BGP Attributes and Path Selection

ISP Training Workshops
BGP Attributes

The “tools” available for the job
What Is an Attribute?

- Part of a BGP Update
- Describes the characteristics of prefix
- Can either be transitive or non-transitive
- Some are mandatory
AS-Path

- Sequence of ASes a route has traversed
- Mandatory transitive attribute
- Used for:
  - Loop detection
  - Applying policy
AS-Path (with 16 and 32-bit ASNs)

- Internet with 16-bit and 32-bit ASNs
  - 32-bit ASNs are 65536 and above
- AS-PATH length maintained

```
180.10.0.0/16   300 23456 23456
170.10.0.0/16   300 23456
170.10.0.0/16   300 23456
180.10.0.0/16   300 23456
170.10.0.0/16   300 23456
150.10.0.0/16   300 400
```

```
AS 70000
180.10.0.0/16
180.10.0.0/16
180.10.0.0/16
```

```
AS 80000
170.10.0.0/16
170.10.0.0/16
```

```
AS 300
```

```
AS 400
150.10.0.0/16
```

```
AS 90000
```

```
AS 300
```

```
AS 400
150.10.0.0/16
```

```
AS 90000
```
AS-Path loop detection

- 180.10.0.0/16 is not accepted by AS100 as the prefix has AS100 in its AS-PATH – this is loop detection in action
Next Hop

- **eBGP** – address of external neighbour
- **iBGP** – NEXT_HOP from eBGP
- Mandatory non-transitive attribute
iBGP Next Hop

- Next hop is ibgp router loopback address
- Recursive route look-up

Diagram:

- AS 300
- Loopback 120.1.254.2/32
- 120.1.1.0/24
- 120.1.2.0/23
- Loopback 120.1.254.3/32
Third Party Next Hop

- eBGP between Router A and Router B
- eBGP between Router B and Router C
- 120.68.1/24 prefix has next hop address of 150.1.1.3 – this is used by Router A instead of 150.1.1.2 as it is on same subnet as Router B
- More efficient
- No extra config needed
Next Hop Best Practice

- Cisco IOS default is for external next-hop to be propagated unchanged to iBGP peers
  - This means that IGP has to carry external next-hops
  - Forgetting means external network is invisible
  - With many eBGP peers, it is unnecessary extra load on IGP

- ISP Best Practice is to change external next-hop to be that of the local router

  `neighbor x.x.x.x next-hop-self`
Next Hop (Summary)

- IGP should carry route to next hops
- Recursive route look-up
- Unlinks BGP from actual physical topology
- Use “next-hop-self” for external next hops
- Allows IGP to make intelligent forwarding decision
Origin

- Conveys the origin of the prefix
- **Historical** attribute
  - Used in transition from EGP to BGP
- Transitive and Mandatory Attribute
- Influences best path selection
- Three values: IGP, EGP, incomplete
  - IGP – generated by BGP network statement
  - EGP – generated by EGP
  - incomplete – redistributed from another routing protocol
Aggregator

- Conveys the IP address of the router or BGP speaker generating the aggregate route
- Optional & transitive attribute
- Useful for debugging purposes
- Does not influence best path selection
- Creating aggregate using “aggregate-address” sets the aggregator attribute:

```
router bgp 100
  aggregate-address 100.1.0.0 255.255.0.0
```
Local Preference

AS 100
160.10.0.0/16

AS 200

AS 300

AS 400

160.10.0.0/16  500
> 160.10.0.0/16  800
Local Preference

- Non-transitive and optional attribute
- Local to an AS only
  - Default local preference is 100 (IOS)
- Used to influence BGP path selection
  - determines best path for *outbound* traffic
- Path with highest local preference wins
Local Preference

- Configuration of Router B:
  
  ```bash
  router bgp 400
  neighbor 120.5.1.1 remote-as 300
  neighbor 120.5.1.1 route-map local-pref in
  
  route-map local-pref permit 10
  match ip address prefix-list MATCH
  set local-preference 800
  
  route-map local-pref permit 20
  
  ip prefix-list MATCH permit 160.10.0.0/16
  ```
Multi-Exit Discriminator (MED)
Multi-Exit Discriminator

- Inter-AS – non-transitive & optional attribute
- Used to convey the relative preference of entry points
  - determines best path for inbound traffic
- Comparable if paths are from same AS
  - `bgp always-compare-med` allows comparisons of MEDs from different ASes
- Path with lowest MED wins
- Absence of MED attribute implies MED value of zero (RFC4271)
**MED & IGP Metric**

- IGP metric can be conveyed as MED
  - *set metric-type internal* in route-map
    - enables BGP to advertise a MED which corresponds to the IGP metric values
    - changes are monitored (and re-advertised if needed) every 600s
  - *bgp dynamic-med-interval* <secs>
Multi-Exit Discriminator

- Configuration of Router B:
  
  ```
  router bgp 400
  
  neighbor 120.5.1.1 remote-as 200
  neighbor 120.5.1.1 route-map set-med out
  
  route-map set-med permit 10
  match ip address prefix-list MATCH
  set metric 1000
  
  route-map set-med permit 20
  
  ip prefix-list MATCH permit 120.68.1.0/24
  ```
Weight

- Not really an attribute – local to router
- Highest weight wins
- Applied to all routes from a neighbour

```bash
neighbor 120.5.7.1 weight 100
```

- Weight assigned to routes based on filter

```bash
neighbor 120.5.7.3 filter-list 3 weight 50
```
Best path to AS4 from AS1 is always via B due to local-pref

But packets arriving at A from AS4 over the direct C to A link will pass the RPF check as that path has a priority due to the weight being set

- If weight was not set, best path back to AS4 would be via B, and the RPF check would fail
Best path to AS4 from AS1 is always via B due to local-pref.

