RFC8950

Save IPv4 addresses on interconnection

Marian Rychtecký, André Grüneberg

NIX·CZ   euro-ix   BCIX

Who are we?

- Euro-IX RFC8950 IXP working group
  - Contributions by several IXPs in the room
  - Open to non-IXPs as well
  - Documentation of work in GitHub (https://github.com/euro-ix/rfc8950-ixp/)
- NIX.cz
  - Prague, Czechia; RFC8950 enabled route server, FENIX project
- BCIX
  - Berlin, Germany; RFC8950 enabled route server

What's it all about?

- Old Topic, new number RFC8950 = RFC5549 (effectively, small diff for VPNs)
- From 2009 and people have been thinking of its use for IXPs

ams*ix*
amsterdam internet exchange

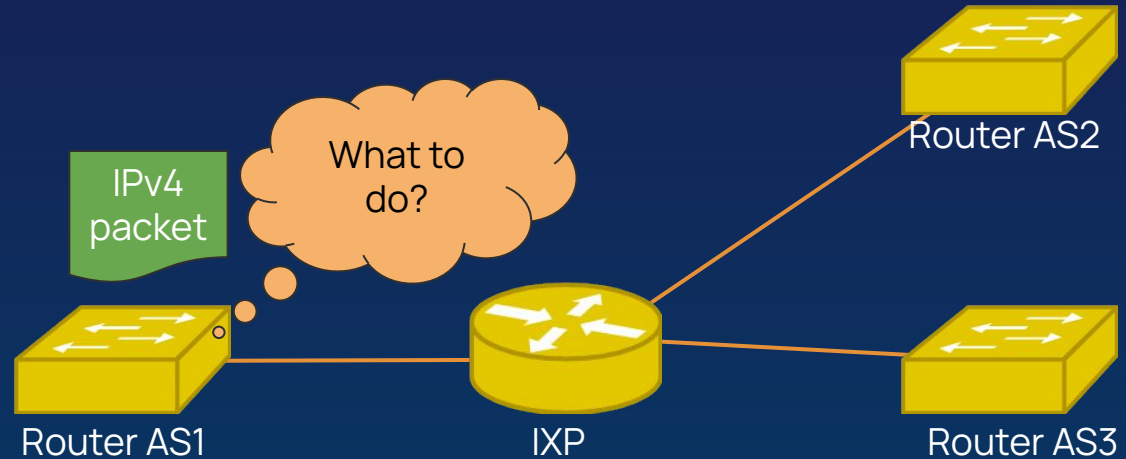RFC 5549
BGP IPv4 NLRIs with an IPv6 next hop

RIPE-65 Amsterdam 25-09-2012
arien.vijn@ams-ix.net
stefan.plug@ams-ix.net

How does it work?

- Router's data plane receives packet
- Lookup destination IP in FIB, Longest-Prefix-Match (TCAM)
- Next-hop: target interface + target MAC

IPv4 packet

What to do?

Router AS1

IXP

Router AS2

Router AS3

How does it work?

- To find target MAC a resolution may be necessary
- In RIB we have a next-hop IP address
  - Next-hop IPv4 → ARP
  - Next-hop IPv6 → IPv6 ND (Neighbor Discovery)

IPv4 packet

What's the target MAC?

Router AS2

Router AS1

IXP

Router AS3
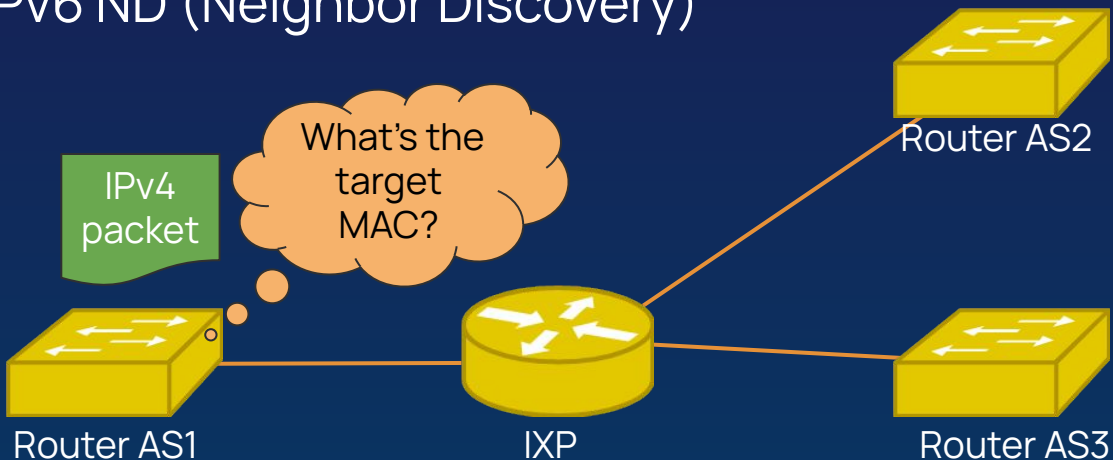
How does it work?

