Features of modern route servers at IXPs

Agenda

- Recap last session
- Bilateral peering
- Multilateral peering with route servers
  - What happens
  - What you can do

Disclaimers

- Interrupt me
- Let’s be interactive
- Very happy to get feedback

Last session

- Efficiency when peering at multiple IXPs
- Questions?

Last session

- Efficiency when peering at multiple IXPs
- Questions?
  - There are different environments.

Motivation

- When peering at an IXP, or more IXPs, one has to control traffic flows
- Outgoing traffic flows everyone can direct themselves, by prioritising routes within their network
- Incoming flows are controlled by managing advertisements
- BGP route servers offer some mechanisms

Before route servers

- We are connected to IXPs
- We don’t have to pull 20 cables to talk to 20 other networks :-)
- At the IXP we can talk BGP with them over the switch, exchange routing info, forward traffic
- That is not an exception, that is normal! Bilateral peering
- Don’t need a TIX Manager for that! :-)

Bilateral peering

- In principle, it should be done
  - Or is configuring one neighbor too much work…?
- This serves as a backup when IXP route servers don’t work
  - Or can be considered the primary choice, with RS as backup

Now we also have the RS

IXP BGP route servers

- not to be confused with DNS root servers

- BGP speaker
- Software, daemon
- Setup not optimised to forward a lot of traffic
- Can and should have additional features
  - That’s why we’re here

IXP RS features – 1

It does not add itself to the AS_PATH

- Peers get the same AS_PATH as if peering bilaterally
- Routes will be preferred in the same way. On other routers in receiving peer’s AS.
- In BGP decision algorithm, when comparing AS_PATH length there’s no difference

IXP RS features – 1 cont.

But some BGP implementations check that the first AS in the AS_PATH is the same as the neighbors. In normal eBGP that makes sense.

For sessions with IXP RS that would reject them.

Cisco:

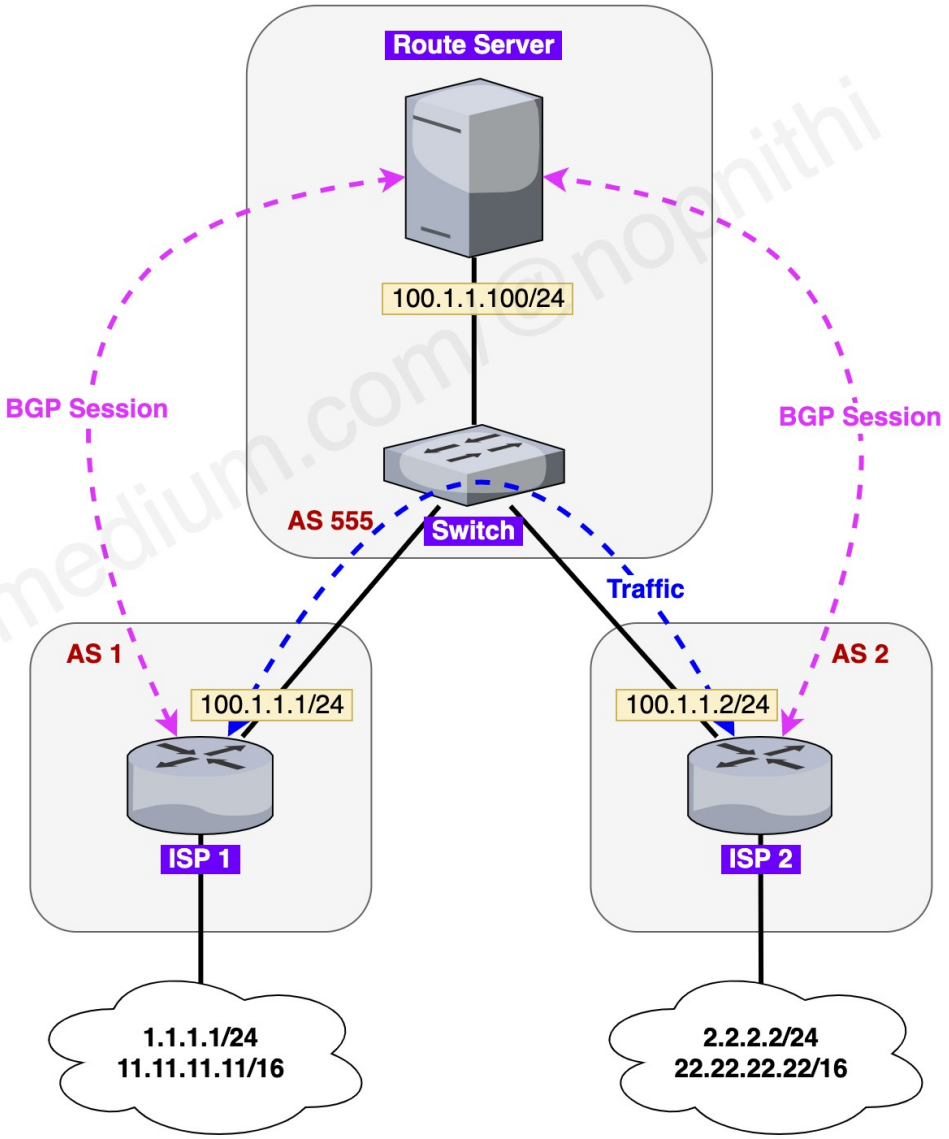
```
router bgp < asn >
no bgp enforce-first-as
```

