Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: February 19, 2021

H. Chen
R. Li
Futurewei
Y. Yang
IBM
A. Kumar S N
RtBrick
Y. Fan
Casa Systems
N. So

V. Liu

M. Toy
Verizon
L. Liu
Fujitsu
K. Makhijani
Futurewei

August 18, 2020

# IS-IS Topology-Transparent Zone
draft-chen-isis-ttz-12.txt

## Abstract

This document presents a topology-transparent zone in an area. A zone is a block/piece of an area, which comprises a group of routers and a number of circuits connecting them. It is abstracted as a virtual entity such as a single virtual node or zone edges mesh. Any router outside of the zone is not aware of the zone. The information about the circuits and routers inside the zone is not distributed to any router outside of the zone.

## Status of This Memo

This Internet-Draft is submitted to IETF 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 February 19, 2021.



Table of Contents

## 1. Introduction

[ISO10589] describes two levels of areas, which are level 1 and level 2 areas in IS-IS. There are scalability issues in using areas as the number of routers in a network becomes larger and larger.

Through splitting the network into multiple areas, we may extend the network further. However, dividing a network from one area into multiple areas or from a number of existing areas to even more areas is a very challenging and time consuming task since it is involved in significant network architecture changes.

These issues can be resolved by using topology-transparent zone (TTZ), which abstracts a zone (i.e., a block/piece of an area) as a single virtual node or zone edges' mesh with minimum efforts and minimum service interruption. Note that a zone can be an area (i.e., the entire piece of an area).

This document presents a topology-transparent zone and describes extensions to IS-IS for supporting the topology-transparent zone.

### 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].

### 1.2. Terminology

LSP: A Link State Protocol Data Unit (PDU) in IS-IS.

LS: A Link Sate, which is short for LSP in IS-IS.

TTZ: A Topology-Transparent Zone.

Zone: A block or piece of an area. In a special case, a zone is an area (i.e., the entire piece of an area).

Zone External Node: A node outside of a zone.

Zone Internal Node: A node of a zone without any connection to a node outside of the zone.

Zone Edge/Border: A node of a zone connecting to a node outside of the zone.

Zone Node: A zone internal node or a zone edge/border node (i.e., a node of a zone).

Zone Link: A link connecting zone nodes (i.e., a link of a zone).

Zone Neighbor: A node outside of a zone that is a neighbor of a zone edge/border.

## 2. Requirements

Topology-Transparent Zone (TTZ) may be deployed for resolving some critical issues such as scalability in existing networks and future networks. The requirements for TTZ are listed as follows:

- TTZ MUST be backward compatible. When a TTZ is deployed on a set of routers in a network, the routers outside of the TTZ in the network do not need to know or support TTZ.
- TTZ MUST support at least one more levels of network hierarchies, in addition to the hierarchies supported by existing routing protocols.
- Abstracting a zone as a virtual entity, which is a single virtual node or zone edges' mesh, SHOULD be smooth with minimum service interruption.
- De-abstracting (or say rolling back) a virtual entity to a zone SHOULD be smooth with minimum service interruption.
- Users SHOULD be able to easily set up an end to end service crossing TTZs.
- The configuration for a TTZ in a network SHOULD be minimum.
- The changes on the existing protocols for supporting TTZ SHOULD be minimum.

## 3. Zone Abstraction

A zone can be abstracted as a single virtual node or the zone edges' full mesh.

When a zone is abstracted as a single virtual node, this single node is connected to all the neighbors of the zone, and is in the same area as the neighbors.

When a zone is abstracted as its edges' full mesh, there is a full mesh connections among the edges and each edge is also connected to its neighbors outside of the zone.

# 4. Topology-Transparent Zone

A Topology-Transparent Zone (TTZ) comprises an Identifier (ID) and a piece/block of an area such as a Level 2 area in IS-IS. It is abstracted as a single virtual node or its edges' full mesh. TTZ and zone will be used exchangably below.

## 4.1. Zone as a Single Node

After a zone is abstracted as a single virtual node having a virtual node ID, every node outside of the zone sees a number of links connected to this single node. Each of these links connects a zone neighbor. The link states inside the zone are not advertised to any node outside of the zone. The virtual node ID may be derived from the zone ID.

### 4.1.1. An Example of Zone as a Single Node

The figure below shows an example of an area containing a TTZ: TTZ 600.

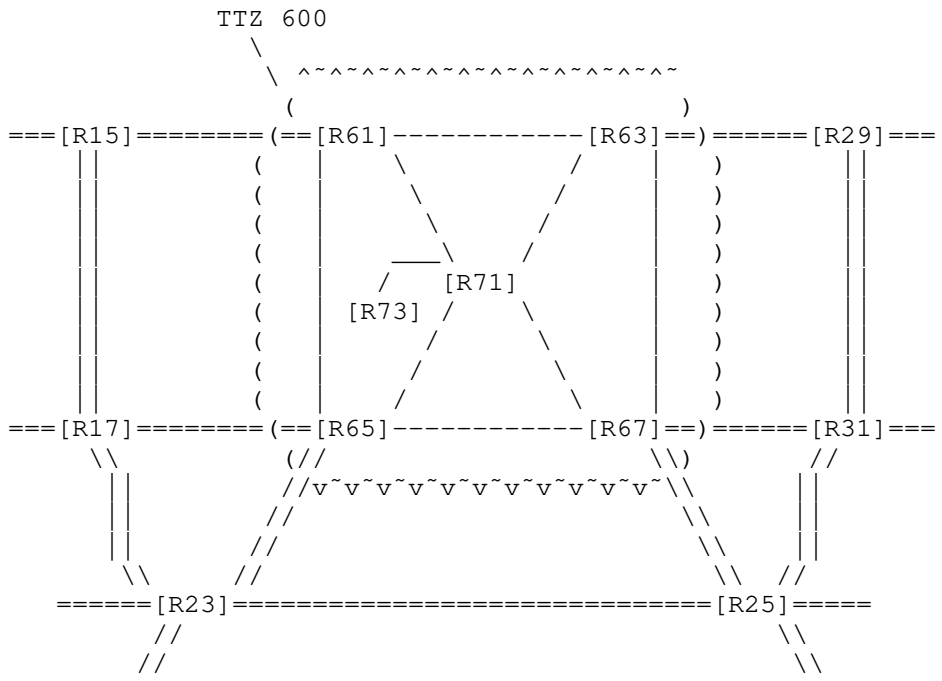



Figure 1: An Example of TTZ 600

The area comprises routers R15, R17, R23, R25, R29 and R31. It also contains TTZ 600, which comprises routers R61, R63, R65, R67, R71 and R73, and the circuits connecting them.

There are two types of routers in a TTZ: TTZ internal routers and TTZ edge/border routers. A TTZ internal router is a router inside the TTZ and its adjacent routers are inside the TTZ. A TTZ edge/border router is a router inside the TTZ and has at least one adjacent router that is outside of the TTZ.

The TTZ in the figure above comprises four TTZ edge/border routers R61, R63, R65 and R67. Each TTZ edge/border router is connected to at least one router outside of the TTZ. For instance, router R61 is a TTZ edge/border router since it is connected to router R15, which is outside of the TTZ.

In addition, the TTZ comprises two TTZ internal routers R71 and R73. A TTZ internal router is not connected to any router outside of the TTZ. For instance, router R71 is a TTZ internal router since it is not connected to any router outside of the TTZ. It is just connected to routers R61, R63, R65, R67 and R73 inside the TTZ.

A TTZ MUST hide the information inside the TTZ from the outside. It MUST NOT directly distribute any internal information about the TTZ to a router outside of the TTZ.

For instance, the TTZ in the figure above MUST NOT send the information about TTZ internal router R71 to any router outside of the TTZ in the routing domain; it MUST NOT send the information about the circuit between TTZ router R61 and R65 to any router outside of the TTZ.

From a router outside of the TTZ, a TTZ is seen as a single node (refer to the Figure below). For instance, router R15, which is outside of TTZ 600, sees TTZ 600 as a single node Rz, which has normal connections to R15, R29, R17 and R23, R25 and R31.

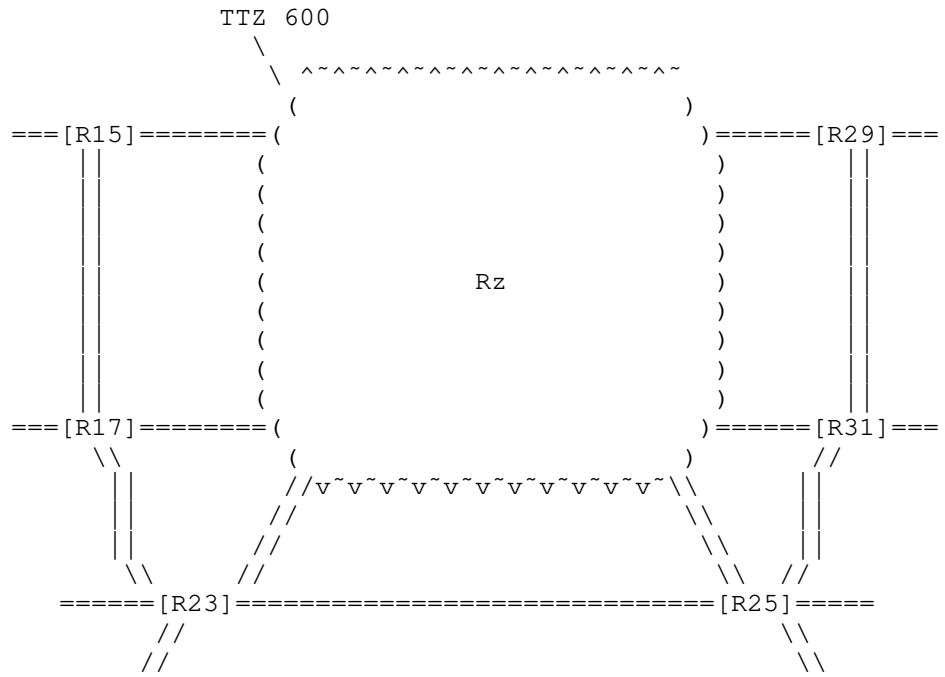



Figure 2: TTZ 600 as Single Node Rz

### 4.1.2. Zone Leader Election

A node in a zone is elected as a leader for the zone, which is the node with the highest priority (and the highest node ID when there are more than one nodes having the same highest priority) in the zone. The leader election mechanism described in

[I-D.ietf-lsr-dynamic-flooding] may be used to elect the leader for the zone.

#### 4.1.3. LS Generation for Zone as a Single Node

The leader for the zone originates an LS (i.e., an LSP in IS-IS) for the zone as a single virtual node and sends it to its neighbors.

The LS comprises all the links connecting the zone neighbors. The LS ID is the ID of the virtual node for the zone. The Source ID or Advertising Node/Router ID is the ID of the virtual node.

In addition, the LS may contain the stub links for the routes such as the loopback addresses inside the zone to be accessed by zone external nodes (i.e., nodes outside of the zone).

#### 4.1.4. Adjacency Establishment and Termination

A zone edge node, acting as a single virtual node for the zone, forms an adjacency with a node outside of the zone in a way described below.

Case 1 for a new adjacency (i.e., no adjacency exists between the edge and the node outside of the zone also called zone neighbor):

The edge node originates and sends the zone neighbor every protocol packet such as Hello, which contains the virtual node ID as Source ID.

When the edge node synchronizes its link state database (LSDB) with the zone neighbor, it sends the zone neighbor the information about all the link states except for the link states belonging to the zone that are hidden from any node outside of the zone.

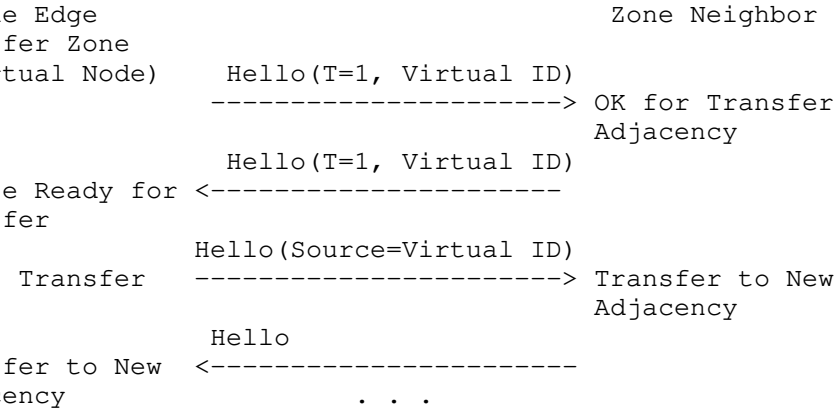
At the end of the LSDB synchronization, the LS for the zone as the single virtual node is originated by the zone leader and distributed to the zone neighbor. This LS contains the links connecting all the zone neighbors, including this newly formed zone neighbor.

Case 2 for an existing adjacency (i.e., an adjacency already exists between the zone edge and the zone neighbor):

At first, the edge acting as virtual node creates a new adjacency between the virtual node for the zone and the zone external node in a normal way. It sends Hellos and other packets containing the virtual node ID as Source ID to the zone external node. The zone external node establishes the adjacency with the virtual in the normal way.

And then, the edge terminates the existing adjacency between the edge and the external node after the zone has been transferred to the virtual node. It stops sending Hellos for the adjacency to the zone external node. Without receiving Hellos from the edge node for a given time such as hold-timer interval, the zone external node removes the adjacency to the edge node. Even though this adjacency terminates, the edge node keeps the link to the external node in its LS.

In another option, the zone edge sends Hellos to the zone neighbor with additional information, including a flag T-bit set to one and a TLV with the virtual node ID. This information requests the zone neighbor to transfer the existing adjacency to the new adjacency smoothly through working together with the zone edge in following steps.




Step 1: When "Transfer Zone to Virtual Node" is triggered, the zone edge sends the zone neighbor a Hello containing additional information T=1 and Virtual node ID.

Step 2: After receiving the Hello with T=1 and virtual node ID from the zone edge, the zone neighbor sends the zone edge a Hello with T=1 and virtual node ID, which means ok for transfer to the new adjacency.

Step 3: The edge sends the zone neighbor a Hello containing the virtual node ID as Source ID after receiving the Hello with T=1 and virtual node ID from the zone neighbor, which starts to transfer to the new adjacency.

Step 4: The zone neighbor changes the existing adjacency to the new adjacency after receiving the Hello containing the virtual node ID as Source ID from the zone edge; and sends the zone edge a Hello

without the additional information, which means that it transferred to the new adjacency.

Step 5: The zone edge changes the existing adjacency to the new adjacency after receiving the Hello without the additional information from the zone neighbor; and continues to send the zone neighbor a Hello containing the virtual node ID as Source ID. At this point, the old adjacency is transferred to the new one.

For the zone neighbor, changing the existing adjacency to the new one includes:

- Changing the existing adjacency ID from the edge node ID to the virtual node ID through either removing the existing adjacency and adding a new adjacency with the virtual node ID or just changing the existing adjacency ID from the edge node ID to the virtual node ID,
- Removing the link to the zone edge node from its LS and adding a new link to the virtual node (or just changing the link to the edge node to the link to the virtual node in its LS), and
- Continuing sending the zone edge Hellos without additional information.

For the zone edge, changing the existing adjacency to the new one includes:

- Keeping the link to the zone neighbor in its LS, and
- Continuing sending the zone neighbor Hellos containing the virtual node ID as Source ID.

#### 4.1.5. Computation of Routes

After a zone edge migrates to zone as a virtual node, it computes the routes (i.e., shortest paths to the destinations) in the zone using the zone topology (i.e., the topology of the zone without the virtual node).

For the routes outside of the zone, it computes them using the zone topology, the topology outside of the zone without the virtual node and the connections between each zone edge and its zone neighbor.

After a zone internal node migrates to zone as a virtual node, it computes the routes using the zone topology, the topology outside of the zone without the virtual node and the connections between each zone edge and its zone neighbor.

#### 4.1.6. Extensions to Protocols

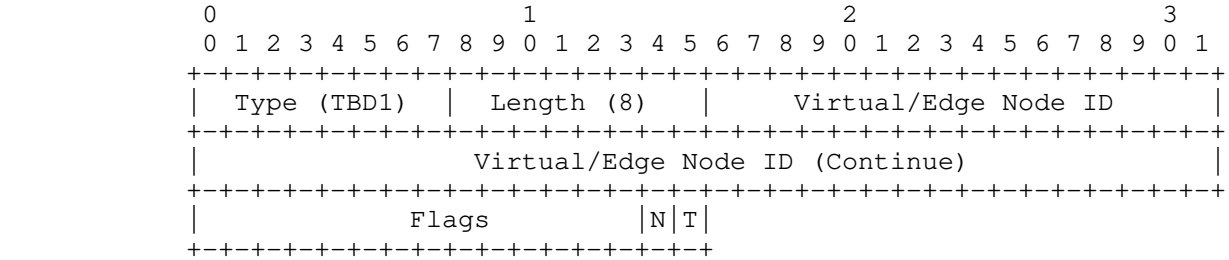
The following TLVs are defined in IS-IS.

- Adjacent Node ID TLV: containing an adjacent node ID, to which an adjacency is transferred or rolled back. In case of transfer, the TLV contains the virtual node ID; in case of roll back, the TLV contains the edge node ID.

- Zone TLV: containing a zone ID, a flags field and optional sub-TLVs.

##### 4.1.6.1. Adjacent Node ID TLV

The format of Adjacent Node ID TLV is illustrated below.




Type (1 byte): To be assigned by IANA.

Length (1 byte): Its value is 8.

Virtual/Edge Node ID (6 bytes): An adjacent node ID, to which an adjacency is transferred or rolled back.

Flags field (16 bits): two new flag bits are defined as follows:

- T-bit: Short for Transfer Adjacency bit. The T-bit set to one indicates a request for transferring to a new 'virtual' adjacency from the existing adjacency and the new adjacency is identified by the virtual node ID (or say abstract node ID).

- N-bit: Short for Roll Back to Normal Adjacency bit. The N-bit set to one indicates a request for rolling back to a Normal adjacency from the existing 'virtual' adjacency and the normal adjacency is identified by the edge node ID.

4.1.6.2. Zone TLV

The format of IS-IS Zone TLV is illustrated below. It may be added into an LSP or a Hello PDU for a zone node. When a node in a zone receives a CLI command triggering zone information distribution for migration, it updates its LSP by adding an IS-IS Zone TLV with T set to 1. When a node in a zone receives a CLI command activating migration zone to an abstracted entity, it sets M to 1 in the Zone TLV in its LSP.

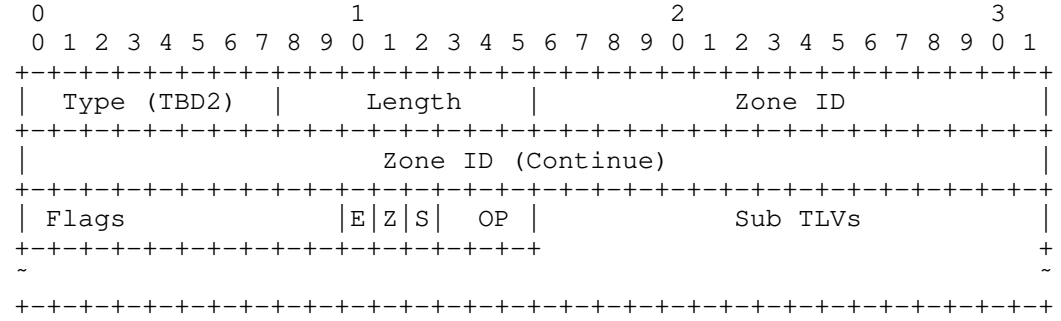



Figure 3: IS-IS Zone TLV

Type (1 byte): To be assigned by IANA.

Length (1 byte): Its value is variable.

Zone ID (6 bytes): It is the identifier (ID) of a zone.

Flags field (16 bits): Three flag bits E, Z and S, and OP of 3 bits are defined.

E = 1: Indicating a node is a zone edge node
Z = 1: Indicating a node has migrated to Zone as a virtual entity
S = 1: Indicating the virtual entity is a Single virtual node

When a zone node originates an LS containing a zone TLV, it MUST set flag E to 1 if it is a zone edge node and to 0 if it is a zone-internal node. It MUST set flag Z to 1 after it has migrated to zone as a virtual entity and to 0 before it migrates zone to the virtual entity or after it rolls back from zone as a virtual entity. When the entity abstracted from a zone is a Single virtual node, flag S MUST be set to 1.

| OP Value | Meaning (Operation) |
|---|---|
| 0x001 (T): | Advertising Zone Topology Information for Migration |
| 0x010 (M): | Migrating Zone to a Virtual Entity |
| 0x011 (N): | Advertising Normal Topology Information for Rollback |
| 0x100 (R): | Rolling Back from the Virtual Entity |

The value of OP indicates one of the four operations above. When any of the other values is received, it is ignored.

When a node in a zone receives a CLI command triggering zone information distribution for migration, it updates its LSP by adding an IS-IS Zone TLV with T set to 1. When a node in a zone receives a CLI command activating migration zone to a virtual entity, it sets M to 1 in the Zone TLV in its LSP.

