BGP Route Servers<draft>

This document is inspired from https://www.inx.net.za/display/pub/BGP+Route+Servers

There are two BGP separate route servers on each peering LAN. It is recommended to **always** peer with both BGP Route Servers at a location, as sessions to both servers ensure that there is no disruption to the advertisement of your prefixes should it be necessary to performance maintenance on a Route Server. The Route Servers do **not** peer with each other by design, so peering with only one server is an unnecessary risk for your network!

Bi-lateral peering is considered best practice !

While the BGP Route Server service is made available as a convenience, it is strongly recommended that, in addition to any sessions you plan to establish with the BGP Route Servers, you still maintain direct bi-lateral peering sessions with peers that you feel are important to your network! BGP Route Servers should be used to pickup quick/easy/additional peers only, and **not** as a replacement for your discrete peering policy!

In particular there are many peers that advertise only a subset of their prefixes to the BGP Route Server. **Always** aim for a bilateral session !

IXP	ASN	Hostname	Type	IPv4	IPv6
MIXP	37324	rs1.mixp.org	<to document>	196.223.0.201	2001:43f8:270: d0::201
		rs2.mixp.org	<to document>	196.223.0.202	2001:43f8:270: d0::202

First ASN Check

Remember that the BGP-RS service at all the MIXP do not include the BGP-RS ASN in BGP update messages, as the RS is not actually a transit network. Ensure that if you do plan on peering with the BGP Route Servers, you understand that the BGP-RS does not attach its ASN to outbound BGP messages.

Please implement the IOS "`no bgp enforce-next-as`" (or IOS-XR "`enforce-first-as disable`"), or appropriate equivalent, for your platform.

Filtering policy and process

MIXP has **always** believed in filtering and we filter all client sessions to the BGP-RS service. We encourage peers to keep their IRR objects accurate to help us to autogenerate these filters.

- Filters are built based on RPKI and IRRDB registered objects.
- We search the AfriNIC, RADB and RIPE registries (in that order).
- We permit only **exact match filters** for both IPv4 and IPv6.
- RPKI invalids are dropped.
- Some prefixes are automatically filtered by the route servers (eg. bogons and martians).
- We do not accept BGP announcements from private ASNs, or with private ASNs in the path.

MIXP's Route Server filtering policy is below:

- Drop small prefixes – longer than /24 for ipv4 and longer than /48 for ipv6.
- Drop all well-known martians and bogons.
- Ensure that there is at least 1 ASN and less than 64 ASNs in the AS path.
- Ensure that the peer AS is the same as the first AS in the AS path.
- Drop any prefix where the next-hop IP address is not the same as the peer IP address. This prevents prefix hijacking.
- Drop any prefix with a transit network ASN in the AS path.
- Ensure that origin AS is in set of ASNs from the client’s IRRDB AS-SET.
- If the prefix is evaluated as RPKI valid, accept.
- If the prefix is evaluated as RPKI invalid, drop.
- If the prefix is evaluated as RPKI unknown, revert to standard IRRDB prefix filtering.

Max-prefix

We recommend that you set the BGP max-prefix to the BGP-RS to 150,000 prefixes for IPv4 and 100,000 prefixes for IPv6

Filtering Frequency

Filter generation happens every 4h starting at 0h45. If you need a filter update done in an emergency, email us via noc@mixp.org, peering@mixp.org

Should you require help and a personal assistance, we can schedule a online conference session via Zoom, Jitsi or IM

BGP Communities for policy control

A simple set of BGP communities are made available for rudimentary policy control. These will be expanded on over time, as the BGP Route Server service is enhanced. We provide both extended and large community (RFC 8092) support. Note that if you intend to effect policy to 32bit ASNs you'll need to make use of the BGP-LC communities. As a general rule, you should implement large community (LC) filtering if your device supports this. Do not mix both types!

Remember to use the correct ASN

Note: The communities example below applies to peers using the MIXP route servers. The appropriate ASN for each MIXP, should be substituted when using the BGP route servers.

