

# IP Engineering

## Lab 1 - The functioning of BGP



Examining BGP peering Relationships

By:  
Roland Szczesny  
Tom Woo  
Stuart MacLean  
Matt Capranos

## **Executive Summary**

The objective of this lab is to understand how a BGP machine interacts with another peer, this lab begins with the investigation of BGP by looking at its fundamentals. This will provide the opportunity to explore the way that BGP peers talk to each other and the contents of its messages. With careful observation, the different stages in the BGP finite state machine can be identified and studied. As part of the BGP basics, three attributes will be analyzed; these attributes are the Origin, AS\_Path, and Next\_Hop.

The lab will involve the configuration of 10 Autonomous systems, each AS is a reflection of one of the tables in the class, and for example Table #3 will be AS 65003. After the AS # has been assigned to each table the, PC's located there will be used for the Zebra routers to run the BGP and IGP sessions. The last step in the configuration the lab is to assign 4 VLANs.

Prior to the sniffing process and during the final configuration phase of the lab, a routing loop in AS 65003 was discovered, the routing loop which will be discussed further in the lab was causing the following Router-id's 3.3.3.1, 3.3.3.2, 3.3.3.3 to point to the wrong next hope when leaving the AS.

After the configuration of the PC's, Switches were completed and the removal of the Routing loop, the analysis of the BGP session begun, during the sniffing process we were interested in examining the following information.

- Full Network view
- Network Configuration
- Route Information crossing AS
- Description of BGP sniffed dialogue
- Description of Attributes
- BGP synchronization
- Points of Interest

## Full network view

The following is the listing of the routes that are viewable from PC 3 on Table 3.

Destination	Gateway	Genmask	Flags	Metric	Ref	Use	Iface
10.10.10.3	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
5.5.5.1	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
10.10.10.2	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
10.10.10.1	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
5.5.5.3	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
5.5.5.2	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
2.2.2.1	192.168.3.1	255.255.255.255	UGH	0	0	0	eth0
8.8.8.3	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
7.7.7.1	172.16.13.7	255.255.255.255	UGH	0	0	0	eth1
2.2.2.3	172.16.13.2	255.255.255.255	UGH	0	0	0	eth1
8.8.8.1	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
7.7.7.3	172.16.13.7	255.255.255.255	UGH	0	0	0	eth1
2.2.2.2	192.168.3.1	255.255.255.255	UGH	0	0	0	eth0
4.4.4.3	172.16.13.4	255.255.255.255	UGH	0	0	0	eth1
1.1.1.2	172.16.13.1	255.255.255.255	UGH	0	0	0	eth1
4.4.4.2	172.16.13.2	255.255.255.255	UGH	0	0	0	eth1
1.1.1.3	172.16.13.2	255.255.255.255	UGH	0	0	0	eth1
4.4.4.1	172.16.13.2	255.255.255.255	UGH	0	0	0	eth1
1.1.1.1	172.16.13.1	255.255.255.255	UGH	0	0	0	eth1
6.6.6.2	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
9.9.9.1	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
6.6.6.3	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
3.3.3.2	192.168.3.2	255.255.255.255	UGH	20	0	0	eth0
9.9.9.2	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
3.3.3.1	192.168.3.1	255.255.255.255	UGH	20	0	0	eth0
9.9.9.3	172.16.13.5	255.255.255.255	UGH	0	0	0	eth1
192.168.7.0	172.16.13.7	255.255.255.0	UG	0	0	0	eth1
192.168.6.0	172.16.13.5	255.255.255.0	UG	0	0	0	eth1
192.168.5.0	172.16.13.5	255.255.255.0	UG	0	0	0	eth1
192.168.4.0	172.16.13.4	255.255.255.0	UG	0	0	0	eth1
192.168.3.0	*	255.255.255.0	U	0	0	0	eth0
192.168.2.0	172.16.13.2	255.255.255.0	UG	0	0	0	eth1
192.168.1.0	172.16.13.2	255.255.255.0	UG	0	0	0	eth1
172.16.13.0	*	255.255.255.0	U	0	0	0	eth1
172.16.13.5	255.255.255.0		UG	0	0	0	eth1
172.16.13.5	255.255.255.0		UG	0	0	0	eth1
172.16.13.5	255.255.255.0		UG	0	0	0	eth1
*	255.0.0.0		U	0	0	0	lo

