Stream: Internet Engineering Task Force (IETF)
RFC: 9229
Category: Experimental
Published: May 2022
ISSN: 2070-1721
Author: J. Chroboczek
IRIF, University of Paris

RFC 9229
IPv4 Routes with an IPv6 Next Hop in the Babel Routing Protocol

Abstract

This document defines an extension to the Babel routing protocol that allows announcing routes to an IPv4 prefix with an IPv6 next hop, which makes it possible for IPv4 traffic to flow through interfaces that have not been assigned an IPv4 address.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

The role of a routing protocol is to build a routing table, a data structure that maps network prefixes in a given family (IPv4 or IPv6) to next hops, which are (at least conceptually) pairs of an outgoing interface and a neighbour's network address. For example:

```
        destination                      next hop
    2001:db8:0:1::/64               eth0, fe80::1234:5678
    203.0.113.0/24                  eth0, 192.0.2.1
```

When a packet is routed according to a given routing table entry, the forwarding plane typically uses a neighbour discovery protocol (the Neighbour Discovery (ND) protocol [RFC4861] in the case of IPv6 and the Address Resolution Protocol (ARP) [RFC0826] in the case of IPv4) to map the next-hop address to a link-layer address (a "Media Access Control (MAC) address"), which is then used to construct the link-layer frames that encapsulate forwarded packets.

It is apparent from the description above that there is no fundamental reason why the destination prefix and the next-hop address should be in the same address family: there is nothing preventing an IPv6 packet from being routed through a next hop with an IPv4 address (in which case the next hop's MAC address will be obtained using ARP) or, conversely, an IPv4 packet from being routed through a next hop with an IPv6 address. (In fact, it is even possible to store link-layer addresses directly in the next-hop entry of the routing table, which is commonly done in networks using the OSI protocol suite).

The case of routing IPv4 packets through an IPv6 next hop is particularly interesting, since it makes it possible to build networks that have no IPv4 addresses except at the edges and still provide IPv4 connectivity to edge hosts. In addition, since an IPv6 next hop can use a link-local address that is autonomously configured, the use of such routes enables a mode of operation where the network core has no statically assigned IP addresses of either family, which significantly reduces the amount of manual configuration required. (See also [RFC7404] for a discussion of the issues involved with such an approach.)

We call a route towards an IPv4 prefix that uses an IPv6 next hop a "v4-via-v6" route. This document describes an extension that allows the Babel routing protocol [RFC8966] to announce v4-via-v6 routes across interfaces that have no IPv4 addresses assigned but are capable of forwarding IPv4 traffic. Section 3 describes procedures that ensure that all routers can originate ICMPv4 packets, even if they have not been assigned any IPv4 addresses.

The extension described in this document is inspired by a previously defined extension to BGP [RFC5549].

1.1. Specification of Requirements

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.

2. Protocol Operation

The Babel protocol fully supports dual-stack operation: all data that represent a neighbour address or a network prefix are tagged by an Address Encoding (AE), a small integer that identifies the address family (IPv4 or IPv6) of the address of prefix and describes how it is encoded. This extension defines a new AE, called "v4-via-v6", which has the same format as the existing AE for IPv4 addresses (AE 1). This new AE is only allowed in TLVs that carry network prefixes: TLVs that carry an IPv6 neighbour address use one of the normal encodings for IPv6 addresses.

2.1. Announcing v4-via-v6 Routes

A Babel node can use a v4-via-v6 announcement to announce an IPv4 route over an interface that has no assigned IPv4 address. In order to do so, it first establishes an IPv6 next-hop address in the usual manner (either by sending the Babel packet over IPv6, or by including a Next Hop TLV containing an IPv6 address and using AE 2 or 3); it then sends an Update, with AE equal to 4 (v4-via-v6) containing the IPv4 prefix being announced.

