Workgroup: SPRING
Internet-Draft:
draft-clad-spring-srv6-srh-compression-
illus-01
Published: 19 April 2022
Intended Status: Informational
Expires: 21 October 2022
Authors: F. Clad, Ed.
Cisco Systems, Inc.
D. Dukes, Ed.
Cisco Systems, Inc.

Illustrations for Compressed SRv6 Segment List Encoding in SRH

Abstract

This document provides illustrations for compressed SRv6 Segment List Encoding in the Segment Routing Header (SRH).

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 21 October 2022.

Table of Contents

1. Introduction

This document provides illustrations for [I-D.filsfilscheng-spring-srv6-srh-compression] compressed SRv6 Segment List Encoding in the Segment Routing Header (SRH).

2. Terminology

This document leverages the terminology introduced in [RFC8402], [RFC8754], and [RFC8986]. The definition of the most important terms is reproduced in this section for convenience.

2.1. From RFC 8402

Segment Routing domain (SR domain): the set of nodes participating in the source-based routing model. These nodes may be connected to the same physical infrastructure (e.g., a Service Provider's network). They may as well be remotely connected to each other (e.g., an enterprise VPN or an overlay). If multiple protocol instances are deployed, the SR domain most commonly includes all of the protocol instances in a network. However, some deployments may wish to subdivide the network into multiple SR domains, each of which includes one or more protocol instances. It is expected that all nodes in an SR domain are managed by the same administrative entity.

2.2. From RFC 8754

SR Source Node (section 3.1): A SR source node is any node that originates an IPv6 packet with a segment (i.e., SRv6 SID) in the destination address of the IPv6 header.

Transit Node (section 3.2): A transit node is any node forwarding an IPv6 packet where the destination address of that packet is not locally configured as a segment or a local interface. A transit node is not required to be capable of processing a segment or SRH.

SR Segment Endpoint Node (section 3.3): An SR segment endpoint node is any node receiving an IPv6 packet where the destination address of that packet is locally configured as a segment or local interface.

2.3. From RFC 8986

SID Format: This document defines an SRv6 SID as consisting of LOC:FUNCT:ARG, where a locator (LOC) is encoded in the L most significant bits of the SID, followed by F bits of function (FUNCT) and A bits of arguments (ARG). L, the locator length, is flexible, and an operator is free to use the locator length of their choice. F and A may be any value as long as L+F+A <= 128. When L+F+A is less than 128, then the remaining bits of the SID MUST be zero. A locator may be represented as B:N where B is the SRv6 SID block (IPv6 prefix allocated for SRv6 SIDs by the operator) and N is the identifier of the parent node instantiating the SID.

3. Intra-SR-Domain Deployment Model

(The content of this section is a partial reproduction of section 5 for [RFC8754].)

The use of the SIDs exclusively within the SR domain and solely for packets of the SR domain is an important deployment model.

This enables the SR domain to act as a single routing system.

3.1. Securing the SR Domain

(The reader can easily understand that the dual measures provided can prevent SR packets from leaving the SR domain.)

Nodes outside the SR domain are not trusted: they cannot directly use the SIDs of the domain. This is enforced by two levels of access control lists:

* Any packet entering the SR domain and destined to a SID within the SR domain is dropped. This may be realized with the following

logic. Other methods with equivalent outcome are considered compliant:

- -Allocate all the SIDs from a block S/s
- -Configure each external interface of each edge node of the domain with an inbound infrastructure access list (IACL) that drops any incoming packet with a destination address in S/s
- -Failure to implement this method of ingress filtering exposes the SR domain to source-routing attacks, as described and referenced in [RFC5095]

*The distributed protection in #1 is complemented with per-node protection, dropping packets to SIDs from source addresses outside the SR domain. This may be realized with the following logic. Other methods with equivalent outcome are considered compliant:

- -Assign all interface addresses from prefix A/a
- -At node k, all SIDs local to k are assigned from prefix Sk/sk
- -Configure each internal interface of each SR node k in the SR domain with an inbound IACL that drops any incoming packet with a destination address in Sk/sk if the source address is not in A/a.

