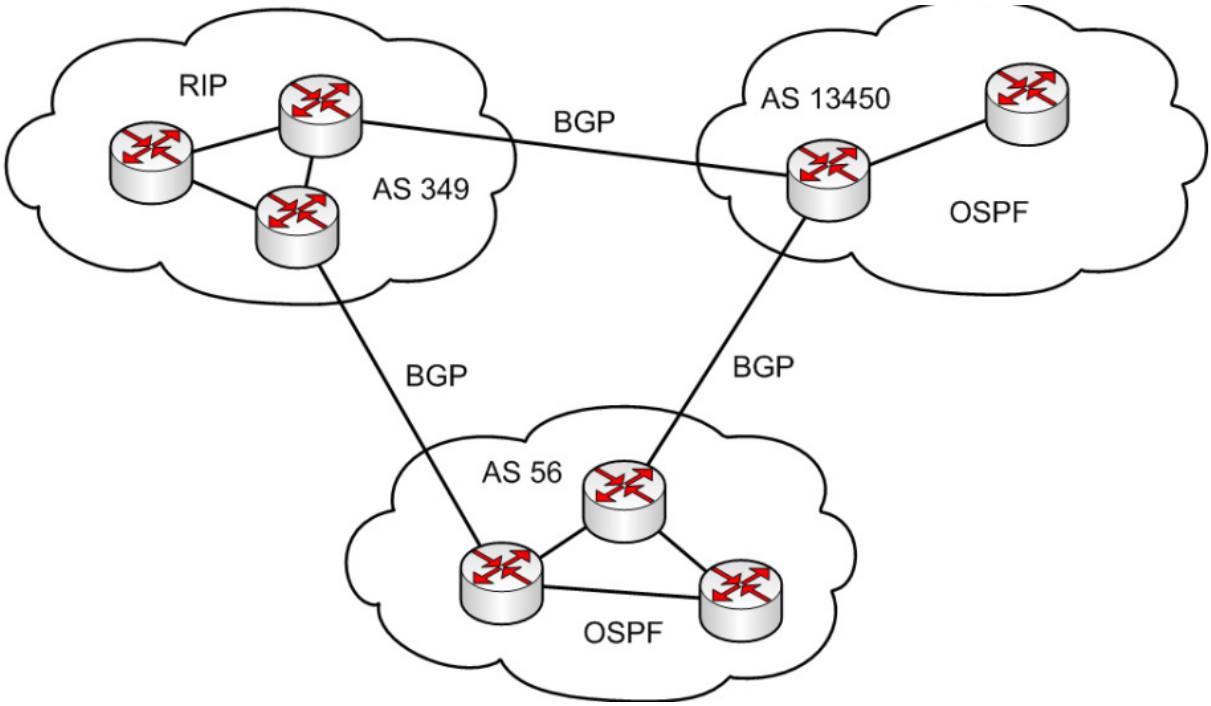
Border Gateway Protocol (BGP)

BGP, or Border Gateway Protocol, is used to exchange routing information between autonomous systems (AS), which are primarily networks operated by Internet Service Providers (ISPs). BGP-4 is the current version, replacing the earlier Exterior Gateway Protocol (EGP) used during the NSFNET era. Unlike internal routing protocols that optimize for the most technically efficient path, BGP also considers non-technical factors like political and commercial agreements.

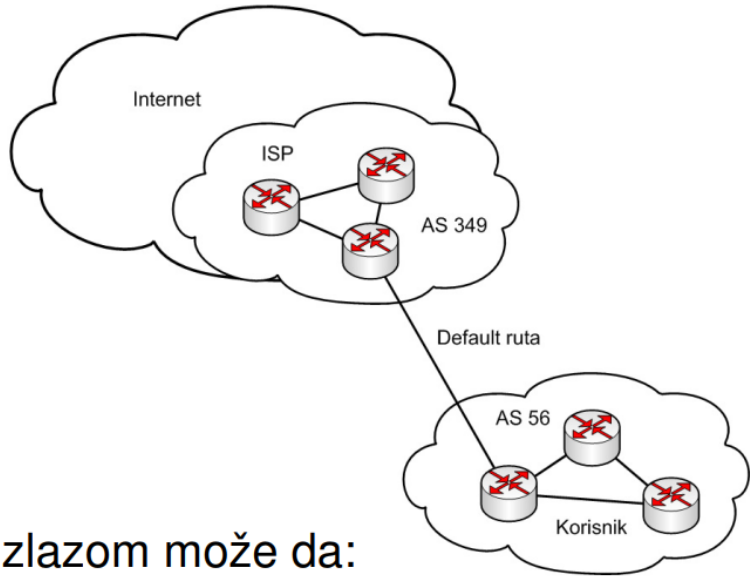
Autonomous Systems



Autonomous systems are identified by unique AS numbers. The range from 64,512 to 65,535 is reserved for private use. Types of AS include:

- Single-homed (one exit point)
- Multi-homed nontransit (multiple exit points, no transit traffic)
- Multi-homed transit (multiple exit points, with transit traffic)

Single-homed AS

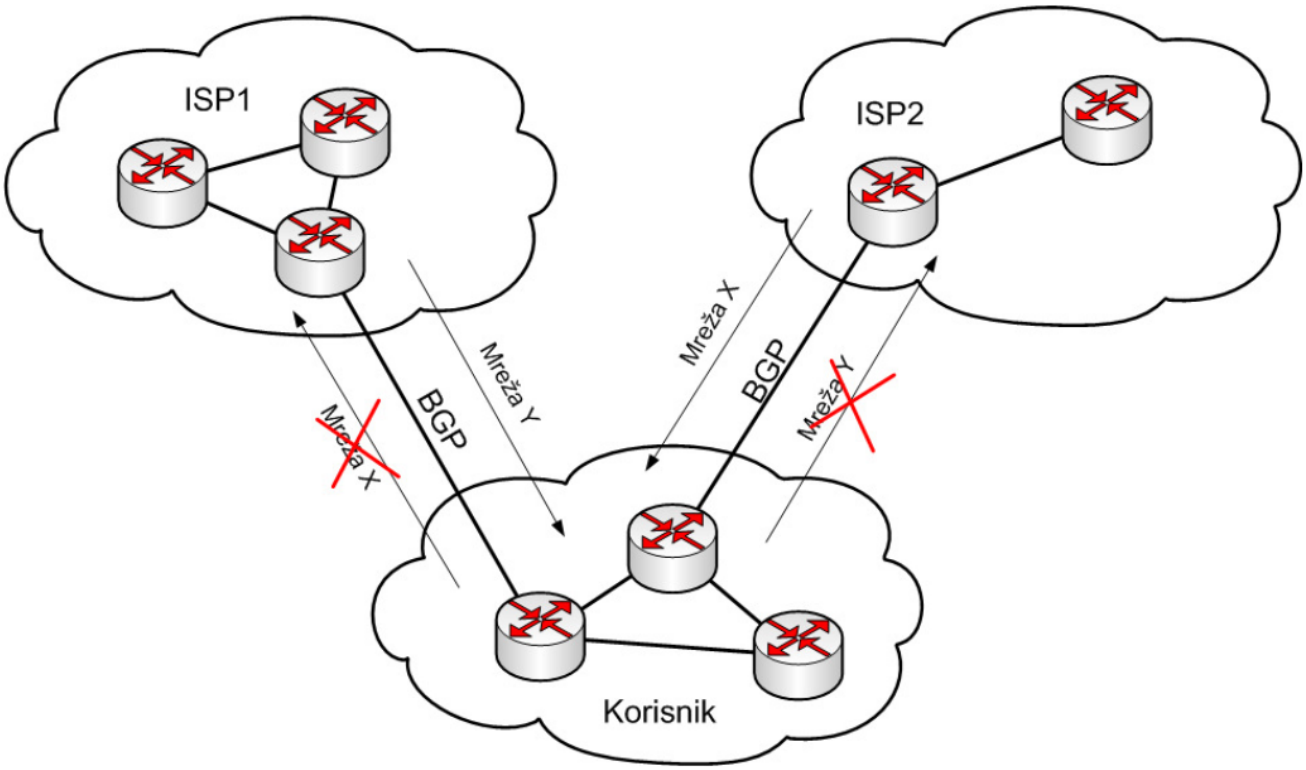


AS sa jednim izlazom može da:

