Workgroup: INTAREA
Internet-Draft:
draft-zhou-intarea-computing-segment-san-03
Published: 18 October 2023
Intended Status: Standards Track
Expires: 20 April 2024
Authors: F. Zhou, ZTE Corporation
D. Yuan, ZTE Corporation
D. Yang, Beijing Jiaotong University

Computing Segment for Service Routing in SAN

Abstract

Since services provisioning requires delicate coordination among the client, network and cloud, this draft defines a new Segment to provide service routing and addressing functions by leveraging SRv6 Segment programming capabilities. With Computing Segments proposed, the network gains its capability to identify and process a SAN header in need and a complete service routing procedure can be achieved.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

1.1. Service Identification in SAN

In order to deliver responsive services to clients, computing resources continuously migrate and spread from central sites to edge nodes. As shown in Figure 1, multiple instances located distributedly in different resource pools are capable of providing services. Compared with applying traditional IP routing protocols, a fine-grained service routing policy is capable of achieving optimal and efficient invocation of both computing power and the network.

```
                                       +-------------+ +---------+
                                     +-+Load Balancer+-+Service 1|
                                     | +-------------+ +---------+
                                     |
+-------+   +-----------+   +----------+ | +-------------+ +---------+
|Client+---+Ingress PE+---+Egress PE+-+-+Load Balancer+-+Service 2|
+-------+   +-----------+   +----------+ | +-------------+ +---------+
                                     |
                                     | +-------------+ |---------+
                                     +-+Load Balancer+-|Service 3|
                                       +-------------+ +---------+
|<-Client->|<---------Network-------->|<----------Cloud---------->|
```

Figure 1: Computing Power Networks

In order to implement service routing, the network should be aware of specific services and a service awareness network framework is introduced in [I-D.huang-service-aware-network-framework]. Within the proposed network framework, a service identification is defined as a SAN ID(Service ID) in [I-D.ma-intarea-identification-header-of-san] to represent a globally unique service semantic identification.

As mentioned in [I-D.ma-intarea-encapsulation-of-san-header], a SAN ID is encapsulated in a SAN header which can be carried as an option in the IPv6 Hop-by-Hop Options Header, Destination Options Header and a type of SRH TLV. Since services provisioning requires delicate coordination among the client, network and cloud and thus simply encapsulating SAN header among packets delivery can hardly satisfy various practical situations:

* The Destination Options header is used to carry optional information that need be examined by the destination of the path which is defined in [RFC8200], SAN header will only be resolved by the destination node. When a multi-layer service routing strategy is applied in the network domain, a quantity of relay nodes besides the destination are also required to identify SAN ID and forward the received packet accordingly as well. Thus, simply carring a SAN header can not fulfill a multi-layer service routing procedure.

* When a SAN header is carried as an option in the IPv6 Hop-by-Hop Options Header, it may be processed by each nodes. Practically, not all nodes along the delivery path of the packet are capable of identifying and processing a SAN header. The SAN header may be changed accidentally and the packet may even be discarded in the forwarding process.

* The Segment Routing Header (SRH) and the SRH TLV is defined in [RFC8754]. Since the segment list is encoded in order, it indicates that the service routing process and successive forwarding behaviours must be orchestrated in advance. However, previous orchestration brings visible restrictions to the flexibility and adaptability of service routing.

To achieve a SAN header being processed in need in the network domain and to preserve its identifiability along the path from the client to the server, a new Segment to specify and standardize node behaviours is urgently required.

1.2. Service Routing in SAN

As shown in Figure 2, a service routing table is designed to establish a mapping relationship between the SAN ID and the conventional IP routing table.

```
                                   +-------+
                                   |  I P  |
                  SAN ID <-------------> |Routing|
                              |          | Table |
                              |          +-------+
                              v
                          +-------+
                          |Service|
                          |Routing|
                          | Table |
                          +-------+

  +--------+        +-----------+          +----------+        +-----+
  | Client +---------+Ingress  PE+----------+Egress  PE+---------+ L B |
  +--------+        +-----------+          +----------+        +-----+
```

Figure 2: Service Routing in SAN

A service routing table can be published from a control and management system to the network domain within a centralized control plane while it can also be calculated and generated by the Ingress PE itself under a distributed control plane.

With considerations of path metrics, computing status and service SLA requirements, a specific service routing table is introduced, including mutiple attributes, SAN ID and outer gateway for instance. Afterwards, a corresponding IP routing table should be indexed which further determines the next hop or an SRv6 policy.

In order to describe and standardize the mentioned behaviours, a new Computing Segment is proposed. With Computing Segments, multiple nodes in the network domain can be informed to identify and resolve SAN header in need and to implement a referred forwarding behaviour through the complete procedure.

