IDR Working Group A. Wang
Internet-Draft China Telecom
Intended status: Standards Track H. Chen
Expires: April 16, 2022 Futurewei
K. Talaulikar
Cisco Systems
S. Zhuang
Huawei Technologies
October 13, 2021

BGP-LS Extension for Inter-AS Topology Retrieval
draft-ietf-idr-bgpls-inter-as-topology-ext-10

Abstract

This document describes the process to build Border Gateway Protocol-Link State (BGP-LS) key parameters in inter-domain scenario, defines one new BGP-LS Network Layer Reachability Information (NLRI) type (Stub Link NLRI) and some new inter Autonomous (inter-AS) Traffic Engineering (TE) related Type-Length-Values (TLVs) for BGP-LS to let Software Definition Network (SDN) controller retrieve the network topology automatically under various inter-AS environments.

Such extension and process can enable the network operator to collect the interconnect information between different domains and then calculate the overall network topology automatically based on the information provided by BGP-LS protocol.

Copyright Notice

Table of Contents

1. Introduction

BGP-LS [RFC7752] describes the methodology that using BGP protocol to transfer the Link-State information. Such method can enable SDN controller to collect the underlay network topology automatically, but normally it can only get the information within one Interior Gateway Protocol (IGP) domain. If the operator has more than one IGP domain, and these domains interconnect with each other, there is no

mechanic within current BGP- LS to transfer the interconnect topology information.

Draft [I-D.ietf-idr-bgpls-segment-routing-epe] defines some extensions for exporting BGP peering node topology information (including its peers, interfaces and peering ASs) in a way that is exploitable in order to compute efficient BGP Peering Engineering policies and strategies. Such information can also be used to calculate the interconnection topology among different IGP domains, but it requires every border router to run BGP-LS protocol and report the information to SDN controller. Considering there will be several border routers on the network boundary, such solution restricts its deployment flexibility.

This draft analysis the situations that the SDN controller needs to get the interconnected topology information between different AS domains, defines the new Stub Link NLRI and some new TLVs within the BGP-LS protocol to transfer the key information related to them. After that, the SDN controller can then deduce the multi-domain topology automatically based on the information from BGP-LS protocol.

2. Conventions used in this document

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] .

3. Terminology

The following terms are defined in this document:

- IDCs: Internet Data Centers

- MAN: Metrio-Area-Network

- SDN: Software Definition Network

4. Inter-AS Domain Scenarios.

Figure 1 illustrates the multi-domain scenarios that this draft discusses. Normally, SDN Controller can get the topology of IGP A and IGP B individually via the BGP-LS protocol, but it can't get the topology connection information between these two IGP domains because there is generally no IGP protocol run on the connected links.

```
                             +-----------------+
                        +----+IP SDN Controller+----+
                        |    +-----------------+    |
                        |                           |
                        |BGP-LS                     |BGP-LS
                        |                           |
        +---------------+-----+               +-----+--------------+
        | +--+        +-++   ++-+           +-++   +|-+        +--+|
        | |S1+--------+S2+---+B1+-----------+B2+---+T1+--------+T2||
        | +-++   N1   +-++   ++-+           +-++   ++++   N2   +-++|
        |   |           |     |               |      ||           | |
        |   |           |     |               |      ||           | |
        | +-++        +-++   ++-+           +-++   ++++        +-++|
        | |S4+--------+S3+---+B3+-----------+B4+---+T3+--------+T4||
        | +--+        +--+   ++-+           +-++   ++-+        +--+|
        |                     |               |                    |
        |                     |               |                    |
        |       IGP A         |               |      IGP B         |
        +---------------------+               +--------------------+
```

Figure 1: Inter-AS Domain Scenarios

5. Stub Link NLRI

[RFC7752] defines four NLRI types(Node NLRI, Link NLRI, IPv4 Topology Prefix NLRI, IPv6 Topology Prefix NLRI) to transfer the topology and prefix information. For inter-as link, the two ends of the link locates in different IGP domains, then it is not appropriate to transfer their information within the current defined NLRI types.

This draft defines one new NLRI type, called Stub Link NLRI, which is coded as 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
       +-+-+-+-+-+-+-+-+
       |  Protocol-ID  |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                           Identifier                          |
       |                            (64 bits)                          |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       //              Local Node Descriptors (variable)             //
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       //               Stub Link Descriptors (variable)             //
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: Stub Link NLRI Format

The "Protocol-ID" should be set to the value that indicates "Direct" protocol.

The semantics of "Local Node Descriptors" and "Stub Link Descriptors" are same as that defined in [RFC7752] for "Node Descriptors" and "Link Descriptor".

This newly defined NLRI can be used to describe the link that has only one end located within the IGP domain, as described in the following sections.

5.1. Inter-AS Native IP Scenario

Draft [RFC8735] describes the situation that operator needs some traffic engineering solution for the inter-as native IP environment. In such situation, different domain may run different IGP protocol. The operator needs to know the inter-as topology first to calculate the end to end optimal path centrally.

