6man R. Bonica
Internet-Draft Juniper Networks
Intended status: Standards Track Y. Kamite
Expires: September 6, 2020 NTT Communications Corporation
T. Niwa
KDDI
A. Alston
Liquid Telecom
L. Jalil
Verizon
March 5, 2020

The IPv6 Compressed Routing Header (CRH)
draft-bonica-6man-comp-rtg-hdr-13

Abstract

This document defines two new Routing header types. Collectively, they are called the Compressed Routing Headers (CRH). Individually, they are called CRH-16 and CRH-32.

Table of Contents

1. Introduction

IPv6 [RFC8200] source nodes use Routing headers to specify the path that a packet takes to its destination. The IETF has defined several Routing header types [IANA-RH]. RH0 [RFC2460] was the first to be defined and was deprecated [RFC5095] because of security vulnerabilities.

This document defines two new Routing header types. Collectively, they are called the Compressed Routing Headers (CRH). Individually, they are called CRH-16 and CRH-32.

The CRH, like RH0, allows IPv6 source nodes to specify the path that a packet takes to its destination. The CRH differs from RH0 because:

- o It can be encoded in fewer bytes than RH0.
- o It addresses the security vulnerabilities that affected RH0.

The following are reasons for encoding the CRH in as few bytes as possible:

- o Many ASIC-based forwarders copy all headers from buffer memory to on-chip memory. As header sizes increase, so does the cost of this copy.
- o Because Path MTU Discovery (PMTUD) [RFC8201] is not entirely reliable, many IPv6 hosts refrain from sending packets larger than the IPv6 minimum link MTU (i.e., 1280 bytes). When packets are small, the overhead imposed by large Routing Headers is excessive.

Section 10 of this document addresses security considerations. Appendix A of this document demonstrates how the CRH can be encoded in fewer bytes than RH0.

2. Requirements Language

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. The Compressed Routing Headers (CRH)

Both CRH versions (i.e., CRH-16 and CRH-32) contain the following fields:

- o Next Header - Defined in [RFC8200].
- o Hdr Ext Len - Defined in [RFC8200].
- o Routing Type - Defined in [RFC8200]. Value TBD by IANA. (For CRH-16, the suggested value is 5. For CRH-32, the suggested value is 6.)
- o Segments Left - Defined in [RFC8200].
- o Type-specific Data - Described in [RFC8200].

In the CRH, the Type-specific data field contains a list of Segment Identifiers (SIDs). Each SID represents both of the following:

o A segment of the path that the packet takes to its destination.

o An entry in the CRH Forwarding Information Base (CRH-FIB) (Section 4).

SIDs are listed in reverse order. So, the first SID in the list represents the final segment in the path. Because segments are listed in reverse order, the Segments Left field can be used as an index into the SID list. In this document, the "current SID" is the SID list entry referenced by the Segments Left field.

The first segment in the path can be omitted from the list. See (Appendix B) for examples.

In the CRH-16 (Figure 1), each SID is encoded in 16-bits. In the CRH-32 (Figure 2), each SID is encoded in 32-bits.

In all cases, the CRH MUST end on a 64-bit boundary. So, the Type-specific data field MUST be padded with zeros if the CRH would otherwise not end on a 64-bit boundary.