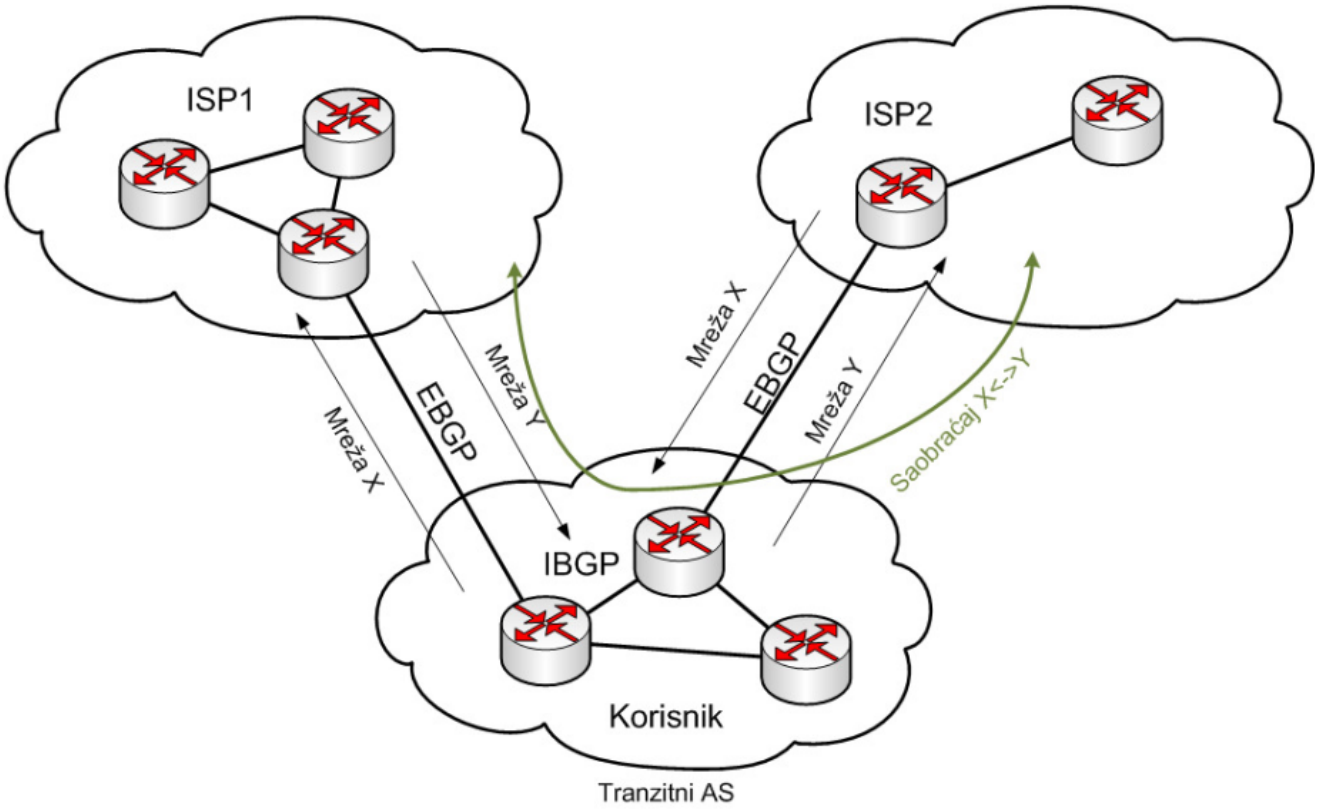
A single-homed AS can:

- Use a default route to forward traffic to the Internet.
- Be part of its provider's IGP protocol.
- Operate as a private AS within the provider's AS.

Multi-homed AS without Transit



Multi-homed AS with Transit



BGP može da se koristi:

- **izvan AS-a i onda je to eksterni BGP – EBGP**
- **unutar AS-a i onda je to interni BGP – IBGP**

When Not to Use BGP

BGP may not be used in situations where:

- There is only one connection to the Internet or ISP.
- The network's routing policy is the same as the ISP's policy.
- Border routers do not support BGP or lack the resources to run BGP.
- The bandwidth between two networks is very low.

Path Length to Internet Destinations

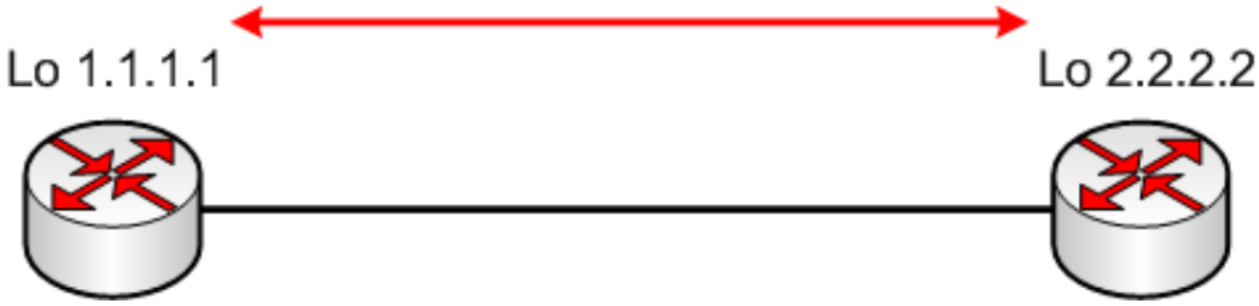
The distribution of AS path lengths to destinations on the Internet can be observed in BGP data from AS 6447 (University of Oregon) and AS 131072 (APNIC R&D).

Basic Characteristics of BGP

1. **Distance-Vector Protocol**: BGP is a distance-vector routing protocol with mechanisms to prevent routing loops.
2. **Routing Decisions**: BGP makes routing decisions based on rules set by the network administrator.
3. **BGP-4**: The current version, defined in RFC 1772 and updated to RFC 4271, supports CIDR and route aggregation.
4. **TCP-Based**: BGP uses TCP port 179 for message exchanges, requiring IP connectivity between two routers.

5. **Route Management**: BGP maintains a table of routes and the autonomous systems from which these routes were learned to prevent loops.
6. **Extensibility**: BGP is easily extendable, supporting functionalities like IPv6, VPNs, and Multicast through various RFCs (e.g., RFC 2283, RFC 2858, RFC 4760).

BGP Peers and Operation



When two routers establish a BGP connection, they become "BGP peers." Each router running BGP is known as a "BGP speaker." The process involves:

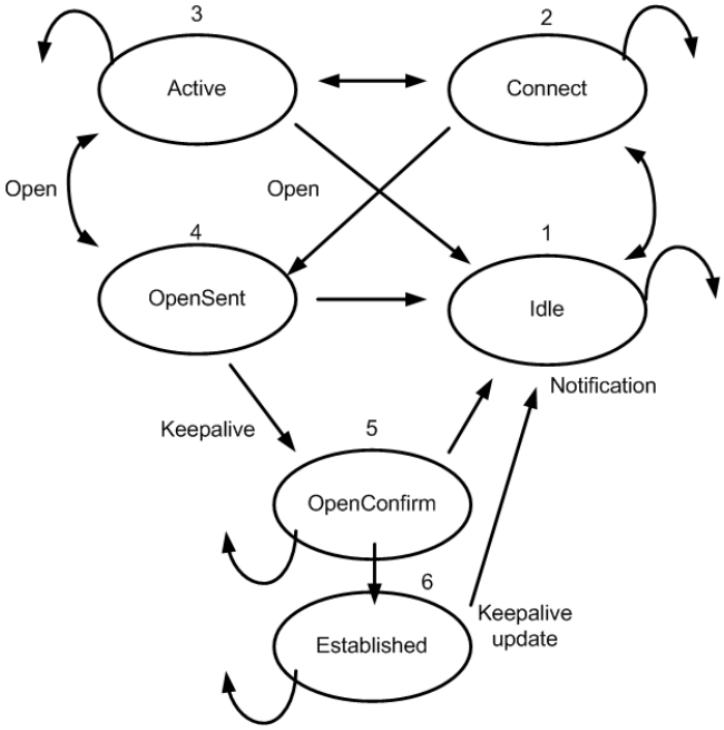
1. **Connection Establishment**: Routers exchange BGP messages to establish the connection (Open).
2. **Configuration Check**: If configuration parameters (AS numbers, IP addresses) do not match, the session fails, and Notification messages are sent.
3. **Route Exchange**: Once a session is established, routers exchange all known routes (Update). Subsequent updates are incremental.
4. **Keepalive Messages**: Sent every 60 seconds to maintain the session. These messages are 19 bytes in size.