Two new sub-TLVs are defined, which may be added into an IS-IS Zone TLV in an LSP. One is Zone IS Neighbor sub-TLV, or Zone ISN sub-TLV for short. The other is Zone ES Neighbor sub-TLV, or Zone ESN sub-TLV for short. A Zone ISN sub-TLV contains the information about a number of IS neighbors in the zone connected to a zone edge router. It has the format below.

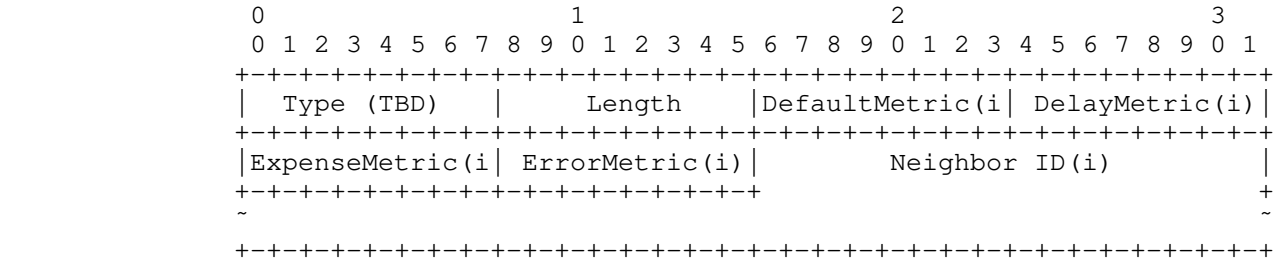



Figure 4: Zone ISN Sub TLV

A Zone ISN Sub TLV has 1 byte of Type, 1 byte of Length of n*(IDLength + 4), which is followed by n tuples of Default Metric, Delay Metric, Expense Metric, Error Metric and Neighbor ID.

A Zone ESN sub-TLV contains the information about a number of ES neighbors in the zone connected to a zone edge node. It has the format below.

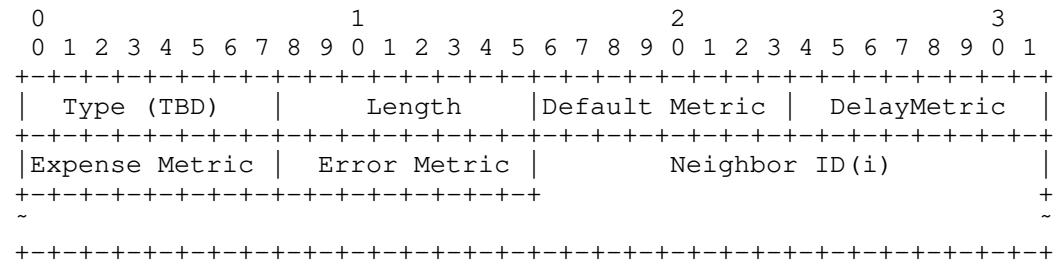



Figure 5: Zone ESN Sub TLV

## 4.2. Zone as Edges Full Mesh

OSPF Topology-Transparent Zone [RFC8099] describes the zone as edges' full mesh and the extensions to OSPF for supporting zone as edges' full mesh. Based on these extensions, IS-IS is extended by a few new TLVs or Sub-TLVs.

### 4.2.1. Extensions to IS-IS

#### 4.2.1.1. Updating LSPs for Zone

A zone internal node adds an IS-IS Zone TLV into its LSP after it receives an LSP containing an IS-IS Zone TLV with T = 1 or a CLI command triggering zone information distribution for migration. The TLV has a zone ID set to the ID of the zone and E bit in Flags set to 0 indicating zone internal node. The node floods its LSP to its neighbors in the zone.

When a node inside the zone receives an LSP containing an IS-IS Zone TLV from a neighboring node in the zone, it stores the LSP and floods the LSP to the other neighboring nodes in the zone.

For every zone edge node, it updates its LSP in three steps and floods the LSP to all its neighbors.

At first, the zone edge node adds an IS-IS Zone TLV into its LSP after it receives an LSP containing an IS-IS Zone TLV with T = 1 or a CLI command triggering zone information distribution for migration. The TLV has a zone ID set to the ID of the zone, E bit in Flags set to 1 indicating zone edge node and a Zone ISN Sub TLV. The Sub TLV contains the information about the zone IS neighbors connected to the zone edge node. In addition, the TLV may has a Zone ESN Sub TLV comprising the information about the zone end systems connected to the zone edge node.

Secondly, it adds each of the other zone edge nodes as an IS neighbor into the Intermediate System Neighbors TLV in the LSP after it receives an LSP containing an IS-IS Zone TLV with M = 1 or a CLI command activating migration zone to an abstracted entity. The metric to the neighbor is the metric of the shortest path to the edge node within the zone.

In addition, it adds a Prefix Neighbors TLV into its LSP. The TLV contains a number of address prefixes in the zone to be reachable from outside of the zone.

And then it removes the IS neighbors corresponding to the IS neighbors in the Zone TLV (i.e., in the Zone ISN sub TLV) from Intermediate System Neighbors TLV in the LSP, and the ES neighbors corresponding to the ES neighbors in the Zone TLV (i.e., in the Zone ESN sub TLV) from End System Neighbors TLV in the LSP. This SHOULD be done after it receives the LSPs for virtualizing zone from the other zone edges for a given time.

### 4.3. Advertisement of LSs

LSs can be divided into a couple of classes according to their Advertisements. The first class of LSs is advertised within a zone. The second is advertised through a zone.

#### 4.3.1. Advertisement of LSs within Zone

Any LS about a link state in a zone is advertised only within the zone. It is not advertised to any router outside of the zone. For example, a router LS generated for a zone internal router is advertised only within the zone.

Any network LS generated for a broadcast network in a zone is advertised only within the zone. It is not advertised outside of the zone.

After migrating to zone as a single virtual node or edges' full mesh, every zone edge MUST NOT advertise any LS belonging to the zone or any information in a LS belonging to the zone to any node outside of the zone. The zone edge determines whether an LS is about a zone internal link state by checking if the advertising router of the LS is a zone internal router.

For any zone LS originated by a node within the zone, every zone edge node MUST NOT advertise it to any node outside of the zone.

#### 4.3.2. Advertisement of LSs through Zone

Any LS about a link state outside of a zone received by a zone edge is advertised using the zone as transit. For example, when a zone edge node receives an LS from a node outside of the zone, it floods the LS to its neighbors both inside and outside of the zone. This LS may be any LS such as a router LSA that is advertised within an OSPF area.

The nodes in the zone continue to flood the LS. When another zone edge receives the LS, it floods the LS to its neighbors both inside and outside of the zone.

## 5. Seamless Migration

This section presents the seamless migration between a zone and its single virtual node. The seamless migration between a zone and its edges' full mesh for IS-IS is similar to that described in OSPF Topology-Transparent Zone [RFC8099] for OSPF.

### 5.1. Transfer Zone to a Single Node

After transfer a Zone to a Single Virtual Node is triggered, the zone is abstracted as a single virtual node in two steps:

Step 1: Every zone edge node works together with each of its zone neighbor nodes to create a new adjacency between the virtual node and the neighbor node in the way described in Section 4.1.4 for Adjacency Establishment and Termination procedure for case 2. After creating the adjacency, each of the zone neighbor nodes update its LS by adding the adjacency/link into its LS.

Step 2: The zone leader originates an LS for the virtual node after receiving the updated LSes originated by all the zone neighbor nodes, where the updated LSes contain all the zone neighbors.

Step 3: After receiving the LS for the virtual node, every zone edge does not send any LS inside the zone to any zone neighbors. It advertises its LS without any links inside the zone to the nodes outside of the zone and terminates its adjacency to each of its zone neighbors in the way described in Section 4.1.4 for Adjacency Establishment and Termination procedure for case 2.

### 5.2. Roll Back from Zone as a Single Node

After roll back from Zone as a Signle Virtual Node is triggered, rolling back is done in following steps:

Step 1: Every zone edge creates an adjacency to each of its zone neighbors in a normal way.

Step 2: After all the adjacencies between the zone edges and the zone neighbors are created, the zone leader updates the LS for the virtual node by changing every link metric to the maximum metric in the LS.

Step 3: Every zone edge sends its LS with the links inside the zone and all the LSes inside the zone to its zone neighbors. Every zone edge acting as the virtual node terminates the adjacency between the virtual node and each of its zone neighbors through stopping Hellos to the neighbors.

In another option, rolling back is done as follows:

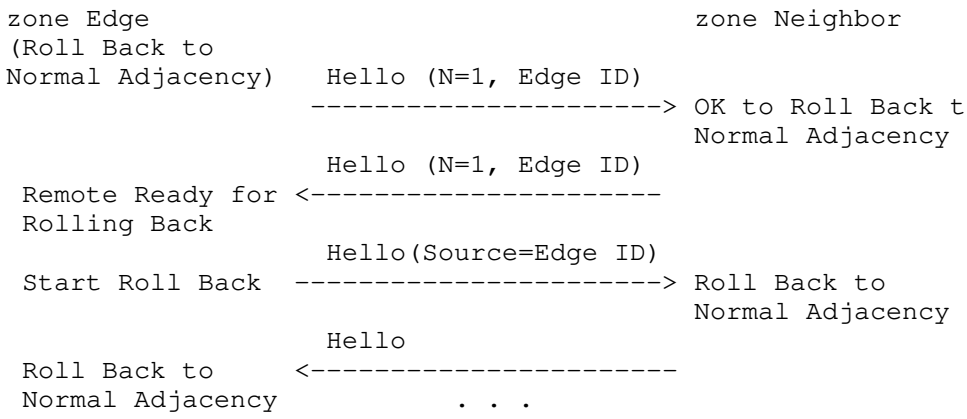
Step 1: Using the procedure described in the following, every zone edge rolls back the existing virtual adjacency between the edge node acting as the virtual node and the zone neighbor node to a normal adjacency between the edge node and the neighbor.

Step 2: The zone leader may flush the LS for the virtual node. Every zone edge sends Hello and other packets to its zone neighbors, where the packets contain the edge node ID as Source ID.

The procedure below smoothly rolls back the existing virtual adjacency between the edge node acting as the virtual node and the zone neighbor node to a normal adjacency between the edge node and the neighbor node.

The edge node sends the neighbor node Hellos with additional information, including a flag N-bit set to one and a TLV with the edge node ID such as the Adjacent Node ID TLV with the edge node ID. This information requests the neighbor node to roll back the existing virtual adjacency to the normal adjacency smoothly through working together with the edge node.

The following steps will roll back the existing virtual adjacency to the normal one:




Step 1: When "Roll Back from Zone as a Single Node" is triggered, the edge node sends the neighbor node a Hello with the additional information N=1 and Edge ID as normal adjacency ID in order to roll back to the normal adjacency from the virtual adjacency.

Step 2: After receiving the Hello with the additional information from the edge node, the neighbor node sends the edge node a Hello with the additional information (i.e., N=1 and Edge ID as normal adjacency ID), which means ok for rolling back to the normal adjacency.

Step 3: The edge sends the neighbor a Hello containing the edge node ID as Source ID after receiving the Hello with the additional information from the neighbor, which starts to roll back to the normal adjacency.

Step 4: The neighbor node changes the existing adjacency to the normal adjacency after receiving the Hello containing the edge node ID as Source ID from the edge node; and sends the edge node a Hello without the additional information, which means that it rolled back to the normal adjacency.

Step 5: The edge node changes the existing adjacency to the normal adjacency after receiving the Hello without the additional information from the neighbor node; and continues to send the neighbor Hello containing the edge node ID as Source ID. At this point, the virtual adjacency is rolled back to the normal adjacency.

For the neighbor node, changing the existing virtual adjacency to the normal one includes:

- Changing the existing adjacency ID from the virtual node ID to the edge node ID through either removing the existing adjacency and adding a new adjacency with the edge node ID or just changing the existing adjacency ID from the virtual node ID to the edge node ID,

- Removing the link to the virtual node from its LS and adding a new link to the edge node (or just changing the link to the virtual node to the link to the edge node in its LS), and

- Continuing sending the edge node Hellos without additional information.

For the edge node, changing the existing virtual adjacency to the normal one includes:

- Sending its LS to the neighbor, and

- Continuing sending the neighbor node Hellos containing the edge node ID as Source ID without additional information.

## 6. Operations

The Operations on TTZ described in OSPF Topology-Transparent Zone [RFC8099] are for Zone as Edges Full Mesh in OSPF. They can be used for Zone as Edges Full Mesh in IS-IS. They can also be used for Zone as a Single Virtual Node in IS-IS.

## 7. Security Considerations

The mechanism described in this document does not raise any new security issues for the IS-IS protocols.

## 8. IANA Considerations

Under the registry name "IS-IS TLV Codepoints", IANA is requested to assign new registry types for Adjacent Node ID, Zone ID and Zone Options as follows:

| TLV Type | TLV Name | reference |
|---|---|---|
| 26(suggested) | Adjacent Node ID | This document |
| 27(suggested) | Zone | This document |

## 9. Contributors

Alvaro Retana
Futurewei
Raleigh, NC
USA

Email: alvaro.retana@futurewei.com

## 10. Acknowledgement

The authors would like to thank Acee Lindem, Abhay Roy, Christian Hopps, Dean Cheng, Russ White, Tony Przygienda, Wenhu Lu, Lin Han, Kiran Makhijani, Padmadevi Pillay Esnault, and Yang Yu for their valuable comments on TTZ.

## 11. References

### 11.1. Normative References

[I-D.ietf-lsr-dynamic-flooding]
Li, T., Psenak, P., Ginsberg, L., Chen, H., Przygienda, T., Cooper, D., Jalil, L., Dontula, S., and G. Mishra, "Dynamic Flooding on Dense Graphs", draft-ietf-lsr-dynamic-flooding-07 (work in progress), June 2020.

[I-D.ietf-spring-segment-routing]
Filsfils, C., Previdi, S., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-15 (work in progress), January 2018.

[ISO10589]
International Organization for Standardization, "Intermediate System to Intermediate System Intra-Domain Routing Exchange Protocol for use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service (ISO 8473)", ISO/IEC 10589:2002, Nov. 2002.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, DOI 10.17487/RFC1195, December 1990, <https://www.rfc-editor.org/info/rfc1195>.

[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>.

[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>.

[RFC7142] Shand, M. and L. Ginsberg, "Reclassification of RFC 1142 to Historic", RFC 7142, DOI 10.17487/RFC7142, February 2014, <https://www.rfc-editor.org/info/rfc7142>.

[RFC8099] Chen, H., Li, R., Retana, A., Yang, Y., and Z. Liu, "OSPF Topology-Transparent Zone", RFC 8099, DOI 10.17487/RFC8099, February 2017, <https://www.rfc-editor.org/info/rfc8099>.

## 11.2. Informative References

[Clos] Clos, C., "A Study of Non-Blocking Switching Networks", The Bell System Technical Journal Vol. 32(2), DOI 10.1002/j.1538-7305.1953.tb01433.x, March 1953, <http://dx.doi.org/10.1002/j.1538-7305.1953.tb01433.x>.

[RFC5307] Kompella, K., Ed. and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, DOI 10.17487/RFC5307, October 2008, <https://www.rfc-editor.org/info/rfc5307>.

## Authors' Addresses

Huaimo Chen
Futurewei
Boston, MA
USA

Email: huaimo.chen@futurewei.com

Richard Li
Futurewei
2330 Central expressway
Santa Clara, CA
USA

Email: richard.li@futurewei.com

Yi Yang
IBM
Cary, NC
United States of America

Email: yyietf@gmail.com

Anil Kumar S N
RtBrick
Bangalore
India

Email: anil.ietf@gmail.com

Yanhe Fan
Casa Systems
USA

Email: yfan@casa-systems.com

Ning So
Plano, TX 75082
USA

Email: ningso01@gmail.com

Vic Liu
USA

Email: liu.cmri@gmail.com

Mehmet Toy
Verizon
USA

Email: mehmet.toy@verizon.com

Lei Liu
Fujitsu
USA

Email: liulei.kddi@gmail.com

Kiran Makhijani
Futurewei
USA

Email: kiranm@futurewei.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 14, 2021

P. Psenak, Ed.
Cisco Systems
S. Hegde
Juniper Networks, Inc.
C. Filsfils
K. Talaulikar
Cisco Systems, Inc.
A. Gulko
Individual
September 10, 2020

# IGP Flexible Algorithm
draft-ietf-lsr-flex-algo-11.txt

## Abstract

IGP protocols traditionally compute best paths over the network based on the IGP metric assigned to the links. Many network deployments use RSVP-TE based or Segment Routing based Traffic Engineering to steer traffic over a path that is computed using different metrics or constraints than the shortest IGP path. This document proposes a solution that allows IGPs themselves to compute constraint-based paths over the network. This document also specifies a way of using Segment Routing (SR) Prefix-SIDs and SRv6 locators to steer packets along the constraint-based paths.

## 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 March 14, 2021.

Table of Contents

## 1. Introduction

An IGP-computed path based on the shortest IGP metric must often be replaced by a traffic-engineered path due to the traffic requirements which are not reflected by the IGP metric. Some networks engineer the IGP metric assignments in a way that the IGP metric reflects the link bandwidth or delay. If, for example, the IGP metric is reflecting the bandwidth on the link and the application traffic is delay sensitive, the best IGP path may not reflect the best path from such an application's perspective.

To overcome this limitation, various sorts of traffic engineering have been deployed, including RSVP-TE and SR-TE, in which case the TE component is responsible for computing paths based on additional metrics and/or constraints. Such paths need to be installed in the forwarding tables in addition to, or as a replacement for, the original paths computed by IGPs. Tunnels are often used to represent

the engineered paths and mechanisms like one described in [RFC3906] are used to replace the native IGP paths with such tunnel paths.

This document specifies a set of extensions to ISIS, OSPFv2, and OSPFv3 that enable a router to advertise TLVs that identify (a) calculation-type, (b) specify a metric-type, and (c) describe a set of constraints on the topology, that are to be used to compute the best paths along the constrained topology. A given combination of calculation-type, metric-type, and constraints is known as a "Flexible Algorithm Definition". A router that sends such a set of TLVs also assigns a Flex-Algorithm value to the specified combination of calculation-type, metric-type, and constraints.

This document also specifies a way for a router to use IGPs to associate one or more SR Prefix-SIDs or SRv6 locators with a particular Flex-Algorithm. Each such Prefix-SID or SRv6 locator then represents a path that is computed according to the identified Flex-Algorithm.

## 2. Requirements notation

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 [BCP14] [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 3. Terminology

This section defines terms that are often used in this document.

Flexible Algorithm Definition (FAD) - the set consisting of (a) calculation-type, (b) metric-type, and (c) a set of constraints.

Flexible Algorithm - a numeric identifier in the range 128-255 that is associated via configuration with the Flexible-Algorithm Definition.

Local Flexible Algorithm Definition - Flexible Algorithm Definition defined locally on the node.

Remote Flexible Algorithm Definition - Flexible Algorithm Definition received from other nodes via IGP flooding.

Flexible Algorithm Participation - per application configuration state that expresses whether the node is participating in a particular Flexible Algorithm.

IGP Algorithm - value from the the "IGP Algorithm Types" registry defined under "Interior Gateway Protocol (IGP) Parameters" IANA registries. IGP Algorithms represents the triplet (Calculation Type, Metric, Constraints), where the second and third elements of the triple MAY not be specified.

ABR - Area Border Router. In ISIS terminology it is also known as L1/L2 router.

ASBR - Autonomous System Border Router.

## 4. Flexible Algorithm

Many possible constraints may be used to compute a path over a network. Some networks are deployed as multiple planes. A simple form of constraint may be to use a particular plane. A more sophisticated form of constraint can include some extended metric as described in [RFC8570]. Constraints which restrict paths to links with specific affinities or avoid links with specific affinities are also possible. Combinations of these are also possible.

To provide maximum flexibility, we want to provide a mechanism that allows a router to (a) identify a particular calculation-type, (b) metric-type, (c) describe a particular set of constraints, and (d) assign a numeric identifier, referred to as Flex-Algorithm, to the combination of that calculation-type, metric-type, and those constraints. We want the mapping between the Flex-Algorithm and its meaning to be flexible and defined by the user. As long as all routers in the domain have a common understanding as to what a particular Flex-Algorithm represents, the resulting routing computation is consistent and traffic is not subject to any looping.

The set consisting of (a) calculation-type, (b) metric-type, and (c) a set of constraints is referred to as a Flexible-Algorithm Definition.

Flexible-Algorithm is a numeric identifier in the range 128-255 that is associated via configuratin with the Flexible-Algorithm Definition.

