Network Shaofu. Peng
Internet-Draft ZTE Corporation
Intended status: Standards Track Ron. Bonica
Expires: 2 September 2024 Juniper Networks
1 March 2024

Deterministic Routing Header
draft-pb-6man-deterministic-crh-00

Abstract

This document introduces a new IPv6 Routing Header used for deterministic forwarding wihch is generally a strict explicit path. This Routing Header will contain the decoupled topology instructions and deterministic forwarding resource indications. The target is low cost of encapasultion and less amount of allocated SIDs.

Table of Contents

1. Introduction 2
1.1. Requirements Language 3
2. The Compressed Routing Headers for DetNet (CRH-20) 3
3. The CRH Forwarding Information Base (CRH-FIB) 6
4. Processing Rules 6
4.1. Processing on the Headend Node 6
4.2. Processing on the Transit or Endpoint Node 7
5. IANA Considerations 8
6. Security Considerations 8
7. Acknowledgements 8
8. Normative References 8
Authors' Addresses 10

1. Introduction

[RFC8655] describes the architecture of deterministic network and defines the QoS goals of deterministic forwarding: Minimum and maximum end-to-end latency from source to destination, timely delivery, and bounded jitter (packet delay variation); packet loss ratio under various assumptions as to the operational states of the nodes and links; an upper bound on out-of-order packet delivery. In order to achieve these goals, deterministic networks use resource reservation, explicit routing, service protection and other means. In general, a deterministic path is generally a strictly explicit path calculated by a centralized controller, and resources are reserved on the nodes along the path. There is such a need that metadata related to each hop used to provide QoS needs to be carried in the packet to avoid network core maintenance flow states and meet large scaling deterministic requirements.

[RFC8200] defines the Routing Header. The source node of the IPv6 packet can include some segment information in the Routing Header to control the packet's access to these segments before reaching the final destination node. How to reduce the cost of Routing Header, especially in scenarios with strict explicit deterministic forwarding paths, is a concern. Generally, there are two categories of optimization methods. One is to use short indexes to map to IPv6 addresses, and the other is to rely on a common prefix for all segments. For example, [I-D.ietf-6man-comp-rtg-hdr] defines two new routing headers called CRH-16 and CRH-32, containing a short index for mapping to 128-bit IPv6 addresses, and retrieving all forwarding information from the CRH-FIB entries matched by the index. And, [RFC6554] defines Routing Type 3, where only the different parts of each segment need to be stored in the segment list, and the common prefix of all elements is stored in the Destination Address field of the IPv6 header. [I-D.ietf-spring-srv6-srh-compression] also describes another common prefixes based compression method suitable for SRv6 network.

This document introduces a CRH varaint, called CRH-20, to meet the requirement of deterministic forwarding. It contains the decoupled topology instructions and deterministic forwarding resource indications. Here, the decoupled means that the SIDs for topology instructions and resource indications for DetNet QoS are defined, signaled in control plane, and forwarded in data plane, independently. Considering that there are many types of resources and the huge amount of resource instances supported by each type, this document does not recommend distinguishing different resource instances by allocating different SIDs. The target of CRH-20 is low cost of encapasultion and less amount of SIDs.

1.1. 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.