BGP Message Types

BGP messages have a common header and include:

- **Open**: Establishes a BGP session.
- **Keepalive**: Maintains the session.
- **Notification**: Indicates errors.
- **Update**: Contains routing information such as prefixes, AS paths, and attributes.

BGP Session State Machine



45

The BGP session establishment involves several states:

1. **Idle**: Initial state waiting for an event to start the connection.
2. **Connect**: Attempts to establish a TCP session.
3. **Active**: Retries TCP session establishment if the initial attempt fails.
4. **OpenSent**: Waits for an Open message from the peer.
5. **OpenConfirm**: Waits for Keepalive or Update messages to transition to Established state.
6. **Established**: Regular operation where Update messages are exchanged.

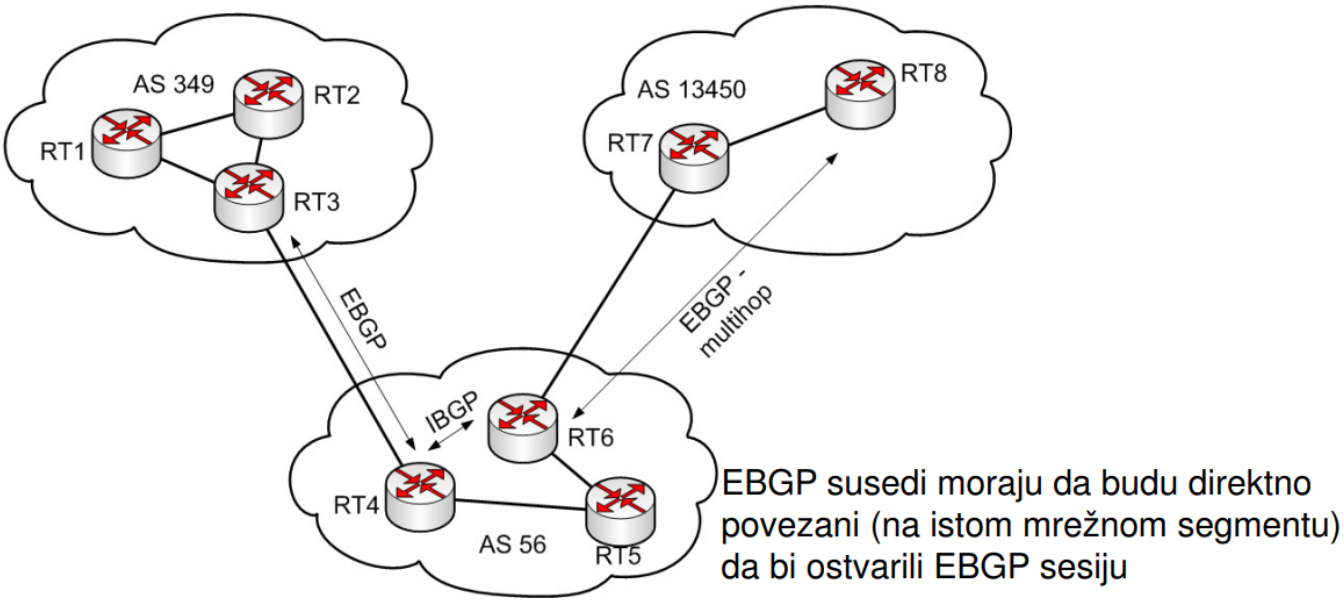
In case of any errors, BGP reverts to the Idle state.

Full Internet Routing Table Size (IPv4, 2023)

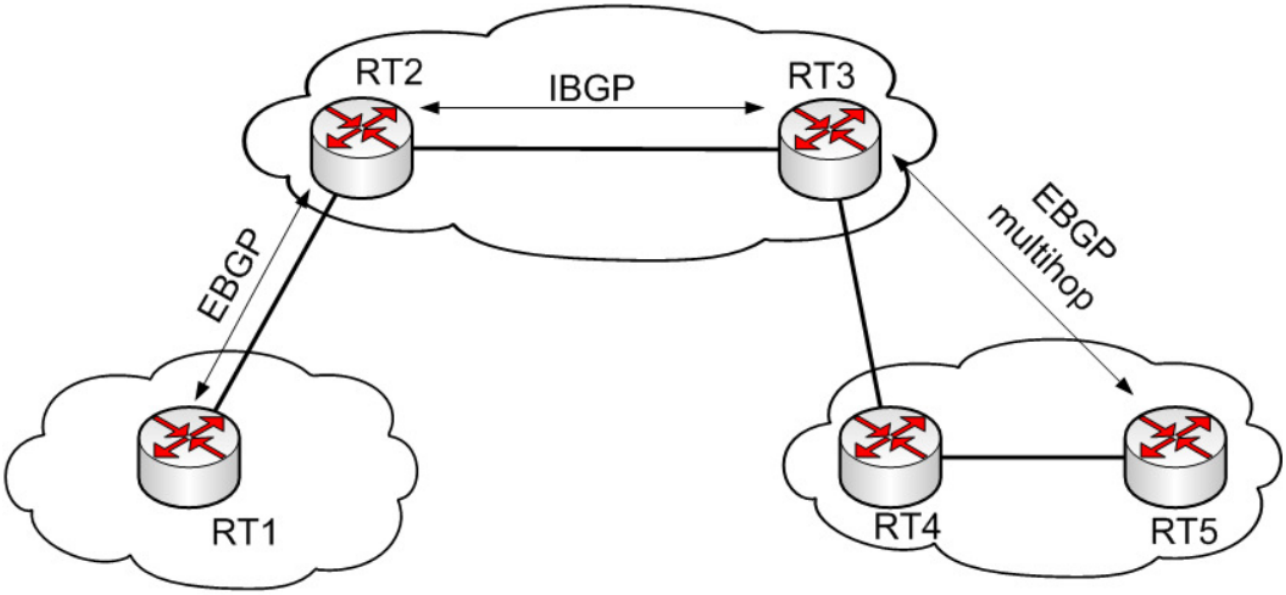
As the number of routes in the full Internet routing table grows, BGP routers must manage increasingly complex routing information to ensure efficient and accurate data delivery across the global Internet.

EBGP Multihop

External BGP (EBGP) neighbors typically need to be directly connected on the same network segment to establish an EBGP session. However, when direct connectivity is not possible, EBGP multihop is used to establish the session over multiple hops, requiring IP connectivity between the routers.



Different Types of BGP Sessions

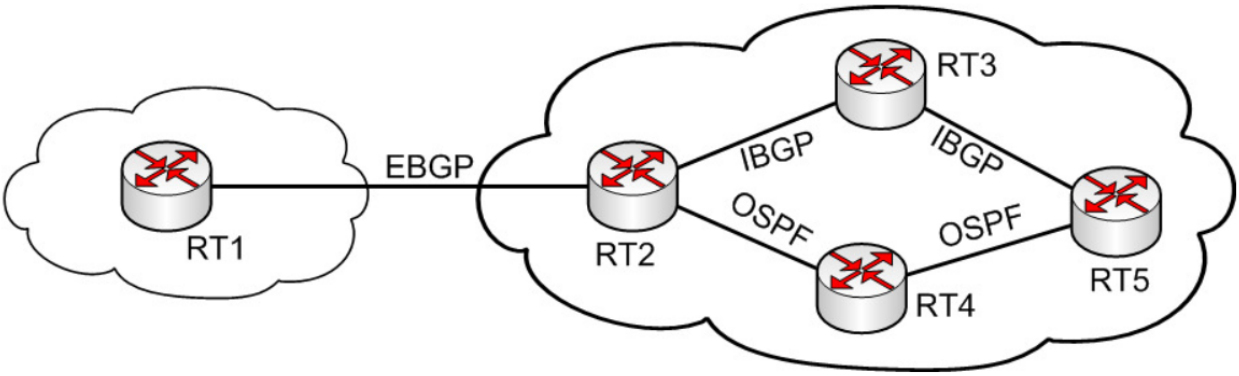


EBGP and IBGP

In BGP, routers do not advertise routes learned from their Internal BGP (IBGP) peers to other IBGP peers. This is to prevent routing loops within the autonomous system (AS).

