Co-funded by the European Union

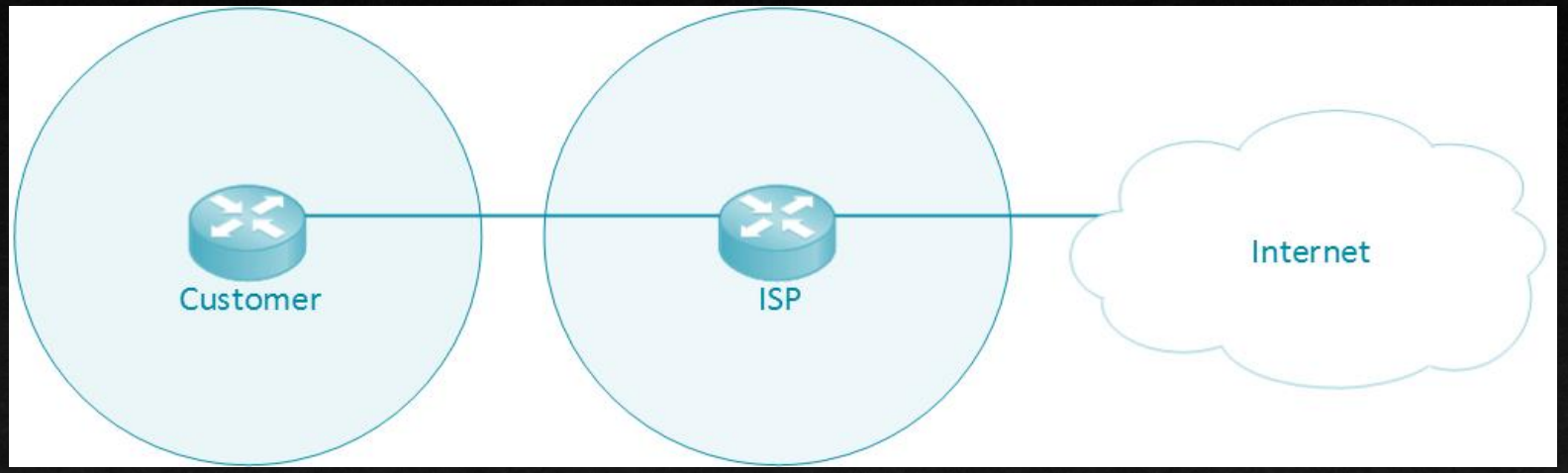
BGP

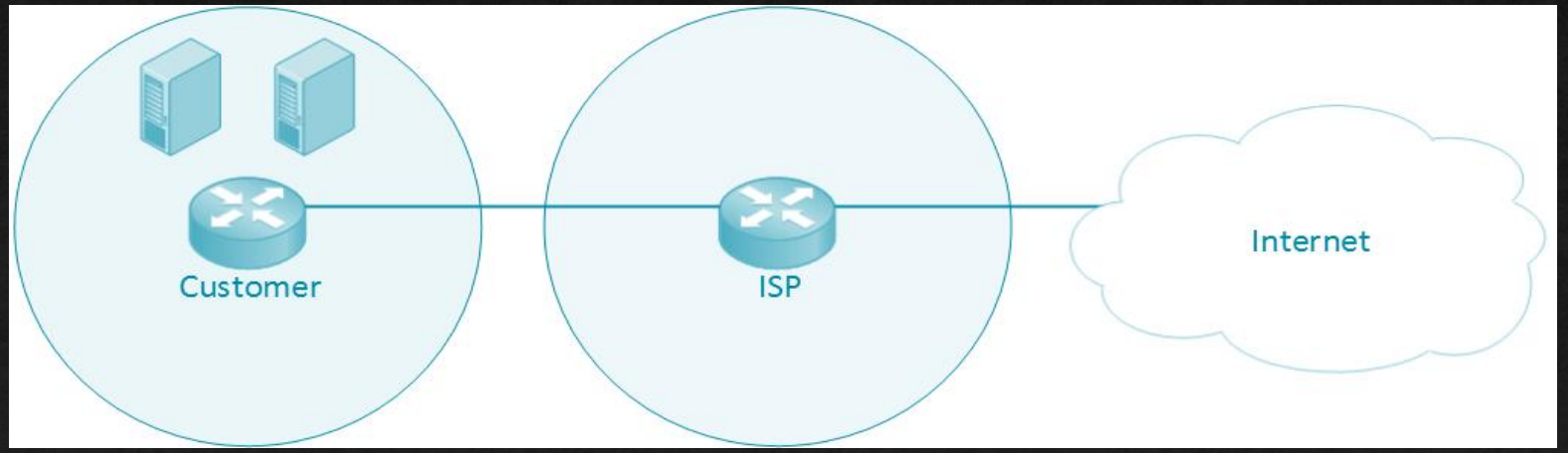
linkedin.com/in/nazrul13

BGP in Short

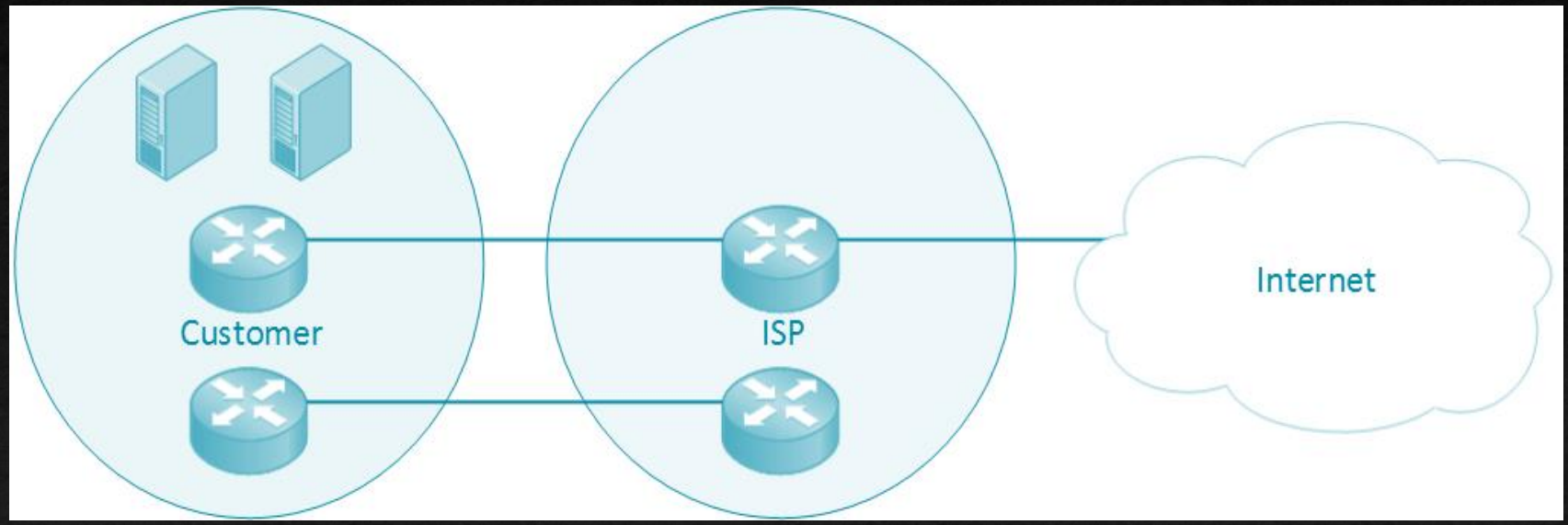
- BGP is, quite literally, the protocol that makes the internet work. BGP is short for Border Gateway Protocol and it is the routing protocol used to route traffic across the internet. Routing Protocols (such as BGP, OSPF, RIP, EIGRP, etc...) are designed to help routers advertise adjacent networks and since the internet is a network of networks, BGP helps to propagate these networks to all BGP Routers across the world.
- BGP is defined by IETF in RFC 4271 and we are currently on version 4 (BGP4 or BGP-4) since 2006. BGP is a Layer 4 Protocol where peers have to be manually configured to form a TCP connection and begin speaking BGP to exchange routing information.
- Some BGP Attributes are-
  - Path Vector Protocol
  - Incremental Updates
  - Many options for policy enforcement
  - Classless Inter Domain Routing (CIDR)
  - Widely used for Internet Backbone
  - Autonomous systems

Why We Need BGP?

