Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 7, 2018

Z. Chen
X. Xu
Huawei Technologies

January 3, 2018

Overheads Reduction for IS-IS Enabled Spine-Leaf Networks
draft-chen-isis-sl-overheads-reduction-02

Abstract

When a Spine-Leaf topology adopts the Intermediate System to Intermediate System (IS-IS) routing protocol, the Leaf node receives Link State Packets (LSPs) from all the other nodes thus having the entire routing information of the topology. This is usually considered unnecessary and costly. This document describes a solution to this problem by utilizing IS-IS's inherent multi-level and area partition features, which requires that an IS-IS router SHOULD check a level-1 LSP's area addresses before advertising it to a neighbor.

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].

Table of Contents

1. Introduction

Spine-Leaf topology (a.k.a., CLOS topology) is widely used in today's datacenter and campus networks. When the Spine-Leaf topology runs the Intermediate System to Intermediate System (IS-IS) routing protocol, each Leaf node receives Link State Packets (LSPs) from all the other nodes thus having the entire routing information of the topology. This is usually considered unnecessary and costly because the Leaf node only needs to know its default gateways (i.e., the Spine nodes it connects to) and the LSPs generated by the other Leaf nodes bring little benefit for it to forward traffic.

To avoid Leaf nodes from learning the unnecessary LSPs from one another, [IS-IS-SL-Extension] proposes a new TLV attached to the IS-IS Hello (IIH) PDU to carry an router's role (i.e., Spine or Leaf) in the topology. The Spine nodes then prevent all LSPs from being sent

to the Leaf nodes, and each Leaf node sets the Spine nodes it connects to as its default gateways.

This document proposes another solution to this problem, which utilizes IS-IS's inherent multi-level and area partition features. In particular, it requires that each Leaf node (configured as L1 router) SHOULD be assigned with a unique area address and each Spine node (configured as L1/L2 router) MUST NOT advertise level-1 LSPs of a given area to neighbors within another area. This prevents Leaf nodes from receiving routing information from one another, without introducing new message formats.

2. Solution Description

2.1. Area Address Assignment

```
            +------------+                      +------------+
            |  Spine-A   |    10.10.10.0/24     |  Spine-B   |
            |  L1/L2     +----------------------+  L1/L2     |
            |  Area10/20 | .1                .2 |  Area10/20 |
            +---+--+-----+                      +---+----+---+
             .1 |  | .1                          .2 |    | .1
                |  |          10.10.40.0/24         |    |
                |  |  +-----------------------------+    |
  10.10.20.0/24 |  |  |                                  | 10.10.30.0/24
                |  +--|-------------------------------+  |
                |     |       10.10.50.0/24           |  |
             .2 |     | .1                         .2 |  | .2
            +---+-----+--+                      +-----+--+---+
            |  Leaf-A    |                      |  Leaf-B    |
            |  L1        |                      |  L1        |
            |  Area10    |                      |  Area20    |
            +-----+------+                      +-----+------+
                  |                                   |
                  |                                   |
            ------+-------                      ------+-------
            192.168.10.0/24                     192.168.20.0/24
```

Figure 1: Topology Example

This section describes how to assign area addresses in the Spine-Leaf topology, and illustrates why IS-IS routers SHOULD check the area addresses before advertising level-1 LSPs. As shown in Figure 1, there are two Spine nodes (i.e., Spine-A and Spine-B) and two Leaf nodes (i.e., Leaf-A and Leaf-B). The System IDs of Spine-A, Spine-B, Leaf-A, and Leaf-B are 1111.1111.1111.00 to 4444.4444.4444.00, respectively.

To prevent a Leaf node from learning the routing information of the other ones, the following configurations are REQUIRED:

a. Leaf nodes SHOULD be configured as L1 routers and each of them SHOULD be assigned a unique area address.

b. Spine nodes SHOULD be configured as L1/L2 routers and SHOULD be assigned multiple area addresses with each being that of a given Leaf node connected to it.