If the outgoing interface has been assigned an IPv4 address, then, in the interest of maximising compatibility with existing routers, the sender **SHOULD** prefer an ordinary IPv4 announcement; even in that case, however, it **MAY** send a v4-via-v6 announcement. A node **SHOULD NOT** send both ordinary IPv4 and v4-via-v6 announcements for the same prefix over a single interface (if the update is sent to a multicast address) or to a single neighbour (if sent to a unicast address), since doing that provides no benefit while doubling the amount of routing traffic.

Updates with infinite metric are retractions: they indicate that a previously announced route is no longer available. Retractions do not require a next hop; therefore, there is no difference between v4-via-v6 retractions and ordinary retractions. A node **MAY** send IPv4 retractions only, or it **MAY** send v4-via-v6 retractions on interfaces that have not been assigned an IPv4 address.

2.2. Receiving v4-via-v6 Routes

Upon reception of an Update TLV with AE equal to 4 (v4-via-v6) and finite metric, a Babel node computes the IPv6 next hop, as described in Section 4.6.9 of [RFC8966]. If no IPv6 next hop exists, then the Update **MUST** be ignored. If an IPv6 next hop exists, then the node **MAY** acquire the route being announced, as described in Section 3.5.3 of [RFC8966]; the parameters of the route are as follows:

- The prefix, plen, router-id, seqno, and metric **MUST** be computed as for an IPv4 route, as described in Section 4.6.9 of [RFC8966].
- The next hop **MUST** be computed as for an IPv6 route, as described in Section 4.6.9 of [RFC8966]. It is taken from the last preceding Next Hop TLV with an AE field equal to 2 or 3; if no such entry exists and if the Update TLV has been sent in a Babel packet carried over IPv6, then the next hop is the network-layer source address of the packet.

An Update TLV with a v4-via-v6 AE and metric equal to infinity is a retraction: it announces that a previously available route is being retracted. In that case, no next hop is necessary, and the retraction is treated as described in Section 4.6.9 of [RFC8966].

As usual, a node **MAY** ignore the update, e.g., due to filtering (see Appendix C of [RFC8966]). If a node cannot install v4-via-v6 routes, e.g., due to hardware or software limitations, then routes to an IPv4 prefix with an IPv6 next hop **MUST NOT** be selected.

2.3. Route and Seqno Requests

Route and seqno requests are used to request an update for a given prefix. Since they are not related to a specific next hop, there is no semantic difference between IPv4 and v4-via-v6 requests. Therefore, a node **SHOULD NOT** send requests of either kind with the AE field being set to 4 (v4-via-v6); instead, it **SHOULD** request IPv4 updates by sending requests with the AE field being set to 1 (IPv4).

When receiving requests, AEs 1 (IPv4) and 4 (v4-via-v6) **MUST** be treated in the same manner: the receiver processes the request as described in Section 3.8 of [RFC8966]. If an Update is sent, then it **MAY** be an ordinary IPv4 announcement (AE = 1) or a v4-via-v6 announcement (AE = 4), as described in Section 2.1, irrespective of which AE was used in the request.

When receiving a request with AE 0 (wildcard), the receiver **SHOULD** send a full route dump, as described in Section 3.8.1.1 of [RFC8966]. Any IPv4 routes contained in the route dump may use either AE 1 (IPv4) or AE 4 (v4-via-v6), as described Section 2.1.

2.4. Other TLVs

The only other TLVs defined by [RFC8966] that carry an AE field are Next Hop and IHU. Next Hop and IHU TLVs **MUST NOT** carry the AE 4 (v4-via-v6).

3. ICMPv4 and PMTU Discovery

The Internet Control Message Protocol (ICMPv4, or simply ICMP) [RFC0792] is a protocol related to IPv4 that is primarily used to carry diagnostic and debugging information. ICMPv4 packets may be originated by end hosts (e.g., the "destination unreachable, port unreachable" ICMPv4 packet), but they may also be originated by intermediate routers (e.g., most other kinds of "destination unreachable" packets).