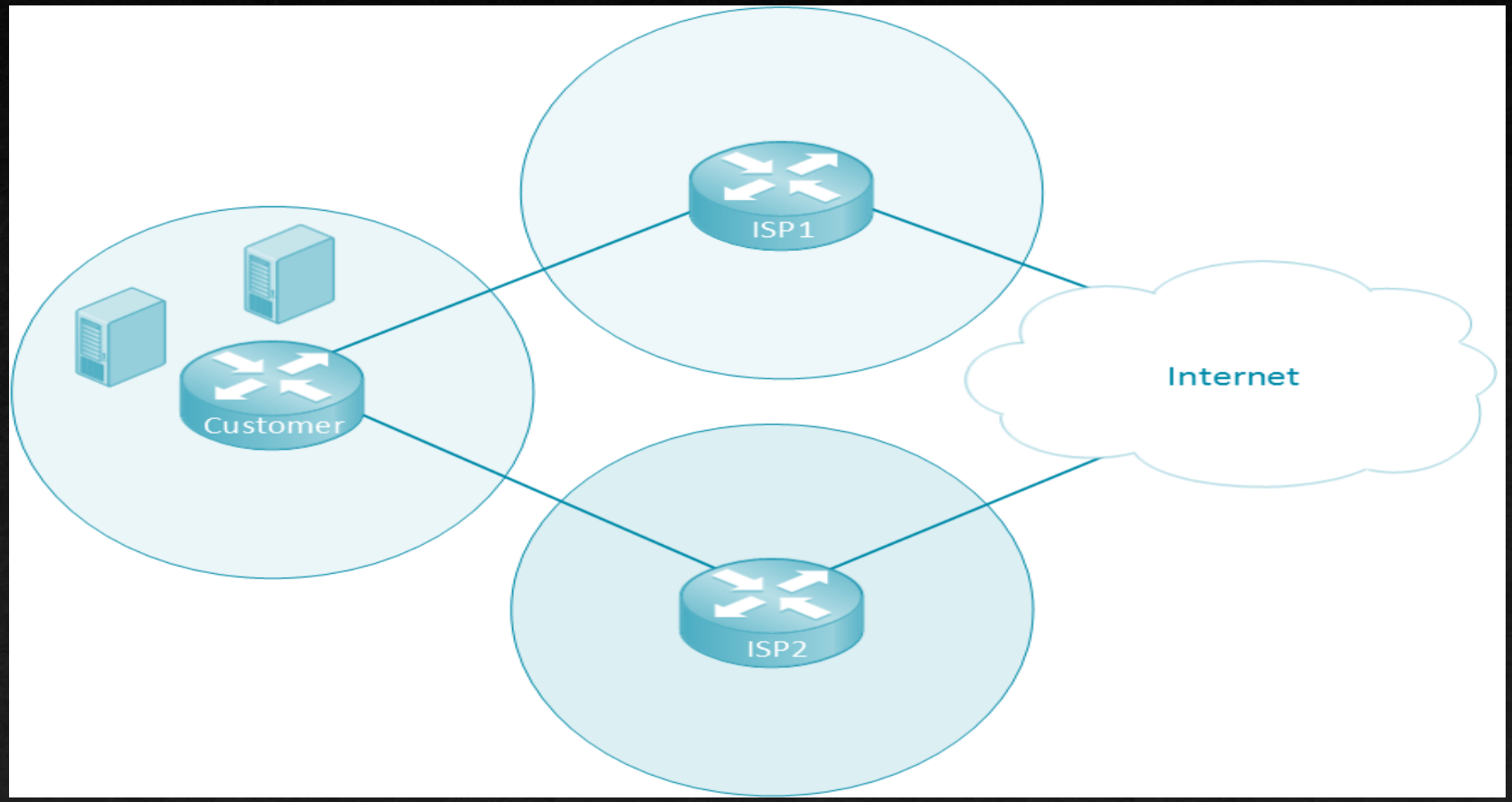




Why We Need BGP? (Cont.)

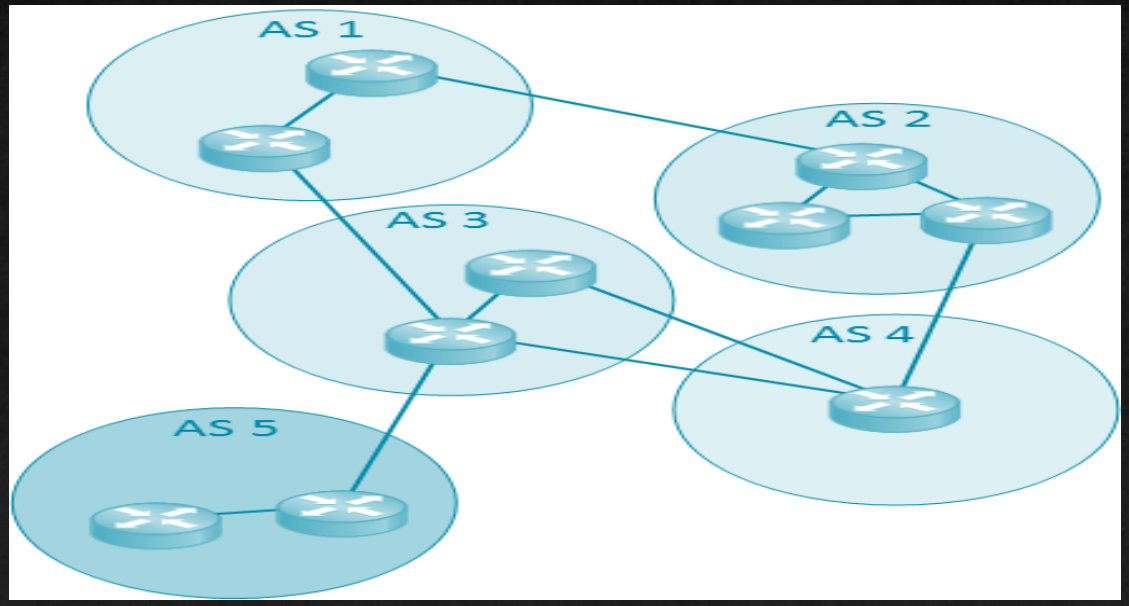


Why We Need BGP? (Cont.)



Autonomous Systems

- An AS is a collection of networks under a single administrative domain.
- The Internet is nothing more but a bunch of autonomous systems that are connected to each other.
- An organization requiring connectivity to the internet must obtain an ASN.
- Range 1 – 64511 are globally unique AS numbers and range 64512 – 65535 are private autonomous system numbers.



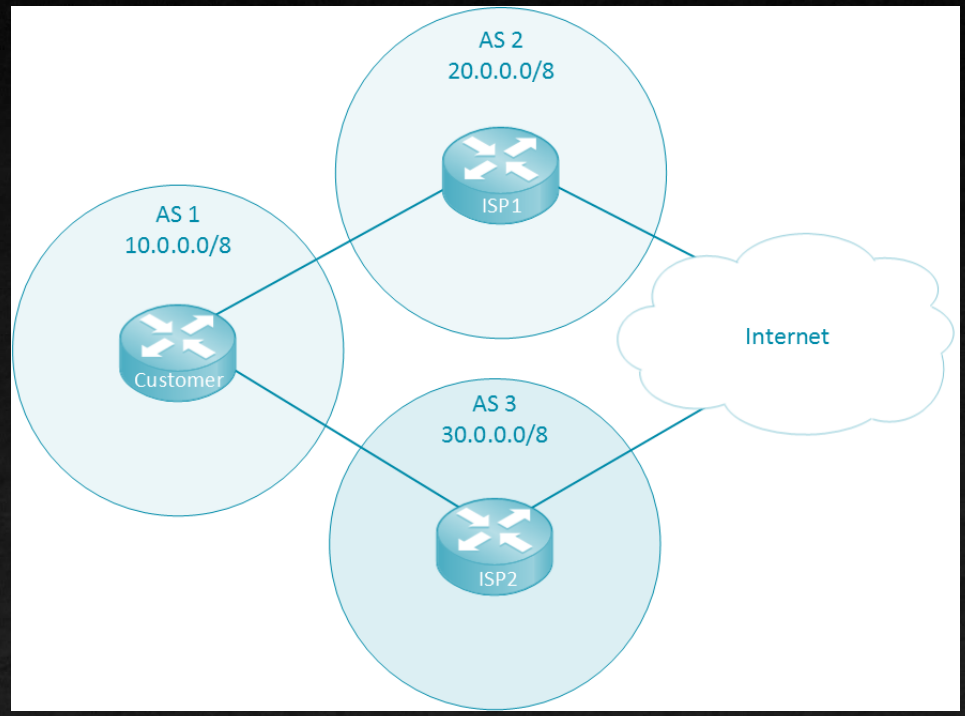
BGP Sessions

- A BGP session is an established adjacency between two BGP routers.
- Multi-hop sessions require that the router use an underlying route installed in the Routing Information Base (RIB) (static or from any routing protocol) to establish the TCP session with the remote endpoint.