2. The Compressed Routing Headers for DetNet (CRH-20)

```
 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 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| ST  | RT  |P| |                   Common RI                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             SID[0]                    |     Individual RI     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             SID[1]                    |     Individual RI     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                            ... ...                          //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             SID[n-1]                  |     Individual RI     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Padding                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: CRH-20

CRH-20 contains the following fields:

* Next Header: Defined in [RFC8200].

* Hdr Ext Len: Defined in [RFC8200].

* Routing Type: Defined in [RFC8200] (TBA for CRH-20)

* Segments Left: Defined in [RFC8200].

* Type-specific Data: Described in [RFC8200].

In the CRH-20, the Type-specific data field contains topology instructions and forwarding resource indications. In detailed:

* ST (SID Type): 3 bits for the index type.

  - 0: the SID field is 20-bit index that is the default type and unrelated to any well-knownn index type. Implementers need to additionally define the announcement, learning, and FIB creation for this default index type.

  - 1: the SID field is 20-bit MPLS label. In this case, CRH-FIB reuses MPLS ILM (Incoming Label Map) table.

  - 2: the SID field is 20-bit SR-MPLS SID index. In this case, CRH-FIB resuses SR-MPLS SID index table.

- 3: the SID field is 20-bit BIER ID. In this case, CRH-FIB resuses BIFT table.

- Other types may be defined in the future.

* RT (Resource Type): 3 bits for the forwarding resource types.

  - 0: undefined. In this case, Common RI field and Individual RI fields may also be 0 and ignored during forwarding procedure of the packet, or other values used for user defined purposes.

  - 1: Indicates that the forwarding resources is timeslot resources. [I-D.peng-detnet-packet-timeslot-mechanism] defines the timeslot queueing and forwarding (TQF) mechanism in IP/MPLS network. Timeslot resources are defined as multiple equally time slots contained within a fixed length of periodic time (known as orchestration period). Different nodes interoperate on the same orchestration period, but for the same orchestration period, different nodes may configure different timeslot lengths. In this case, Common RI field contains orchestration period length (OPL) information, and Individual RI field contains the specific timeslot for the corresponding segment.

  - 2: Indicates that the forwarding resources is delay resources. [I-D.peng-detnet-deadline-based-forwarding] defines the deadline based forwarding mechanism in IP/MPLS network. The network may provide multiple delay levels each provided a bounded latency according to the schedulability condition of EDF (earliest deadline first) scheduling. In this case, Common RI field contains the latency deviation (E) information, and Individual RI field contains the planned residence time (D) for the corresponding segment. All segments may have the same planned residence time (D) by evenly dividing the total residence time budget, but this is not mandatory.

  - 3: Indicate that forwarding resources is network slice resources, also known as Network Resource Partition (NRP) resources. [I-D.ietf-teas-ns-ip-mpls] defines a network slicing mechanism in IP/MPLS network. The physical network may be partitioned into multiple slices dedicated to specific services or customers. It is possible to directly reuse network slices to get deterministic QoS. In this case, Common RI field contains the default NRP-ID, and Individual RI field contains the overrided NRP-ID when necessary.

  - Other types may be defined in the future.

* P (Pad) flag: 1 bit to indicate whether there are padding field. If P flag is 0, there are no padding. If P flag is 1, there are padding with 4 bytes.

* Common RI (Resource Indication): 24 bits for the common forwarding resource information that all segments shared. The information contained in Common RI depends on RT. Please see the above explanation of Resource Type field.

* SID (Segment Identifier): 20 bits for the topology instruction. Each SID identifies an entry in the CRH-FIB. Each CRH-FIB entry identifies an interface on the path that the packet takes to its destination.

* Individual RI (Resource Indication): 12 bits for the forwarding resource that is specific for the corresponding segment. The information contained in Individual RI depends on RT. Please see the above explanation of Resource Type field.

* Padding: 0 or 4 bytes to ensure 8-octet alignment.

In CRH-20, each segment is constructed by 32-bit tuple <SID, Individual RI>. Segments are listed in reverse order. So, the first segment in the list represents the final interface 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 Segment" is the Segment list entry referenced by the Segments Left field.

The first segment in the path can be omitted from the list.

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

Please refer to [I-D.ietf-6man-comp-rtg-hdr]. There are no additional considerations for the topology instructions, except that FIB may be organized by specific type of index.

4. Processing Rules

4.1. Processing on the Headend Node

The headend node is responsible for encapsulating the CRH-20 for the packet. For a logical Segment List <S1, S2, ..., Sn>, set the SID and RI information corresponding to each segment in CRH-20.

According to the CRH-FIB entry matched with the SID in the segment S1, the headend obtain the corresponding destination IPv6 address, and copy it to the DA field of the IPv6 header. In addition, set Segment Left to n-1, indicating that there are still n-1 segment elements to be processed in the segment list.

In order to further save the cost of CRH-20, because the destination IPv6 address corresponding to the first segment has already been copied to the DA field of the IPv6 header, the headend node may exclude the first segment from the segment list of CRH-20. If this is the case, only n-1 segments (from [0] to [n-2]) need to be included in the list, and the segment [n-2] field stores the segment S2. The Segment Left is still set to n-1, indicating that there are still n-1 segment elements to be processed in the segment list. The headend node must ensure that when forwarding packets to the first segment S1, it accesses the specified forwarding resource (although it is also not included in the list). Note that this attempting to reduce overhead may not be meaningful due to the need for 8-octet alignment.

4.2. Processing on the Transit or Endpoint Node

When a transit or endpoint node receives an IPv6 packet, if the DA in the IPv6 header matches the local IP address, and the Next Header field of the IPv6 header indicates that the next layer header is CRH-20, then continue to process CRH-20 as below:

```
S01. When a CRH-20 is processed {
S02.   if (Segments Left == 0) {
S03.      Stop processing the CRH-20, and proceed to process the
            next header in the packet, whose type is identified by
            the Next Header field in the routing header.
S04.   }
S05.   If (IPv6 Hop Limit <= 1) {
S06.      Send an ICMP Time Exceeded message to the Source Address
            with Code 0 (Hop limit exceeded in transit), interrupt
            packet processing, and discard the packet.
S07.   }
S08.   Decrement IPv6 Hop Limit by 1
S09.   Decrement Segments Left by 1
S10.   Read the next 32-bit segment by Segment List[Segments Left].
       Match CRH-FIB entry by ST field and segment.SID field.
S11.   If there is no matched FIB entry, discard the packet and send
         an ICMPv6 Parameter Problem, Code 0, message to the Source
         Address, pointing to the current segment.
S12.   Get the destinatoin IPv6 address and output interface from
         the matched FIB entry, and copy the destinatoin IPv6
         address to the DA field in the IPv6 header.
S13.   Transmit the packet to the output interface, and access the
         forwarding resource indicated by Common RI field and
         Individual RI field.
S14. }
```

5. IANA Considerations

TBD

6. Security Considerations

TBD

7. Acknowledgements

TBD

8. Normative References

[I-D.ietf-6man-comp-rtg-hdr]
Bonica, R., Kamite, Y., Alston, A., Henriques, D., and L. Jalil, "The IPv6 Compact Routing Header (CRH)", Work in Progress, Internet-Draft, draft-ietf-6man-comp-rtg-hdr-03, 18 January 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-6man-comp-rtg-hdr-03>.

[I-D.ietf-spring-srv6-srh-compression]
Cheng, W., Filsfils, C., Li, Z., Decraene, B., and F. Clad, "Compressed SRv6 Segment List Encoding", Work in Progress, Internet-Draft, draft-ietf-spring-srv6-srh-compression-13, 29 February 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-srv6-srh-compression-13>.

[I-D.ietf-teas-ns-ip-mpls]
Saad, T., Beeram, V. P., Dong, J., Wen, B., Ceccarelli, D., Halpern, J. M., Peng, S., Chen, R., Liu, X., Contreras, L. M., Rokui, R., and L. Jalil, "Realizing Network Slices in IP/MPLS Networks", Work in Progress, Internet-Draft, draft-ietf-teas-ns-ip-mpls-03, 26 November 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-ns-ip-mpls-03>.

[I-D.peng-detnet-deadline-based-forwarding]
Peng, S., Du, Z., Basu, K., cheng, Yang, D., and C. Liu, "Deadline Based Deterministic Forwarding", Work in Progress, Internet-Draft, draft-peng-detnet-deadline-based-forwarding-08, 14 December 2023, <https://datatracker.ietf.org/doc/html/draft-peng-detnet-deadline-based-forwarding-08>.

[I-D.peng-detnet-packet-timeslot-mechanism]
Peng, S., Liu, P., Basu, K., Liu, A., Yang, D., and G. Peng, "Timeslot Queueing and Forwarding Mechanism", Work in Progress, Internet-Draft, draft-peng-detnet-packet-timeslot-mechanism-05, 14 December 2023, <https://datatracker.ietf.org/doc/html/draft-peng-detnet-packet-timeslot-mechanism-05>.

[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>.

[RFC6554] Hui, J., Vasseur, JP., Culler, D., and V. Manral, "An IPv6 Routing Header for Source Routes with the Routing Protocol for Low-Power and Lossy Networks (RPL)", RFC 6554, DOI 10.17487/RFC6554, March 2012, <https://www.rfc-editor.org/info/rfc6554>.

[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>.

[RFC8655] Finn, N., Thubert, P., Varga, B., and J. Farkas, "Deterministic Networking Architecture", RFC 8655, DOI 10.17487/RFC8655, October 2019, <https://www.rfc-editor.org/info/rfc8655>.

Authors' Addresses

Shaofu Peng
ZTE Corporation
China
Email: peng.shaofu@zte.com.cn

Ron Bonica
Juniper Networks
United States of America
Email: rbonica@juniper.net