As a result, Leaf-A and Leaf-B in Figure 1 are configured as L1 routers and are assigned 10 and 20 as their area addresses, respectively. Spine-A and Spine-B are configured as L1/L2 routers and are assigned both 10 and 20 as their area addresses.

Level-1 Link State Database (Spine-A):

LSPID	Seq Num	Checksum	Holdtime	Length	ATT/P/OL	Area
1111.1111.1111.00-00	0x0000006c	0x540b	743	124	0/0/0	10/20
2222.2222.2222.00-00	0x0000006d	0x933b	1068	124	0/0/0	10/20
3333.3333.3333.00-00	0x0000006b	0x1815	402	122	0/0/0	10
4444.4444.4444.00-00	0x0000006a	0xf543	431	122	0/0/0	20

Level-2 Link State Database (Spine-A):

LSPID	Seq Num	Checksum	Holdtime	Length	ATT/P/OL	Area
1111.1111.1111.00-00	0x0000006f	0x682f	743	150	0/0/0	10/20
2222.2222.2222.00-00	0x00000063	0x30eb	1068	150	0/0/0	10/20

Figure 2: Link State Database of Spine-A

Under such configurations, however, Leaf-A still receives Leaf-B's LSPs (and vice versa) even though they are in different areas. This is because of the IS-IS definition that all routers in a specific area SHOULD share the same level-1 Link State Database (LSDB). In other words, IS-IS routers check area addresses during neighbor establishment, but are regardless of area addresses when advertising LSPs to a neighbor.

The example in Figure 1 and the LSDB of Spine-A (in Figure 2) further illustrate this. Since Spine-A and Leaf-B are both in area 20, Spine-A will receive LSP 4444.4444.4444.00-00 from Leaf-B and store the LSP into its level-1 LSDB. On the other hand, since Spine-A and Leaf-A are both in area 10, Spine-A will advertise LSP 4444.4444.4444.00-00 to Leaf-A although Leaf-A and Leaf-B (generator of the LSP) are in different areas. As a result, Leaf-A installs the route 192.168.20.0/24 into its routing table (Figure 3), even though it is an external area route.

Leaf-A Routing Table:

Destination	Proto	Pre	Cost	Flags	NextHop	Interface
10.10.10.0/24	ISIS-L1	15	20	D	10.10.20.1	Ethernet0/0/0
	ISIS-L1	15	20	D	10.10.40.2	Ethernet0/0/1
10.10.20.0/24	Direct	0	0	D	127.0.0.1	Ethernet0/0/0
10.10.30.0/24	ISIS-L1	15	20	D	10.10.40.2	Ethernet0/0/1
10.10.40.0/24	Direct	0	0	D	127.0.0.1	Ethernet0/0/1
10.10.50.0/24	ISIS-L1	15	20	D	10.10.20.1	Ethernet0/0/0
192.168.10.0/24	Direct	0	0	D	127.0.0.1	GEthernet0/0/0
192.168.20.0/24	ISIS-L1	15	30	D	10.10.20.1	Ethernet0/0/0
	ISIS-L1	15	30	D	10.10.40.2	Ethernet0/0/1
127.0.0.0/8	Direct	0	0	D	127.0.0.1	InLoopBack0
0.0.0.0/0	ISIS-L1	15	10	D	10.10.20.1	Ethernet0/0/0
	ISIS-L1	15	10	D	10.10.40.2	Ethernet0/0/1

Figure 3: Routing Table of Leaf-A

Therefore, the solution proposed in this document requires that an IS-IS router SHOULD check a level-1 LSP's area addresses before advertising it to a neighbor (see Section 2.2).

2.2. Area Address Checking

Before advertising a level-1 LSP to a neighbor, an IS-IS router SHOULD compare the area addresses associated with the LSP and the ones associated with the neighbor. If they have at least one area

address in common, the router SHOULD advertise the LSP to the neighbor. Otherwise, the router MUST NOT advertise the LSP to the neighbor.

In the former case, the router SHOULD remove every area addresse in the LSP except the ones associated with the neighbor before the advertisement. This makes the solution more compatible since the Leaf nodes can be unaltered (see Section 3.2).

