IDR Working Group   J. Tantsura
Internet-Draft   Apstra, Inc.
Intended status: Standards Track   U. Chunduri
Expires: August 31, 2020   Futurewei Technologies
                                                       K. Talaulikar
                                                       Cisco Systems
                                                           G. Mirsky
                                                           ZTE Corp.
                                                   N. Triantafillis
                                                 Amazon Web Services
                                                   February 28, 2020

Signaling MSD (Maximum SID Depth) using Border Gateway Protocol - Link State
draft-ietf-idr-bgp-ls-segment-routing-msd-11

Abstract

   This document defines a way for a Border Gateway Protocol - Link
   State (BGP-LS) speaker to advertise multiple types of supported
   Maximum SID Depths (MSDs) at node and/or link granularity.

   Such advertisements allow entities (e.g., centralized controllers) to
   determine whether a particular Segment Identifier (SID) stack can be
   supported in a given network.

Status of This Memo

Copyright Notice

Table of Contents

1. **Introduction**

   When Segment Routing (SR) [RFC8402] paths are computed by a
   centralized controller, it is critical that the controller learn the
   Maximum SID Depth (MSD) that can be imposed at each node/link on a
   given SR path. This ensures that the Segment Identifier (SID) stack
   depth of a computed path doesn't exceed the number of SIDs the node
   is capable of imposing.

   [RFC8664] defines how to signal MSD in the Path Computation Element
   Protocol (PCEP). The OSPF and IS-IS extensions for signaling of MSD
   are defined in [RFC8476] and [RFC8491] respectively.

However, if PCEP is not supported/configured on the head-end of a SR
tunnel or a Binding-SID anchor node, and controller does not
participate in IGP routing, it has no way of learning the MSD of
nodes and links. BGP-LS [RFC7752] defines a way to expose topology
and associated attributes and capabilities of the nodes in that
topology to a centralized controller.

This document defines extensions to BGP-LS to advertise one or more
types of MSDs at node and/or link granularity. Other types of MSD
are known to be useful. For example, [I-D.ietf-ospf-mpls-elc] and
[I-D.ietf-isis-mpls-elc] define Readable Label Depth Capability
(RLDC) that is used by a head-end to insert an Entropy Label (EL) at
a depth that can be read by transit nodes.

In the future, it is expected that new MSD-Types will be defined to
signal additional capabilities, e.g., ELs, SIDs that can be imposed
through recirculation, or SIDs associated with another data plane
such as IPv6. MSD advertisements MAY be useful even if SR itself is
not enabled. For example, in a non-SR MPLS network, MSD defines the
maximum label depth.

1.1. Conventions used in this document

1.1.1. Terminology

BGP-LS: Distribution of Link-State and TE Information using Border
Gateway Protocol

MSD: Maximum SID Depth

PCC: Path Computation Client

PCE: Path Computation Element

PCEP: Path Computation Element Protocol

SID: Segment Identifier

SR: Segment routing

Label Imposition: Imposition is the act of modifying and/or adding
labels to the outgoing label stack associated with a packet. This
includes:

o replacing the label at the top of the label stack with a new
   label.

   o pushing one or more new labels onto the label stack. The number
      of labels imposed is then the sum of the number of labels that are
      replaced and the number of labels that are pushed. See [RFC3031]
      for further details.

1.1.2. 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. Advertisement of MSD via BGP-LS

   This document describes extensions that enable BGP-LS speakers to
   signal the MSD capabilities (described in [RFC8491]) of nodes and
   their links in a network to a BGP-LS consumer of network topology
   such as a centralized controller. The centralized controller can
   leverage this information in computation of SR paths and their
   instantiation on network nodes based on their MSD capabilities. When
   a BGP-LS speaker is originating the topology learnt via link-state
   routing protocols like OSPF or IS-IS, the MSD information for the
   nodes and their links is sourced from the underlying extensions as
   defined in [RFC8476] and [RFC8491] respectively.

