Global Routing Operations
Internet-Draft
Updates: 6890, 7947 (if approved)
Intended status: Standards Track
Expires: 4 April 2026

M. Matejka
CZ.NIC
D. Wagner
DE-CIX

1 October 2025

Route Server Next Hop Translation
draft-marenamat-grow-route-server-nh-translation-00

Abstract

With the advent of RFC8950, Internet Exchang Points (IXPs) are enabled to rely solely on IPv6 addresses for adressing in their peering LANs. However, routers not supporting RFC8950 are a technical roadblock.

It is easier to extend the capabilities of the IXP Route Server (RS) instead of those of every unsupporting router. Thus, this document introduces the concept of Specific Local Address Tables (SLATs). SLATs translate BGP next hops between all IXP members, regardless of their RFC8950 support, paving the way for IPv6-only IXPs.

This document updates RFC 6890 by registering a special-purpose address, and RFC 7947 by specifying an allowed route modification at the route server.

About This Document

This note is to be removed before publishing as an RFC.

Status information for this document may be found at https://datatracker.ietf.org/doc/draft-marenamat-grow-route-server-nh-translation/.

Discussion of this document takes place on the Global Routing Operations Working Group mailing list (mailto:grow@ietf.org), which is archived at https://mailarchive.ietf.org/arch/browse/grow/. Subscribe at https://www.ietf.org/mailman/listinfo/grow/.

Source for this draft and an issue tracker can be found at https://github.com/marenamat/ietf-draft-marenamat-grow-route-server-nh-translation.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

This Internet-Draft will expire on 4 April 2026.

Copyright Notice

Table of Contents

1. Introduction 3
2. Conventions and Definitions 3
3. Providing reachability between Legacy and Unnumbered speakers 4
3.1. Client Address Assignment 4
3.1.1. MAC Address Assignment 4
3.1.2. IPv6 Address Assignment 4
3.1.3. IPv4 Address Assignment 5
3.2. ARP and ND Proxy Configuration 6
3.3. NEXT_HOP Attribute Management at Route Servers 6
4. IXP Interconnection Space 6
5. Operational and Management Considerations 7
5.1. Step-by-Step Rollout 8
5.2. Bilateral Peerings 8
5.3. Address Translation Transparency 8
6. Security Considerations 8
7. IANA Considerations 9
8. References 9
8.1. Normative References 9
8.2. Informative References 10

Acknowledgments 11
Authors' Addresses 11

1. Introduction

Traditionally, Internet Exchange Point (IXP) Border Gateway Protocol (BGP) Route Servers (RS) [RFC7947] serve IPv6 Network Layer Reachability Information (NRLI) with IPv6 next hops, and IPv4 NLRI with IPv4 next hops to the BGP speakers in their peering LAN. On the one hand, this dual-stack operation allows both IPv4 and IPv6 supporting BGP speakers to exchange NLRI with another and the route server. On the other hand, this requires them to have next hop addresses of the same Address Familiy (AF) as well.

With the depletion of available IPv4 address space, solutions have emerged to support forwarding of IPv4 traffic over IPv6-only intermediate hosts [I-D.chroboczek-intarea-v4-via-v6]. In the IXP environment, however, these networks would still require an IPv4 address to be assigned to allow for routing from and to legacy-only networks where IPv6 next hops for IPv4 NLRIs [RFC8950] are not supported.

This document specifies how to extend the Address Resolution Protocol (ARP) Proxy [RFC9161] functionality to allow deployment of IPv6 next hops for IPv4 NLRIs [RFC8950], without the need to assign public IPv4 addresses to any of the BGP speakers at IXPs.

This document does not cover IPv6 NLRIs with IPv4 next hops.

2. Conventions and Definitions

The terminology of [RFC9161], [RFC7947] and [RFC4271] applies.

Client: A BGP speaker which is connected to the IXP's Route Server. The Client may be a Legacy speaker, Supporting speaker or Unnumbered speaker.