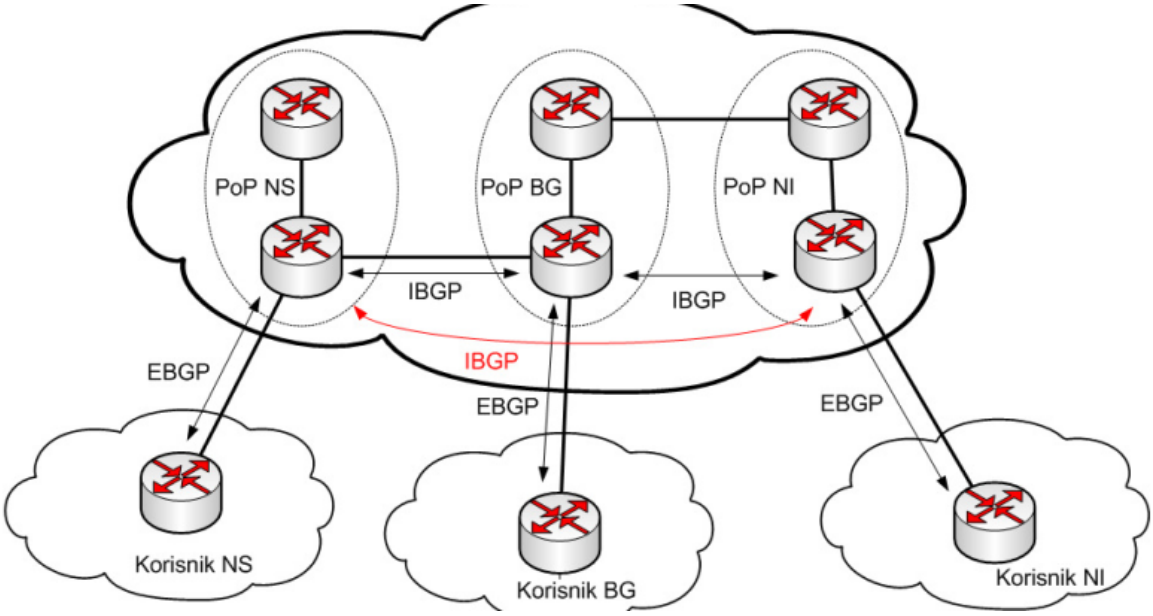
- **EBGP** is used for routing between different ASes.

- **IBGP** is used for routing synchronization and coordination within the same AS.



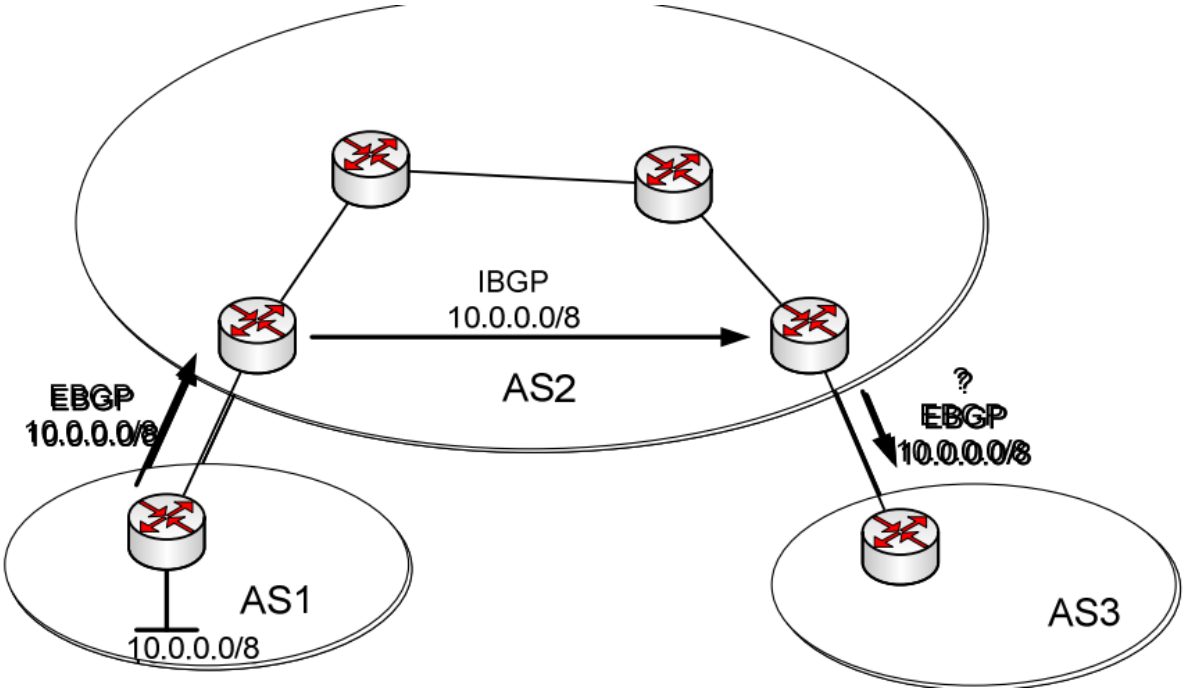
Continuity of BGP within an AS

For IBGP routers to learn all routes within an AS, they need to be fully meshed, meaning each IBGP router must have a direct connection to every other IBGP router in the AS. This full IBGP mesh can be logical rather than physical.



Synchronization within an AS

BGP must be synchronized with the Interior Gateway Protocol (IGP) within an AS before it can advertise routes learned from IBGP to external peers. This synchronization ensures that all internal routes are known to the internal routers, avoiding potential routing issues.



Solutions for Synchronization

1. **Redistribution of all routes into the IGP**: This can cause scalability and performance issues in the IGP.
2. **Internal routers using default routes to an exit router**: This method can lead to suboptimal routing.
3. **Full IBGP mesh within the AS**: All routers have direct IBGP sessions with each other, and synchronization can be disabled.

Example BGP Commands

To check the current BGP routing information, the following commands can be used:

- `show ip bgp`: Displays the current BGP routing table.
- `show ip bgp paths`: Shows the paths for the BGP routes.

sh ip bgp

```
cisco6509#sh ip bgp
BGP table version is 5011434, local router ID is 147.91.0.112
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*  0.0.0.0          195.178.34.57          150             0 8400 8400 i
*>                  195.178.35.17                          0 8400 8400 i
*> 3.0.0.0          195.178.35.17                          0 8400 8400 702 703 80 i
*                   195.178.34.57          150             0 8400 8400 702 703 80 i
*                   195.251.4.44                           0 34771 5408 20965 3356 701 703 80 i
*  4.0.0.0          195.178.34.57          150             0 8400 8400 5400 3356 i
*                   195.178.35.17                          0 8400 8400 5400 3356 i
*>                  195.251.4.44                           0 34771 5408 20965 3356 i
*  4.23.84.0/22     195.178.34.57          150             0 8400 8400 5400 6461 20171 i
*>                  195.178.35.17                          0 8400 8400 5400 6461 20171 i
*                   195.251.4.44                           0 34771 5408 20965 1299 6461 20171 i
*  4.23.112.0/22    195.178.34.57          150             0 8400 8400 5400 174 21889 i
*>                  195.178.35.17                          0 8400 8400 5400 174 21889 i
*                   195.251.4.44                           0 34771 5408 20965 1299 174 21889 i
*  4.23.180.0/24    195.178.34.57          150             0 8400 8400 5400 6128 30576 i
*>                  195.178.35.17                          0 8400 8400 5400 6128 30576 i
*                   195.251.4.44                           0 34771 5408 20965 1299 6128 30576 i
```

