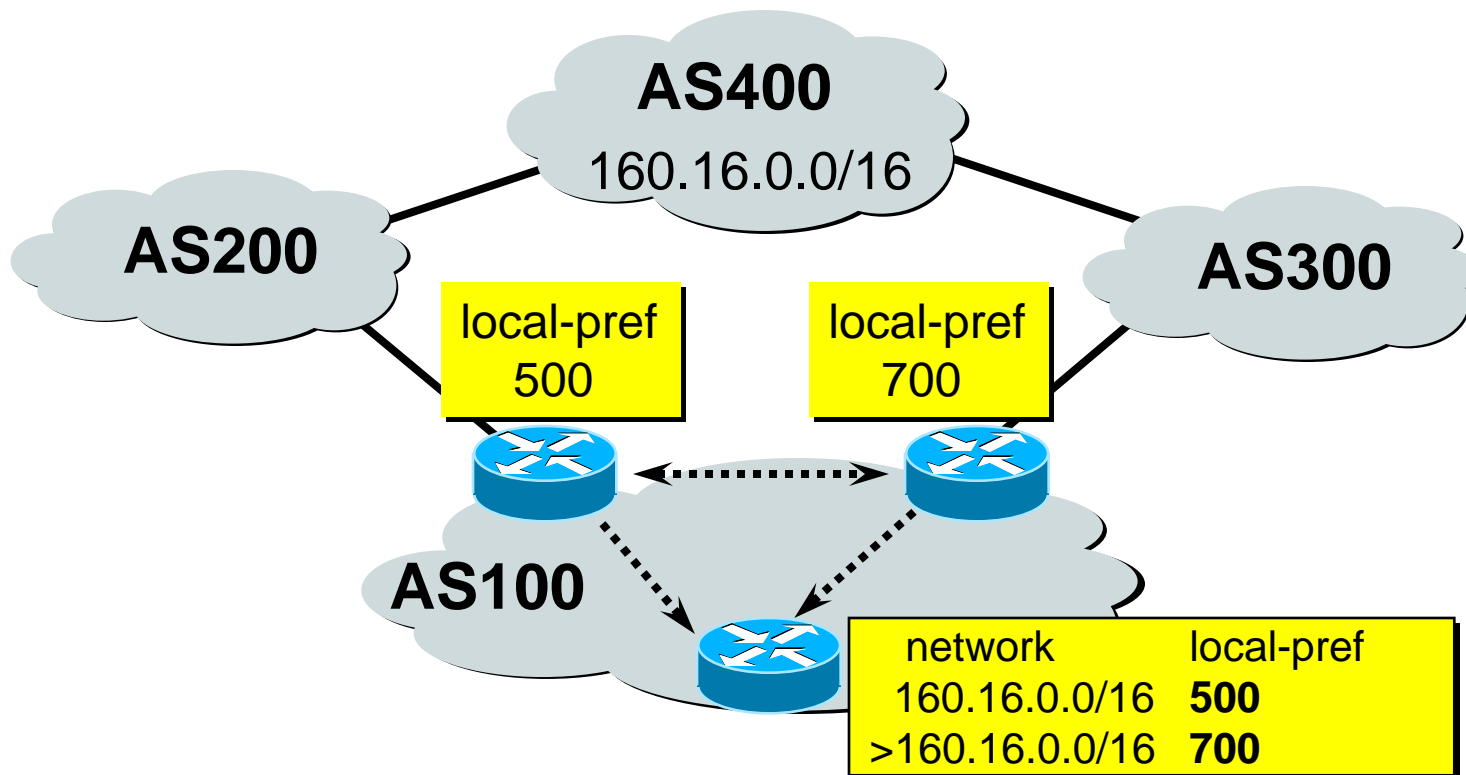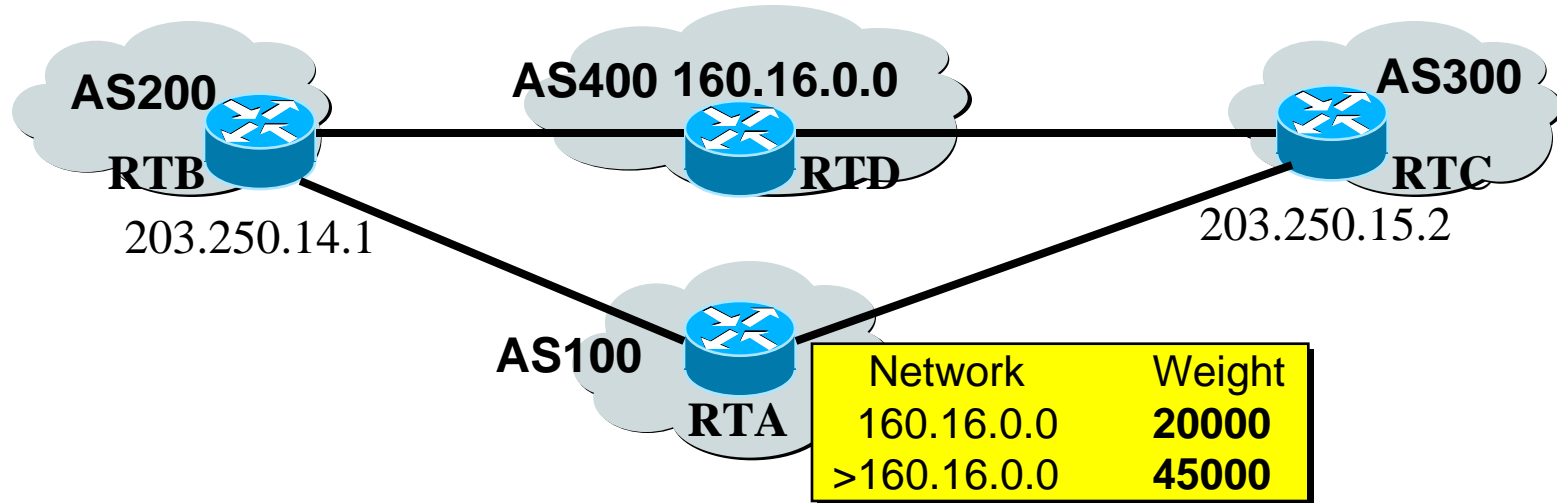
```
RTA#
router bgp 100
  neighbor 203.250.14.1 remote-as 200
  neighbor 203.250.14.1 weight 20000
  neighbor 203.250.15.2 remote-as 300
  neighbor 203.250.15.2 weight 45000
```

