Network Working Group D. Voyer, Ed.
Internet-Draft Bell Canada
Intended status: Standards Track C. Filsfils
Expires: May 2, 2021 R. Parekh
Cisco Systems, Inc.
H. Bidgoli
Nokia
Z. Zhang
Juniper Networks
October 29, 2020

SR Replication Segment for Multi-point Service Delivery
draft-ietf-spring-sr-replication-segment-03

Abstract

This document describes the SR Replication segment for Multi-point service delivery. A SR Replication segment allows a packet to be replicated from a replication node to downstream nodes.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Copyright Notice

Table of Contents

1. Introduction

We define a new type of segment for Segment Routing [RFC8402], called Replication segment, which allows a node (henceforth called as Replication Node) to replicate packets to a set of other nodes (called Downstream Nodes) in a Segment Routing Domain. Replication segments provide building blocks for Point-to-Multipoint Service delivery via SR Point-to-Multipoint (SR P2MP) policy. A Replication segment can replicate packet to directly connected nodes or to downstream nodes (without need for state on the transit routers). This document focuses on the Replication Segment building block. The use of one or more stitched Replication Segments constructed for SR P2MP Policy tree is specified in [I-D.voyer-pim-sr-p2mp-policy].

2. Replication Segment

In a Segment Routing Domain, a Replication segment is a logical construct which connects a Replication Node to a set of Downstream Nodes. A Replication segment is a local segment instantiated at a Replication node. It can be either provisioned locally on a node or programmed by a PCE. Replication segments apply equally to both SR-MPLS and SRv6 instantiations of Segment Routing.

A Replication segment is identified by the tuple <Replication-ID, Node-ID>, where:

o Replication-ID: An identifier for a Replication segment that is unique in context of the Replication Node.

o Node-ID: The address of the Replication Node that the Replication segment is for. Note that the root of a multi-point service is also a replication node.

In simplest case, Replication-ID can be a 32-bit number, but it can be extended or modified as required based on specific use of a Replication segment. When the PCE signals a Replication segment to its node, the <Replication-ID, Node-ID> tuple identifies the segment. Examples of such signaling and extension are described in [I-D.voyer-pim-sr-p2mp-policy].

A Replication segment includes the following elements:

o Replication SID: The Segment Identifier of a Replication segment. This is a SR-MPLS label or a SRv6 SID [RFC8402].

o Downstream Nodes: Set of nodes in Segment Routing domain to which a packet is replicated by the Replication segment.

o Replication State: See below.

The Downstream Nodes and Replication State of a Replication segment can change over time, depending on the network state and leaf nodes of a multi-point service that the segment is part of.

Replication State is a list of replication branches to the Downstream Nodes. In this document, each branch is abstracted to a <Downstream Node, Downstream Replication SID> tuple. A Downstream Node is represented by a SID-list or a Segment Routing Policy [I-D.ietf-spring-segment-routing-policy] that specifies the explicit path from the Replication Node to the Downstream Node, or even represented by another Replication segment. The SID-list MAY just

have one SID. If a downstream node is adjacent to a Replication node, it MAY also be represented by an interface.

Replication SID identifies the Replication segment in the forwarding plane. At a Replication node, the Replication SID is the equivalent of Binding SID [I-D.ietf-spring-segment-routing-policy] of a Segment Routing Policy.

A packet steered into a Replication segment at a Replication node is replicated to each Downstream Node with the Downstream Replication SID that is relevant at that node. A packet is steered into a Replication Segment in two ways:

- o When the Active Segment [RFC8402] is the Replication SID. In this case, the operation is NEXT followed by a PUSH for a replicated copy.

- o On the root of a multi-point service, based on local policy-based routing. In this case, the operation for a replicated copy is PUSH.

If a Downstream Node is an egress (aka leaf) of the multi-point service, i.e. no further replication is needed, then that leaf node's Replication segment will not have any Replication State and the operation is NEXT. At an egress node, the Replication SID MAY be used to identify that portion of the multi-point service. Notice that the segment on the leaf node is still referred to as a Replication segment for the purpose of generalization.

A node can be a bud node, i.e. it is a replication node and a leaf node of a multi-point service at the same time [I-D.voyer-pim-sr-p2mp-policy]. In this case, the Replication segment's Replication State includes a branch with the Downstream Node being itself and the operation for the replicated copy is NEXT.

The Replication SID MUST be the last SID (at the bottom of stack for SR-MPLS) in a packet that is steered out from a Replication node of a Replication Segment. The behavior at Downstream nodes of a Replication Segment is undefined If there are any SIDs after the Replication SID and is outside the scope of this document.

3. Use Cases

In the simplest use case, a single Replication segment includes the root node of a multi-point service and the egress/leaf nodes of the the service as all the Downstream Nodes. This achieves Ingress Replication [RFC7988] that has been widely used for MVPN [RFC6513] and EVPN [RFC7432] BUM (Broadcast, Unknown and Multicast) traffic.

Replication segments can also be used as building blocks for replication trees when Replication segments on the root, intermediate replication nodes and leaf nodes are stitched together to achieve efficient replication. That is specified in [I-D.voyer-pim-sr-p2mp-policy].

4. IANA Considerations

This document makes no request of IANA.

5. Security Considerations

There are no additional security risks introduced by this design.

6. Acknowledgements

The authors would like to acknowledge Siva Sivabalan, Mike Koldychev, Vishnu Pavan Beeram, Alexander Vainshtein, Bruno Decraene and Joel Halpern for their valuable inputs.