IANA "IGP Algorithm Types" registry defines the set of values for IGP Algorithms. We propose to allocate the following values for Flex-Algorithms from this registry:

128-255 - Flex-Algorithms

## 5. Flexible Algorithm Definition Advertisement

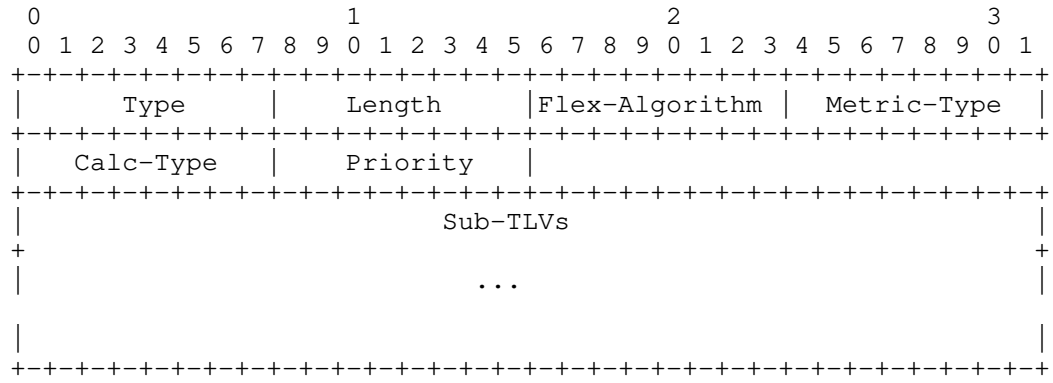
To guarantee the loop-free forwarding for paths computed for a particular Flex-Algorithm, all routers that (a) are configured to participate in a particular Flex-Algorithm, and (b) are in the same Flex-Algorithm definition advertisement scope MUST agree on the definition of the Flex-Algorithm.

### 5.1. ISIS Flexible Algorithm Definition Sub-TLV

The ISIS Flexible Algorithm Definition Sub-TLV (FAD Sub-TLV) is used to advertise the definition of the Flex-Algorithm.

The ISIS FAD Sub-TLV is advertised as a Sub-TLV of the ISIS Router Capability TLV-242 that is defined in [RFC7981].

ISIS FAD Sub-TLV has the following format:




where:

Type: 26

Length: variable, dependent on the included Sub-TLVs

Flex-Algorithm: Single octet value between 128 and 255 inclusive.

Metric-Type: Type of metric to be used during the calculation. Following values are defined:

0: IGP Metric

1: Min Unidirectional Link Delay as defined in [RFC8570], section 4.2, encoded as application specific link attribute as

specified in [I-D.ietf-isis-te-app] and Section 11 of this document.

2: Traffic Engineering Default Metric as defined in [RFC5305], section 3.7, encoded as application specific link attribute as specified in [I-D.ietf-isis-te-app] and Section 11 of this document.

Calc-Type: value from 0 to 127 inclusive from the "IGP Algorithm Types" registry defined under "Interior Gateway Protocol (IGP) Parameters" IANA registries. IGP algorithms in the range of 0-127 have a defined triplet (Calculation Type, Metric, Constraints). When used to specify the Calc-Type in the FAD Sub-TLV, only the Calculation Type defined for the specified IGP Algorithm is used. The Metric/Constraints MUST NOT be inherited. If the required calculation type is Shortest Path First, the value 0 SHOULD appear in this field.

Priority: Value between 0 and 255 inclusive that specifies the priority of the advertisement.

Sub-TLVs - optional sub-TLVs.

The ISIS FAD Sub-TLV MAY be advertised in an LSP of any number, but a router MUST NOT advertise more than one ISIS FAD Sub-TLV for a given Flexible-Algorithm. A router receiving multiple ISIS FAD Sub-TLVs for a given Flexible-Algorithm from the same originator SHOULD select the first advertisement in the lowest numbered LSP.

The ISIS FAD Sub-TLV has an area scope. The Router Capability TLV in which the FAD Sub-TLV is present MUST have the S-bit clear.

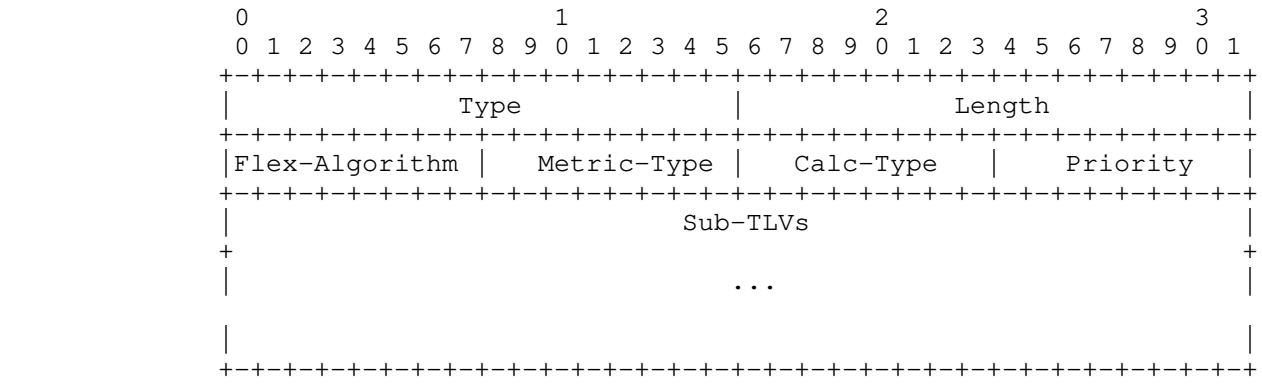
ISIS L1/L2 router MAY be configured to re-generate the winning FAD from level 2, without any modification to it, to level 1 area. The re-generation of the FAD Sub-TLV from level 2 to level 1 is determined by the L1/L2 router, not by the originator of the FAD advertisement in the level 2. In such case, the re-generated FAD Sub-TLV will be advertised in the level 1 Router Capability TLV originated by the L1/L2 router.

L1/L2 router MUST NOT re-generate any FAD Sub-TLV from level 1 to level 2.

### 5.2. OSPF Flexible Algorithm Definition TLV

OSPF FAD TLV is advertised as a top-level TLV of the RI LSA that is defined in [RFC7770].

OSPF FAD TLV has the following format:




where:

Type: 16

Length: variable, dependent on the included Sub-TLVs

Flex-Algorithm:: Flex-Algorithm number. Value between 128 and 255 inclusive.

Metric-Type: Type of metric to be used during the calculation. Following values are defined:

0: IGP Metric

1: Min Unidirectional Link Delay as defined in [RFC7471], section 4.2, encoded as application specific link attribute as specified in [I-D.ietf-ospf-te-link-attr-reuse] and Section 11 of this document.

2: Traffic Engineering metric as defined in [RFC3630], section 2.5.5, encoded as application specific link attribute as specified in [I-D.ietf-ospf-te-link-attr-reuse] and Section 11 of this document.

Calc-Type: as described in Section 5.1

Priority: as described in Section 5.1

Sub-TLVs - optional sub-TLVs.

When multiple OSPF FAD TLVs, for the same Flexible-Algorithm, are received from a given router, the receiver MUST use the first occurrence of the TLV in the Router Information LSA. If the OSPF FAD TLV, for the same Flex-Algorithm, appears in multiple Router Information LSAs that have different flooding scopes, the OSPF FAD TLV in the Router Information LSA with the area-scoped flooding scope MUST be used. If the OSPF FAD TLV, for the same algorithm, appears in multiple Router Information LSAs that have the same flooding scope, the OSPF FAD TLV in the Router Information (RI) LSA with the numerically smallest Instance ID MUST be used and subsequent instances of the OSPF FAD TLV MUST be ignored.

The RI LSA can be advertised at any of the defined opaque flooding scopes (link, area, or Autonomous System (AS)). For the purpose of OSPF FAD TLV advertisement, area-scoped flooding is REQUIRED. The Autonomous System flooding scope SHOULD not be used by default unless local configuration policy on the originating router indicates domain wide flooding.

### 5.3. Common Handling of Flexible Algorithm Definition TLV

This section describes the protocol-independent handling of the FAD TLV (OSPF) or FAD Sub-TLV (ISIS). We will refer to it as FAD TLV in this section, even though in case of ISIS it is a Sub-TLV.

The value of the Flex-Algorithm MUST be between 128 and 255 inclusive. If it is not, the FAD TLV MUST be ignored.

Only a subset of the routers participating in the particular Flex-Algorithm need to advertise the definition of the Flex-Algorithm.

Every router, that is configured to participate in a particular Flex-Algorithm, MUST select the Flex-Algorithm definition based on the following ordered rules. This allows for the consistent Flex-Algorithm definition selection in cases where different routers advertise different definitions for a given Flex-Algorithm:

1. From the advertisements of the FAD in the area (including both locally generated advertisements and received advertisements) select the one(s) with the highest priority value.

2. If there are multiple advertisements of the FAD with the same highest priority, select the one that is originated from the router with the highest System-ID, in the case of ISIS, or Router ID, in the case of OSPFv2 and OSPFv3. For ISIS, the System-ID is described in [ISO10589]. For OSPFv2 and OSPFv3, standard Router ID is described in [RFC2328] and [RFC5340] respectively.

A router that is not configured to participate in a particular Flex-Algorithm MUST ignore FAD Sub-TLVs advertisements for such Flex-Algorithm.

A router that is not participating in a particular Flex-Algorithm is allowed to advertise FAD for such Flex-Algorithm. Receiving routers MUST consider FAD advertisement regardless of the Flex-Algorithm participation of the FAD originator.

Any change in the Flex-Algorithm definition may result in temporary disruption of traffic that is forwarded based on such Flex-Algorithm paths. The impact is similar to any other event that requires network-wide convergence.

If a node is configured to participate in a particular Flexible-Algorithm, but the selected Flex-Algorithm definition includes calculation-type, metric-type, constraint, flag, or Sub-TLV that is not supported by the node, it MUST stop participating in such Flexible-Algorithm. That implies that it MUST NOT announce participation for such Flexible-Algorithm as specified in Section 10 and it MUST remove any forwarding state associated with it.

Flex-Algorithm definition is topology independent. It applies to all topologies that a router participates in.

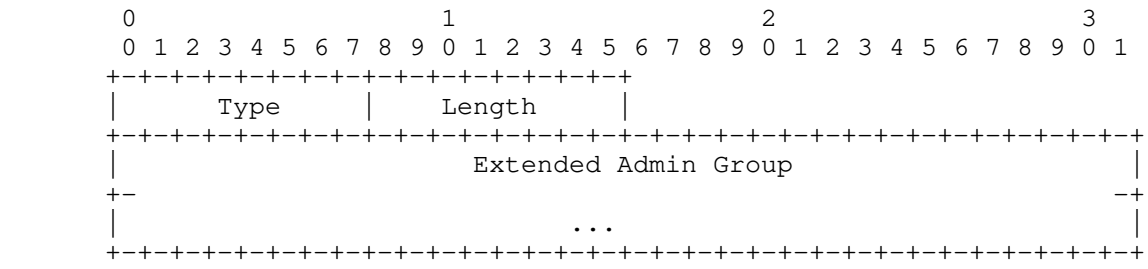
## 6. Sub-TLVs of ISIS FAD Sub-TLV

### 6.1. ISIS Flexible Algorithm Exclude Admin Group Sub-TLV

The Flexible Algorithm definition can specify 'colors' that are used by the operator to exclude links during the Flex-Algorithm path computation.

The ISIS Flexible Algorithm Exclude Admin Group Sub-TLV is used to advertise the exclude rule that is used during the Flex-Algorithm path calculation as specified in Section 12.

The ISIS Flexible Algorithm Exclude Admin Group Sub-TLV (FAEAG Sub-TLV) is a Sub-TLV of the ISIS FAD Sub-TLV. It has the following format:




where:

Type: 1

Length: variable, dependent on the size of the Extended Admin Group. MUST be a multiple of 4 octets.

Extended Administrative Group: Extended Administrative Group as defined in [RFC7308].

The ISIS FAEAG Sub-TLV MAY NOT appear more then once in an ISIS FAD Sub-TLV. If it appears more then once, the ISIS FAD Sub-TLV MUST be ignored by the receiver.

## 6.2. ISIS Flexible Algorithm Include-Any Admin Group Sub-TLV

The Flexible Algorithm definition can specify 'colors' that are used by the operator to include links during the Flex-Algorithm path computation.

The ISIS Flexible Algorithm Include-Any Admin Group Sub-TLV is used to advertise include-any rule that is used during the Flex-Algorithm path calculation as specified in Section 12.

The format of the ISIS Flexible Algorithm Include-Any Admin Group Sub-TLV is identical to the format of the FAEAG Sub-TLV in Section 6.1.

The ISIS Flexible Algorithm Include-Any Admin Group Sub-TLV Type is 2.

The ISIS Flexible Algorithm Include-Any Admin Group Sub-TLV MAY NOT appear more then once in an ISIS FAD Sub-TLV. If it appears more then once, the ISIS FAD Sub-TLV MUST be ignored by the receiver.

### 6.3. ISIS Flexible Algorithm Include-All Admin Group Sub-TLV

The Flexible Algorithm definition can specify 'colors' that are used by the operator to include link during the Flex-Algorithm path computation.

The ISIS Flexible Algorithm Include-All Admin Group Sub-TLV is used to advertise include-all rule that is used during the Flex-Algorithm path calculation as specified in Section 12.

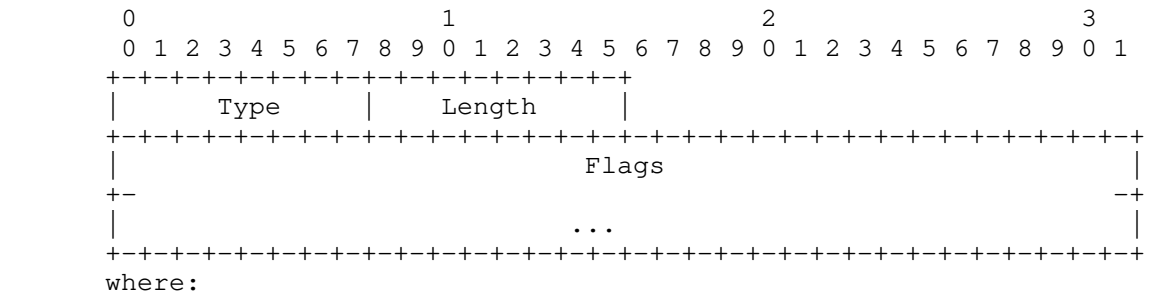
The format of the ISIS Flexible Algorithm Include-All Admin Group Sub-TLV is identical to the format of the FAEAG Sub-TLV in Section 6.1.

The ISIS Flexible Algorithm Include-All Admin Group Sub-TLV Type is 3.

The ISIS Flexible Algorithm Include-All Admin Group Sub-TLV MAY NOT appear more then once in an ISIS FAD Sub-TLV. If it appears more then once, the ISIS FAD Sub-TLV MUST be ignored by the receiver.

### 6.4. ISIS Flexible Algorithm Definition Flags Sub-TLV

The ISIS Flexible Algorithm Definition Flags Sub-TLV (FADF Sub-TLV) is a Sub-TLV of the ISIS FAD Sub-TLV. It has the following format:

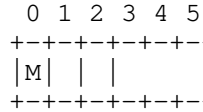



where:

Type: 4

Length: variable, non-zero number of octets of the Flags field

Flags:



M-flag: when set, the Flex-Algorithm specific prefix metric MUST be used, if advertised with the prefix. This flag is not applicable to prefixes advertised as SRv6 locators.

Bits are defined/sent starting with Bit 0 defined above. Additional bit definitions that may be defined in the future SHOULD be assigned in ascending bit order so as to minimize the number of bits that will need to be transmitted.

Undefined bits MUST be transmitted as 0.

Bits that are NOT transmitted MUST be treated as if they are set to 0 on receipt.

The ISIS FADF Sub-TLV MAY NOT appear more then once in an ISIS FAD Sub-TLV. If it appears more then once, the ISIS FAD Sub-TLV MUST be ignored by the receiver.

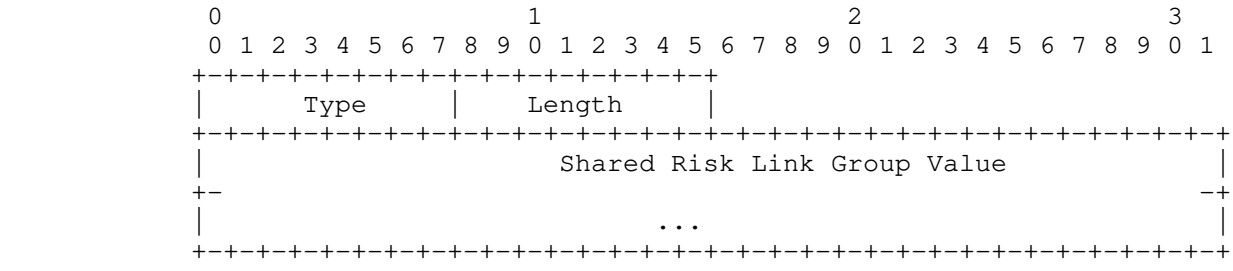
If the ISIS FADF Sub-TLV is not present inside the ISIS FAD Sub-TLV, all the bits are assumed to be set to 0.

## 6.5. ISIS Flexible Algorithm Exclude SRLG Sub-TLV

The Flexible Algorithm definition can specify Shared Risk Link Groups (SRLGs) that the operator wants to exclude during the Flex-Algorithm path computation.

The ISIS Flexible Algorithm Exclude SRLG Sub-TLV (FAESRLG) is used to advertise the exclude rule that is used during the Flex-Algorithm path calculation as specified in Section 12.

The ISIS FAESRLG Sub-TLV is a Sub-TLV of the ISIS FAD Sub-TLV. It has the following format:




where:

Type: 5

Length: variable, dependent on number of SRLG values. MUST be a multiple of 4 octets.

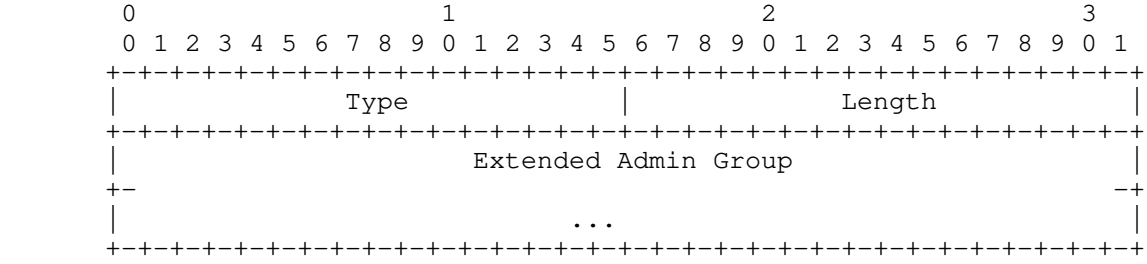
Shared Risk Link Group Value: SRLG value as defined in [RFC5307].

The ISIS FAESRLG Sub-TLV MAY NOT appear more then once in an ISIS FAD Sub-TLV. If it appears more then once, the ISIS FAD Sub-TLV MUST be ignored by the receiver.

# 7. Sub-TLVs of OSPF FAD TLV

## 7.1. OSPF Flexible Algorithm Exclude Admin Group Sub-TLV

The Flexible Algorithm Exclude Admin Group Sub-TLV (FAEAG Sub-TLV) is a Sub-TLV of the OSPF FAD TLV. It's usage is described in Section 6.1. It has the following format:




where:

Type: 1

Length: variable, dependent on the size of the Extended Admin Group. MUST be a multiple of 4 octets.

Extended Administrative Group: Extended Administrative Group as defined in [RFC7308].

The OSPF FAEAG Sub-TLV MAY NOT appear more then once in an OSPF FAD TLV. If it appears more then once, the OSPF FAD TLV MUST be ignored by the receiver.

## 7.2. OSPF Flexible Algorithm Include-Any Admin Group Sub-TLV

The usage of this Sub-TLVs is described in Section 6.2.

The format of the OSPF Flexible Algorithm Include-Any Admin Group Sub-TLV is identical to the format of the OSPF FAEAG Sub-TLV in Section 7.1.

The OSPF Flexible Algorithm Include-Any Admin Group Sub-TLV Type is 2.

The OSPF Flexible Algorithm Include-Any Admin Group Sub-TLV MAY NOT appear more then once in an OSPF FAD TLV. If it appears more then once, the OSPF FAD TLV MUST be ignored by the receiver.

### 7.3. OSPF Flexible Algorithm Include-All Admin Group Sub-TLV

The usage of this Sub-TLVs is described in Section 6.3.

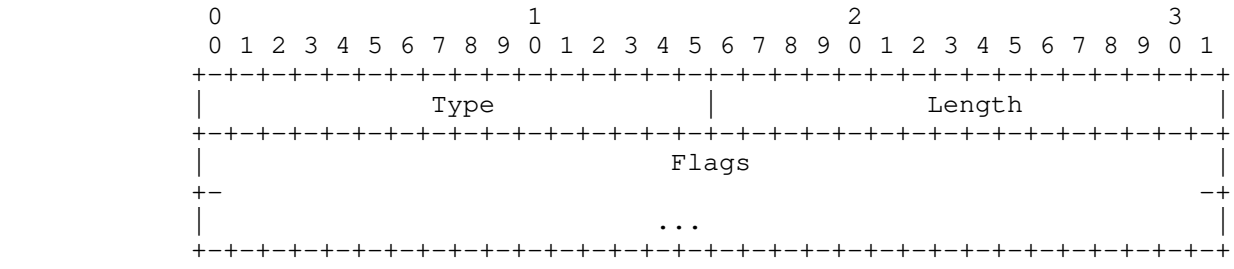
The format of the OSPF Flexible Algorithm Include-Any Admin Group Sub-TLV is identical to the format of the OSPF FAEAG Sub-TLV in Section 7.1.

