The IPv6 Compressed Routing Header (CRH)
draft-bonica-6man-comp-rtg-hdr-04

Abstract

This document defines the Compressed Routing Header (CRH). The CRH, like any other Routing header, contains a list of segment identifiers (SID). The CRH differs from other Routing headers in that its segment identifiers can be 8, 16 or 32 bits long.

Status of This Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on November 7, 2019.
1. Introduction

An IPv6 [RFC8200] source node can steer a packet through a specific path to its destination. The source node defines the path as an ordered list of segments and encodes the path in an IPv6 Routing header. As per [RFC8200], all Routing headers include the following fields:
o Next Header - Identifies the header immediately following the Routing header.

o Hdr Ext Len - Length of the Routing header.

o Routing Type - Identifies the particular Routing header variant.

o Segments Left - The number of segments still to be traversed before reaching the destination.

o Type-specific Data - Syntax and semantics are defined by the Routing Type.

The following Routing types are currently defined:

o Source Route (i.e., RH0) [RFC5095] (deprecated)

o Type 2 Routing Header [RFC6275]

o RPL Source Route Header [RFC6554]

o Segment Routing Header (SRH) [I-D.ietf-6man-segment-routing-header]

In each of the above-mentioned Routing Types, Type-specific Data contains a list of one or more segment identifiers (SID). Typically, a SID is an IPv6 address that identifies a segment endpoint. In the SRH, the SID may carry additional semantics.

In all cases, the SID is 128 bits long. Therefore, routing headers can be very large. For example, an 88-byte Source Route header is required to specify a path that contains six segments. The same can be said of the SRH.

Large Routing headers are undesirable for the following reasons:

o Many ASIC-based forwarders copy the entire IPv6 extension header chain from buffer memory to on-chip memory. As the size of the IPv6 extension header chain increases, 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 becomes pronounced.
This document defines the Compressed Routing Header (CRH). The CRH, like any other Routing header, contains a list of SIDs. The CRH differs from other Routing headers in that its SIDs can be 8, 16, or 32 bits long.

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 Header (CRH)

Figure 1 depicts the CRH.

```
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 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Last Entry  |Com|              Reserved                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           SID List   ........             |
+-+-+-+-+-+-+-+-+-+-+
| SID List   ........             |
+-+-+-+-+-+-+-+-+-+
```

Figure 1: Compressed Routing Header (CRH)

The CRH contains the following fields:

- **Next Header** - Defined in [RFC8200].
- **Hdr Ext Len** - Defined in [RFC8200].
- **Routing Type** - Defined in [RFC8200]. Value TBD by IANA. (Suggested value: 5)
- **Segments Left (SL)** - Defined in [RFC8200].
- **Last Entry** - 8 bits. Represents the index (zero based), in the Segment List, of the last element of the Segment List.
- **Com (Compression)** - 2 bits. Represents the length of each entry in the SID List. Values are eight bits (0), sixteen bits (1), thirty-two bits (2), and reserved (3).
o Reserved - SHOULD be set to zero by the sender. MUST be ignored by the receiver.

o SID List - A zero-indexed list of SIDs. SIDs are listed in reverse order, with SID[0] representing the packet’s ultimate destination, SID[1] representing the segment endpoint prior to the ultimate destination, and so forth. SIDs are listed in reverse order so that Segments Left can be used as an index to the SID List. See Section 5 for SID details.

Figure 2 through Figure 4 illustrate CRH encodings with Com equal to 0, 1 and 2. In all cases, the CRH MUST end on a 64-bit boundary. Therefore, the CRH MAY be padded with zeros.

```
0                   1                   2                   3
+-------------------+-------------------+-------------------+-------------------+
| Next Header | Hdr Ext Len | Routing Type | Segments Left |
| Reserved     | Reserved     | Reserved     | Reservd         |
| SID[0] | SID[1] |              |                |
| ........       | ........     |             |           |
+-------------------+-------------------+-------------------+-------------------+
```

Figure 2: Eight-bit Encoding (Com equals 0)

```
0                   1                   2                   3
+-------------------+-------------------+-------------------+-------------------+
| Next Header | Hdr Ext Len | Routing Type | Segments Left |
| Reserved     | Reserved     | Reserved     | Reservd         |
| SID[0] | SID[1] |              |                |
| ........       | ........ |             |           |
+-------------------+-------------------+-------------------+-------------------+
```

Figure 3: Sixteen-bit Encoding (Com equals 1)
Figure 4: Thirty-two bit Encoding (Com equals 2)

4. The CRH Domain

A CRH domain is a collection of IPv6 devices. On the control plane, CRH domain members exchange SID information with one another. On the forwarding plane, CRH domain members accept packets containing the CRH from one another.