   The BGP-LS speaker may also advertise the MSD information for the
   local node and its links when not running any link-state IGP protocol
   e.g. when running BGP as the only routing protocol. The Protocol-ID
   field should be set to BGP since the link and node attributes have
   BGP based identifiers. Deployment model for such case would be: a
   limited number (meeting resiliecy requirements) of BGP-LS speakers
   exposing the topology to the controller, full-mesh/RouteReflectors
   for iBGP(Internal Border Gateway Protocol) or regular eBGP(External
   Border Gateway Protocol) connectivity between nodes in the topology.

   The extensions introduced in this document allow for advertisement of
   different MSD-Types. This document does not define these MSD-Types
   but leverages the definition, guidelines and the code-point registry
   specified in [RFC8491]. This enables sharing of MSD-Types that may
   be defined in the future by the IGPs in BGP-LS.

3. Node MSD TLV

   Node MSD is encoded in a new Node Attribute TLV [RFC7752] using the
   following format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type              |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   MSD-Type    |  MSD-Value    |  MSD-Type... |  MSD-Value... |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
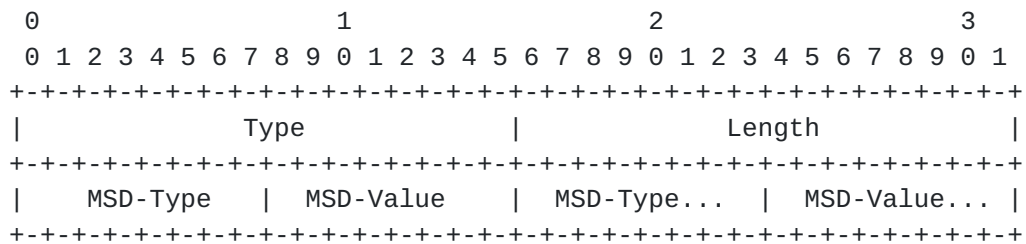
                   Figure 1: Node MSD TLV Format

   Where:

   o Type: 266

   o Length: variable (multiple of 2); represents the total length of
      the value field in octets.

   o Value : consists of one or more pairs of a 1-octet MSD-Type and
      1-octet MSD-Value.

      * MSD-Type : one of the values defined in the IANA registry
         titled "IGP MSD-Types" under the "Interior Gateway Protocol
         (IGP) Parameters" registry created by [RFC8491].

      * MSD-Value : a number in the range of 0-255. For all MSD-Types,
         0 represents the lack of ability to impose an MSD stack of any
         depth; any other value represents that of the node. This value
         MUST represent the lowest value supported by any link
         configured for use by the advertising protocol instance.

4. Link MSD TLV

   Link MSD is encoded in a new Link Attribute TLV [RFC7752] using the
   following format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type              |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   MSD-Type    |  MSD-Value    |  MSD-Type... |  MSD-Value... |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

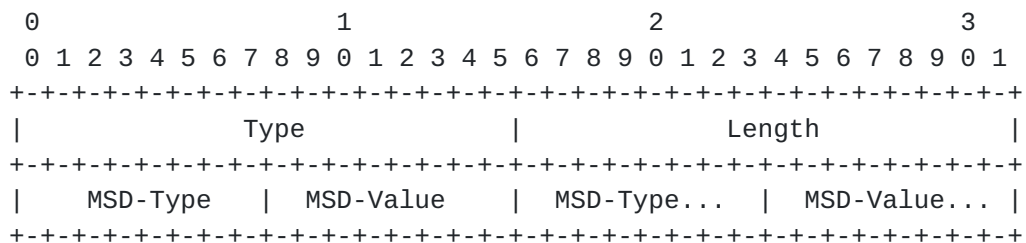   Figure 2: Link MSD TLV Format

   Where:

   o Type: 267

o Length: variable (multiple of 2); represents the total length of
   the value field in octets.

o Value : consists of one or more pairs of a 1-octet MSD-Type and
   1-octet MSD-Value.