- Router learns next-hop IP from BGP
  - It's a variable length field
- RFC8950 re-defines interpretation:
  - 4 octets = IPv4
  - 16/32/48 octets = IPv6

IPv4 packet

What's next-hop IP?

Router AS2

Router AS1

IXP

Router AS3
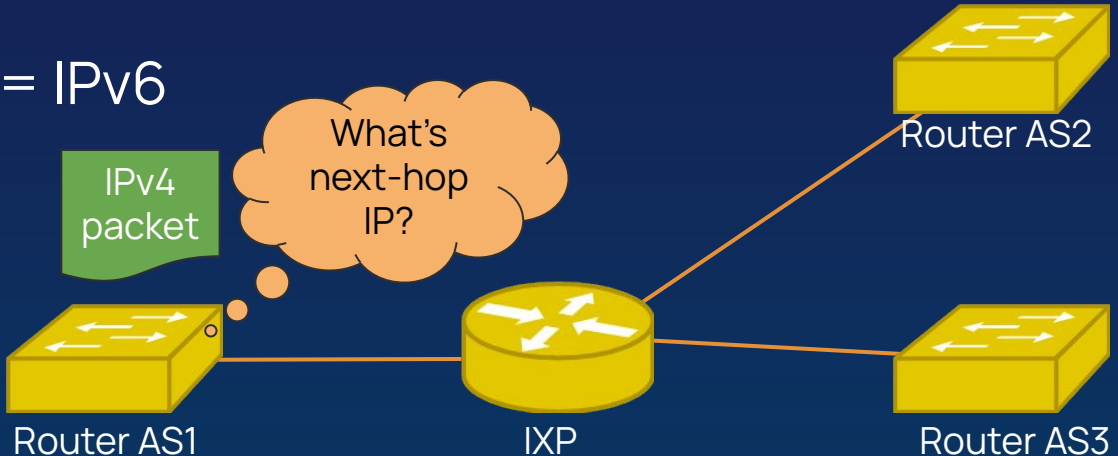
How does it work?

- Previously the address family of the next-hop was defined by the destination's AFI
- So we need a Capability advertised in BGP OPEN
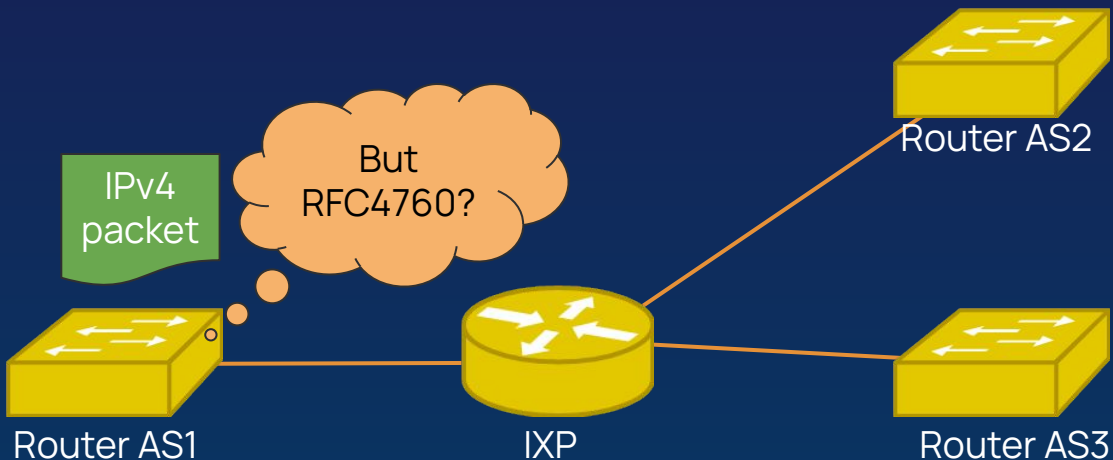
How does it work?

- Summary:
  - BGP capability exchanged during session start
  - Next-hop IPv6 address in BGP NLRI → RIB → FIB
  - MAC lookup from IPv6 next-hop
- Just a few pieces!
- No tunneling!

IPv4 packet

Go packet! Go!

Router AS1

IXP

Router AS2

Router AS3
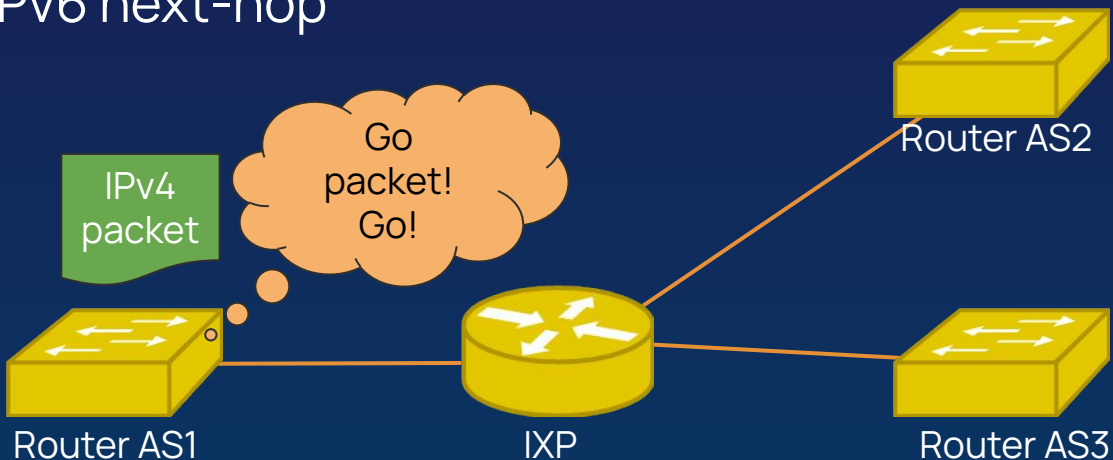
Why should network operators bother?