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. Terminology

* SAN: Service Aware Network
* SAN ID: Service Aware Network Identification, an identification designed to indicate the fundamental and common service types
* SAN header: Encapsulation format of the SAN ID
* DOH: Destination Options Header
* HBH: Hop-by-Hop Options Header
* SRH: Segment Routing Header
* SID: Segment Identifier
* FIB: Forwarding Information Base
* DA: Destination Address
* LB: Load Balancer

4. Computing Segment

This draft introduces a new SRv6 Segment, namely Computing Segment, aiming to describe the behaviour of querying service routing table and corresponding packet forwarding.

Computing Segment is the identifier of packets in which a corresponding SAN header should be identified and further being forwarded via the matched service routing table entity, indicating the following operations:

* Identify the SAN ID encapsulated in DOH, HBH or SRH TLV.
* Query the service routing table indexed by SAN ID.
* Update destination address accordingly.
* Push a new IPv6 header with possible SRH containing the list of segments of the SRv6 policy.
* Forward the packet.

In the case of SRv6, a new behavior End.C for Computing Segment is defined. Behaviours of End.C are described in the following sections.

4.1. When the SAN ID is encapsulated in the DOH

When an IPv6 node (N) receives an IPv6 packet whose destination address matches a local IPv6 address instantiated as a SID (S), and S is a Computing SID, N does:

(1) If the traffic is steered into a tunnel, an SRv6 policy for instance:

```
S01.  If (IPv6 Hop Limit <= 1) {
S02.    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.
S03.  }
S04.  Decrement IPv6 Hop Limit by 1
S05.  Resolve the SAN ID encapsulated in the DOH
S06.  Maintain the SAN Header in the DOH
S07.  Query the service routing table indexed by SAN ID to determine
      an outer gateway and an according SRv6 policy
S08.  If an SRH is carried in the IPv6 header {
S09.    If (Segments Left == 0) {
S10.      Stop processing the SRH, and proceed to process the next
          header in the packet, whose type is identified by
          the Next Header field in the routing header.
S11.    }
S12.    max_LE = (Hdr Ext Len / 2) - 1
S13.    If ((Last Entry > max_LE) or (Segments Left > Last Entry+1)) {
S14.      Send an ICMP Parameter Problem to the Source Address
          with Code 0 (Erroneous header field encountered)
          and Pointer set to the Segments Left field,
          interrupt packet processing, and discard the packet.
S15.    }
S16.    Decrement Segments Left by 1
S17.    Update IPv6 DA with Segment List[Segments Left]
S18.  }
S19.  else {
S20.      Update IPv6 DA with the queried gateway
S21.  }
S22.  Push a new IPv6 header with its own SRH containing the list of
      segments of the SRv6 policy
S23.  Set the outer IPv6 SA to itself
S24.  Set the outer IPv6 DA to the first SID of the SRv6 policy
S25.  Set the outer Payload Length, Traffic Class, Flow Label and
      Next Header fields
S26.  Submit the packet to the egress IPv6 FIB lookup for transmission
      to the new destination
```

Figure 3: When the SAN ID is encapsulated in the DOH: Part 1

(2) If the traffic is steered in a BE manner:

The line S07 and lines from S22 to S24 are replaced by the following:

```
S07.    Query the service routing table indexed by SAN ID to determine
        an outer gateway

S22.    Push a new IPv6 header
S23.    Set the outer IPv6 SA to itself
S24.    Set the outer IPv6 DA to the queried outer gateway
```

Figure 4: When the SAN ID is encapsulated in the DOH: Part 2

4.2. When the SAN ID is encapsulated in the HBH

When an IPv6 node (N) receives an IPv6 packet whose destination address matches a local IPv6 address instantiated as a SID (S), and S is a Computing SID, N does:

(1) If the traffic is steered into a tunnel, an SRv6 policy for instance:

```
S01.  If (IPv6 Hop Limit <= 1) {
S02.    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.
S03.  }
S04.  Decrement IPv6 Hop Limit by 1
S05.  Resolve the SAN ID encapsulated in the HBH
S06.  Maintain the SAN Header in the HBH
S07.  Query the service routing table indexed by SAN ID to determine
      an outer gateway and an according SRv6 policy
S08.  If an SRH is carried in the IPv6 header {
S09.    If (Segments Left == 0) {
S10.      Stop processing the SRH, and proceed to process the next
          header in the packet, whose type is identified by
          the Next Header field in the routing header.
S11.    }
S12.    max_LE = (Hdr Ext Len / 2) - 1
S13.    If ((Last Entry > max_LE) or (Segments Left > Last Entry+1)) {
S14.      Send an ICMP Parameter Problem to the Source Address
          with Code 0 (Erroneous header field encountered)
          and Pointer set to the Segments Left field,
          interrupt packet processing, and discard the packet.
S15.    }
S16.    Decrement Segments Left by 1
S17.    Update IPv6 DA with Segment List[Segments Left]
S18.  }
S19.  else {
S20.      Update IPv6 DA with the queried gateway
S21.  }
S22.  Push a new IPv6 header with its own SRH containing the list of
      segments of the SRv6 policy
S23.  Set the outer IPv6 SA to itself
S24.  Set the outer IPv6 DA to the first SID of the SRv6 policy
S25.  Set the outer Payload Length, Traffic Class, Flow Label and
      Next Header fields
S26.  Submit the packet to the egress IPv6 FIB lookup for transmission
      to the new destination
```