Huawei:

```
bgp < asn >
 undo check-first-as
```

IXP RS features – 2

- It keeps the next-hop
- It tells advert recipient how to send traffic to other peer directly
- Traffic flows from peer via switch to peer
- Not via the RS!



IXP RS features – 3 – Filtering

Filtering

- RS are configured to filter incoming advertisements from peers
- RS should not accept unwanted advertisements
  - default route: 0.0.0.0/0 and ::/0
  - too specifics, like /28
- RS should check advertisements per IRR
- RS should do ROV using RPKI

Can show examples, later.

IXP RS features – 4 – Controlling

Controlling advertisements

- Seacom and Liquid don’t peer with the route server
- If we tell them that they MUST, they will leave
- This is because RS would normally advertise what they get from a peer to all the others.
- It is a good thing for the operator of the IXP, the operator of the RS, to give peers more options to control this.
- The BGP software can be configured to do this.
- BGP communities can be used by peers to signal the intend to the RS.

IXP RS features – 4 – Controlling

Controlling advertisements

- Normal BGP communities are 32-bit
- One can’t communicate **what** to do **and** 32bit info of the ASN concerned inside 32bit

1. With BGP community 0:37027 you can tell the RS to not advertise this to peer AS 37027
2. With community 0:33791 you can tell the TIX RS (using itself 33791) to not advertise to any other peer
3. But with community 33791:15399 you can tell the RS to advertise to peer AS 15399 , despite #2

- Above is talking about other peers with 16-bit ASNs

IXP RS features – 4 – Controlling

Controlling advertisements

- What if we want to talk about peers with 32-bit ASNs (above 65535)
- We can use “Large Communities” – RFC 8092
- From RFC 8195 section 4.4:

BGP Large Community	Description
64511:0:peer-as	Explicitly prevent announcement of route to peer-as
64511:1:peer-as	Explicitly announce route to peer-as
64511:0:0	Do not announce route to any peers by default
64511:1:0	Announce route to all peers by default

IXP RS features – Conclusion

- Peers should know this behaviour
- It is a feature by the IXP to support requirements by some peers

Show in config file

Other issues

- Peers should be better filtering what is advertised to the RS.
  - We see routes that peers are getting from their upstreams – these should not get advertised to the RS
- Some peer is not advertising all they could

Not optimal...

- Traffic from TZ to TZ via Kenya :-(

```
[frank@fisi ~]$ traceroute -q 1 41.75.217.2
traceroute to 41.75.217.2 (41.75.217.2), 30 hops max, 60 byte packets
 1  41.221.41.1 (41.221.41.1)  9.305 ms
 2  155.12.1.133 (155.12.1.133)  6.689 ms
 3  155.12.1.157 (155.12.1.157)  0.512 ms
 4  155.12.1.170 (155.12.1.170)  1.696 ms
 5  ae-2-37.er-02-dar.tz.seacomnet.com (105.21.168.45)  1.885 ms
 6  ae-2.cr-02-dar.tz.seacomnet.com (105.16.22.2)  14.978 ms
 7  xe-0-0-0-2.cr-02-mba.ke.seacomnet.com (105.16.8.17)  14.488 ms
 8  xe-0-1-0-7.cr-01-nbo.ke.seacomnet.com (105.16.15.101)  14.740 ms
 9  xe-0-2-0.pp-01-nbo.ke.seacomnet.com (105.16.19.8)  13.831 ms
10  196.223.21.100 (196.223.21.100)  20.128 ms
11  102.134.23.29 (102.134.23.29)  20.339 ms
12  10.88.35.149 (10.88.35.149)  13.821 ms
13  10.88.35.148 (10.88.35.148)  14.454 ms
14  *
15  *
```

The End

Thanks!

Any questions?

More meetings like this?

- What should we be talking about?
  - DNS resolvers?
  - Policies?
  - Internet routing registries?
  - RPKI?
  - VAS / AS112