LSR Working Group   A. Wang
Internet-Draft   China Telecom
Intended status: Standards Track   A. Lindem
Expires: April 18, 2019   Cisco Systems
                                                             J. Dong
                                                 Huawei Technologies
                                                        K. Talaulikar
                                                           P. Psenak
                                                       Cisco Systems
                                                    October 15, 2018

OSPF Extension for Prefix Originator
draft-wang-lsr-ospf-prefix-originator-ext-00

Abstract

   This document describes method to transfer the source router id of
   inter-area prefixes for OSPFv2 [RFC7684]and OSPFv3 [RFC8362], which
   is needed in several use cases in OSPF inter-area scenario.

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 April 18, 2019.

Copyright Notice

Table of Contents

1. Introduction

   Draft [I-D.ietf-ospf-mpls-elc] defines mechanisms to signal Entropy
   Label Capability(ELC) and Entropy Readable Label Depth (ERLD) for the
   ingress LSR to know each LSR's capability of reading the maximum
   label stack depth and performing EL-based load-balancing in MPLS
   network. After knowing this information, the ingress LSR can push
   the appropriate label stack for specific FEC trafffic, especially in
   segment routing environment or in other stacked LSPs scenarios.

   But in OSPF inter-area scenario, the prefixes originator information
   in another area is omitted by ABR router, all prefixes are attached
   behind the ABR. Router in one area doesn't know where the prefixes
   really come from, can't decide the associated router of the inter-
   area prefixes and then can't judge the ELC and ERLD capabilities of
   the destination. It is necessary to transfer the originator
   information of these inter-area prefixes to assist the ingress LSR
   does the right Label stack action.

   More generally, draft [I-D.ietf-ospf-segment-routing-msd] defines the
   mechanism to advertise multiple types of supported Maximum SID
   Depths(MSD) at node and/or link granularity. These information will
   be referred when the head end router starts to send the traffic to
   destination prefixes. In inter-area scenario, it is also necessary

for the sender to knows the capabilities of the receivers that
associated with the inter-area prefixes.

There is also other scenario that the originator of inter-area
prefixes are useful. For example, 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 if the underlay network is divided into multi area and running
OSPF protocol, it is not easy for the SDN controller to rebuild the
multi-area topology, because normally the ABR that locates on the
boundary of different area will hide the detail topology information
in non-backbone area, and the router in backbone area that runs BGP-
LS protocol can only get and report the summary network information
in non-backbone area. If the SDN controller can get the originator
of the inter-area prefixes, it is easy for them to rebuild the inter-
area topology automatically.

[RFC7794] introduces "IPv4/IPv6 Source Router IDs" TLV to label the
source of the prefixes redistributed from different Level, this TLV
can be used in the above scenarios. Such solution can also be
applied into network that run OSPF protocol, but the related LSP
messages must be extended.

This draft gives such solution for the OSPF v2 and OSPF v3 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 [RFC2119] .

3. Scenario Description

Fig.1 illustrates the topology scenario when OSPF is running in
multi-area. R0-R4 are routers in backbone area, S1-S4,T1-T4 are
internal router in area 1 and area 2 respectively. R1 and R3 are
border routers between area 0 and area 1; R2 and R4 are border
routers between area 0 and area 2. N1 is the network between router
S1 and S2, N2 is the network between router T1 and T2. Ls1 is the
loopback address of Node S1, Lt1 is the loopback address of Node T1.

```
                       +-----------------+
                       |IP SDN Controller|
                       +--------+--------+
                                |
                                |BGP-LS
                                |
     +--------------------+------+--------+-----+--------------+
     | +--+        +--+  ++-+  ++-+   +-++   + -+        +--+|
     | |S1+--------+S2+---+R1+---|R0+----+R2+---+T1+--------+T2||
     | +-++   N1   +-++  ++-+  +--+   +-++   ++++   N2  +-++|
     | |  |         |     |              |     ||         | |
     | |  |         |     |              |     ||         | |
     | +-++        +-++  ++-+           +-++   ++++       +-++|
     | |S4+--------+S3+---+R3+-----------+R4+---+T3+--------+T4||
     | +--+        +--+  ++-+           +-++   ++-+        +--+|
     |  |            |             |             |            |
     |  |            |             |             |            |
     |        Area 1 |     Area 0  |     Area 2  |            |
     +--------------------+--------------+-------------------+
```
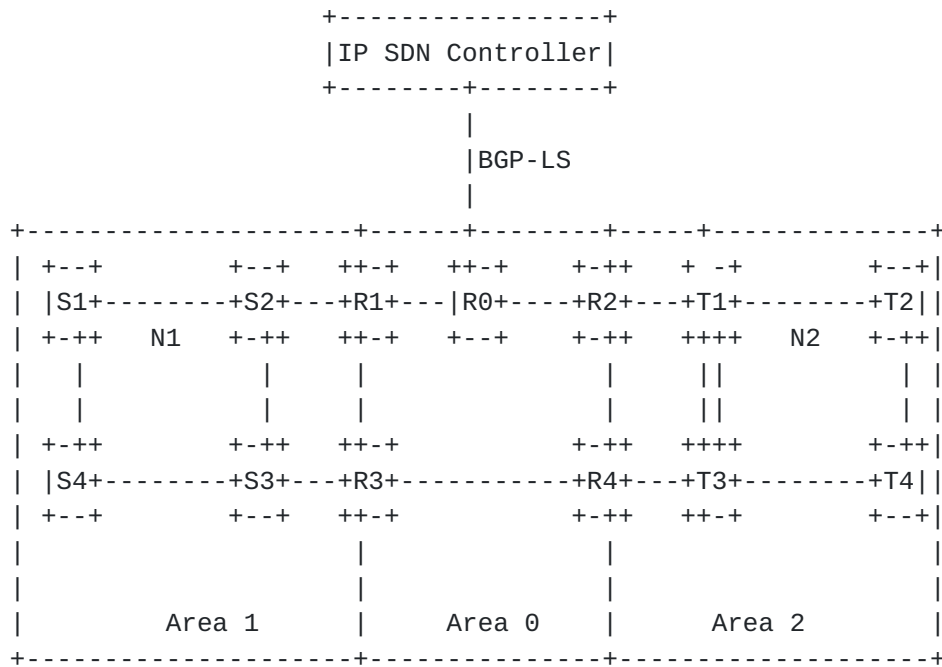
         Fig.1 OSPF Inter-Area Prefix Originator Scenario

