

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



RFC 5549

BGP IPv4 NLRIs with an IPv6 next hop

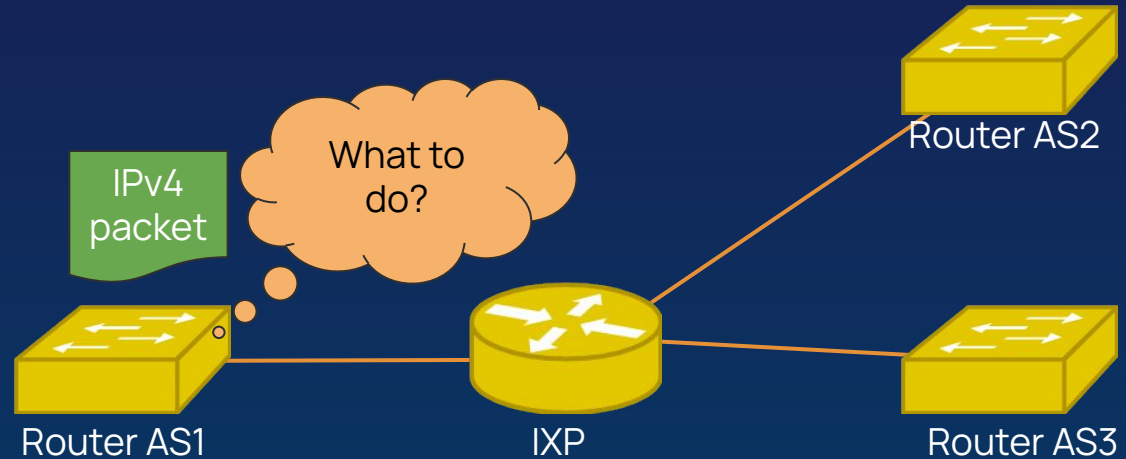
RIPE-65 Amsterdam 25-09-2012

arien.vijn@ams-ix.net

stefan.plug@ams-ix.net

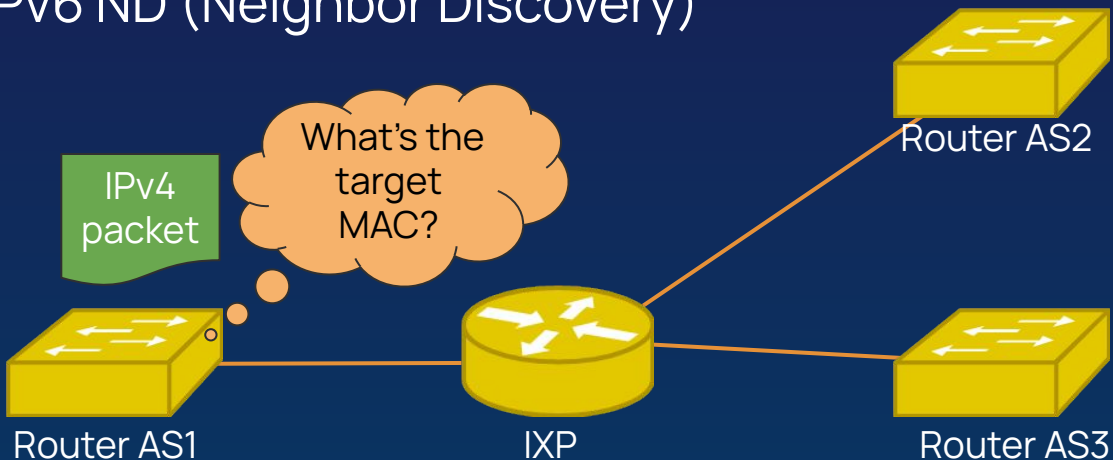
How does it work?