IBGP Mesh Solution

- A router running IBGP doesn't relay routing updates from another IBGP router

- This characteristic require full IBGP mesh between IBGP routers

- Problems : If there exists n peers, then [n(n-1)/2] session is needed

- Solutions
  - Route Reflector
  - Confederation

Route Reflector(RR)

- Terminology
  - Route Reflector(RR)
  - Rout Reflector Client
  - Cluster : RR + Clients
  - Cluster ID
- Configuration
  - On RR : make neighbor relationship with clients
  - On clients : make neighbor relationship with only RR

Route Reflector(cont.)

Cluster

RTB(RR)
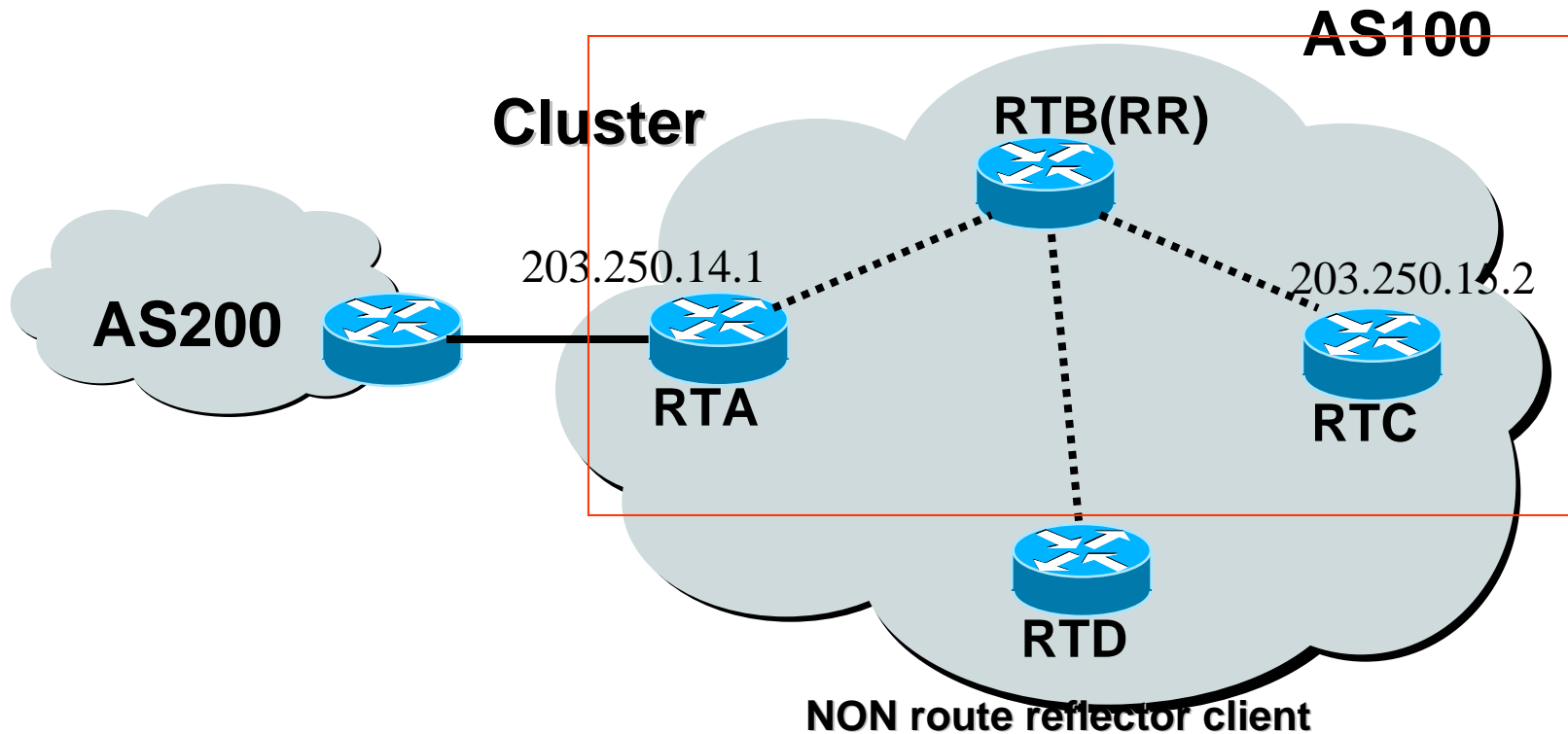
203.250.14.1

AS200

AS100

203.250.15.2

RTA

RTC
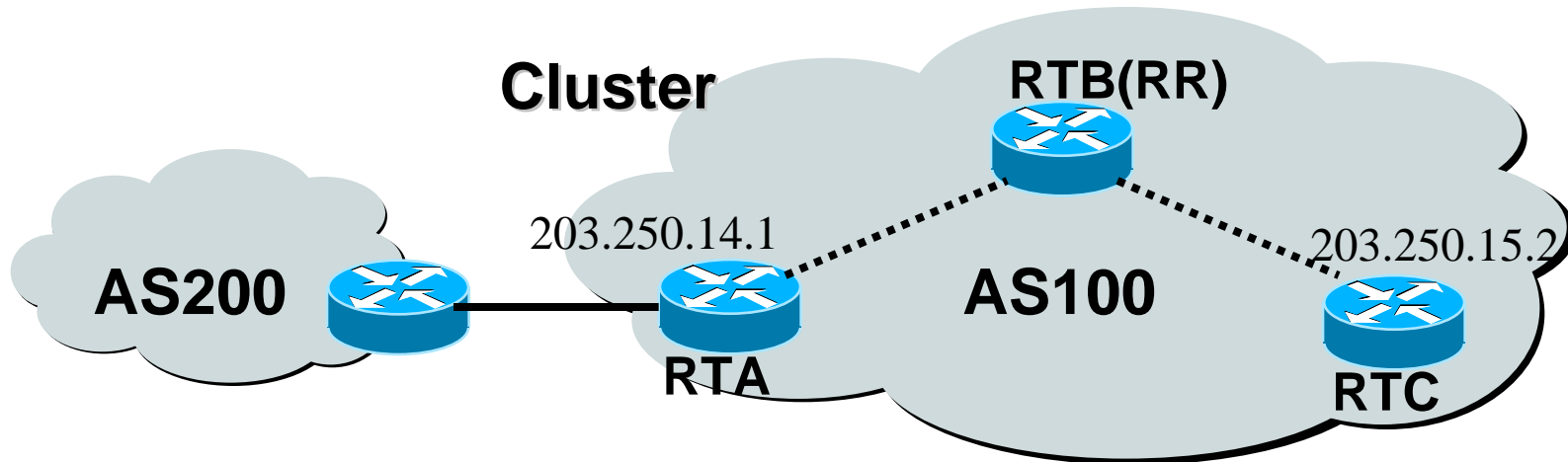
```
RTB#
router bgp 100
  neighbor 203.250.15.2 remote-as 100
  neighbor 203.250.15.2 route-reflector-client
  neighbor 203.250.14.1 remote-as 100
  neighbor 203.250.14.1 route-reflector-client
