PCE Working Group D. Dhody
Internet-Draft U. Palle
Intended status: Standards Track Huawei Technologies India Pvt Ltd
Expires: September 1, 2012
February 29, 2012

OSPF Protocol Extensions for Boundary Node Discovery (BND)
draft-dhody-pce-bn-discovery-ospf-04

Abstract

There are various circumstances where it is highly desirable to be able to dynamically and automatically discover a set of Boundary Nodes (BN) along with their domain information. For that purpose, this document defines extensions to the Open Shortest Path First (OSPF) routing protocol for the advertisement of Boundary Node (BN) Discovery information within an OSPF area or within the entire OSPF routing domain.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

This document defines extensions to OSPFv2 [RFC2328] and OSPFv3 [RFC5340] to allow a boundary node in an OSPF routing domain to advertise its location, along with domain information.

Generic capability advertisement mechanisms for OSPF are defined in [RFC4970]. These allow a router to advertise its capabilities within an OSPF area or an entire OSPF routing domain. This document leverages this generic capability advertisement mechanism to fully satisfy the dynamic BN discovery.

This document defines a new TLV (named the Boundary Node Discovery TLV (BND TLV)) to be carried within the OSPF Router Information LSA ([RFC4970]).

The Boundary Node information advertised is detailed in Section 3. Protocol extensions and procedures are defined in Section 9 and Section 10.

A detailed description about the need for auto discovery of Boundary Nodes (BN) and thier domains is also provided in this document.

The OSPF extensions defined in this document allow for Boundary Node discovery within an OSPF routing domain. Boundary Node can be an ABR or ASBR.

1.1. 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 [RFC2119].

2. Terminology

The following terminology is used in this document.

ABR: OSPF Area Border Router. Routers used to connect two IGP areas.

AS: Autonomous System.

ASBR: Autonomous System Border Router. Router used to connect together ASes of the same or different service providers via one or more inter-AS links

BN: A boundary node is either an ABR in the context of inter-area Traffic Engineering or an ASBR in the context of inter-AS Traffic Engineering.

BND: Boundary Node Discovery

BRPC: Backward Recursive Path Computation

Domain: Any collection of network elements within a common sphere of address management or path computational responsibility. Examples of domains include Interior Gateway Protocol (IGP) areas and Autonomous Systems (ASs).

H-PCE: Hierarchical PCE.

IGP: Interior Gateway Protocol. Either of the two routing protocols, Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS).

LSA: Link State Advertisement.

OSPF: Open Shortest Path First.

PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

TLV: Type-Length-Variable data encoding.

3. Overview

3.1. Boundary Node (BN) Discovery Information

The BN discovery information is composed of:

- o The BN location: an IPv4 and/or IPv6 address that is used to reach the BN. It is RECOMMENDED to use an address that is always reachable from all connected domains;

- o The set of two or more Domain(s) into which the BN has connectivity;

Changes in BN discovery information may occur as a result of BN configuration update or domain status change.

3.2. Flooding Scope

The flooding scope for BN information advertised through OSPF can be limited to OSPF area(s) the BN belongs to, or can be extended across the entire OSPF routing domain.

4. Applications

BRPC procedure as defined in [RFC5441], requires Path Computation Element (PCE) [RFC4655] to be aware of the BNs for the inter-domain path computation. This information would be either statically configured at PCE or learned via some mechanism, as listed in Section 5.

In case of static configuration, as shown in the Figure 1, incase of OSPF Area0, configuration of BNs at PCE5 is extensive. BRPC procedure guarantees a best path only if BNs are selected correctly, any change in BNs at run time may lead to sub-optimal path. Also Administrator need to configure ABR / ASBR ID in such a way that it is reachable from all the domains, BND Tlv can take care of this automatically.

```
                           +--------------------+
                           |             +-----+|
                           |      Area 2 | PCE2||
                           |             +-----+|
                           |                    |
                           |                    |
                           |     BN4+----+      |
                           +--------+----+------+
                         +----------+----+---------+
                         |          +----+         |
                         |                         |
                         |                         |
                         |                         |
       +---------------+ |                         |+----------------+
       |           BN1 | |                         ||  BN5           |
       |             +-+-++                       +++--+             |
       |             | | ||                       |||  |             |
       |             +-+-++                       +++--+             |
       |               | |                         ||                |
       |           BN2 | |                         ||  BN6           |
       |             +-+-++       Area 0          +++--+             |
       |  Area 1     | | ||                       |||  |  Area 3     |
       |             +-+-++                       +++--+             |
       |               | |                         ||                |
       |           BN3 | |          +-----+        ||  BN7           |
       |+-----+      +-+-++         | PCE5|       +++--+     +-----+ |
       || PCE1|      | | ||         +-----+       |||  |     | PCE3| |
       |+-----+      +-+-++                       +++--+     +-----+ |
       |               | |                         ||                |
       +---------------+ |                         |+----------------+
                         |                         |
                         |                         |
                         |          +----+         |
                         +----------+----+---------+
                           +--------+----+------+
                           |    BN8 +----+      |
                           |                    |
                           |                    |
                           |      Area 4 +-----+|
                           |             | PCE4||
                           |             +-----+|
                           +--------------------+
```

Figure 1: OSPF Area Topology

The problems with existing mechanism to discover Boundary nodes are listed in the next section.