Some protocols deployed in the Internet rely on ICMPv4 packets sent by intermediate routers. Most notably, Path MTU Discovery (PMTUD) [RFC1191] is an algorithm executed by end hosts to discover the maximum packet size that a route is able to carry. While there exist variants of PMTUD that are purely end-to-end [RFC4821], the variant most commonly deployed in the Internet has a hard dependency on ICMPv4 packets originated by intermediate routers: if intermediate routers are unable to send ICMPv4 packets, PMTUD may lead to persistent blackholing of IPv4 traffic.

Due to this kind of dependency, every Babel router that is able to forward IPv4 traffic **MUST** be able originate ICMPv4 traffic. Since the extension described in this document enables routers to forward IPv4 traffic received over an interface that has not been assigned an IPv4 address, a router implementing this extension **MUST** be able to originate ICMPv4 packets even when the outgoing interface has not been assigned an IPv4 address.

In such a situation, if a Babel router has an interface that has been assigned an IPv4 address (other than a loopback address) or if an IPv4 address has been assigned to the router itself (to the "loopback interface"), then that IPv4 address may be used as the source of originated ICMPv4 packets. If no IPv4 address is available, a Babel router could use the experimental mechanism described in Requirement R-22 of Section 4.8 of [RFC7600], which consists of using the dummy address 192.0.0.8 as the source address of originated ICMPv4 packets. Note, however, that using the same address on multiple routers may hamper debugging and fault isolation, e.g., when using the "traceroute" utility.

4. Protocol Encoding

This extension defines the v4-via-v6 AE, whose value is 4. This AE is solely used to tag network prefixes and **MUST NOT** be used to tag neighbour addresses, e.g., in Next Hop or IHU TLVs.

This extension defines no new TLVs or sub-TLVs.

4.1. Prefix Encoding

Network prefixes tagged with AE 4 (v4-via-v6) **MUST** be encoded and decoded just like prefixes tagged with AE 1 (IPv4), as described in Section 4.1.5 of [RFC8966].

A new compression state for AE 4 (v4-via-v6) distinct from that of AE 1 (IPv4) is introduced and **MUST** be used for address compression of prefixes tagged with AE 4, as described in Sections 4.5 and 4.6.9 of [RFC8966]

4.2. Changes to Existing TLVs

The following TLVs **MAY** be tagged with AE 4 (v4-via-v6):

- Update (Type = 8)
- Route Request (Type = 9)
- Seqno Request (Type = 10)

As AE 4 (v4-via-v6) is suitable only for network prefixes, IHU (Type = 5) and Next Hop (Type = 7) TLVs are never sent with AE 4. Such (incorrect) TLVs **MUST** be ignored upon reception.

4.2.1. Update

An Update (Type = 8) TLV with AE 4 (v4-via-v6) is constructed as described in Section 4.6.9 of [RFC8966] for AE 1 (IPv4), with the following specificities:

- The Prefix field is constructed according to Section 4.1.
- The Next Hop field is built and parsed as described in Sections 2.1 and 2.2.

4.2.2. Requests

When tagged with the AE 4 (v4-via-v6), Route Request and Seqno Request TLVs **MUST** be constructed and decoded as described in Section 4.6 of [RFC8966], and the network prefixes contained within them MUST be decoded as described in Section 4.1 (see also Section 2.3).

5. Backwards Compatibility

This protocol extension adds no new TLVs or sub-TLVs.

This protocol extension uses a new AE. As discussed in Appendix D of [RFC8966] and specified in the same document, implementations that do not understand the present extension will silently ignore the various TLVs that use this new AE. As a result, incompatible versions will ignore v4-via-v6 routes. They will also ignore requests with AE 4 (v4-via-v6), which, as stated in Section 2.3, are not recommended.

Using a new AE introduces a new compression state, which is used to parse the network prefixes. As this compression state is separate from the states of other AEs, it will not interfere with the compression state of unextended nodes.