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



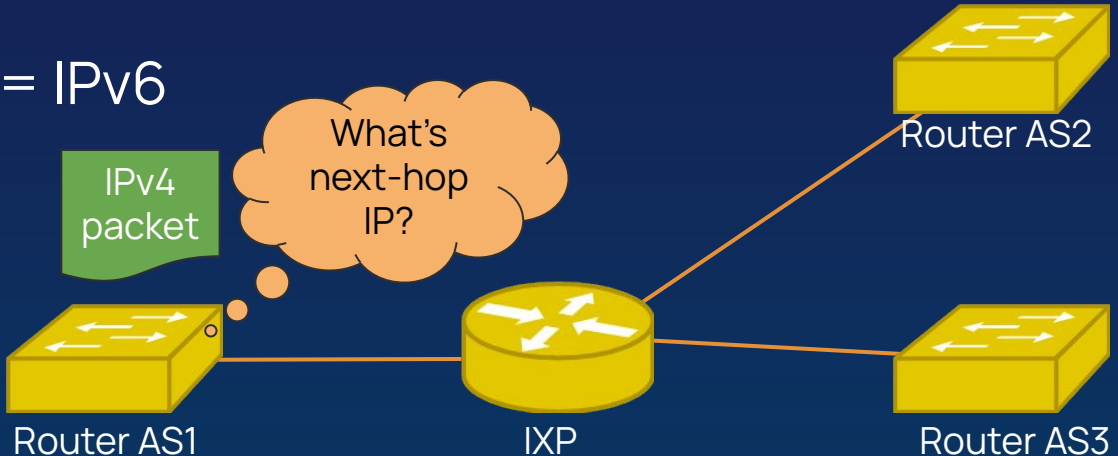
How does it work?

- 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)



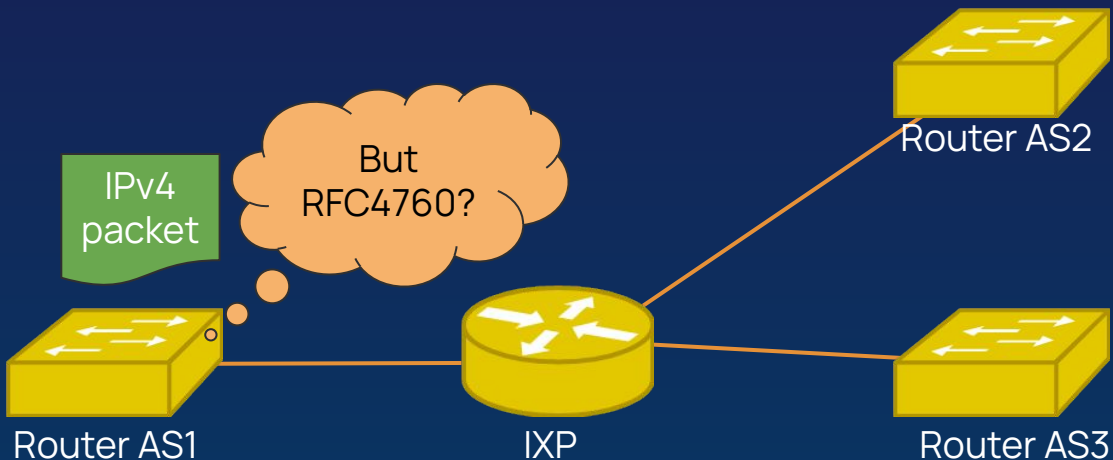
How does it work?

- 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



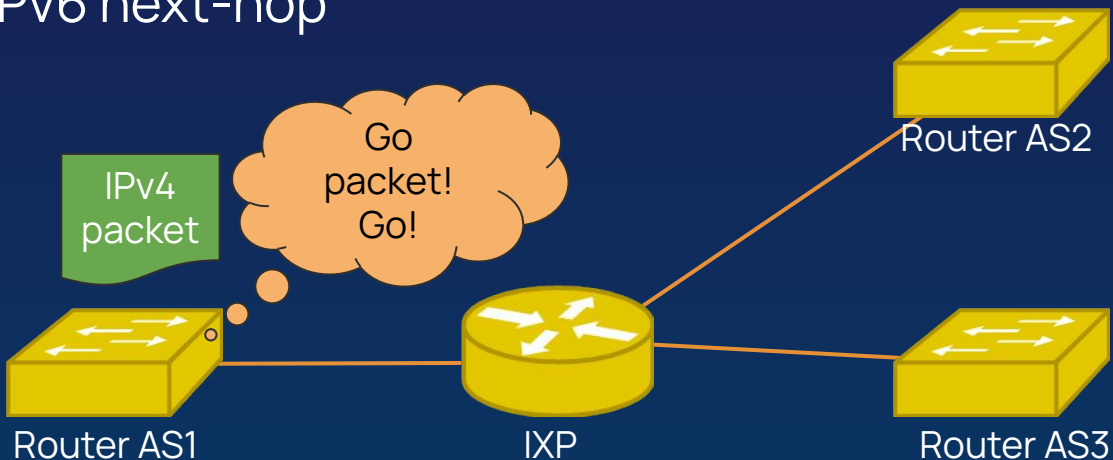
How does it work?