Figure 5: When the SAN ID is encapsulated in the HBH: Part 1

(2) If the traffic is steered in a BE manner:

The line S07 and lines from S22 to S24 are replaced by the following:

```
 S07.    Query the service routing table indexed by SAN ID to determine
         an outer gateway

 S22.    Push a new IPv6 header
 S23.    Set the outer IPv6 SA to itself
 S24.    Set the outer IPv6 DA to the queried outer gateway
```

Figure 6: When the SAN ID is encapsulated in the HBH: Part 2

4.3. When the SAN ID is encapsulated in the SRH TLV

When an IPv6 node (N) receives an IPv6 packet whose destination address matches a local IPv6 address instantiated as a SID (S), and S is a Computing SID, N does:

(1) If the traffic is steered into a tunnel, an SRv6 policy for instance:

```
S01.  When an SRH is processed {
S02.    If (Segments Left == 0) {
S03.      Stop processing the SRH, 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.    max_LE = (Hdr Ext Len / 2) - 1
S09.    If ((Last Entry > max_LE) or (Segments Left > Last Entry+1)) {
S10.      Send an ICMP Parameter Problem to the Source Address
          with Code 0 (Erroneous header field encountered)
          and Pointer set to the Segments Left field,
          interrupt packet processing, and discard the packet.
S11.    }
S12.    Decrement IPv6 Hop Limit by 1
S13.    Decrement Segments Left by 1
S14.    Update IPv6 DA with Segment List[Segments Left]
S15.    Resolve the SAN ID encapsulated in the HBH, DOH or a type of
        SRH TLV
S16.    Maintain the SAN Header in the HBH, DOH or a type of SRH TLV
S17.    Query the service routing table indexed by SAN ID to determine
        an outer gateway and an according SRv6 policy
S18.    Push a new IPv6 header with its own SRH containing the list of
        segments of the SRv6 policy
S19.    Set the outer IPv6 SA to itself
S20.    Set the outer IPv6 DA to the first SID of the SRv6 policy
S21.    Set the outer Payload Length, Traffic Class, Flow Label and
        Next Header fields
S22.    Submit the packet to the egress IPv6 FIB lookup for transmission
        to the new destination
S23.  }
```

Figure 7: When the SAN ID is encapsulated in the SRH TLV: Part 1

(2) If the traffic is steered in a BE manner:

The lines from S17 to S20 are replaced by the following:

S17.   Query the service routing table indexed by SAN ID to determine
        an outer gateway
S18.   Push a new IPv6 header
S19.   Set the outer IPv6 SA to itself
S20.   Set the outer IPv6 DA to the queried outer gateway