Section 9 of this document requires the CRH domain to be contained within an operator’s trust domain.

5. Segment Identifiers (SID)

This document defines the following SID types:

- Loosely routed
- Strictly routed

All SIDs, regardless of type, map to exactly one IPv6 address. The mapped address identifies an interface or set of interfaces (in the case of multicast) that terminate the segment. The address MUST be one of the following:

- A globally scoped IPv6 unicast address [RFC4291].
- A Unique Local IPv6 Unicast Address (ULA) [RFC4193].
- A Multicast address [RFC4291].
A strictly routed SID also maps to a link interface. Nodes send packets through that interface in order to access the segment endpoint.

Loosely routed SIDs have domain-wide significance. This means that within a CRH domain, a loosely routed SID MUST map to exactly one IPv6 address. By contrast, strictly routed SIDs have node-local significance. This means that within a CRH domain, one node can map a strictly routed SID to one address while another node maps the same strictly routed SID to a different address. See Appendix A for an example.

Forwarding nodes can learn the above-mentioned mappings from a central controller, from a distributed routing protocol or using any other means. The mechanisms that forwarding nodes use to learn the above-mentioned mappings are beyond the scope of this document.

6. Processing Rules

6.1. General

[RFC8200] defines rules that apply to IPv6 extension headers, in general, and IPv6 Routing headers, in particular. All of these rules apply to the CRH.

For example:

- Extension headers (except for the Hop-by-Hop Options header) are not processed, inserted, or deleted by any node along a packet’s delivery path, until the packet reaches the node (or each of the set of nodes, in the case of multicast) identified in the Destination Address field of the IPv6 header.

- If, while processing a received packet, a node encounters a Routing header with an unrecognized Routing Type value, the required behavior of the node depends on the value of the Segments Left field. If Segments Left is zero, the node must ignore the Routing header and proceed to process the next header in the packet, whose type is identified by the Next Header field in the Routing header. If Segments Left is non-zero, the node must discard the packet and send an ICMP [RFC4443] Parameter Problem, Code 0, message to the packet’s Source Address, pointing to the unrecognized Routing Type.

- If, after processing a Routing header of a received packet, an intermediate node determines that the packet is to be forwarded onto a link whose link MTU is less than the size of the packet,
the node must discard the packet and send an ICMP Packet Too Big message to the packet’s Source Address.

6.2. CRH Specific

When a node recognizes and processes a CRH, it executes the following procedure:

- If the IPv6 Source Address is a link-local address, discard the packet.
- If the IPv6 Source Address is a multicast address, discard the packet.
- If Segments Left equal 0, skip over the CRH and process the next header in the packet.
- If Segments Left is greater than Last Entry plus one, send an ICMP Parameter Problem, Code 0, message to the Source Address, pointing to the Segments Left field, and discard the packet.
- If Com is equal to (3) Reserved, send an ICMP Parameter Problem, Code 0, message to the Source Address, pointing to the Com field, and discard the packet.
- If the IPv6 Hop Limit is less than or equal to 1, send an ICMP Time Exceeded -- Hop Limit Exceeded in Transit message to the Source Address and discard the packet.
- Compute L, the minimum CRH length (See Section 6.2.1)
- If L is greater than Hdr Ext Len, send an ICMP Parameter Problem, Code 0, message to the Source Address, pointing to the Last Entry field, and discard the packet.
- Decrement Segments Left (i.e., SL).
- Search for SID[SL] in the table that maps SID’s to IPv6 addresses and interfaces. If SID[SL] cannot be found in that table, send an ICMP Parameter Problem, Code 0, message to the Source Address, pointing to SID[SL], and discard the packet.
- Copy the address associated with SID[SL] to the IPv6 Destination Address.
- If the IPv6 Destination address is a multicast address and SL is greater than zero, send an ICMP Parameter Problem, Code 0, message
to the Source Address, pointing to Segment List \([SL]\), and discard the packet.

- Decrement the IPv6 Hop Limit.

- If SID\([SL]\) is a loosely routed segment, resubmit the packet to the IPv6 module for transmission to the new destination.

- If SID\([SL]\) is a strictly routed segment, forward the packet through the interface that is associated with SID\([SL]\).

The above stated rules are demonstrated in Appendix A.

6.2.1. Computing Minimum CRH Length

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

- Compression (Com).

- Last Entry.

