

Networking Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: September 9, 2019

N. Shen  
L. Ginsberg  
Cisco Systems  
S. Thyamagundalu  
March 8, 2019

IS-IS Routing for Spine-Leaf Topology  
draft-ietf-lsr-isis-spine-leaf-ext-01

## Abstract

This document describes a mechanism for routers and switches in a Spine-Leaf type topology to have non-reciprocal Intermediate System to Intermediate System (IS-IS) routing relationships between the leafs and spines. The leaf nodes do not need to have the topology information of other nodes and exact prefixes in the network. This extension also has application in the Internet of Things (IoT).

## 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 September 9, 2019.

## Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in [Section 4.e](#) of

Internet-Draft

IS-IS SL Extension

March 2019

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

<a href="#">1.</a>	<a href="#">Introduction</a>	<a href="#">2</a>
<a href="#">1.1.</a>	<a href="#">Requirements Language</a>	<a href="#">3</a>
<a href="#">2.</a>	<a href="#">Motivations</a>	<a href="#">3</a>
<a href="#">3.</a>	<a href="#">Spine-Leaf (SL) Extension</a>	<a href="#">4</a>
<a href="#">3.1.</a>	<a href="#">Topology Examples</a>	<a href="#">4</a>
<a href="#">3.2.</a>	<a href="#">Applicability Statement</a>	<a href="#">5</a>
<a href="#">3.3.</a>	<a href="#">Spine-Leaf TLVs</a>	<a href="#">6</a>
<a href="#">3.3.1.</a>	<a href="#">Spine-Leaf TLV</a>	<a href="#">6</a>
<a href="#">3.3.2.</a>	<a href="#">Leaf-Set TLV</a>	<a href="#">7</a>
<a href="#">3.3.2.1.</a>	<a href="#">Leaf-Set Sub-TLVs</a>	<a href="#">7</a>
<a href="#">3.3.3.</a>	<a href="#">Advertising IPv4/IPv6 Reachability</a>	<a href="#">8</a>
<a href="#">3.3.4.</a>	<a href="#">Advertising Connection to RF-Leaf Node</a>	<a href="#">8</a>
<a href="#">3.4.</a>	<a href="#">Mechanism</a>	<a href="#">9</a>
<a href="#">3.4.1.</a>	<a href="#">Pure CLOS Topology</a>	<a href="#">10</a>
<a href="#">3.5.</a>	<a href="#">Implementation and Operation</a>	<a href="#">11</a>
<a href="#">3.5.1.</a>	<a href="#">CSNP PDU</a>	<a href="#">11</a>
<a href="#">3.5.2.</a>	<a href="#">Leaf to Leaf connection</a>	<a href="#">12</a>
<a href="#">3.5.2.1.</a>	<a href="#">Local traffic only</a>	<a href="#">12</a>
<a href="#">3.5.2.2.</a>	<a href="#">Transit traffic allowed</a>	<a href="#">12</a>
<a href="#">3.5.3.</a>	<a href="#">Spine Node Hostname</a>	<a href="#">13</a>
<a href="#">3.5.4.</a>	<a href="#">IS-IS Reverse Metric</a>	<a href="#">13</a>
<a href="#">3.5.5.</a>	<a href="#">Spine-Leaf Traffic Engineering</a>	<a href="#">13</a>
<a href="#">3.5.6.</a>	<a href="#">Other End-to-End Services</a>	<a href="#">13</a>
<a href="#">3.5.7.</a>	<a href="#">Address Family and Topology</a>	<a href="#">14</a>
<a href="#">3.5.8.</a>	<a href="#">Migration</a>	<a href="#">14</a>
<a href="#">4.</a>	<a href="#">IANA Considerations</a>	<a href="#">14</a>
<a href="#">5.</a>	<a href="#">Security Considerations</a>	<a href="#">15</a>
<a href="#">6.</a>	<a href="#">Acknowledgments</a>	<a href="#">15</a>
<a href="#">7.</a>	<a href="#">References</a>	<a href="#">15</a>
<a href="#">7.1.</a>	<a href="#">Normative References</a>	<a href="#">15</a>
<a href="#">7.2.</a>	<a href="#">Informative References</a>	<a href="#">17</a>
	<a href="#">Authors' Addresses</a>	<a href="#">17</a>

## [1.](#) Introduction

The IS-IS routing protocol defined by [[ISO10589](#)] has been widely deployed in provider networks, data centers and enterprise campus environments. In the data center and enterprise switching networks,

a Spine-Leaf topology is commonly used. This document describes a mechanism where IS-IS routing can be optimized for a Spine-Leaf topology.

