Workgroup: Network Working Group
Internet-Draft:
draft-ietf-spring-cs-sr-policy-02
Published: 18 April 2024
Intended Status: Informational
Expires: 20 October 2024
Authors: C. Schmutzer, Ed.
Cisco Systems, Inc.
Z. Ali, Ed.
Cisco Systems, Inc.
P. Maheshwari
Airtel India
R. Rokui
Ciena
A. Stone
Nokia

Circuit Style Segment Routing Policies

Abstract

This document describes how Segment Routing (SR) policies can be used to satisfy the requirements for bandwidth, end-to-end recovery and persistent paths within a segment routing network. SR policies satisfying these requirements are called "circuit-style" SR policies (CS-SR policies).

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 20 October 2024.

Table of Contents

1. Introduction

Segment routing does allow for a single network to carry both typical IP (connection-less) services and connection-oriented transport services commonly referred to as "private lines". IP services typically require ECMP and TI-LFA, while transport services delivered via pseudowires (defined by the PWE3 and PALS workgroups) do require:

*Persistent end-to-end traffic engineered paths that provide predictable and identical latency in both directions

*A requested amount of bandwidth per path to ensure no impact on the Service Level Agreement (SLA) due to changing network load from other services

*Fast end-to-end protection and restoration mechanisms

*Monitoring and maintenance of path integrity

*Data plane remaining up while control plane is down

Such a "transport centric" behavior is referred to as "circuit-style" in this document.

This document describes how SR policies [I-D.ietf-spring-segment-routing-policy] and the use of adjacency-SIDs defined in the SR architecture [RFC8402] together with a stateful Path Computation Element (PCE) [RFC8231] can be used to satisfy those requirements. It includes how end-to-end recovery and path integrity monitoring can be implemented.

SR policies that satisfy those requirements are called "circuit-style" SR policies (CS-SR policies).

2. Terminology

*BSID : Binding Segment Identifier

*CS-SR : Circuit-Style Segment Routing

*ID : Identifier

*LSP : Label Switched Path

*LSPA : LSP attributes

*OAM : Operations, Administration and Maintenance

*OF : Objective Function

*PCE : Path Computation Element

*PCEP : Path Computation Element Communication Protocol

*PT : Protection Type

*SID : Segment Identifier

*SLA : Service Level Agreement

*SR : Segment Routing

*STAMP : Simple Two-Way Active Measurement Protocol

*TI-LFA : Topology Independent Loop Free Alternate

*TLV : Type Length Value

3. Reference Model

The reference model for CS-SR policies is following the Segment Routing Architecture [RFC8402] and SR Policy Architecture [I-D.ietf-spring-segment-routing-policy] and is depicted in Figure 1.

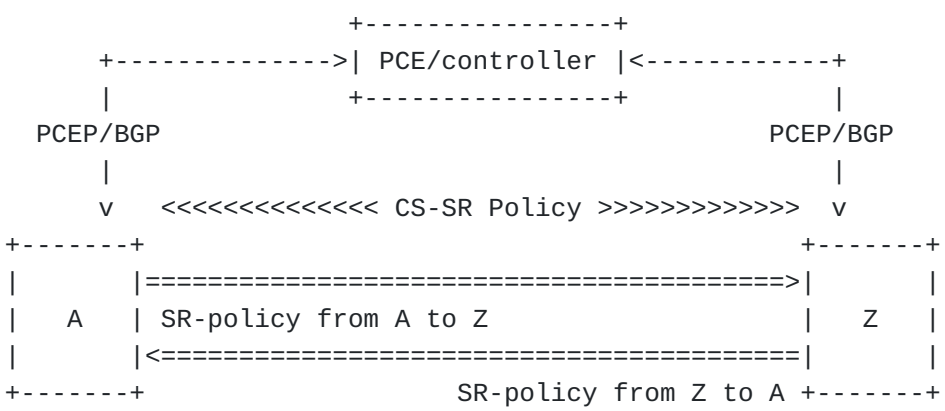

Figure 1: Circuit-style SR Policy Reference Model

By nature of CS-SR policies, paths will be computed and maintained by a centralized entity providing a consistent simple mechanism for initializing the co-routed bidirectional end to end paths, performing bandwidth allocation control, as well as monitoring facilities to ensure SLA compliance for the live of the CS-SR Policy.

When using PCEP as the communication protocol on the endpoints, the centralized entity is a stateful PCE defined in [RFC8231]. When using a MPLS data plane [RFC8660], PCEP extensions defined in [RFC8664] will be used. When using a SRv6 data plane [RFC8754], PCEP extensions defined in [I-D.ietf-pce-segment-routing-ipv6] will be used.

When using BGP as the communication protocol on the endpoints, the centralized entity does perform the same role. BGP extensions defined in [I-D.ietf-idr-segment-routing-te-policy] will be used.