The OSPF Flexible Algorithm Include-Any Admin Group Sub-TLV Type is 3.

The OSPF Flexible Algorithm Include-All Admin Group Sub-TLV MAY NOT appear more then once in an OSPF FAD TLV. If it appears more then once, the OSPF FAD TLV MUST be ignored by the receiver.

### 7.4. OSPF Flexible Algorithm Definition Flags Sub-TLV

The OSPF Flexible Algorithm Definition Flags Sub-TLV (FADF Sub-TLV) is a Sub-TLV of the OSPF FAD TLV. It has the following format:

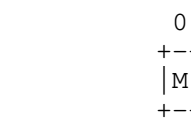



where:

Type: 4

Length: variable, dependent on the size of the Flags field. MUST be a multiple of 4 octets.

Flags:




M-flag: when set, the Flex-Algorithm specific prefix metric MUST be used, if advertised with the prefix. This flag is not applicable to prefixes advertised as SRv6 locators.

Bits are defined/sent starting with Bit 0 defined above. Additional bit definitions that may be defined in the future SHOULD be assigned in ascending bit order so as to minimize the number of bits that will need to be transmitted.

Undefined bits MUST be transmitted as 0.

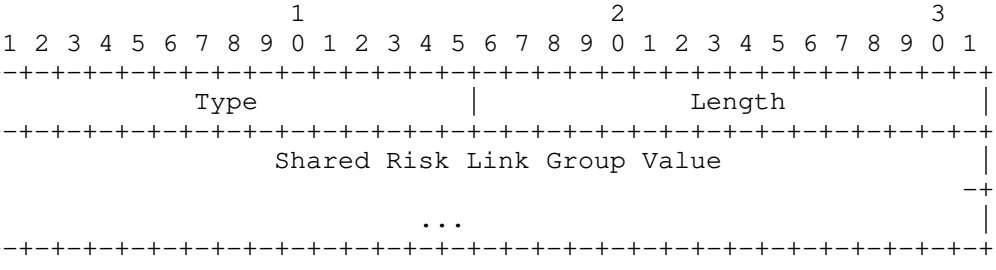
Bits that are NOT transmitted MUST be treated as if they are set to 0 on receipt.

The OSPF FADF Sub-TLV MAY NOT appear more then once in an OSPF FAD TLV. If it appears more then once, the OSPF FAD TLV MUST be ignored by the receiver.

If the OSPF FADF Sub-TLV is not present inside the OSPF FAD TLV, all the bits are assumed to be set to 0.

### 7.5. OSPF Flexible Algorithm Exclude SRLG Sub-TLV

The OSPF Flexible Algorithm Exclude SRLG Sub-TLV (FAESRLG Sub-TLV) is a Sub-TLV of the OSPF FAD TLV. Its usage is described in Section 6.5. It has the following format:




where:

Type: 5

Length: variable, dependent on the number of SRLGs. MUST be a multiple of 4 octets.

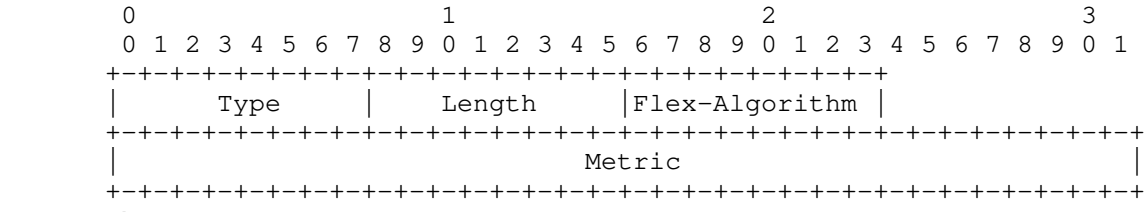
Shared Risk Link Group Value: SRLG value as defined in [RFC4203].

The OSPF FAESRLG Sub-TLV MAY NOT appear more then once in an OSPF FAD TLV. If it appears more then once, the OSPF FAD TLV MUST be ignored by the receiver.

## 8. ISIS Flexible Algorithm Prefix Metric Sub-TLV

The ISIS Flexible Algorithm Prefix Metric (FAPM) Sub-TLV supports the advertisement of a Flex-Algorithm specific prefix metric associated with a given prefix advertisement.

The ISIS FAPM Sub-TLV is a sub-TLV of TLVs 135, 235, 236, and 237 and has the following format:




where:

Type: 6

Length: 5 octets

Flex-Algorithm: Single octet value between 128 and 255 inclusive.

Metric: 4 octets of metric information

The ISIS FAPM Sub-TLV MAY appear multiple times in its parent TLV. If it appears more then once with the same Flex-Algorithm value, the first instance MUST be used and any subsequent instances MUST be ignored.

If a prefix is advertised with a Flex-Algorithm prefix metric larger then MAX_PATH_METRIC as defined in [RFC5305] this prefix MUST NOT be considered during the Flexible-Algorithm computation.

The usage of the Flex-Algorithm prefix metric is described in Section 12.

The ISIS FAPM Sub-TLV MUST NOT be advertised as a sub-TLV of the ISIS SRv6 Locator TLV [I-D.ietf-lsr-isis-srv6-extensions]. The ISIS SRv6 Locator TLV includes the Algorithm and Metric fields which MUST be used instead. If the FAPM Sub-TLV is present as a sub-TLV of the

ISIS SRv6 Locator TLV in the received LSP, such FAPM Sub-TLV MUST be ignored.

## 9. OSPF Flexible Algorithm Prefix Metric Sub-TLV

The OSPF Flexible Algorithm Prefix Metric (FAPM) Sub-TLV supports the advertisement of a Flex-Algorithm specific prefix metric associated with a given prefix advertisement.

The OSPF Flex-Algorithm Prefix Metric (FAPM) Sub-TLVis a Sub-TLV of the:

- OSPFv2 Extended Prefix TLV [RFC7684]

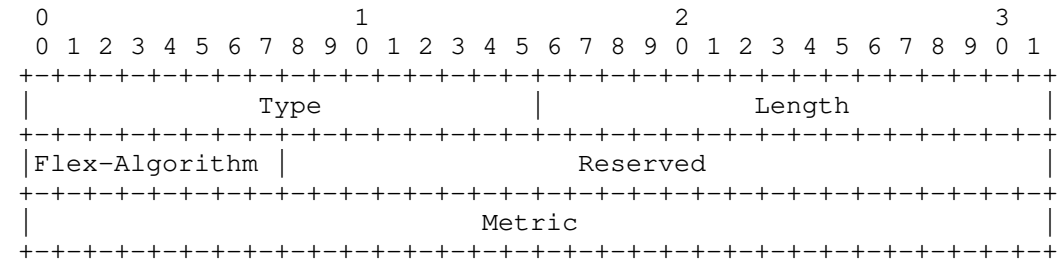
- Following OSPFv3 TLVs as defined in [RFC8362]:

  Intra-Area Prefix TLV

  Inter-Area Prefix TLV

  External Prefix TLV

OSPF FAPM Sub-TLV has the following format:




where:

Type: 3 for OSPFv2, 26 for OSPFv3

Length: 8 octets

Flex-Algorithm: Single octet value between 128 and 255 inclusive.

Reserved: Must be set to 0, ignored at reception.

Metric: 4 octets of metric information

The OSPF FAPM Sub-TLV MAY appear multiple times in its parent TLV. If it appears more then once with the same Flex-Algorithm value, the

first instance MUST be used and any subsequent instances MUST be ignored.

The usage of the Flex-Algorithm prefix metric is described in Section 12.

## 10. Advertisement of Node Participation in a Flex-Algorithm

When a router is configured to support a particular Flex-Algorithm, we say it is participating in that Flex-Algorithm.

Paths computed for a specific Flex-Algorithm MAY be used by various applications, each potentially using its own specific data plane for forwarding traffic over such paths. To guarantee the presence of the application specific forwarding state associated with a particular Flex-Algorithm, a router MUST advertise its participation for a particular Flex-Algorithm for each application specifically.

### 10.1. Advertisement of Node Participation for Segment Routing

[RFC8667], [RFC8665], and [RFC8666] (IGP Segment Routing extensions) describe how the SR-Algorithm is used to compute the IGP best path.

Routers advertise the support for the SR-Algorithm as a node capability as described in the above mentioned IGP Segment Routing extensions. To advertise participation for a particular Flex-Algorithm for Segment Routing, including both SR MPLS and SRv6, the Flex-Algorithm value MUST be advertised in the SR-Algorithm TLV (OSPF) or sub-TLV (ISIS).

Segment Routing Flex-Algorithm participation advertisement is topology independent. When a router advertises participation in an SR-Algorithm, the participation applies to all topologies in which the advertising node participates.

### 10.2. Advertisement of Node Participation for Other Applications

This section describes considerations related to how other applications can advertise their participation in a specific Flex-Algorithm.

Application-specific Flex-Algorithm participation advertisements MAY be topology specific or MAY be topology independent, depending on the application itself.

Application-specific advertisement for Flex-Algorithm participation MUST be defined for each application and is outside of the scope of this document.

## 11. Advertisement of Link Attributes for Flex-Algorithm

Various link attributes may be used during the Flex-Algorithm path calculation. For example, include or exclude rules based on link affinities can be part of the Flex-Algorithm definition as defined in Section 6 and Section 7.

Link attribute advertisements that are to be used during Flex-Algorithm calculation MUST use the Application-Specific Link Attribute (ASLA) advertisements defined in [I-D.ietf-isis-te-app] or [I-D.ietf-ospf-te-link-attr-reuse]. In the case of IS-IS, this includes use of the L-flag as defined in [I-D.ietf-isis-te-app] Section 4.2 subject to the constraints discussed in Section 6 of the [I-D.ietf-isis-te-app]. The mandatory use of ASLA advertisements applies to link attributes specifically mentioned in this document (Min Unidirectional Link Delay, TE Default Metric, Administrative Group, Extended Administrative Group and Shared Risk Link Group) and any other link attributes that may be used in support of Flex-Algorithm in the future.

A new Application Identifier Bit is defined to indicate that the ASLA advertisement is associated with the Flex-Algorithm application. This bit is set in the Standard Application Bit Mask (SABM) defined in [I-D.ietf-isis-te-app] or [I-D.ietf-ospf-te-link-attr-reuse]:

Bit-3: Flexible Algorithm (X-bit)

ASLA Admin Group Advertisements to be used by the Flexible Algorithm Application MAY use either the Administrative Group or Extended Administrative Group encodings. If the Administrative Group encoding is used, then the first 32 bits of the corresponding FAD sub-TLVs are mapped to the link attribute advertisements as specified in RFC 7308.

## 12. Calculation of Flexible Algorithm Paths

A router MUST be configured to participate in a given Flex-Algorithm K and MUST select the FAD based on the rules defined in Section 5.3 before it can compute any path for that Flex-Algorithm.

As described in Section 10, participation for any particular Flex-Algorithm MUST be advertised on a per-application basis. Calculation of the paths for any particular Flex-Algorithm MUST be application specific.

The way applications handle nodes that do not participate in Flexible-Algorithm is application specific. If the application only wants to consider participating nodes during the Flex-Algorithm calculation, then when computing paths for a given Flex-Algorithm,

all nodes that do not advertise participation for that Flex-Algorithm in their application-specific advertisements MUST be pruned from the topology. Segment Routing, including both SR MPLS and SRv6, are applications that MUST use such pruning when computing Flex-Algorithm paths.

When computing the path for a given Flex-Algorithm, the metric-type that is part of the Flex-Algorithm definition (Section 5) MUST be used.

When computing the path for a given Flex-Algorithm, the calculation-type that is part of the Flex-Algorithm definition (Section 5) MUST be used.

Various link include or exclude rules can be part of the Flex-Algorithm definition. To refer to a particular bit within an AG or EAG we uses term 'color'.

Rules, in the order as specified below, MUST be used to prune links from the topology during the Flex-Algorithm computation.

For all links in the topology:

1. Check if any exclude rule is part of the Flex-Algorithm definition. If such exclude rule exists, check if any color that is part of the exclude rule is also set on the link. If such a color is set, the link MUST be pruned from the computation.

2. Check if any exclude SRLG rule is part of the Flex-Algorithm definition. If such exclude rule exists, check if the link is part of any SRLG that is also part of the SRLG exclude rule. If the link is part of such SRLG, the link MUST be pruned from the computation.

3. Check if any include-any rule is part of the Flex-Algorithm definition. If such include-any rule exists, check if any color that is part of the include-any rule is also set on the link. If no such color is set, the link MUST be pruned from the computation.

4. Check if any include-all rule is part of the Flex-Algorithm definition. If such include-all rule exists, check if all colors that are part of the include-all rule are also set on the link. If all such colors are not set on the link, the link MUST be pruned from the computation.

5. If the Flex-Algorithm definition uses other than IGP metric (Section 5), and such metric is not advertised for the particular

link in a topology for which the computation is done, such link MUST be pruned from the computation. A metric of value 0 MUST NOT be assumed in such case.

### 12.1. Multi-area and Multi-domain Considerations

Any IGP Shortest Path Tree calculation is limited to a single area. This applies to Flex-Algorithm calculations as well. Given that the computing router does not have visibility of the topology of the next areas or domain, the Flex-Algorithm specific path to an inter-area or inter-domain prefix will be computed for the local area only. The egress L1/L2 router (ABR in OSPF), or ASBR for inter-domain case, will be selected based on the best path for the given Flex-Algorithm in the local area and such egress ABR or ASBR router will be responsible to compute the best Flex-Algorithm specific path over the next area or domain. This may produce an end-to-end path, which is sub-optimal based on Flex-Algorithm constraints. In cases where the ABR or ASBR has no reachability to a prefix for a given Flex-Algorithm in the next area or domain, the traffic may be dropped by the ABR/ASBR.

To allow the optimal end-to-end path for an inter-area or inter-domain prefix for any Flex-Algorithm to be computed, the FAPM has been defined in Section 8 and Section 9.

If the FAD selected based on the rules defined in Section 5.3 includes the M-flag, an ABR or ASBR MUST include the FAPM (Section 8, Section 9) when advertising the prefix between areas or domains. Such metric will be equal to the metric to reach the prefix for a given Flex-Algorithm in a source area or domain. This is similar in nature to how the metric is set when prefixes are advertised between areas or domains for the default algorithm.

If the FAD selected based on the rules defined in Section 5.3 includes the M-flag, the FAPM MUST be used during calculation of prefix reachability for the inter-area and external prefixes. If the FAPM for the Flex-Algorithm is not advertised with the inter-area or external prefix reachability advertisement, the prefix MUST be considered as unreachable for that Flex-Algorithm.

Flex-Algorithm prefix metrics MUST NOT be used during the Flex-Algorithm computation unless the FAD selected based on the rules defined in Section 5.3 includes the M-Flag, as described in (Section 6.4 or Section 7.4).

If the FAD selected based on the rules defined in Section 5.3 does not includes the M-flag, it is NOT RECOMMENDED to use the Flex-Algorithm for inter-area or inter-domain prefix reachability. The

reason is that without the explicit Flex-Algorithm Prefix Metric advertisement, it is not possible to conclude whether the ABR or ASBR has reachability to the inter-area or inter-domain prefix for a given Flex-Algorithm in the next area or domain. Sending the Flex-Algoritm traffic for such prefix towards the ABR or ASBR may result in traffic looping or black-holing.

The FAPM MUST NOT be advertised with ISIS L1 or L2 intra-area, OSPFv2 intra-area, or OSPFv3 intra-area routes. If the FAPM is advertised for these route-types, it MUST be ignored during the prefix reachability calculation.

The M-flag in FAD is not applicable to prefixes advertised as SRv6 locators. The ISIS SRv6 Locator TLV includes the Algorithm and Metric fields [I-D.ietf-lsr-isis-srv6-extensions]. When the ISIS SRv6 Locator is advertised between areas or domains, the metric field in the Locator TLV MUST be used irrespective of the M-flag in the FAD advertisement.

## 13. Flex-Algorithm and Forwarding Plane

This section describes how Flex-Algorithm paths are used in forwarding.

### 13.1. Segment Routing MPLS Forwarding for Flex-Algorithm

This section describes how Flex-Algorithm paths are used with SR MPLS forwarding.

Prefix SID advertisements include an SR-Algorithm value and, as such, are associated with the specified SR-Algorithm. Prefix-SIDs are also associated with a specific topology which is inherited from the associated prefix reachability advertisement. When the algorithm value advertised is a Flex-Algorithm value, the Prefix SID is associated with paths calculated using that Flex-Algorithm in the associated topology.

A Flex-Algorithm path MUST be installed in the MPLS forwarding plane using the MPLS label that corresponds to the Prefix-SID that was advertised for that Flex-algorithm. If the Prefix SID for a given Flex-algorithm is not known, the Flex-Algorithm specific path cannot be installed in the MPLS forwarding plane.

Traffic that is supposed to be routed via Flex-Algorithm specific paths, MUST be dropped when there are no such paths available.

Loop Free Alternate (LFA) paths for a given Flex-Algorithm MUST be computed using the same constraints as the calculation of the primary

paths for that Flex-Algorithm. LFA paths MUST only use Prefix-SIDs advertised specifically for the given algorithm. LFA paths MUST NOT use an Adjacency-SID that belongs to a link that has been pruned from the Flex-Algorithm computation.

If LFA protection is being used to protect a given Flex-Algorithm paths, all routers in the area participating in the given Flex-Algorithm SHOULD advertise at least one Flex-Algorithm specific Node-SID. These Node-SIDs are used to steer traffic over the LFA computed backup path.

### 13.2. SRv6 Forwarding for Flex-Algorithm

This section describes how Flex-Algorithm paths are used with SRv6 forwarding.

In SRv6 a node is provisioned with topology/algorithm specific locators for each of the topology/algorithm pairs supported by that node. Each locator is an aggregate prefix for all SIDs provisioned on that node which have the matching topology/algorithm.

The SRv6 locator advertisement in IGPs ([I-D.ietf-lsr-isis-srv6-extensions] [I-D.ietf-lsr-ospfv3-srv6-extensions]) includes the MTID value that associates the locator with a specific topology. SRv6 locator advertisements also includes an Algorithm value that explicitly associates the locator with a specific algorithm. When the algorithm value advertised with a locator represents a Flex-Algorithm, the paths to the locator prefix MUST be calculated using the specified Flex-Algorithm in the associated topology.

Forwarding entries for the locator prefixes advertised in IGPs MUST be installed in the forwarding plane of the receiving SRv6 capable routers when the associated topology/algorithm is participating in them. Forwarding entries for locators associated with Flex-Algorithms in which the node is not participating MUST NOT be installed in the forwarding palne.

When the locator is associated with a Flex-Algorithm, LFA paths to the locator prefix MUST be calculated using such Flex-Algorithm in the associated topology, to guarantee that they follow the same constraints as the calculation of the primary paths. LFA paths MUST only use SRv6 SIDs advertised specifically for the given Flex-Algorithm.

If LFA protection is being used to protect locators associated with a given Flex-Algorithm, all routers in the area participating in the given Flex-Algorithm SHOULD advertise at least one Flex-Algorithm

specific locator and END SID per node and one END.X SID for every link that has not been pruned from such Flex-Algorithm computation. These locators and SIDs are used to steer traffic over the LFA-computed backup path.

### 13.3. Other Applications' Forwarding for Flex-Algorithm

Any application that wants to use Flex-Algorithm specific forwarding needs to install some form of Flex-Algorithm specific forwarding entries.

Application-specific forwarding for Flex-Algorithm MUST be defined for each application and is outside of the scope of this document.

## 14. Operational considerations

### 14.1. Inter-area Considerations

The scope of the FA computation is an area, so is the scope of the FAD. In ISIS, the Router Capability TLV in which the FAD Sub-TLV is advertised MUST have the S-bit clear, which prevents it to be flooded outside of the level in which it was originated. Even though in OSPF the FAD Sub-TLV can be flooded in an RI LSA that has AS flooding scope, the FAD selection is performed for each individual area in which it is being used.

There is no requirement for the FAD for a particular Flex-Algorithm to be identical in all areas in the network. For example, traffic for the same Flex-Algorithm may be optimized for minimal delay (e.g., using delay metric) in one area or level, while being optimized for available bandwidth (e.g., using IGP metric) in another area or level.

As described in Section 5.1, ISIS allows the re-generation of the winning FAD from level 2, without any modification to it, into a level 1 area. This allows the operator to configure the FAD in one or multiple routers in the level 2, without the need to repeat the same task in each level 1 area, if the intent is to have the same FAD for the particular Flex-Algorithm across all levels. This can similarly be achieved in OSPF by using the AS flooding scope of the RI LSA in which the FAD Sub-TLV for the particular Flex-Algoritm is advertised.