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  Next Header  |  Hdr Ext Len  | Routing Type  | Segments Left |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |             SID[0]            |          SID[1]               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-|
    |                          .........
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
```

Figure 1: CRH-16

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  Next Header  |  Hdr Ext Len  | Routing Type  | Segments Left |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    +                             SID[0]                            +
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    +                             SID[1]                            +
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    //                                                              //
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    +                             SID[n]                            +
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: CRH-32

4. The CRH Forwarding Information Base (CRH-FIB)

Each SID identifies a CRH-FIB entry.

Each CRH-FIB entry contains:

- o A IPv6 address.
- o A forwarding method.
- o Method-specific parameters (optional).

The IPv6 address represents an interface on the next segment endpoint. It MUST NOT be a link-local address. While the IPv6 address represents an interface on the next segment endpoint, it does not necessarily represent the interface through which the packet will arrive at the next segment endpoint.

The forwarding method specifies how the processing node will forward the packet to the next segment endpoint. The following are examples:

- o Forward the packet to the next-hop along the least-cost path to the next segment endpoint.
- o Forward the packet through a specified interface to the next segment endpoint.

Some forwarding methods require method-specific parameters. For example, a forwarding method might require a parameter that identifies the interface through which the packet should be forwarded.

The CRH-FIB can be populated:

- o By an operator, using a Command Line Interface (CLI).
- o By a controller, using the Path Computation Element (PCE) Communication Protocol (PCEP) [RFC5440] or the Network Configuration Protocol (NETCONF) [RFC6241].
- o By a distributed routing protocol [ISO10589-Second-Edition], [RFC5340], [RFC4271].

5. Processing Rules

The following rules apply to packets that contain a CRH:

- o If the IPv6 Source Address is a link-local address, discard the packet.
- o If the IPv6 Source Address is a multicast address, discard the packet.
- o If the IPv6 Destination Address is a link-local address, discard the packet.
- o If the IPv6 Hop Limit is less than or equal to 1, discard the packet and send an ICMPv6 Time Exceeded message to the Source Address.

The following rules describe CRH processing:

- o If Segments Left equals 0, skip over the CRH and process the next header in the packet.
- o If Hdr Ext Len indicates that the CRH is larger than the implementation can process, discard the packet and send an ICMPv6 Parameter Problem, Code 0, message to the Source Address, pointing to the Hdr Ext Len field.
- o Compute L, the minimum CRH length (See (Section 5.1)).
- o If L is greater than Hdr Ext Len, discard the packet and send an ICMPv6 Parameter Problem, Code 0, message to the Source Address, pointing to the Segments Left field.
- o Decrement Segments Left.

- Search for the current SID in the CRH-FIB. In this document, the "current SID" is the SID list entry referenced by the Segments Left field.

- If the search does not return a CRH-FIB entry, discard the packet and send an ICMPv6 Parameter Problem, Code 0, message to the Source Address, pointing to the current SID.

- If the CRH-FIB entry contains a link-local address, discard the packet and send an ICMPv6 Parameter Problem, Code 0, message to the Source Address, pointing to the current SID.

- If Segments Left is greater than 0 and the CRH-FIB entry contains a multicast address, discard the packet and send an ICMPv6 Parameter Problem, Code 0, message to the Source Address, pointing to the current SID.

- Copy the IPv6 address from the CRH-FIB entry to the Destination Address field in the IPv6 header.

- Decrement the IPv6 Hop Limit.

- Resubmit the packet to the IPv6 module for transmission to the new destination, ensuring that it executes the forwarding method specified by the CRH-FIB entry.

5.1. Computing Minimum CRH Length

The algorithm described in this section accepts the following CRH fields as its input parameters:

- Routing Type (i.e., CRH-16 or CRH-32).

- Segments Left.

It yields L, the minimum CRH length. The minimum CRH length is measured in 8-octet units, not including the first 8 octets.

```
<CODE BEGINS>

switch(Routing Type) {
    case CRH-16:
        if (Segments Left <= 2)
            return(0)
        sidsBeyondFirstWord = Segments Left - 2;
        sidPerWord = 4;
    case CRH-32:
        if (Segments Left <= 1)
            return(0)
        sidsBeyondFirstWord = Segments Left - 1;
        sidsPerWord = 2;
    case default:
        return(0xFF);
    }

words = sidsBeyondFirstWord div sidsPerWord;
if (sidsBeyondFirstWord mod sidsPerWord)
    words++;

return(words)


