Inter-Domain Routing
Internet-Draft
Intended status: Standards Track
Expires: August 18, 2014

H. Gredler, Ed.
Juniper Networks, Inc.
S. Ray, Ed.
S. Previdi
C. Filsfils
Cisco Systems, Inc.
M. Chen
Huawei Technologies
J. Tantsura
Ericsson

February 14, 2014

BGP Link-State extensions for Segment Routing
draft-gredler-bgp-ls-segment-routing-extensions-00

Abstract

Segment Routing (SR) allows for a flexible definition of end-to-end paths within link-state graphs by encoding paths as sequences of topological sub-paths, called "segments".

The link-state routing protocols (IS-IS, OSPF and OSPFv3) have been extended to advertise the segments. But flooding based propagation of path segments using IGPs is limited by the perimeter of the IGP domain. For building paths which span across IGP domains (e.g. multiple ASes), the Border Gataway Protocol (BGP) is better suited as its propagation perimeter is not limited like the IGPs.

This draft defines extensions to the BGP Link-state address-family to carry path segment information via BGP.

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].

This Internet-Draft will expire on August 18, 2014.

Copyright Notice

Table of Contents

1. Introduction 2
2. BGP-LS Extensions for Segment Routing 5
  2.1. Node Attribute TLVs 6
  2.2. Link Attribute TLVs 6
  2.3. Prefix Attribute TLVs 7
3. IANA Considerations 7
4. Manageability Considerations 7
  4.1. Operational Considerations 7
    4.1.1. Operations 7
5. TLV/Sub-TLV Code Points Summary 7
6. Security Considerations 8
7. Acknowledgements 8
8. References 8
  8.1. Normative References 8
  8.2. Informative References 9
Authors' Addresses 10

1. Introduction

Segment Routing (SR) allows for a flexible definition of end-to-end paths within link-state topologies by encoding paths as sequences of topological sub-paths, called "segments". Segment routing is an amalgamation of source routing and MPLS. In Segment Routing, the ingress node prepends a sequence of instructions, called "segments",

to the packet. The SR capable nodes sequentially execute the instructions on the packet to achieve packet forwarding via required topological paths as well as service paths.

The segments can be thought of, in a simple way, to represent instructions such as "go to node N using the shortest path", "follow the shortest path for prefix P", "use link/node/ERO L", etc. Each segment is identified by a 32 bit Segment Identifier (SID) (when unmodified MPLS data-plane is used, the SIDs are restricted to 20 bits). There are "global" segments that are known to all SR nodes in the local domain, and there are local segments whose semantics are known only to the nodes that originate them. The segment routing architecture is described in [I-D.filsfils-rtgwg-segment-routing] and segment routing use-cases in [I-D.filsfils-rtgwg-segment-routing-use-cases].

Segment routing is enabled in a network by advertising the segments (including the associated SIDs) to the nodes in the network. The IGP link-state routing protocols (IS-IS [I-D.previdi-isis-segment-routing-extensions], OSPFv2 [I-D.psenak-ospf-segment-routing-extensions] and OSPFv3 [I-D.psenak-ospf-segment-routing-ospfv3-extension]) have been extended to advertise the segments. Using these extensions, segment routing can be enabled within an IGP domain.

```
                           +------------+
                           |  Consumer  |
                           +------------+
                                 ^
                                 |
                                 v
                       +-------------------+
                       |    BGP Speaker    |         +-----------+
                       | (Route-Reflector) |         | Consumer  |
                       +-------------------+         +-----------+
                             ^   ^   ^                       ^
                             |   |   |                       |
             +---------------+   |   +-------------------+   |
             |                   |                       |   |
             v                   v                       v   v
       +-----------+       +-----------+             +-----------+
       |    BGP    |       |    BGP    |             |    BGP    |
       |  Speaker  |       |  Speaker  |    . . .    |  Speaker  |
       +-----------+       +-----------+             +-----------+
             ^                   ^                         ^
             |                   |                         |
            IGP                 IGP                       IGP
```

Figure 1: Link State info collection

Segment Routing (SR) allows advertisement of single or multi-hop paths. The flooding scope for the IGP extensions for Segment routing is IGP area-wide. Consequently, the contents of a Link State Database (LSDB) or a Traffic Engineering Database (TED) has the scope of an IGP area and therefore by using the IGP alone it is not possible to construct segments across an IGP Area or AS boundaries.