Re-generation of FAD from a level 1 area to the level 2 area is not supported in ISIS, so if the intent is to regenerate the FAD between ISIS levels, the FAD MUST be defined on router(s) that are in level 2. In OSPF, the FAD definition can be done in any area and be

propagated to all routers in the OSPF routing domain by using the AS flooding scope of the RI LSA.

### 14.2. Usage of SRLG Exclude Rule with Flex-Algorithm

There are two different ways in which SRLG information can be used with Flex-Algorithm:

In a context of a single Flex-Algorithm, it can be used for computation of backup paths, as described in [I-D.ietf-rtgwg-segment-routing-ti-lfa]. This usage does not require association of any specific SRLG constraint with the given Flex-Algorithm definition.

In the context of multiple Flex-Algorithms, it can be used for creating disjoint sets of paths by pruning the links belonging to a specific SRLG from the topology on which a specific Flex-Algorithm computes its paths. This usage:

- Facilitates the usage of already deployed SRLG configurations for setup of disjoint paths between two or more Flex-Algorithms.

- Requires explicit association of a given Flex-Algorithm with a specific set of SRLG constraints as defined in Section 6.5 and Section 7.5.

The two usages mentioned above are orthogonal.

### 14.3. Max-metric consideration

Both ISIS and OSPF have a mechanism to set the IGP metric on a link to a value that would make the link either non-reachable or to serve as the link of last resort. Similar functionality would be needed for the Min Unidirectional Link Delay and TE metric, as these can be used to compute Flex-Algorithm paths.

The link can be made un-reachable for all Flex-Algorithms that use Min Unidirectional Link Delay as metric, as described in Section 5.1, by removing the Flex-Algorithm ASLA Min Unidirectional Link Delay advertisement for the link. The link can be made the link of last resort by setting the delay value in the Flex-Algorithm ASLA delay advertisement for the link to the value of 16,777,215 (2^24 - 1).

The link can be made un-reachable for all Flex-Algorithms that use TE metric, as described in Section 5.1, by removing the Flex-Algorithm ASLA TE metric advertisement for the link. The link can be made the link of last resort by setting the TE metric value in the Flex-

Algorithm ASLA delay advertisement for the link to the value of (2^24 - 1) in ISIS and (2^32 - 1) in OSPF.

## 15. Backward Compatibility

This extension brings no new backward compatibility issues.

## 16. Security Considerations

This draft adds two new ways to disrupt IGP networks:

An attacker can hijack a particular Flex-Algorithm by advertising a FAD with a priority of 255 (or any priority higher than that of the legitimate nodes).

An attacker could make it look like a router supports a particular Flex-Algorithm when it actually doesn't, or vice versa.

Both of these attacks can be addressed by the existing security extensions as described in [RFC5304] and [RFC5310] for ISIS, in [RFC2328] and [RFC7474] for OSPFv2, and in [RFC5340] and [RFC4552] for OSPFv3.

## 17. IANA Considerations

### 17.1. IGP IANA Considerations

#### 17.1.1. IGP Algorithm Types Registry

This document makes the following registrations in the "IGP Algorithm Types" registry:

Type: 128-255.

Description: Flexible Algorithms.

Reference: This document (Section 4).

#### 17.1.2. IGP Metric-Type Registry

IANA is requested to set up a registry called "IGP Metric-Type Registry" under a "Interior Gateway Protocol (IGP) Parameters" IANA registries. The registration policy for this registry is "Standards Action" ([RFC8126] and [RFC7120]).

Values in this registry come from the range 0-255.

This document registers following values in the "IGP Metric-Type Registry":

Type: 0

Description: IGP metric

Reference: This document (Section 5.1)

Type: 1

Description: Min Unidirectional Link Delay as defined in [RFC8570], section 4.2, and [RFC7471], section 4.2.

Reference: This document (Section 5.1)

Type: 2

Description: Traffic Engineering Default Metric as defined in [RFC5305], section 3.7, and Traffic engineering metric as defined in [RFC3630], section 2.5.5

Reference: This document (Section 5.1)

### 17.2. Flexible Algorithm Definition Flags Registry

IANA is requested to set up a registry called "ISIS Flexible Algorithm Definition Flags Registry" under a "Interior Gateway Protocol (IGP) Parameters" IANA registries. The registration policy for this registry is "Standards Action" ([RFC8126] and [RFC7120]).

This document defines the following single bit in Flexible Algorithm Definition Flags registry:

| Bit # | Name |
|---|---|
| 0 | Prefix Metric Flag (M-flag) |

Reference: This document (Section 6.4, Section 7.4).

### 17.3. ISIS IANA Considerations

#### 17.3.1. Sub TLVs for Type 242

This document makes the following registrations in the "sub-TLVs for TLV 242" registry.

Type: 26.

Description: Flexible Algorithm Definition.

Reference: This document (Section 5.1).

#### 17.3.2. Sub TLVs for for TLVs 135, 235, 236, and 237

This document makes the following registrations in the "Sub-TLVs for for TLVs 135, 235, 236, and 237" registry.

Type: 6

Description: Flexible Algorithm Prefix Metric.

Reference: This document (Section 8).

#### 17.3.3. Sub-Sub-TLVs for Flexible Algorithm Definition Sub-TLV

This document creates the following Sub-Sub-TLV Registry:

Registry: Sub-Sub-TLVs for Flexible Algorithm Definition Sub-TLV

Registration Procedure: Expert review

Reference: This document (Section 5.1)

This document defines the following Sub-Sub-TLVs in the "Sub-Sub-TLVs for Flexible Algorithm Definition Sub-TLV" registry:

Type: 1

Description: Flexible Algorithm Exclude Admin Group

Reference: This document (Section 6.1).

Type: 2

Description: Flexible Algorithm Include-Any Admin Group

Reference: This document (Section 6.2).

Type: 3

Description: Flexible Algorithm Include-All Admin Group

Reference: This document (Section 6.3).

Type: 4

Description: Flexible Algorithm Definition Flags

Reference: This document (Section 6.4).

Type: 5

Description: Flexible Algorithm Exclude SRLG

Reference: This document (Section 6.5).

### 17.4. OSPF IANA Considerations

#### 17.4.1. OSPF Router Information (RI) TLVs Registry

This specification updates the OSPF Router Information (RI) TLVs Registry.

Type: 16

Description: Flexible Algorithm Definition TLV.

Reference: This document (Section 5.2).

#### 17.4.2. OSPFv2 Extended Prefix TLV Sub-TLVs

This document makes the following registrations in the "OSPFv2 Extended Prefix TLV Sub-TLVs" registry.

Type: 3

Description: Flexible Algorithm Prefix Metric.

Reference: This document (Section 9).

#### 17.4.3. OSPFv3 Extended-LSA Sub-TLVs

This document makes the following registrations in the "OSPFv3 Extended-LSA Sub-TLVs" registry.

Type: 26

Description: Flexible Algorithm Prefix Metric.

Reference: This document (Section 9).

#### 17.4.4. OSPF Flexible Algorithm Definition TLV Sub-TLV Registry

This document creates the following registry:

Registry: OSPF Flexible Algorithm Definition TLV sub-TLV

Registration Procedure: Expert review

Reference: This document (Section 5.2)

The "OSPF Flexible Algorithm Definition TLV sub-TLV" registry will define sub-TLVs at any level of nesting for the Flexible Algorithm TLV and should be added to the "Open Shortest Path First (OSPF) Parameters" registries group. New values can be allocated via IETF Review or IESG Approval.

This document registers following Sub-TLVs in the "TLVs for Flexible Algorithm Definition TLV" registry:

Type: 1

Description: Flexible Algorithm Exclude Admin Group

Reference: This document (Section 7.1).

Type: 2

Description: Flexible Algorithm Include-Any Admin Group

Reference: This document (Section 7.2).

Type: 3

Description: Flexible Algorithm Include-All Admin Group

Reference: This document (Section 7.3).

Type: 4

Description: Flexible Algorithm Definition Flags

Reference: This document (Section 7.4).

Type: 5

Description: Flexible Algorithm Exclude SRLG

Reference: This document (Section 7.5).

Types in the range 32768-33023 are for experimental use; these will not be registered with IANA, and MUST NOT be mentioned by RFCs.

Types in the range 33024-65535 are not to be assigned at this time. Before any assignments can be made in the 33024-65535 range, there MUST be an IETF specification that specifies IANA Considerations that covers the range being assigned.

#### 17.4.5. Link Attribute Applications Registry

This document registers following bit in the Link Attribute Applications Registry:

Bit-3

Description: Flexible Algorithm (X-bit)

Reference: This document (Section 11).

## 18. Acknowledgements

This draft, among other things, is also addressing the problem that the [I-D.gulkohegde-routing-planes-using-sr] was trying to solve. All authors of that draft agreed to join this draft.

Thanks to Eric Rosen, Tony Przygienda for their detailed review and excellent comments.

Thanks to Cengiz Halit for his review and feedback during initial phase of the solution definition.

Thanks to Kenji Kumaki for his comments.

Thanks to William Britto A J. for his suggestions.

Thanks to Acee Lindem for editorial comments.

## 19. References

### 19.1. Normative References

[BCP14] , <https://tools.ietf.org/html/bcp14>.

[I-D.ietf-isis-te-app]
Ginsberg, L., Psenak, P., Previdi, S., Henderickx, W., and J. Drake, "IS-IS Application-Specific Link Attributes", draft-ietf-isis-te-app-19 (work in progress), June 2020.

[I-D.ietf-lsr-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extension to Support Segment Routing over IPv6 Dataplane", draft-ietf-lsr-isis-srv6-extensions-09 (work in progress), September 2020.

[I-D.ietf-lsr-ospf-reverse-metric]
Talaulikar, K., Psenak, P., and H. Johnston, "OSPF Reverse Metric", draft-ietf-lsr-ospf-reverse-metric-01 (work in progress), June 2020.

[I-D.ietf-lsr-ospfv3-srv6-extensions]
Li, Z., Hu, Z., Cheng, D., Talaulikar, K., and P. Psenak, "OSPFv3 Extensions for SRv6", draft-ietf-lsr-ospfv3-srv6-extensions-01 (work in progress), August 2020.

[I-D.ietf-ospf-te-link-attr-reuse]
Psenak, P., Ginsberg, L., Henderickx, W., Tantsura, J., and J. Drake, "OSPF Application-Specific Link Attributes", draft-ietf-ospf-te-link-attr-reuse-16 (work in progress), June 2020.

[ISO10589]
International Organization for Standardization, "Intermediate system to Intermediate system intra-domain routeing 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>.

[RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, DOI 10.17487/RFC4203, October 2005, <https://www.rfc-editor.org/info/rfc4203>.

[RFC5307] Kompella, K., Ed. and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, DOI 10.17487/RFC5307, October 2008, <https://www.rfc-editor.org/info/rfc5307>.

[RFC7308] Osborne, E., "Extended Administrative Groups in MPLS Traffic Engineering (MPLS-TE)", RFC 7308, DOI 10.17487/RFC7308, July 2014, <https://www.rfc-editor.org/info/rfc7308>.

[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>.

[RFC7770] Lindem, A., Ed., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 7770, DOI 10.17487/RFC7770, February 2016, <https://www.rfc-editor.org/info/rfc7770>.

[RFC7981] Ginsberg, L., Previdi, S., and M. Chen, "IS-IS Extensions for Advertising Router Information", RFC 7981, DOI 10.17487/RFC7981, October 2016, <https://www.rfc-editor.org/info/rfc7981>.

[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>.

[RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <https://www.rfc-editor.org/info/rfc8665>.

[RFC8666] Psenak, P., Ed. and S. Previdi, Ed., "OSPFv3 Extensions for Segment Routing", RFC 8666, DOI 10.17487/RFC8666, December 2019, <https://www.rfc-editor.org/info/rfc8666>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

### 19.2. Informative References

[I-D.gulkohegde-routing-planes-using-sr]
Hegde, S. and a. arkadiy.gulko@thomsonreuters.com, "Separating Routing Planes using Segment Routing", draft-gulkohegde-routing-planes-using-sr-00 (work in progress), March 2017.

[I-D.ietf-rtgwg-segment-routing-ti-lfa]
Litkowski, S., Bashandy, A., Filsfils, C., Decraene, B., Francois, P., Voyer, D., Clad, F., and P. Camarillo, "Topology Independent Fast Reroute using Segment Routing", draft-ietf-rtgwg-segment-routing-ti-lfa-04 (work in progress), August 2020.

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

[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>.

[RFC3906] Shen, N. and H. Smit, "Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels", RFC 3906, DOI 10.17487/RFC3906, October 2004, <https://www.rfc-editor.org/info/rfc3906>.

[RFC4552] Gupta, M. and N. Melam, "Authentication/Confidentiality for OSPFv3", RFC 4552, DOI 10.17487/RFC4552, June 2006, <https://www.rfc-editor.org/info/rfc4552>.

[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>.

[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>.

[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>.

[RFC7120] Cotton, M., "Early IANA Allocation of Standards Track Code Points", BCP 100, RFC 7120, DOI 10.17487/RFC7120, January 2014, <https://www.rfc-editor.org/info/rfc7120>.

[RFC7471] Giacalone, S., Ward, D., Drake, J., Atlas, A., and S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions", RFC 7471, DOI 10.17487/RFC7471, March 2015, <https://www.rfc-editor.org/info/rfc7471>.

[RFC7474] Bhatia, M., Hartman, S., Zhang, D., and A. Lindem, Ed., "Security Extension for OSPFv2 When Using Manual Key Management", RFC 7474, DOI 10.17487/RFC7474, April 2015, <https://www.rfc-editor.org/info/rfc7474>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[RFC8570] Ginsberg, L., Ed., Previdi, S., Ed., Giacalone, S., Ward, D., Drake, J., and Q. Wu, "IS-IS Traffic Engineering (TE) Metric Extensions", RFC 8570, DOI 10.17487/RFC8570, March 2019, <https://www.rfc-editor.org/info/rfc8570>.

## Authors' Addresses

Peter Psenak (editor)
Cisco Systems
Apollo Business Center
Mlynske nivy 43
Bratislava, 82109
Slovakia

Email: ppsenak@cisco.com

Shraddha Hegde
Juniper Networks, Inc.
Embassy Business Park
Bangalore, KA, 560093
India

Email: shraddha@juniper.net

Clarence Filsfils
Cisco Systems, Inc.
Brussels
Belgium

Email: cfilsfil@cisco.com

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

Email: ketant@cisco.com

Arkadiy Gulko
Individual

Email: arkadiy.gulko@gmail.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 8, 2020

H. Chen
Futurewei
M. Toy
Verizon
Y. Yang
IBM
A. Wang
China Telecom
X. Liu
Volta Networks
Y. Fan
Casa Systems
L. Liu
Fujitsu
June 6, 2020

# Flooding Topology Minimum Degree Algorithm
# draft-ietf-lsr-flooding-topo-min-degree-00

## Abstract

This document proposes an algorithm for a node to compute a flooding topology, which is a subgraph of the complete topology per underline physical network. When every node in an area automatically calculates a flooding topology by using a same algorithm and floods the link states using the flooding topology, the amount of flooding traffic in the network is greatly reduced. This would reduce convergence time with a more stable and optimized routing environment.

## 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].

## 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/.

Table of Contents



## 1. Introduction

For some networks such as dense Data Center (DC) networks, the existing Link State (LS) flooding mechanism is not efficient and may have some issues. The extra LS flooding consumes network bandwidth. Processing the extra LS flooding, including receiving, buffering and decoding the extra LSs, wastes memory space and processor time. This

may cause scalability issues and affect the network convergence negatively.

This document proposes an algorithm for a node to compute a flooding topology, which is a subgraph of the complete topology per underline physical network. The physical network can be any network, including clos leaf spine network. It can be used in the distributed mode of flooding topology computation for flooding reduction and the centralized mode, which are described in [I-D.ietf-lsr-dynamic-flooding]. When the distributed mode is selected, every node in an area automatically calculates a flooding topology by using a same algorithm and floods the link states using the flooding topology, the amount of flooding traffic in the network is greatly reduced. This would reduce convergence time with a more stable and optimized routing environment.

There may be multiple algorithms for computing a flooding topology. Users can select one they prefer, and smoothly switch from one to another.

## 2. Terminology

LSA: A Link State Advertisement in OSPF.

LSP: A Link State Protocol Data Unit (PDU) in IS-IS.

LS: A Link Sate, which is an LSA or LSP.

FT: Flooding Topology.

FTC: Flooding Topology Computation.

## 3. Flooding Topology

For a given network topology, a flooding topology is a sub-graph or sub-network of the given network topology that has the same reachability to every node as the given network topology. Thus all the nodes in the given network topology MUST be in the flooding topology. All the nodes MUST be inter-connected directly or indirectly. As a result, LS flooding will in most cases occur only on the flooding topology, that includes all nodes but a subset of links. Note even though the flooding topology is a sub-graph of the original topology, any single LS MUST still be disseminated in the entire network.

### 3.1. Flooding Topology Construction

Many different flooding topologies can be constructed for a given network topology. For example, a chain connecting all the nodes in the given network topology is a flooding topology. A circle connecting all the nodes is another flooding topology. A tree connecting all the nodes is a flooding topology. In addition, the tree plus the connections between some leaves of the tree and branch nodes of the tree is a flooding topology.

The following parameters need to be considered for constructing a flooding topology:

- Degree: The degree of the flooding topology is the maximum degree among the degrees of the nodes on the flooding topology. The degree of a node on the flooding topology is the number of connections on the flooding topology it has to other nodes.

- Number of links: The number of links on the flooding topology is a key factor for reducing the amount of LS flooding. In general, the smaller the number of links, the less the amount of LS flooding.

- Diameter: The diameter of the flooding topology is the shortest distance between the two most distant nodes on the flooding topology. It is a key factor for reducing the network convergence time. The smaller the diameter, the less the convergence time.

- Redundancy: The redundancy of the flooding topology means a tolerance to the failures of some links and nodes on the flooding topology. If the flooding topology is split by some failures, it is not tolerant to these failures. In general, the larger the number of links on the flooding topology is, the more tolerant the flooding topology to failures.

Note that the flooding topology constructed by a node is dynamic in nature, that means when the base topology (the entire topology graph) changes, the flooding topology (the sub-graph) MUST be re-computed/ re-constructed to ensure that any node that is reachable on the base topology MUST also be reachable on the flooding topology.

## 4. Algorithms to Compute Flooding Topology

There are many algorithms to compute a flooding topology. A simple and efficient one is briefed, which comprises:

- Selecting a node R0 with the smallest node ID;

- Building a tree using R0 as root in breadth first; and then
- Connecting each node whose degree is one to another node to have a flooding topology.

### 4.1. Algorithm with Considering Degree

The algorithm is described below, where a variable MaxD with an initial value 3, data structures candidate queue Cq and flooding topology FT are used. Cq and FT comprise elements of form (N, D, PHs), where N represents a Node, D is the Degree of node N, and PHs contains the Previous Hops of node N. The detailed FT computation by the algorithm is illustrated in Appendix A through an example.

The algorithm starts from node R0 as root with a maximum degree MaxD of value 3, a candidate queue Cq = {(R0, D = 0, PHs = { })}, and an empty flooding topology FT = { }. Cq contains one element (R0, D = 0, PHs = { }), where node R0 is the root, D = 0 indicates that the Degree (D for short) of R0 is 0 (i.e., the number of links on the flooding topology connected to R0 is 0), PHs = { } indicates that the Previous Hops (PHs for short) of R0 is empty.

1. Finding and removing the first element with node A in Cq that is not on FT and one PH's D in PHs < MaxD.

   If A is root R0, then add the element into FT

   otherwise (i.e., A != R0 with one PH's D in PHs < MaxD. Assume that PH is the first one in PHs whose D < MaxD), PH's D++, and add A with D = 1 and PHs = {PH} into FT.

   Note: if no element in Cq satisfies the conditions, algorithm is restarted from R0, ++MaxD, Cq = {(R0,D=0,PHs={ })}, FT = { };

2. If all the nodes are on the FT, then goto step 4;

3. Suppose that node Xi (i = 1, 2,..., n) is connected to node A and not on FT, and X1, X2,..., Xn are in an increasing order by their IDs (i.e., X1's ID < X2's ID < ... < Xn's ID). If Xi is not in Cq, then add it into the end of Cq with D = 0 and PHs = {A}; otherwise (i.e., Xi is in Cq), add A into the end of Xi's PHs; Goto step 1.

4. For each node B on FT whose D is one (from minimum to maximum node ID), find a link L attached to B such that L's remote node R has minimum D and ID, add link L between B and R into FT and increase B's D and R's D by one. Return FT.

### 4.2. Algorithm with Considering Others

There may be some constraints on some nodes in a network. For example, in a spine-and-leaf network, there may be a constraint on the degree of every leaf node on the flooding topology, which is that the degree of every leaf node is not greater than a given number ConMaxD of value 2. For each of the other nodes such as the spine nodes, there is no such constraint, that is that ConMaxD is a huge number for each of these nodes.

