

# RFC8950 Translation

André Grüneberg



# Short recap

- At 40th forum I explained RFC8950 and how to implement it on a route servers
- It's simple! Do it!
- BCIX did it, it's working
- But: No Legacy-Interop

## Interoperability of legacy IPv4 and RFC8950

- Q: Can legacy peers interact with RFC8950 peers?
  - A: Not really
- Q: Can route servers help?
  - A: Only in one direction (translating next-hop IPv4→IPv6), even then you cannot be sure the IPv6 next-hop works
- Recommendation: DO NOT ATTEMPT

**DON'T**

# Short recap

- At 40th forum I explained RFC8950 and how to implement it on a route servers
- It's simple! Do it!
- BCIX did it, it's working
- But: No Legacy-Interop
- I stand corrected!

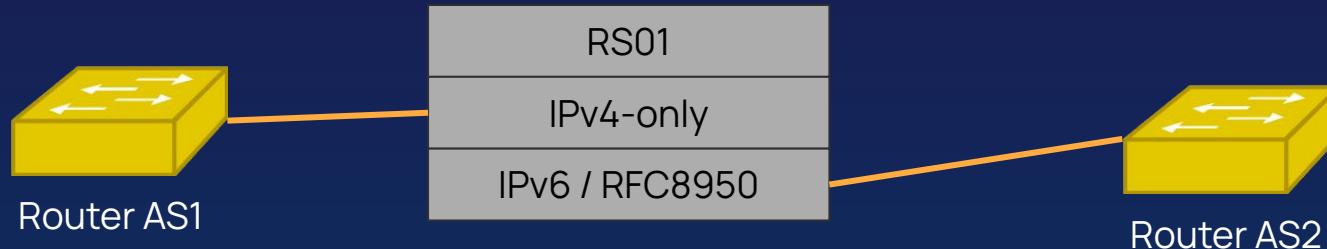
## Interoperability of legacy IPv4 and RFC8950

- Q: Can legacy routers interact with RFC8950 routers?
  - A: Not directly
- Q: Can route servers interwork?
  - A: Only in one direction, even then you need to do next-hop IPv4→IPv6), so next-hop works
- Recommended: use BGP AT

**DON'T**

# After deploying RFC8950

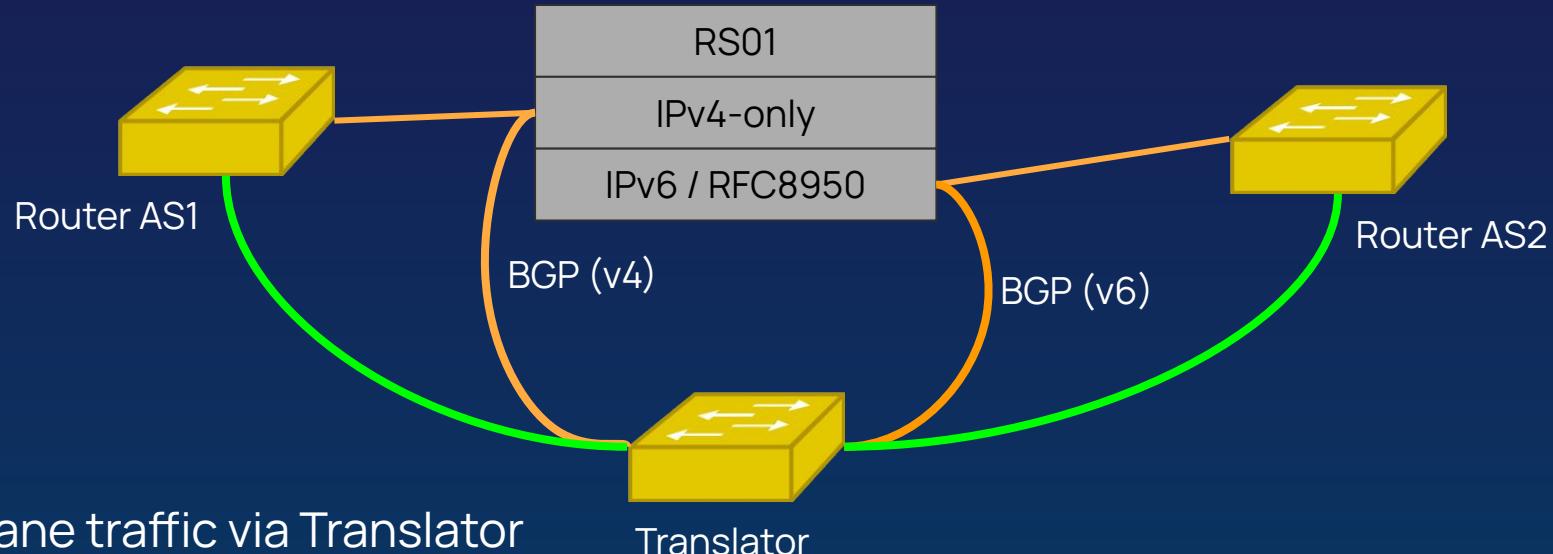
- There is IPv4 in IPv4-only and in RFC8950 part of RS