show ip bgp summary

```
cisco6509#sh ip bgp summary
BGP router identifier 147.91.0.112, local AS number 13092
BGP table version is 5011825, main routing table version 5011825
174488 network entries using 17623288 bytes of memory
866789 path entries using 41605872 bytes of memory
158814 BGP path attribute entries using 8898456 bytes of memory
86369 BGP AS-PATH entries using 2797716 bytes of memory
484 BGP community entries using 24374 bytes of memory
1 BGP extended community entries using 24 bytes of memory
226170 BGP route-map cache entries using 7237440 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 78187170 total bytes of memory
346919 received paths for inbound soft reconfiguration
BGP activity 485542/310513 prefixes, 3874396/3007065 paths, scan interval 60 secs

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
147.91.0.3      4 13092   24571   25008  5011797    0    0 1d21h           32
147.91.8.77     4  6701   27950   24407  5011765    0    0 1w0d            27
160.99.1.12     4 13303   24378   25004  5011768    0    0 2w2d             1
193.254.1.242   4  8214       0       0        0    0    0 never    Active
194.141.252.13  4  6802       0       0        0    0    0 never    Active
194.149.130.249 4  5379   24297   24365  5011768    0    0 1d05h           5
194.177.210.40  4  5408       0       0        0    0    0 never    Active
195.111.106.254 4  1955   24389   24402  5011792    0    0 2w2d           15
195.178.34.57   4  8400 1115639   24417  5011768    0    0 2d16h      172845
195.178.35.17   4  8400 1084803   24411  5011768    0    0 2w2d       172854
195.251.4.44    4 34771  655799   48809  5011792    0    0 1w1d       174088
```

show ip bgp paths

```
cisco6509#sh ip bgp paths
Address    Hash Refcount Metric Path
0x579E7DD0    0        1      0 8400 8400 1299 3343 2895 2895 2587 i
0x48A29898    0        3      0 8400 8400 1299 8928 31222 i
0x52468890    0        1    150 8400 8400 5400 209 15194 i
0x4611BA50    0        2      0 8400 702 30829 i
0x53F54BC8    0        5      0 8400 702 20485 6767 i
0x581F5450    0        3      0 8400 8400 1299 19962 30444 i
0x581EDE78    0        2      0 8400 8400 1299 3549 26315 i
0x5421C6A0    0        2    150 8400 1299 2828 5725 i
0x53CE44E8    0        2      0 8400 1299 7911 16905 1832 i
0x53CE63D8    0        4    150 8400 8400 5400 7018 16609 i
0x57BF0018    0        1    150 8400 8400 5400 7018 16609 i
0x46A26C70    0        1    150 8400 8400 5400 5511 6505 21862 i
0x568B45B0    0        1    150 8400 8400 5400 174 27429 i
0x48A2CE98    0        3      0 34771 5408 20965 1299 1239 13228 25465
```

Path Attributes in BGP

In BGP, path attributes are essential components in route and path configuration. Each route in BGP has specific attributes which fall into four categories:

1. **Well-known mandatory**: These attributes must be included in every BGP update message associated with a specific Network Layer Reachability Information (NLRI). Their absence results in an error.
2. **Well-known discretionary**: Recognized by all BGP implementations but not required to be included with every NLRI.
3. **Optional transitive**: These attributes may not be recognized by all BGP implementations but are forwarded to other BGP peers if they are not recognized.
4. **Optional nontransitive**: These attributes may not be recognized by all BGP implementations and are not forwarded if not recognized.

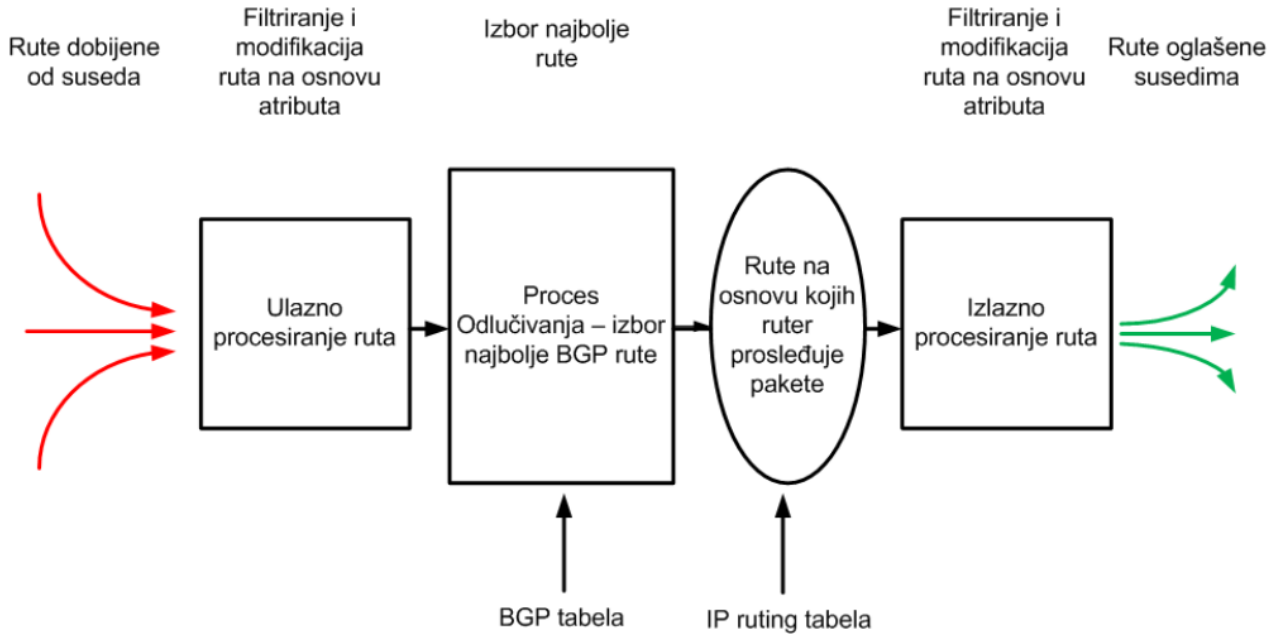
Common Path Attributes

Below are common BGP path attributes and their types:

- **1-ORIGIN**: Well-known mandatory
- **2-NEXT-HOP**: Well-known mandatory
- **3-AS_PATH**: Well-known mandatory
- **4-MULTI_EXIT_DISC**: Optional nontransitive
- **5-LOCAL_PREF**: Well-known discretionary
- **6-ATOMIC_AGGREGATE**: Well-known discretionary
- **7-AGGREGATOR**: Well-known discretionary
- **8-COMMUNITY**: Optional transitive
- **9-ORIGINATOR_ID**: Optional nontransitive
- **10-Cluster list**: Optional nontransitive

BGP Routing Process

When there are multiple BGP routes to a destination, BGP advertises only the best route to its peers. The selection of the best route can be controlled using various BGP attributes such as Next Hop, AS_Path, Atomic Aggregate, Aggregator, Local Preference, Weight, Multiple Exit Discriminator (MED), and Origin.

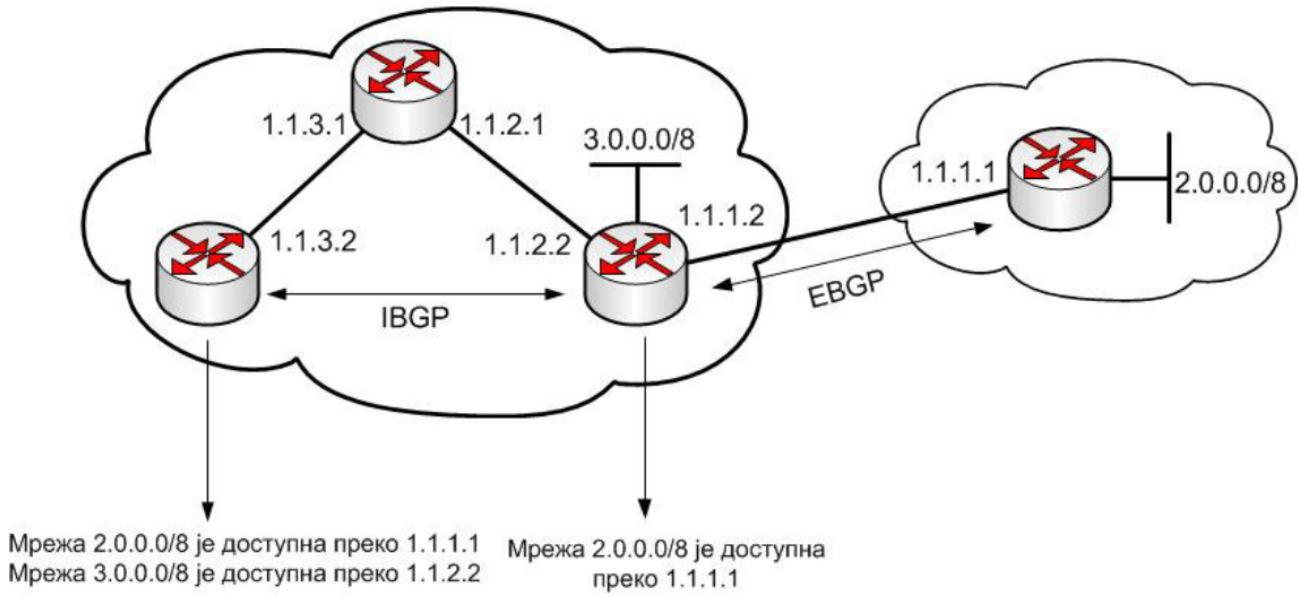


Next Hop Attribute

The Next Hop attribute does not necessarily need to be on the directly connected network segment. The rules for the Next Hop attribute are as follows:

- In EBGP sessions, the Next Hop is the IP address of the EBGP neighbor that advertised the route.
- In IBGP sessions, if routes are advertised within the same AS, the Next Hop is the IP address of the router within the AS that advertised the route.
- In IBGP sessions, if routes are advertised into the AS from another AS via EBGP, the Next Hop received via EBGP is used unchanged in IBGP.

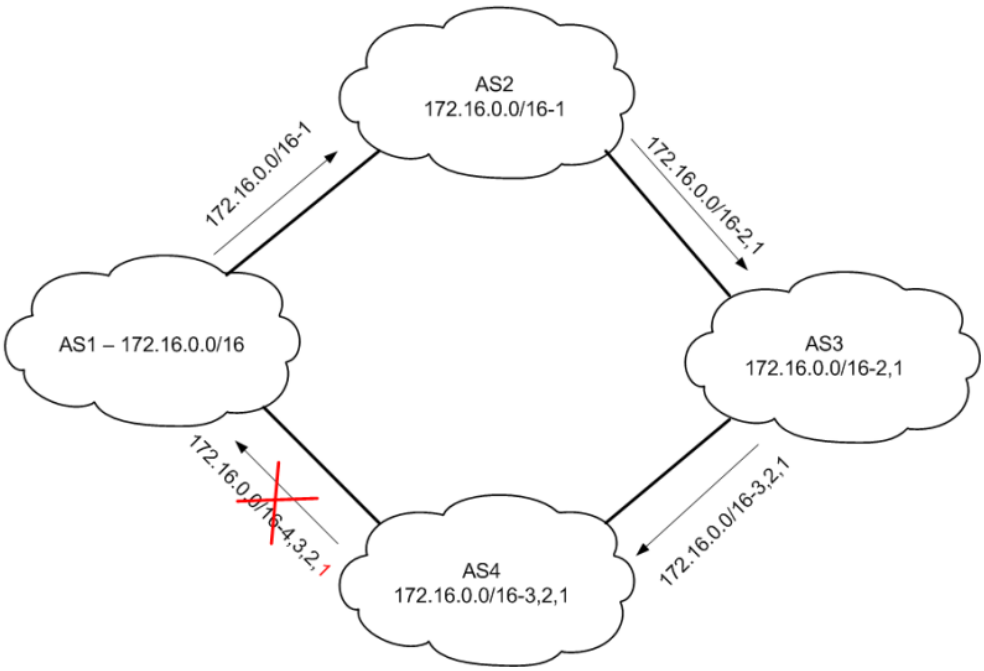
If a router does not have a route to the Next Hop attribute for a given route, that route will not be included in the routing table.