   If S1 want to send traffic to prefix Lt1 that is connected T1 in
   another area, it should know the ELC, ERLD and MSD values that are
   associated with the node T1, and then select the right label action
   at the ingress node for the target traffic.

   On the other hand, If R0 has some methods to know the originator of
   network N1 and reports such information to IP SDN controller, then it
   is possible for the controller to retrieval the topology in non-
   backbone area. The topology retrieval process and its usage
   limitation are described in the Appendix A and Appendix B.

   From the above scenarios we can conclude it is useful to introduce
   and define the prefix originator sub TLV within OSPF.

4. Prefix Source Router ID sub TLV

   [RFC7684] and [RFC8362] define the TLV format extension for OSPFv2
   and OSPFv3 respectively. These documents give the flexibility to add
   new attributes for the prefixes and links. Based on these formats,
   we can define new sub TLV to transfer the "Prefix Source Router ID",
   as that defined in [RFC7794].

   The proposed "Prefix Source Router ID" format is the following:

```
 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         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Prefix Source Router ID (variable)               |
+--------------------------------------------------------------+
```

For IPv4 network, it is the following:

o IPv4 Source Router ID Type: TBD

o Length: 4

o Value: IPv4 Router ID of the source of the advertisement

This sub TLV should be included in the "OSPFv2 Extended Prefix Opaque
LSA" that defined in [RFC7684]

For IPv6 network, it is the following:

o IPv6 Source Router ID Type: TBD

o Length: 16

o Value: IPv6 Router ID of the source of the advertisement

This sub TLV should be included in "E-Inter-Area-Prefix-LSA" that
defined in [RFC8362]

5. Extend LSA Operation Procedure

When ABR(for example R2 in Fig.1)receives the "Router LSA"
announcement in area 2, it should generate the corresponding "OSPFv2
Extended Prefix Opaque LSA" for OSPFv2 or "E-Inter-Area-Prefix-LSA"
for OSPFv3 that includes the sub TLV "Source Router ID" of the
network prefixes(for example, for prefix Lt1, N2 etc.), which labels
the source router of the corresponding link.

When S1 in another area receives such LSA, it then can know the
prefix Lt1 is associated with node T1, check the ELC, ERLD or MSD
value according to its necessary, and select the right label action
at the ingress node S1 for the traffic target to Lt1.

When R0 receives such LSA, it then strips this additional
information, put it into the corresponding part in BGP-LS protocol as
described in[I-D.ietf-idr-bgp-ls-segment-routing-ext] and reports
them to the IP SDN Controller, the SDN controller can then use such

information to build the inter-area topology according to the process described in the [Appendix A](#). The topology retrieval process may not suitable for some special environment as that stated in [Appendix B](#)

[6](#). Security Considerations

TBD.

[7](#). IANA Considerations

TBD.

[8](#). Acknowledgement

Very thanks Les Ginsberg for their valuable suggestions on the contents of this draft. And also thanks Jeff Tantsura, Rob Shakir for their valuable comments on this draft.

[9](#). References

[9.1](#). Normative References

[I-D.ietf-idr-bgp-ls-segment-routing-ext]
          Previdi, S., Talaulikar, K., Filsfils, C., Gredler, H.,
          and M. Chen, "BGP Link-State extensions for Segment
          Routing", [draft-ietf-idr-bgp-ls-segment-routing-ext-08](#)
          (work in progress), May 2018.

