

Network Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: April 13, 2020

D. Voyer, Ed.  
Bell Canada  
C. Filsfils  
R. Parekh  
Cisco Systems, Inc.  
H. Bidgoli  
Nokia  
Z. Zhang  
Juniper Networks  
October 11, 2019

**Segment Routing Point-to-Multipoint Policy**  
**draft-voyer-pim-sr-p2mp-policy-00**

Abstract

This document describes an architecture to construct a Point-to-Multipoint (P2MP) tree to deliver Multi-point services in a Segment Routing domain. A SR P2MP tree is constructed by stitching a set of Replication segments together. A SR Point-to-Multipoint (SR P2MP) Policy is used to define and instantiate a P2MP tree which is computed by a PCE.

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](#)].

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 April 13, 2020.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](https://trustee.ietf.org/license-info) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- [1. Introduction](#) . . . . . [3](#)
- [2. P2MP Tree](#) . . . . . [3](#)
  - [2.1. Sharing Replication segments across P2MP trees](#) . . . . . [4](#)
- [3. SR P2MP Policy](#) . . . . . [5](#)
- [4. Using Controller to build a P2MP Tree](#) . . . . . [6](#)
  - [4.1. Provisioning SR P2MP Policy Creation](#) . . . . . [6](#)
    - [4.1.1. API](#) . . . . . [6](#)
    - [4.1.2. Invoking API](#) . . . . . [7](#)
  - [4.2. P2MP Tree Computation](#) . . . . . [7](#)
    - [4.2.1. Topology Discovery](#) . . . . . [8](#)
    - [4.2.2. Capability and Attribute Discovery](#) . . . . . [8](#)
  - [4.3. Instantiating P2MP tree on nodes](#) . . . . . [8](#)
    - [4.3.1. PCEP](#) . . . . . [8](#)
    - [4.3.2. BGP](#) . . . . . [8](#)
    - [4.3.3. NetConf](#) . . . . . [8](#)
  - [4.4. Protection](#) . . . . . [9](#)
    - [4.4.1. Local Protection](#) . . . . . [9](#)
    - [4.4.2. Path Protection](#) . . . . . [9](#)
- [5. IANA Considerations](#) . . . . . [9](#)
- [6. Security Considerations](#) . . . . . [9](#)
- [7. Acknowledgements](#) . . . . . [9](#)
- [8. Contributors](#) . . . . . [9](#)
- [9. References](#) . . . . . [10](#)
  - [9.1. Normative References](#) . . . . . [10](#)
  - [9.2. Informative References](#) . . . . . [11](#)
- [Authors' Addresses](#) . . . . . [11](#)



## 1. Introduction

A Multi-point service delivery could be realized via P2MP trees in a Segment Routing domain [[RFC8402](#)]. A P2MP tree spans from a Root node to a set of Leaf nodes via intermediate Replication nodes. It consists of a Replication segment [[I-D.voyer-spring-sr-replication-segment](#)] at the root node, one or more Replication segments at Leaf nodes and intermediate Replication nodes. The Replication segments are stitched together.

A Segment Routing P2MP policy, a variant of the SR Policy [[I-D.ietf-spring-segment-routing-policy](#)], is used to define a P2MP tree. A PCE is used to compute the tree from the Root node to the set of Leaf nodes via a set of replication nodes. The PCE then instantiates the P2MP tree in the SR domain by signaling Replication segments to Root, replication and Leaf nodes using various protocols (PCEP, BGP, NetConf etc.).

## 2. P2MP Tree

A P2MP tree in a SR domain connects a Root to a set of Leaf nodes via a set of intermediate Replication nodes. It consists of a Replication segment at the root stitched to Replication segments at intermediate Replication nodes eventually reaching the Leaf nodes.

The Replication SID of the Replication Segment at Root node is called Tree-SID. The Tree-SID SHOULD also be used as Replication SID of Replication segments at Replication and Leaf nodes. The Replication segments at Replication and Leaf nodes MAY use Replication SIDs that are not same as the Tree-SID.

The Replication segment at Root of a P2MP tree MUST be associated with that P2MP tree (i.e. <Root, Tree-ID> identifier in SR P2MP policy section below) to map a Multi-point service to the tree. A Replication segment that terminates a P2MP tree at a Leaf node MUST be associated with the P2MP tree to determine the context for a Multi-point service. The information that can be used to derive this association is specific to encoding of the protocol (PCEP, BGP, NetConf etc.) used to instantiate the Replication segment for a P2MP tree. Replication segments at intermediate Replication nodes of a tree are also associated with that tree.