To address the need for applications that require visibility into LSDB across IGP areas, or even across ASes, the BGP-LS address-family /sub-address-family have been defined that allows BGP to carry LSDB information. The BGP Network Layer Reachability Information (NLRI) encoding format for BGP-LS and a new BGP Path Attribute called BGP-LS attribute are defined in [I-D.ietf-idr-ls-distribution]. The identifying key of each LSDB object, namely a node, a link or a prefix, is encoded in the NLRI and the properties of the object are encoded in the BGP-LS attribute. Figure Figure 1 describes a typical deployment scenario. In each IGP area, one or more nodes are configured with BGP-LS. These BGP speakers form an IBGP mesh by connecting to one or more route-reflectors. This way, all BGP speakers - specifically the route-reflectors - obtain LSDB information from all IGP areas (and from other ASes from EBGP peers). An external component connects to the route-reflector to obtain this

information (perhaps moderated by a policy regarding what information is sent to the external component, and what information isn't).

This document describes extensions to BGP-LS to carry the segments. An external component - a Controller - then can collect segment information in the "northbound direction" across IGP areas/autonomous systems and construct the segment stack that need to be added to an incoming packet to achieve the desired end-to-end forwarding.

2. BGP-LS Extensions for Segment Routing

The BGP-LS NLRI can be a node NLRI, a link NLRI or a prefix NLRI. The corresponding BGP-LS attribute is a node attribute, a link attribute or a prefix attribute. BGP-LS [I-D.ietf-idr-ls-distribution] defines the TLVs that map link-state information to BGP-LS NLRI and BGP-LS attribute. This document adds additional BGP-LS attribute TLVs to encode SR information.

[I-D.previdi-isis-segment-routing-extensions] defines the following TLVs to encode SR information.

- o TLV for Prefix-SID
- o TLV for Adjacency-SID between two nodes as well as between nodes in a LAN
- o TLV for SID/Label binding for advertising paths from other protocols (and their optional ERO)
- o TLV for SR Capabilities
- o TLV for SR Algorithm