No dataplane traffic

# Adding Translator

- Translator peers with RS “instances”, next-hop self



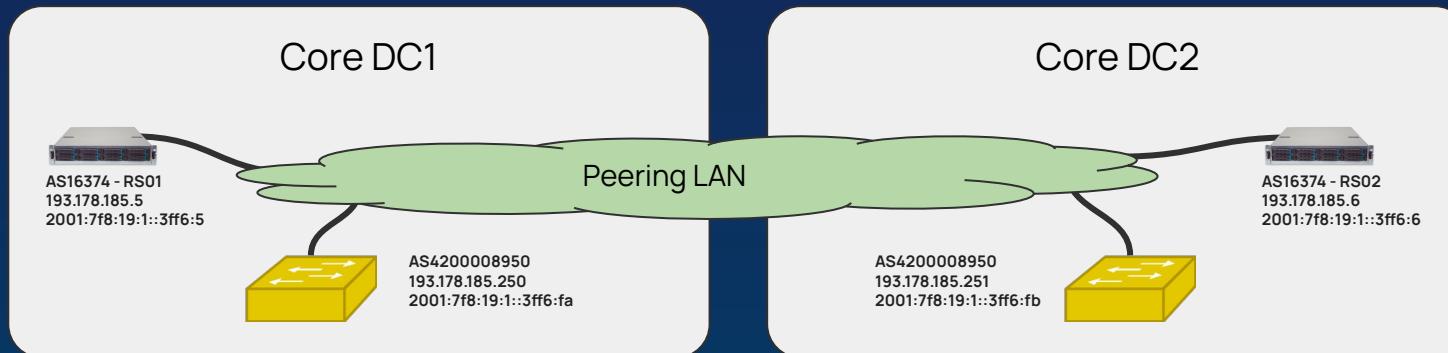
# The plan

- 2 Route servers → we need 2 routers
- ... connected with sufficient bandwidth
- Translator AS to be removed by route server
- Translated routes need lower priority than “native”
- No OTC added by RS towards translator (role: provider)
- Keep RS TE communities for handling on other end



# Our implementation (BCIX)

- We have 2x MX204 for the management network
- Located in same DC like route servers
- These have 2x 100G available → should be sufficient
- New routing instance, add IPv4+IPv6 addresses
- Configure BGP sessions to RS, no filters



# Our implementation (BCIX)

- Create special peer config in route server
  - First we tried iBGP RR, but required too many hacks
  - Eventually eBGP: AS4200008950
  - No filters (almost)
  - Ingress filter: remove first AS in Path
  - Egress filter: global do-not-export communities
- I.e. not an ordinary RS client in IXP Manager



# Let's turn it on!

- We did that on a friday ~noon
- We had high router IDs, so BIRD prefers native routes
- We did not expect any traffic
- An RFC8950-dual peer started to send traffic (5Mbps)



# Let's turn it on!

- We did that on a friday ~noon
- We had high router IDs, so BIRD prefers native routes
- We did not expect any traffic
- An RFC8950-dual peer started to send traffic (5Mbps)
- Junos reason “Always Compare MED”
- Learning: MED is reset on export via eBGP
- We also need to set high MED = 4200008950



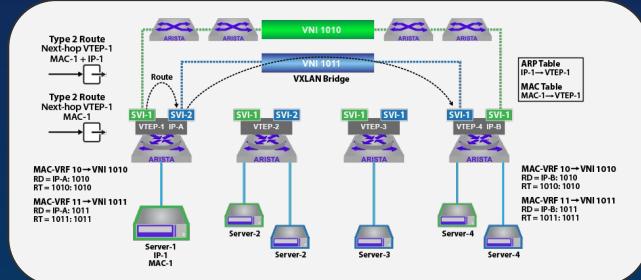
# Caveats

- Only applies to routes available on RS
- Only one best path
  - “Don’t export” TE communities may cause route hiding
  - Add-Path not applicable
- We cannot scale bandwidth



# Scalability

- Idea: Integration into EVPN fabric
  - Employ Asymmetric IRB
  - Each edge switch uses virtual IPv4/IPv6 address on SVI
  - Challenge how to get routes from RS to all edge switches
    - IP reachability of edge switches via Peering LAN? Security?
    - Nokia SR-OS can leak BGP routes incl. next-hop
    - Arista cannot (yet?)



# Summary

- Find 2 routers with sufficient bandwidth
- Have 2 sets of IP addresses available
- Add simple BGP configuration
- Use eBGP to connect Route servers