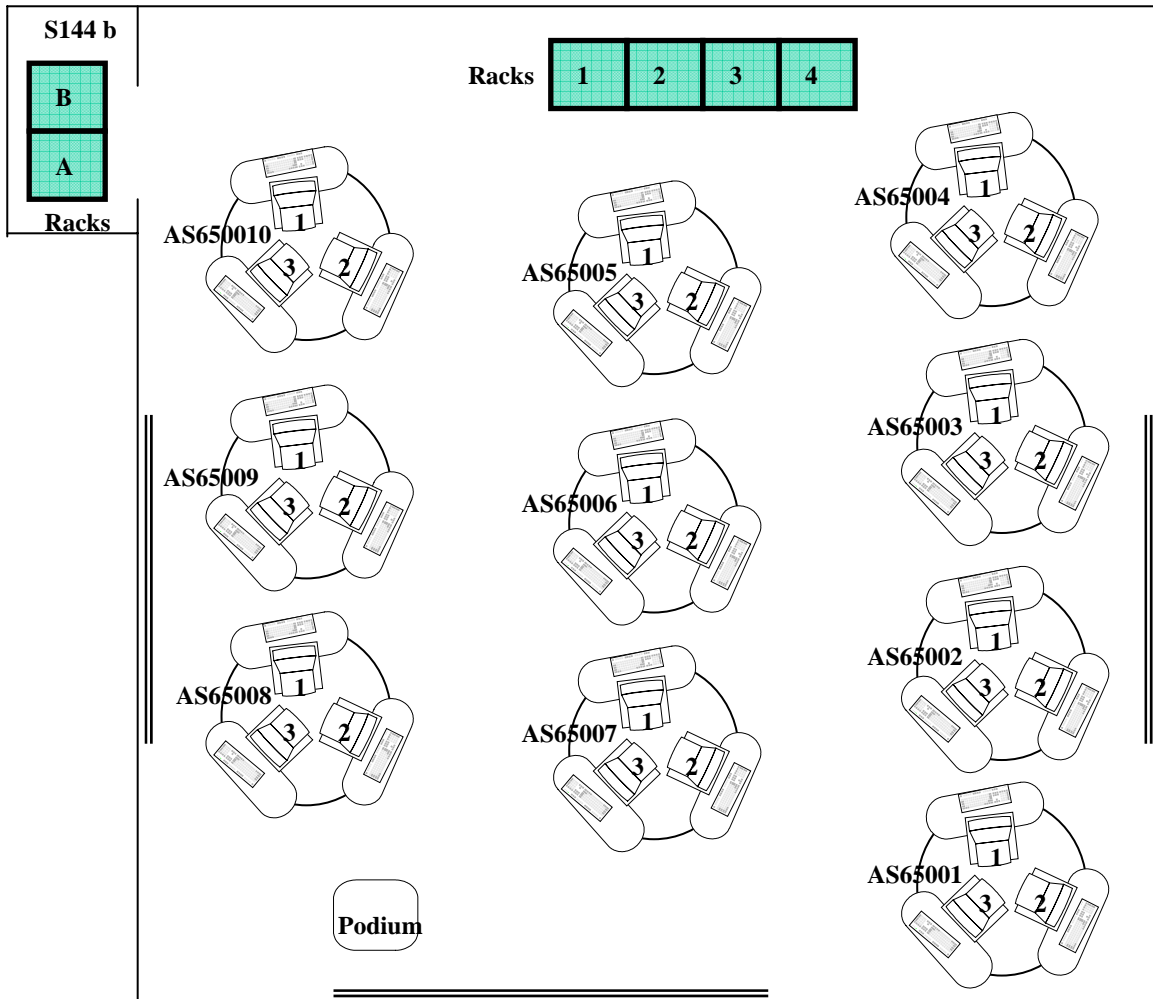
The network view shows on which interfaces routes are accessible and the gateway used to access those routes. In order to access internal routes for our network on table 3 (192.168.3.0) interface eth0 is used. For access to the external routes to AS 65001, 65004-10 interface eth1 on Router 3.3.3.3 is used, additionally in the Route map is the gateway IP address that is used to access the routes. On an interesting note, after the route map was produced, it was noticed that Table 2's routes were being advertised in to Table 3's AS, using Router 3.3.3.1. What is interesting is the fact that the default IP gateway is router 3.3.3.1 eth0 interface. This problem was due to a minor BGP configuration issue and was corrected.