<CODE ENDS>
```

6. Mutability

In the CRH, the Segments Left field is mutable. All remaining fields are immutable.

7. Applications And SIDs

A CRH contains one or more SIDs. Each SID is processed by exactly one node.

Therefore, a SID is not required to have domain-wide significance. Applications can:

- o Allocate SIDs so that they have domain-wide significance.

- o Allocate SIDs so that they have node-local significance.

8. Management Considerations

PING and TRACEROUTE [RFC2151] both operate correctly in the presence of the CRH.

9. ICMPv6 Considerations

Implementations MUST comply with the ICMPv6 processing rules specified in Section 2.4 of [RFC4443]. For example:

- An implementation MUST NOT originate an ICMPv6 error message in response to another ICMPv6 error message.

- An implementation MUST rate limit the ICMPv6 messages that it originates.

10. Security Considerations

Networks that process the CRH MUST mitigate the security vulnerabilities described in [RFC5095]. Their border routers SHOULD discard packets that satisfy the following criteria:

- The packet contains a CRH

- The Segments Left field in the CRH has a value greater than 0

- The Destination Address field in the IPv6 header represents an interface that resides inside of the network.

Many border routers cannot filter packets based upon the Segments Left value. These border routers MAY discard packets that satisfy the following criteria:

- The packet contains a CRH

- The Destination Address field in the IPv6 header represents an interface that resides inside of the network.

11. Implementation and Deployment Status

Juniper Networks has produced experimental implementations of the CRH on:

- A LINUX-based software platform

- The MX-series (ASIC-based) router

Liquid Telecom has deployed the CRH, on a limited basis, in their network. Other experimental deployments are in progress.

12. IANA Considerations

IANA is requested to make the following entries in the Internet Protocol Version 6 (IPv6) Parameters "Routing Type" registry [IANA-RH]:

Suggested Value	Description	Reference
5	Compressed Routing Header (16-bit) (CRH-16)	This document
6	Compressed Routing Header (32-bit) (CRH-32)	This document

13. Acknowledgements

Thanks to Dr. Vanessa Ameen, Fernando Gont, Naveen Kottapalli, Joel Halpern, Tony Li, Gerald Schmidt, Nancy Shaw, and Chandra Venkatraman for their contributions to this document.

14. Contributors

Daniam Henriques

Liquid Telecom

Johannesburg, South Africa

Email: daniam.henriques@liquidtelecom.com

Gang Chen

Baidu

No.10 Xibeiwang East Road Haidian District

Beijing 100193 P.R. China

Email: phdgang@gmail.com

Yifeng Zhou

ByteDance

Building 1, AVIC Plaza, 43 N 3rd Ring W Rd Haidian District

Beijing 100000 P.R. China

Email: yifeng.zhou@bytedance.com

15. References

15.1. Normative References

[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>.

[RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", STD 89, RFC 4443, DOI 10.17487/RFC4443, March 2006, <https://www.rfc-editor.org/info/rfc4443>.

[RFC5095] Abley, J., Savola, P., and G. Neville-Neil, "Deprecation of Type 0 Routing Headers in IPv6", RFC 5095, DOI 10.17487/RFC5095, December 2007, <https://www.rfc-editor.org/info/rfc5095>.

[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>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[RFC8201] McCann, J., Deering, S., Mogul, J., and R. Hinden, Ed., "Path MTU Discovery for IP version 6", STD 87, RFC 8201, DOI 10.17487/RFC8201, July 2017, <https://www.rfc-editor.org/info/rfc8201>.

15.2. Informative References

[IANA-RH] IANA, "Routing Headers", <https://www.iana.org/assignments/ipv6-parameters/ipv6-parameters.xhtml#ipv6-parameters-3>.

[ISO10589-Second-Edition]
International Organization for Standardization, ""Intermediate system to Intermediate system intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473)", ISO/IEC 10589:2002, Second Edition,", November 2001.

[RFC2151] Kessler, G. and S. Shepard, "A Primer On Internet and TCP/IP Tools and Utilities", FYI 30, RFC 2151, DOI 10.17487/RFC2151, June 1997, <https://www.rfc-editor.org/info/rfc2151>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <https://www.rfc-editor.org/info/rfc2460>.

[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/info/rfc4271>.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, DOI 10.17487/RFC5340, July 2008, <https://www.rfc-editor.org/info/rfc5340>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <https://www.rfc-editor.org/info/rfc5440>.

[RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, <https://www.rfc-editor.org/info/rfc6241>.

Appendix A. CRH Processing Examples

The CRH-16 and CRH-32 encode information more efficiently than RH0.

SIDs	RH0	CRH-16	CRH-32
1	24	8	8
2	40	8	16
3	56	16	16
4	72	16	24
5	88	16	24
6	104	16	32
7	120	24	32
8	136	24	40
9	152	24	40
10	168	24	48
11	184	32	48
12	200	32	52
13	216	32	52
14	232	32	56
15	248	40	56
16	264	40	60
17	280	40	60
18	296	40	64

Table 1: Routing Header Size (in Bytes) As A Function Of Routing Header Type and Number Of SIDs

(Table 1) reflects Routing header size as a function of Routing header type and number of SIDs contained by the Routing header.

Appendix B. CRH Processing Examples

This appendix demonstrates CRH processing in the following scenarios:

- o The SID list contains one entry for each segment in the path (Appendix B.1).

- o The SID list omits the first entry in the path (Appendix B.2).