5. Existing Mechanisms

5.1. OSPF LSA

o E bit and B bit of Router LSA defined in [RFC2328] can help in finding a router acting as ABR/ASBR but there is no way to find out the domain information of this ABR/ASBR. As stated in Section 7, Selection of correct BN is based on domain and thus it is ineffective.

o Selection of ABR based on summary LSA or ASBR based on AS-external LSA is not a good idea, first it requires PCE to look into LSDB thus adding to coupling, second it MAY require BGP routes to be redistributed into OSPF which is also not a good network design principle.

5.2. Inter-AS TE Link

o [RFC5392] specifies how to advertise TE properties of inter-AS links; through which ASBR and remote AS can be discovered, but ABR and their domain information cannot be discovered via above RFC.

o AS is made up of multiple Area, there maybe a need to clearly identify a BN by combination of both AS number and Area-id. Refer Section 3.5 of [DOMAIN-SEQ].

o AS shown in below figure, just the knowledge of AS 100 and AS 200 is not sufficient, the BN should have both AS and Area information. The area information cannot be provided by [RFC5392].

```
                    |
                    |  +-------------+                +----------------+
                    |  |Area 2       |                |Area 4          |
                    |  |         +--+|                |          +--+  |
                    |  |         |  ||                |          |  |  |
                    |  |  +--+   +--+|                |   +--+   +--+  |
                    |  |  |  |       |                |   |  |         |
                    |  |  *--+       |                |   +--+         |
                    |  | /      +--+ |                |          +--+  |
                    |  |/       |  | |                |          |  |  |
                    |  /        +--+ |                |   +--+   +--+  |
                    | /|  +--+       |+--------------+|   |  |         |
                    |/ |  |  |       ++-+          +-++   +--+         |
     +-------------+/  |  +--+       || |          | ||                |
     |             /|  |             ++-+          +-++                |
     |         +--*||  +-------------+|              |+----------------+
     |         |  |||                 |     +--+     |
     |         +--+||                 |     |  |     |
     |    +--+     ||                 |     +--+     |
     |    |  |     ||                 |              |
     |    +--+     ||                 |              |
     |             ||                 |     +--+     |
     |+--+         ||                 |     |  |     |
     ||  |         ||                 |     +--+     |
     |+--+         ||                 |              |
     |             ||                 |     +--+     |
     |    +--+     ||  +------------+ |     |  |     |+----------------+
     |    |  |     ||  |Area 3      +-++    +--+   +-++ Area 5         |
     |    +--+     ||  |            | ||           | ||                |
     |             ||  |            +-++           +-++                |
     |         +--+||  |       +--+ | |  Area 0      ||   +--+         |
     |         |  |||  |       |  | | +--------------+|   |  |         |
     |         +--*||  |       +--+ |                 |   +--+         |
     |             \|  |            |                 |          +--+  |
     |Area 1       |\  |   +--+     |                 |   +--+   |  |  |
     +-------------+|\ |   |  |     |                 |   |  |   +--+  |
                    | \|   +--+  +--+                 |   +--+         |
                    |  \         |  |                 |                |
                    |  |\        +--+                 |          +--+  |
                    |  | \ +--+     |                 |          |  |  |
                    |  |  \|  |     |                 |          +--+  |
                    |  |   *--+     |                 |                |
                    |  |            |                 |                |
                    |  +------------+                 +----------------+
                    |
                    |
         As 100     |  AS 200
                    |
```

5.3. OSPF Area Topology

- o Consider OSPF topology as shown in Figure 1, One can make a generalization that all ABR connects to Area 0, and one doesnt need the area information.

- o But as per [RFC3509], there are alternative interpretaion of ABR, which needs to be considered. If there is a connectivity between Area 1 and Area 2 directly that needs to be considered. In the scope of path computation. it is not correct to force all paths to go through area 0, irrespective of actual topology.

```
         .                .
          .    Area 0    .
           +--+      +--+
         ..|R1|..  ..|R2|..
        .  +--+  ..  +--+  .
        .        ..        .
        .       +--+       .
        . Area1 |R3| Area2 .
        .       +--+  +--+ .
        .        ..   |R4| .
        .       .  .  +--+ .
         .......    .......
```

5.4. HPCE

- o Section 4 of [H-PCE] specifies each child PCE should know the identity of the domains that neighbor its own domain and advertises the same to the parent PCE. No method exist to find the neighbor domain which need to be carried in NEIG-PCE-DOMAIN Sub-TLV and BN discovery along with neighbor domain information can help in generating NEIG-PCE-DOMAIN Sub-TLV.