## Network Configuration

In order to get the network operation several configuration steps were required. The steps are as follows:

1. Assignment of Autonomous System number
2. Configuration of VLAN information on the switch
3. Configuration of Zebra (BGP and OSPF) on desktops

In the first step of the lab configuration, was the assigning of AS numbers to each of the tables. In this lab private AS numbers were being utilized, the number was assigned based on table numbers (AS 6500#table). Below is a diagram showing the AS numbers for the class.

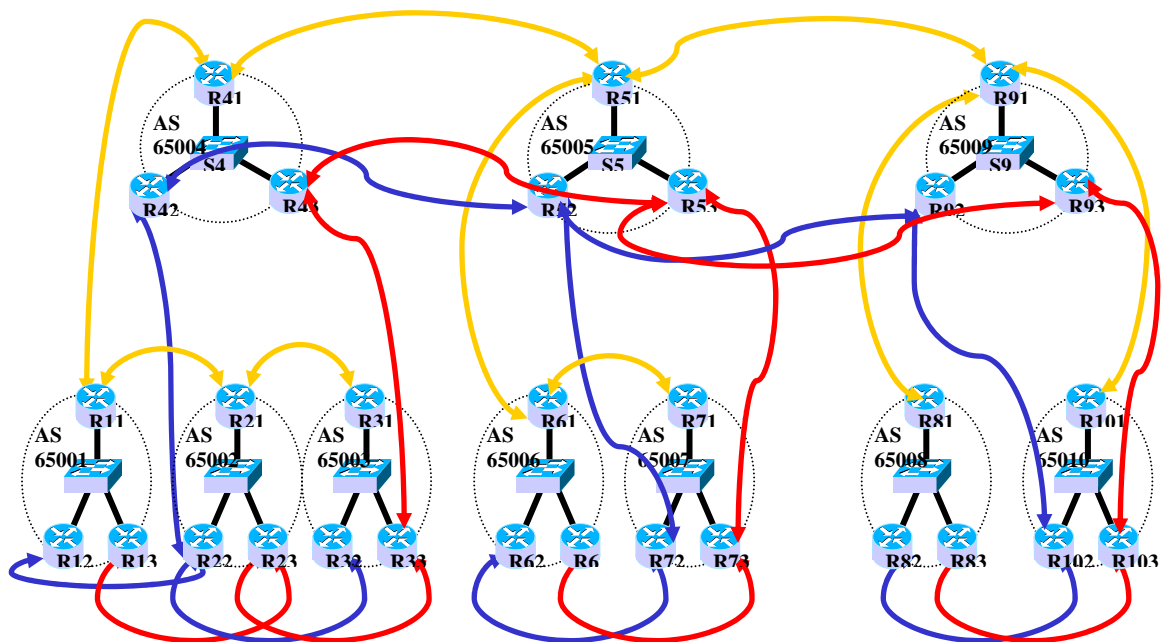


After the AS numbers were assigned to each table, the configuration of the switches and PC's was completed. The switch for AS 65003 required 4 VLANs to be configured, the VLAN assignment and the interfaces that will be assigned to those interfaces are found in the table below.