In a Spine-Leaf topology, normally a leaf node connects to a number of spine nodes. Data traffic going from one leaf node to another leaf node needs to pass through one of the spine nodes. Also, the decision to choose one of the spine nodes is usually part of equal cost multi-path (ECMP) load sharing. The spine nodes can be considered as gateway devices to reach destinations on other leaf nodes. In this type of topology, the spine nodes have to know the topology and routing information of the entire network, but the leaf nodes only need to know how to reach the gateway devices to which are the spine nodes they are uplinked.

This document describes the IS-IS Spine-Leaf extension that allows the spine nodes to have all the topology and routing information, while keeping the leaf nodes free of topology information other than the default gateway routing information. The leaf nodes do not even need to run a Shortest Path First (SPF) calculation since they have no topology information.

### [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 [RFC 2119](#) [[RFC2119](#)].

## [2](#). Motivations

- o The leaf nodes in a Spine-Leaf topology do not require complete topology and routing information of the entire domain since their forwarding decision is to use ECMP with spine nodes as default gateways
- o The spine nodes in a Spine-Leaf topology are richly connected to leaf nodes, which introduces significant flooding duplication if they flood all Link State PDUs (LSPs) to all the leaf nodes. It saves both spine and leaf nodes' CPU and link bandwidth resources if flooding is blocked to leaf nodes. For small Top of the Rack



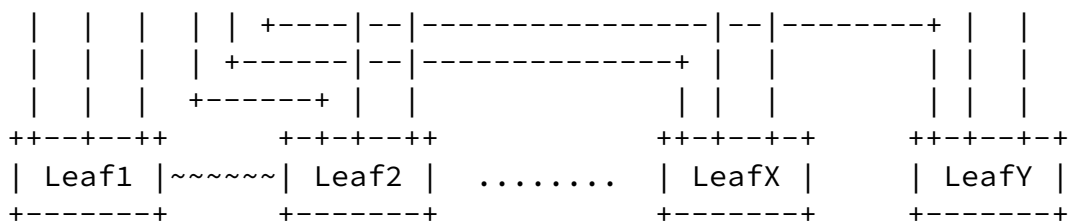


Figure 1: A Spine-Leaf Topology

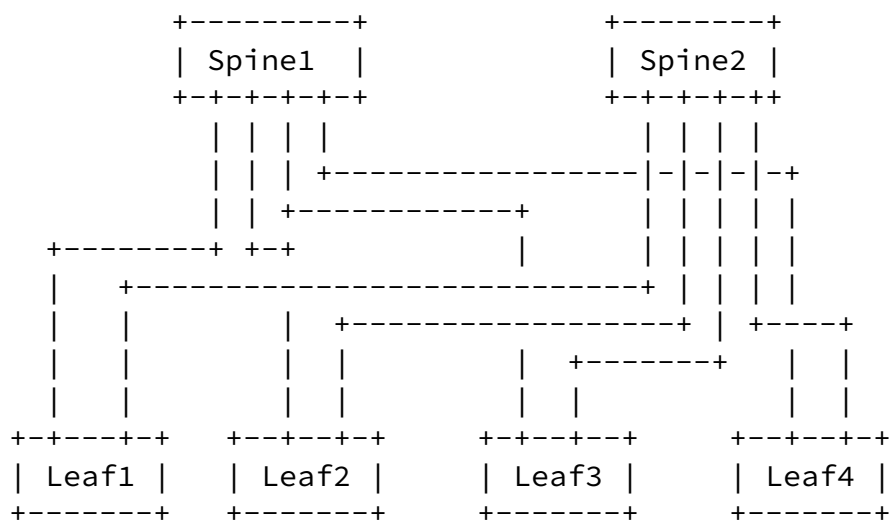


Figure 2: A CLOS Topology

### [3.2.](#) Applicability Statement

This extension assumes the network is a Spine-Leaf topology, and it should not be applied in an arbitrary network setup. The spine nodes can be viewed as the aggregation layer of the network, and the leaf

nodes as the access layer of the network. The leaf nodes use a load sharing algorithm with spine nodes as nexthops in routing and forwarding.

