Workgroup: IDR Working Group
Internet-Draft:
draft-zhou-idr-bgp-srmpls-elp-03
Published: 30 August 2021
Intended Status: Standards Track
Expires: 3 March 2022
Authors: Yao. Liu   Shaofu. Peng
         ZTE Corp.   ZTE Corp.

BGP Extension for SR-MPLS Entropy Label Position

Abstract

   This document proposes extensions for BGP to indicate the entropy
   label position in the SR-MPLS label stack when delivering SR Policy
   via BGP.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Segment Routing (SR) leverages the source routing paradigm. Segment
Routing can be instantiated on MPLS data plane which is referred to
as SR-MPLS [RFC8660]. SR-MPLS leverages the MPLS label stack to
construct the SR path.

Entropy labels (ELs) [RFC6790] are used in the MPLS data plane to
provide entropy for load-balancing. The idea behind the entropy
label is that the ingress router computes a hash based on several
fields from a given packet and places the result in an additional
label named "entropy label". Then, this entropy label can be used as
part of the hash keys used by an LSR. Using the entropy label as
part of the hash keys reduces the need for deep packet inspection in
the LSR while keeping a good level of entropy in the load-balancing.

[RFC8662] proposes to use entropy labels for SR-MPLS networks and
multiple < ELI, EL> pairs may be inserted in the SR-MPLS label
stack. The ingress node may decide the number and position of the
ELI/ELs which need to be inserted into the label stack, that is
termed as ELP (Entropy Label Position) in this document. But in some
cases,the the controller (e.g. PCE) can be used to perform the TE
path computation as well as the Entropy Label Position which is
useful for inter-domain scenarios.

[I-D.ietf-idr-segment-routing-te-policy] specifies the way to use
BGP to distribute one or more of the candidate paths of an SR Policy
to the headend of that policy.

This document proposes extensions for BGP to indicate the ELP in the
segment list when delivering SR Policy via BGP in SR-MPLS networks.

2. Conventions used in this document

2.1. 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.

2.2. Terminology and Acronyms

EL: Entropy Label

ELI: Entropy Label Indicator

ELC: Entropy Label Capability

ERLD: Entropy Readable Label Depth

ELP: Entropy Label Position

MSD: Maximum SID Depth

3. Entropy Label Position in SR-MPLS with the Controller

As described in [RFC8662] section 7, ELI/EL placement is not an easy decision, multiple criteria may be taken into account.

First is the Maximum SID Depth (MSD), it defines the maximum number of labels that a particular node can impose on a packet, and it is a limit when the ingress node imposing ELI/EL pairs on the SR label stack.

The Entropy Readable Label Depth(ERLD) value is an important parameter to consider when inserting an ELI/EL. The ERLD is defined as the number of labels a router can both read in an MPLS packet received on its incoming interface(s) and use in its load-balancing function. An ELI/EL pair must be within the ERLD of the LSR in order for the LSR to use the EL during load-balancing. It's necessary to get the ERLD of the nodes along the SR path to achieve efficient load-balancing.

An implementation MAY try to evaluate if load-balancing is really expected at a particular node based on the segment type of its label, which also influences the ELP of a segment list.

Other criteria includes maximizing number of LSRs that will load-balance, preference for a part of the path, and etc. Using which

criteria and how to decide the ELP based on the criteria is a matter of implementation.

As shown in Figure 1, in the inter-domain scenario, a path from A to Z is required, a centralized controller performs the computation of the end-to-end path, along which traffic load-balancing is required.

```
....................   ....................    .....................
.                 .   .                 .    .                   .
.+---+     +---+ .   . +---+     +---+ .    .+---+     +----+   .
.| A |-------| B |------ | C |------| X |-------| Y |------| Z  |   .
.+---+     +---+ .   . +---+     +---+ .    .+---+     +----+   .
.    domain 1    .   .     domain 2    .    .     domain 3      .
....................   ....................    .....................
```
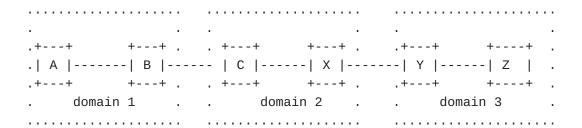
  Figure 1: Figure 1: Entropy Labels in SR-MPLS Inter-Domain Scenario

 When the headend node in the first domain can't get the information of the nodes/SIDs in other domains, e.g, the ERLD of each node or the type of the SID bounded to a node/link, it's difficult for the headend node to decide the ELP of the segment list for the path.

 Performing the computation of the ELP by the controller is an alternate, since it's easier for the controller to get the required information along the segment list prescribed by itself.

 For example, the ERLD value can be advertised via IS-IS[I-D.ietf-isis-mpls-elc] and OSPF[I-D.ietf-ospf-mpls-elc] with the domain, in each domain, one or more nodes are configured with BGP-LS so the controller can get the ERLD value of all the nodes through BGP-LS[RFC9085]. The controller can acquire the MSD of the headend node or the Binding SID anchor node via BGP-LS[RFC8814] or PCEP[RFC8664].