[I-D.ietf-ospf-mpls-elc]
          Xu, X., Kini, S., Sivabalan, S., Filsfils, C., and S.
          Litkowski, "Signaling Entropy Label Capability and Entropy
          Readable Label-stack Depth Using OSPF", [draft-ietf-ospf-mpls-elc-07](#) (work in progress), September 2018.

[I-D.ietf-ospf-segment-routing-msd]
          Tantsura, J., Chunduri, U., Aldrin, S., and P. Psenak,
          "Signaling MSD (Maximum SID Depth) using OSPF", [draft-ietf-ospf-segment-routing-msd-23](#) (work in progress),
          October 2018.

[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>.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, [RFC 2328](#),
          DOI 10.17487/RFC2328, April 1998,
          <https://www.rfc-editor.org/info/rfc2328>.

   [RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF
              for IPv6", RFC 5340, DOI 10.17487/RFC5340, July 2008,
              <https://www.rfc-editor.org/info/rfc5340>.

   [RFC7684] Psenak, P., Gredler, H., Shakir, R., Henderickx, W.,
              Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute
              Advertisement", RFC 7684, DOI 10.17487/RFC7684, November
              2015, <https://www.rfc-editor.org/info/rfc7684>.

   [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>.

   [RFC7794] Ginsberg, L., Ed., Decraene, B., Previdi, S., Xu, X., and
              U. Chunduri, "IS-IS Prefix Attributes for Extended IPv4
              and IPv6 Reachability", RFC 7794, DOI 10.17487/RFC7794,
              March 2016, <https://www.rfc-editor.org/info/rfc7794>.

   [RFC8362] Lindem, A., Roy, A., Goethals, D., Reddy Vallem, V., and
              F. Baker, "OSPFv3 Link State Advertisement (LSA)
              Extensibility", RFC 8362, DOI 10.17487/RFC8362, April
              2018, <https://www.rfc-editor.org/info/rfc8362>.

9.2. Informative References

   [I-D.wang-idr-bgpls-inter-as-topology-ext]
              Wang, A. and H. Chen, "BGP-LS Extension for Inter-AS
              Topology Retrieval", draft-wang-idr-bgpls-inter-as-
              topology-ext-02 (work in progress), August 2018.

Appendix A. Inter-Area Topology Retrieval Process

   When IP SDN Controller receives this information, it should compare
   the prefix NLRI that included in the BGP-LS packet. When it
   encounters the same prefix but with different source router ID, it
   should extract the corresponding area ID, rebuild the link between
   these two different source router in non-backbone area. Belows is
   one example that based on the Fig.1:

   Assuming we want to rebuild the connection between router S1 and
   router S2 that locates in area 1:

   a. Normally, router S1 will advertise prefix N1 within its router
       LSA

b. When this router LSA reaches the ABR router R1, it will convert
    it into summary LSA, add the "Source Router Information", which
    is router id of S1 in this example, as proposed in this draft.

c. R1 then floods this extension summary LSA to R0, which is running
    BGP-LS protocol with IP SDN Controller. The controller then
    knows the prefixes of N1 is from S1.

d. Router S2 will do the similar process, and the controller will
    also knows the prefixes N1 is also from S2

e. Then it can reconstruct the connection between S1 and S2, which
    prefix is N1. The topology within Area 1 can then be recovered
    accordingly.

Iterating the above process continuously, the IP SDN controller can
then retrieve the detail topology that span multi-area.

Appendix B. Special Considerations on Inter-Area Topology Retrieval

The above topology retrieval process can be applied in general
situation, where each prefix of the link between two nodes is planned
and allocated in normal space. But there are some situations not
belong to this and needs to be considered specially, for example when
the link is unnumbered and there are anycast prefixes deployed within
the network etc.

When ABR receives the unnumbered LSA, it will not advertise such LSA
into another area in the current OSPF specification. Considering
such situation is seldom exist in real network, here we will only
state explicitly that the above retrieval process is not suitable for
the network that deploys unnumbered links.

When there are anycast prefixes deployment within the network, if the
anycast prefix length is equal to 32, the controller can bypass them
easily because no prefix of the link will use such prefix length. If
the ansycast prefixes length is less than 32, it is acceptable that
connects the nodes advertising these anycast prefixes logically. Or,
if these anycast prefixes come from more than two nodes, the
controller can also detect such situation and label it explicitly.

Authors' Addresses

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

Email: wangaj.bri@chinatelecom.cn

Acee Lindem
Cisco Systems
301 Midenhall Way
Cary, NC 27513
USA

Email: acee@cisco.com

Jie Dong
Huawei Technologies
Beijing
China

Email: jie.dong@huawei.com

Ketan Talaulikar
Cisco Systems
S.No. 154/6, Phase I, Hinjawadi
Pune 411 057
India

Email: ketant@cisco.com

Peter Psenak
Cisco Systems
Pribinova Street 10
Bratislava, Eurovea Centre, Central 3 81109
Slovakia

Email: ppsenak@cisco.com