```

NON-Route Reflector

AS100

Cluster

RTB(RR)

203.250.14.1   203.250.15.2

AS200

RTA   RTC

RTD

NON route reflector client

- When the non-RRC send update information to RR, the RR send it only to RRC
- When RRC sends update information to RR, RR send it to other RRC and non-RRC
- When RR gets the update information through EBGP, RR sends it to all routers.

Confederation

- Make sub(mini) AS inside public AS
- It appears just one public AS outside the AS
- IBGP peering inside sub AS
- Pseudo EBGP peering between sub ASs
- Advantages
    - can minimize the number of IBGP sessions dramatically

Confederation

sub-AS 65200 RTB

sub-AS 65100 RTA

sub-AS 65300 RTC

AS 100
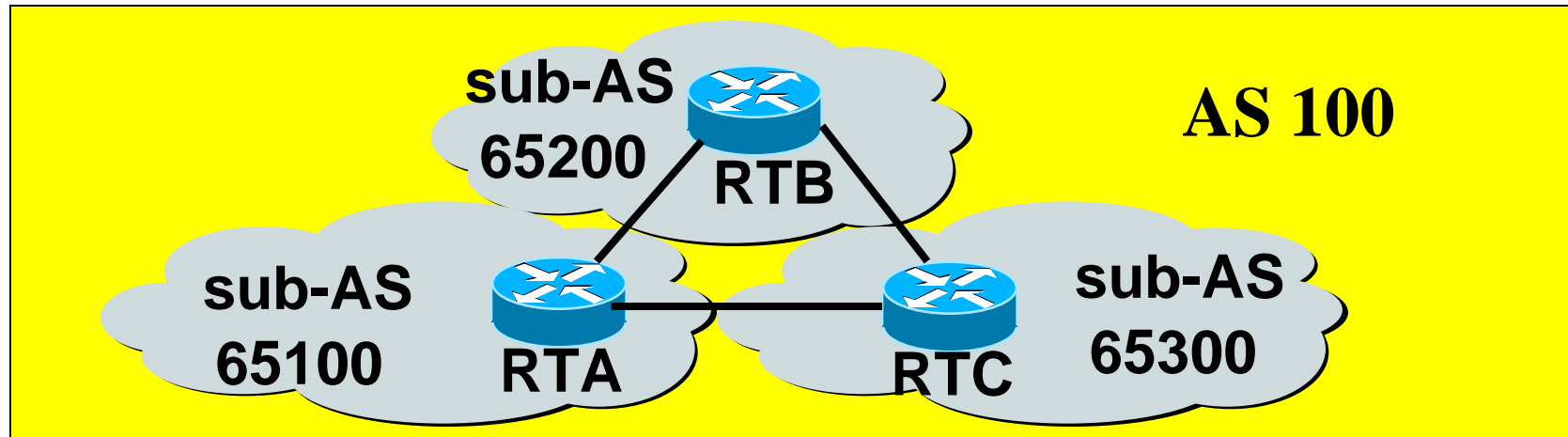
RTB#
router bgp 65200
  bgp confederation identifier 100
  bgp confederation peers 65100 65300
  neighbor 203.250.14.1 remote-as 65100
  neighbor 203.250.15.2 remote-as 65300