Step 1 of the algorithm described above is updated below to consider this constraint. In addition to checking constraint PH's D < MaxD, step 1 checks another constraint PH's D < PH's ConMaxD.

1. Finding and removing the first element with node A in Cq that is not on FT and one PH's D in PHs < MaxD and PH's D < PH's ConMaxD.

   If A is root R0, then add the element into FT

   otherwise (i.e., A != R0 with one PH's D in PHs < MaxD and PH's D < PH's ConMaxD. Assume that PH is the first one in PHs whose D < MaxD and PH's D < PH's ConMaxD), PH's D++, and add A with D = 1 and PHs = {PH} into FT.

   Note: if no element in Cq satisfies the conditions, algorithm is restarted from R0, ++MaxD, Cq = {(R0,D=0,PHs={ })}, FT = { };

## 5. Security Considerations

This document does not introduce any new security issue.

## 6. IANA Considerations

Under Registry Name: "IGP Algorithm Type For Computing Flooding Topology" under an existing "Interior Gateway Protocol (IGP) Parameters" IANA registries (refer to Section 7.3. IGP [I-D.ietf-lsr-dynamic-flooding]), IANA is requested to assign one value of IGP Algorithm Type For Computing Flooding Topology as follows:

| Type Value | Type Name | reference |
|---|---|---|
| 1 | Breadth First Minimum Degree Algorithm | This document |
| 2 | Breadth First Leaf Constraint Algorithm | This document |

## 7. Acknowledgements

The authors would like to thank Dean Cheng, Acee Lindem, Zhibo Hu, Robin Li, Stephane Litkowski and Alvaro Retana for their valuable suggestions and comments on this draft.

## 8. References

### 8.1. Normative References

[I-D.ietf-lsr-dynamic-flooding]
Li, T., Psenak, P., Ginsberg, L., Chen, H., Przygienda, T., Cooper, D., Jalil, L., and S. Dontula, "Dynamic Flooding on Dense Graphs", draft-ietf-lsr-dynamic-flooding-06 (work in progress), May 2020.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, DOI 10.17487/RFC1195, December 1990, <https://www.rfc-editor.org/info/rfc1195>.

[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>.

### 8.2. Informative References

[I-D.ietf-rtgwg-spf-uloop-pb-statement]
Litkowski, S., Decraene, B., and M. Horneffer, "Link State protocols SPF trigger and delay algorithm impact on IGP micro-loops", draft-ietf-rtgwg-spf-uloop-pb-statement-10 (work in progress), January 2019.

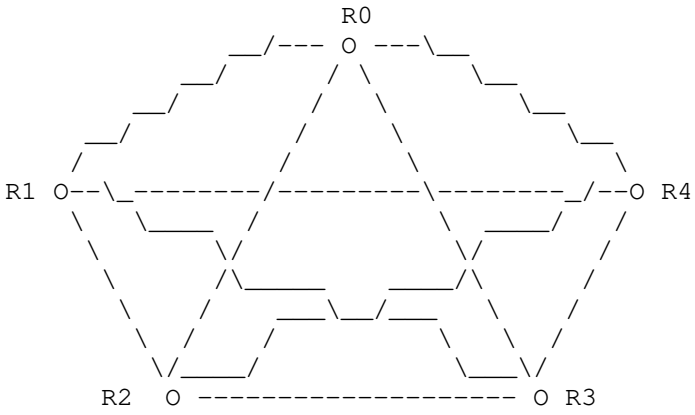
[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

## Appendix A. FT Computation Details through Example

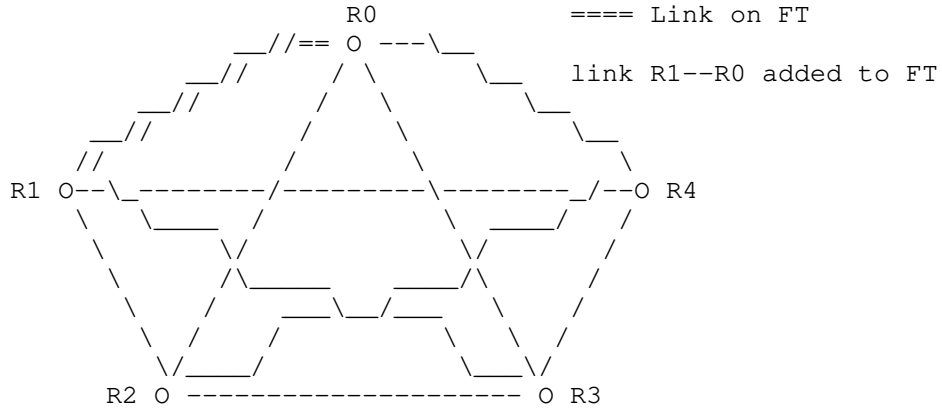
This section presents the details on FT computation by the algorithm through an example. The detailed procedure of computing a FT for a network of five nodes with full mess connections is illustrated. Suppose that the network has five nodes R0, R1, R2, R3 and R4; R0's

ID < R1's ID < R2's ID < R3's ID < R4's ID. The algorithm starts with Cq = {(R0, D=0, PH={})}, FT = {}, MaxD = 3.

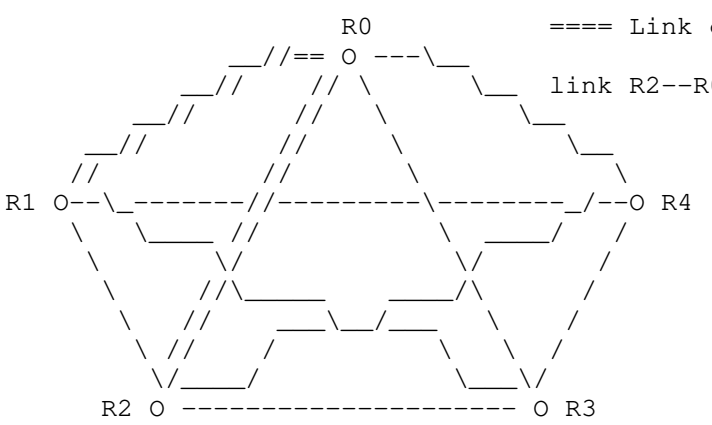
0. // remove the first element containing root R0 from Cq
   Cq = { };
   // add the element into FT
   FT = { (R0,D=0,PHs={ }) }; // root R0 on FT
   // for each Ri connected to R0 (not in Cq), add it to the end of Cq
   Cq = { (R1,D=0,PHs={R0}), (R2,D=0,PHs={R0}), (R3,D=0,PHs={R0}),
          ^^^^^^^^^^^^^^^^^  (R4,D=0,PHs={R0}) }



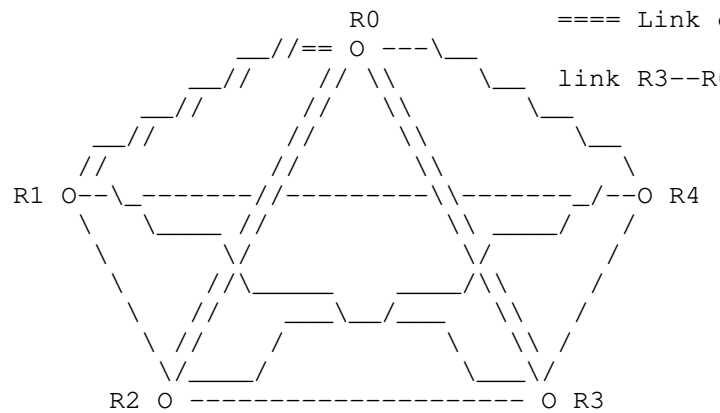

1. //remove first element (R1,D=0,PHs={R0}) from Cq, R0's D=0 < MaxD
   Cq = { (R2,0,{R0}), (R3,0,{R0}), (R4,0,{R0}) };
   // add (R1,1,{R0}) into FT, increase PH R0's D by one
   FT = { (R0,1, { }), (R1,1, {R0}) }; // Link R1--R0 on FT
           ^^^       ^^^^^^^^^^^^
   // for Ri connected to R1 (in Cq) not on FT, append R1 to Ri's PHs
   Cq = { (R2,0, {R0,R1}), (R3,0, {R0,R1}), (R4,0,{R0,R1}) }.
                  ^^              ^^              ^^



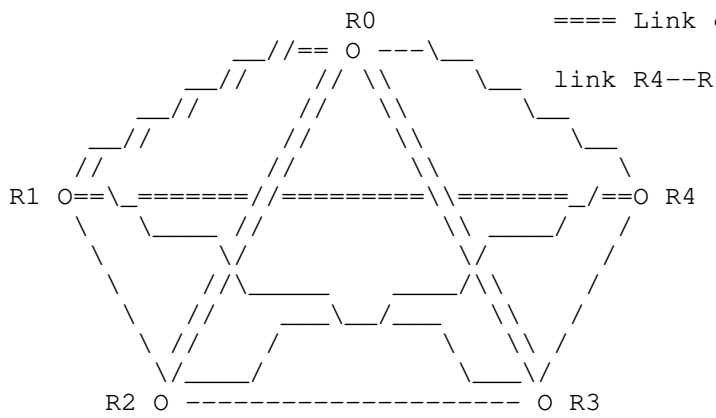
2. // remove the first element (R2,0, {R0,R1}) from Cq, R0's D=1 < MaxD
   Cq = { (R3,0, {R0,R1}), (R4,0,{R0,R1}) }
   // add (R2,1,{R0}) into FT, increase R0's D by one
   FT = { (R0,2,{ }), (R1,1,{R0}), (R2,1,{R0}) } //Link R2--R0 on FT
   // for Ri connected to R2 (in Cq) not on FT, append R2 to Ri's PHs
   Cq = { (R3,0, {R0,R1,R2}), (R4,0,{R0,R1,R2}) }




3. //remove the 1st element (R3,0,{R0,R1,R2}) from Cq, R0's D=2 < MaxD
   Cq = { (R4,0,{R0,R1,R2}) }
   // add (R3,1,{R0}) into FT, increase R0's D by one
   FT = { (R0,3,{}), (R1,1,{R0}), (R2,1,{R0}), (R3,1,{R0}) }
   // for Ri connected to R3 (in Cq) not on FT, append R3 to Ri's PHs
   Cq = { (R4,0,{R0,R1,R2,R3}) }.

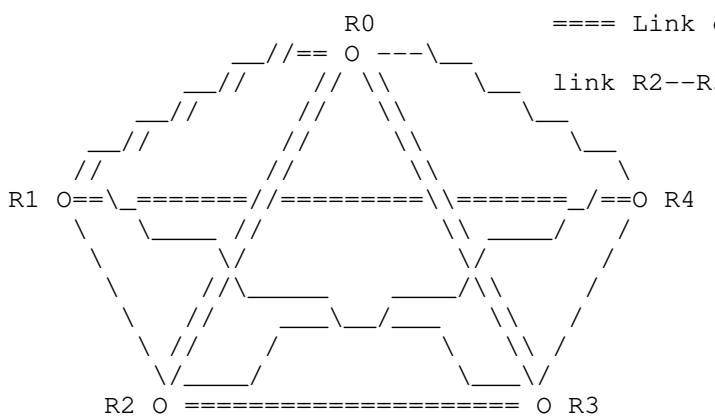


4. //remove the 1st element (R4,0,{R0,R1,R2,R3}) from Cq,R1's D=1 < MaxD
   Cq = { }
   // add (R4,1,{R1}) into FT, increase R1's D by one
   FT = {(R0,3,{}), (R1,2,{R0}), (R2,1,{R0}), (R3,1,{R0}), (R4,1,{R1})}
                    ^^^                                  ^^^^^^^^^^^^

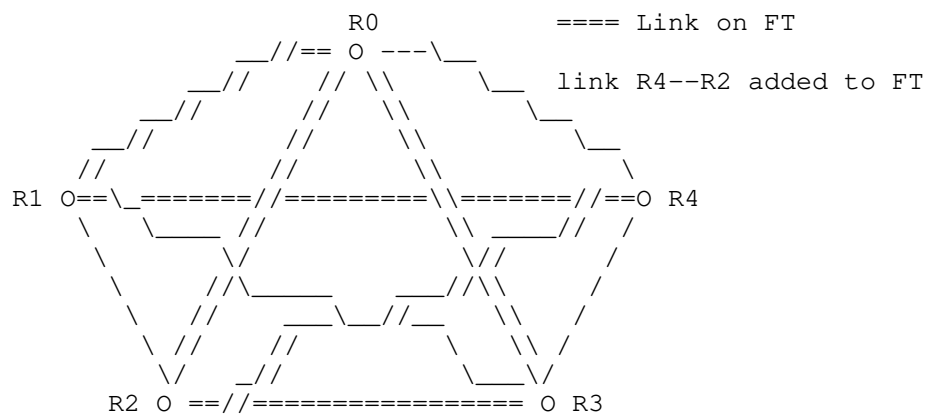



   All nodes are on FT now. In the following, for each node on FT whose D = 1 (from minimum to maximum ID), link L attached to it and not on FT is found such that L's remote node has minimum D and ID. L is added into FT.

5. // On FT, get node R2 with smallest ID whose D=1
   FT = {(R0,3,{}),(R1,2,{R0}),(R2,1,{R0}),(R3,1,{R0}), (R4,1,{R1})}
   // Add link R2--R3 to FT,   ^^^^^^^^^^^^
   // where R2--R3 is not on FT, R3's D=1 is minimum first and then
   // R3's ID is minimum (R3 and R4 tie for D), R2's D++ and R3's D++
   FT = {(R0,3,{}),(R1,2,{R0}),(R2,2,{R0,R3}),(R3,2,{R0}),(R4,1,{R1})}
                               ^^^    ^^      ^^^



6. // On FT, get node R4 with smallest ID whose D=1
   FT = {(R0,3,{}),(R1,2,{R0}),(R2,2,{R0,R3}),(R3,2,{R0}),(R4,1,{R1})}
   // Add link R4--R2 to FT, where ^^^^^^^^^^^
   // R4--R2 is not on FT, R2's D=2 is minimum first and then R2's ID is
   // minimum (R2 and R3 tie for D), increase R2's D and R4's D by one
   FT = {(R0,3,{}),(R1,2,{R0}),(R2,3,{R0,R3}),(R3,2,{R0}),(R4,2,{R1,R2})}
   ^^^ ^^^ ^^




   FT is computed, which has Degree of 3 and Diameter of 2.

Authors' Addresses

Huaimo Chen
Futurewei
Boston
USA

Email: huaimo.chen@futurewei.com

Mehmet Toy
Verizon
USA

Email: mehmet.toy@verizon.com

Yi Yang
IBM
Cary, NC
United States of America

Email: yyietf@gmail.com

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

Email: wangaj3@chinatelecom.cn

Xufeng Liu
Volta Networks
McLean, VA
USA

Email: xufeng.liu.ietf@gmail.com

Yanhe Fan
Casa Systems
USA

Email: yfan@casa-systems.com

Lei Liu
Fujitsu
USA

Email: liulei.kddi@gmail.com

LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 1, 2021

A. Wang
China Telecom
A. Lindem
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.

## Copyright Notice

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:

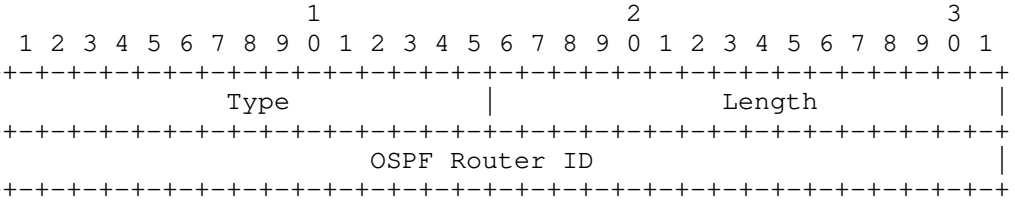



Figure 1: Prefix Source Router-ID Sub-TLV Format

Where:

- Type: 4 for OSPFv2 and 27 for OSPFv3
- Length: 4
- 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:

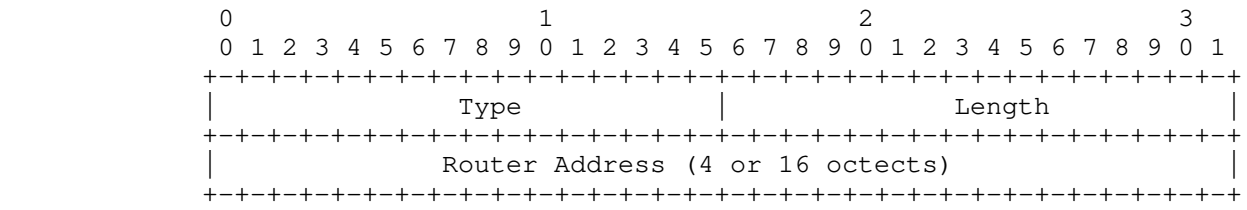



Figure 2: Prefix Originator Sub-TLV Format

Where:

- Type: TBD1 for OSPFv2 and TBD2 for OSPFv3
- Length: 4 or 16
- 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 Point | Description | IANA Allocation 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 Point | Description | IANA Allocation 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.

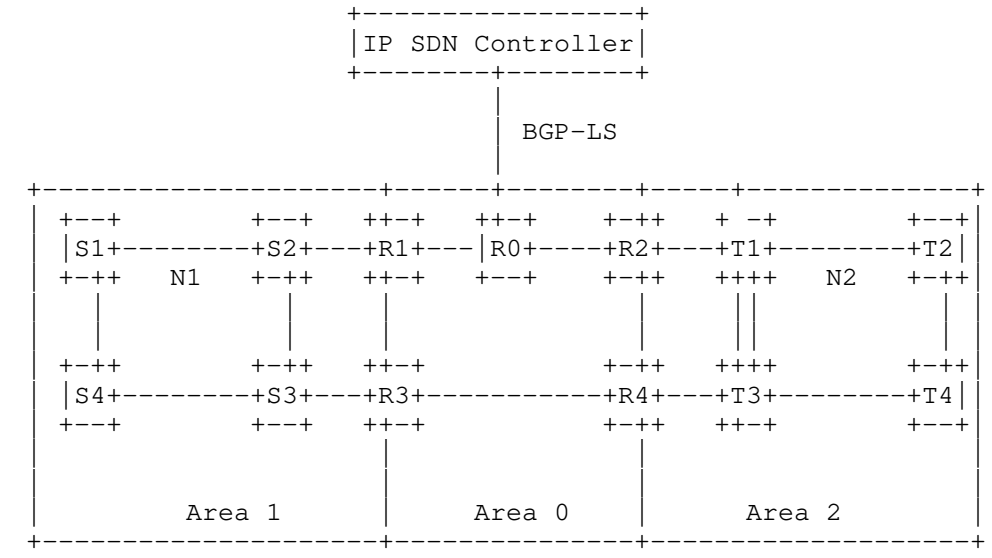



Figure 5: OSPF Inter-Area Prefix Originator Scenario

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

LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 11, 2021

Yao. Liu
Zheng. Zhang
ZTE Corporation

July 10, 2020

# IGP Extensions for Segment Routing Service Segment
## draft-lz-lsr-igp-sr-service-segments-02

Abstract

This document defines extensions to the link-state routing protocols (IS-IS and OSPF) in order to carry service segment information via IGP.

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 11, 2021.

Table of Contents



## 1. Introduction

Segments are introduced in the SR architecture [RFC8402]. Segment Routing (SR) allows for a flexible definition of end-to-end paths by encoding paths as sequences of topological sub-paths, called "segments".

Service Function Chaining (SFC) [RFC7665] provides support for the creation of composite services that consist of an ordered set of Service Functions (SF) that are to be applied to packets and/or frames selected as a result of classification.

[I-D.ietf-spring-sr-service-programming] describes how a service can be associated with a SID and how to achieve service funtion chaining in SR-enabled MPLS and IPv6 networks. It also defines SR-aware and SR-unaware services. For a SR-unaware service ,there has to be a SR proxy handling the SR processing on behalf of the service .

[I-D.dawra-idr-bgp-ls-sr-service-segments] propose extensions to BGP-LS for Service Chaining to distribute the service segment information to SR Controller.

The network topology is shown in figure 1.

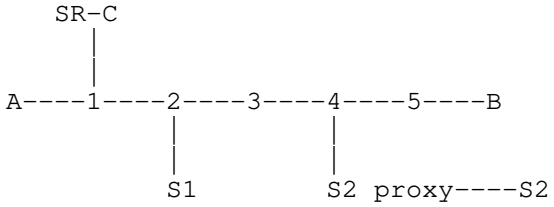



Figure 1: Network with Services

Node 1-5 are nodes capital of segment routing. A and B are two end hosts. S1 is an SR-aware Service. S2 is an SR-unaware Service.

SR Controller (SR-C) is connected to node 1, but may be attached to any node 1-5 in the network.

SR-C is capable of receiving BGP-LS updates to discover topology, and calculating constrained paths between 1 and 5.

Node 1 can use the BGP-LS extensions [I-D.ietf-spring-sr-service-programming] to advertise the service segment information to the SR-C, but it must get the information from other nodes at first.