A PCE MAY decide not instantiate Replication segments at Leaf nodes of a P2MP tree if it is known a priori that Multi-point services mapped to the P2MP tree can be identified using a context that is globally unique in SR domain. Multi-point service contexts assigned from "Domain-wide Common Block" (DCB) [[I-D.ietf-bess-mvpn-evpn-aggregation-label](#)] are an example of such



globally unique contexts. A Segment Routing Global Block (SRGB) [[RFC8402](#)] MAY be used to allocate globally unique Multi-point service contexts, but it is NOT RECOMMENDED to do so as the service contexts only need to be unique at service edge nodes. In this case, Replication nodes connecting to Leaf nodes SHOULD use Penultimate-Hop Pop (PHP) behavior to pop Tree-SID from a packet.

A packet steered into a P2MP tree is replicated by the Replication segment at Root node to each downstream node in the Replication segment, with the Replication SID of the Replication Segment at the downstream node. A downstream node could be a Leaf node or an intermediate Replication node. In the latter case, replication continues with the Replication segments until all Leaf nodes are reached. A packet is steered into a P2MP tree in two ways:

- o Based on a local policy-based routing at the Root node.
- o Based on steering via the Tree-SID at the Root node.

### **2.1. Sharing Replication segments across P2MP trees**

Two or more P2MP trees MAY share a Replication segment at Root or Replication nodes if at minimum as the first condition below is satisfied. A tree always has its own Replication segment at its root even if shares another Replication segment. A tree that shares another Replication segment may or may not have its own Replication segment on its Leaf nodes. If not, the second and third conditions apply to such situations.

1. The Leaf nodes reached via a shared Replication segment must be subset of Leaf or Replication nodes of the P2MP trees that shares this segment. Note if a Replication segment is shared, all its downstream Replication segments are also shared.
2. Some Multi-point services realized by the P2MP trees may need service context (e.g. packets are for certain VPNs, and/or from certain nodes). If the trees do not have their own Replication segments at their Leaf nodes then the packets transported on the P2MP trees MUST carry a service context that does not rely on the tree or root identification, e.g. a service label assigned from Domain-wide Common Block or common SRGB.
3. For some Multi-point services using P2MP trees that share Replication segments, packets transported on these trees MAY require a Tree context (e.g. MVPN Extranet [[RFC7900](#)] to avoid certain ambiguities - see [Section 2.3.1 of RFC 7900](#)). In this case, the trees MUST have their own Replication segments on the Leaf nodes. This is similar to "tunnel stacking" concept.



Sharing of a Replication segment for P2MP trees is OPTIONAL. Exact procedures to ensure validity of above conditions across PM2P services on nodes of a Segment Routing domain are outside the scope of this document.

### 3. SR P2MP Policy

The SR P2MP policy is a variant of an SR policy[I-D.ietf-spring-segment-routing-policy] and is used to instantiate SR P2MP trees.

A SR P2MP Policy is identified by the tuple <Root, Tree-ID>, where:

- o Root: The address of Root node of P2MP tree instantiated by the SR P2MP Policy
- o Tree-ID: A identifier that is unique in context of the Root. This is an unsigned 32-bit number.

A SR P2MP Policy is defined by following elements:

- o Leaf nodes: A set of nodes that terminate the P2MP trees.
- o Candidate Paths: See below.

A SR P2MP policy is provisioned on a PCE to instantiate the P2MP tree. The Tree-SID SHOULD be used as Binding SID of the P2MP policy. A PCE computes the P2MP tree and instantiates Replication segments at Root, Replication and Leaf nodes. When Replication segments are not shared across P2MP trees, the Root and Tree-ID of the SR P2MP policy are mapped to Replication-ID element of the Replication segment identifier i.e the SR Replication segment identifier is <Root, Tree-ID, Node-ID>. A shared Replication segment MAY be identified with zero Root-ID address (0.0.0.0 for IPv4 and :: for IPv6) and a Replication-ID that is unique in context of Node address where the Replication segment is instantiated when it is not associated a particular tree.

A SR P2MPPolicy has one or more Candidate paths. The active Candidate path is selected based on the tie breaking rules amongst the candidate-paths as specified in[I-D.ietf-spring-segment-routing-policy]. Each candidate path has a set of topological/resource constraints and/or optimization objectives which determine the P2MP tree for that Candidate path. Tree-SID is an identifier of the P2MP tree of the candidate path in the forwarding plane. It is instantiated in the forwarding plane at Root node, intermediate Replication nodes and Leaf nodes. The Tree-SID MAY be different at Replication and Leaf nodes.





