LSR Working Group   A. Wang
Internet-Draft   China Telecom
Intended status: Standards Track   A. Lindem
Expires: January 1, 2021   Cisco Systems
                                                        J. Dong
                                            Huawei Technologies
                                                      P. Psenak
                                                   K. Talaulikar
                                                   Cisco Systems
                                                  June 30, 2020

OSPF Prefix Originator Extensions
draft-ietf-lsr-ospf-prefix-originator-06

Abstract

   This document defines OSPF extensions to include information
   associated with the node originating a prefix along with the prefix
   advertisement.

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 January 1, 2021.

Table of Contents

1. Introduction

   Prefix attributes are advertised in OSPFv2 [RFC2328] using the
   Extended Prefix Opaque Link State Advertisement (LSA) [RFC7684] and
   in OSPFv3 [RFC5340] using the various Extended Prefix LSA types
   [RFC8362].

   The identification of the originating router for a prefix in OSPF
   varies by the type of the prefix and is currently not always
   possible. For intra-area prefixes, the originating router is
   identified by the advertising Router ID field of the area-scoped LSA
   used for those prefix advertisements. However, for the inter-area
   prefixes advertised by the Area Border Router (ABR), the advertising
   Router ID field of their area-scoped LSAs is set to the ABR itself
   and the information about the router originating the prefix
   advertisement is lost in this process of prefix propagation across
   areas. For Autonomous System (AS) external prefixes, the originating
   router may be considered as the Autonomous System Border Router
   (ASBR) and is identified by the advertising Router ID field of the
   AS-scoped LSA used. However, the actual originating router for the
   prefix may be a remote router outside the OSPF domain. Similarly,
   when an ABR performs translation of Not-So-Stubby Area (NSSA)
   [RFC3101] LSAs to AS-external LSAs, the information associated with

the NSSA ASBR (or the router outside the OSPF domain) is not conveyed across the OSPF domain.

While typically the originator of information in OSPF is identified by its OSPF Router ID, it does not necessarily represent a reachable address for the router. The IPv4/IPv6 Router Address as defined in [RFC3630] and [RFC5329] for OSPFv2 and OSPFv3 respectively provide an address to reach that router.

The primary use case for the extensions proposed in this document is to be able to identify the originator of the prefix in the network. In cases where multiple prefixes are advertised by a given router, it is also useful to be able to associate all these prefixes with a single router even when prefixes are advertised outside of the area in which they originated. It also helps to determine when the same prefix is being originated by multiple routers across areas.

This document proposes extensions to the OSPF protocol for inclusion of information associated with the router originating the prefix along with the prefix advertisement. These extensions do not change the core OSPF route computation functionality. They provide useful information for topology analysis and traffic engineering, especially on a controller when this information is advertised as an attribute of the prefixes via mechanisms such as Border Gateway Protocol Link-State (BGP-LS) [RFC7752].

Applications related to use of the prefix originating node information for topology reconstruction process on a controller and the associated limitations are described in Appendix A and Appendix B.

1.1. 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. Protocol Extensions

This document defines the Prefix Source Router-ID and the Prefix Originator Sub-TLVs for inclusion of the Router ID and a reachable address information for the router originating the prefix as a prefix attribute.

2.1. Prefix Source Router-ID Sub-TLV

   For OSPFv2, the Prefix Source Router-ID Sub-TLV is an optional Sub-
   TLV of the OSPFv2 Extended Prefix TLV [RFC7684]. For OSPFv3, the
   Prefix Source Router-ID Sub-TLV is an optional Sub-TLV of the Intra-
   Area-Prefix TLV, Inter-Area-Prefix TLV, and External-Prefix TLV
   [RFC8362] when originating either an IPv4 [RFC5838] or an IPv6 prefix
   advertisement.

   The Prefix Source Router-ID Sub-TLV has 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             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                        OSPF Router ID                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

           Figure 1: Prefix Source Router-ID Sub-TLV Format

    Where:

   o Type: 4 for OSPFv2 and 27 for OSPFv3

   o Length: 4

   o OSPF Router ID : the OSPF Router ID of the OSPF router that
      originated the prefix advertisement in the OSPF domain.

   A prefix advertisement MAY include more than one Prefix Source
   Router-ID sub-TLV, one corresponding to each of the Equal-Cost Multi-
   Path (ECMP) nodes that originated the given prefix.

   A received Prefix Source Router-ID Sub-TLV with OSPF Router ID set to
   0 MUST be considered invalid and ignored. Additionally, reception of
   such Sub-TLV SHOULD be logged as an error (subject to rate-limiting).

