

BGP communities for service providers

Marco d'Itri

<md@seeweb.it>

@rfc1036

Seeweb s.r.l.

Wholesale Winery Tour - May 6, 2022



Attributes of BGP routes

BGP attributes are metadata attached to the prefixes.

Well-known mandatory attributes:

- as-path
- next-hop
- origin

Some optional attributes:

- Community: a 32 bit (or more...) integer, can have multiple values.
- Local preference: a 32 bit integer.
- Multi-exit discriminator (MED): a 32 bit integer.

How routers choose the best path

In this order:

- Highest local preference.
- Shortest as-path.
- Lowest origin¹.
- Lowest MED.

(And some other parameters which are less important and may be implementation-dependent.)

¹ It should be rewritten as egp on ingress anyway!

MED and local preference

Multi-exit discriminator

If received over eBGP then it will not be sent to other eBGP neighbors.
Often accepted only from customers.
May be reset or changed on ingress for traffic engineering purposes.

Local preference

Never sent to eBGP neighbors.
Used to create your local routing policy.

Types of communities

Regular: 32 bit, represented as 16+16 bit.

Extended: 64 bit.

Large: 96 bit, represented as 32+32+32 bit.

Nobody cares about extended communities, so they will not be discussed here.

Representation of communities

Regular: 16+16 bit

- 12637:1234
- 0:12637
- 1:2

Large: 32+32+32 bit

- 43369:1:250000
- 1:2:3
- 250000:1:2

Usually administrators encode their own ASN in the first pseudo-field for communities intended to be propagated to other ASes.

32 bit ASNs can be represented only in large communities!

How are communities used?

- Received from neighbors (or your other routers).
 - Advertised to neighbors (or your other routers).
-
- Provide information about the route (e.g. geographical origin).
 - Directly influence the neighbor routing policy (e.g. announce selectively).

Acting on communities is a social agreement between peers!

Well-known communities

The semantics of some communities are defined by RFCs. Some examples:

- `no-export (65535:65281)`: do not announce to other ASes.
- `graceful_shutdown (65535:0)`: prefer any other path.
- `blackhole (65535:666)`: drop traffic.

Informational communities (inbound)

Route received, e.g.:

- from a customer.
- from a peer.
- in a specific city/country/region.

This kind of community can be evaluated in your own routing policy, e.g. to set an appropriate local preference.

Traffic engineering communities (outbound)

Generally, transit providers provide TE communities to their customers but peers do not.

Example actions:

- Prepend N times.
- Do not announce.
- Change the local preference.
- Blackhole all traffic to the prefix.

Actions can be limited, e.g.:

- To a specific transit provider or peer of your transit provider.
- To a specific IXP.
- In a specific country or region.

By IXPs route servers

IXPs often implement these communities in their route servers to allow customers to control how the RSeS should readvertise the routes.

If the RS ASN is 43369, then:

- `0:PEERASN` or `43369:0:PEERASN`: do not advertise to PEERASN.
- `43369:PEERASN` or `43369:1:PEERASN`: advertise to PEERASN.
- `0:43369` or `43369:0:0`: do not advertise to any peer.

The outbound filtering design pattern

To prevent route leaks, only own and customer routes must be announced to peers and transits.

Stupid solution

- Maintain prefix lists containing all customer routes on your edge routers.

Correct solution

- Tag routes when they are generated or enter your network.
- Only re-announce routes tagged as your own or customers'.

Also: please filter your customers' announces with a prefix-lists!

Any questions?



`https://www.linux.it/~md/text/communities-ftwwt2022.pdf`
(Google ... Marco d'Itri ... I feel lucky)