In order to satisfy the requirements of CS-SR policies, each link in the topology MUST have:

*An adjacency-SID which is:

-Manually allocated or persistent : to ensure that its value does not change after a node reload

-Non-protected : to avoid any local TI-LFA protection to happen upon interface/link failures

*The bandwidth available for CS-SR policies specified

*A per-hop behavior ([RFC3246] or [RFC2597]) that ensures that the specified bandwidth is available to CS-SR policies at all times independent of any other traffic

When using a MPLS data plane [RFC8660] existing IGP extensions defined in [RFC8667] and [RFC8665] and BGP-LS defined in [RFC9085] can be used to distribute the topology information including those persistent and unprotected adjacency-SIDs.

When using a SRv6 data plane [RFC8754] the IGP extensions defined in [I-D.ietf-lsr-isis-srv6-extensions] and [I-D.ietf-lsr-ospfv3-srv6-extensions] and BGP-LS extensions in [I-D.ietf-idr-bgpls-srv6-ext] apply.

3.1. Managing Bandwidth

In a network, resources are represented by links of certain bandwidth. In a circuit switched network such as SONET/SDH, OTN or DWDM resources (timeslots or a wavelength) are allocated for a provisioned connection at the time of reservation even if no communication is present. In a packet switched network resources are only allocated when communication is present, i.e. packets are to be sent. This allows for the total reservations to exceed the link bandwidth as well in general for link congestion.

To satisfy the bandwidth requirement for CS-SR policies it must be ensured that packets carried by CS-SR policies can be at all times sent up to the reserved bandwidth on each hop along the path. This is done by:

*Firstly, CS-SR policy bandwidth reservations per link must be limited to euqal or less than the physical link bandwidth.

*Secondly, ensuring traffic for each CS-SR policy is limited to the bandwidth reserved for that CS-SR policy by traffic policing or shaping

*Thirdly, ensuring that during times of link congestion only non-CS-SR policy traffic is being buffered or dropped.

For the later several approaches can be considered:

*Allocate a dedicated physical link of bandwidth P to CS-SR policies and allow CS-SR reservations up to bandwidth C. Consider bandwidth N allocated for network control, ensure that P - N >= C

*Allocate a dedicate logical link (i.e. 801.q VLAN on ethernet) to CS-SR policies on a physical link of bandwidth P. Limit the total utilization across all other logical links to bandwidth O by traffic policing or shaping and ensure that P - N - O >= C

*Allocate a dedicated Diffserv codepoint and queue to CS-SR policies and limit the total utilization across all other queues to bandwidth O by traffic policing or shaping and ensure that P -N - O >= C

*Allocate a dedicate Diffserv codepoint and strict priority queue to CS-SR policies and limit the total utilization across all priority queues of higher or equal priority to bandwidth O by traffic policing or shaping and ensure that P - N - O >= C

*Allocate a dedicate Diffserv codepoint and a strict priority queue with a priority higher than all other queues to CS-SR policies and limit the utilization of that priority queue by traffic policing to C <= P - N

In addition CS-SR policy telemetry collection can be used to raise alarms when bandwidth utilization thresholds are passed or to request the reserved bandwidth to be adjusted.

4. CS-SR Policy Characteristics

A CS-SR policy has the following characteristics:

*Requested bandwidth : bandwidth to be reserved for the CS-SR policy

*Bidirectional co-routed : a CS-SR policy between A and Z is an association of an SR-Policy from A to Z and an SR-Policy from Z to A following the same path(s)

*Deterministic and persistent paths : segment lists with strict hops using unprotected adjacency-SIDs

*Not automatically recomputed or reoptimized : the SID list of a candidate path must not change automatically to a SID list representing a different path (for example upon topology change)

*Multiple candidate paths in case of protection/restoration:

-Following the SR policy architecture, the highest preference valid path is carrying traffic

-Depending on the protection/restoration scheme (Section 6), lower priority candidate paths

omay be pre-computed

omay be pre-programmed

omay have to be disjoint

*Connectivity verification and performance measurement is activated on each candidate path (Section 7)

5. CS-SR Policy Creation

5.1. Policy Creation when using PCEP

Considering the scenario illustrated in Figure 1 a CS-SR policy between A and Z is configured both on A (with Z as endpoint) and Z (with A as endpoint).