This extension works when the spine nodes are inter-connected, and it works with a pure CLOS or Fat Tree topology based network where the spines are NOT horizontally interconnected.

Although the example diagram in Figure 1 shows a fully meshed Spine-Leaf topology, this extension also works in the case where they are partially meshed. For instance, leaf1 through leaf10 may be fully meshed with spine1 through spine5 while leaf11 through leaf20 is fully meshed with spine4 through spine8, and all the spines are inter-connected in a redundant fashion.

This extension can also work in multi-level spine-leaf topology. The lower level spine node can be a 'leaf' node to the upper level spine node. A spine-leaf 'Tier' can be exchanged with IS-IS hello packets to allow tier X to be connected with tier X+1 using this extension. Normally tier-0 will be the TOR routers and switches if provisioned.

This extension also works with normal IS-IS routing in a topology with more than two layers of spine and leaf. For instance, in example diagrams Figure 1 and Figure 2, there can be another Core layer of routers/switches on top of the aggregation layer. From an IS-IS routing point of view, the Core nodes are not affected by this

extension and will have the complete topology and routing information just like the spine nodes. To make the network even more scalable, the Core layer can operate as a level-2 IS-IS sub-domain while the Spine and Leaf layers operate as stays at the level-1 IS-IS domain.

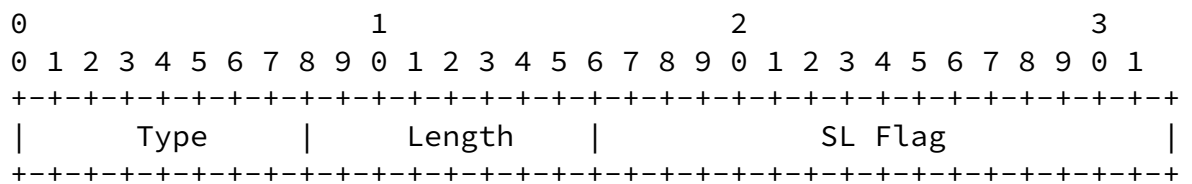
This extension assumes the link between the spine and leaf nodes are point-to-point, or point-to-point over LAN [[RFC5309](#)]. The links connecting among the spine nodes or the links between the leaf nodes can be any type.

### [3.3.](#) Spine-Leaf TLVs

This extension introduces two new TLVs, the Spine-Leaf TLV and the Leaf-Set TLV. The Spine-Leaf TLV may be advertised in IS-IS Hello (IIH) PDUs; the Leaf-Set TLV may be advertised in IS-IS Circuit

Scoped Link State PDUs (CS-LSP) [RFC7356]. They are used by both spine and leaf nodes in this Spine-Leaf mechanism.

### 3.3.1. Spine-Leaf TLV

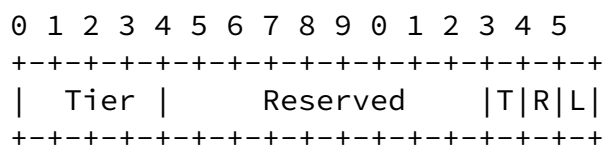


The fields of this TLV are defined as follows:

Type: 1 octet Suggested value 151 (to be assigned by IANA)

Length: 1 octet (2 + length of sub-TLVs).

SL Flags: 16 bits



Tier: A value from 0 to 15. It represents the spine-leaf tier level. The value 15 is reserved to indicate the tier level is unknown. This value is only valid when the 'T' bit (see below) is set. If the 'T' bit is

clear, this value MUST be set to zero on transmission, and it MUST be ignored on receipt.

L bit (0x01): Only leaf node sets this bit. If the L bit is set in the SL flag, the node indicates it is in 'Leaf-Mode'.

R bit (0x02): Only Spine node sets this bit. If the R bit is

set, the node indicates to the leaf neighbor that it can be used as the default route gateway.

T bit (0x04): If set, the value in the "Tier" field (see above) is valid.

### 3.3.2. Leaf-Set TLV

```

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      |      .. Optional Sub-TLVs
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+....