 Another benefit of utilizing the controller to calculate ELP is that if the criteria or calculation algorithm is changed, the corresponding modification only needs to be made on the controller instead of each headend node in the network.

 When the controller performs the computation of the the ELP for a segment list, the considerations for the placement of ELI/ELs introduced in [RFC8662] are still applicable. How the controller computes the ELP is out of scope of the document.

 After the ELP of an SR path is decided, the controller SHOULD inform the result to the headend node of the path, so the node knows where to insert the ELI/ELs when needed. Section 4 proposes the detailed extensions for BGP to carry this information.

4. BGP Extensions for ELP in SR Policy

The Segment Flags for Segment Sub-TLVs are defined in Section
2.4.4.2.12 of [I-D.ietf-idr-segment-routing-te-policy]. In this
document, the ELP information is transmitted by extending the flags
of Segment Sub-TLVs.

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|V|A|S|B|E|     |
+-+-+-+-+-+-+-+-+
```

E-Flag: This flag, when set, indicates that presence of < ELI, EL>
label pairs which are inserted after this segment. E-Flag is
applicable to Segment Types A, C, D, E, F, G and H. If E-Flag
appears with Segment Types B, I, J and K, it MUST be ignored.

5. Operations

Node A receives an SR Policy NLRI with an Segment List sub-TLV from
the controller. The Segment List sub-TLV contains multiple Segment
sub-TLVs, e.g, <S1, S2, S3, S4, S5, S6>, the E-Flags of S3 and S6
are set, it indicates that if load-balancing is required, two <ELI,
EL> pairs SHOULD be inserted into the label stack of the SR-TE
forwarding entry, respectively after the Label for S3 and Label for
S6.

The value of EL is supplemented by the ingress node according to
load-balancing function of the appropriate keys extracted from a
given packet. After inserting ELI/ELs, the label stack on the
ingress node would be <S1, S2, S3, ELI, EL, S4, S5, S6, ELI, EL>.

6. Security Considerations

Procedures and protocol extensions defined in this document do not
introduce any new security considerations beyond those already
listed in [RFC8662] and [I-D.ietf-idr-segment-routing-te-policy].

7. IANA Considerations

This document requests bit 4 for Entropy Label Flag in "SR Policy
Segment Flags" under the "BGP Tunnel Encapsulation" registry.

```
 Bit     Description                               Reference
-------------------------------------------------------------------
  4      Entropy Label Position Flag(E-Flag)       This document
```

8. References

8.1. Normative References

[I-D.ietf-idr-segment-routing-te-policy]
          Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P.,
          Rosen, E., Jain, D., and S. Lin, "Advertising Segment
          Routing Policies in BGP", Work in Progress, Internet-
          Draft, draft-ietf-idr-segment-routing-te-policy-13, 7
          June 2021, <https://www.ietf.org/archive/id/draft-ietf-
          idr-segment-routing-te-policy-13.txt>.

[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>.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and
          L. Yong, "The Use of Entropy Labels in MPLS Forwarding",
          RFC 6790, DOI 10.17487/RFC6790, November 2012, <https://
          www.rfc-editor.org/info/rfc6790>.

[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>.

[RFC8662] Kini, S., Kompella, K., Sivabalan, S., Litkowski, S.,
          Shakir, R., and J. Tantsura, "Entropy Label for Source
          Packet Routing in Networking (SPRING) Tunnels", RFC 8662,
          DOI 10.17487/RFC8662, December 2019, <https://www.rfc-
          editor.org/info/rfc8662>.

8.2. Informative References

[I-D.ietf-isis-mpls-elc] Xu, X., Kini, S., Psenak, P., Filsfils, C.,
          Litkowski, S., and M. Bocci, "Signaling Entropy Label
          Capability and Entropy Readable Label Depth Using IS-IS",
          Work in Progress, Internet-Draft, draft-ietf-isis-mpls-
          elc-13, 28 May 2020, <https://www.ietf.org/archive/id/
          draft-ietf-isis-mpls-elc-13.txt>.

[I-D.ietf-ospf-mpls-elc] Xu, X., Kini, S., Psenak, P., Filsfils, C.,
          Litkowski, S., and M. Bocci, "Signaling Entropy Label
          Capability and Entropy Readable Label Depth Using OSPF",
          Work in Progress, Internet-Draft, draft-ietf-ospf-mpls-
          elc-15, 1 June 2020, <https://www.ietf.org/archive/id/
          draft-ietf-ospf-mpls-elc-15.txt>.

[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/info/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/info/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/info/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/info/rfc8664>.

[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/info/rfc8814>.

[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/info/rfc9085>.

Authors' Addresses

Liu Yao
ZTE Corp.

Email: liu.yao71@zte.com.cn

Peng Shaofu
ZTE Corp.

Email: peng.shaofu@zte.com.cn