Both nodes A and Z act as PCC and delegate path computation to the PCE using PCEP with the extensions defined in [RFC8664] and the procedure described in Section 5.7.1 of [RFC8231]. The PCRpt message sent from the headends to the PCE contains the following parameters:

*BANDWIDTH object (Section 7.7 of [RFC5440]) : to indicate the requested bandwidth

*LSPA object (section 7.11 of [RFC5440]) : to indicate that no local protection requirements

-L flag set to 0 : no local protection

-E flag set to 1 : protection enforcement (section 5 of [I-D.ietf-pce-local-protection-enforcement])

*ASSOCIATION object ([RFC8697]) :

-Type : Double-sided Bidirectional with Reverse LSP Association ([I-D.ietf-pce-sr-bidir-path])

-Bidirectional Association Group TLV ([RFC9059]) :

oR flag is always set to 0 (forward path)

oC flag is always set to 1 (co-routed)

If the SR-policies are configured with more than one candidate path, a PCEP request is sent per candidate path. Each PCEP request does include the "SR Policy Association" object (type 6) as defined in

[I-D.ietf-pce-segment-routing-policy-cp] to make the PCE aware of the candidate path belonging to the same policy.

The signaling extensions described in [I-D.sidor-pce-circuit-style-pcep-extensions] are used to ensure that

- Path determinism is achieved by the PCE only using segment lists representing a strict hop by hop path using unprotected adjacency-SIDs.

- Path persistency across node reloads in the network is achieved by the PCE only including manually configured adjacency-SIDs in its path computation response.

- Persistency across network changes is achieved by the PCE not performing periodic nor network event triggered re-optimization.

Bandwidth adjustment can be requested after initial creation by signaling both requested and operational bandwidth in the BANDWIDTH object but the PCE is not allowed to respond with a changed path.

As discussed in section 3.2 of [I-D.ietf-pce-multipath] it may be necessary to use load-balancing across multiple paths to satisfy the bandwidth requirement of a candidate path. In such a case the PCE will notify the PCC to install multiple segment lists using the signaling procedures described in section 5.3 of [I-D.ietf-pce-multipath].

5.2. Policy Creation when using BGP

Again considering the scenario illustrated in Figure 1, there is no CS-SR policy configuration required on A nor Z in order to create the CS-SR policy between A and Z.

The centralized controller is instructed (i.e. by an application via a API call) to create the CS-SR policy, for which the controller does perform path computation and is requesting A via BGP to instante a SR-policy (with Z as endpoint) and requesting Z via BGP to instantiate a SR-policy (with Z as endpoint).

To instantiate the SR-policies in A and Z the BGP extensions defined in [I-D.ietf-idr-segment-routing-te-policy] are used.

No signaling extensions are required for the following:

- Path determinism is achieved by the controller only using segment lists representing a strict hop by hop path using unprotected adjacency-SIDs.

*Path persistency across node reloads in the network is achieved by the controller only including manually configured adjacency-SIDs in its path computation response.

*Persistency across network changes is achieved by the controller not performing periodic nor network event triggered re-optimization.

If there are more than one candidate paths per SR-policy required, multiple NLRIs with different distinguisher values (see section 2.1 of [I-D.ietf-idr-segment-routing-te-policy]) have to be included in the BGP UPDATE message.

To achieve load-balancing across multiple paths to satisfy the bandwidth requirement of a candidate path, multiple Segment List Sub-TLVs have to be included in the SR Policy Sub-TLV. See section 2.1 of [I-D.ietf-idr-segment-routing-te-policy]

The endpoints A and Z report the SR-policy states back to the centralized controller via BGP-LS using the extension defined in [I-D.ietf-idr-bgp-ls-sr-policy].

5.3. Maximum Segment Depth

A Segment Routed path defined by a segment list is constrained by maximum segment depth (MSD), which is the maximum number of segments a router can impose onto a packet. [RFC8491], [RFC8476], [RFC8814] and [RFC8664] provide the necessary capabilities for a PCE to determine the MSD capability of a router. The MSD constraint is typically resolved by leveraging a label stack reduction technique, such as using Node SIDs and/or BSIDs (SR architecture [RFC8402]) in a segment list, which represents one or many hops in a given path.

As described in Section 4, adjacency-SIDs without local protection are to be used for CS-SR policies to ensure no ECMP, no rerouting due to topological changes nor localized protection is being invoked on the traffic, as the alternate path may not be providing the desired SLA.

If a CS-SR Policy path requires SID List reduction, a Node SID cannot be utilized as it is eligible for traffic rerouting following IGP re-convergence. However, a BSID can be programmed to a transit node, if the following requirements are met:

*The BSID is unprotected, hence only has one candidate path

*The BSID follows the rerouting and optimization characteristics defined in Section 4 which implies the SID list of the candidate path MUST only use unprotected adjacency-SIDs.

This ensures that any CS-SR policies in which the BSID provides transit for do not get rerouted due to topological changes or protected due to failures. A BSID may be pre-programmed in the network or automatically injected in the network by a PCE.