```

The Type is suggested value of 152 (to be assigned by IANA). This TLV and associated Sub-TLVs MAY appear in CS-LSP PDUs. Multiple TLVs MAY be sent.

#### 3.3.2.1. Leaf-Set Sub-TLVs

If the data center topology is a pure CLOS or Fat Tree, there are no link connections among the spine nodes. If we also assume there is not another Core layer on top of the aggregation layer, then the traffic from one leaf node to another may have a problem if there is a link outage between a spine node and a leaf node. For instance, in the diagram of Figure 2, if Leaf1 sends data traffic to Leaf3 through Spine1 node, and the Spine1-Leaf3 link is down, the data traffic will be dropped on the Spine1 node.

To address this issue spine and leaf nodes may use the sub-TLVs defined below to obtain more specific reachability information.

Two Leaf-Set sub-TLVs are defined. The Leaf-Neighbors sub-TLV and the Reachability-Req sub-TLV.

##### 3.3.2.1.1. Leaf-Neighbors Sub-TLV

This sub-TLV is used by spine nodes to advertise the current set of Leaf neighbors to Leaf nodes. The fields of this sub-TLV are defined as follows:



Length: 1 octet MUST be a multiple of 6 octets.

Leaf-Neighbors A list of IS-IS System-IDs of the leaf node neighbors of this spine node.

#### [3.3.2.1.2](#). Reachability-Req Sub-TLV

This sub-TLV is used by leaf nodes to request the advertisement of more specific prefix information from one or more selected spine node(s). The list of leaf nodes in this sub-TLV reflects the current set of leaf-nodes for which not all spine node neighbors have indicated the presence of connectivity in the Leaf-Neighbors sub-TLV (See [Section 3.3.2.1.1](#)). The fields of this sub-TLV are defined as follows:

Type: 1 octet Suggested value 2 (to be assigned by IANA)

Length: 1 octet. It MUST be a multiple of 6 octets.

Leaf Nodes List of IS-IS System-IDs of leaf nodes for which reachability information is being requested.

#### [3.3.3](#). Advertising IPv4/IPv6 Reachability

In cases where connectivity between a leaf node and a spine node is down, the leaf node MAY request reachability information from a spine node as described in [Section 3.3.2.1.2](#). The spine node utilizes TLVs 135 [[RFC5305](#)] and TLVs 236 [[RFC5308](#)] to advertise this information. These TLVs MAY be included in CS-LSPs [[RFC7356](#)] sent from the spine to the requesting leaf node.

#### [3.3.4](#). Advertising Connection to RF-Leaf Node

For links between Spine and Leaf Nodes on which the Spine Node has set the R-bit and the Leaf node has set the L-bit in their respective Spine-Leaf TLVs, spine nodes MAY advertise the link with a bit in the "link-attribute" sub-TLV [[RFC5029](#)] to indicate that this link is not used for LSP flooding. This bit is named the Connect-to-RF-Leaf Node bit. This information can be used by nodes computing a flooding topology e.g., [[DYNAMIC-FLOODING](#)], to exclude the RF-Leaf nodes from the computed flooding topology.

For links between Spine and Leaf Nodes on which the Spine Node has set the R-bit and the Leaf node has set the L-bit in their respective

Spine-Leaf TLVs, leaf nodes MAY advertise the link with a bit in the "link-attribute" sub-TLV [[RFC5029](#)] to indicate that this link is to a Spine Node neighbor. This bit is named the Connect-to-RF-Spine Node bit. This information can be used by leaf nodes when deciding whether a leaf to leaf link can be used as an alternate default path when a leaf node has no connectivity to any spines. See [Section 3.5.2](#).

### [3.4](#). Mechanism

Leaf nodes in a spine-leaf application using this extension are provisioned with two attributes:

- 1) Tier level of 0. This indicates the node is a Leaf Node. The value 0 is advertised in the Tier field of Spine-Leaf TLV defined above.
- 2) Flooding reduction enabled/disabled. If flooding reduction is enabled the L-bit is set to one in the Spine-Leaf TLV defined above

A spine node does not need explicit configuration. Spine nodes can dynamically discover their tier level by computing the number of hops to a leaf node. Until a spine node determines its tier level it MUST advertise level 15 (unknown tier level) in the Spine-Leaf TLV defined above. Each tier level can also be statically provisioned on the node.