These TLVs are mapped to BGP-LS attribute TLVs in the following way.

```
 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            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                        Value (variable)                     //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: TLV format

The 2 octet Type field values are defined in Table 1, Table 2, and Table 3. The next 2 octet Length field encodes length of the rest of the TLV. The Value portion of the TLV is variable and is equal to

the corresponding Value portion of the TLV defined in [I-D.previdi-isis-segment-routing-extensions].

In each case, multiple TLVs for a given type are allowed to be added. The semantics of multiple such values are determined by [I-D.previdi-isis-segment-routing-extensions].

2.1. Node Attribute TLVs

The following 'Node Attribute' TLVs are defined:

TLV Code Point	Description	Length	IS-IS SR TLV/sub-TLV
1033	SID/Label Binding	variable	149 (section 2.4)
1034	SR Capabilities	variable	2 (section 3.1)
1035	SR Algorithm	variable	15 (section 3.2)

Table 1: Node Attribute TLVs

The sections refer to [I-D.previdi-isis-segment-routing-extensions].

These TLVs can ONLY be added to the Node Attribute associated with the Node NLRI that originates the corresponding SR TLV.

2.2. Link Attribute TLVs

The following 'Link Attribute' TLVs are are defined:

TLV Code Point	Description	Length	IS-IS SR TLV /sub-TLV
1099	Adjacency Segment Identifier (Adj-SID) TLV	variable	31 (section 2.3.1)
1100	LAN Adjacency Segment Identifier (Adj-SID) TLV	variable	32 (section 2.3.2)

Table 2: Link Attribute TLVs

The sections refer to [I-D.previdi-isis-segment-routing-extensions].

These TLVs can ONLY be added to the Link Attribute associated with the link whose local node originates the corresponding SR TLV.

For a LAN, normally a node only announces its adjacency to the pseudo-node. [I-D.previdi-isis-segment-routing-extensions] allows a node to announce adjacency to all other nodes attached to the LAN. In such a case, the corresponding BGP-LS link NLRI must be originated for each additional link in order to add the SR TLVs to the Link Attribute.

2.3. Prefix Attribute TLVs

The following 'Prefix Attribute' TLVs are defined:

TLV Code Point	Description	Length	IS-IS SR TLV/sub-TLV
1158	Prefix SID	variable	3 (section 2.2)

Table 3: Prefix Attribute TLVs

The sections refer to [I-D.previdi-isis-segment-routing-extensions].

These TLVs can ONLY be added to the Prefix Attribute whose local node in the corresponding prefix NLRI is the node that originates the corresponding SR TLV.

3. IANA Considerations

This document requests assigning code-points from the registry for BGP-LS attribute TLVs based on table Table 4.

4. Manageability Considerations

This section is structured as recommended in [RFC5706].

4.1. Operational Considerations

4.1.1. Operations

Existing BGP and BGP-LS operational procedures apply. No new operation procedures are defined in this document.

5. TLV/Sub-TLV Code Points Summary

This section contains the global table of all TLVs/Sub-TLVs defined in this document.

TLV Code Point	Description	Length	IS-IS SR TLV /sub-TLV
1033	SID/Label Binding	variable	149 (section 2.4)
1034	SR Capabilities	variable	2 (section 3.1)
1035	SR Algorithm	variable	15 (section 3.2)
1099	Adjacency Segment Identifier (Adj-SID) TLV	variable	31 (section 2.3.1)
1100	LAN Adjacency Segment Identifier (Adj-SID) TLV	variable	32 (section 2.3.2)
1158	Prefix SID	variable	3 (section 2.2)

Table 4: Summary Table of TLV/Sub-TLV Codepoints

6. Security Considerations

Procedures and protocol extensions defined in this document do not affect the BGP security model. See the 'Security Considerations' section of [RFC4271] for a discussion of BGP security. Also refer to [RFC4272] and [I-D.ietf-karp-routing-tcp-analysis] for analysis of security issues for BGP.

7. Acknowledgements

TBD.

8. References

8.1. Normative References

[I-D.ietf-idr-ls-distribution]
Gredler, H., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and TE Information using BGP", draft-ietf-idr-ls-distribution-04 (work in progress), November 2013.

[I-D.previdi-isis-segment-routing-extensions]
Previdi, S., Filsfils, C., Bashandy, A., Gredler, H., and S. Litkowski, "IS-IS Extensions for Segment Routing", draft-previdi-isis-segment-routing-extensions-04 (work in progress), October 2013.

[I-D.psenak-ospf-segment-routing-extensions]
Psenak, P., Previdi, S., Filsfils, C., Gredler, H., Shakir, R., and W. Henderickx, "OSPF Extensions for Segment Routing", draft-psenak-ospf-segment-routing-extensions-03 (work in progress), October 2013.

[I-D.psenak-ospf-segment-routing-ospfv3-extension]
Psenak, P., Previdi, S., Filsfils, C., Gredler, H., Shakir, R., and W. Henderickx, "OSPFv3 Extensions for Segment Routing", draft-psenak-ospf-segment-routing-ospfv3-extension-00 (work in progress), October 2013.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, January 2006.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

8.2. Informative References

[I-D.filsfils-rtgwg-segment-routing-use-cases]
Filsfils, C., Francois, P., Previdi, S., Decraene, B., Litkowski, S., Horneffer, M., Milojevic, I., Shakir, R., Ytti, S., Henderickx, W., Tantsura, J., and E. Crabbe, "Segment Routing Use Cases", draft-filsfils-rtgwg-segment-routing-use-cases-01 (work in progress), July 2013.

[I-D.filsfils-rtgwg-segment-routing]
Filsfils, C., Previdi, S., Bashandy, A., Decraene, B., Litkowski, S., Horneffer, M., Milojevic, I., Shakir, R., Ytti, S., Henderickx, W., Tantsura, J., and E. Crabbe, "Segment Routing Architecture", draft-filsfils-rtgwg-segment-routing-01 (work in progress), October 2013.

[I-D.ietf-karp-routing-tcp-analysis]
Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP and MSDP Issues According to KARP Design Guide", draft-ietf-karp-routing-tcp-analysis-07 (work in progress), April 2013.

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

[RFC5152] Vasseur, JP., Ayyangar, A., and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, February 2008.

[RFC5706] Harrington, D., "Guidelines for Considering Operations and Management of New Protocols and Protocol Extensions", RFC 5706, November 2009.

Authors' Addresses

Hannes Gredler (editor)
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: hannes@juniper.net

Saikat Ray (editor)
Cisco Systems, Inc.
170, West Tasman Drive
San Jose, CA 95134
US

Email: sairay@cisco.com

Stefano Previdi
Cisco Systems, Inc.
Via Del Serafico, 200
Rome 00142
Italy

Email: sprevidi@cisco.com

Clarence Filsfils
Cisco Systems, Inc.
Brussels
Belgium

Email: cfilsfil@cisco.com

Mach(Guoyi) Chen
Huawei Technologies
Huawei Building, No. 156 Beiqing Rd.
Beijing 100095
China

Email: mach.chen@huawei.com

Jeff Tantsura
Ericsson
300 Holger Way
San Jose, CA 95134
US

Email: jeff.tantsura@ericsson.com