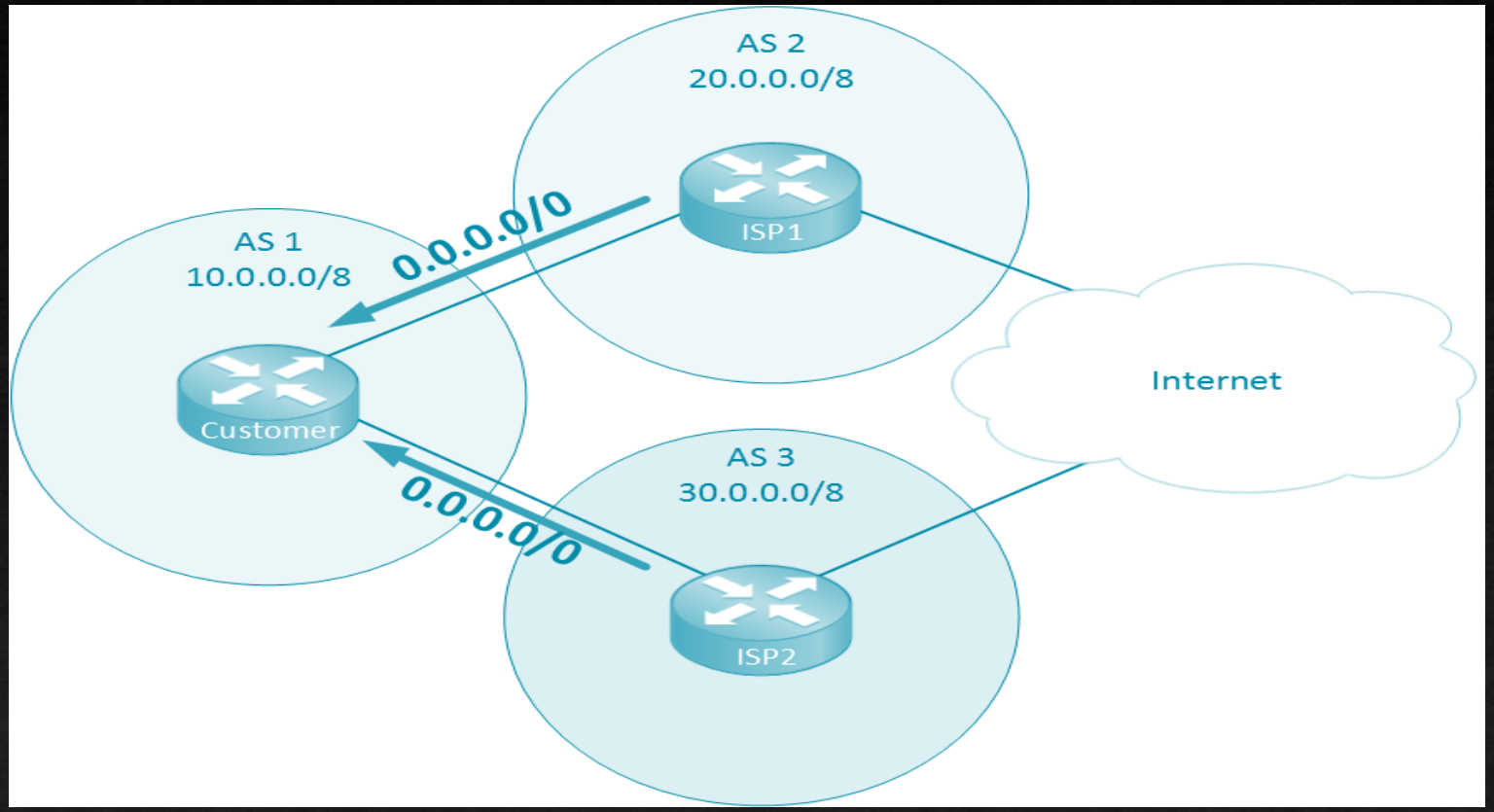
- BGP sessions are categorized into two types:
  - Internal BGP (iBGP) - Sessions established with an iBGP router that are in the same AS or that participate in the same BGP confederation. AD Value is 200.
  - External BGP (eBGP) - Sessions established with a BGP router that is in a different AS. AD value is 20.

BGP Advertisements

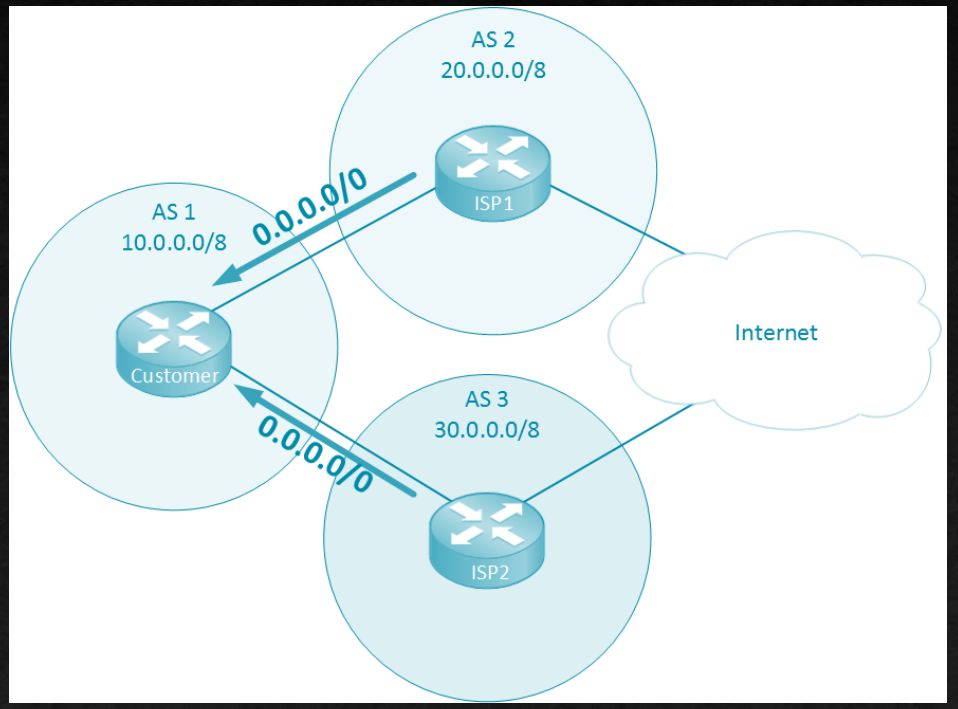
- Downside:
- Customer can use BGP to advertise his address space to the ISPs
- Upside:
- ISPS can Advertise 3 things to customer through BGP-
  - They advertise only a default route.
  - They advertise a default route and a partial routing table.
  - They advertise the full Internet routing table.