4. General Addressing

The illustrations in this document use the IPv6 documentation prefix 2001:db8::/32.

Loopback interface addresses are allocated from the prefix 2001:db8:a::/48.

SRv6 SIDs are allocated from the prefix 2001:db8:b::/48.

An operator deploying this solution could instead select any sub-prefix out of the prefix allocated by their Regional Internet Registry (RIR) to this operator or from the Unique Local Unicast (ULA) prefix. ULA provides the uniqueness and privacy characteristics defined in Section 1 of [RFC4193].

5. NEXT-C-SID Flavor

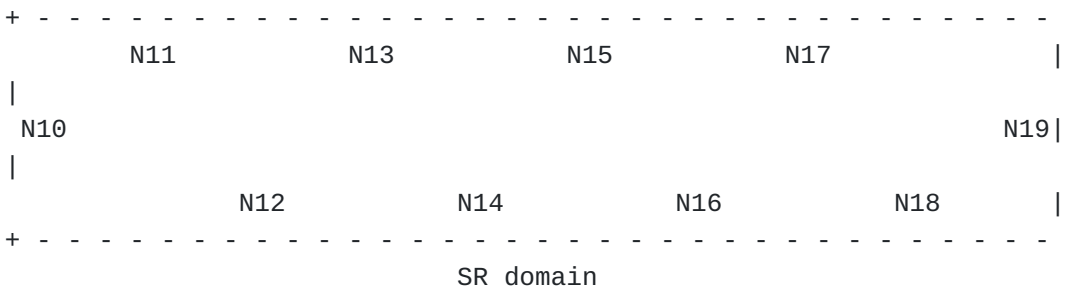

Figure 1: Reference topology

N10 to N19 represent the potential SR source and SR segment endpoint nodes in the SR domain.

The SR domain may include any number of transit nodes (not shown) between the nodes that are represented in this figure.

5.1. Addressing and SRv6 SID allocation

Nodes N10 to N19 have a loopback interface configured with the address 2001:db8:a:NN00::, where NN is the node identifier.

Nodes N10 to N19 instantiate the SID 2001:db8:b:NN00::, where NN is the node identifier, with Locator-Block length (LBL) = 48, Locator-Node length (LNL)= 16, Function length (FL) = 0, Argument length (AL) = 64, and bound to the End behavior with the NEXT-C-SID and USD flavors.

The "Endpoint" (or "End") behavior is the most basic operation that can be performed by an SR segment endpoint node (i.e., a node that identifies the destination address of a received packet as matching a locally instantiated SID). It updates the destination address of the packet with the next SID in the segment list. The pseudocode of the End behavior with the NEXT-C-SID and USD flavors is specified in section 4.1.1 of [I-D.filsfilscheng-spring-srv6-srh-compression].

5.2. Routing

Nodes N10 to N19 advertise the prefixes 2001:db8:a:NN00::/64 and 2001:db8:b:NN00::/64, where NN is the node identifier, in the IGP.

5.3. Case 1: Intra-domain Traffic Engineering

Let us assume that a centralized controller programs N11 to classify the traffic from 2001:db8:a:1000:: to 2001:db8:a:1900:: into an SR Policy encoded through an IPv6 encapsulation with:

- IPv6
  - Source address 2001:db8:a:1100::

```
      -Destination address 2001:db8:b:1200:1300:1400:1500:1600

      -Next Header = 43 (Routing header)

    *SRH

      -Segment List < 2001:db8:b:1200:1300:1400:1500:1600,
       2001:db8:b:1700:1800:: >

      -Segments Left = 1

      -Next Header = 41 (IPv6)
```

For illustration purposes, we use SID allocation that allows for a straightforward human reading of a compressed segment list. Indeed, < 2001:db8:b:1200:1300:1400:1500:1600, 2001:db8:b:1700:1800:: > means: within the domain 2001:db8:b::, go first through node N12 then N13, N14, N15, and N16, then retrieve the next segment list entry from the SRH and go through node N17 before decapsulating the packet at node N18.