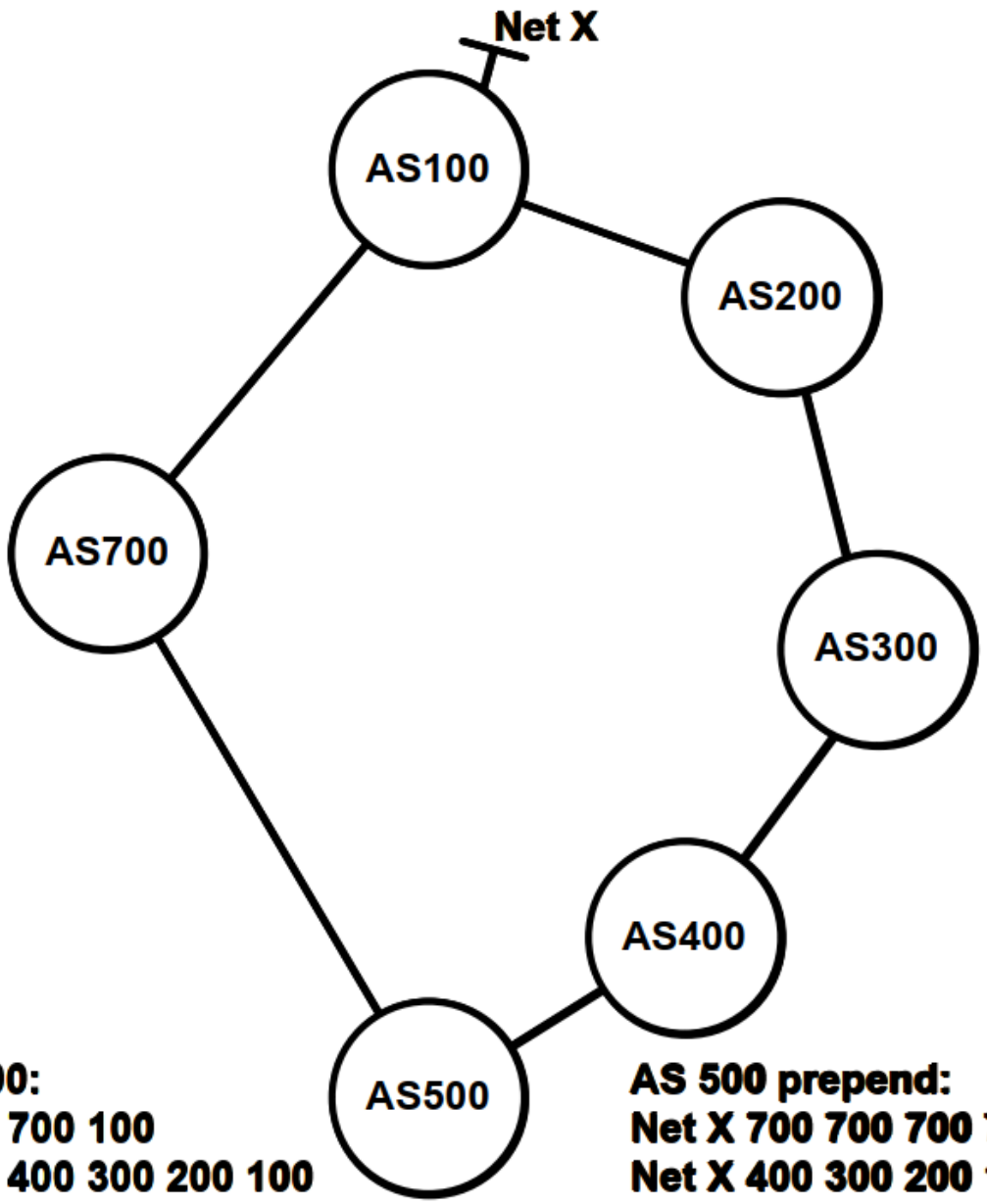
AS Path Attribute

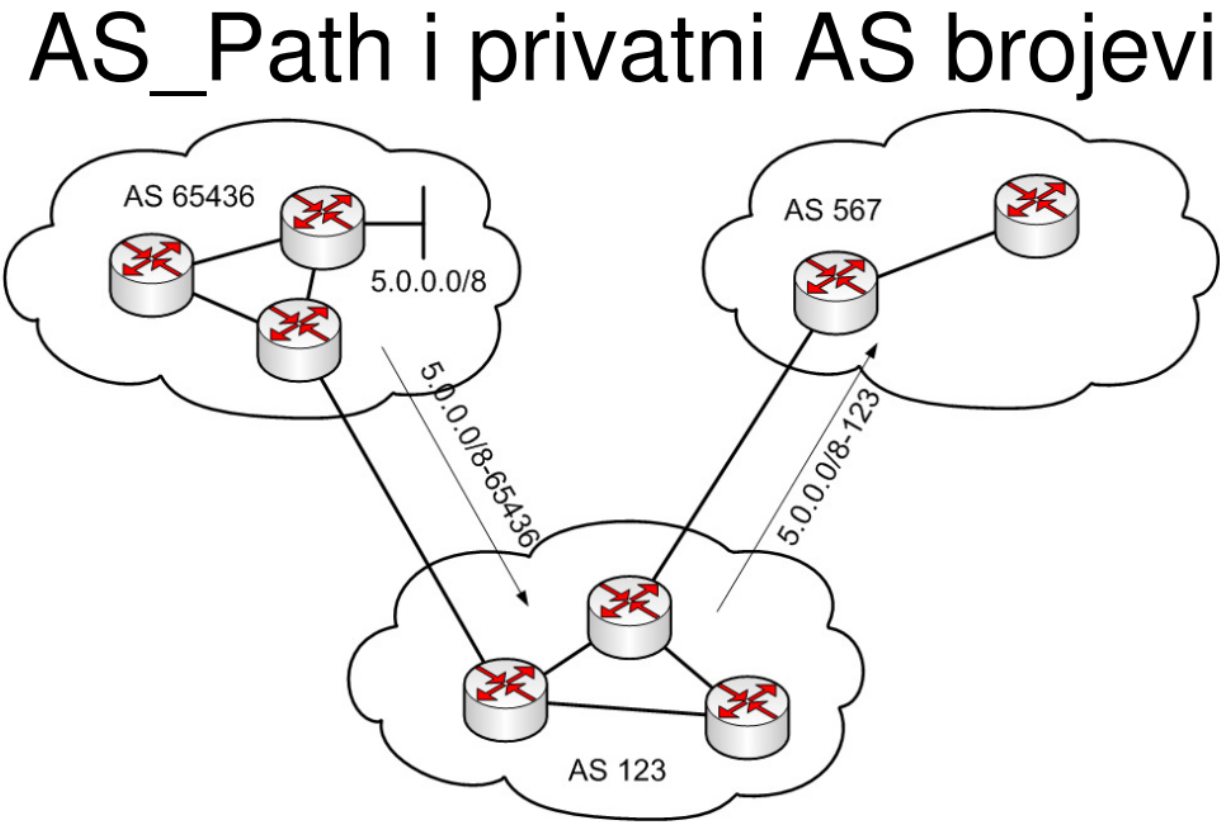
- Each AS that advertises a route adds its AS number to the AS_PATH attribute.
- If a router sees its own AS number in the AS_PATH attribute, it discards the route to prevent loops.
- BGP uses the AS_PATH attribute to help select the best path, with shorter AS paths typically being preferred.

Sometimes AS path prepending is used to artificially lengthen the AS_PATH attribute to influence route selection.



AS Path prepending





Private AS numbers must be taken off the AS Path atribute before the route hits the rest of the internet.

Origin Attribute

The Origin attribute indicates the origin of the route or prefix and is used in route selection. There are three types of Origin attributes:

1. **IGP**: The prefix was learned from the IGP within the AS.
2. **EGP**: The prefix was learned from BGP.
3. **Incomplete**: The prefix was learned through redistribution.

Routes with a lower Origin attribute type are preferred (IGP < EGP < Incomplete).

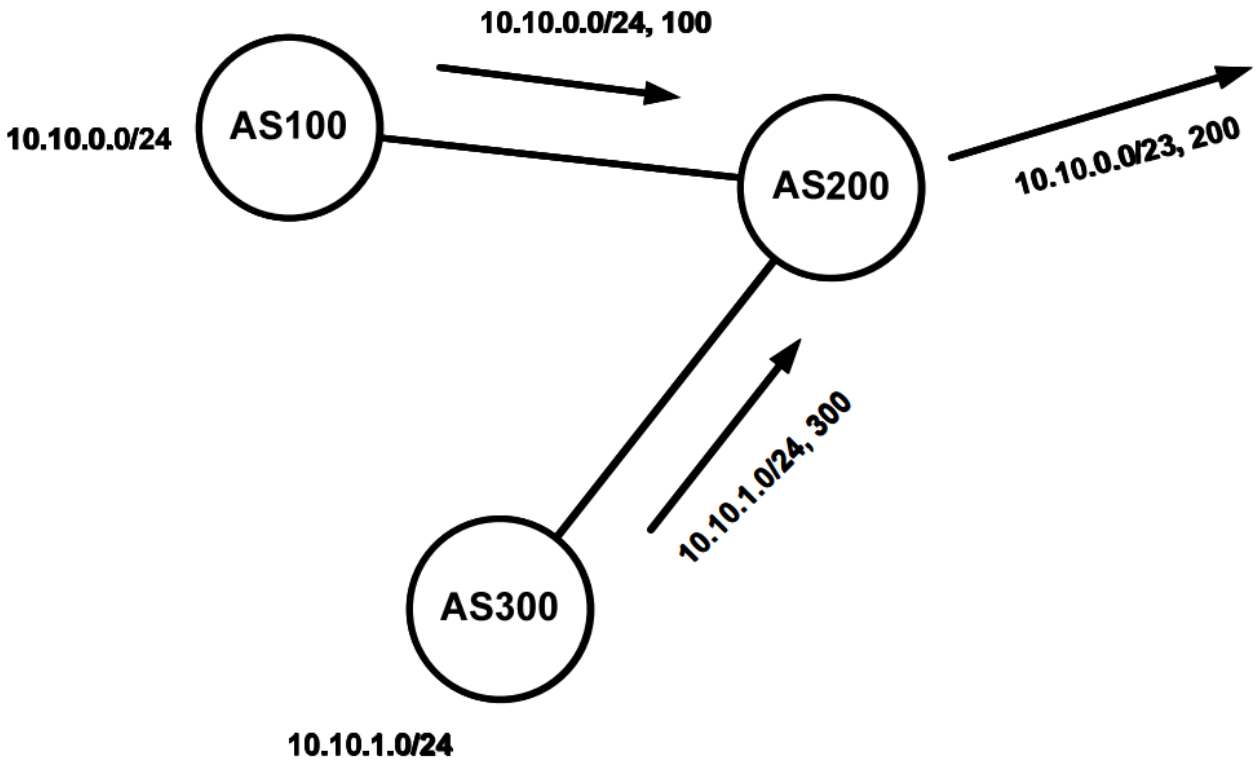
Atomic Aggregate Attribute

The Atomic Aggregate attribute indicates route aggregation and loss of information in the AS_PATH attribute. It can be either True or False:

- **True**: The prefix was aggregated from multiple prefixes.
- **False**: No aggregation was performed.