This document proposes extensions for IGP to advertise service segment information so that there is only one SR node needed per Autonomous System to be connected with the SR-C through BGP-LS to advertise the information to it.

This extension works for both SR-MPLS and SRv6.

## 2. IGP Extensions for Service Segments

After an SFF like node 2 or node 4 get the service segment information, it needs to advertise the information to other SR nodes in the domain through IGP.

How can SFFs like node 2 and node 4 get the service segment information from S1 and S2 proxy will be discussed further.

There may be other alternate mechanisms and are outside of scope of this document.

### 2.1. IS-IS Extensions

This document introduces new sub-sub-TLVs for SRv6 End SID sub-TLV [I-D.ietf-lsr-isis-srv6-extensions] and Prefix Segment Identifier (Prefix-SID) Sub-TLV [RFC8667] for SR-MPLS to associate the Service SID Value with Service-related Information.

One of the new TLVs is Service Chaining (SC) TLV, the TLV is defined as follows :

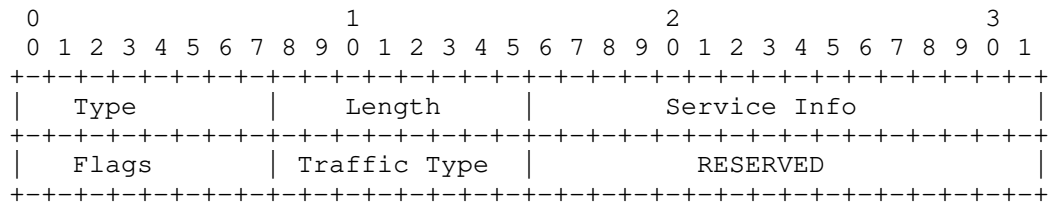



Figure 2:Service Chaining (SC) TLV

where:

Type: 8 bit field. TBD

Length: 8 bit field indicating the length of the remainder of the TLV

The Flags, Traffic Type and RESERVED fields are the same as in the SC TLV defined in [I-D.dawra-idr-bgp-ls-sr-service-segments] chapter 2.

Flags: 8 bit field. Bits SHOULD be 0 on transmission and MUST be ignored on reception.

Traffic Type: 8 Bit field. A bit to identify if Service is IPv4 OR IPv6 OR L2 Ethernet Capable.

Bit 0(LSB): Set to 1 if Service is IPv4 Capable

Bit 1: Set to 1 if Service is IPv6 Capable

Bit 2: Set to 1 if Service is Ethernet Capable

RESERVED: 16bit field. SHOULD be 0 on transmission and MUST be ignored on reception.

Service Info: 16-bits field. The right most 12 bits categorize the Service Type: (such as "Firewall", "Classifier" etc).

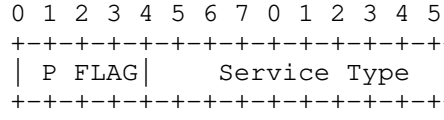



Figure 3: Service Info Field

The first 4 bits are P FLAG which is used to indicate the SR proxy type with the following values:

0000:SR-aware function.

0001:Static proxy.

0010:Dynamic proxy.

0011:Masquerading proxy(for SRv6 only).

0100:Shared memory proxy.

Other values are reserved.

The P FLAG is mainly defined for SR-MPLS.

In SRv6, although the SR proxy type can be represented by the END functions[I-D.ietf-spring-sr-service-programming] which can be advertised in Endpoint Behavior field of End SID sub-TLV [I-D.ietf-lsr-isis-srv6-extensions], there may be situations that the proxy with certain type cannot be associated with a network programming function(for example, Shared memory proxy),or an user want to define a new type of proxy for private use, or the SR proxy node does not support network programming, so the P flag is still useful.

In the IS-IS notification message, when both SR proxy END function and P FLAG exist, the proxy type represented by P FLAG shall prevail.

Another Optional Opaque Metadata(OM) TLV is defined in figure 4. The definition and structure are the same as the OM TLV defined in [I-D.dawra-idr-bgp-ls-sr-service-segments] chapter 2.

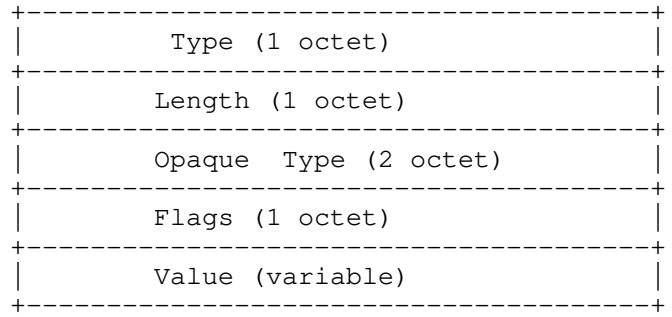



Figure 4:Opaque Metadata(OM) TLV

## 2.2. OSPFv2 and OSPFv3 Extensions

This document introduces new sub-sub-TLVs for SRv6 End SID sub-TLV [I-D.li-ospf-ospfv3-srv6-extensions] and Prefix-SID Sub-TLV [RFC8665] [RFC8665] for SR-MPLS to associate the Service SID Value with Service-related Information.

One of the new TLVs is Service Chaining (SC) TLV,

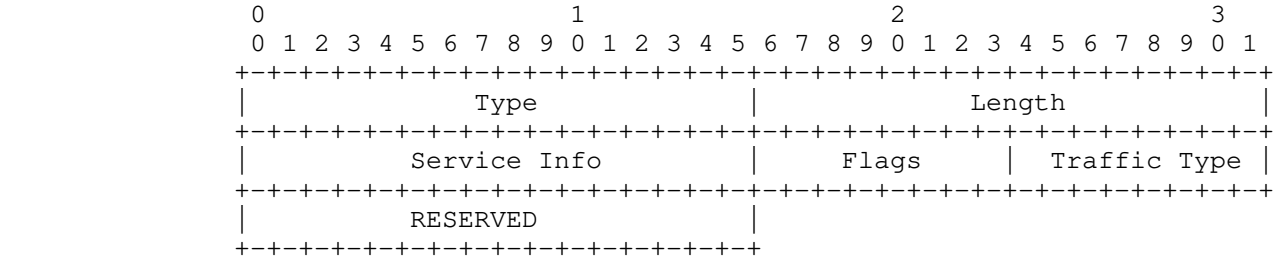



Figure 5:Service Chaining (SC) TLV

where:

Type: 16 bit field. TBD

Length: 16 bit field indicating the length of the remainder of the TLV

Flags, Traffic Type and RESERVED are the same as that in SC TLV defined in [I-D.dawra-idr-bgp-ls-sr-service-segments] chapter 2.

The definition and use principle of the Service Type field is the same as that defined in the IS-IS extension above.

Another Optional Opaque Metadata(OM) TLV is defined in figure 6. The definition and structure are the same as the OM TLV defined in [I-D.dawra-idr-bgp-ls-sr-service-segments] chapter 2.

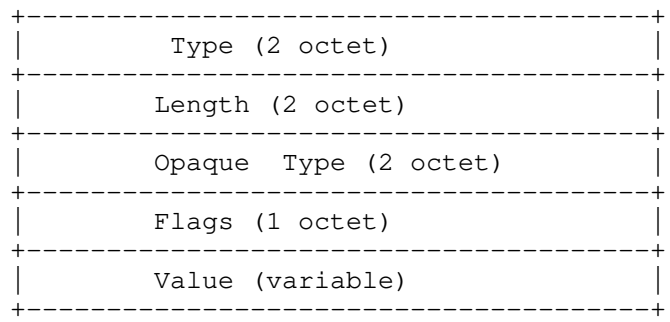



Figure 6:Opaque Metadata(OM) TLV

## 3. Security Considerations

Procedures and protocol extensions defined in this document do not affect the IS-IS and OSPF security model

## 4. IANA Considerations

TBD

## 5. References

### 5.1. Normative References

[I-D.dawra-idr-bgp-ls-sr-service-segments]
Dawra, G., Filsfils, C., Talaulikar, K., Clad, F., daniel.bernier@bell.ca, d., Uttaro, J., Decraene, B., Elmalky, H., Xu, X., Guichard, J., and C. Li, "BGP-LS Advertisement of Segment Routing Service Segments", draft-dawra-idr-bgp-ls-sr-service-segments-03 (work in progress), January 2020.

[I-D.ietf-lsr-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extension to Support Segment Routing over IPv6 Dataplane", draft-ietf-lsr-isis-srv6-extensions-08 (work in progress), April 2020.

[I-D.ietf-spring-sr-service-programming]
Clad, F., Xu, X., Filsfils, C., daniel.bernier@bell.ca, d., Li, C., Decraene, B., Ma, S., Yadlapalli, C., Henderickx, W., and S. Salsano, "Service Programming with Segment Routing", draft-ietf-spring-sr-service-programming-02 (work in progress), March 2020.

[I-D.li-ospf-ospfv3-srv6-extensions]
Li, Z., Hu, Z., Cheng, D., Talaulikar, K., and P. Psenak, "OSPFv3 Extensions for SRv6", draft-li-ospf-ospfv3-srv6-extensions-07 (work in progress), November 2019.

[RFC7665] Halpern, J., Ed. and C. Pignataro, Ed., "Service Function Chaining (SFC) Architecture", RFC 7665, DOI 10.17487/RFC7665, October 2015, <https://www.rfc-editor.org/info/rfc7665>.

[RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <https://www.rfc-editor.org/info/rfc8665>.

[RFC8666] Psenak, P., Ed. and S. Previdi, Ed., "OSPFv3 Extensions for Segment Routing", RFC 8666, DOI 10.17487/RFC8666, December 2019, <https://www.rfc-editor.org/info/rfc8666>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

## 5.2. Informative References

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

## Authors' Addresses

Liu Yao
ZTE Corporation
No. 50 Software Ave, Yuhuatai Distinct
Nanjing
China

Email: liu.yao71@zte.com.cn

Zhang Zheng
ZTE Corporation
No. 50 Software Ave, Yuhuatai Distinct
Nanjing
China

Email: zzhang_ietf@hotmail.com

L
Internet-Draft
Intended status: Informational
Expires: October 1, 2020

P. Shaofu
C. Ran
ZTE Corporation
G. Mirsky
ZTE Corp.

March 30, 2020

# IGP Flexible Algorithm with L2bundles
# draft-peng-lsr-flex-algo-l2bundles-01

## Abstract

IGP Flex Algorithm proposes a solution that allows IGPs themselves to compute constraint based paths over the network, and it also specifies a way of using Segment Routing (SR) Prefix-SIDs and SRv6 locators to steer packets along the constraint-based paths. This document describes how to create Flex-algo plane with L2bundles 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 October 1, 2020.

Table of Contents



## 1. Introduction

IGP Flex Algorithm [I-D.ietf-lsr-flex-algo] proposes a solution that allows IGPs themselves to compute constraint based paths over the network, and it also specifies a way of using Segment Routing (SR) Prefix-SIDs and SRv6 locators to steer packets along the constraint-based paths. It specifies a set of extensions to ISIS, OSPFv2 and OSPFv3 that enable a router to send TLVs that identify (a) calculation-type, (b) specify a metric-type, and (c )describe a set of constraints on the topology, that are to be used to compute the best paths along the constrained topology. A given combination of calculation-type, metric-type, and constraints is known as an FAD (Flexible Algorithm Definition).

[RFC8668] and [I-D.ketant-lsr-ospf-l2bundles] introduces the ability for IS-IS and OSPF respectively to advertise the link attributes of Layer 2 (L2) Bundle Members. Especially, the link attribute "Administrative Group" and "Extended Administrative Group" could be individual to each L2 Bundle Member for purpose of Flex-algo plane construction, where multiple Flex-algo planes share the same Layer 3 parent interface and each Flex-algo plane has dedicated L2 Bundle Member.

This document describes how to create Flex-algo plane with L2bundles scenario.

## 2. 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.

## 3. Color set on L2 Bundle Member

Traffic Engineering affinity (also termed as Color) is often to be set on the Layer 3 interface and be flooded by IGP-TE. However, when the Layer 3 interface is a Layer 2 interface bundle, operators can config individual color for each L2 Bundle Member. So that IGP link-state database will contain the TE affinity attribute of L2 Bundle Member, as well as Layer 3 parrent interface.

Note that Layer 3 interface can join to IGP instance explicitly, but L2 Bundle Member not.

The TE affinity of the Layer 3 parrent interface can be a combined value of all L2 Bundle Members. For example, if the Layer 3 parrent interface contains three L2 Bundle Members, each with color "RED", "GREEN", "BLUE" respectively, the Layer 3 parrent interface will have color "RED|GREEN|BLUE".

## 4. Flex-algo plane with L2 link resource

### 4.1. Best-effort

[I-D.ietf-lsr-flex-algo] defines the color-based link resource selection rules in FAD to construct the expected Flex-algo plane. Each node in the Flex-algo plane will establish the SPT with self as root node, to maintain the best path to other nodes and get the FIB entry based on that. The root node need check the outgoing Layer 2 interface bundle interface, to see which L2 Bundle Member does exactly belong to the Flex-algo plane. The forwarding information of the FIB entry with outgoing Layer 2 interface bundle interface will exactly select the L2 Bundle Member that belongs to the Flex-algo plane to forward packets.

For example, three Flex-algo plane share the same Layer 3 parrent interface including three L2 Bundle Members each with color "RED", "GREEN", "BLUE" respectively, and each Flex-algo plane with link selection rule "Include-Any RED", "Include-Any GREEN", "Include-Any BLUE" respectively, Flex-algo SHOULD not simply select the Layer 3 parrent interface to all Flex-algo plane, but need continue to select individual L2 Bundle Member to the specific Flex-algo plane. As a

reslut, the FIB entry within Flex-algo RED plane will exactly choose the L2 Bundle Members with color "RED" to forward packets, the FIB entry within Flex-algo GREEN plane will exactly choose the L2 Bundle Members with color "GREEN" to forward packets, and the FIB entry within Flex-algo BLUE plane will exactly choose the L2 Bundle Members with color "BLUE" to forward packets.

### 4.2. Traffic Engineering

A segment list contains SIDs advertised specifically for the given algorithm is possible, such as an inter-domain path contains multiple Flex-algo planes, a TI-LFA backup path within the Flex-algo plane, or an optimized TE path avoiding congested link within the Flex-algo plane. In these cases, an Adjacency segment could be used to steer the packets along the expected L2 Bundle Member that belongs to the specific Flex-algo plane.

[RFC8668] and [I-D.ketant-lsr-ospf-l2bundles] have defined Adjacency-SID for each L2 Bundle Member, that can be used to isolate flows among multiple Flex-algo planes, when these Flex-algo planes share the same Layer 3 parrent interface. A specific Adjacency-SID for a specific L2 Bundle Member will steer the packets to that member.

## 5. IGP L2 Bundle Member EAG advertisement

### 5.1. ISIS L2 Bundle Member EAG advertisement

[RFC8668] defined TLV-25 for ISIS to advertise the link attributes of L2 Bundle Members, and mentioned that the traditional "Administrative group (color) Sub-TLV" and "Extended Administrative Group Sub-TLV" may appear in TLV-25 and MAY be shared by multiple L2 Bundle Members. If we want to advertise unique EAG values for each bundle member, we can use multiple L2 Bundle Attribute Descriptors with each specify a single bundle member.

### 5.2. OSPF L2 Bundle Member EAG advertisement

[I-D.ketant-lsr-ospf-l2bundles] defined "L2 Bundle Member Attributes sub-TLV" for OSPF/OSPFv3 to advertise the link attributes of L2 Bundle Members, and mentioned that the traditional "Administrative group (color) Sub-TLV" and "Extended Administrative Group Sub-TLV" are applicable in "L2 Bundle Member Attributes sub-TLV". Because there is "L2 Bundle Member Attributes sub-TLV" per L2 Bundle Member, it is also sufficient to construct Flex-algo plane to select L2 link resource.

## 6. IANA Considerations

This document need not define new sub-TLV to IGP for Flex-algo combined with l2bundles.

## 7. Security Considerations

There are no new security issues introduced by the extensions in this document.

## 8. Acknowledgements

TBD

## 9. Normative References

[I-D.ietf-lsr-flex-algo]
Psenak, P., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", draft-ietf-lsr-flex-algo-06 (work in progress), February 2020.

[I-D.ketant-lsr-ospf-l2bundles]
Talaulikar, K. and P. Psenak, "Advertising L2 Bundle Member Link Attributes in OSPF", draft-ketant-lsr-ospf-l2bundles-01 (work in progress), January 2020.

[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>.

[RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P. Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF", RFC 4915, DOI 10.17487/RFC4915, June 2007, <https://www.rfc-editor.org/info/rfc4915>.

[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>.

[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>.

[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>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8668] Ginsberg, L., Ed., Bashandy, A., Filsfils, C., Nanduri, M., and E. Aries, "Advertising Layer 2 Bundle Member Link Attributes in IS-IS", RFC 8668, DOI 10.17487/RFC8668, December 2019, <https://www.rfc-editor.org/info/rfc8668>.

## Authors' Addresses

Peng Shaofu
ZTE Corporation
No.68 Zijinghua Road, Yuhuatai District
Nanjing
China

Email: peng.shaofu@zte.com.cn

Chen Ran
ZTE Corporation
No.50 Software Avenue, Yuhuatai District
Nanjing
China

Email: chen.ran@zte.com.cn

Greg Mirsky
ZTE Corp.

Email: gregimirsky@gmail.com

LSR
Internet-Draft
Intended status: Standards Track
Expires: July 25, 2020

Shaofu. Peng
Ran. Chen
Gregory. Mirsky
ZTE Corporation

January 22, 2020

# ISIS Extension to Support Network Slicing over IPv6 Dataplane
## draft-peng-lsr-isis-network-slicing-srv6-00

Abstract

[I-D.peng-teas-network-slicing] defines an unified TN-slice identifier, i.e., AII(administrative instance identifier) and related processing combined with Segment Routing technology, for the purpose of unified identification of topology, computing, storage resources, unified partition of L2 and L3 link resources, unified basis of underlay path selection within or between domains, and unified flow steering rule with SR policy color scheme. This document describes the ISIS extensions required to support Packet Network Slicing over IPv6 dataplane.

Table of Contents



## 1. Introduction

For a packet network, network slicing requires the underlying network to support partitioning of the network resources to provide the client with dedicated (private) networking, computing, and storage resources drawn from a shared pool. [I-D.peng-teas-network-slicing] defines an unified TN-slice identifier, i.e., AII(administrative instance identifier) and related processing combined with Segment Routing technology, for the purpose of unified identification of topology, computing, storage resources, unified partition of L2 and L3 link resources, unified basis of underlay path selection within or between domains, and unified flow steering rule with SR policy color scheme.

[I-D.zch-lsr-isis-network-slicing] describes the ISIS extensions required to support Packet Network Slicing over SR-MPLS dataplane. This document continues to describe the ISIS extensions required to

support Packet Network Slicing over IPv6 dataplane, it will supplement on the basis of [I-D.ietf-lsr-isis-srv6-extensions].

For SRv6 case, IPv6 address resource is directly used to represent SID, so that different IPv6 block could be allocated to different TN-slice. There are two possible ways to advertise TN-slice specfic IPv6 block:

- o Traditional prefix reachability, for default AII (0) specific IPv6 block.
- o New SRv6 Locator per TN-slice advertisement, for nonzero TN-slice specific IPv6 block.

## 2. Advertising SRv6 Capabilities and Maximum SID Depths

SRv6 capable router can advertise the SRv6 Capabilities sub-TLV and Maximum SRv6 SID Depths (MSD) as defined in [I-D.ietf-lsr-isis-srv6-extensions].

## 3. Advertising Participated TN-slice

SRv6 capable router indicates participated TN-slice by advertising the TN-slice identifier Participation sub-TLV as defined in [I-D.zch-lsr-isis-network-slicing].

## 4. TN-slice specific SRv6 SIDs and Reachability

A node is provisioned with TN-slice specific locators for each of the TN-slice identified by AII participated by that node. Each locator is a covering prefix for all SIDs provisioned on that node which have the matching AII.

Locators MUST be advertised in the SRv6 Locator per TN-slice TLV (see Section 6.1). Forwarding entries for the locators advertised in the SRv6 Locator per TN-slice TLV MUST be installed in the forwarding plane of receiving SRv6 capable routers when the associated AII is supported by the receiving node.

Locators are routable and MAY also be advertised in Prefix Reachability TLVs (236 or 237). But locators associated with non default AII SHOULD NOT be advertised in Prefix Reachability TLVs (236 or 237).

Locators associated with default AII (0) SHOULD be advertised in a Prefix Reachability TLV (236 or 237) so that legacy routers (i.e., routers which do NOT support SRv6) will install a forwarding entry for default AII (0) SRv6 traffic.

In cases where a locator advertisement is received in both in a Prefix Reachability TLV and an SRv6 Locator per TN-slice TLV, the Prefix Reachability advertisement MUST be preferred when installing entries in the forwarding plane. This is to prevent inconsistent forwarding entries on SRv6 capable/SRv6 incapable routers.