6. Recovery Schemes

Various protection and restoration schemes can be implemented. The terms "protection" and "restoration" are used with the same subtle distinctions outlined in section 1 of [RFC4872], [RFC4427] and [RFC3386] respectively.

- *Protection : another candidate path is computed and fully established in the data plane and ready to carry traffic

- *Restoration : a candidate path may be computed and may be partially established but is not ready to carry traffic

The term "failure" is used to represent both "hard failures" such complete loss of connectivity detected by Section 7.1 or degradation, a packet loss ratio, beyond a configured acceptable threshold.

6.1. Unprotected

In the most basic scenario no protection nor restoration is required. The CS-SR policy has only one candidate path configured. This candidate path is established, activated and is carrying traffic.

When using PCEP, a PCRpt message is sent from the PCC to the PCE with the O field in the LSP object set to 2.

When using BGP, a BGP-LS update with a SR Policy Candidate Path NLRI is sent from the endpoint to the centralized controller having

- *C flag set to 1 to indicate the candidate path was provisioned by the controller

- *A flag set to 1 to indicate the candidate path is active and carrying traffic

In case of a failure along the path the CS-SR policy will go down and traffic will not be recovered.

Typically two CS-SR policies are deployed either within the same network with disjoint paths or in two completely separate networks and the overlay service is responsible for traffic recovery.

6.2. 1:1 Protection

For fast recovery against failures the CS-SR policy has two candidate paths. Both paths are established but only the candidate with higher preference is activated and is carrying traffic.

When using PCEP, the PCRpt message for the candidate path with higher preference will have the O field in the LSP object set to 2. For the candidate path with the lower preference the O field in the LSP object is set to 1.

Appropriate routing of the protect path diverse from the working path can be requested from the PCE by using the "Disjointness Association" object (type 2) defined in [RFC8800] in the PCRpt messages. The disjoint requirements are communicated in the "DISJOINTNESS-CONFIGURATION TLV"

- *L bit set to 1 for link diversity
- *N bit set to 1 for node diversity
- *S bit set to 1 for SRLG diversity
- *T bit set to enforce strict diversity

The P bit may be set for first candidate path to allow for finding the best working path that does satisfy all constraints without considering diversity to the protect path.

The "Objective Function (OF) TLV" as defined in section 5.3 of [RFC8800] may also be added to minimize the common shared resources.

When using BGP, the controller is already aware of the disjoint requirements and does consider them while computing both paths. Two NLRIs with different distinguisher values and different preference values are included in the BGP UPDATE sent to the headend routers.

A BGP-LS update is sent to the controller with a SR Policy Candidate Path NLRI for the candidate path with higher preference with

- *C flag set to 1 to indicate that candidate path was provisioned by the controller
- *A flag set to 1 to indicate the candidate path is active and carrying traffic

and another SR Policy Candidate Path NLRI for the candidate path with lower preference with

- *C flag set to 1 to indicate the candidate path was provisioned by the controller

- *B flag set to 1 to indicate the role of backup path

Upon a failure impacting the candidate path with higher preference carrying traffic, the candidate path with lower preference is activated immediately and traffic is now sent across it.

When using PCEP a PCRpt message for the higher preference candidate path is sent to the PCE with the O field changed from 2 to 0 and a PCRpt message for the lower preference candidate path with the O field change from 1 to 2.

When using BGP a BGP-LS update is sent to the controller with a SR Policy Candidate Path NLRI for the candidate path with higher preference with the A flag cleared and for another BGP-LS update for the candidate path with lower preference with the B flag cleared and A flag set to 1.

Protection switching is bidirectional. As described in Section 7.1, both headends will generate and receive their own loopback mode test packets, hence even a unidirectional failure will always be detected by both headends without protection switch coordination required.

6.2.1. Reversion

Two cases are to be considered when the failure(s) impacting a candidate path with higher preference are cleared:

- *Revertive switching : re-activate the highere preference candidate path and start sending traffic over it

- *Non-revertive switching : do not activate the higher preference candidate path and keep sending traffic via the lower preference candidate path

When using PCEP, for revertive switching a PCRpt message for the recovered higher preference candidate path is sent to the PCE with the O field changed from 0 to 2 and send a PCRpt message for the lower preference candidate path with the O field changed from 2 to 1. For non-revertive switching only a PCRpt message for the recovered higher preference candidate path with the O field set to 1 is sent.

When using BGP and revertive switching a BGP-LS update is sent to the controller with a SR Policy Candidate Path NLRI for the recovered higher preference candidate path with the A flag set to 1 and another

BGP-LS update for the lower preference candidate path with the A flag cleared and B flag set to 1. For non-revertive switching only a BGP-LS update for the higher preference candidate path with the B flag set to 1 is sent.