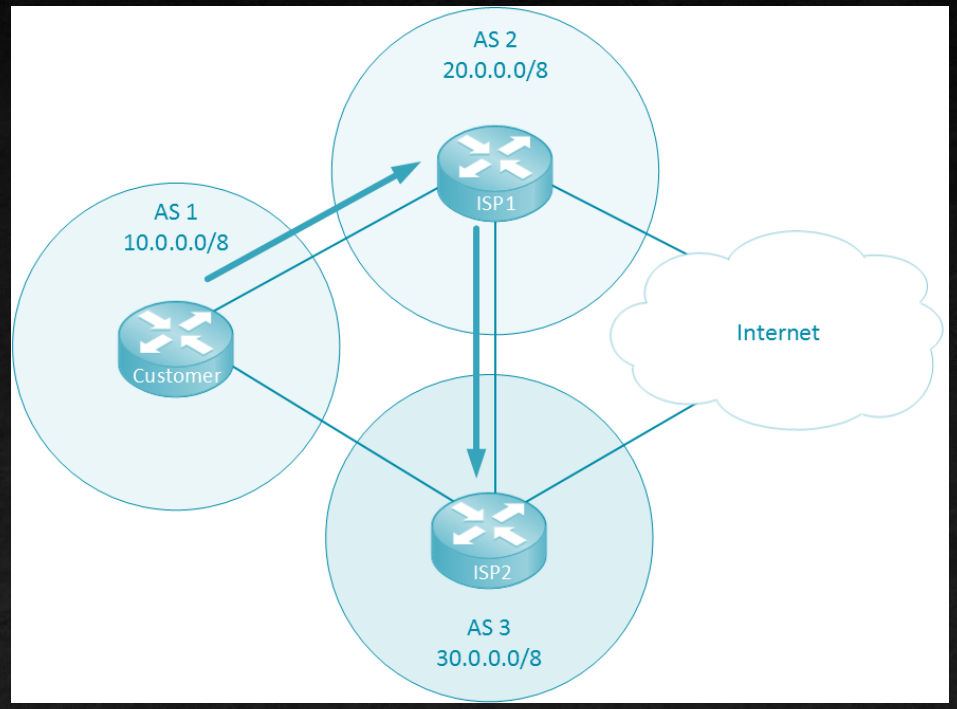


Default Route

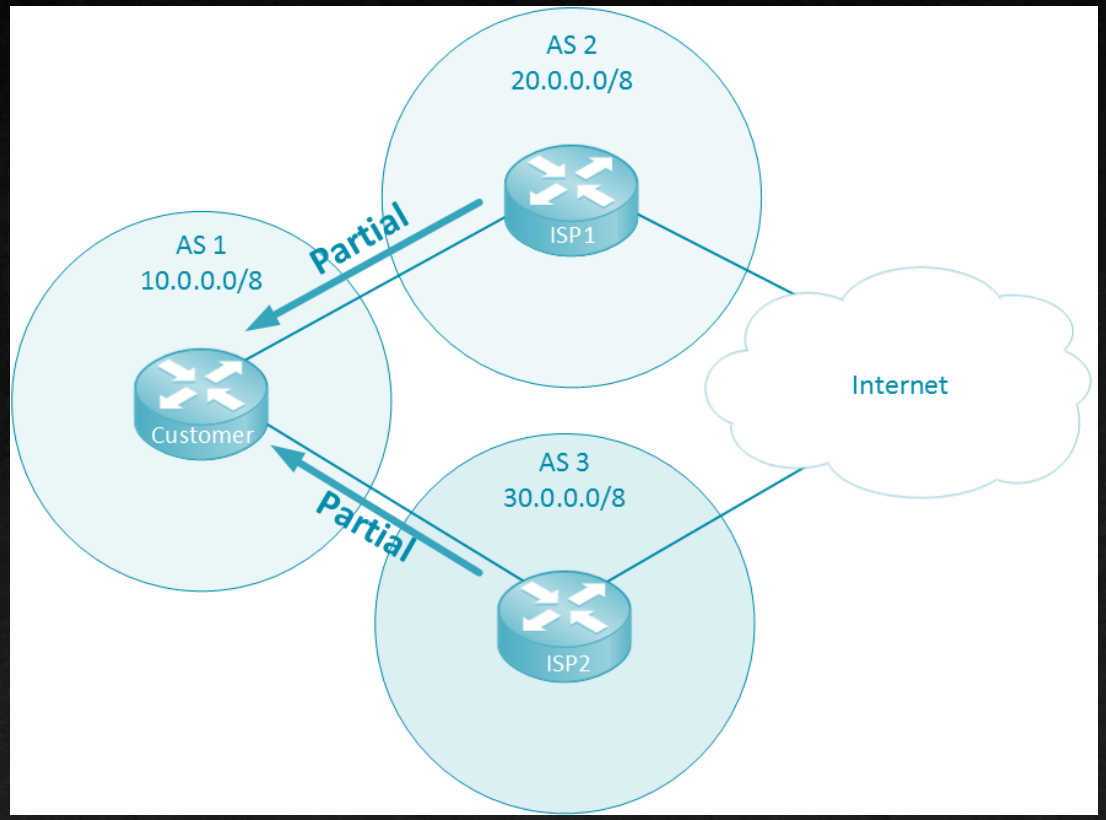


Default Route



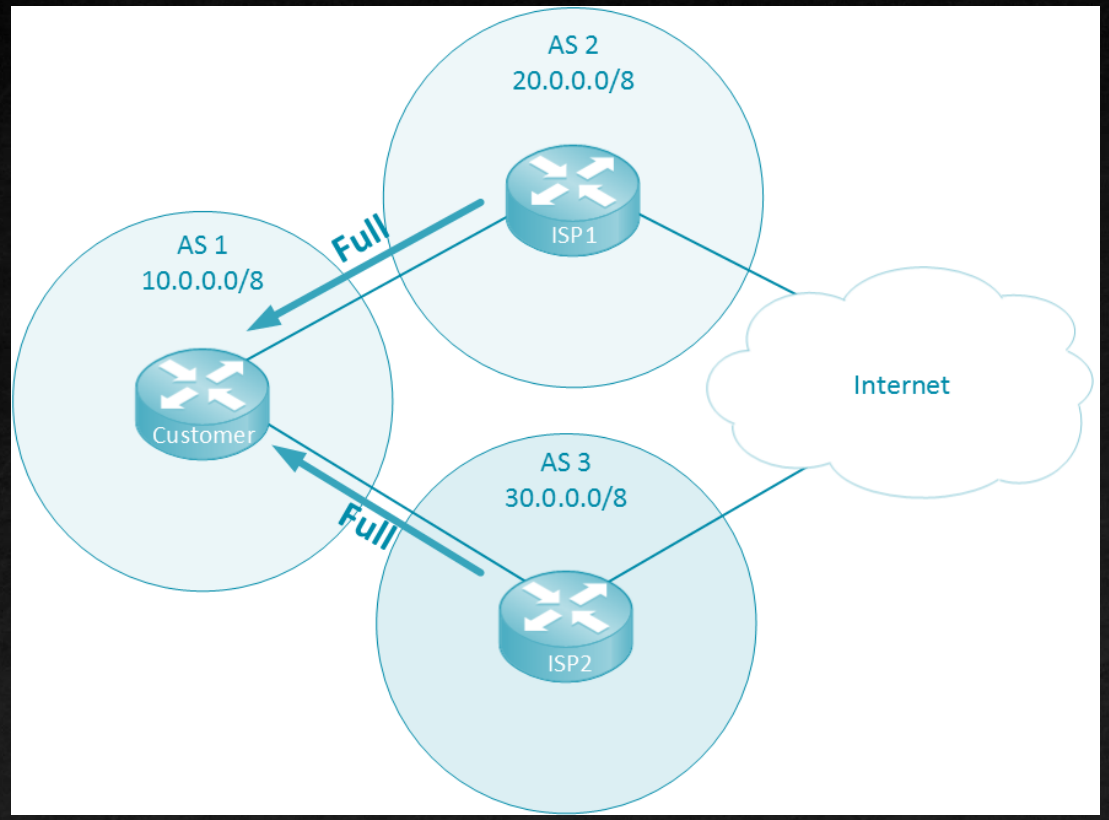


Partial Route



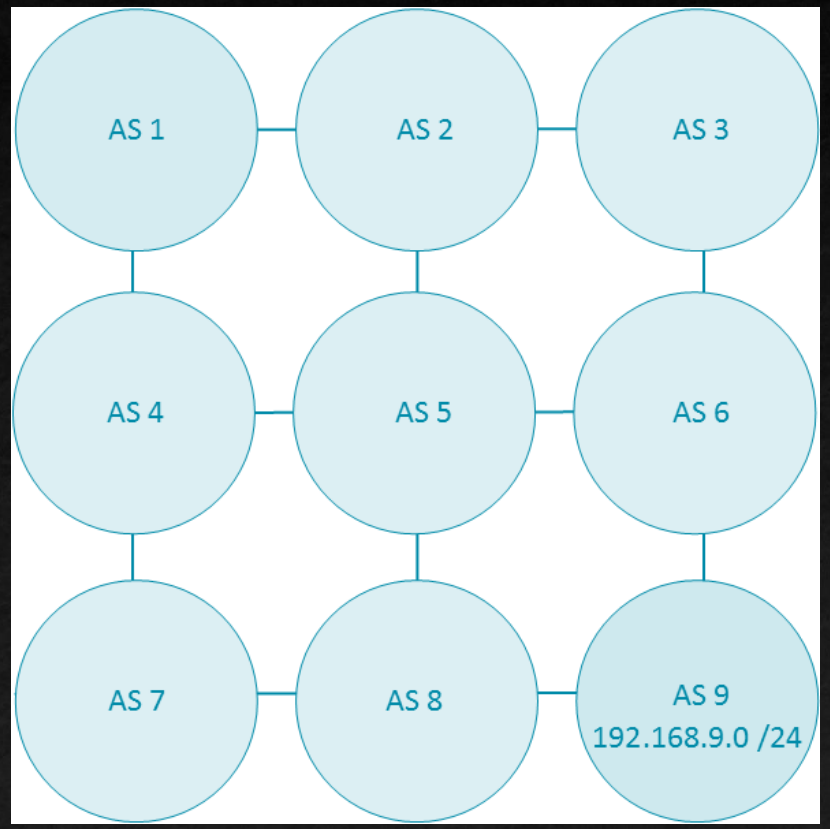
- ISPs can send a partial routes plus a default route to the customer.
- This partial update might include all the IP address space that the ISPs have assigned to their customers.

Full Internet Routing Table



- The last option ISPs can send full route table to the customer
- This requires more resources, but customer will be able to make the best routing decisions.

BGP Path Vector and Route Selection



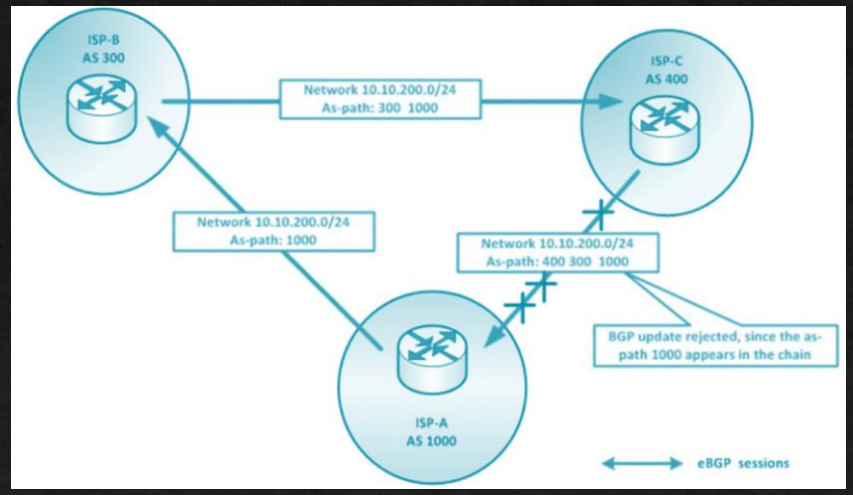
- BGP is called a path vector routing protocol.
- It store the prefix but also the paths it has to cross in order to get the destination.
- Paths are the autonomous systems we have to get through in order to get the destination.
- We can use only exit path to make it best path.
- BGP uses a set of BGP attributes to select a path.