When a spine node receives an IIH which includes the Spine-Leaf TLV with Tier level 0 and 'L' bit set, it labels the point-to-point interface and adjacency to be a 'Reduced Flooding Leaf-Peer (RF-Leaf)'. IIHs sent by a spine node on a link to an RF-Leaf include the Spine-Leaf TLV with the 'R' bit set in the flags field. The 'R' bit indicates to the RF-Leaf neighbor that the spine node can be used as a default routing nexthop.

There is no change to the IS-IS adjacency bring-up mechanism for Spine-Leaf peers.

A spine node blocks LSP flooding to RF-Leaf adjacencies, except for the LSP PDUs in which the IS-IS System-ID matches the System-ID of the RF-Leaf neighbor. This exception is needed since when the leaf node reboots, the spine node needs to forward to the leaf node non-purged LSPs from the RF-Leaf's previous incarnation.

Leaf nodes will perform IS-IS LSP flooding as normal to send the LSPs over all of its IS-IS adjacencies. In the case of RF-Leafs only

self-originated LSPs will exist in its LSP database, and in the case

of leaf-leaf connections, there will be neighbor leaf nodes LSPs in the LSP database in addition to the self-originated LSPs.

Spine nodes will receive all the LSP PDUs in the network, including all the spine nodes and leaf nodes. It will perform Shortest Path First (SPF) as a normal IS-IS node does. There is no change to the route calculation and forwarding on the spine nodes.

The LSPs of a node only floods north bound towards the upper layer spine nodes. The default route is generated with loadsharing also towards the upper layer spine nodes.

RF-Leaf nodes do not have any LSP in the network except for its own. Therefore there is no need to perform SPF calculation on the RF-Leaf node. It only needs to download the default route with the nexthops of those Spine Neighbors which have the 'R' bit set in the Spine-Leaf TLV in IIH PDUs. IS-IS can perform equal cost or unequal cost load sharing while using the spine nodes as nexthops. The aggregated metric of the outbound interface and the 'Reverse Metric' [[RFC8500](#)] can be used for this purpose.

#### [3.4.1.](#) Pure CLOS Topology

In a data center where the topology is pure CLOS or Fat Tree, there is no interconnection among the spine nodes, and there is not another Core layer above the aggregation layer with reachability to the leaf nodes. When flooding reduction to RF-Leafs is in use, if the link between a spine and a leaf goes down, there is then a possibility of black holing the data traffic in the network.

As in the diagram Figure 2, if the link Spine1-Leaf3 goes down, there needs to be a way for Leaf1, Leaf2 and Leaf4 to avoid the Spine1 if the destination of data traffic is to Leaf3 node.

In the above example, the Spine1 and Spine2 are provisioned to advertise the Leaf-Set sub-TLV of the Spine-Leaf TLV. Originally both Spines will advertise Leaf1 through Leaf4 as their Leaf-Set. When the Spine1-Leaf3 link is down, Spine1 will only have Leaf1, Leaf2 and Leaf4 in its Leaf-Set. This allows the other leaf nodes to

know that Spine1 has lost connectivity to the leaf node of Leaf3.

Each RF-Leaf node can select another spine node to request for some prefix information associated with the lost leaf node. In this diagram of Figure 2, there are only two spine nodes (Spine-Leaf topology can have more than two spine nodes in general). Each RF-Leaf node can independently select a spine node for the leaf information. The RF-Leaf nodes will include the Info-Req sub-TLV in

the Spine-Leaf TLV in hellos sent to the selected spine node, Spine2 in this case.

The spine node, upon receiving the request from one or more leaf nodes, will find the IPv6/IPv4 prefixes advertised by the leaf nodes listed in the Info-Req sub-TLV. The spine node will use the mechanism defined in [Section 3.3.2](#) to advertise these prefixes to the RF-Leaf node. For instance, it will include the IPv4 loopback prefix of leaf3 based on the policy configured or administrative tag attached to the prefixes. When the leaf nodes receive the more specific prefixes, they will install the advertised prefixes towards the other spine nodes (Spine2 in this example).