0:peer-asn	deny to peer-asn	block announcement of prefix to pe as
0:37324	block all	block announcement of prefix to al peers
37324:peer-asn	allow to peer-asn	announce prefix to specific peer-as conjunction with block all)
37324:37324	allow all	announce prefix to all peers (implic default)

We honour the well-known no-export and no-advertise communities as if they were sent to us as a regular peer. If you would specifically like us to propagate these, then please tag as below:

37324:65281	add no-export	adds the well known no-export community to all routes sent to pe
37324:65282	add no-advertise	adds the well known no-advertise community to all routes sent to pe

BGP Large Community Support for policy control

37324:0:peer-asn	deny to peer-asn	block announcement of prefix to pe asn
37324:0:0	block all	block announcement of prefix to al peers
37324:1:peer-asn	allow to peer-as	announce prefix to specific peer-as conjunction with block all)
37324:1:0	allow all	announce prefix to all peers (implic default)

We also support path prepending using the following policy:

37324:101:peer-asn	Prepend to peer AS once
37324:102:peer-asn	Prepend to peer AS twice
37324:103:peer-asn	Prepend to peer AS three times

Communities returned for filtered routes

If your prefix is filtered by the BGP-RS, we'll return one of the BGP communities below, that should help aid in the debugging process.

Filtered community List

```
PREFIX_LEN_TOO_LONG      = ( routeserverasn, 1101, 1  )
PREFIX_LEN_TOO_SHORT     = ( routeserverasn, 1101, 2  )
BOGON                    = ( routeserverasn, 1101, 3  )
BOGON_ASN                = ( routeserverasn, 1101, 4  )
AS_PATH_TOO_LONG         = ( routeserverasn, 1101, 5  )
AS_PATH_TOO_SHORT        = ( routeserverasn, 1101, 6  )
FIRST_AS_NOT_PEER_AS     = ( routeserverasn, 1101, 7  )
NEXT_HOP_NOT_PEER_IP     = ( routeserverasn, 1101, 8  )
IRRDB_PREFIX_FILTERED    = ( routeserverasn, 1101, 9  )
IRRDB_ORIGIN_AS_FILTERED = ( routeserverasn, 1101, 10 )
PREFIX_NOT_IN_ORIGIN_AS  = ( routeserverasn, 1101, 11 )
RPKI_UNKNOWN             = ( routeserverasn, 1101, 12 )
RPKI_INVALID             = ( routeserverasn, 1101, 13 )
TRANSIT_FREE_ASN         = ( routeserverasn, 1101, 14 )
TOO_MANY_COMMUNITIES     = ( routeserverasn, 1101, 15 )
```

Prefixes auto-filtered by the Route Servers

For the overall safety and security of our participants, we actively filter the following prefixes at the Route Servers. That is, advertisements from peers, containing the following networks, will be dropped, and not onward announced.

IPv4 prefixes filtered by the BGP-RS (RFC6890)

```
martians = [
          10.0.0.0/8+,
          100.64.0.0/10+,
          169.254.0.0/16+,
          172.16.0.0/12+,
          192.0.0.0/24+,
          192.0.2.0/24+,
          192.168.0.0/16+,
          198.18.0.0/15+,
          198.51.100.0/24+,
          203.0.113.0/24+,
          224.0.0.0/4+,
          240.0.0.0/4+,
          0.0.0.0/32-,
          0.0.0.0/0{25,32},
          0.0.0.0/0{0,7}
];
```

IPv6 prefixes filtered by the BGP-RS

```
martians = [
          ::/0,                    # Default (can be advertised as a route in BGP to peers if desired)
          ::/96,                   # IPv4-compatible IPv6 address - deprecated by RFC4291
          ::/128,                  # Unspecified address
          ::1/128,                 # Local host loopback address
          ::ffff:0.0.0.0/96+,      # IPv4-mapped addresses
          ::224.0.0.0/100+,        # Compatible address (IPv4 format)
          ::127.0.0.0/104+,        # Compatible address (IPv4 format)
          ::0.0.0.0/104+,          # Compatible address (IPv4 format)
          ::255.0.0.0/104+,        # Compatible address (IPv4 format)
          0000::/8+,               # Pool used for unspecified, loopback and embedded IPv4 addresses
          0200::/7+,               # OSI NSAP-mapped prefix set (RFC4548) - deprecated by RFC4048
          3ffe::/16+,              # Former 6bone, now decommissioned
          2001:db8::/32+,          # Reserved by IANA for special purposes and documentation
          2002:e000::/20+,         # Invalid 6to4 packets (IPv4 multicast)
          2002:7f00::/24+,         # Invalid 6to4 packets (IPv4 loopback)
          2002:0000::/24+,         # Invalid 6to4 packets (IPv4 default)
          2002:ff00::/24+,         # Invalid 6to4 packets
          2002:0a00::/24+,         # Invalid 6to4 packets (IPv4 private 10.0.0.0/8 network)
          2002:ac10::/28+,         # Invalid 6to4 packets (IPv4 private 172.16.0.0/12 network)
          2002:c0a8::/32+,         # Invalid 6to4 packets (IPv4 private 192.168.0.0/16 network)
          fc00::/7+,               # Unicast Unique Local Addresses (ULA) - RFC 4193
          fe80::/10+,              # Link-local Unicast
          fec0::/10+,              # Site-local Unicast - deprecated by RFC 3879 (replaced by ULA)
          ff00::/8+,               # Multicast
          ::/0{49,128}             # Filter small prefixes
];
```