6. Static Configurations

A simple solution would be to configure BNs [ABR and ASBR] at PCE(s) along with their domain information. As this information is fairly static this could work in simple situations. But as PCE are being used in bigger and multiple domains, any sort of static configurations would put extra effort on the system administrator. Selection of correct BNs is the core of the BRPC procedure, we feel this information should be dynamically learned and maintained.

7. Importance of Domain Information along with BNs

There are methods to learn BNs dynamically from IGP, but the knowledge of neighboring-domains is not possible to obtain. Without

this the correct BN based on the domain-path can not be selected. [RFC5441] mentions:

"Note that PCE(i) only considers the entry BNs of domain(i), i.e., only the BNs that provide connectivity from domain(i-1). In other words, the set BN-en(k,i) is only made of those BNs that provide connectivity from domain (i-1) to domain(i)."

This selection of correct BNs providing connectivity between correct domains cannot be made by the information obtained from IGP. Without the correct selection we would not be following [RFC5441].

8. Relationship to Domain-Sequence

[DOMAIN-SEQ] provides a standard representation of Domain Sequence in all deployment scenarios. The Domain Information carried in the BN-DOMAIN sub-tlv is same as the sub-objects inside the domain sequence.

9. The OSPF BND TLV

The OSPF BN Discovery TLV (BND TLV) contains a non-ordered set of sub-TLVs.

The format of the OSPF BND TLV and its sub-TLVs is identical to the TLV format used by the Traffic Engineering Extensions to OSPF [RFC3630]. That is, the TLV is composed of 2 octets for the type, 2 octets specifying the TLV length, and a value field. The Length field defines the length of the value portion in octets.

The OSPF BND TLV has the following format:

```
                     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            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                            sub-TLVs                         //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

```
   Type:      To be assigned by IANA (suggested value 8)
   Length:    Variable
   Value:     This comprises of following sub-TLVs

   Two sub-TLVs are defined:
         Sub-TLV type  Length            Name
               1      variable     BN-ADDRESS sub-TLV
               2         4         BN-DOMAIN sub-TLV
```

The BN-ADDRESS and BN-DOMAIN sub-TLVs MUST always be present within the BND TLV.

Malformed BND TLVs or sub-TLVs not explicitly described in this document MUST cause the LSA to be treated as malformed according to the normal procedures of OSPF.

Any unrecognized sub-TLV MUST be silently ignored.

The BND TLV is carried within an OSPF Router Information LSA defined in [RFC4970].

The following sub-sections describe the sub-TLVs which are carried within the BND TLV.

9.1. BN-ADDRESS Sub-TLV

The BN-ADDRESS sub-TLV specifies an IP address that can be used to reach the BN. It is RECOMMENDED to make use of an address that is always reachable, provided that the BN is alive and reachable.

The BN-ADDRESS sub-TLV is mandatory; it MUST be present within the BND TLV. It MAY appear twice, when the BN has both an IPv4 and IPv6 address. It MUST NOT appear more than once for the same address type. If it appears more than once for the same address type, only the first occurrence is processed and any others MUST be ignored.

The format of the BN-ADDRESS sub-TLV is as follows:

```
                         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 = 1           |             Length            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |          address-type         |            Reserved           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    //                       BN IP Address                         //
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

BN-ADDRESS sub-TLV format

Type: 1
Length: 8 (IPv4) or 20 (IPv6)

Address-type:
- 1   IPv4
- 2   IPv6

Reserved: SHOULD be set to zero on transmission and MUST be ignored on receipt.

BN IP Address: The IP address to be used to reach the BN.

9.2. BN-DOMAIN Sub-TLV

The BN-DOMAIN sub-TLV specifies a BN-Domain (area or AS) where the BN has topology connectivity.

The BN-DOMAIN sub-TLV is mandatory; it MUST be present within the BND TLV.

A BND TLV MUST include two or more BN-DOMAIN sub-TLVs as the BN has connectivity into multiple BN-Domains.