<p><b><u>VLAN Table 100 + table number</u></b>  <b><i>IP 192.168.T#.0 / 24</i></b></p> <p>Ports 1,2,3,4,5,6</p> <p>Connect:          Desktop # 1, NIC card # 1 to port 1          Desktop # 2, NIC card # 1 to port 2          Desktop # 3, NIC card # 1 to port 3</p>	<p><b><u>VLAN for all machines 1</u></b>          = VLAN 11  <b><i>IP 172.16.11.0/ 24</i></b></p> <p>Ports 7, 8 of each switch</p> <p>Connect: Desktop # 1,          interface # 2 to P8</p>	<p><b><u>VLAN for all machines 2</u></b>          = VLAN 12  <b><i>IP 172.16.12.0/ 24</i></b></p> <p>Ports 9 and 10 of each switch</p> <p>Connect: Desktop # 2,          Interface # 2 to P9</p>	<p><b><u>VLAN for all machines 3</u></b>          = VLAN 13  <b><i>IP 172.16.13.0/ 24</i></b></p> <p>Ports 11 and 12 of each switch</p> <p>Connect: Desktop # 3,          interface # 3 to P11</p>
--	--	--	--

A diagram showing the overall network topology can be found below.

### Whole Internet: Peer relationships



After the configuration of the physical links, we moved on to the configuration of Zebra (BGP and OSPF) on the PCs, below shows the configuration of our BGP and OSPF routers.

OSPF Configuration

```
router ospf
ospf router-id 3.3.3.3
network 3.3.3.3/32 area 0.0.0.0
network 192.168.3.0/24 area 0.0.0.0
```

BGP Configuration

```
router bgp 65003
bgp router-id 3.3.3.3
network 3.3.3.3/32
network 192.168.3.0/24
neighbor 172.16.13.2 remote-as 65002
neighbor 172.16.13.4 in AS 65004 remote-as 65004
neighbor 192.168.3.1 remote-as 65003
neighbor 192.168.3.1 next-hop-self
neighbor 192.168.3.2 remote-as 65003
neighbor 192.168.3.2 next-hop-self
```

The next-hop-self command was used on our BGP routers to force BGP to use a specified IP address in this case, the IP address of the other PC's in our AS as the next hop rather than letting the protocol choose the next\_hop.

## **Route information crossing AS**

Below is a sample of the route information crossing autonomous systems:

Network	Next Hop	Metric	LocPrf	Weight	Path
1.1.1.1/32	172.16.13.1			0	65002 65001 i
2.2.2.3/32	172.16.13.2	0		0	65002 i
3.3.3.3/32	0.0.0.0	0		32768	i
192.168.3.0	0.0.0.0	0		32768	i

From the table above we see the routing information that has come from neighboring AS. An example of routing information crossing AS would be the advertisement for network 1.1.1.1/32, the Next Hop for this network 172.16.13.1, and the route has traversed AS path AS65002 and 65001

## Description of BGP dialog sniffs

In this section we will discuss the BGP dialogue information that was captured during the sniffing process. The first process that was noticed after configuring the BGP sessions and bringing up our BGP router was the SYN, SYN ACK, ACK process between BGP peers.

No. *	Time	Source	Destination	Protocol	Info
23	38.435657	172.16.13.3	172.16.13.4	TCP	36875 > bgp [SYN] Seq=0 Ack=0 Win=5840 Len=0 MSS=1460 TSV=12611
24	38.435841	172.16.13.4	172.16.13.3	TCP	bgp > 36875 [SYN, ACK] Seq=0 Ack=1 Win=5792 Len=0 MSS=1460 TSV=
25	38.435857	172.16.13.3	172.16.13.4	TCP	36875 > bgp [ACK] Seq=1 Ack=1 Win=5840 Len=0 TSV=1261461 TSER=

The above picture shows the SYN, SYN ACK, ACK process between BPG peers 172.16.13.3 and 172.16.3.4. After the process of sending the SYN, SYN ACK, ACK messages to the BGP peer, OPEN MESSAGES are sent by BGP peers 172.16.13.3, 172.16.13.4 in AS 65004 to one another, as shown in the picture below.