ASNs filtered by the Route Servers (Tier-1 networks/peer-locking)

We filter a regular set of networks that are known to be transit-free (ie. we do not expect a peer to send us a prefix with one of these ASNs in the path).

IPv6 prefixes filtered by the BGP-RS

```
TRANSIT_FREE_ASNS = [ 174, 701, 1299, 2914, 3257, 3320, 3356, 3491, 5511, 6453, 6461, 6762, 6830, 7018 ];
```

Filtering of the Route Servers (ingress to a peer)

AS-Path Stripping

The BGP route servers do not add their own ASN in the advertised path, so if you're planning on constructing a filter list to filter the BGP Route servers, do not use the BGP route servers ASN in the path!

We publish IRR record to show networks that peer with the BGP-RS service. Peers are are so inclined, may use this to create their own filters that they can then elect to use to filter the BGP-RS in question. These are available as AS-SETs for:

- MIXP: AFRINIC::AS-MIXP-RS

This is also published at PeeringDB

FAQ

1. BGP peering with route server is up but I did not receive any prefixes and I receive weird messages/errors

%Log packet overrun, PC 0xAAB1DCE504, format:

unsupported or mal-formatted message received from %s:

%s

*Oct 4 04:52:29.830: %BGP-5-NBR_RESET: Neighbor 196.223.0.202 active reset (Peer closed the session)

*Oct 4 04:52:29.831: %BGP_SESSION-5-ADJCHANGE: neighbor 196.223.0.202 IPv4 Unicast topology base removed from session Peer closed the session

*Oct 4 04:52:31.880: %BGP-5-ADJCHANGE: neighbor 196.223.0.201 Up

*Oct 4 04:52:31.886: %BGP-6-MSGDUMP_LIMIT: unsupported or mal-formatted message received from 196.223.0.201:

FFFF FFFF FFFF FFFF FFFF FFFF FFFF FFFF 025A 0200 0000 1440 0101 0040 0206 0201

0000 7917 4003 04C4 DF00 0818 C4C0 5212 66A0 8018 9A47 0618 66A3 6618 294C 2918

C4C0 5318 9A47 0718 294C 2A16 66A3 6412 66A3 0018 9A47 0818 294C 2B18 9A47 0916

66A3 4014 66A1 D018 294C 2C18 66A3 1E14 66A1 7018 69EB 9F14 66A2 C014 66A0 0012

66A0 4016 66A3 6814 66A1 5018 66A3 2018 C4C0 5012 66A0 0018 9A47 0418 66A3 6414

66A2 E018 C4C0 5116 66A3 4412 66A2 C018 9A47 0518 66A3 6518 294C 2816 66A3 5C14

66A2 700E 66A0 1869 EB9B 1466 A100 1266 A140 1869 EB9C 1466 A**MSG 00002 TRUNCATED**

MSG 00002 CONTINUATION #01250 1266 A100 1466

A140 1869 EB9D 1666 A37C 189A 4700 1869 EB9E 1266 A3C0 1466 A120 1466 A0C0 1866

A316 1266 A380 1466 A210 1869 EB97 1866 A338 1866 A317 1866 A339 1866 A318 1869

EB99 1866 A33A 1466 A0E0 1466 A230 1266 A0C0 139A 4700 189A 471E 1869 EB9A 1469

EB90 1866 A333 1466 A1C0 1666 A354 1869 EB93 1866 A334 1666 A374 1669 EB90 1866

A3FE 189A 4718 1869 EB94 1866 A335 1866 A3FF 189A 4719 1869 EB95 1866 A39B 1266

A1C0 1869 EB96 1866 A337 189A 4712 1866 A32F 1666 A358 1266 A180 1466 A160 189A

4713 1866 A330 1869 EB90 1866 A331 1666 A378 1669 EB94 1869 EB91 1466 A030 1866

A332 1466 A180 1466 A2D0 1869 EB92 189A 4**MSG 00002 TRUNCATED**

MSG 00002 CONTINUATION #0270E 1266 A280 1866 A32B 1466 A110 1666

A34C 1466 A260 189A 470F 1866 A32C 1666 A36C 189A 4710 1866 A32D 189A 4711 1666

A348 1466 A130 1866 A32E 189A 470A 1829 4C2D 1866 A327 1666 A350 1466 A0B0 1666

A370 189A 470B 1466 A200 1266 A240 1829 4C2E 1866 A328 1466 A0F0 189A 470C 1266

A200 1466 A240 1829 4C2F 1866 A329 189A 470D 1866 A32A 1466 A220

*Oct 4 04:52:31.888: %BGP-6-MALFORMEDATTR: Malformed attribute in (BGP(0) Prefixes: 196.192.82.0/24 102.160.128.0/18 154.71.6.0/24) received from 196.223.0.201,

%Log packet overrun, PC 0xAAB1DCE504, format:

unsupported or mal-formatted message received from %s:

%s

*Oct 4 04:52:40.072: %BGP-5-ADJCHANGE: neighbor 196.223.0.202 Up

*Oct 4 04:52:40.078: %BGP-6-MSGDUMP_LIMIT: unsupported or mal-formatted message received from 196.223.0.202:

FFFF FFFF FFFF FFFF FFFF FFFF FFFF FFFF 025A 0200 0000 1440 0101 0040 0206 0201

0000 7917 4003 04C4 DF00 0818 C4C0 5212 66A0 8018 9A47 0618 66A3 6618 294C 2918

C4C0 5318 9A47 0718 294C 2A16 66A3 6412 66A3 0018 9A47 0818 294C 2B18 9A47 0916

66A3 4014 66A1 D018 294C 2C18 66A3 1E14 66A1 7018 69EB 9F14 66A2 C014 66A0 0012

66A0 4016 66A3 6814 66A1 5018 66A3 2018 C4C0 5012 66A0 0018 9A47 0418 66A3 6414

66A2 E018 C4C0 5116 66A3 4412 66A2 C018 9A47 0518 66A3 6518 294C 2816 66A3 5C14

66A2 700E 66A0 1869 EB9B 1466 A100 1266 A140 1869 EB9C 1466 A**MSG 00003 TRUNCATED**

MSG 00003 CONTINUATION #01250 1266 A100 1466

A140 1869 EB9D 1666 A37C 189A 4700 1869 EB9E 1266 A3C0 1466 A120 1466 A0C0 1866

A316 1266 A380 1466 A210 1869 EB97 1866 A338 1866 A317 1866 A339 1866 A318 1869

EB99 1866 A33A 1466 A0E0 1466 A230 1266 A0C0 139A 4700 189A 471E 1869 EB9A 1469

EB90 1866 A333 1466 A1C0 1666 A354 1869 EB93 1866 A334 1666 A374 1669 EB90 1866

A3FE 189A 4718 1869 EB94 1866 A335 1866 A3FF 189A 4719 1869 EB95 1866 A39B 1266

A1C0 1869 EB96 1866 A337 189A 4712 1866 A32F 1666 A358 1266 A180 1466 A160 189A

4713 1866 A330 1869 EB90 1866 A331 1666 A378 1669 EB94 1869 EB91 1466 A030 1866

A332 1466 A180 1466 A2D0 1869 EB92 189A 4**MSG 00003 TRUNCATED**

MSG 00003 CONTINUATION #0270E 1266 A280 1866 A32B 1466 A110 1666

A34C 1466 A260 189A 470F 1866 A32C 1666 A36C 189A 4710 1866 A32D 189A 4711 1666

A348 1466 A130 1866 A32E 189A 470A 1829 4C2D 1866 A327 1666 A350 1466 A0B0 1666

A370 189A 470B 1466 A200 1266 A240 1829 4C2E 1866 A328 1466 A0F0 189A 470C 1266

A200 1466 A240 1829 4C2F 1866 A329 189A 470D 1866 A32A 1466 A220

Resolution: Run command on cisco “no bgp enforce-first-as”

Revision #5
Created 21 November 2022 11:58:07 by Keessun Fokeerah
Updated 4 October 2023 12:03:20 by Keessun Fokeerah