For instance, before Spine-A advertises LSP 1111.1111.1111.00-00 to Leaf-A, it compares the LSP's area addresses (i.e., 10 and 20) with Leaf-A's area address (i.e., 10). Since they have a common area address 10, Spine-A SHOULD remove area address 20 from the LSP and advertise the LSP to Leaf-A. On the other hand, before Spine-A advertises LSP 4444.4444.4444.00-00 to Leaf-A, it checks their area addresses and finds that they have no area address in common. So Spine-A MUST NOT advertise the LSP to Leaf-A. As a result, Leaf-A would not learn any routing information of Leaf-B, as shown in Figure 4.

Leaf-A Routing Table:

Destination	Proto	Pre	Cost	Flags	NextHop	Interface
10.10.10.0/24	ISIS-L1	15	20	D	10.10.20.1	Ethernet0/0/0
	ISIS-L1	15	20	D	10.10.40.2	Ethernet0/0/1
10.10.20.0/24	Direct	0	0	D	127.0.0.1	Ethernet0/0/0
10.10.30.0/24	ISIS-L1	15	20	D	10.10.40.2	Ethernet0/0/1
10.10.40.0/24	Direct	0	0	D	127.0.0.1	Ethernet0/0/1
10.10.50.0/24	ISIS-L1	15	20	D	10.10.20.1	Ethernet0/0/0
192.168.10.0/24	Direct	0	0	D	127.0.0.1	GEthernet0/0/0
127.0.0.0/8	Direct	0	0	D	127.0.0.1	InLoopBack0
0.0.0.0/0	ISIS-L1	15	10	D	10.10.20.1	Ethernet0/0/0
	ISIS-L1	15	10	D	10.10.40.2	Ethernet0/0/1

Figure 4: Routing Table of Leaf-A

2.3. Default Route Advertising

As defined in [RFC 1195], a L1/L2 router will indicate in its LSPs that it is "attached" by setting the ATT bits. Therefore, each Leaf node would set the Spine nodes as its default gateways and install a default route in its routing table, as shown in Figure 4.

However, a specific IS-IS implementation in this case may not let the L1/L2 router set the ATT bits, because it may speculate that the L1/L2 router has lost connectivity to the level-2 backbone. To solve this problem, operators can manually configure the L1/L2 router to advertise a default route.

3. Compatibility

3.1. Overlapping Areas Use Case

In most deployments, an IS-IS router is assigned only one area address, which will not be influenced by the area checking mechanism proposed in this document. However, an IS-IS router might be assigned more than one area addresses in some practical deployments for the following reasons: 1) it is desirable to change the area address of an area, 2) to merge two areas into one area, or 3) to partition an area into two areas.

For instance, to change an area's address from X to Y, one can simply add area address Y to all routers in the area, and then remove X from them. Note that such operations would not disrupt live traffic in the network.

Although the solution in this document requires IS-IS router to check LSP's area addresses before advertising it, the above use cases are still applicable and no compatible issue rises.

3.2. Maximum Area Addresses

The maximumAreaAddresses parameter in today's IS-IS implementation is set to be 3 (or 0 which indicates 3) on consensus. Therefore, the solution in this document also requires that Spine node SHOULD be modified for supporting more area addresses. However, as LSPs sent to a given neighbor only carry the area address(es) of the neighbor (see Section 2.2), the solution does not require to modify Leaf nodes.

4. IANA Considerations

TBD.

5. Security Considerations

TBD.

6. Acknowledgements

TBD.

7. Normative References

[IS-IS-SL-Extension]
Shen, N. and S. Thyamagundalu, "IS-IS Routing for Spine-Leaf Topology", draft-shen-isis-spine-leaf-ext-03 (work in progress) , March 2017.

[RFC1195] Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments", RFC 1195 , December 1990.

[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>.

Authors' Addresses

Zhe Chen
Huawei Technologies
No. 156 Beiqing Rd
Beijing 100095
China

Email: chenzhe17@huawei.com

Xiaohu Xu
Huawei Technologies
No. 156 Beiqing Rd
Beijing 100095
China

Email: xuxiaohu@huawei.com