No. *	Time	Source	Destination	Protocol	Info
26	38.436008	172.16.13.3	172.16.13.4	BGP	OPEN Message
27	38.436194	172.16.13.4	172.16.13.3	TCP	bgp > 36875 [ACK] Seq=1 Ack=46 Win=5792 Len=0 TSV=274764110 TS
28	38.436676	172.16.13.4	172.16.13.3	BGP	OPEN Message

This process of sending OPEN Message is one of the basic steps of the BGP protocol in establishing sessions between BGP peers, in the OPEN message the “My AS number, Protocol Version, Hold Timer, BGP Identifier and Optional Parameter” information is sent. The picture below shows the contents of the OPEN message being sent from 172.16.13.3 to 172.16.13.4 in AS 65004 and the message being sent from 172.16.13.4 in AS 65004 to 172.16.13.3.

*Message sent from 172.16.13.3 to 172.16.13.4 in AS 65004	*Message sent from 172.16.13.4 in AS 65004 to 172.16.13.3
<ul style="list-style-type: none"><li>Border Gateway Protocol<ul style="list-style-type: none"><li>OPEN Message<ul style="list-style-type: none"><li>Marker: 16 bytes</li><li>Length: 45 bytes</li><li>Type: OPEN Message (1)</li><li>Version: 4</li><li>My AS: 65003</li><li>Hold time: 180</li><li>BGP identifier: 3.3.3.3</li><li>Optional parameters length: 16 bytes</li></ul></li><li>Optional parameters<ul style="list-style-type: none"><li>Capabilities Advertisement (8 bytes)</li></ul></li></ul></li></ul>	<ul style="list-style-type: none"><li>Border Gateway Protocol<ul style="list-style-type: none"><li>OPEN Message<ul style="list-style-type: none"><li>Marker: 16 bytes</li><li>Length: 45 bytes</li><li>Type: OPEN Message (1)</li><li>Version: 4</li><li>My AS: 65004</li><li>Hold time: 180</li><li>BGP identifier: 4.4.4.3</li><li>Optional parameters length: 16 bytes</li></ul></li><li>Optional parameters</li></ul></li></ul>

After the OPEN message has been sent between the BGP peers, KEEPALIVE messages are then sent between the BGP peers until the first UPDATE messages are received. An example of the KEEPALIVE message being sent between the BGP peers is shown in the picture on the next page.

No. *	Time	Source	Destination	Protocol	Info
33	38.476752	172.16.13.4	172.16.13.3	BGP	KEEPALIVE Message
34	38.476763	172.16.13.3	172.16.13.4	BGP	KEEPALIVE Message

Shortly after the KEEPALIVE messages are exchanged between the BGP peers, the first UPDATE message is sent from our BGP router to BGP peer 172.16.13.4 in AS 65004, as shown in the picture below.

No. *	Time	Source	Destination	Protocol	Info
34	38.476763	172.16.13.3	172.16.13.4	BGP	KEEPALIVE Message
36	39.437523	172.16.13.3	172.16.13.4	BGP	UPDATE Message
38	39.437866	172.16.13.4	172.16.13.3	BGP	UPDATE Message
39	39.438935	172.16.13.3	172.16.13.4	BGP	UPDATE Message
40	39.439267	172.16.13.4	172.16.13.3	BGP	UPDATE Message, UPDATE Message, UPDATE Message, UPDATE Message

At almost the same time an UPDATE message is received from the BGP peer 172.16.13.4 in AS 65004. The contents of the UPDATE messages are shown below. The initial update messages exchange only locally attached routes within the AS.