Path Attributes

- BGP uses path attributes (PAs) associated with each network path. The PAs provide BGP with granularity and control of routing policies in BGP.
- The BGP prefix PAs are classified as any of the following:
  - Well-known mandatory: must be included with every prefix advertisement
  - Well-known discretionary: may or may not be included with the prefix advertisement
  - Optional transitive: stays with the route advertisement from AS to AS
  - Optional non-transitive: cannot be shared from AS to AS
- In BGP, the Network Layer Reachability Information (NLRI) is the routing update that consists of the network prefix, prefix length, and any BGP PAs for that specific route.

Loop Prevention

- BGP is a path vector routing protocol and does not contain a complete topology of the network, as do link-state routing protocols.
- BGP behaves like distance vector protocols, ensuring that a path is loop free.
- The BGP attribute AS_Path is a well-known mandatory attribute and includes a complete list of all the ASNs that the prefix advertisement has traversed from its source AS.
- AS_Path is used as a loop-prevention mechanism in BGP. If a BGP router receives a prefix advertisement with its AS listed in AS_Path, it discards the prefix because the router thinks the advertisement forms a loop.

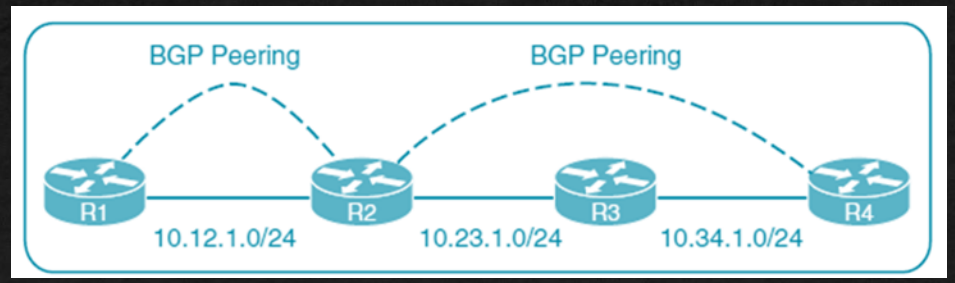


Address Families

- BGP was intended for only routing IPv4 prefixes, but RFC 2858 added Multi-Protocol BGP (MP-BGP) capability by adding an extension called the address family identifier (AFI).
- An address family correlates to IPv4 or IPv6, and additional granularity is provided through a subsequent address family identifier (SAFI), such as unicast or multicast.
- Multiprotocol BGP (MP-BGP) achieves this separation by using the BGP path attributes (PAs) MP_REACH_NLRI and MP_UNREACH_NLRI. These attributes are held inside BGP update messages and are used to carry network reachability information for different address families.
- Every address family maintains a separate database and configuration for each protocol (address family plus sub-address family) in BGP. This allows for a routing policy in one address family to be different from a routing policy in a different address family.
- BGP includes an AFI and a SAFI with every route advertisement to differentiate between the AFI and SAFI databases.

Inter-Router Communication

- BGP does not use hello packets to discover neighbors, and it cannot discover neighbors dynamically.
- A BGP session refers to the established adjacency between two BGP routers.
- BGP neighbors are defined by IP address and BGP uses TCP port 179 to communicate with other routers which can cross networks (multi-hop capable).
- BGP can form directly connected neighbor adjacencies as well as adjacencies that are multiple hops away.
- Multi-hop sessions require that the router use an underlying route installed in the RIB (static or from any routing protocol) to establish the TCP session with the remote endpoint.



BGP States

- Idle State
  - Refueses all BGP connections
  - Initiates a TCP connection to its configured BGP peers.
  - Listen for incoming TCP connections
  - ConnectRetry time is started. (120Seconds – cannot be changed)
- Connect state
  - Waits for the TCP connection to be completed
  - If TCP is successful:
    - Sends a OPEN message to its peers
    - Changes state to OpenSend and clears the CR timer.
  - If TCP is unsuccessful:
    - ConnectRetry timer is reset
    - BGP goes into active state

- Active State
  - The BGP speaker tries to acquire its Peer by initiating another TCP connection.
  - IF TCP is successful:
    - ConnectRetry Timer is clear
    - Sends OPEN Message to Peer and changes state to OpenSent
  - If TCP is Unsuccessful:
    - ConnectRetry timer is reset
    - BGP goes into idle state
- OpenSent State
  - Waits for OPEN messages from its peer and verifies all the fields once its received.
    - Contain BGP version, AS Number, Hold time & BGP ID
  - If is successful:
    - BGP starts sending keepalive and sets its hold and keep alive timers.
  - If is unsuccessful:
    - ConnectRetry timer is reset
    - BGP goes into idle state

- OpenConfirm State
  - BGP Peers wait for the keepalives to be received
  - If is successful:
    - State is changed to established
  - If is unsuccessful:
    - ConnectRetry times is reset
    - BGP goes into idle state
- Established State
  - Exchange of update, notification and keepalives messages takes place with peer.
  - Each update or keepalive received results in the hold time reset. Default Keepalive is 60sec and holdtime is 180sec.

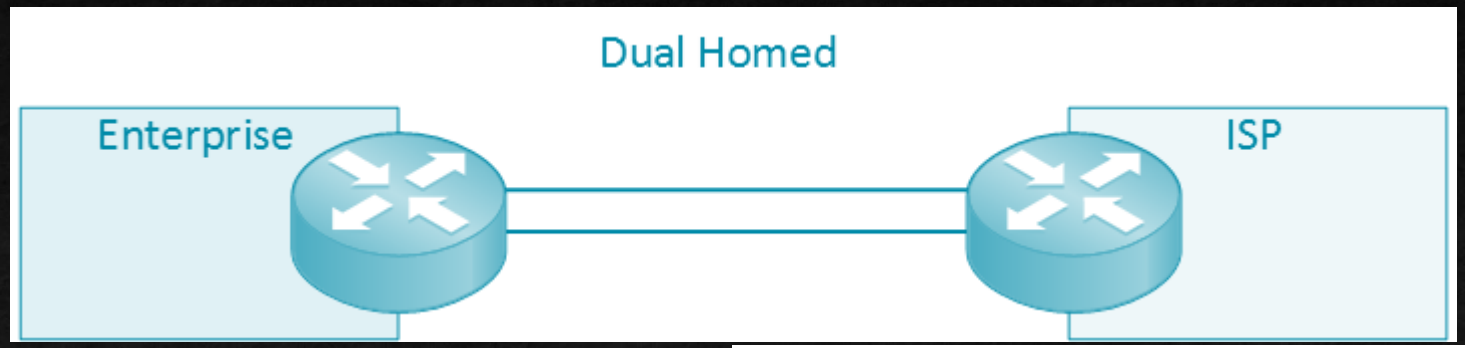
BGP Messages

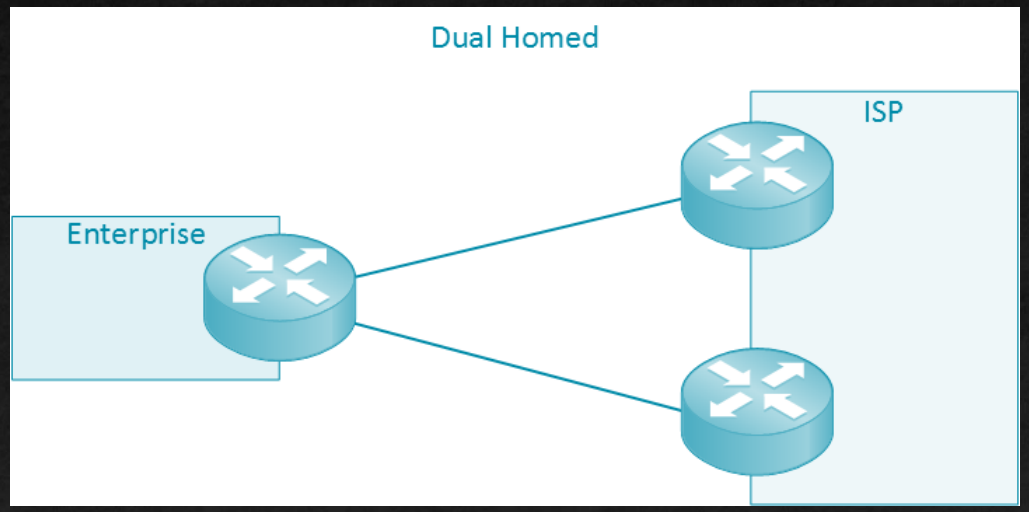
- BGP communication uses four message types: OPEN, UPDATE, NOTIFICATION and KEEPALIVE
- OPEN – OPEN is used to set up and establish a BGP adjacency. The OPEN message contains the BGP version number, the ASN of the originating router, the hold time, the BGP identifier, and other optional parameters that establish the session capabilities.
- KEEPALIVE - These messages are exchanged ever one-third of the hold timer agreed upon between the two BGP routers. Cisco devices have a default hold time of 180 seconds, and a default keepalive interval of 60 seconds. If the hold time is set to 0, no KEEPALIVE messages are sent between the BGP neighbors.
- UPDATE – This message advertises any feasible routes, withdraws previously advertised routes, or both. The UPDATE message holds the NLRI, which includes the prefix and associated BGP PAs when advertising prefixes. Withdrawn NLRI routes include only the prefix. An UPDATE message can act as a keepalive to reduce unnecessary traffic.
- NOTIFICATION – This message is sent when an error is detected with the BGP session, such as a hold timer expiring, neighbor capabilities changing, or a BGP session reset being requested. This causes the BGP connection to close.