- IPv4 addresses are scarce
  → Just one IPv4 address per device
- You need IPv6 interconnection anyway
  → No double maintenance
- Simplification of IPAM
- It's the end of ARP
- No more renumbering on IXPs

I'M SO EXCITED

What's the catch?

- Router vendors need to implement
  - IPv6 next-hop for IPv4 routes
  - BGP capability for Extended Next-hop
- Networks may need to upgrade their equipment
- Admins need to learn it
  - Good news: uptake of BGP unnumbered in DC fabrics
- Traceroute without link specific IPv4 address
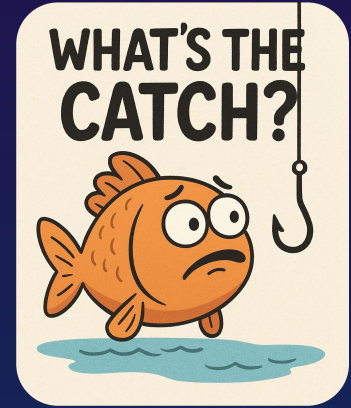
How about router vendors' support?

- We tested a few (virtual environment)
- Issues with old platforms
- Updates necessary
- Example Config:

  https://github.com/euro-ix/rfc8950-ixp

Huawei?
Anyone?

Vendor	OS	Status
Arista	EOS	Since 4.22.1F
Cisco	IOS XE	Not supported
Cisco	IOS XR	Since 7.3.3
Cisco	NX-OS	Supported, not tested
CZNIC	Bird (+ Linux)	Since Bird 2.0.8, Linux 5.2
Exa	ExaBGP	Since 4.1.0, only RIB
Extreme Networks	IronWare, SLX-OS	Not supported
FRR community	FRR (+ Linux)	FRR 9.1.3 (7.0.0), Linux 5.2
IPInfusion	OcNOS	Not fully supported
Juniper	JunOS	Since 21.2
Mikrotik	ROS	Since 7.20 (beta)
Nokia	SR-OS	Since 19.5R1
Nokia	SR Linux	Since 20.06, not tested
OpenBSD	OpenBGPd	8.8, only RIB
OSRG	GoBGP	Supported, only RIB
RtBrick	RBFS	All versions
Vyatta	VyOS	Since 1.4.3 (1.2.2)

Where can you apply it?

- Any type of interconnect
  - Your datacenter fabric / underlay
  - iBGP (you may remove IPv4 addresses on links after full mesh migration)
  - eBGP PNI
  - eBGP IXP (full migration may take a while)

ANYWHERE IN THE NETWORK

Status at some IXPs (in Euro-IX area)

- RFC8950-only
  - TREX Turku

- RFC8950 on test Route servers

  - TREX Tampere, …

- RFC8950 on production Route servers
  - BCIX
  - NIX.CZ + NIX.SK

NIX.cz: Adding RFC8950 and security

- Support IPv4 channel over IPv6 BGP session only

- Added mandatory GTSM support (RFC 6720)

- "Added" TCP-AO support (waiting for stable kernel)

- Not opening session without RFC8950 support:

  ```
  "require extended next hop on;"
  ```

Anyone: RFC8950 and security

- Filters, Filters and Filters!
    - don't forget you have two AFIs in the channel now
    - RPKI filtering
    - IRRDB filtering
    - next-hop check (only v6, not v4)
    - Max prefix per AFI (or global)

One session - two channels

RFC8950 at NIX.CZ – Looking glass

Facing the problems

- Session shut down due to
   "bad nhop len: 32 for afi 1, safi 1"
   (even RFC8950 support was signalled)

- JunOS 19.4R3-S7.3

- Resolved by upgrading to JunOS 23.4R2-S3

It works ... some stats, next-steps

- A few networks removed IPv4 session completely

- Recently 145 v4 prefixes with v6 next-hop on RS

- In the trial period with BIRD3

- After the trial (Q4 2025) plans for online workshop / education

Spread the word!
Questions?
Ideas?

ag@bcix.de   mr@nix.cz

Backup

Real-life issues

- Some peers may have IPv4 AFI enabled on IPv6 BGP session without Extended Next-hop capability and may try to send you a full-table → filter!
- Filters and max-prefix limits may be per BGP session (e.g. JunOS) or per address family
- IPv4 forwarding may need to be enabled explicitly → else blackhole