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>

if (Com == 0) { /* Eight bit encoding */
    L = ( ( Last Entry + 1 ) / 8 );
    if ( ( Last Entry + 1 ) % 8 )
        L++;
}
elsif (Com == 1) { /* Sixteen bit encoding */
    L = ( ( Last Entry + 1 ) / 4 );
    if ( ( Last Entry + 1 ) % 4 )
        L++;
}
elsif (Com == 2) { /* Thirty-two bit encoding */
    L = ( ( Last Entry + 1 ) / 2 );
    if ( ( Last Entry + 1 ) % 2 )
        L++;
}
else { /* Invalid Com */
    L = 0xFF
}
return(L)

<CODE ENDS>

7. Mutability

The Segments Left field is mutable and MAY be decremented by processing nodes. All remaining fields are immutable.

8. Management Considerations

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

9. Security Considerations

The CRH can be used within trusted domains only. In order to enforce this requirement, domain edge routers MUST do one of the following:

- Discard all inbound packets that contain a CRH
- Authenticate [RFC4302] [RFC4303] all inbound packets that contain a CRH
10. IANA Considerations

This document makes the following registration in the Internet Protocol Version 6 (IPv6) Parameters "Routing Type" registry maintained by IANA:

<table>
<thead>
<tr>
<th>Suggested Value</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Compressed Routing Header (CRH)</td>
<td>This document</td>
</tr>
</tbody>
</table>

11. Acknowledgements

Thanks to Joel Halpern, Tony Li, Gerald Schmidt, Nancy Shaw and Chandra Venkatraman for their comments.

12. References

12.1. Normative References

12.2. Informative References

[I-D.ietf-6man-segment-routing-header]


Appendix A. CRH Processing Examples

This appendix provides examples of CRH processing in the following applications:

- Loose source routing (Appendix A.1)
o Loose source routing preserving the first SID (Appendix A.2)

o Strict source routing (Appendix A.3)

```
+--------------------+-----+--------------+
| Instantiating Node | SID | IPv6 Address |
+--------------------+-----+--------------+
|        All         | 1   | 2001:db8::1 |
|        All         | 2   | 2001:db8::2 |
|        All         | 3   | 2001:db8::3 |
|        All         | 10  | 2001:db8::a |
|        All         | 11  | 2001:db8::b |
+--------------------+-----+--------------+
```

Table 1: Loosely Routed SIDs

Table 1 provides mappings for loosely routed SIDs. These mappings are instantiated on all nodes in the reference topology.

Figure 5: Reference Topology

Figure 5 provides a reference topology that is used in all examples.
Table 2: Strictly Routed SIDs

Table 2 provides mappings for strictly routed SIDs. These mappings are available on the instantiating node only.

A.1. Loose Source Routing

In this example, Node S sends a packet to Node D, specifying loose source route through Node I3. In this example, the first node in the path, I3, does not appear in the CRH segment list. Therefore, the destination node may not be able to send return traffic through the same path.

A.2. Loose Source Routing Preserving The First SID

In this example, Node S sends a packet to Node D, specifying loose source route through Node I3. In this example, the first node in the path, I3, appears in the CRH segment list. Therefore, the destination node can send return traffic through the same path.
A.3. Strict Source Routing

In this example, Node S sends a packet to Node D, specifying the strict source route through I1 and I3.
As the packet travels from I3 to D:

<table>
<thead>
<tr>
<th>Source Address = 2001:db8::a</th>
<th>Last Entry = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Destination Address = 2001:db8:0:b::2</td>
<td>Segments Left = 0</td>
</tr>
<tr>
<td></td>
<td>SID[0] = 129</td>
</tr>
<tr>
<td></td>
<td>SID[1] = 129</td>
</tr>
</tbody>
</table>

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
Daniam Henriques  
Liquid Telecom  
Johannesburg  
South Africa  

Email: daniam.henriques@liquidtelecom.com

Ning So  
Reliance Jio  
3010 Gaylord PKWY, Suite 150  
Frisco, Texas 75034  
USA  

Email: Ning.So@ril.com

Fengman Xu  
Reliance Jio  
3010 Gaylord PKWY, Suite 150  
Frisco, Texas 75034  
USA  

Email: Fengman.Xu@ril.com

Gang Chen  
Baidu  
No.10 Xibeiwang East Road Haidian District  
Beijing 100193  
P.R. China  

Email: phdgang@gmail.com

Yongqing Zhu  
China Telecom  
109 West Zhongshan Ave, Tianhe District  
Guangzhou  
P.R. China  

Email: zhuyq.gd@chinatelecom.cn
Guangming Yang  
China Telecom  
109 West Zhongshan Ave, Tianhe District  
Guangzhou  
P.R. China  

Email: yanggm.gd@chinatelecom.cn

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