When IGP A or IGP B in Figure 1 runs native IS-IS/OSPF protocol, the operator can use passive feature for the inter-domain links to let the routers within the IGP domain know these links. Such stub links information can then be carried within the Stub Link NLRI reported via the BGP-LS protocol to the SDN controller.

For OSPFv2, when the interface is configured as passive, the "Linktype" field in corresponding Router LSA will be set to 3, to indicate it connects with stub network. Other routers in the IGP domain can identify such interfaces via this characteristics, and report them via the newly defined "Stub Link NLRI".

For OSPFv3 and ISIS, [I-D.wang-lsr-stub-link-attributes] describes the method to label the stub link within the network. The router that runs BGP-LS can extract these stub link from other interfaces that participate in the IGP protocol and report them via the newly defined "Stub Link NLRI".

The "Local Node Descriptors" should describe the characteristics of ASBRs that are connected these stub links.

When such information is reported via the BGP-LS protocol, the SDN controller can construct the underlay inter-domain topology according to procedure described in Section 7

5.2. Inter-AS TE Scenario

When IGP A or IGP B in Figure 1 runs IS-IS TE/OSPF-TE protocol,[RFC5316] and [RFC5392] define IS-IS and OSPF extensions respectively to deal with the situation for inter-AS traffic engineering. Three new sub-TLVs(Remote AS Number、IPv4 Remote ASBR ID、IPv6 Remote ASBR ID) which are associated with the inter-AS TE link are defined.

These TLVs are flooded within the IGP domain automatically. They should be carried within the newly defined Stub Link NLRI within the BGP-LS protocol, as the descriptors for the inter-AS stub link.

The "Local Node Descriptors" should describe the the characteristics of ASBRs that are connected these inter-AS TE links.

If the SDN controller knows these information via one of the interior router that runs BGP-LS protocol, the SDN controller can rebuild the inter-AS TE topology correctly according to the procedure described in Section 7

6. Inter-AS TE NLRI related TLVs

This draft proposes to add three new TLVs that is included within the Stub Link NLRI to transfer the information via BGP-LS, which are required to build the inter-AS TE related topology by the SDN controller.

The following Link Descriptor TLVs are added into the BGP-LS protocol :

TLV Code Point	Description	IS-IS/OSPF TLV /Sub-TLV	Reference (RFC/Section)
TBD	Remote AS Number	24/21	[RFC5316]/3.3.1 [RFC5392]/3.3.1
TBD	IPv4 Remote ASBR ID	25/22	[RFC5316]/3.3.2 [RFC5392]/3.3.2
TBD	IPv6 Remote ASBR ID	26/24	[RFC5316]/3.3.3 [RFC5392]/3.3.3

Figure 3: Link Descriptor TLVs

Detail encoding of these TLVs are synchronized with the corresponding parts in [RFC5316] and [RFC5392], which keeps the BGP-LS protocol agnostic to the underly protocol.

6.1. Remote AS Number TLV

A new TLV, the remote AS number TLV, is defined for inclusion in the link descriptor when advertising inter-AS TE links. The remote AS number TLV specifies the AS number of the neighboring AS to which the advertised link connects.

The remote AS number TLV is TLV type TBD (see Section 9) and is 4 octets in length. The format is as follows:

```
 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            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Remote AS Number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: Remote AS Number TLV Format

The Remote AS number field has 4 octets. When only 2 octets are used for the AS number, as in current deployments, the left (high-order) 2 octets MUST be set to 0. The remote AS number TLV MUST be included when a router advertises an inter-AS TE link.

6.2. IPv4 Remote ASBR ID

A new TLV, which is referred to as the IPv4 remote ASBR ID TLV, is defined for inclusion in the link descriptor when advertising inter-AS TE links. The IPv4 remote ASBR ID TLV specifies the IPv4 identifier of the remote ASBR to which the advertised inter-AS link connects. This could be any stable and routable IPv4 address of the remote ASBR. Use of the TE Router ID as specified in the Traffic Engineering router ID TLV [RFC5316] is RECOMMENDED.

The IPv4 remote ASBR ID TLV is TLV type TBD (see Section 9) and is 4 octets in length. The format of the IPv4 remote ASBR ID sub-TLV is as follows:

```
 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            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Remote ASBR ID                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: IPv4 Remote ASBR ID TLV Format

The IPv4 remote ASBR ID TLV MUST be included if the neighboring ASBR has an IPv4 address. If the neighboring ASBR does not have an IPv4 address (not even an IPv4 TE Router ID), the IPv6 remote ASBR ID TLV MUST be included instead. An IPv4 remote ASBR ID TLV and IPv6 remote ASBR ID TLV MAY both be present in an inter-AS TE link NLRI.