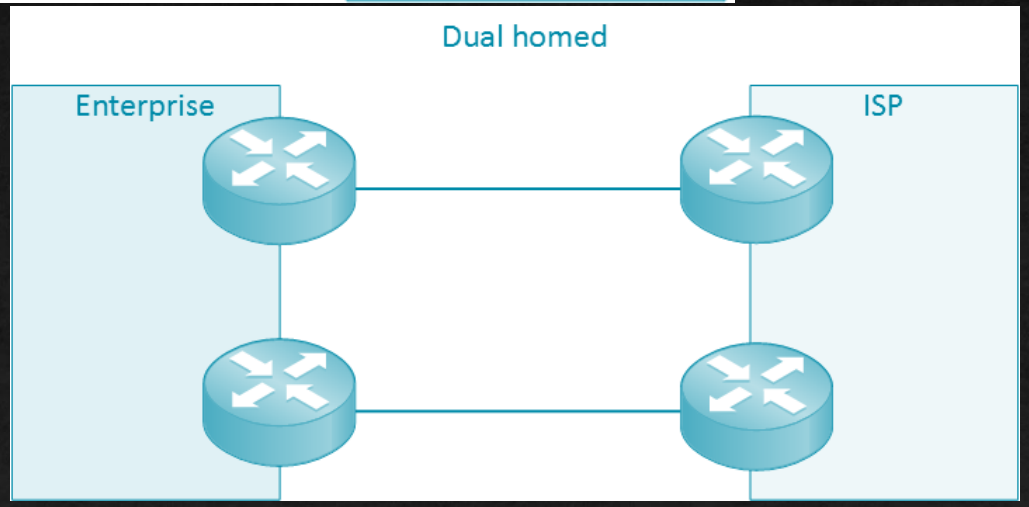
Single/Dual Homed and Multi-homed Designs

- Single Homed: Connected to a single ISP using a single link

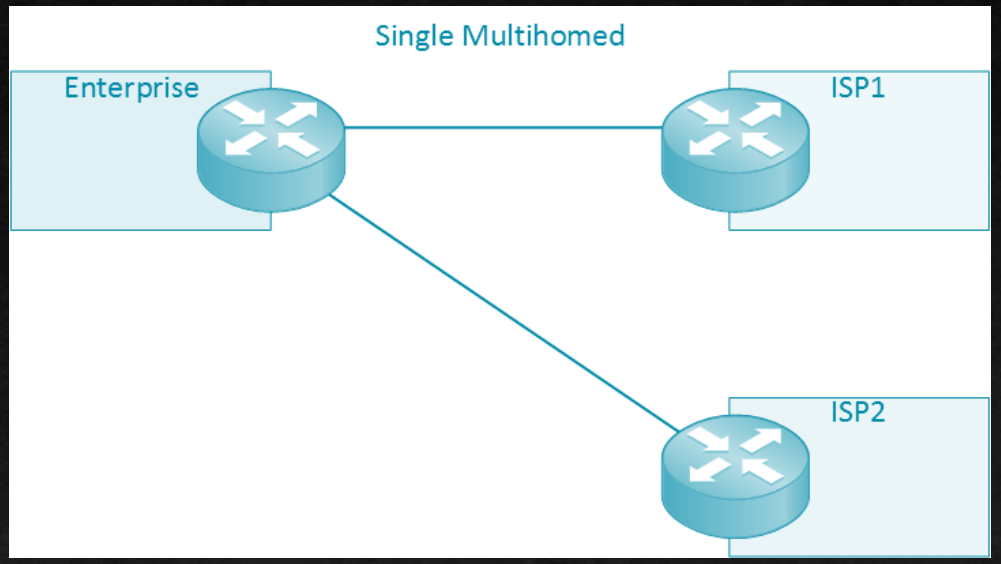
- Dual homed: Connected to a single ISP using dual links.