   * MSD-Type : one of the values defined in the IANA registry
      titled "IGP MSD-Types" under the "Interior Gateway Protocol
      (IGP) Parameters" registry created by [RFC8491].

   * MSD-Value : a number in the range of 0-255. For all MSD-Types,
      0 represents the lack of ability to impose an MSD stack of any
      depth; any other value represents that of the link when used as
      an outgoing interface.

5. Procedures for Defining and Using Node and Link MSD Advertisements

   When Link MSD is present for a given MSD-type, the value of the Link
   MSD MUST take precedence over the Node MSD. When a Link MSD-type is
   not signaled but the Node MSD-type is, then the Node MSD-type value
   MUST be considered as the MSD value for that link.

   In order to increase flooding efficiency, it is RECOMMENDED that
   routers with homogenous link MSD values advertise just the Node MSD
   value.

   The meaning of the absence of both Node and Link MSD advertisements
   for a given MSD-type is specific to the MSD-type. Generally it can
   only be inferred that the advertising node does not support
   advertisement of that MSD-type. However, in some cases the lack of
   advertisement might imply that the functionality associated with the
   MSD-type is not supported. The correct interpretation MUST be
   specified when an MSD-type is defined in [RFC8491].

6. IANA Considerations

   This document requests assigning code-points from the registry "BGP-
   LS Node Descriptor, Link Descriptor, Prefix Descriptor, and Attribute
   TLVs" based on table below. Early allocation for these code-points
   have been done by IANA.

Code Point	Description	IS-IS TLV/Sub-TLV
266	Node MSD	242/23
267	Link MSD	(22,23,25,141,222,223)/15

7. Manageability Considerations

   The new protocol extensions introduced in this document augment the
   existing IGP topology information that is distributed via [RFC7752].
   Procedures and protocol extensions defined in this document do not
   affect the BGP protocol operations and management other than as
   discussed in the Manageability Considerations section of [RFC7752].
   Specifically, the malformed attribute tests for syntactic checks in
   the Fault Management section of [RFC7752] now encompass the new BGP-
   LS Attribute TLVs defined in this document. The semantic or content
   checking for the TLVs specified in this document and their
   association with the BGP-LS NLRI types or their BGP-LS Attribute is
   left to the consumer of the BGP-LS information (e.g. an application
   or a controller) and not the BGP protocol.

   A consumer of the BGP-LS information retrieves this information over
   a BGP-LS session (refer Section 1 and 2 of [RFC7752]). The handling
   of semantic or content errors by the consumer would be dictated by
   the nature of its application usage and hence is beyond the scope of
   this document.

   This document only introduces new Attribute TLVs and any syntactic
   error in them would result in the BGP-LS Attribute being discarded
   with an error log. The MSD information introduced in BGP-LS by this
   specification, may be used by BGP-LS consumer applications like a SR
   path computation engine (PCE) to learn the SR SID-stack handling
   capabilities of the nodes in the topology. This can enable the SR
   PCE to perform path computations taking into consideration the size
   of SID Stack that the specific headend node may be able to impose.
   Errors in the encoding or decoding of the MSD information may result
   in the unavailability of such information to the SR PCE or incorrect
   information being made available to it. This may result in the
   headend node not being able to instantiate the desired SR path in its
   forwarding and provide the SR based optimization functionality. The
   handling of such errors by applications like SR PCE may be
   implementation specific and out of scope of this document.

   The extensions specified in this document, do not specify any new
   configuration or monitoring aspects in BGP or BGP-LS. The
   specification of BGP models BGP and BGP-LS models is an ongoing work
   based on the [I-D.ietf-idr-bgp-model]. The management of the MSD
   features within an ietf segment-routing stack is also an ongoing work
   based on the [I-D.ietf-spring-sr-yang]. Management of the segment
   routing in IGPs is ongoing work for ISIS [I-D.ietf-isis-sr-yang] ,
   and OSPF [I-D.ietf-ospf-sr-yang].