BGP Decision Process

- route with a reachable next hop

- route with largest weight

- route with largest local preference

- route locally originated

- route with shortest as-path

BGP Decision Process(cont.)

- Route with lowest origin type
    - igp>egp>incomplete
- Route with lowest MED
- EBGP, next Confederation External, next IBGP
- Route with nearest IGP neighbor
- Route with the lowest BGP router ID

Neighboring negotiation

- Idle
  - Searching routing table to see if a route exists to reach the neighbor
- Connect
- Opensent
  - Sent open message
- Active
  - When no response for 5sec, go back to idle
- Openconfirm
- established

Neighboring negotiation

```
3d21h: BGP: 10.0.0.2 went from Idle to Active
3d21h: BGP: 10.0.0.2 open active, delay 21531ms
3d21h: BGP: 10.0.0.2 open active, local address 10.0.0.1
3d21h: BGP: 10.0.0.2 open failed: Connection refused by remote host
3d21h: BGP: 10.0.0.2 passive open
3d21h: BGP: 10.0.0.2 went from Active to Idle
3d21h: BGP: 10.0.0.2 went from Idle to Connect
3d21h: BGP: 10.0.0.2 rcv message type 1, length (excl. header) 26
3d21h: BGP: 10.0.0.2 rcv OPEN, version 4
3d21h: BGP: 10.0.0.2 went from Connect to OpenSent
3d21h: BGP: 10.0.0.2 sending OPEN, version 4, my as: 100
3d21h: BGP: 10.0.0.2 rcv OPEN w/ OPTION parameter len: 16
3d21h: BGP: 10.0.0.2 rcvd OPEN w/ optional parameter type 2 (Capability) len 6
3d21h: BGP: 10.0.0.2 OPEN has CAPABILITY code: 1, length 4
3d21h: BGP: 10.0.0.2 OPEN has MP_EXT CAP for afi/safi: 1/1
3d21h: BGP: 10.0.0.2 rcvd OPEN w/ optional parameter type 2 (Capability) len 2
3d21h: BGP: 10.0.0.2 OPEN has CAPABILITY code: 128, length 0
3d21h: BGP: 10.0.0.2 OPEN has ROUTE-REFRESH capability(old) for all address-families
3d21h: BGP: 10.0.0.2 rcvd OPEN w/ optional parameter type 2 (Capability) len 2
3d21h: BGP: 10.0.0.2 OPEN has CAPABILITY code: 2, length 0
3d21h: BGP: 10.0.0.2 OPEN has ROUTE-REFRESH capability(new) for all address-families
3d21h: BGP: 10.0.0.2 went from OpenSent to OpenConfirm
3d21h: BGP: 10.0.0.2 send message type 1, length (incl. header) 45
3d21h: BGP: 10.0.0.2 went from OpenConfirm to Established
3d21h: %BGP-5-ADJCHANGE: neighbor 10.0.0.2 Up
```

BGP Operation

- Establish TCP connection
- Negotiate parameter(ex: version number) between peers
- Exchange entire routing table at initial phase
- Exchange incremental updates after initial phase
- Send keepalives to confirm connectivity between peers

BGP Operation(cont.)