The BN-DOMAIN sub-TLV has the following format:

```
                        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 = 2            |             Length            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Domain-type          |            Reserved           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           Domain ID                           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

BN-DOMAIN sub-TLV format

Type: 2
Length: 8

Two domain-type values are defined:
1 OSPF Area ID
2 AS Number

Domain ID: With the domain-type set to 1, this indicates the 32-bit Area ID of an area where the BN (ABR) has connectivity. With domain-type set to 2, this indicates an AS number of an AS where the BN (ASBR) has connectivity. When the AS number is coded in two octets, the AS Number field MUST have its first two octets set to 0.

10. Elements of Procedure

The BND TLV is advertised within OSPFv2 Router Information LSAs (Opaque type of 4 and Opaque ID of 0) or OSPFv3 Router Information LSAs (function code of 12), which are defined in [RFC4970]. As such, elements of procedure are inherited from those defined in [RFC4970].

In OSPFv2, the flooding scope is controlled by the opaque LSA type(as defined in [RFC5250]) and in OSPFv3, by the S1/S2 bits (as defined in [RFC5340]). If the flooding scope is area local, then the BND TLV MUST be carried within an OSPFv2 type 10 router information LSA or an OSPFV3 Router Information LSA with the S1 bit set and the S2 bit clear. If the flooding scope is the entire IGP domain, then the BND TLV MUST be carried within an OSPFv2 type 11 Router Information LSA or OSPFv3 Router Information LSA with the S1 bit clear and the S2 bit set.

When the BN function is deactivated, the OSPF speaker MUST originate a new Router Information LSA that no longer includes the corresponding BND TLV, provided there are other TLVs in the LSA. If there are no other TLVs in the LSA, it MUST either send an empty Router Information LSA or purge it by prematurely aging it.

The BN address (i.e., the address indicated within the BN-ADDRESS sub-TLV) SHOULD be reachable via some prefixes advertised by OSPF.

The BND TLV information regarding a specific BN is only considered current and useable when the router advertising this information is itself reachable via OSPF calculated paths in the same area of the LSA in which the BND TLV appears.

A change in the state of a BN (activate, deactivate, domain change) MUST result in a corresponding change in the BND TLV information advertised by an OSPF router (inserted, removed, updated)in its LSA. The way BNs determine the information they advertise, and how that information is made available to OSPF, is out of the scope of this document. Some information may be configured and other information may be automatically determined by the OSPF.

A change in information in the BND TLV MUST NOT trigger any SPF computation at a receiving router.

11. Backward Compatibility

The BND TLV defined in this document does not introduce any interoperability issues.

A router not supporting the BND TLV will just silently ignore the TLV as specified in [RFC4970].

12. Impact on Network

The routers acting as BNs will originate Opaque LSA with BND Tlv; As there are only few BNs exist in the network, the performance impact in flooding is very less.

13. IANA Considerations

13.1. OSPF TLV

IANA has defined a registry for TLVs carried in the Router Information LSA defined in [RFC4970]. A number of IANA considerations have been highlighted in previous sections of this document. IANA is requested to make the following allocations.

Value	TLV Name	Reference
To be assigned by IANA	BND	(this document)

14. Security Considerations

This document defines OSPF extensions for BN discovery within an administrative domain. Hence the security of the BN discovery relies on the security of OSPF.

Mechanisms defined to ensure authenticity and integrity of OSPF LSAs [RFC2154], and their TLVs, can be used to secure the BN Discovery information as well.

OSPF provides no encryption mechanism for protecting the privacy of LSAs and, in particular, the privacy of the BN discovery information.

15. Manageability Considerations

TBD

16. Acknowledgments

We would like to thank Quintin Zhao, Daniel King, Adrian Ferral, Suresh babu, Pradeep Shastry, Saravana Kumar and Srinivasan for their useful comments and suggestions.

17. References

17.1. Normative References

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

[RFC2154] Murphy, S., Badger, M., and B. Wellington, "OSPF with Digital Signatures", RFC 2154, June 1997.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

[RFC4970] Lindem, A., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 4970, July 2007.

[RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, July 2008.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF

for IPv6", RFC 5340, July 2008.

17.2. Informative References

[RFC3509] Zinin, A., Lindem, A., and D. Yeung, "Alternative Implementations of OSPF Area Border Routers", RFC 3509, April 2003.

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

[RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5392, January 2009.

[RFC5441] Vasseur, JP., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.

[H-PCE] King, D. and A. Farrel , "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS & GMPLS. (draft-ietf-pce-hierarchy-fwk-00)", October 2011.

[DOMAIN-SEQ] Dhody, D., Palle, U., and R. Casellas, "Standard Representation Of Domain Sequence (draft-dhody-pce-pcep-domain-sequence-02)", February 2012.

Authors' Addresses

Dhruv Dhody
Huawei Technologies India Pvt Ltd
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: dhruv.dhody@huawei.com

Udayasree Palle
Huawei Technologies India Pvt Ltd
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: udayasree.palle@huawei.com