6.3. Restoration

6.3.1. 1+R Restoration

Compared to 1:1 protection described in Section 6.2, this restoration scheme avoids pre-allocating protection bandwidth in steady state, while still being able to recover traffic flow in case of a network failure in a deterministic way (maintain required bandwidth commitment)

When using PCEP, the CS-SR policy is configured with two candidate paths. The candidate path with higher preference is established, activated (O field in LSP object is set to 2) and is carrying traffic.

The second candidate path with lower preference is only established and activated (PCRpt message to the PCE with O field in LSP object is set to 2) upon a failure impacting the first candidate path in order to send traffic over an alternate path through the network around the failure with potentially relaxed constraints but still satisfying the bandwidth commitment.

The second candidate path is generally only requested from the PCE and activated after a failure, but may also be requested and pre-established during CS-SR policy creation with the downside of bandwidth being set aside ahead of time.

As soon as failure(s) that brought the first candidate path down are cleared, the second candidate path is getting deactivated (PCRpt message to the PCE with O field in LSP object is set to 1) or torn down. The first candidate path is activated (PCRpt message to the PCE with O field in LSP object is set to 2) and traffic sent across it.

When using BGP, the controller does compute one path and does include one NLRI in the BGP UPDATE message sent to the headend routers to instantiate the CS-SR policy with one candidate path active and carrying traffic.

A BGP-LS update with a SR Policy Candidate Path NLRI is sent to the controller with

* C flag set to 1 to indicate the candidate path was provisioned by the controller

*A flag set to 1 to indicate the candidate path is active and carrying traffic

Upon the controller detecting the failure of the CS-SR policy's candidate path, another path is computed and added as second candidate path to the CS-SR policy by sending a BGP UPDATE message to the headend routers with a NLRI distinguisher value being different and preference being lower compared to the first candidate path.

A BGP-LS update with a SR Policy Candidate Path NLRI for the candidate path with higher preference is sent to the controller with

*A flag is cleared to indicate the candidate path is no longer active and not carrying traffic anymore

and another SR Policy Candidate Path NLRI for the candidate path with lower preference with

*C flag set to 1 to indicate the candidate path was provisioned by the controller

*A flag set to 1 to indicate the candidate path is active and carrying traffic

The second candidate path is generally only instantiated by the controller and activated after a failure, but may also be instantiated and pre-established during CS-SR policy creation with the downside of bandwidth being set aside ahead of time. If so, a BGP-LS update with a SR Policy Candidate Path NLRI is sent to the controller with

*C flag set to 1 to indicate the candidate path was provisioned by the controller

*B flag set to 1 to indicate the role of backup path

Once the controller has detected the failure(s) that brought the first candidate path down are cleared, a BGP-LS update with a SR Policy Candidate Path NLRI for the first candidate path is sent to the controller with

*A flag set to 1 to indicate the candidate path became active and is carrying traffic again

The second candidate path is getting removed by a BGP UPDATE message withdrawing the NLRI of the second candidate path.

Restoration and reversion behavior is bidirectional. As described in Section 7.1, both headends use connectivity verification in loopback mode and therefore even in case of unidirectional failures both

headends will detect the failure or clearance of the failure and switch traffic away from the failed or to the recovered candidate path.

6.3.2. 1:1+R Restoration

For further resiliency in case of multiple concurrent failures that could affect both candidate paths of 1:1 protection described in Section 6.2, a third candidate path with a preference lower than the other two candidate paths is added to the CS-SR policy to enable restoration.

When using PCEP, the third candidate path will generally only be established, activated (PCRpt message to the PCE with O field in LSP object is set to 2) and carry traffic after failure(s) have impacted both the candidate path with highest and second highest preference.

The third candidate path may also be requested and pre-computed already whenever either the first or second candidate path went down due to a failure with the downside of bandwidth being set aside ahead of time.

As soon as failure(s) that brought either the first or second candidate path down are cleared the third candidate path is getting deactivated (PCRpt message to the PCE with O field in LSP object is set to 1), the candidate path that recovered is activated (PCRpt message to the PCE with O field in LSP object is set to 2) and traffic sent across it.

When using BGP, the third candidate path will generally only be instantiated by the controller and activated after failure(s) have impacted both the candidate path with highest and second highest preference, but may also be instantiated and pre-established during CS-SR policy creation with the downside of bandwidth being set aside ahead of time.