- Get path information about destination prefix from internal and external BGP peers

- Register the best route in the routing table

- Can use routing policy when select the best route

BGP Command

❑ **Enable BGP routing protocol**

Router(config)#

    router bgp *autonomous-system*

❑ **Activate the BGP session**

Router(config-router)#

    neighbor *ip-address* remote-as *autonomous-system*

BGP Command(cont.)

❑ **Declare network to advertise**

Router(config-router)#

network *network-number*

❑ **Reset BGP connection to update BGP information, but be cautious**
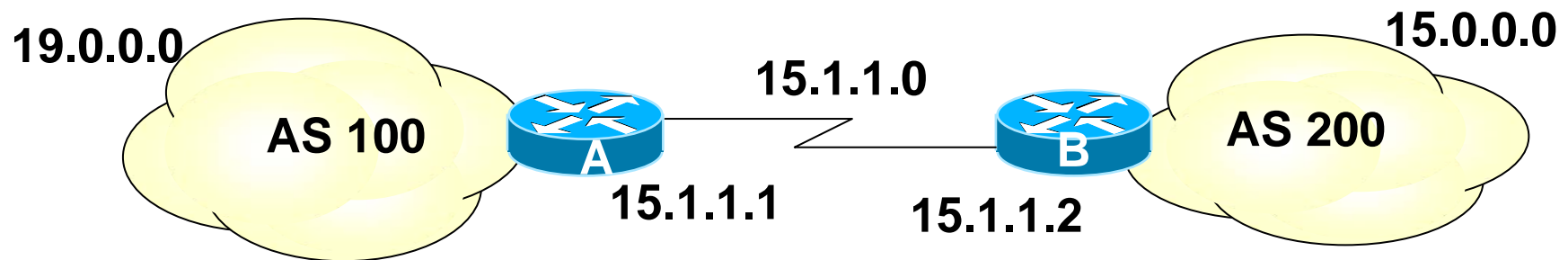
Router#

clear ip bgp { * | address }

BGP Command(cont.)

❑ **Disable synchronization**

Router(config-router)#

no synchronization

BGP Configuration Example

19.0.0.0

AS 100

15.0.0.0

15.1.1.0

AS 200

15.1.1.1

15.1.1.2

Router A

```
router bgp 100
network 19.0.0.0
neighbor 15.1.1.2 remote-as 200
```
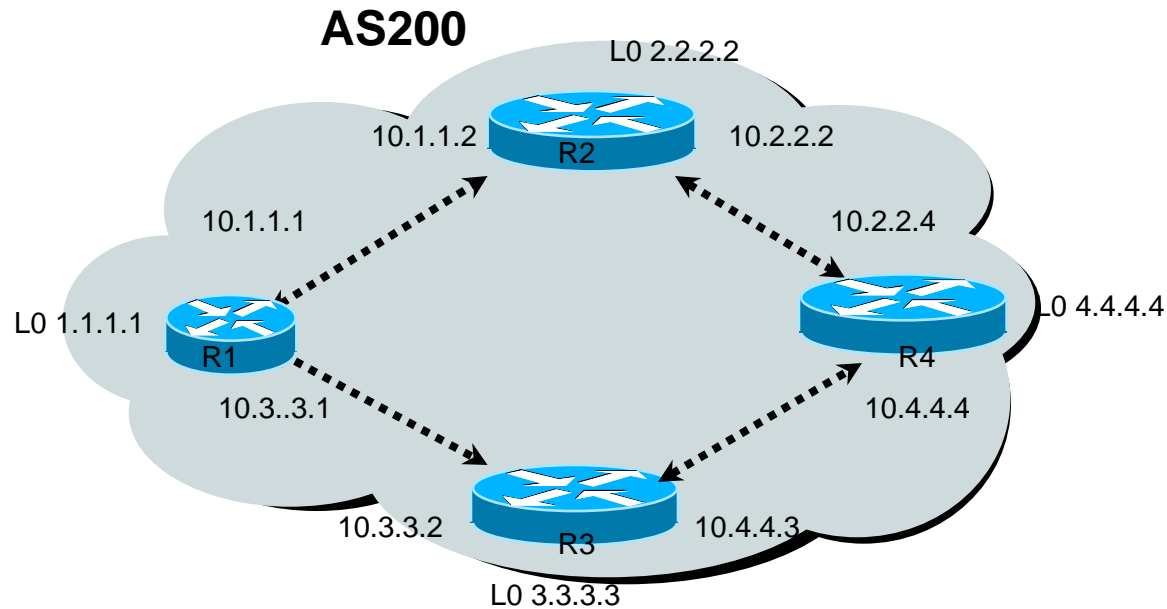
Router B

```
router bgp 200
network 15.0.0.0
neighbor 15.1.1.1 remote-as 100
```