4. Using Controller to build a P2MP Tree

A P2MP tree can be built using a Path Computation Element (PCE). This section outlines a high-level architecture for such an approach.

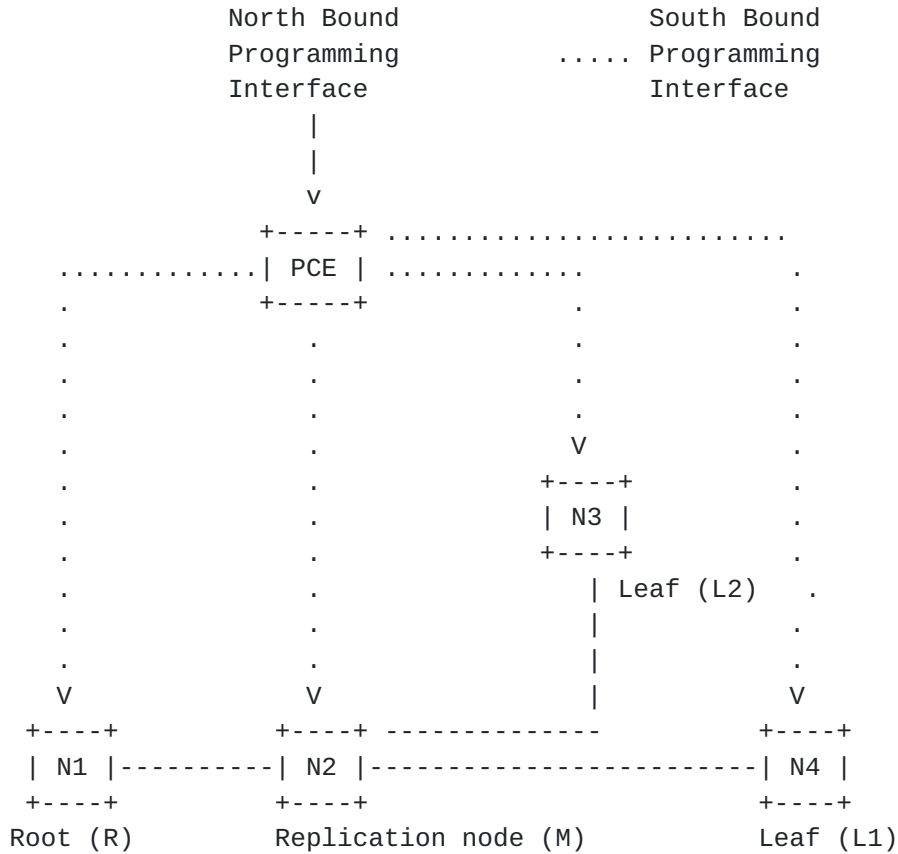


Figure 1: Centralized Control Plane Model

4.1. Provisioning SR P2MP Policy Creation

A SR P2MP policy can be instantiated and maintained in a centralized fashion using a Path Computation Element (PCE).

4.1.1. API

North-bound APIs on a PCE can be used to:

1. Create SR P2MP policy
2. Delete SR P2MP policy
3. Update SR P2MP policy



4. Create a Candidate Path for SR P2MP policy
5. Update a Candidate Path for SR P2MP policy
6. Delete a Candidate Path for SR P2MP policy

#### **4.1.2. Invoking API**

Interaction with a PCE can be via PCEP, REST, Netconf, gRPC, CLI. Yang model shall be developed for this purpose as well.

#### **4.2. P2MP Tree Computation**

An entity (an operator, a network node or a machine) provisions a SR P2MP policy by specifying the addresses of the root (R) and set of leaves {L} as well as Traffic Engineering (TE) attributes of Candidate paths via a suitable North-Bound API. The PCE computes the tree of Active candidate path. The PCE MAY compute P2MP trees for all Candidate paths., If tree computation is successful, PCE instantiates the P2MP tree(s) using Replication segments on Root, Replication, and Leaf nodes.

Candidate path constraints shall include link color affinity, bandwidth, disjointness (link, node, SRLG), delay bound, link loss, etc. Candidate path shall be optimized based on IGP or TE metric or link latency.

The Tree SID of Candidate path of a SR P2MP policy can be either dynamically allocated by the PCE or statically assigned by entity provisioning the SR P2MP policy. Ideally, same Tree-SID SHOULD be used for Replication segments at Root, Replication, and Leaf nodes. Different Tree-SIDs MAY be used at replication node(s) if it is not feasible to use same Tree SID.