*UPDATE Message to 172.16.13.4 in AS 65004	*UPDATE Message from 172.16.13.4 in AS 65004
<ul style="list-style-type: none"> <li>Internet Protocol, Src: 172.16.13.3 (172.16.13.3), Dst: 172.16.13.4 (172.16.13.4)</li> <li>Transmission Control Protocol, Src Port: 36875 (36875), Dst Port: bgp (179)</li> <li>Border Gateway Protocol <ul style="list-style-type: none"> <li>UPDATE Message <ul style="list-style-type: none"> <li>Marker: 16 bytes</li> <li>Length: 58 bytes</li> <li>Type: UPDATE Message (2)</li> <li>Unfeasible routes length: 0 bytes</li> <li>Total path attribute length: 25 bytes</li> <li>Path attributes <ul style="list-style-type: none"> <li>ORIGIN: INCOMPLETE (4 bytes)</li> <li>AS_PATH: 65003 (7 bytes)</li> <li>NEXT_HOP: 172.16.13.3 (7 bytes)</li> <li>MULTI_EXIT_DISC: 20 (7 bytes)</li> </ul> </li> <li>Network layer reachability information: 10 bytes <ul style="list-style-type: none"> <li>3.3.3.1/32</li> <li>3.3.3.2/32</li> </ul> </li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Internet Protocol, Src: 172.16.13.4 (172.16.13.4), Dst: 172.16.13.3 (172.16.13.3)</li> <li>Transmission Control Protocol, Src Port: bgp (179), Dst Port: 36875 (36875)</li> <li>Border Gateway Protocol <ul style="list-style-type: none"> <li>UPDATE Message <ul style="list-style-type: none"> <li>Marker: 16 bytes</li> <li>Length: 57 bytes</li> <li>Type: UPDATE Message (2)</li> <li>Unfeasible routes length: 0 bytes</li> <li>Total path attribute length: 25 bytes</li> <li>Path attributes <ul style="list-style-type: none"> <li>ORIGIN: IGP (4 bytes)</li> <li>AS_PATH: 65004 (7 bytes)</li> <li>NEXT_HOP: 172.16.13.4 (7 bytes)</li> <li>MULTI_EXIT_DISC: 0 (7 bytes)</li> </ul> </li> <li>Network layer reachability information: 9 bytes <ul style="list-style-type: none"> <li>4.4.4.3/32</li> <li>192.168.4.0/24</li> </ul> </li> </ul> </li> </ul> </li> </ul>

After the initial UPDATE messages being sent from BGP peers 172.16.13.3 and 172.16.13.4 in AS 65004, additional KEEPALIVE messages were sent. After a peering session was established with BGP peer 172.16.13.4 in AS 65004, a peering session with 172.16.13.2 in AS 65002 began, the peering followed the same process as with 172.16.13.4 in AS 65004.

Once the initial peering between AS 65002 and 65004 was complete, additional routes that were known to those AS were flooded to our AS 65003. A series of update messages were received from our peers as shown in the picture below.

No. -	Time	Source	Destination	Protocol	Info
58	41.476601	172.16.13.3	172.16.13.2	BGP	UPDATE Message, UPDATE Message, UPDATE Message, UPDATE Message
60	41.477128	172.16.13.2	172.16.13.3	BGP	UPDATE Message
78	69.437964	172.16.13.3	172.16.13.4	BGP	UPDATE Message
80	69.477627	172.16.13.3	172.16.13.4	BGP	UPDATE Message, UPDATE Message, UPDATE Message, UPDATE Message
83	71.437655	172.16.13.3	172.16.13.2	BGP	UPDATE Message

These update messages include the Network Layer Reachability Information from the other Autonomous systems found in the classroom, below is the information that was received in an update from AS 65004.