Source IP address

- **Neighboring process**
  - ◆ Receive BGP pkt
  - ◆ Compare the source address of the packet with the list of neighbor statements
    - Match: neighboring is established
    - No match: the packet is ignored
- The source IP address of BGP must be listed in the neighbor statement of the other routers

IBGP peering: source address

- To Establish the IBGP sesseion between R1 and R4
  - R1: ip address in the neighbor statement ?
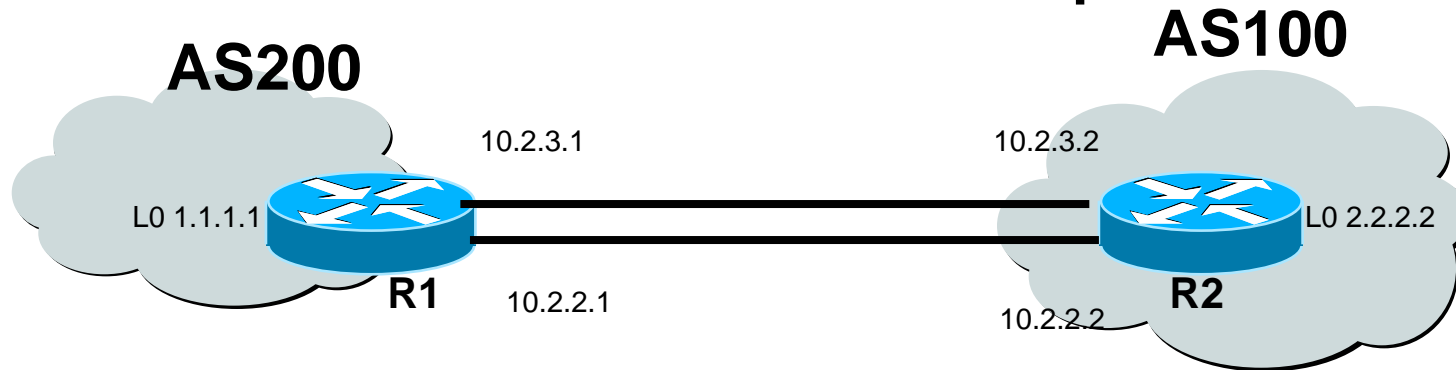 - R4: ip address in the neighbor statement ?

BGP Command

❑ **override source IP addr for BGP pkt**

Router(config-router)#

Neighbor {ip-address | peer-group-name} update-source
Interface-type interface-number

◆ Loopback interface is usually used
◆ Normally used only with IBGP neighbors

EBGP-multihop

AS200

AS100

10.2.3.1   10.2.3.2

L0 1.1.1.1   R1   R2   L0 2.2.2.2

10.2.2.1

10.2.2.2

- Need to use loopback address for neighbor
- Static route for loopback address
- Ebgp-multihop