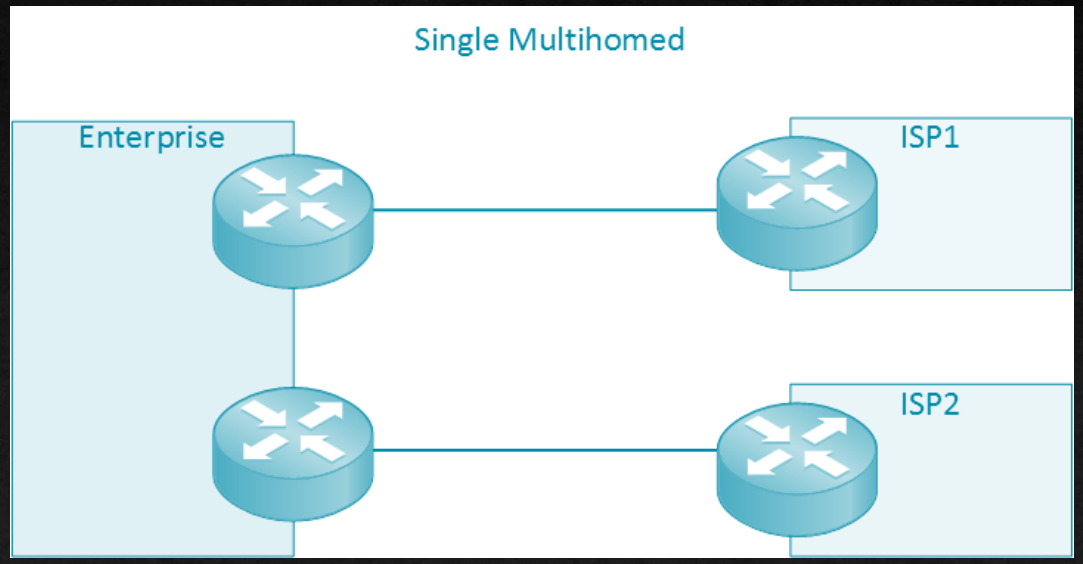




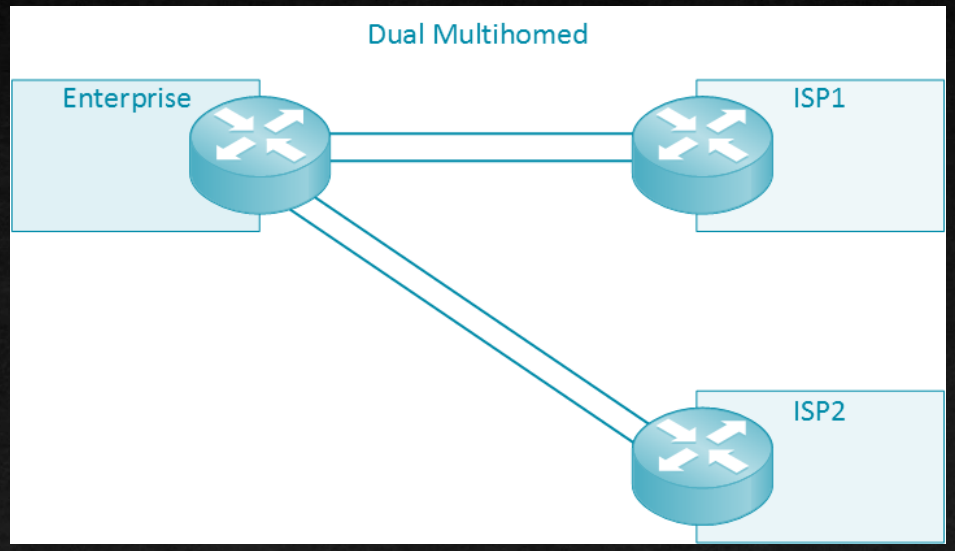


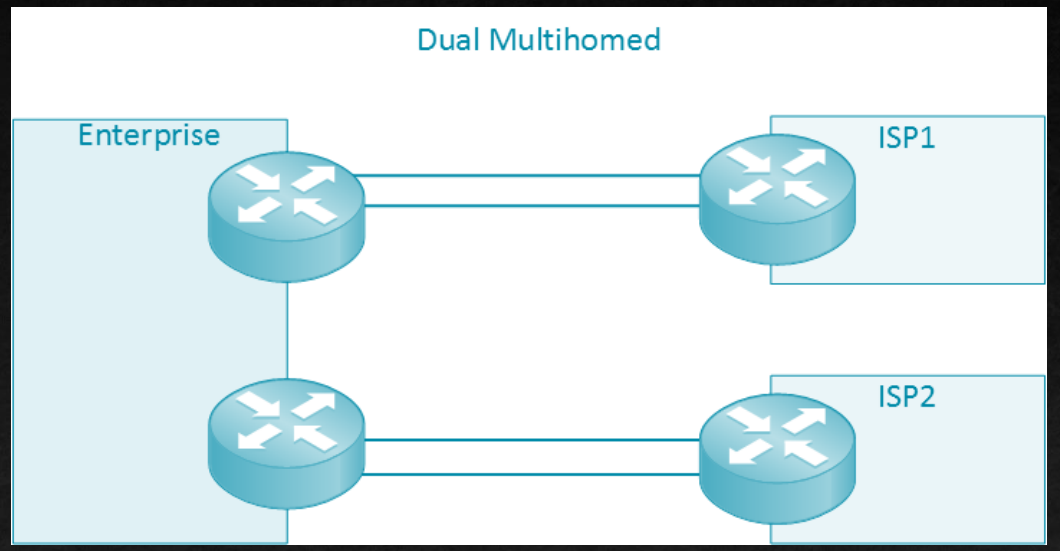
- Single multi-homed: Connected to two ISPs using single links.



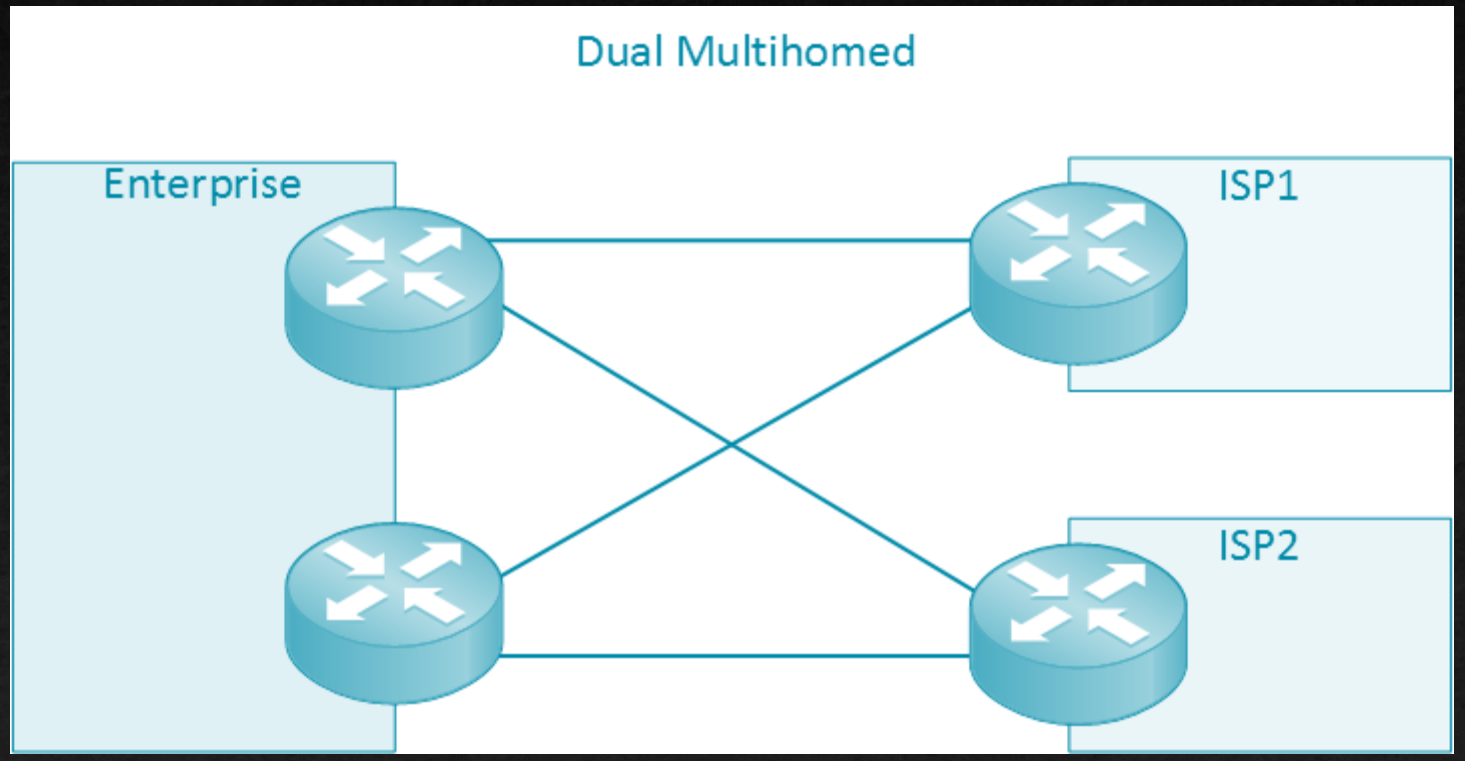


- Dual multi-homed: Connected to two ISPs using dual links.



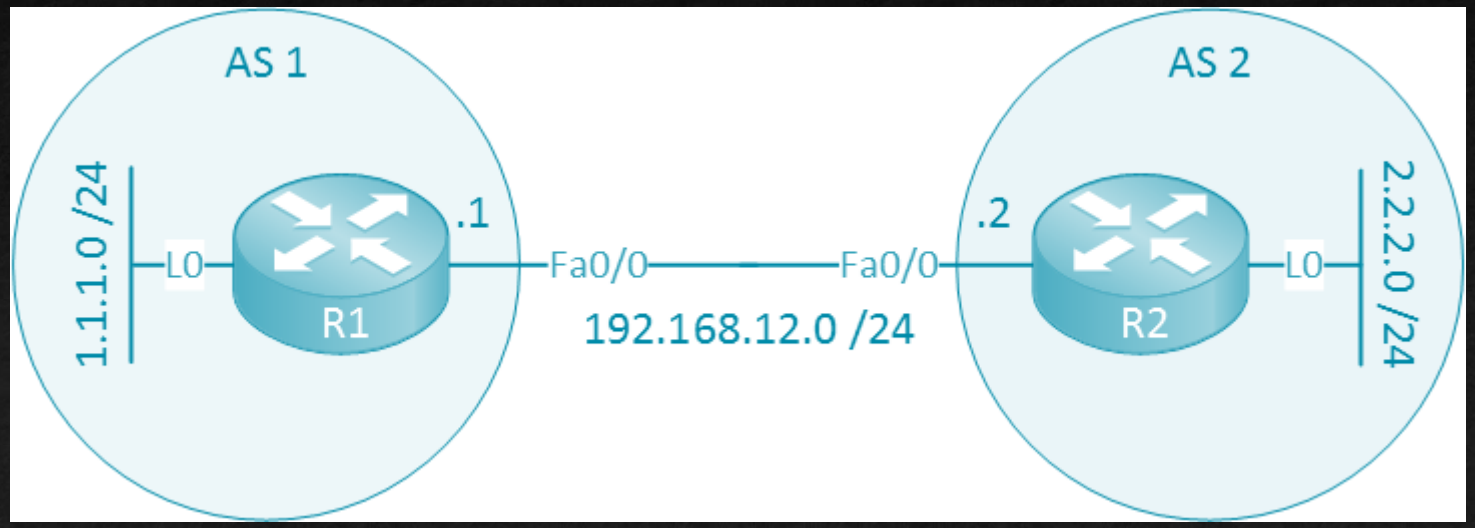


- Dual multi-homed: Connected to two ISPs using dual links.



EBGP

- External Border Gateway Protocol (EBGP) is a Border Gateway Protocol (BGP) extension that is used for communication between distinct autonomous systems (AS). EBGP enables network connections between autonomous systems and autonomous systems implemented with BGP.