For instance in the data center overlay scenario, when any IP destination or MAC destination uses the leaf3's loopback as the tunnel nexthop, the overlay tunnel from leaf nodes will only select Spine2 as the gateway to reach leaf3 as long as the Spine1-Leaf3 link is still down.

In cases where multiple links or nodes fail at the same time, the RF-leaf node may need to send the Info-Req to multiple upper layer spine nodes in order to obtain reachability information for all the partially connected nodes.

This negative routing is more useful between tier 0 and tier 1 spine-leaf levels in a multi-level spine-leaf topology when the reduced flooding extension is in use. Nodes in tiers 1 or greater may have much richer topology information and alternative paths.

## [3.5.](#) Implementation and Operation

### [3.5.1.](#) CSNP PDU

In Spine-Leaf extension, Complete Sequence Number PDUs (CSNP) do not need to be transmitted over the Spine-Leaf link to an RF-Leaf. Some IS-IS implementations send periodic CSNPs after the initial adjacency bring-up over a point-to-point interface. There is no need for this optimization here since the RF-Leaf does not need to receive any other LSPs from the network, and the only LSPs transmitted across the Spine-Leaf link are the leaf node LSPs.

Also in the graceful restart case[RFC5306], for the same reason, there is no need to send the CSNPs over the Spine-Leaf interface to an RF-Leaf. Spine nodes only need to set the SRMflag on the LSPs belonging to the RF-Leaf that has restarted.

### [3.5.2.](#) Leaf to Leaf connection

Leaf to leaf node links are useful in host redundancy cases in switching networks. There are no flooding extensions required in this case. Leaf node LSPs will be exchanged over this link using the normal operation of the IS-IS Update process. In the example diagram Figure 1, Leaf1 will receive Leaf2's LSPs and Leaf2 will receive Leaf1's LSPs. Each of the Leaf nodes will in turn flood the LSPs they receive from their leaf node neighbor to their spine neighbors. Prefix reachability advertisements received from the leaf neighbor will result in the installation of more specific routes using this local Leaf-Leaf link. SPF will be performed in this case just like when the entire network only involves with those two IS-IS nodes. This does not affect the normal Spine-Leaf mechanism they perform toward the spine nodes.

Leaf to leaf connections SHOULD be limited to a single leaf neighbor.

Two modes of operation for the Leaf-Leaf link are possible and are described in the following sub-sections.

#### [3.5.2.1.](#) Local traffic only

The leaf node sets the 'overload' bit in its LSP PDU so that spine nodes will not send traffic destined for the neighboring leaf node

via its leaf node neighbor. The Leaf-Leaf link will then be used solely for local traffic between the two Leaf Nodes.

#### [3.5.2.2.](#) Transit traffic allowed

If a leaf node becomes disconnected from all spine nodes, it is possible for spine nodes to route traffic destined for the disconnected leaf node via its leaf node neighbor. However the leaf to leaf link SHOULD be the link of last resort. To support this mode the leaf nodes do NOT set the overload bit in their LSPs and they advertise a high metric for the leaf to leaf link( $(2^{24} - 2)$  is recommended). This signals to the Spine Nodes that the leaf to leaf link may be used for transit traffic, but also insures that it will not be used unless the spine node has no other path to a given leaf node.

When the leaf node is disconnected from all spine nodes it MAY install a default route towards its leaf-node neighbor in support of return traffic to the spine nodes. When doing so the leaf should validate that its leaf neighbor has at least one spine neighbor. This can be done by looking for the Connect-to-RF-Spine Node bit in the Link Attributes sub-TLVs [[RFC5029](#)] advertised in the LSPs of its leaf node neighbor.

#### [3.5.3.](#) Spine Node Hostname

This extension creates a non-reciprocal relationship between the spine node and leaf node. The spine node will receive leaf's LSP and will know the leaf's hostname, but the leaf does not have spine's LSP. This extension allows the Dynamic Hostname TLV [[RFC5301](#)] to be optionally included in spine's IIH PDU when sending to a 'Leaf-Peer'. This is useful in troubleshooting cases.

#### [3.5.4.](#) IS-IS Reverse Metric

This metric is part of the aggregated metric for leaf's default route installation with load sharing among the spine nodes. When a spine node is in 'overload' condition, it should use the IS-IS Reverse Metric TLV in IIH [[RFC8500](#)] to set this metric to maximum to discourage the leaf using it as part of the loadsharing.