Legacy speaker: Any Client with no support for IPv4 NLRIs with IPv6 next hops in context of an IXP.

Supporting speaker: Any Client with support for IPv4 NLRIs with IPv6 next hops, while still capable of producing and receiving IPv4 next hops.

Unnumbered speaker: Any Client with support for IPv4 NLRIs with IPv6 next hops, and with no support for IPv4 next hops.

Production IPv4 prefix: The IPv4 prefix used by the IXP operator to

assign IPv4 addresses to the Clients.

Production IPv6 prefix: The IPv6 prefix used by the IXP operator to assign IPv6 addresses to the Clients.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Providing reachability between Legacy and Unnumbered speakers

All IPv4 routes announced to and from Legacy speakers MUST have IPv4 next hops, while all IPv4 routes announced to and from Unnumbered speakers MUST have IPv6 next hops. To facilitate reachability between these Clients, we need to translate between IPv4 and IPv6 next hops in BGP, IPv6 Neighbor Discovery (ND) and ARP.

3.1. Client Address Assignment

3.1.1. MAC Address Assignment

All Clients SHOULD have a fixed MAC address set and registered with the IXP.

3.1.2. IPv6 Address Assignment

All Clients MUST have their IPv6 link-local address (LLA) and IPv6 globally unicast address (GUA) assigned by the IXP. They MAY set these addresses up on the respective interfaces while their already established BGP sessions are still able to run.

These assignments MUST be unique, such that for any two triples (MAC, LLA, GUA) and (MAC', LLA', GUA') it holds that MAC != MAC', LLA != LLA' and GUA != GUA'.

IPv6 addresses from the Production IPv6 prefix of the IXP MAY be used for GUA allocation if there are unused addresses available and the above requirement holds.

The resulting set of triples is stored in a Local Address Table (LAT). This table is maintained by the IXP and used to translate next hops to MAC addresses for Unnumbered speakers.

MAC Address	Link Local Address	Global Unicast Address
00-00-5E-00-53-10	FE80::10	2001:db8::10
00-00-5E-00-53-20	FE80::20	2001:db8::20
...	...	...

Table 1: Local Address Table (LAT)

3.1.3. IPv4 Address Assignment

Due to IPv4 scarcity, IXP are typically assigned much less spacious Production IPv4 prefixes than Production IPv6 prefixes. Therefore, the IXP, in cooperation with every Supporting Speaker and Legacy Speaker, MUST decide on an IPv4 prefix (or a set of IPv4 prefixes) short enough to accommodate the number of Clients in the IXP network. This prefix MAY be different for different Clients. This prefix is called Client-specific local prefix (CSLP).

For every Supporting and Legacy Speaker, the IXP then adds another column for every CSLP to the LAT, completing it to a Specific Local Address Table (SLAT). These columns then hold a unique IPv4 address assigned from the respective CSLP for every triple in the LAT. These entries are used to translate next hops to MAC addresses for Legacy speakers.

MAC Address	Link Local Address	Global Unicast Address	CSLP 1	CLSP 2	...
MAC Address	Link Local Address	Global Unicast Address	CSLP 1	CLSP 2	...
00-00-5E-00-53-10	FE80::10	2001:db8::10	10.0.0.10	192.0.2.10	...
00-00-5E-00-53-20	FE80::20	2001:db8::20	10.0.0.20	192.0.2.20	...
...	...	...	...	...	...

Table 2: Specific Local Address Table (SLAT)

The Unnumbered Speakers need no such prefix negotiation and therefore have no risk of adding another CSLP to bloat the SLAT.

Legacy Speakers SHOULD set up their NEXT_HOP attribute handling so that they never propagate the IPv4 addresses from the SLAT outside any communication with the RS.

3.2. ARP and ND Proxy Configuration

For each Client, the IXP MUST set up ARP and ND snooping. The IXP MUST NOT forward neither ARP nor ND traffic between Clients. The IXP MUST answer all ARP and ND requests from the Clients themselves using the respective SLAT column for that Client.