```
    -----------                 -----------                 -----------
   |Node: S    |               |Node: I1   |               |Node: I2   |
   |Loopback:  |---------------|Loopback:  |---------------|Loopback:  |
   |2001:db8::a|               |2001:db8::1|               |2001:db8::2|
    -----------                 -----------                 -----------
         |                                                        |
         |                      -----------                       |
         |                     |Node: D    |                      |
          ---------------------|Loopback:  |---------------------
                               |2001:db8::b|
                                -----------
```

Figure 3: Reference Topology

Figure 3 provides a reference topology that is used in all examples.

SID	IPv6 Address	Forwarding Method
2	2001:db8::2	Least-cost path
11	2001:db8::b	Least-cost path

Table 2: Node SIDs

Table 2 describes two entries that appear in each node's CRH-FIB.

B.1. The SID List Contains One Entry For Each Segment In The Path

In this example, Node S sends a packet to Node D, via I2. In this example, I2 appears in the CRH segment list.

```
+-------------------------------------+-------------------+
| As the packet travels from S to I2: |                   |
+-------------------------------------+-------------------+
| Source Address = 2001:db8::a        | Segments Left = 1 |
| Destination Address = 2001:db8::2   | SID[0] = 11       |
|                                     | SID[1] = 2        |
+-------------------------------------+-------------------+

+-------------------------------------+-------------------+
| As the packet travels from I2 to D: |                   |
+-------------------------------------+-------------------+
| Source Address = 2001:db8::a        | Segments Left = 0 |
| Destination Address = 2001:db8::b   | SID[0] = 11       |
|                                     | SID[1] = 2        |
+-------------------------------------+-------------------+
```

B.2. The SID List Omits The First Entry In The Path

In this example, Node S sends a packet to Node D, via I2. In this example, I2 does not appear in the CRH segment list.

```
+-------------------------------------+-------------------+
| As the packet travels from S to I2: |                   |
+-------------------------------------+-------------------+
| Source Address = 2001:db8::a        | Segments Left = 1 |
| Destination Address = 2001:db8::2   | SID[0] = 11       |
+-------------------------------------+-------------------+

+-------------------------------------+-------------------+
| As the packet travels from I2 to D: |                   |
+-------------------------------------+-------------------+
| Source Address = 2001:db8::a        | Segments Left = 0 |
| Destination Address = 2001:db8::b   | SID[0] = 11       |
+-------------------------------------+-------------------+
```

Authors' Addresses

Ron Bonica
Juniper Networks
2251 Corporate Park Drive
Herndon, Virginia 20171
USA

Email: rbonica@juniper.net

Yuji Kamite
NTT Communications Corporation
3-4-1 Shibaura, Minato-ku
Tokyo 108-8118
Japan

Email: y.kamite@ntt.com

Tomonobu Niwa
KDDI
3-22-7, Yoyogi, Shibuya-ku
Tokyo 151-0053
Japan

Email: to-niwa@kddi.com

Andrew Alston
Liquid Telecom
Nairobi
Kenya

Email: Andrew.Alston@liquidtelecom.com

Luay Jalil
Verizon
Richardson, Texas
USA

Email: luay.jalil@one.verizon.com