6.3. IPv6 Remote ASBR ID

A new TLV, which is referred to as the IPv6 remote ASBR ID TLV, is defined for inclusion in the link descriptor when advertising inter-AS links. The IPv6 remote ASBR ID TLV specifies the IPv6 identifier of the remote ASBR to which the advertised inter-AS link connects. This could be any stable and routable IPv6 address of the remote ASBR. Use of the TE Router ID as specified in the IPv6 Traffic Engineering router ID TLV [RFC5316] is RECOMMENDED.

The IPv6 remote ASBR ID TLV is TLV type TBD (see Section 9) and is 16 octets in length. The format of the IPv6 remote ASBR ID TLV is as follows:

```
 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            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Remote ASBR ID                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Remote ASBR ID (continued)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Remote ASBR ID (continued)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Remote ASBR ID (continued)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: IPv6 Remote ASBR ID TLV Format

The IPv6 remote ASBR ID TLV MUST be included if the neighboring ASBR has an IPv6 address. If the neighboring ASBR does not have an IPv6 address, the IPv4 remote ASBR ID TLV MUST be included instead. An IPv4 remote ASBR ID TLV and IPv6 remote ASBR ID TLV MAY both be present in an inter-AS TE link NLRI.

7. Topology Reconstruction.

When SDN controller gets such information from BGP-LS protocol, it should compares the proximity of these stub links. If they are under the same network scope and in different AS, then it should find the corresponding associated router information, build the link between these two border routers.

If the prefixes reported via the "Stub Link" NLRI are under the same network scope, and in the same AS, the SDN controller can then determine there is some IGP adjacency irregular. The usage of such information is out of scope of this draft.

After iterating the above procedures for all of the stub links, the SDN controller can then retrieve the connection topology between different domains automatically.

8. Security Considerations

It is common for one operator to occupy several IGP domains that are composited by its backbone network and several MAN(Metrio-Area-Network)s/Internet Data Centers (IDCs). When they do traffic engineering which spans MAN, Backbone and IDC, they need to know the inter-as topology via the process described in this draft. Using the passive interface features or configuring the Traffic Engineering (TE) parameters on the interconnect links will not spread the topology fluctuation across each other domain.

9. IANA Considerations

This document defines:

- A new BGP NLRI Type: Stub Link NLRI. The codepoint is from the "BGP-LS NLRI Types"

- Three new Link Descriptors TLV: Remote AS Number TLV, IPv4 Remote ASBR ID, IPv6 Remote ASBR ID. The codepoint are from "BGP-LS Node Descriptor, Link Descriptor, Prefix Descriptor, and Attribute TLVs" registry.

9.1. New BGP-LS NLRI type

This document defines a new value in the registry "BGP-LS NLRI Types":

Code Point	Description	Status
TBD	Stub Link NLRI	Allocation from IANA

Figure 7: Stub Link NLRI Codepoint

9.2. New Link Descriptors

This document defines three new values in the registry "BGP-LS Node Descriptor, Link Descriptor, Prefix Descriptor, and Attribute TLVs":

Code Point	Description	Status
TBD	Remote AS Number	Allocation from IANA
TBD	IPv4 Remote ASBR ID	Allocation from IANA
TBD	IPv6 Remote ASBR ID	Allocation from IANA

Figure 8: BGP-LS Link Descriptors TLV

10. Acknowledgement

The author would like to thank Acee Lindem, Jie Dong, Shaowen Ma, Jeff Tantsura and Dhruv Dhody for their valuable comments and suggestions.

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>.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <https://www.rfc-editor.org/info/rfc5305>.

[RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, DOI 10.17487/RFC5316, December 2008, <https://www.rfc-editor.org/info/rfc5316>.

[RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5392, DOI 10.17487/RFC5392, January 2009, <https://www.rfc-editor.org/info/rfc5392>.

[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>.

[RFC8735] Wang, A., Huang, X., Kou, C., Li, Z., and P. Mi, "Scenarios and Simulation Results of PCE in a Native IP Network", RFC 8735, DOI 10.17487/RFC8735, February 2020, <https://www.rfc-editor.org/info/rfc8735>.

11.2. Informative References

[I-D.ietf-idr-bgpls-segment-routing-epe] Previdi, S., Talaulikar, K., Filsfils, C., Patel, K., Ray, S., and J. Dong, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing BGP Egress Peer Engineering", draft-ietf-idr-bgpls-segment-routing-epe-19 (work in progress), May 2019.

[I-D.wang-lsr-stub-link-attributes] Wang, A., Hu, Z., Mishra, G. S., Lindem, A., and J. Sun, "Advertisement of Stub Link Attributes", draft-wang-lsr-stub-link-attributes-02 (work in progress), September 2021.

Authors' Addresses

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing, Beijing 102209
China

Email: wangaj3@chinatelecom.cn

Huaimo Chen
Futurewei
Boston, MA
USA

Email: hchen@futurewei.com

Ketan Talaulikar
Cisco Systems

Email: ketant@cisco.com

Shunwan Zhuang
Huawei Technologies
Huawei Building, No.156 Beiqing Rd.
Beijing 100095
China

Email: zhuangshunwan@huawei.com