2.2. Prefix Originator Sub-TLV

   For OSPFv2, the Prefix Originator Sub-TLV is an optional Sub-TLV of
   the OSPFv2 Extended Prefix TLV [RFC7684]. For OSPFv3, the Prefix
   Originator Sub-TLV is an optional Sub-TLV of the Intra-Area-Prefix
   TLV, Inter-Area-Prefix TLV, and External-Prefix TLV [RFC8362] when
   originating either an IPv4 [RFC5838] or an IPv6 prefix advertisement.
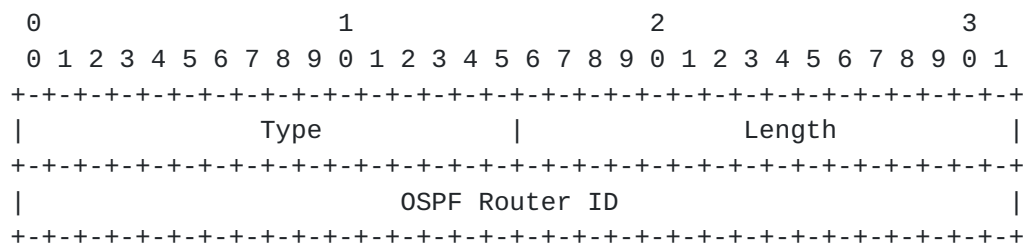
   The Prefix Originator Sub-TLV has the following format:

```
  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            |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |             Router Address (4 or 16 octects)                 |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

             Figure 2: Prefix Originator Sub-TLV Format

   Where:

   o Type: TBD1 for OSPFv2 and TBD2 for OSPFv3

   o Length: 4 or 16

   o Router Address: A reachable IPv4 or IPv6 router address for the
      router that originated the IPv4 or IPv6 prefix advertisement.
      Such an address would be semantically equivalent to what may be
      advertised in the OSPFv2 Router Address TLV [RFC3630] or in the
      OSPFv3 Router IPv6 Address TLV [RFC5329].

   A prefix advertisement MAY include more than one Prefix Originator
   sub-TLV, one corresponding to each of the Equal-Cost Multi-Path
   (ECMP) nodes that originated the given prefix.

   A received Prefix Originator Sub-TLV that has an invalid length (not
   4 or 16) or a Reachable Address containing an invalid IPv4 or IPv6
   address (dependent on address family of the associated prefix) MUST
   be considered invalid and ignored. Additionally, reception of such
   Sub-TLV SHOULD be logged as an error (subject to rate-limiting).

   [RFC7794] provides similar functionality for the Intermediate System
   to Intermediate System (IS-IS) protocol.

3. Elements of Procedure

   This section describes the procedure for advertisement of the Prefix
   Source Router-ID and Prefix Originator Sub-TLVs along with the prefix
   advertisement.

   The OSPF Router ID of the Prefix Source Router-ID is set to the OSPF
   Router ID of the node originating the prefix in the OSPF domain.

   If the originating node is advertising an OSPFv2 Router Address TLV
   [RFC3630] or an OSPFv3 Router IPv6 Address TLV [RFC5329], then that
   value is set in the Router Address field of the Prefix Originator
   Sub-TLV. When the orignating node is not advertising such an

address, implementations MAY support mechanisms to determine a
reachable address belonging to the originating node to set in the
Router Address field. Such mechanisms are outside the scope of this
document.

Implementations MAY support the selection of specific prefixes for
which the originating node information needs to be included with
their prefix advertisements.

When an ABR generates inter-area prefix advertisements into its non-
backbone areas corresponding to an inter-area prefix advertisement
from the backbone area, the only way to determine the originating
node information is based on the Prefix Source Router-ID and Prefix
Originator Sub-TLVs present in the inter-area prefix advertisement
originated into the backbone area by an ABR for another non-backbone
area. The ABR performs its prefix calculation to determine the set
of nodes that contribute to the best prefix reachability. It MUST
use the prefix originator information only from this set of nodes.
The ABR MUST NOT include the Prefix Source Router-ID or the Prefix
Originator Sub-TLVs when it is unable to determine the information of
the best originating node.

Implementations MAY provide control on ABRs to selectively disable
the propagation of the originating node information across area
boundaries.

Implementations MAY support the propagation of the originating node
information along with a redistributed prefix into the OSPF domain
from another routing domain. The details of such mechanisms are
outside the scope of this document. Such implementations MAY also
provide control on whether the Router Address in the Prefix
Originator Sub-TLV is set as the ABSR node address or as the address
of the actual node outside the OSPF domain that owns the prefix.

When translating the NSSA prefix advertisements [RFC3101] to the AS
external prefix advertisements, the NSSA ABR, follows the same
procedures as an ABR generating inter-area prefix advertisements for
the propagation of the originating node information.

4. Security Considerations

Since this document extends the OSPFv2 Extended Prefix LSA, the
security considerations for [RFC7684] are applicable. Similarly,
since this document extends the OSPFv3 E-Intra-Area-Prefix-LSA, E-
Inter-Area-Prefix-LSA, E-AS-External LSA and E-NSSA-LSA, the security
considerations for [RFC8362] are applicable.

5. IANA Considerations

   This document requests IANA for the allocation of the codepoint from
   the "OSPFv2 Extended Prefix TLV Sub-TLVs" registry under the "Open
   Shortest Path First v2 (OSPFv2) Parameters" registry.

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Code  |              Description            |   IANA Allocation   |
| Point |                                     |       Status        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   4   |  Prefix Source Router-ID Sub-TLV    | early allocation done |
|  TBD1 |    Prefix Originator Sub-TLV        |       pending       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

      Figure 3: Code Points in OSPFv2 Extended Prefix TLV Sub-TLVs

   This document requests IANA for the allocation of the codepoint from
   the "OSPFv3 Extended Prefix TLV Sub-TLVs" registry under the "Open
   Shortest Path First v3 (OSPFv3) Parameters" registry.

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Code  |              Description            |   IANA Allocation   |
| Point |                                     |       Status        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  27   |  Prefix Source Router-ID Sub-TLV    | early allocation done |
|  TBD2 |    Prefix Originator Sub-TLV        |       pending       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

      Figure 4: Code Points in OSPFv3 Extended-LSA Sub-TLVs

6. Acknowledgement

   Many thanks to Les Ginsberg for his suggestions on this draft. Also
   thanks to Jeff Tantsura, Rob Shakir, Gunter Van De Velde, Goethals
   Dirk, Smita Selot, Shaofu Peng, and John E Drake for their valuable
   comments.

7. References

7.1. 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>.

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

   [RFC3101] Murphy, P., "The OSPF Not-So-Stubby Area (NSSA) Option",
              RFC 3101, DOI 10.17487/RFC3101, January 2003,
              <https://www.rfc-editor.org/info/rfc3101>.

   [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>.

   [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>.

   [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>.

   [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>.

7.2. Informative References

   [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering
              (TE) Extensions to OSPF Version 2", RFC 3630,
              DOI 10.17487/RFC3630, September 2003,
              <https://www.rfc-editor.org/info/rfc3630>.

   [RFC5329] Ishiguro, K., Manral, V., Davey, A., and A. Lindem, Ed.,
              "Traffic Engineering Extensions to OSPF Version 3",
              RFC 5329, DOI 10.17487/RFC5329, September 2008,
              <https://www.rfc-editor.org/info/rfc5329>.

   [RFC5838] Lindem, A., Ed., Mirtorabi, S., Roy, A., Barnes, M., and
              R. Aggarwal, "Support of Address Families in OSPFv3",
              RFC 5838, DOI 10.17487/RFC5838, April 2010,
              <https://www.rfc-editor.org/info/rfc5838>.

   [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>.

Appendix A. Inter-Area Topology Retrieval Process

   When an IP SDN Controller receives BGP-LS [RFC7752] information, it
   should compare the prefix Network Layer Reachability Information
   (NLRI) that is included in the BGP-LS NLRI. 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 source
   routers in the non-backbone area. Below is one example that based on
   the Figure 5 which illustrates a topology where OSPF is running in
   multiple areas.