\*The following tables are the NLRI that was advertised from AS 65004

No. -	Time	Source	Destination	Prot	Info
					<ul style="list-style-type: none"> <li>Border Gateway Protocol               <ul style="list-style-type: none"> <li>UPDATE Message                   <ul style="list-style-type: none"> <li>Marker: 16 bytes</li> <li>Length: 62 bytes</li> <li>Type: UPDATE Message (2)</li> <li>Unfeasible routes length: 0 bytes</li> <li>Total path attribute length: 20 bytes</li> <li>Path attributes                       <ul style="list-style-type: none"> <li>ORIGIN: IGP (4 bytes)</li> <li>AS_PATH: 65004 65005 (9 bytes)</li> <li>NEXT_HOP: 172.16.13.5 (7 bytes)</li> </ul> </li> <li>Network layer reachability information: 19 bytes                       <ul style="list-style-type: none"> <li>5.5.5.1/32</li> <li>192.168.5.0/24</li> <li>5.5.5.3/32</li> <li>5.5.5.2/32</li> </ul> </li> </ul> </li> </ul> </li> </ul>
					<ul style="list-style-type: none"> <li>Border Gateway Protocol               <ul style="list-style-type: none"> <li>UPDATE Message                   <ul style="list-style-type: none"> <li>Marker: 16 bytes</li> <li>Length: 61 bytes</li> <li>Type: UPDATE Message (2)</li> <li>Unfeasible routes length: 0 bytes</li> <li>Total path attribute length: 24 bytes</li> <li>Path attributes                       <ul style="list-style-type: none"> <li>ORIGIN: IGP (4 bytes)</li> <li>AS_PATH: 65004 65005 65009 65008 (13 bytes)</li> <li>NEXT_HOP: 172.16.13.5 (7 bytes)</li> <li>Network layer reachability information: 14 bytes                       <ul style="list-style-type: none"> <li>8.8.8.1/32</li> <li>192.168.8.0/24</li> <li>8.8.8.3/32</li> </ul> </li> </ul> </li> </ul> </li> </ul> </li></ul>

A similar update is received from AS65002 advertising the NLRI information that is reachable through that AS.

## Description of Attributes

Attributes are used to define routing policies and maintain a stable routing environment. Routes learned using BGP have properties that are used to determine the best route to a destination when multiple paths exist to a particular destination. BGP attributes influence route selection.

The **Origin** attribute shows how BGP learned about a particular route. This attribute is used for route selection and can be one of the following:

**IGP**-The route is interior to the originating AS. This value is set when the **network** router configuration command is used to inject the route into BGP.

**EGP**-The route is learned via the Exterior Border Gateway Protocol (EBGP).

**Incomplete**-The origin of the route is unknown or learned in some other way. An origin of incomplete occurs when a route is redistributed into BGP.

**AS\_path**: When a route advertisement passes through an autonomous system, the AS number is added to a sequential list of AS numbers that the route advertisement has traversed. BGP used this technique to avoid routing loops.

**Next-hop**: The EBGP *next-hop* attribute is the IP address that is used to reach the advertising router. For EBGP peers, the next-hop address is the IP address of the connection between the peers. For IBGP, the EBGP next-hop address is carried into the local AS.

When an EBGP router sends routes learned from another EBGP (from another AS) router, into its own AS, the EBGP next-hop information is preserved. If the routers within the AS do not have routing information regarding the next hop, the route will be discarded. This is why an IGP running in the AS to propagate next-hop routing information is needed.

BGP can receive multiple advertisements for the same route from multiple peers. BGP selects only one path as the best path. When the path is selected, BGP puts the selected path in the IP routing table and sends the path to its neighbors.