8. Security Considerations

   The advertisement of an incorrect MSD value may have negative
   consequences. If the value is smaller than supported, path
   computation may fail to compute a viable path. If the value is
   larger than supported, an attempt to instantiate a path that can't be
   supported by the head-end (the node performing the SID imposition)
   may occur. The presence of this information may also inform an
   attacker of how to induce any of the aforementioned conditions.

   The document does not introduce additional security issues beyond
   discussed in [RFC7752], [RFC8476] and [RFC8491]. However, [RFC7752]
   is being revised in [I-D.ietf-idr-rfc7752bis] to provide additional
   clarification in several portions of the specification after
   receiving feedback from implementers. One of the places that is
   being clarified is the error handling and security. It is expected
   that after [I-D.ietf-idr-rfc7752bis] is released that implementers
   will update all BGP-LS base implementations improving the error
   handling for protocol work (including this document) that depend on
   this function.

9. Contributors

   Siva Sivabalan
   Cisco Systems Inc.
   Canada

   Email: msiva@cisco.com

10. Acknowledgements

   We like to thank Acee Lindem, Stephane Litkowski and Bruno Decraene
   for their reviews and valuable comments.

11. References

11.1. 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/info/rfc2119>.

   [RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and
              S. Ray, "North-Bound Distribution of Link-State and
              Traffic Engineering (TE) Information Using BGP", RFC 7752,
              DOI 10.17487/RFC7752, March 2016,
              <https://www.rfc-editor.org/info/rfc7752>.

   [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>.

   [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>.

11.2. Informative References

   [I-D.ietf-idr-bgp-model]
               Jethanandani, M., Patel, K., Hares, S., and J. Haas, "BGP
               YANG Model for Service Provider Networks", draft-ietf-idr-
               bgp-model-07 (work in progress), October 2019.

   [I-D.ietf-idr-rfc7752bis]
               Talaulikar, K., "Distribution of Link-State and Traffic
               Engineering Information Using BGP", draft-ietf-idr-
               rfc7752bis-02 (work in progress), November 2019.

   [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", draft-ietf-
               isis-mpls-elc-10 (work in progress), October 2019.

   [I-D.ietf-isis-sr-yang]
               Litkowski, S., Qu, Y., Sarkar, P., Chen, I., and J.
               Tantsura, "YANG Data Model for IS-IS Segment Routing",
               draft-ietf-isis-sr-yang-07 (work in progress), January
               2020.

   [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-stack Depth Using OSPF", draft-
               ietf-ospf-mpls-elc-12 (work in progress), October 2019.

   [I-D.ietf-ospf-sr-yang]
              Yeung, D., Qu, Y., Zhang, Z., Chen, I., and A. Lindem,
              "YANG Data Model for OSPF SR (Segment Routing) Protocol",
              draft-ietf-ospf-sr-yang-11 (work in progress), February
              2020.

   [I-D.ietf-spring-sr-yang]
              Litkowski, S., Qu, Y., Lindem, A., Sarkar, P., and J.
              Tantsura, "YANG Data Model for Segment Routing", draft-
              ietf-spring-sr-yang-15 (work in progress), January 2020.

   [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol
              Label Switching Architecture", RFC 3031,
              DOI 10.17487/RFC3031, January 2001,
              <https://www.rfc-editor.org/info/rfc3031>.

   [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>.

   [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>.

Authors' Addresses

   Jeff Tantsura
   Apstra, Inc.

   Email: jefftant.ietf@gmail.com

   Uma Chunduri
   Futurewei Technologies

   Email: umac.ietf@gmail.com

   Ketan Talaulikar
   Cisco Systems

   Email: ketant@cisco.com

Greg Mirsky
ZTE Corp.

Email: gregimirsky@gmail.com

Nikos Triantafillis
Amazon Web Services

Email: nikost@amazon.com