Router(config-router)#

Neighbor {ip-address | peer-group-name} ebgp-multihop 2

Monitoring BGP

- show ip bgp

- show ip bgp neighbor

- show ip bgp paths

- show ip bgp summary

- show ip route

BGP peering

- Command
  - Show ip bgp summary

```
t3#sh ip bgp sum
BGP router identifier 200.0.3.1, local AS number 100
BGP table version is 14, main routing table version 14
5 network entries using 505 bytes of memory
5 path entries using 240 bytes of memory
2 BGP path attribute entries using 120 bytes of memory
1 BGP AS-PATH entries using 24 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 889 total bytes of memory
BGP activity 9/4 prefixes, 9/4 paths, scan interval 60 secs

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.0.0.1        4   200      72      76       14    0    0 01:05:17           1
```

Show ip BGP command

- Command
  - Show ip bgp

```
t3#sh ip bgp
BGP table version is 14, local router ID is 200.0.3.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 10.0.0.0         0.0.0.0                130          32768 ?
*> 20.0.0.0/24      10.0.0.1                 0              0 200 i
*> 50.0.0.0         0.0.0.0                130          32768 ?
*> 200.0.5.0        10.0.1.2               130          32768 ?
*> 200.0.7.0        10.0.1.2               130          32768 ?
```

BGP session establishment

Router#

<div style="border:1px solid black; padding:10px;">
<center>Debug ip bgp events</center>
</div>

```
3d21h: BGP: 10.0.0.2 went from Idle to Active
3d21h: BGP: 10.0.0.2 open active, delay 21531ms
3d21h: BGP: 10.0.0.2 open active, local address 10.0.0.1
3d21h: BGP: 10.0.0.2 open failed: Connection refused by remote host
3d21h: BGP: 10.0.0.2 passive open
3d21h: BGP: 10.0.0.2 went from Active to Idle
3d21h: BGP: 10.0.0.2 went from Idle to Connect
3d21h: BGP: 10.0.0.2 rcv message type 1, length (excl. header) 26
3d21h: BGP: 10.0.0.2 rcv OPEN, version 4
3d21h: BGP: 10.0.0.2 went from Connect to OpenSent
3d21h: BGP: 10.0.0.2 sending OPEN, version 4, my as: 100
3d21h: BGP: 10.0.0.2 rcv OPEN w/ OPTION parameter len: 16
3d21h: BGP: 10.0.0.2 rcvd OPEN w/ optional parameter type 2 (Capability) len 6
3d21h: BGP: 10.0.0.2 OPEN has CAPABILITY code: 1, length 4
3d21h: BGP: 10.0.0.2 OPEN has MP_EXT CAP for afi/safi: 1/1
3d21h: BGP: 10.0.0.2 rcvd OPEN w/ optional parameter type 2 (Capability) len 2
3d21h: BGP: 10.0.0.2 OPEN has CAPABILITY code: 128, length 0
3d21h: BGP: 10.0.0.2 OPEN has ROUTE-REFRESH capability(old) for all address-families
3d21h: BGP: 10.0.0.2 rcvd OPEN w/ optional parameter type 2 (Capability) len 2
3d21h: BGP: 10.0.0.2 OPEN has CAPABILITY code: 2, length 0
3d21h: BGP: 10.0.0.2 OPEN has ROUTE-REFRESH capability(new) for all address-families
3d21h: BGP: 10.0.0.2 went from OpenSent to OpenConfirm
3d21h: BGP: 10.0.0.2 send message type 1, length (incl. header) 45
3d21h: BGP: 10.0.0.2 went from OpenConfirm to Established
3d21h: %BGP-5-ADJCHANGE: neighbor 10.0.0.2 Up
```

BGP idle and established states

- Idle
  - The router cannot find the address of the neighbor in the routing table.
    - Check for an IGP problem.

- Established
  - The proper state for BGP
  - In the Show ip bgp summary, the state column is blank or number.

BGP active state troubleshooting

- Active
  - The router sent out an open packet and is waiting for a response.
  - This state may cycle between active and idle.
  - Reasons maybe :
    - Neighbor peering with the wrong ip address
    - Neighbor does not have neighbor statement for this router
    - Neighbor does not have a route to the source ip address of the BGP open packet generated by this router