This is compliant with the [RFC8986] because the SID meets the Locator:Function:Argument format definition (Section 3.1 of [RFC8986]). For example, the packet sent by node N11 has a destination address 2001:db8:b:1200:1300:1400:1500:1600 where 2001:db8:b:1200/64 is the Locator and 0x1300140015001600 is the Argument.

A packet in transit towards a given SID (e.g. 2001:db8:b:1200:1300:1400:1500:1600), is forwarded by transit nodes via a longest-match lookup on the destination address of the packet. This results in a match of the SID locator (in this case, 2001:db8:b:1200::/64), the transit node then forwards the packet accordingly. The SID function and argument bits are opaque to transit nodes. The function is only identified at the SR segment endpoint node (represented by the SID locator in the destination address) which further processes the argument.

Also note the source N11 performs IPv6 header encapsulation with SRH, and the selected SID list containing function/arguments to be processed at some endpoints, because we are in a source routed domain within a secured SR domain.

The remainder of this section details the packet journey.

The packet Px transmitted by a node Nn is identified as "@Nn Px".

```
@N10 P1:(IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)
```

N11 (as programmed by the centralized controller) encapsulates the packet P1 and submits the updated packet (P2) to the IPv6 module for transmission. It performs an IP lookup on the destination address, matching an entry for the prefix 2001:db8:b:1200::/64 advertised by N12. N11 forwards the packet on its shortest path towards to node N12.

```
@N11 P2:(IPv6 2001:db8:a:1100::, 2001:db8:b:1200:1300:1400:1500:1600)
        (SRH 2001:db8:b:1700:1800::,
             2001:db8:b:1200:1300:1400:1500:1600;
             SL=1)
        (IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)
```

The transit nodes between N11 and N12 forward P1 as per their route 2001:db8:b:1200::/64 to N12. Similarly, the transit nodes between each subsequent pair of consecutive SR segment endpoint nodes forwards the packet as per their IPv6 routes for the destination address. Those transit nodes are plain IPv6 routers with the plain IPv6 dataplane, they do not need to have any knowledge of SRv6.

The hop limit of packet P1 is decremented at every transit node and every SR segment endpoint node.

When the packet reaches the first SR segment endpoint node N12 (i.e., the first TE waypoint), this performs a longest-prefix-match lookup on the IPv6 destination address. This lookup returns a FIB entry that represents a locally instantiated SRv6 SID bound to the End behavior with the NEXT-C-SID flavor. N12 processes the packet accordingly, resulting in a new destination address. It then submits the updated packet to the IPv6 module for transmission. This triggers an IP lookup on the destination address, matching an entry for the prefix 2001:db8:b:1300::/64 advertised by N13. The packet is forwarded on the shortest path towards N13.

```
@N12 P2:(IPv6 2001:db8:a:1100::, 2001:db8:b:1300:1400:1500:1600:0000)
        (SRH 2001:db8:b:1700:1800::,
             2001:db8:b:1200:1300:1400:1500:1600;
             SL=1)
        (IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)
```

The subsequent SR segment endpoint nodes N13 to N17 process the packet similarly.

```
@N13 P2:(IPv6 2001:db8:a:1100::, 2001:db8:b:1400:1500:1600:0000:0000)
        (SRH 2001:db8:b:1700:1800::,
             2001:db8:b:1200:1300:1400:1500:1600;
             SL=1)
        (IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)
```

```
@N14 P2:(IPv6 2001:db8:a:1100::, 2001:db8:b:1500:1600:0000:0000:0000)
        (SRH 2001:db8:b:1700:1800::,
             2001:db8:b:1200:1300:1400:1500:1600;
             SL=1)
        (IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)

@N15 P2:(IPv6 2001:db8:a:1100::, 2001:db8:b:1600:0000:0000:0000:0000)
        (SRH 2001:db8:b:1700:1800::,
             2001:db8:b:1200:1300:1400:1500:1600;
             SL=1)
        (IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)
```