3.3. NEXT_HOP Attribute Management at Route Servers

When a route with IPv4 NLRI and IPv4 NEXT_HOP Attribute is announced from any Client, the RS MUST rewrite the NEXT_HOP according to the Client’s IPv6 GUA or LLA entry in the SLAT.

When the RS sends a route to a Legacy speaker, it MUST rewrite the NEXT_HOP according to the IPv4 address assigned for the sender in the receiver’s CSLP column of the SLAT.

When the Route Server sends a route to a Supporting speaker, it SHOULD NOT rewrite the NEXT_HOP.

When the Route server sends a route to an Unnumbered speaker, it MUST NOT rewrite the NEXT_HOP.

The Route Server MUST NOT propagate any route where the NEXT_HOP attribute holds an address not assigned to any Clients in the SLAT.

Section 2.2.1 of [RFC7947] does not apply.

4. IXP Interconnection Space

This document requests an allocation of an IPv4 IXP Interconnection Space from the experimental range. By previous efforts [I-D.schoen-intarea-unicast-240], it has already been shown that these addresses are technically feasible to be used in limited environments. Here, the use is limited for local next hop resolution and possibly BGP session addressing.

It is RECOMMENDED that this prefix is used as the CLSP for every Client that does not use this prefix for other purposes. Having the same prefix as the IXP Interconnection Space for many Clients helps to reduce the size of the SLAT.

Clients MUST NOT propagate any routes with IPv4 NLRI from the IXP Interconnection Space.

Section 2.2.2 of [RFC6890] is updated by adding the following record:

Attribute	Value
Address Block	TBD
Name	IXP Interconnection Space
RFC	TBD
Allocation Date	TBD
Termination Date	N/A
Source	False
Destination	False
Forwardable	False
Global	False
Reserved-by-Protocol	False

Table 3: Shared Address Space

The allocation is probably not strictly needed, as most of the Legacy Speakers will still have some of the private IPv4 addresses [RFC1918] available to use for the SLAT. Yet, these available ranges may be different between networks. To reduce complexity, this allocation will help IXPs to have a shared SLAT for most of the Legacy Speakers.

Some large networks have also claimed recently Section 6.1 of [I-D.schoen-intarea-unicast-240] that they are already using the experimental range for their internal purposes because they are already out of the private IPv4 addresses. These networks would have probably needed to negotiate a custom CLSP with the IXP anyway, with or without the allocation.

5. Operational and Management Considerations

5.1. Step-by-Step Rollout

This setup should be possible to be rolled out in steps. First, the ARP and ND snooping is not dependent on anything else in this document. Then, setting up a new route server supporting IPv6 next hops for IPv4 NLRI, and allowing Supporting speakers to use that server while keeping also the traditional one.

The SLAT may be started as uniform for every Client reflecting the current address assignment from the Production IPv4 prefix. This allows the Legacy Speakers into the new route server, and gradual renumbering may occur later, Client-by-Client, when the Production IPv4 prefix starts being exhausted.

The Clients have to properly assess which address range is suitable for them to use for IXP interconnection. If using the IXP Interconnection Space, they also have to check whether these addresses are considered eligible as next hops by their routing equipment.

5.2. Bilateral Peerings

There might be reasons why any two Clients do not want to use the IXP’s RS to exchange their routing information. Hence, the information from the SLAT should be made publicly available and kept up to date. Clients can then perform the next hop translation themselves.

An alternative is to introduce a new BGP community that tells the RS to exclude the routing information exchanged via such bilateral peerings from the Looking Glasses (LG). This can be useful if privacy is of a concern and no self-translation can be performed.

5.3. Address Translation Transparency

The IXPs may have to rethink how they are displaying the route next hops in their human-facing interfaces (Looking Glasses). It may be handy to display the original next hop (if it was IPv4), the actual IPv6 next hop, and also the result of the egress translation for a selected Client.