- 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!



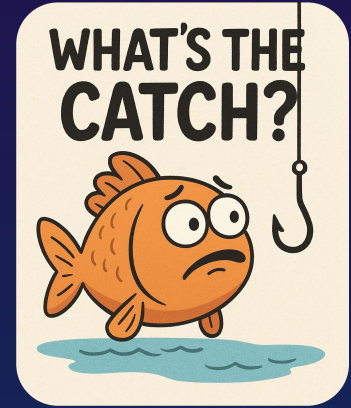
Why should network operators bother?

- 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



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

```
NIX.CZ          T64_6881x2    2001:7f8:14::4    6881    Est

R6_6881x2  BGP      ---      up      2025-03-11 16:45:54  Established
Description:    NIX.CZ - 2001:7f8:14::4 - (2)
Created:        2025-03-11 16:45:49
BGP state:      Established
Neighbor address: 2001:7f8:14::4
Neighbor AS:    6881
Local AS:       47200
Neighbor ID:    93.190.135.253
Local capabilities
Multiprotocol
  AF announced: ipv4 ipv6
Route refresh
  Extended next hop
  IPv6 nexthop: ipv4
Graceful restart
4-octet AS numbers
Enhanced refresh
Long-lived graceful restart
Neighbor capabilities
Multiprotocol
  AF announced: ipv4 ipv6
Route refresh
Extended next hop
  IPv6 nexthop: ipv4 ipv4-mc vpn4-mpls
Graceful restart
Restart time: 120
AF supported: ipv4 ipv6
AF preserved:
4-octet AS numbers
Session:        external route-server AS4
```

```
Channel ipv6
State:      UP
Import state: UP
Export state: READY
Table:      T6_6881x2
Preference: 100
Input filter: bgp_in_AS6881x2
Output filter: bgp_peer_export6
Import limit: 20
Action:     block
Routes:      8 imported, 0 filtered, 0 exported, 8 preferred
Route change stats:  received  rejected  filtered  ignored  RX limit  IN limit  accepted
Import updates:      8           0         0         0         0         0         8
Import withdraws:    0           0        ---         0        ---         0
Export updates:       0           8       3479        ---         0         0
Export withdraws:    0           ---        ---        ---         0

BGP Next hop:  2001:7f8:14::100 fe80::e03a:53ff:fe30:3a80
Pending 0 attribute sets with total 0 prefixes to send
```

```
Channel ipv4
State:      UP
Import state: UP
Export state: READY
Table:      T64_6881x2
Preference: 100
Input filter: bgp_in_AS6881x2
Output filter: bgp_peer_export4
Import limit: 5000
Action:     block
Routes:      10 imported, 0 exported, 10 preferred
Route change stats:  received  rejected  filtered  ignored  RX limit  IN limit  accepted
Import updates:      10           0         0         0         0         0       10
Import withdraws:    0           0        ---         0        ---         0
Export updates:       0          10        352        ---         0         0
Export withdraws:    0           ---        ---        ---         0

BGP Next hop:  2001:7f8:14::100 fe80::e03a:53ff:fe30:3a80
Pending 0 attribute sets with total 0 prefixes to send
```


RFC8950 at NIX.CZ – Looking glass

Imported Routes

Routes imported from protocol: R6_6881x3 NIX.CZ on server: NIX SECRS-1

Show 100 entries

Search:

Copy

Print

CSV

Filter

Reset

Visibility

Network	Next Hop	Flags	ΣC	ΣLC	ΣEC	AS Path	Action
91.207.231.0/24	2001:7 (NIX.CZ) ::2	S V	4	0	0	6881 6881 6881 6881 6881	Show
93.190.128.0/21	2001:7 (NIX.CZ) ::2	S V	4	0	0	6881 6881 6881 6881 6881	Show
93.190.130.10/32	2001:7 (NIX.CZ) ::2	S V RTBH	3	0	0	6881 6881 6881 6881 6881	Show
195.47.235.0/24	2001:7 (NIX.CZ) ::2	S V	4	0	0	6881 6881 6881 6881 6881	Show
2a02:38::/32	2001:7 (NIX.CZ) ::2	S V	4	0	0	6881 6881 6881 6881 6881	Show
2a02:38:bbbb::10/128	2001:7 (NIX.CZ) ::2	S V RTBH	3	0	0	6881 6881 6881 6881 6881	Show

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