In some cases, certain spine nodes may have less bandwidth in link

provisioning or in real-time condition, and it can use this metric to signal to the leaf nodes dynamically.

In other cases, such as when the spine node loses a link to a particular leaf node, although it can redirect the traffic to other spine nodes to reach that destination leaf node, but it MAY want to increase this metric value if the inter-spine connection becomes over utilized, or the latency becomes an issue.

#### [3.5.5.](#) Spine-Leaf Traffic Engineering

Besides using the IS-IS Reverse Metric by the spine nodes to affect the traffic pattern for leaf default gateway towards multiple spine nodes, the IPv6/IPv4 Info-Advertise sub-TLVs can be selectively used by traffic engineering controllers to move data traffic around the data center fabric to alleviate congestion and to reduce the latency of a certain class of traffic pairs. By injecting more specific leaf node prefixes, it will allow the spine nodes to attract more traffic on some underutilized links.

#### [3.5.6.](#) Other End-to-End Services

Losing the topology information will have an impact on some of the end-to-end network services, for instance, MPLS TE or end-to-end segment routing. Some other mechanisms such as those described in PCE [[RFC4655](#)] based solution may be used. In this Spine-Leaf extension, the role of the leaf node is not too much different from the multi-level IS-IS routing while the level-1 IS-IS nodes only have the default route information towards the node which has the Attach Bit (ATT) set, and the level-2 backbone does not have any topology

information of the level-1 areas. The exact mechanism to enable certain end-to-end network services in Spine-Leaf network is outside the scope of this document.

#### [3.5.7.](#) Address Family and Topology

IPv6 Address families[RFC5308], Multi-Topology (MT)[[RFC5120](#)] and Multi-Instance (MI)[[RFC8202](#)] information is carried over the IIH PDU. Since the goal is to simplify the operation of IS-IS network, for the simplicity of this extension, the Spine-Leaf mechanism is applied the same way to all the address families, MTs and MIs.

### 3.5.8. Migration

For this extension to be deployed in existing networks, a simple migration scheme is needed. To support any leaf node in the network, all the involved spine nodes have to be upgraded first. So the first step is to migrate all the involved spine nodes to support this extension, then the leaf nodes can be enabled with 'Leaf-Mode' one by one. No flag day is needed for the extension migration.

## 4. IANA Considerations

Two new TLV codepoint is defined in this document and needs to be assigned by IANA from the "IS-IS TLV Codepoints" registry. They are referred to as the Spine-Leaf TLV and the suggested value is 151, and Leaf-Set TLV and suggested value is 152. The Spine-Leaf TLV is only to be optionally inserted in the IIH PDU, and the Leaf-Set TLV is only to be optionally inserted in Circuit Flooding Scoped LSP PDU. IANA is also requested to maintain the SL-flag bit values in the Spine-Leaf TLV, and 0x01, 0x02 and 0x04 bits are defined in this document.

Value	Name	IIH	LSP	SNP	Purge	CS-LSP
151	Spine-Leaf	y	n	n	n	n
152	Leaf-Set	n	n	n	n	y

This document also proposes to have the Dynamic Hostname TLV, already assigned as code 137, to be allowed in IIH PDU.

Value	Name	IIH	LSP	SNP	Purge
137	Dynamic Name	y	y	n	y

This documents requests IANA to create a new registry under the IS-IS TLV Codepoints registry. The suggested name of the registry is "Sub-

TLVs for TLV 152 (Leaf-Set TLV)". Initial contents of the new registry is defined below:

Value	Name
-------	------



-----	-----
0	Reserved
1	Leaf Neighbors
2	Reachability Req
3-255	Unassigned

This document also requests that IANA allocate from the registry of link-attribute two new bit values for sub-TLV 19 of TLV 22 (Extended IS reachability TLV).

Value	Name	Reference
-----	-----	-----
0x4	Connect to RF-Leaf Node	This document
0x8	Connect to RF-Spine Node	This document

## 5. Security Considerations

Security concerns for IS-IS are addressed in [[ISO10589](#)], [[RFC5304](#)], [[RFC5310](#)], and [[RFC7602](#)]. This extension does not raise additional security issues.