Here is how BGP decides which path to select:

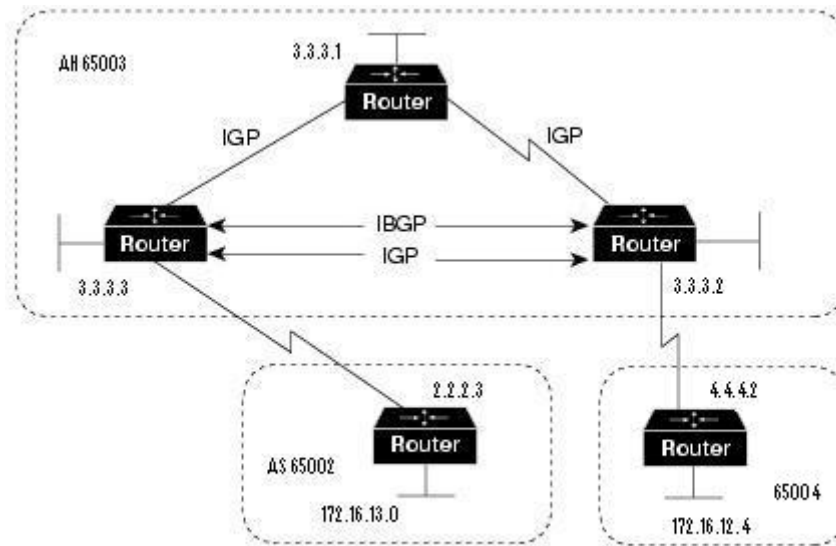
1. If the path specifies a next hop that is inaccessible, drop the update.
2. Prefer the path with the largest weight.
3. If the weights are the same, prefer the path with the largest local preference.
4. If the local preferences are the same, prefer the path that was originated by BGP running on this 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.

## BGP Synchronization

For this lab we were required to have BGP running synchronized. BGP Synchronization is used to maintain internal private full connectivity. When BGP Synchronization is enabled, internal private networks must learn of all routes being re-distributed via IBGP before the routers are able to distribute these networks to their neighbours.

With synchronization enabled, router 3.3.3.3 must send routers 3.3.3.2 and 3.3.3.1 all of the routes which were learned from an external source before passing these routes to its next neighbours.



Router 3.3.3.3 sends updates about network 172.16.13.0 to Router 3.3.3.2. Routers 3.3.3.3 and 3.3.3.2 are running IBGP, so Router 3.3.3.2 receives updates about network 172.16.13.0 via IBGP.

If the link between Router 3.3.3.3 and 3.3.3.2 fails, IGP will still be able to send information between all of the AS's (65002, 65003, 65004). After the link failure if Router 3.3.3.2 wants to reach network 172.16.13.0, it sends traffic to Router 3.3.3.1. If Router 3.3.3.3 does not redistribute network 172.16.13.0 into an IGP, Router 3.3.3.1 has no way of knowing that network 172.16.13.0 exists and will drop the packets.

If Router 3.3.3.2 advertises to AS 65004 that it can reach 172.16.13.0 before Router 3.3.3.1 learns about the network via IGP (assuming the link between Router 3.3.3.3 and Router 3.3.3.2 is down), traffic coming into the network to Router 3.3.3.2 with a destination of 172.16.13.0 will flow to Router 3.3.3.1 and be dropped.

## **Points of Interest**

During the final phase of the lab we observed a point of interest which will be discussed below.

### **Routing Loop**

A routing loop occurred during the test phase of our set up. When any router in AS65003 tried to ping any of the other routers in the AS in the lab, a routing loop would occurred. For example, if router 3.3.3.1 tried to ping any external AS router, the responses would come from router 3.3.3.2 then 3.3.3.3 then 3.3.3.2 etc. This loop would continue until the TTL expired.

We determined that the routing loop occurred because router 3.3.3.2 did not have a dummy interface and the loopback address was set to 3.3.3.2. The next hop self was also not configured in router 3.3.3.2(next hop self concerning 3.3.3.1 and 3.3.3.3). When the dummy interface was set to 3.3.3.2 and the next hop self was configured in router's 3.3.3.2 bgpd config, the routing loop no longer occurred.