6. Security Considerations

Implementing the ARP and ND snooping should improve the overall security of IXPs by blocking possible ARP or ND spoofing, both inadvertent and intended [DE-CIX-EVPN].

Mistakes in the MAC address registration and manual management of IP address assignment may lead to inadvertent invalid route announcement. It’s recommended to run automated address management with a single source of truth.

Mistakes in the next hop address translation may lead to inadvertent invalid route announcement. It’s recommended to run periodic automated checks whether the next hops actually resolve to the same address by the appropriate SLAT.

Mistakes in route announcements are contained to the route not being propagated further.

Mistakes in the Client setup may lead to spreading unreachable routes across their autonomous systems, causing inefficient routing.

It is recommended to log rogue GARP or IPv6 DAD communication to detect possible misconfigurations.

7. IANA Considerations

IANA is asked to record the allocation of an IPv4 /8 from the 240/4 range for use as IXP Interconnection Space as requested in Section 4.

The IXP Interconnection Space address range is: x.0.0.0/8.

[Note to RFC Editor: this address range to be added before publication]

8. References

8.1. Normative References

[I-D.chroboczek-intarea-v4-via-v6]
Chroboczek, J., Kumari, W., and T. H淡iland-J淡rgensen, "IPv4 routes with an IPv6 next hop", Work in Progress, Internet-Draft, draft-chroboczek-intarea-v4-via-v6-03, 20 January 2025, <https://datatracker.ietf.org/doc/html/draft-chroboczek-intarea-v4-via-v6-03>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/rfc/rfc2119>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/rfc/rfc4271>.

[RFC6890] Cotton, M., Vegoda, L., Bonica, R., Ed., and B. Haberman, "Special-Purpose IP Address Registries", BCP 153, RFC 6890, DOI 10.17487/RFC6890, April 2013, <https://www.rfc-editor.org/rfc/rfc6890>.

[RFC7947] Jasinska, E., Hilliard, N., Raszuk, R., and N. Bakker, "Internet Exchange BGP Route Server", RFC 7947, DOI 10.17487/RFC7947, September 2016, <https://www.rfc-editor.org/rfc/rfc7947>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/rfc/rfc8174>.

[RFC8950] Litkowski, S., Agrawal, S., Ananthamurthy, K., and K. Patel, "Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop", RFC 8950, DOI 10.17487/RFC8950, November 2020, <https://www.rfc-editor.org/rfc/rfc8950>.

[RFC9161] Rabadan, J., Ed., Sathappan, S., Nagaraj, K., Hankins, G., and T. King, "Operational Aspects of Proxy ARP/ND in Ethernet Virtual Private Networks", RFC 9161, DOI 10.17487/RFC9161, January 2022, <https://www.rfc-editor.org/rfc/rfc9161>.

8.2. Informative References

[DE-CIX-EVPN] King, T., "Peering LAN 2.0 — Introduction of EVPN at DE-CIX", 23 August 2023, <https://blog.apnic.net/2023/08/16/peering-lan-2-0-introduction-of-evpn-at-de-cix/>.

[I-D.schoen-intarea-unicast-240] Schoen, S. D., Gilmore, J. I., and D. M. Tht, "Unicast Use of the Formerly Reserved 240/4", Work in Progress, Internet-Draft, draft-schoen-intarea-unicast-240-09, 23 June 2025, <https://datatracker.ietf.org/doc/html/draft-schoen-intarea-unicast-240-09>.

[RFC1918] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G. J., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, DOI 10.17487/RFC1918, February 1996, <https://www.rfc-editor.org/rfc/rfc1918>.

Acknowledgments

TODO

Authors' Addresses

Maria Matejka
CZ.NIC
Milesovska 1136/5
13000 Praha
Czechia
Email: maria.matejka@nic.cz, mq@jmq.cz

Daniel Wagner
DE-CIX
Lindleystrae 12
60314 Frankfurt am Main
Germany
Email: daniel.wagner@de-cix.net