Assuming the case where both candidate paths are down, a BGP-LS update is sent with SR Policy Candidate Path NLRIs for the first and second candidate path with

- A flag cleared

and a SR Policy Candidate Path NLRI for the third candidate path with

- C flag set to 1 to indicate the candidate path was provisioned by the controller

- A flag set to 1 to indicate the candidate path is active and carrying traffic

Assuming the case where only one candidate path is down, a BGP-LS update is sent with a SR Policy Candidate Path NLRI for the failed candidate path with

- *A flag cleared

a SR Policy Candidate Path NLRI for the second candidate path with

- *A flag set to 1 to indicate it is active and carrying traffic network

and another SR Policy Candidate Path NLRI for the newly installed third candidate path with

- *C flag set to 1 to indicate the candidate path was provisioned by the controller
- *B flag set to 1 to indicate the role of backup path

Once the controller has detected the failure(s) that brought either the first or the second candidate path down are cleared, a BGP-LS update with a SR Policy Candidate Path NLRI for the recovered candidate path is sent to the controller with

- *A flag set to 1 to indicate the candidate path became active and is carrying traffic again

The third candidate path is getting removed by a BGP UPDATE message withdrawing the NLRI of the third candidate path.

Again restoration and reversion behavior is bidirectional. As described in Section 7.1, both headends use connectivity verification in loopback mode and therefore even in case of unidirectional failures both headends will detect the failure or clearance of the failure and switch traffic away from the failed or to the recovered candidate path.

7. Operations, Administration, and Maintenance (OAM)

7.1. Connectivity Verification

The proper operation of each segment list is validated by both headends using STAMP in loopback measurement mode as described in section 4.2.3 of [I-D.ietf-spring-stamp-srpm].

As the STAMP test packets are including both the segment list of the forward and reverse path, standard segment routing data plane operations will make those packets get switched along the forward path to the tailend and along the reverse path back to the headend.

When using PCEP, the headend forms the bidirectional SR Policy association using the procedure described in [I-D.ietf-pce-sr-bidir-path] and receives the information about the reverse segment list from the PCE as described in section 4.5 of [I-D.ietf-pce-multipath]

When using BGP, the controller does inform the headend routers about the reverse segment list using the Reverse Segment List Sub-TLV defined in section 4.1 of [I-D.ietf-idr-sr-policy-path-segment].

7.2. Performance Measurement

The same STAMP session is used to estimate round-trip loss as described in section 5 of [I-D.ietf-spring-stamp-srpm].

The same STAMP session used for connectivity verification can be used to measure delay. As loopback mode is used only round-trip delay is measured and one-way has to be derived by dividing the round-trip delay by two.

7.3. Candidate Path Validity Verification

A stateful PCE/controller is in sync with the network topology and the CS-SR Policies provisioned on the headend routers. As described in Section 4 a path must not be automatically recomputed after or optimized for topology changes. However there may be a requirement for the stateful PCE/controller to tear down a path if the path no longer satisfies the original requirements, detected by stateful PCE/controller, such as insufficient bandwidth, diversity constraint no longer met or latency constraint exceeded.

The headend may measure the actual bandwidth utilization of a CS-SR policy to take local action and/or report it as requested bandwidth via PCEP or BGP-LS to the stateful PCE/controller. Typical actions are raising alarms or adjusting the reserved bandwidth.

For a CS-SR policy configured with multiple candidate paths, a headend may switch to another candidate path if the stateful PCE/controller decided to tear down the active candidate path.

8. External Commands

8.1. Candidate Path Switchover

It is very common to allow operators to trigger a switch between candidate paths even if no failure is present. I.e. to proactively drain a resource for maintenance purposes. Operator triggered switching between candidate paths is unidirectional and has to be requested on both headends.

8.2. Candidate Path Re-computation

While no automatic re-optimization or pre-computation of CS-SR policy candidate paths is allowed as specified in Section 4, network operators trying to optimize network utilization may explicitly request a candidate path to be re-computed at a certain point in time.

9. Security Considerations

TO BE ADDED

10. IANA Considerations

This document has no IANA actions.

11. Acknowledgements

The author's want to thank Samuel Sidor, Mike Koldychev, Rakesh Gandhi and Tarek Saad for providing their review comments and all contributors for their inputs and support.

12. References

12.1. Normative References

[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/rfc/rfc2119>.

12.2. Informative References

[I-D.ietf-idr-bgp-ls-sr-policy] Previdi, S., Talaulikar, K., Dong, J., Gredler, H., and J. Tantsura, "Advertisement of Segment Routing Policies using BGP Link-State", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-ls-sr-policy-04, 20 March 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-ls-sr-policy-04>.

[I-D.ietf-idr-bgpls-srv6-ext] Dawra, G., Filsfils, C., Talaulikar, K., Chen, M., Bernier, D., and B. Decraene, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing over IPv6 (SRv6)", Work in Progress, Internet-Draft, draft-ietf-idr-bgpls-srv6-ext-14, 17 February 2023, <https://

datatracker.ietf.org/doc/html/draft-ietf-idr-bgpls-srv6-ext-14>.

[I-D.ietf-idr-segment-routing-te-policy]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-segment-routing-te-policy-26, 23 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-segment-routing-te-policy-26>.