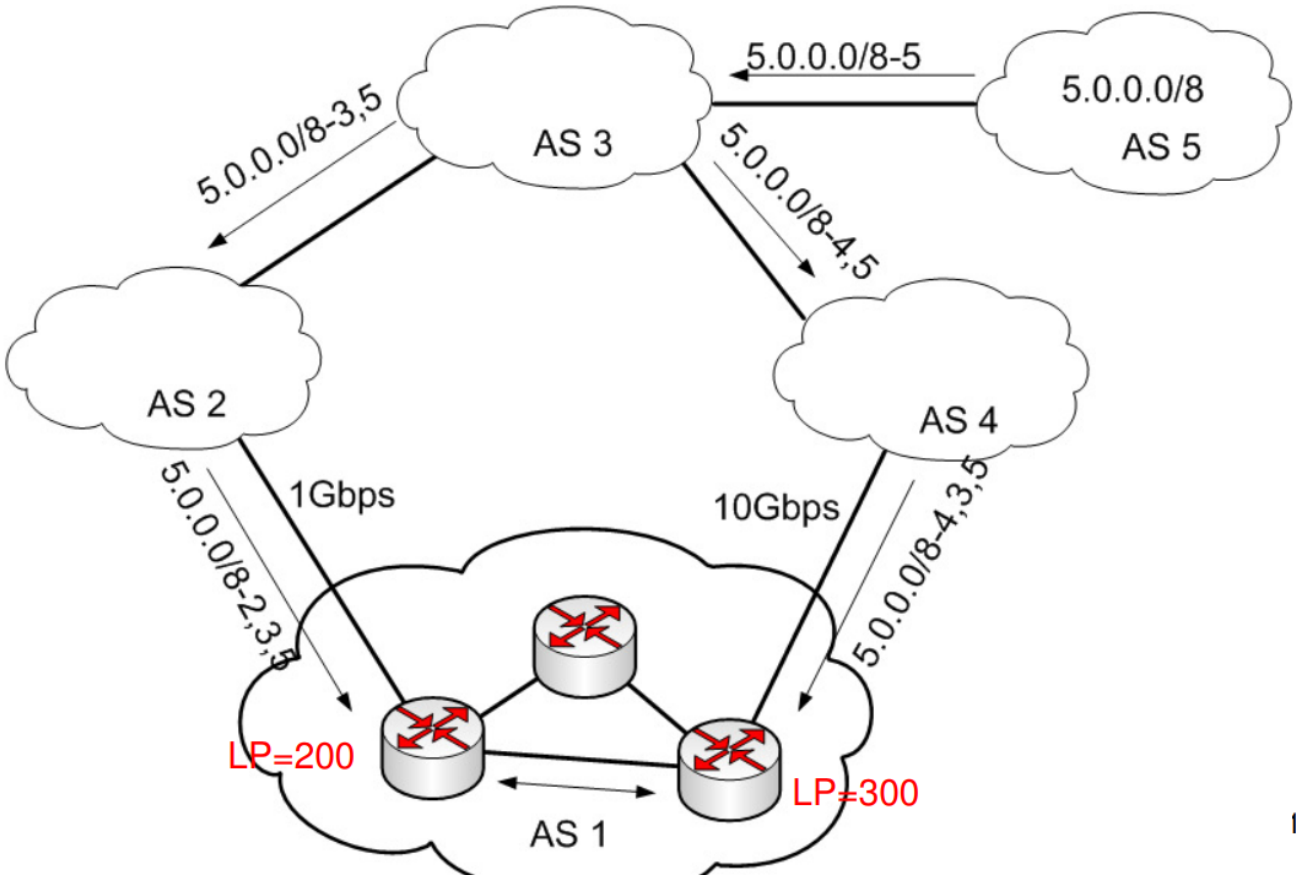
Aggregator Attribute

This attribute identifies the router that performed the aggregation and includes the Router ID of the aggregating router.



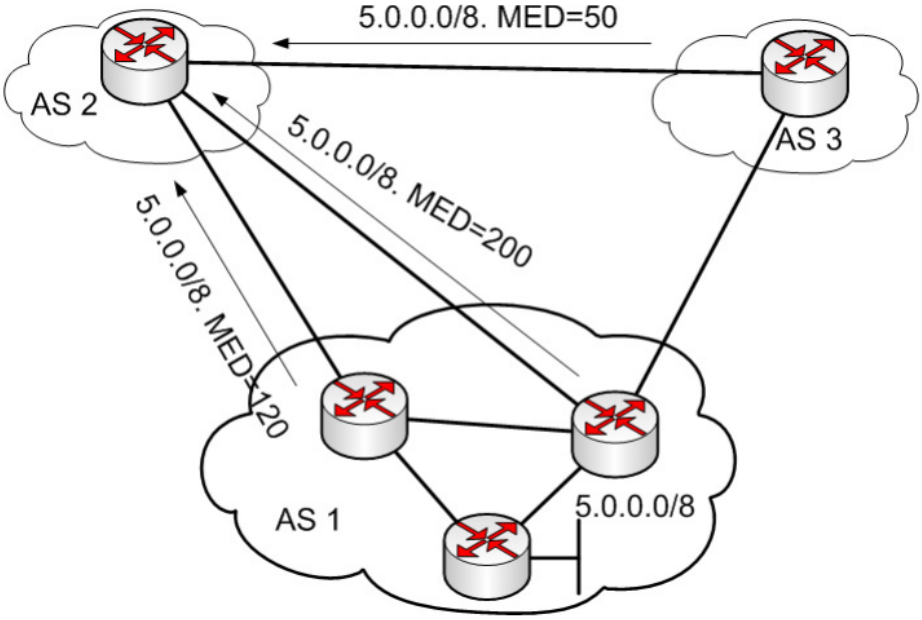
Local Preference Attribute

Local Preference indicates the priority of a route within an AS. Routes with higher Local Preference values are preferred. This attribute is only exchanged within an AS and influences outbound traffic from the AS.



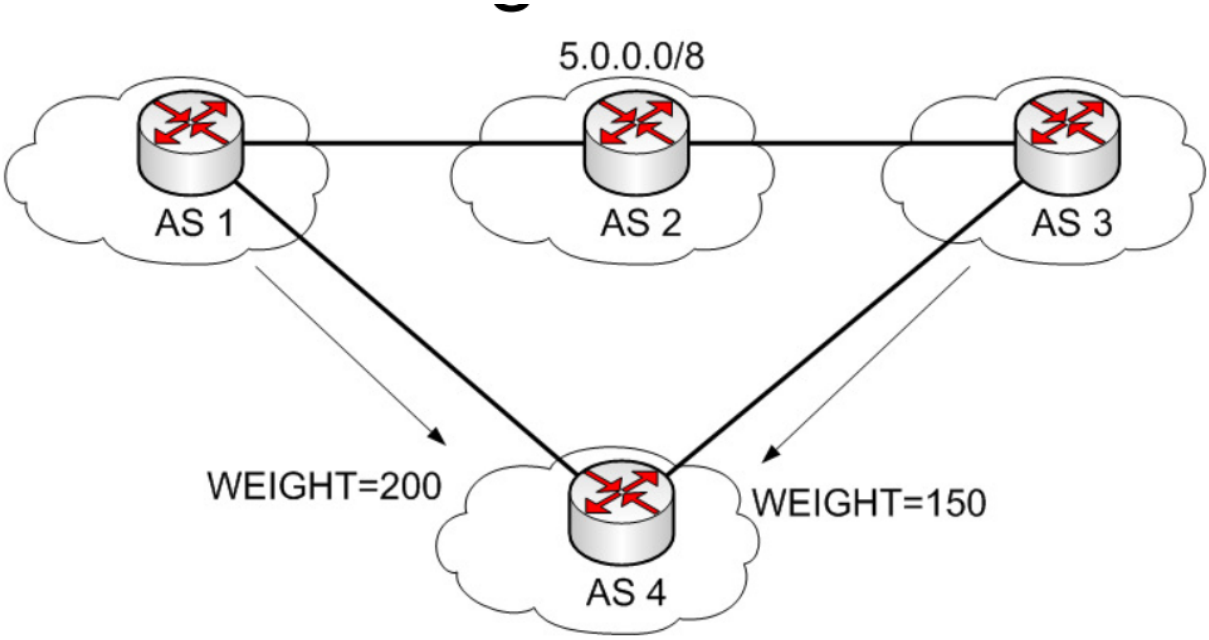
Multiple Exit Discriminator (MED) Attribute

The MED attribute suggests the preferred path into an AS when there are multiple connections with another AS. Lower MED values are preferred, and the attribute is nontransitive, meaning it is not passed beyond the receiving AS.



Weight Attribute

The Weight attribute is Cisco-specific, local to a router, and not exchanged with other routers. It influences the selection of outbound routes within an AS.



Best Path Selection Process in BGP

1. If the Next Hop attribute is not in the routing table, the route is ignored.
2. (If the Weight attribute exists, the route with the highest Weight is selected.)
3. If Weight values are the same, the route with the highest Local Preference is selected.
4. If Local Preference values are the same, the route advertised by the current router is chosen.
5. If previous criteria are the same, the route with the shortest AS_PATH is selected.
6. If AS_PATH lengths are the same, the route with the lower Origin type is chosen.
7. If Origin types are the same, the route with the lower MED is selected.

8. If MED values are the same, routes learned via EBGP are preferred over those learned via IBGP.
9. The route with the lower IGP metric to the BGP Next Hop is preferred.
10. If all previous criteria are the same, the route that was received first is chosen.
11. If still tied, the route from the neighbor with the lower Router ID is preferred.
12. The path with the shorter cluster length is chosen.
13. The route from the neighbor with the lower IP address is preferred.