When the packet is processed by the SR segment endpoint node N16, the SID argument value is 0. As per the pseudocode of the End behavior with the NEXT-C-SID and USD flavors, N16 retrieves the next SID by decrementing the value of segments left in the SRH and copying the next entry from the SRH segment list into the destination address.

```
@N16 P2:(IPv6 2001:db8:a:1100::, 2001:db8:b:1700:1800::)
        (SRH 2001:db8:b:1700:1800::,
             2001:db8:b:1200:1300:1400:1500:1600;
             SL=0)
        (IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)

@N17 P2:(IPv6 2001:db8:a:1100::, 2001:db8:b:1800:0000::)
        (SRH 2001:db8:b:1700:1800::,
             2001:db8:b:1200:1300:1400:1500:1600;
             SL=0)
        (IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)
```

When the packet reaches the final SR segment endpoint node N18, both the SID argument value and the segments left value in the SRH are 0. As per the pseudocode of the End behavior with the NEXT-C-SID and USD flavors, N18 decapsulates the packet and sends the inner packet P1 towards its destination 2001:db8:a:1900::.

```
@N18 P1:(IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)
```

5.4. Case 2: ICMPv6 error generation at a transit node

Let us assume in the previous example that the hop limit expires on a transit node N141, located on the path between the SR segment endpoint nodes N14 and N15.

The packet sent by node N14 is as follows (reproduced from the previous section).

```
@N14 P2:(IPv6 2001:db8:a:1100::, 2001:db8:b:1500:1600:0000:0000:0000)
        (SRH 2001:db8:b:1700:1800::,
             2001:db8:b:1200:1300:1400:1500:1600;
             SL=1)
        (IPv6 2001:db8:a:1000::, 2001:db8:a:1900::)
```

Node N141 generates an ICMPv6 time exceeded error message as follows.

```
@N141 P3: (IPv6 <any address of node N141>, 2001:db8:a:1100::)
          (ICMPv6 time exceeded error
            (IPv6 2001:db8:a:1100::,
                  2001:db8:b:1500:1600:0000:0000:0000)
            (SRH 2001:db8:b:1700:1800::,
                 2001:db8:b:1200:1300:1400:1500:1600;
                 SL=1)
            (IPv6 2001:db8:a:1000::, 2001:db8:a:1900::))
```

Node N11 receives the ICMP error packet transmitted by N141. Section 5.4 of [RFC8754] indicates that a destination address of the invoking packet is determined by looking at Segment List[0].

5.5. Case 3: Ping a SID

The operator wants to ping the End with NEXT-C-SID flavor SID 2001:db8:b:1200:: of N12 from the SR source node N10.

The ICMP echo request is sent by N10 as follows.

```
@N10 P1:(IPv6 2001:db8:a:1000::, 2001:db8:b:1200::)
        (ICMPv6 echo request)
```

This results in an ICMP echo reply from N12 to N10.

```
@N12 P2:(IPv6 2001:db8:b:1200::, 2001:db8:a:1000::)
        (ICMPv6 echo reply)
```

6. REPLACE-C-SID Flavor

TBD

7. Acknowledgements

TBD

8. References

8.1. Normative References

[I-D.filsfilscheng-spring-srv6-srh-compression]

Cheng, W., Filsfils, C., Li, Z., Decraene, B., Cai, D., Voyer, D., Clad, F., Zadok, S., Guichard, J. N., Aihua, L., Raszuk, R., and C. Li, "Compressed SRv6 Segment List Encoding in SRH", Work in Progress, Internet-Draft, draft-filsfilscheng-spring-srv6-srh-compression-02, 28 July 2021, <https://www.ietf.org/archive/id/draft-filsfilscheng-spring-srv6-srh-compression-02.txt>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[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>.

[RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <https://www.rfc-editor.org/info/rfc8986>.

8.2. Informative References

[RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, DOI 10.17487/RFC4193, October 2005, <https://www.rfc-editor.org/info/rfc4193>.

[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>.

Authors' Addresses

Francois Clad (editor)
Cisco Systems, Inc.
France

Email: fclad@cisco.com

Darren Dukes (editor)
Cisco Systems, Inc.
Canada

Email: ddukes@cisco.com