[I-D.ietf-idr-sr-policy-path-segment] Li, C., Li, Z., Yin, Y., Cheng, W., and K. Talaulikar, "SR Policy Extensions for Path Segment and Bidirectional Path", Work in Progress, Internet-Draft, draft-ietf-idr-sr-policy-path-segment-09, 19 February 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-policy-path-segment-09>.

[I-D.ietf-lsr-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extensions to Support Segment Routing over the IPv6 Data Plane", Work in Progress, Internet-Draft, draft-ietf-lsr-isis-srv6-extensions-19, 14 November 2022, <https://datatracker.ietf.org/doc/html/draft-ietf-lsr-isis-srv6-extensions-19>.

[I-D.ietf-lsr-ospfv3-srv6-extensions] Li, Z., Hu, Z., Talaulikar, K., and P. Psenak, "OSPFv3 Extensions for Segment Routing over IPv6 (SRv6)", Work in Progress, Internet-Draft, draft-ietf-lsr-ospfv3-srv6-extensions-15, 21 June 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-lsr-ospfv3-srv6-extensions-15>.

[I-D.ietf-pce-local-protection-enforcement] Stone, A., Aissaoui, M., Sidor, S., and S. Sivabalan, "Local Protection Enforcement in the Path Computation Element Communication Protocol (PCEP)", Work in Progress, Internet-Draft, draft-ietf-pce-local-protection-enforcement-11, 23 June 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-pce-local-protection-enforcement-11>.

[I-D.ietf-pce-multipath]
Koldychev, M., Sivabalan, S., Saad, T., Beeram, V. P., Bidgoli, H., Yadav, B., Peng, S., and G. S. Mishra, "PCEP Extensions for Signaling Multipath Information", Work in Progress, Internet-Draft, draft-ietf-pce-multipath-11, 8 April 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-pce-multipath-11>.

[I-D.ietf-pce-segment-routing-ipv6]
Li, C., Kaladharan, P., Sivabalan, S., Koldychev, M., and Y. Zhu, "Path Computation Element Communication Protocol (PCEP) Extensions for IPv6 Segment Routing", Work in Progress, Internet-Draft, draft-ietf-pce-segment-routing-ipv6-25, 4 April 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-pce-segment-routing-ipv6-25>.

[I-D.ietf-pce-segment-routing-policy-cp]
Koldychev, M., Sivabalan, S., Barth, C., Peng, S., and H. Bidgoli, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing (SR) Policy Candidate Paths", Work in Progress, Internet-Draft, draft-ietf-pce-segment-routing-policy-cp-15, 17 March 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-pce-segment-routing-policy-cp-15>.

[I-D.ietf-pce-sr-bidir-path] Li, C., Chen, M., Cheng, W., Gandhi, R., and Q. Xiong, "Path Computation Element Communication Protocol (PCEP) Extensions for Associated Bidirectional Segment Routing (SR) Paths", Work in Progress, Internet-Draft, draft-ietf-pce-sr-bidir-path-13, 13 February 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-pce-sr-bidir-path-13>.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", Work in Progress, Internet-Draft, draft-ietf-spring-segment-routing-policy-22, 22 March 2022, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-segment-routing-policy-22>.

[I-D.ietf-spring-stamp-srpm] Gandhi, R., Filsfils, C., Voyer, D., Chen, M., and R. F. Foote, "Performance Measurement Using Simple Two-Way Active Measurement Protocol (STAMP) for Segment Routing Networks", Work in Progress, Internet-Draft, draft-ietf-spring-stamp-srpm-14, 4 April 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-stamp-srpm-14>.

[I-D.sidor-pce-circuit-style-pcep-extensions]
Sidor, S., Maheshwari, P., Stone, A., Jalil, L., and S. Peng, "PCEP extensions for Circuit Style Policies", Work in Progress, Internet-Draft, draft-sidor-pce-circuit-style-pcep-extensions-06, 15 December 2023, <https://datatracker.ietf.org/doc/html/draft-sidor-pce-circuit-style-pcep-extensions-06>.

[RFC1925] Callon, R., "The Twelve Networking Truths", RFC 1925, DOI 10.17487/RFC1925, April 1996, <https://www.rfc-editor.org/rfc/rfc1925>.

[RFC2597] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, DOI 10.17487/RFC2597, June 1999, <https://www.rfc-editor.org/rfc/rfc2597>.

[RFC3246] Davie, B., Charny, A., Bennet, J.C.R., Benson, K., Le Boudec, J.Y., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", RFC 3246, DOI 10.17487/RFC3246, March 2002, <https://www.rfc-editor.org/rfc/rfc3246>.