## 6. Acknowledgments

The authors would like to thank Tony Przygienda and Lukas Krattiger for their discussion and contributions. The authors also would like to thank Acee Lindem, Russ White, Christian Hopps and Aijun Wang for their review and comments of this document.

## 7. References

### 7.1. Normative References

- [ISO10589] ISO "International Organization for Standardization", "Intermediate system to Intermediate system intra-domain routing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473), ISO/IEC 10589:2002, Second Edition.", Nov 2002.
- [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>>.

- [RFC5029] Vasseur, JP. and S. Previdi, "Definition of an IS-IS Link Attribute Sub-TLV", [RFC 5029](#), DOI 10.17487/RFC5029, September 2007, <<https://www.rfc-editor.org/info/rfc5029>>.
- [RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", [RFC 5120](#), DOI 10.17487/RFC5120, February 2008, <<https://www.rfc-editor.org/info/rfc5120>>.
- [RFC5301] McPherson, D. and N. Shen, "Dynamic Hostname Exchange Mechanism for IS-IS", [RFC 5301](#), DOI 10.17487/RFC5301, October 2008, <<https://www.rfc-editor.org/info/rfc5301>>.
- [RFC5304] Li, T. and R. Atkinson, "IS-IS Cryptographic Authentication", [RFC 5304](#), DOI 10.17487/RFC5304, October 2008, <<https://www.rfc-editor.org/info/rfc5304>>.
- [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>>.
- [RFC5306] Shand, M. and L. Ginsberg, "Restart Signaling for IS-IS", [RFC 5306](#), DOI 10.17487/RFC5306, October 2008, <<https://www.rfc-editor.org/info/rfc5306>>.
- [RFC5308] Hopps, C., "Routing IPv6 with IS-IS", [RFC 5308](#), DOI 10.17487/RFC5308, October 2008, <<https://www.rfc-editor.org/info/rfc5308>>.
- [RFC5310] Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", [RFC 5310](#), DOI 10.17487/RFC5310, February 2009, <<https://www.rfc-editor.org/info/rfc5310>>.
- [RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding Scope Link State PDUs (LSPs)", [RFC 7356](#), DOI 10.17487/RFC7356, September 2014, <<https://www.rfc-editor.org/info/rfc7356>>.
- [RFC7602] Chunduri, U., Lu, W., Tian, A., and N. Shen, "IS-IS Extended Sequence Number TLV", [RFC 7602](#), DOI 10.17487/RFC7602, July 2015, <<https://www.rfc-editor.org/info/rfc7602>>.
- [RFC8202] Ginsberg, L., Previdi, S., and W. Henderickx, "IS-IS Multi-Instance", [RFC 8202](#), DOI 10.17487/RFC8202, June

2017, <<https://www.rfc-editor.org/info/rfc8202>>.

Shen, et al.

Expires September 9, 2019

[Page 16]

---

Internet-Draft

IS-IS SL Extension

March 2019

- [RFC8500] Shen, N., Amante, S., and M. Abrahamsson, "IS-IS Routing with Reverse Metric", [RFC 8500](#), DOI 10.17487/RFC8500, February 2019, <<https://www.rfc-editor.org/info/rfc8500>>.

## [7.2.](#) Informative References

### [DYNAMIC-FLOODING]

Li, T., "Dynamic Flooding on Dense Graphs", [draft-li-dynamic-flooding](#) (work in progress), 2018.

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", [RFC 4655](#), DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.

- [RFC5309] Shen, N., Ed. and A. Zinin, Ed., "Point-to-Point Operation over LAN in Link State Routing Protocols", [RFC 5309](#), DOI 10.17487/RFC5309, October 2008, <<https://www.rfc-editor.org/info/rfc5309>>.

## Authors' Addresses

Naiming Shen  
Cisco Systems  
560 McCarthy Blvd.  
Milpitas, CA 95035  
US

Email: [naiming@cisco.com](mailto:naiming@cisco.com)

Les Ginsberg  
Cisco Systems  
821 Alder Drive  
Milpitas, CA 95035  
US

Email: [ginsberg@cisco.com](mailto:ginsberg@cisco.com)

Sanjay Thyamagundalu

Email: [tsanjay@gmail.com](mailto:tsanjay@gmail.com)

Shen, et al.

Expires September 9, 2019

[Page 17]