7. Contributors

Clayton Hassen
Bell Canada
Vancouver
Canada

Email: clayton.hassen@bell.ca

Kurtis Gillis
Bell Canada
Halifax
Canada

Email: kurtis.gillis@bell.ca

Arvind Venkateswaran
Cisco Systems, Inc.
San Jose
US

Email: arvvenka@cisco.com

Zafar Ali
Cisco Systems, Inc.
US

Email: zali@cisco.com

Swadesh Agrawal
Cisco Systems, Inc.
San Jose
US

Email: swaagraw@cisco.com

Jayant Kotalwar
Nokia
Mountain View
US

Email: jayant.kotalwar@nokia.com

Tanmoy Kundu
Nokia
Mountain View
US

Email: tanmoy.kundu@nokia.com

Andrew Stone
Nokia
Ottawa
Canada

Email: andrew.stone@nokia.com

Tarek Saad
Juniper Networks
Canada

Email:tsaad@juniper.net

8. References

8.1. Normative References

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-08 (work in progress), July 2020.

[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>.

[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>.

8.2. Informative References

[I-D.voyer-pim-sr-p2mp-policy] Voyer, D., Filsfils, C., Parekh, R., Bidgoli, H., and Z. Zhang, "Segment Routing Point-to-Multipoint Policy", draft-voyer-pim-sr-p2mp-policy-02 (work in progress), July 2020.

[RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February 2012, <https://www.rfc-editor.org/info/rfc6513>.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

[RFC7988] Rosen, E., Ed., Subramanian, K., and Z. Zhang, "Ingress Replication Tunnels in Multicast VPN", RFC 7988, DOI 10.17487/RFC7988, October 2016, <https://www.rfc-editor.org/info/rfc7988>.

Appendix A. Illustration of a Replication Segment

This section illustrates an example of a single Replication Segment. Examples showing Replication Segment stitched together to form P2MP tree (based on SR P2MP policy) are in [I-D.voyer-pim-sr-p2mp-policy].

Consider the following topology:

```
                 R3------R6
                /         \
       R1----R2----R5-----R7
                \         /
                 +--R4---+
```

Figure 1

In this example, the Node-SID of a node Rn is N-SIDn and Adjacency-SID from node Rm to node Rn is A-SIDmn. Interface between Rm and Rn is Lmn.

Assume a Replication Segment identified with R-ID at replication node R1 and downstream Nodes R2, R6 and R7. The Replication SID at node n is R-SIDn. A packet replicated from R1 to R7 has to traverse R4.

The Replication Segment state at nodes R1, R2, R6 and R7 is shown below. Note nodes R3, R4 and R5 do not have state for the Replication Segment.

Replication Segment at R1:

```
Replication Segment <R-ID,R1>:
 Replication SID: R-SID1
 Replication State:
   R2: <R-SID2->L12>
   R6: <N-SID6, R-SID6>
   R7: <N-SID4, A-SID47, R-SID7>
```

Replication to R2 steers packet directly to R2 on interface L12. Replication to R6, using N-SID6, steers packet via IGP shortest path to that node. Replication to R7 is steered via R4, using N-SID4 and then adjacency SID A-sID47 to R7.

Replication Segment at R2:

```
Replication Segment <R-ID,R2>:
 Replication SID: R-SID2
 Replication State:
   R2: <Leaf>
```

Replication Segment at R6:

```
Replication Segment <R-ID,R6>:
 Replication SID: R-SID6
 Replication State:
   R6: <Leaf>
```

Replication Segment at R7:

```
Replication Segment <R-ID,R7>:
 Replication SID: R-SID7
 Replication State:
   R7: <Leaf>
```

When a packet is steered into the replication segment at R1:

o Since R1 is directly connected to R2, R1 performs PUSH operation with just <R-SID2> label for the replicated copy and sends it to

R2 on interface L12. R2, as Leaf, performs NEXT operation, pops R-SID2 label and delivers the payload.

- R1 performs PUSH operation with <N-SID6, R-SID6> label stack for the replicated copy to R6 and sends it to R2, the nexthop on IGP shortest path to R6. R2 performs CONTINUE operation on N-SID6 and forwards it to R3. R3 is the penultimate hop for N-SID6; it performs penultimate hop popping, which corresponds to the NEXT operation and the packet is then sent to R6 with <R-SID6> in the label stack. R6, as Leaf, performs NEXT operation, pops R-SID6 label and delivers the payload.

- R1 performs PUSH operation with <N-SID4, A-SID47, R-SID7> label stack for the replicated copy to R7 and sends it to R2, the nexthop on IGP shortest path to R4. R2 is the penultimate hop for N-SID4; it performs penultimate hop popping, which corresponds to the NEXT operation and the packet is then sent to R4 with <A-SID47, R-SID1> in the label stack. R4 performs NEXT operation, pops A-SID47, and delivers packet to R7 with <R-SID7> in the label stack. R7, as Leaf, performs NEXT operation, pops R-SID7 label and delivers the payload.

Authors' Addresses

Daniel Voyer (editor)
Bell Canada
Montreal
CA

Email: daniel.voyer@bell.ca

Clarence Filsfils
Cisco Systems, Inc.
Brussels
BE

Email: cfilsfil@cisco.com

Rishabh Parekh
Cisco Systems, Inc.
San Jose
US

Email: riparekh@cisco.com

Hooman Bidgoli
Nokia
Ottawa
CA

Email: hooman.bidgoli@nokia.com

Zhaohui Zhang
Juniper Networks

Email: zzhang@juniper.net