Figure 8: When the SAN ID is encapsulated in the SRH TLV: Part 2

5. Use Case

When a SAN header is carried as an option in the DOH, a typical service routing procedure is shown in Figure 9.

```
+--------+         +-----------+        +----------+         +-----+
| Client +---------+Ingress  PE+--------+Egress  PE+---------+ L B |
+--------+         +-----------+        +----------+         +-----+

Inner IPv6 Packet:

        +-----------+        +-----------+        +-----------+
        |    SIP    |        |    SIP    |        |    SIP    |
        +-----------+        +-----------+        +-----------+
        |END.C(SID1)|        |END.C(SID2)|        |    DIP    |
        +-----------+        +-----------+        +-----------+
        |    DOH    |        |    DOH    |        |    DOH    |
        +-----------+        +-----------+        +-----------+
        |  PAYLOAD  |        |  PAYLOAD  |        |  PAYLOAD  |
        +-----------+        +-----------+        +-----------+

  DOH:
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |  Next Header  | Hdr Ext Len   |  Opt  Length  |Opt Data Length|
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                          SAN Header                           |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                                |
  Service Routing Table:        v
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |       SAN ID      |      Gateway      |        Interface      |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |        ID 1       |      Egress 1     |         Policy 1      |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |        ID 2       |      Egress 2     |         Policy 2      |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                             ......                            |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 9: Typical Service Routing Procedure with Service ID Encapsulated in the DOH

Suppose the Endpoint behaviour of END.C is configured at Ingress PE and Egress PE, namely SID 1 and SID 2 respectively. SID1 and SID2 are advertised in the network domain by IGP. The client registers with the management and operation system to acquire a SAN ID and encapsulates it in the packet. The initial destination is END.C (SID 1) which may be configured in a static routing manner. The service addressing procedure from the client to the cloud is described below:

* The Computing SID of Ingress PE (SID1) is configured as DA. The packet carrying the SAN header as the option of the DOH is forwarded to Ingress PE.

* When Ingress PE receives the packet, it identifys that DA is a Computing SID (SID1). As defined in 4.2, the Ingress PE determines the next hop for service routing which is END.C (SID 2) and updates DA. Ingress PE encapsulates an outer IPv6 header and continues to forward the packet carrying the DOH.

* When Egress PE receives the packet, it identifys that DA is a Computing SID (SID2). As defined in 4.2, the Egress PE determines the next hop for service routing which is DIP which represents a specific service instance and updates DA. Egress PE further continues to forward the packet carrying the DOH.

* When an intra-cloud LB receives the packet, the packet can be forwarded in a service routing manner or be processed in a native IP way, depending on the practical circumstances.

```
|<-Client->|<--------------------Network----------------->|<-Cloud->|
+------+     +----------+                   +---------+     +-----+
|Client+-----+Ingress PE+-------------------+Egress PE+-----+ L B |
+------+     +----------+         |         +---------+     +-----+
                          BE:     v      TE:
                      +-----------+  +-----------+
                      |    IIP    |  |    IIP    |
                      +-----------+  +-----------+
                      |    SID    |  |    SID    |
                      +-----------+  +-----------+
                      |    SIP    |  |    SRH    |
                      +-----------+  +-----------+
                      |END.C(SID2)|  |    SIP    |
                      +-----------+  +-----------+
                      |    DOH    |  |END.C(SID2)|
                      +-----------+  +-----------+
                      |  PAYLOAD  |  |    DOH    |
                      +-----------+  +-----------+
                                     |  PAYLOAD  |
                                     +-----------+
```

Figure 10: Outer Headers Encapsulated between Ingress PE and Egress PE

As shown in Figure 10, between Ingress PE and Egress PE, an outer header including SRH should be encapsulated when the traffic follows a specific SRv6 TE policy. Otherwise, a normal IPv6 header should be encapsulated under a BE condition.

6. Security Considerations

Security has always been an indispensable and significant consideration for design and innovation in the fields of communication and services provisioning. A Computing Segment as END.C defined in this draft may be given security semantics and according behaviours, including encryption and decryption, etc. Security considerations may be studied in the future work.

7. Acknowledgements

TBA.

8. IANA Considerations

This document requires registration of End.C behavior in "SRv6 Endpoint Behaviors" sub-registry of "Segment Routing Parameters" registry.

9. Normative References

[I-D.huang-service-aware-network-framework] Huang, D., Tan, B., and D. Yang, "Service Aware Network Framework", Work in Progress, Internet-Draft, draft-huang-service-aware-network-framework-01, 22 November 2022, <https://datatracker.ietf.org/doc/html/draft-huang-service-aware-network-framework-01>.

[I-D.ma-intarea-encapsulation-of-san-header] Ma, L., Zhao, D., Zhou, F., and D. Yang, "Encapsulation of SAN Header", Work in Progress, Internet-Draft, draft-ma-intarea-encapsulation-of-san-header-00, 23 October 2022, <https://datatracker.ietf.org/doc/html/draft-ma-intarea-encapsulation-of-san-header-00>.

[I-D.ma-intarea-identification-header-of-san] Ma, L., , Zhou, F., lihesong, and D. Yang, "Service Identification Header of Service Aware Network", Work in Progress, Internet-Draft, draft-ma-intarea-identification-header-of-san-01, 4 May 2023, <https://datatracker.ietf.org/doc/html/draft-ma-intarea-identification-header-of-san-01>.

[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>.

[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>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

Authors' Addresses

Fenlin Zhou
ZTE Corporation
No.50 Software Avenue
Nanjing
Jiangsu, 210012
China

Email: zhou.fenlin@zte.com.cn

Dongyu Yuan
ZTE Corporation
No.50 Software Avenue
Nanjing
Jiangsu, 210012
China

Email: yuan.dongyu@zte.com.cn

Dong Yang
Beijing Jiaotong University
No.3 Shangyuancun Haidian District
Beijing
100044
China

Email: dyang@bjtu.edu.cn