A PCE can modify a P2MP tree following network element failure or in case a better path can be found based on the new network state. In this case, the PCE may want to setup the new instance of the tree and remove the old instance of the tree from the network in order to minimize traffic loss. In this case, the instances of trees for all the Candidate paths of a P2MP policy can be identified by an Instance-ID which is unique in context of the P2MP policy. As such, the identifier of non-shared Replication segments used to instantiate these trees becomes <Root-ID, Tree-ID, Node-ID, Instance-ID>.

A PCE shall be capable of computing paths across multiple IGP areas or levels as well as Autonomous Systems (ASs).



#### **4.2.1. Topology Discovery**

A PCE shall learn network topology, TE attributes of link/node as well as SIDs via dynamic routing protocols (IGP and/or BGP-LS). It may be possible for entities to pass topology information to PCE via north-bound API.

#### **4.2.2. Capability and Attribute Discovery**

It shall be possible for a node to advertise SR P2MP tree capability via IGP and/or BGP-LS. Similarly, a PCE can also advertise its P2MP tree computation capability via IGP and/or BGP-LS. Capability advertisement allows a network node to dynamically choose one or more PCE(s) to obtain services pertaining to SR P2MP policies, as well a PCE to dynamically identify SR P2MP tree capable nodes.

#### **4.3. Instantiating P2MP tree on nodes**

Once a PCE computes a P2MP tree for Candidate path of SR P2MP policy, it needs to instantiate the tree on the relevant network nodes via Replication segments. The PCE can use various protocols to program the Replication segments as described below.

##### **4.3.1. PCEP**

PCE Protocol (PCEP)has been traditionally used:

1. For a head-end to obtain paths from a PCE.
2. A PCE to instantiate SR policies.

PCEP protocol can be stateful in that a PCE can have a stateful control of an SR policy on a head-end which has delegated the control of the SR policy to the PCE. PCEP shall be extended to provision and maintain SR P2MP trees in a stateful fashion.

##### **4.3.2. BGP**

BGP has been extended to instantiate and report SR policies. It shall be extended to instantiate and maintain P2MP trees for SR P2MP policies.

##### **4.3.3. NetConf**

TBD



## **4.4. Protection**

### **4.4.1. Local Protection**

A network link, node or path on the tree of a P2MP tree can be protected using SR policies computed by PCE. The backup SR policies shall be programmed in forwarding plane in order to minimize traffic loss when the protected link/node fails. It is also possible to use node local Fast Re-Route protection mechanisms (LFA) to protect link/nodes of P2MP tree.

### **4.4.2. Path Protection**

It is possible for PCE create a disjoint backup tree for providing end-to-end path protection.

## **5. IANA Considerations**

This document makes no request of IANA.

## **6. Security Considerations**

There are no additional security risks introduced by this design.

## **7. Acknowledgements**

The authors would like to acknowledge Siva Sivabalan, Mike Koldychev and Vishnu Pavan Beeram for their valuable inputs..

## **8. 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

Tarek Saad  
Juniper Networks  
Canada

Email: tsaad@juniper.net

## **9. References**

### **9.1. Normative References**

[I-D.ietf-spring-segment-routing-policy]  
Filsfils, C., Sivabalan, S., daniel.voyer@bell.ca, d.,  
bogdanov@google.com, b., and P. Mattes, "Segment Routing  
Policy Architecture", [draft-ietf-spring-segment-routing-policy-03](#) (work in progress), May 2019.



- [I-D.voyer-spring-sr-replication-segment]  
daniel.voyer@bell.ca, d., Filsfils, C., Parekh, R.,  
Bidgoli, H., and Z. Zhang, "SR Replication Segment for  
Multi-point Service Delivery", [draft-voyer-spring-sr-  
replication-segment-00](#) (work in progress), October 2019.
- [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>>.

## 9.2. Informative References

- [I-D.ietf-bess-mvpn-evpn-aggregation-label]  
Zhang, Z., Rosen, E., Lin, W., Li, Z., and I. Wijnands,  
"MVPN/EVPN Tunnel Aggregation with Common Labels", [draft-  
ietf-bess-mvpn-evpn-aggregation-label-02](#) (work in  
progress), December 2018.
- [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>>.
- [RFC7900] Rekhter, Y., Ed., Rosen, E., Ed., Aggarwal, R., Cai, Y.,  
and T. Morin, "Extranet Multicast in BGP/IP MPLS VPNs",  
[RFC 7900](#), DOI 10.17487/RFC7900, June 2016,  
<<https://www.rfc-editor.org/info/rfc7900>>.

### 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