```
                         +-----------------+
                         |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    |
      +--------------------+--------------+-------------------+
```

              Figure 5: OSPF Inter-Area Prefix Originator Scenario

   R0-R4 are routers in the backbone area, S1-S4 are internal routers in
   area 1, and T1-T4 are internal routers in area 2. R1 and R3 are ABRs
   between area 0 and area 1. R2 and R4 are ABRs between area 0 and
   area 2. N1 is the network between router S1 and S2 and N2 is the
   network between router T1 and T2. Ls1 is the loopback address of
   Node S1 and Lt1 is the loopback address of Node T1.

Assuming we want to rebuild the connection between router S1 and
router S2 located 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-ID Sub-TLV and the
    Prefix Originator Sub-TLV, as described in Section 3.

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

d. Router S2 will perform a similar process, and the controller will
    also learn that prefix N1 is also from S2.

e. Then it can reconstruct the link between S1 and S2, using the
    prefix N1. The topology within Area 1 can then be reconstructed
    accordingly.

Iterating the above process continuously, the IP SDN controller can
retrieve a detailed topology that spans multiple areas.

Appendix B. Special Considerations on Inter-Area Topology Retrieval

The above topology retrieval process can be applied in the case where
each point-to-point or multi-access link connecting routers is
assigned a unique prefix. However, there are some situations where
this heuristic cannot be applied. Specifically, the cases where the
link is unnumbered or the prefix corresponding to the link is an
anycast prefix.

The Appendix A heuristic to rebuild the topology can normally be used
if all links are numbered. For anycast prefixes, if it corresponds
to the loopback interface and has a host prefix length, i.e., 32 for
IPv4 prefixes and 128 for IPv6 prefixes, Appendix A can also applied
since these anycast prefixes are not required to reconstruct the
topology.

Authors' Addresses

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

Email: wangaj3@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

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

Email: ppsenak@cisco.com

Ketan Talaulikar
Cisco Systems
India

Email: ketant@cisco.com