This extension reuses the next-hop state from AEs 2 and 3 (IPv6) but makes no changes to the way in which it is updated. Therefore, it causes no compatibility issues.

As mentioned in Section 2.1, ordinary IPv4 announcements are preferred to v4-via-v6 announcements when the outgoing interface has an assigned IPv4 address; doing otherwise would prevent routers that do not implement this extension from learning the route being announced.

6. IANA Considerations

IANA has allocated value 4 in the "Babel Address Encodings" registry as follows:

AE	Name	Reference
4	v4-via-v6	RFC 9229

Table 1

7. Security Considerations

The extension defined in this document does not fundamentally change the security properties of the Babel protocol. However, by allowing IPv4 routes to be propagated across routers that have not been assigned IPv4 addresses, it might invalidate the assumptions made by network administrators, which could conceivably lead to security issues.

For example, if an island of IPv4-only hosts is separated from the IPv4 Internet by routers that have not been assigned IPv4 addresses, a network administrator might reasonably assume that the IPv4-only hosts are unreachable from the IPv4 Internet. This assumption is broken if the intermediary routers implement the extension described in this document, which might expose the IPv4-only hosts to traffic from the IPv4 Internet. If this is undesirable, the flow of IPv4 traffic must be restricted by the use of suitable filtering rules (see Appendix C of [RFC8966]) together with matching packet filters in the data plane.

8. References

8.1. Normative References

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, DOI 10.17487/RFC0792, September 1981, <https://www.rfc-editor.org/info/rfc792>.

[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/info/rfc2119>.

[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/info/rfc8174>.

[RFC8966] Chroboczek, J. and D. Schinazi, "The Babel Routing Protocol", RFC 8966, DOI 10.17487/RFC8966, January 2021, <https://www.rfc-editor.org/info/rfc8966>.

8.2. Informative References

[RFC0826] Plummer, D., "An Ethernet Address Resolution Protocol: Or Converting Network Protocol Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware", STD 37, RFC 826, DOI 10.17487/RFC0826, November 1982, <https://www.rfc-editor.org/info/rfc826>.

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, DOI 10.17487/RFC1191, November 1990, <https://www.rfc-editor.org/info/rfc1191>.

[RFC4821] Mathis, M. and J. Heffner, "Packetization Layer Path MTU Discovery", RFC 4821, DOI 10.17487/RFC4821, March 2007, <https://www.rfc-editor.org/info/rfc4821>.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, DOI 10.17487/RFC4861, September 2007, <https://www.rfc-editor.org/info/rfc4861>.

[RFC5549] Le Faucheur, F. and E. Rosen, "Advertising IPv4 Network Layer Reachability Information with an IPv6 Next Hop", RFC 5549, DOI 10.17487/RFC5549, May 2009, <https://www.rfc-editor.org/info/rfc5549>.

[RFC7404] Behringer, M. and E. Vyncke, "Using Only Link-Local Addressing inside an IPv6 Network", RFC 7404, DOI 10.17487/RFC7404, November 2014, <https://www.rfc-editor.org/info/rfc7404>.

[RFC7600] Despres, R., Jiang, S., Ed., Penno, R., Lee, Y., Chen, G., and M. Chen, "IPv4 Residual Deployment via IPv6 - A Stateless Solution (4rd)", RFC 7600, DOI 10.17487/RFC7600, July 2015, <https://www.rfc-editor.org/info/rfc7600>.

Acknowledgments

This protocol extension was originally designed, described, and implemented in collaboration with Theophile Bastian. Margaret Cullen pointed out the issues with ICMP and helped coin the phrase "v4-via-v6". The author is also indebted to Donald Eastlake, Toke Høiland-Jørgensen, David Schinazi, and Donald Sharp.

Author's Address

Juliusz Chroboczek
IRIF, University of Paris
Case 7014
75205 Paris Cedex 13
France
Email: jch@irif.fr