EBGP Configuration

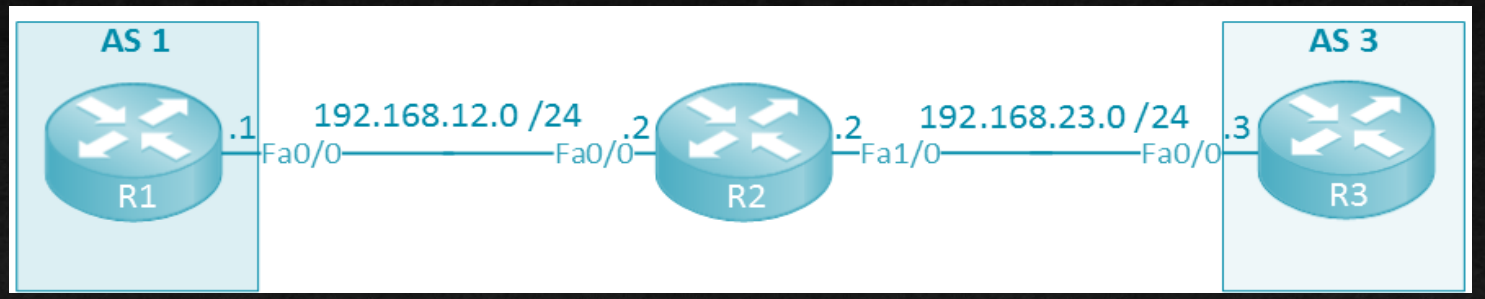
- R1(config)#router bgp 1
- R1(config-router)#neighbor 192.168.12.2 remote-as 2
- R1(config-router)#neighbor 192.168.12.2 password LAB
- R1(config-router)#network 1.1.1.0 mask 255.255.255.0

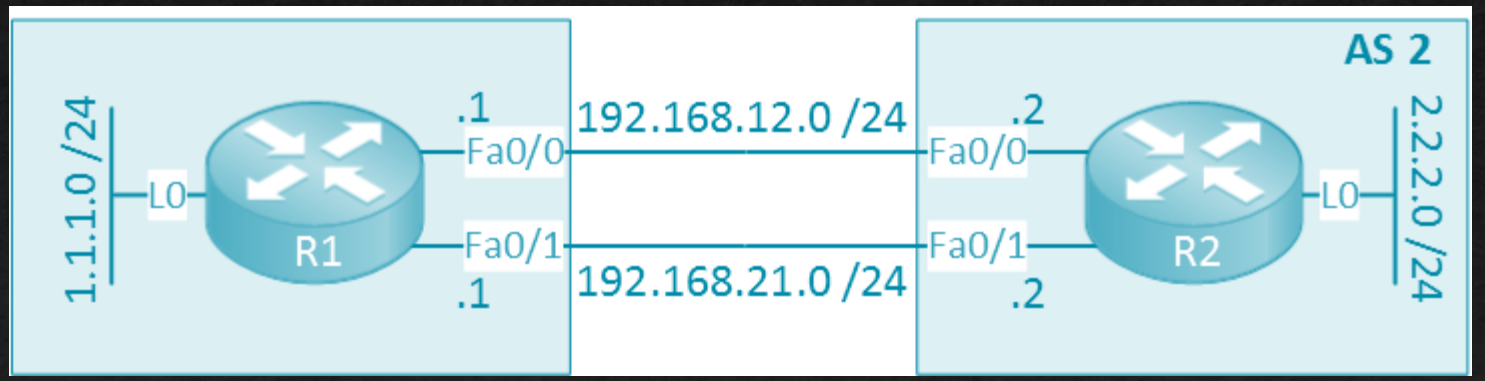
- R2(config)#router bgp 2
- R2(config-router)#neighbor 192.168.12.1 remote-as 1
- R2(config-router)#neighbor 192.168.12.1 password LAB
- R2(config-router)#network 2.2.2.0 mask 255.255.255.0

- #show ip bgp summary
- #show ip bgp
- #show ip route bgp

EBGP Multihop

- eBGP (external BGP) by default requires two routers to be directly connected to each other in order to establish a neighbor adjacency. This is because eBGP routers use a TTL of one for their BGP packets. When the BGP neighbor is more than one hop away, the TTL will decrement to 0 and it will be discarded.





EBGP Multihop Configuration

- R1(config)#ip route 2.2.2.0 255.255.255.0 192.168.12.2
- R1(config)#ip route 2.2.2.0 255.255.255.0 192.168.21.2

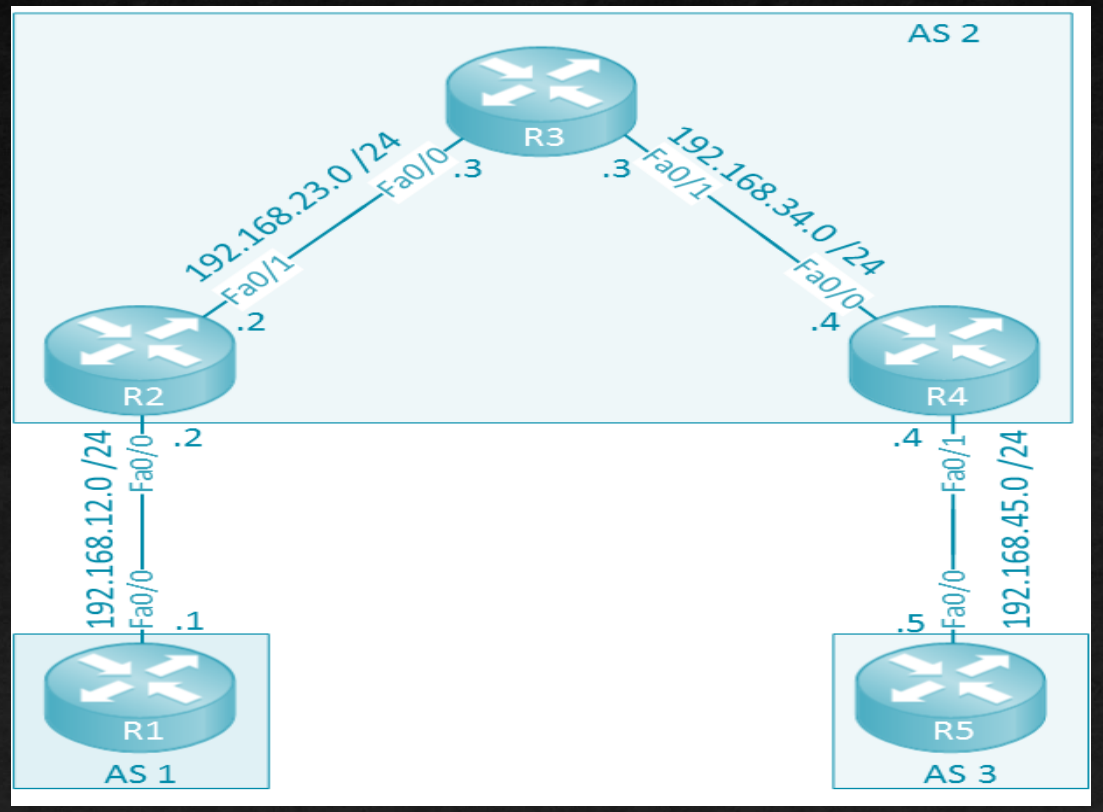
- R2(config)#ip route 1.1.1.0 255.255.255.0 192.168.12.1
- R2(config)#ip route 1.1.1.0 255.255.255.0 192.168.21.1

- R1(config)#router bgp 1
- R1(config-router)#neighbor 2.2.2.2 remote-as 2
- R1(config-router)#neighbor 2.2.2.2 update-source loopback 0
- R1(config-router)#neighbor 2.2.2.2 ebgp-multihop 2

- R2(config)#router bgp 2
- R2(config-router)#neighbor 1.1.1.1 remote-as 1
- R2(config-router)#neighbor 1.1.1.1 update-source loopback 0
- R2(config-router)#neighbor 1.1.1.1 ebgp-multihop 2

iBGP

- iBGP is used inside the autonomous systems. It is used to provide information to your internal routers. It requires all the devices in same autonomous systems to form full mesh topology or either of Route reflectors and Confederation for prefix learning.



iBGP Sample Config

- On R2

```
router bgp 2
 bgp log-neighbor-changes
 neighbor 3.3.3.3 remote-as 2
 neighbor 3.3.3.3 update-source Loopback0
 neighbor 3.3.3.3 next-hop-self
 neighbor 4.4.4.4 remote-as 2
 neighbor 4.4.4.4 update-source Loopback0
 neighbor 4.4.4.4 next-hop-self
```

- On R3

```
router bgp 2
 bgp log-neighbor-changes
 neighbor 2.2.2.2 remote-as 2
 neighbor 2.2.2.2 update-source Loopback0
 neighbor 4.4.4.4 remote-as 2
 neighbor 4.4.4.4 update-source Loopback0
```

Understanding BGP Table

- The * means that this is a valid route and that BGP is able to use it.
- The > means that this entry has been selected as the best path.
- The next hop of 0.0.0.0 means that this network originated on this router.
- Path will show the AS path.
- The ‘i’ is the origin code and indicates that this network was advertised into BGP using the network command. And redistribute something into BGP it will show up with the ? symbol.
- supressed: BGP knows the network but won’t advertise it, this can occur when the network is part of a summary.
- damped: BGP doesn’t advertise this network because it was flapping too often (network appears, disapears, appears, etc.) so it got a penalty.
- history: BGP learned this network but doesn’t have a valid route at the moment.
- RIB-failure: BGP learned this network but didn’t install it in the routing table. This occurs when another routing protocol with a lower administrative distance also learned it.
- stale: this is used for non-stop forwarding, this entry has to be refreshed when the remote BGP neighbor has returned.

QnA