But customers connected directly to Router A use the link to AS7 as best outbound path because of the high weight applied to routes heard from AS7.

- If the A to D link goes down, then the Router A customers see best path via Router B and AS4.
Community

- Communities are described in RFC1997
  - Transitive and Optional Attribute

- 32 bit integer
  - Represented as two 16 bit integers (RFC1998)
  - Common format is <local-ASN>:xx
  - 0:0 to 0:65535 and 65535:0 to 65535:65535 are reserved

- Used to group destinations
  - Each destination could be member of multiple communities

- Very useful in applying policies within and between ASes
Community Example (before)

- Peer AS1
  - 100.10.0.0/16
  - permit 100.10.0.0/16 in

- AS 100
  - 160.10.0.0/16
  - permit 160.10.0.0/16 in

- AS 200
  - 170.10.0.0/16
  - permit 170.10.0.0/16 in

- AS 300
  - ISP 1
  - permit 160.10.0.0/16 out
  - permit 170.10.0.0/16 out

- Upstream AS 400
  - permit 160.10.0.0/16 out
  - permit 170.10.0.0/16 out
Community Example
(after)

Peer AS1
100.10.0.0/16
100.10.0.0/16 300:9

ISP 1

Upstream AS 400

AS 300

160.10.0.0/16 300:1
170.10.0.0/16 300:1

AS 100
160.10.0.0/16

AS 200
170.10.0.0/16
Well-Known Communities

- Several well known communities
  - [www.iana.org/assignments/bgp-well-known-communities](http://www.iana.org/assignments/bgp-well-known-communities)
- no-export 65535:65281
  - do not advertise to any eBGP peers
- no-advertise 65535:65282
  - do not advertise to any BGP peer
- no-export-subconfed 65535:65283
  - do not advertise outside local AS (only used with confederations)
- no-peer 65535:65284
  - do not advertise to bi-lateral peers (RFC3765)
No-Export Community

- AS100 announces aggregate and subprefixes
  - Intention is to improve loadsharing by leaking subprefixes
- Subprefixes marked with no-export community
- Router G in AS200 does not announce prefixes with no-export community set
No-Peer Community

Sub-prefixes marked with **no-peer** community are not sent to bi-lateral peers
- They are only sent to upstream providers
What about 4-byte ASNs?

- Communities are widely used for encoding ISP routing policy
  - 32 bit attribute
- RFC1998 format is now “standard” practice
  - ASN:number
- Fine for 2-byte ASNs, but 4-byte ASNs cannot be encoded

Solutions:
- Use “private ASN” for the first 16 bits
## Summary

### Attributes in Action

Router6>sh ip bgp
BGP table version is 30, local router ID is 10.0.15.246

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal, r RIB-failure, S Stale

Origin codes: i - IGP, e - EGP, ? - incomplete

<table>
<thead>
<tr>
<th>Network</th>
<th>Next Hop</th>
<th>Metric</th>
<th>LocPrf</th>
<th>Weight</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>*i10.0.0.0/26</td>
<td>10.0.15.241</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>i</td>
</tr>
<tr>
<td>*i10.0.0.64/26</td>
<td>10.0.15.242</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>i</td>
</tr>
<tr>
<td>*i10.0.0.128/26</td>
<td>10.0.15.243</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>i</td>
</tr>
<tr>
<td>*i10.0.0.192/26</td>
<td>10.0.15.244</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>i</td>
</tr>
<tr>
<td>*i10.0.1.0/26</td>
<td>10.0.15.245</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>i</td>
</tr>
<tr>
<td>* 10.0.1.64/26</td>
<td>0.0.0.0</td>
<td>0</td>
<td></td>
<td>32768</td>
<td>i</td>
</tr>
</tbody>
</table>

...
BGP Path Selection Algorithm

Why is this the best path?
BGP Path Selection Algorithm for Cisco IOS: Part One

1. Do not consider path if no route to next hop
2. Do not consider iBGP path if not synchronised (Cisco IOS)
3. Highest weight (local to router)
4. Highest local preference (global within AS)
5. Prefer locally originated route
6. Shortest AS path
BGP Path Selection Algorithm for Cisco IOS: Part Two

7. Lowest origin code
   - IGP < EGP < incomplete

8. Lowest Multi-Exit Discriminator (MED)
   - If `bgp deterministic-med`, order the paths by AS number before comparing
   - If `bgp always-compare-med`, then compare for all paths
   - Otherwise MED only considered if paths are from the same AS (default)
BGP Path Selection Algorithm for Cisco IOS: Part Three

9. Prefer eBGP path over iBGP path
10. Path with lowest IGP metric to next-hop
11. For eBGP paths:
   - If multipath is enabled, install N parallel paths in forwarding table
   - If router-id is the same, go to next step
   - If router-id is not the same, select the oldest path
12. Lowest router-id (originator-id for reflected routes)

13. Shortest cluster-list
   - Client must be aware of Route Reflector attributes!

14. Lowest neighbour address
BGP Attributes and Path Selection

ISP Training Workshops