[RFC3386] Lai, W., Ed. and D. McDysan, Ed., "Network Hierarchy and Multilayer Survivability", RFC 3386, DOI 10.17487/RFC3386, November 2002, <https://www.rfc-editor.org/rfc/rfc3386>.

[RFC4427] Mannie, E., Ed. and D. Papadimitriou, Ed., "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4427, DOI 10.17487/RFC4427, March 2006, <https://www.rfc-editor.org/rfc/rfc4427>.

[RFC4872] Lang, J.P., Ed., Rekhter, Y., Ed., and D. Papadimitriou, Ed., "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, DOI 10.17487/RFC4872, May 2007, <https://www.rfc-editor.org/rfc/rfc4872>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <https://www.rfc-editor.org/rfc/rfc5440>.

[RFC8231] Crabbe, E., Minei, I., Medved, J., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", RFC 8231, DOI 10.17487/RFC8231, September 2017, <https://www.rfc-editor.org/rfc/rfc8231>.

[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/rfc/rfc8402>.

[RFC8476] Tantsura, J., Chunduri, U., Aldrin, S., and P. Psenak, "Signaling Maximum SID Depth (MSD) Using OSPF", RFC 8476, DOI 10.17487/RFC8476, December 2018, <https://www.rfc-editor.org/rfc/rfc8476>.

[RFC8491] Tantsura, J., Chunduri, U., Aldrin, S., and L. Ginsberg, "Signaling Maximum SID Depth (MSD) Using IS-IS", RFC 8491, DOI 10.17487/RFC8491, November 2018, <https://www.rfc-editor.org/rfc/rfc8491>.

[RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <https://www.rfc-editor.org/rfc/rfc8660>.

[RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <https://www.rfc-editor.org/rfc/rfc8664>.

[RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <https://www.rfc-editor.org/rfc/rfc8665>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/rfc/rfc8667>.

[RFC8697] Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "Path Computation Element Communication Protocol (PCEP) Extensions for Establishing Relationships between Sets of Label Switched Paths (LSPs)", RFC 8697, DOI 10.17487/RFC8697, January 2020, <https://www.rfc-editor.org/rfc/rfc8697>.

[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/rfc/rfc8754>.

[RFC8800] Litkowski, S., Sivabalan, S., Barth, C., and M. Negi, "Path Computation Element Communication Protocol (PCEP) Extension for Label Switched Path (LSP) Diversity Constraint Signaling", RFC 8800, DOI 10.17487/RFC8800, July 2020, <https://www.rfc-editor.org/rfc/rfc8800>.

[RFC8814] Tantsura, J., Chunduri, U., Talaulikar, K., Mirsky, G., and N. Triantafillis, "Signaling Maximum SID Depth (MSD) Using the Border Gateway Protocol - Link State", RFC 8814, DOI 10.17487/RFC8814, August 2020, <https://www.rfc-editor.org/rfc/rfc8814>.

[RFC9059] Gandhi, R., Ed., Barth, C., and B. Wen, "Path Computation Element Communication Protocol (PCEP) Extensions for Associated Bidirectional Label Switched Paths (LSPs)", RFC 9059, DOI 10.17487/RFC9059, June 2021, <https://www.rfc-editor.org/rfc/rfc9059>.

[RFC9085] Previdi, S., Talaulikar, K., Ed., Filsfils, C., Gredler, H., and M. Chen, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing", RFC 9085, DOI 10.17487/RFC9085, August 2021, <https://www.rfc-editor.org/rfc/rfc9085>.

Contributors

Daniel Voyer
Bell Canada

Email: daniel.voyer@bell.ca

Luay Jalil
Verizon

Email: luay.jalil@verizon.com

Shuping Peng
Huawei Technologies

Email: pengshuping@huawei.com

Clarence Filsfils
Cisco Systems, Inc.

Email: cfilsfil@cisco.com

Francois Clad

Cisco Systems, Inc.

Email: fclad@cisco.com

Tarek Saad
Cisco Systems, Inc.

Email: tsaad.net@gmail.com

Brent Foster
Cisco Systems, Inc.

Email: brfoster@cisco.com

Bertrand Duvivier
Cisco Systems, Inc.

Email: bduvivie@cisco.com

Stephane Litkowski
Cisco Systems, Inc.

Email: slitkows@cisco.com

Jie Dong
Huawei Technologies

Email: jie.dong@huawei.com

Authors' Addresses

Christian Schmutzer (editor)
Cisco Systems, Inc.

Email: cschmutz@cisco.com

Zafar Ali (editor)
Cisco Systems, Inc.

Email: zali@cisco.com

Praveen Maheshwari
Airtel India

Email: Praveen.Maheshwari@airtel.com

Reza Rokui
Ciena

Email: rrokui@ciena.com

Andrew Stone
Nokia

Email: andrew.stone@nokia.com