TN-slice specific SRv6 SIDs are advertised as sub-TLVs in the SRv6 Locator per TN-slice TLV except for TN-slice specific SRv6 End.X SIDs/LAN End.X SIDs which are associated with a specific Neighbor/ Link and are therefore advertised as sub-TLVs in TLVs 22, 23, 222, 223, and 141.

TN-slice specific SRv6 SIDs are not directly routable and MUST NOT be installed in the forwarding plane. Reachability to TN-slice specific SRv6 SIDs depends upon the existence of a covering TN-slice specific locator.

Adherence to the rules defined in this section will assure that TN-slice specific SRv6 SIDs associated with a supported AII will be forwarded correctly, while SRv6 SIDs associated with an unsupported AII will be dropped. NOTE: The drop behavior depends on the absence of a default/summary route covering a given locator.

## 5. Advertising Anycast Property

The same prefix/SRv6 Locator can be advertised by multiple routers. See A-flag defined in [I-D.ietf-lsr-isis-srv6-extensions] to advertise the anycast property.

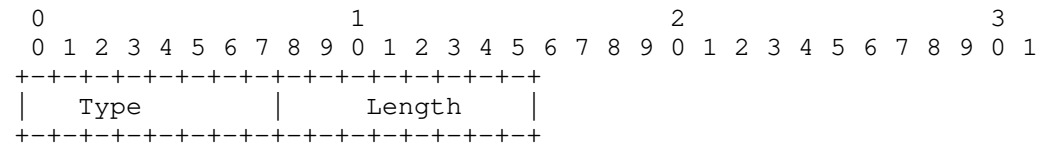
## 6. Advertising Locators and End SIDs for specific TN-slice

The SRv6 Locator per TN-slice TLV is introduced to advertise SRv6 Locators and End SIDs associated with each locator for specific TN-slice.

This new TLV shares the sub-TLV space defined for TLVs 135, 235, 236 and 237.

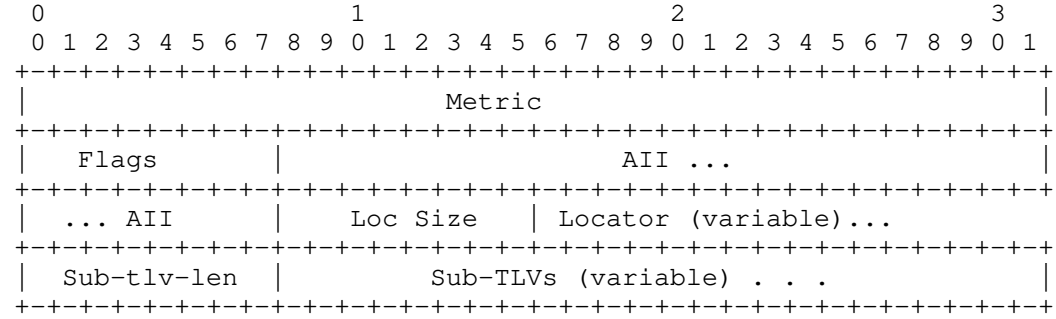
### 6.1. SRv6 Locator per TN-slice TLV Format

The SRv6 Locator per TN-slice TLV has the following format:




SRv6 Locator per TN-slice format

Followed by one or more locator entries of the form:




SRv6 Locator entry per TN-slice format

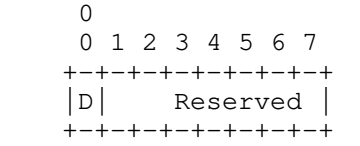
Type: TBA1

Length: variable.

Locator entry:

Metric: 4 octets. As described in [RFC5305].

Flags: 1 octet. The following flags are defined




where:

D bit: When the Locator is leaked from level-2 to level-1, the D bit MUST be set. Otherwise, this bit MUST be clear. Locators with the D

bit set MUST NOT be leaked from level-1 to level-2. This is to prevent looping.

The remaining bits are reserved for future use. They SHOULD be set to zero on transmission and MUST be ignored on receipt.

AII: 4 octet. Administrative Instance Identifier, As defined in [I-D.peng-teas-network-slicing], represented as the TN-slice Identifier.

Loc-Size: 1 octet. Number of bits in the Locator field. (1 - 128)

Locator: 1-16 octets. This field encodes the advertised SRv6 Locator. The Locator is encoded in the minimal number of octets for the given number of bits.

Sub-TLV-length: 1 octet. Number of octets used by sub-TLVs

Optional sub-TLVs.

### 6.2. SRv6 End SID for specific TN-slice

The SRv6 End SID sub-TLV defined in [I-D.ietf-lsr-isis-srv6-extensions] can be reused in this document to advertise TN-slice specific SRv6 Segment Identifiers (SID) with Endpoint behaviors.

The SRv6 End SID sub-TLV is advertised in the SRv6 Locator per TN-slice TLV defined in the previous section. SRv6 End SIDs inherit the AII from the parent locator.

The SRv6 End SID MUST be a subnet of the associated Locator. SRv6 End SIDs which are NOT a subnet of the associated locator MUST be ignored.

Multiple SRv6 End SIDs MAY be associated with the same locator. In cases where the number of SRv6 End SID sub-TLVs exceeds the capacity of a single TLV, multiple Locator per TN-slice TLVs for the same TN-slice specific locator MAY be advertised. For a given AII/Locator the AII value MUST be the same in all TLVs. If this restriction is not met all TLVs for that AII/Locator MUST be ignored.

## 7. Advertising SRv6 Adjacency SIDs for specific TN-slice

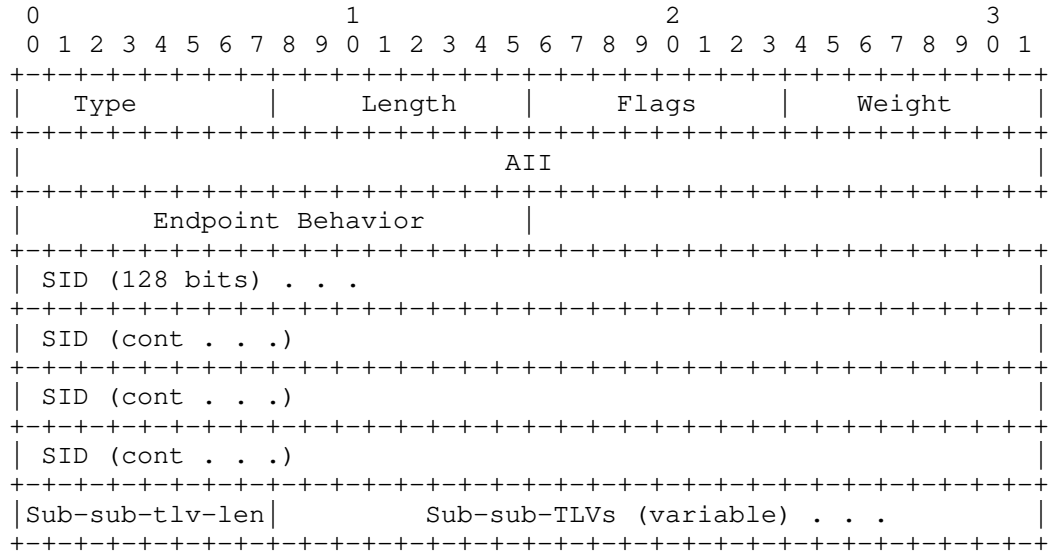
This document defines two new sub-TLVs of TLV 22, 23, 222, 223, and 141 - namely "SRv6 End.X SID per TN-slice" and "SRv6 LAN End.X SID per TN-slice".

All End.X SIDs for specific TN-slice MUST be a subnet of a Locator with matching AII which is advertised by the same node in an SRv6 Locator per TN-slice TLV. End.X SIDs which do not meet this requirement MUST be ignored.

### 7.1. SRv6 End.X SID per TN-slice sub-TLV

This sub-TLV is used to advertise a TN-slice specific SRv6 SID associated with a point to point adjacency. Multiple SRv6 End.X SID per TN-slice sub-TLVs MAY be associated with the same adjacency.

The SRv6 End.X SID per TN-slice sub-TLV has the following format:

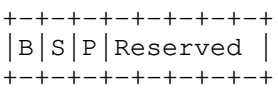



SRv6 End.X SID per TN-slice format

Type: TBA2

Length: variable.

Flags: 1 octet.




where:

B-Flag: Backup flag. If set, the End.X SID is eligible for protection (e.g., using IPFRR) as described in [RFC8355].

S-Flag. Set flag. When set, the S-Flag indicates that the End.X SID refers to a set of adjacencies (and therefore MAY be assigned to other adjacencies as well).

P-Flag. Persistent flag. When set, the P-Flag indicates that the End.X SID is persistently allocated, i.e., the End.X SID value remains consistent across router restart and/or interface flap.

Other bits: MUST be zero when originated and ignored when received.

Weight: 1 octet. The value represents the weight of the End.X SID for the purpose of load balancing. The use of the weight is defined in [RFC8402].

AII: 4 octet. Administrative Instance Identifier, As defined in [I-D.peng-teas-network-slicing], represented as TN-slice Identifier.

Endpoint Behavior: 2 octets. As defined in [I-D.ietf-spring-srv6-network-programming]. Legal behavior values for this sub-TLV are defined in Section 8.

SID: 16 octets. This field encodes the advertised SRv6 SID.

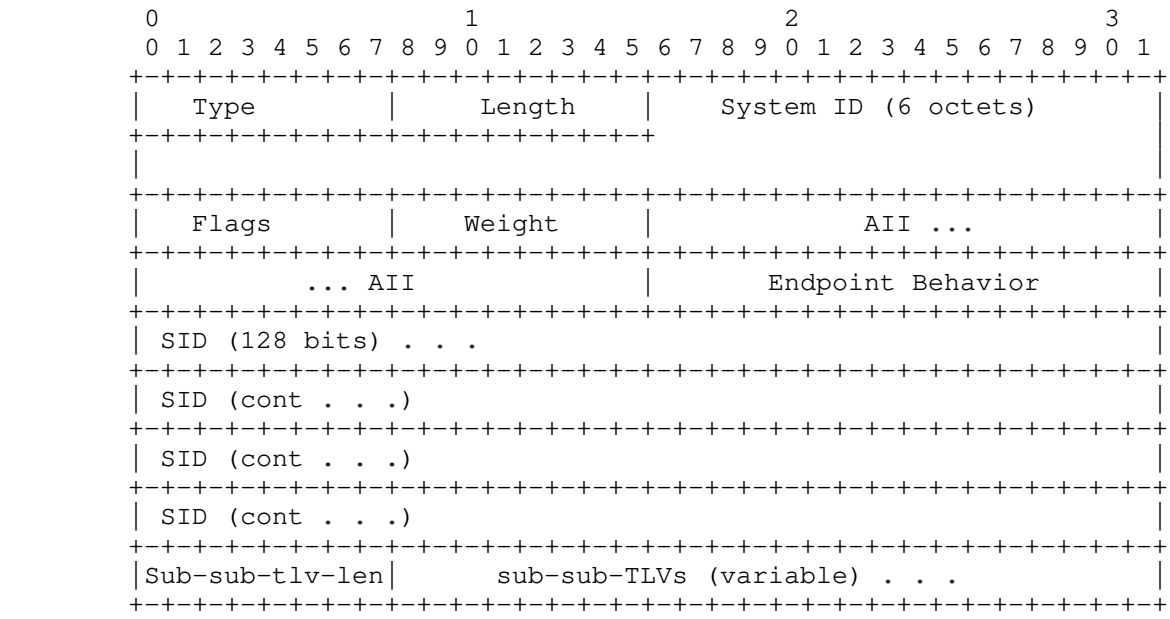
Sub-sub-TLV-length: 1 octet. Number of octets used by sub-sub-TLVs

Note that multiple TLVs for the same neighbor may be required in order to advertise all of the SRv6 End.X SIDs associated with that neighbor.

### 7.2. SRv6 LAN End.X SID per TN-slice sub-TLV

This sub-TLV is used to advertise a TN-slice specific SRv6 SID associated with a LAN adjacency. Since the parent TLV is advertising an adjacency to the Designated Intermediate System(DIS) for the LAN, it is necessary to include the System ID of the physical neighbor on the LAN with which the SRv6 SID is associated. Given that a large number of neighbors may exist on a given LAN a large number of SRv6 LAN END.X SID per TN-slice sub-TLVs may be associated with the same LAN. Note that multiple TLVs for the same DIS neighbor may be required in order to advertise all of the TN-slice specific SRv6 End.X SIDs associated with that neighbor.

The SRv6 LAN End.X SID per TN-slice sub-TLV has the following format:




SRv6 LAN End.X SID per TN-slice format

Type: TBA3

Length: variable.

System-ID: 6 octets of IS-IS System-ID of length "ID Length" as defined in [ISO10589].

Flags: 1 octet.


where B,S, and P flags are as described in Section 7.1. Other bits: MUST be zero when originated and ignored when received.

Weight: 1 octet. The value represents the weight of the End.X SID for the purpose of load balancing. The use of the weight is defined in [RFC8402].

AII: 4 octet. Administrative Instance Identifier, As defined in [I-D.peng-teas-network-slicing], represented as TN-slice Identifier.

Endpoint Behavior: 2 octets. As defined in [I-D.ietf-spring-srv6-network-programming]. Legal behavior values for this sub-TLV are defined in Section 8.

SID: 16 octets. This field encodes the advertised SRv6 SID.

Sub-sub-TLV-length: 1 octet. Number of octets used by sub-sub-TLVs.

## 8. SRv6 SID Structure Indication

The SRv6 SID Structure Sub-Sub-TLV defined in [I-D.ietf-lsr-isis-srv6-extensions] can be reused in this document to indicate the SRv6 SID Structure information. The SRv6 SID Structure Sub-Sub-TLV could be an optional Sub-Sub-TLV of:

SRv6 End SID Sub-TLV (Section 6.2)

SRv6 End.X SID per TN-slice Sub-TLV (Section 7.1)

SRv6 LAN End.X SID per TN-slice Sub-TLV (Section 7.2)

## 9. Advertising Endpoint Behaviors

The Endpoint behaviors and their codepoints, which MAY be advertised by IS-IS and the SID sub-TLVs in which each type MAY Appear, are consistent with that in [I-D.ietf-lsr-isis-srv6-extensions].

## 10. Security Considerations

Security concerns for IS-IS are addressed in [ISO10589], [RFC5304], and [RFC5310].

## 11. Acknowledgements

TBD

## 12. IANA Considerations

This document requests allocation for the following TLVs and sub-TLVs in the ISIS TLV registry.

### 12.1. SRv6 Locator per TN-slice TLV

This document adds one new TLV to the IS-IS TLV Codepoints registry.

Value: TBA1

Name: SRv6 Locator for specific TN-slice

### 12.2. SRv6 End.X SID per TN-slice and SRv6 LAN End.X SID per TN-slice sub-TLVs

This document adds the definition of two new sub-TLVs in the "sub-TLVs for TLV 22, 23, 25, 141, 222 and 223 registry".

Type: TBA2

Description: SRv6 End.X SID per TN-slice

Type: TBA3

Description: SRv6 LAN End.X SID per TN-slice

## 13. Normative References

[I-D.ietf-lsr-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extension to Support Segment Routing over IPv6 Dataplane", draft-ietf-lsr-isis-srv6-extensions-04 (work in progress), January 2020.

[I-D.ietf-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-ietf-spring-srv6-network-programming-08 (work in progress), January 2020.

[I-D.peng-teas-network-slicing]
Peng, S., Chen, R., Mirsky, G., and F. Qin, "Packet Network Slicing using Segment Routing", draft-peng-teas-network-slicing-02 (work in progress), December 2019.

[I-D.zch-lsr-isis-network-slicing]
Zhu, Y., Chen, R., Peng, S., and F. Qin, "IS-IS Extensions to Support Packet Network Slicing using Segment Routing", draft-zch-lsr-isis-network-slicing-03 (work in progress), December 2019.

[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>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

Authors' Addresses

Shaofu Peng
ZTE Corporation

Email: peng.shaofu@zte.com.cn

Ran Chen
ZTE Corporation

Email: chen.ran@zte.com.cn

Gregory Mirsky
ZTE Corporation

Email: gregimirsky@gmail.com

LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 28, 2021

A. Wang
China Telecom
Z. Hu
Y. Xiao
Huawei Technologies
July 27, 2020

# Prefix Unreachable Announcement
# draft-wang-lsr-prefix-unreachable-annoucement-03

## Abstract

This document describes the mechanism that can be used to announce the unreachable prefixes for service fast convergence.

## 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 28, 2021.

## Copyright Notice

Table of Contents



## 1. Introduction

OSPF and IS-IS have the summary route and default route mechanism on area border router or L1L2 border router, which is used to increase the scalability of these IGP protocols. Such summary mechanism can also reduce the SPF calculation time when the link oscillation occurs in another area.

The summary route and the default route may cover the host route or link prefixes of intra area or inter area. But in some situations, the router needs to know the exact reachability information about prefix in other area, especially when the prefix is unreachable but it is located within the summary range.

With the introduction of SRv6, more and more services are migrated from the MPLS data plane to the IPv6 data plane. The biggest difference between IPv6 and MPLS is that IPv6 has aggregation ability, so we need to reconsider how to know the prefix reachability in the case of aggregation.

This document introduces the mechanism that can be used in such situation, to announce the unreachable prefixes which are located in the summary address range.

## 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

Figure1 illustrates the topology scenario when OSPF is running in multi-area. R0-R4 are routers in backbone area, S1-S4,T1-T4 are internal routers in area 1 and area 2 respectively. R1 and R3 are area border routers between area 0 and area 1. R2 and R4 are area border routers between area 0 and area 2. Ps2 is the host address of S2 and Pt2 is the host address T2.

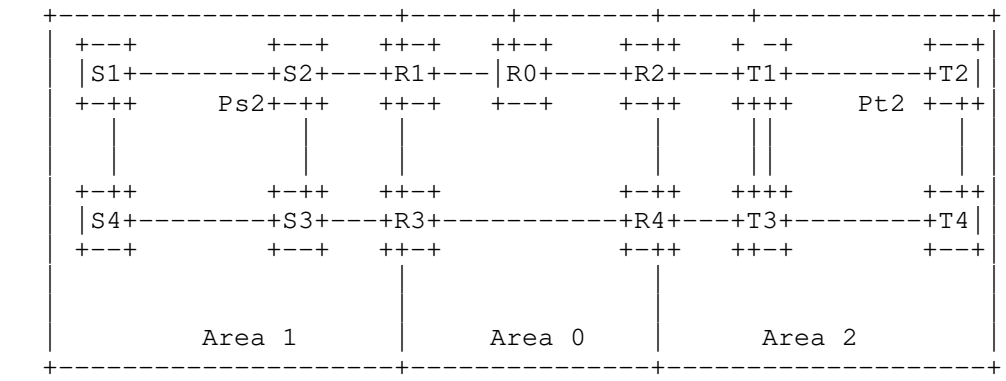



Figure 1: OSPF Inter-Area Prefix Unreachable Announcement Scenario

### 3.1. Inter-Area Node Failure Scenario

If the area border router R2/R4 does the summary action, then one summary address that cover the prefixes of area 2 will be announced to area 0 and area 1, instead of the detail address. When the node T2 is down, Pt2 become unreachable. But there will be no change to the summary prefix. Except the border router R2/R4, the other routers within area 0 and area 1 do not know the unreachable status of this prefix. When these routers send traffic to prefix Pt2, the traffic will be dropped.

### 3.2. Inter-Area Links Failure Scenario

In other situation, if the link between T1/T2 and T1/T3 are broken, R2 will not be able to reach node T2. But as R2 and R4 do the summary announcement, and the summary address covers the prefix of Pt2, other nodes in area 0 area 1 will still send traffic to T2 via

the border router R2. When R2 receives such traffic, it will drop the packet.

In such situation, the border router R2 should notify other routers that it can't reach the prefix Pt2, and lets the other routers to select R4 as the bypass router to reach prefix Pt2.

### 3.3. Intra-Area Node Failure Scenario

For intra area, there are some situations that the border routers, for example R1/R3 in Figure 1, announces the summary address that cover also the prefix addresses in area 1. In this situation, when node S2 failures, node S1 should send traffic to the back up path that bypass the failure node. But the back up path can't be triggered because node S1 still think it can reach the prefix Ps2 because it has the summary route that announced by the border router R1/R3.

From the above scenarios, we can conclude that in such situations, the mechanism for Prefix Unreachable Announcement (PUA) should be designed to alleviate the traffic loss.

## 4. Inter-area prefix unreachable solution

[RFC7794] and [I-D.ietf-lsr-ospf-prefix-originator] both define one sub-TLV "Prefix Source Router ID" to announce the originator router information of one prefix. This TLV can be used to announce the prefix unreachable information when the link or node is down.

According to the procedure described in section 5 of [I-D.ietf-lsr-ospf-prefix-originator], the ABR has the responsibility to add the prefix originator information when it receive the Router LSA from other routers in the same area. When the ABR does the summary work and receives one updated LSA that omits the prefix belong to failed link which is within the range of summary address, the ABR should announce one new Summary LSA, which includes the information about this prefix, but with the prefix originator set to NULL(all 0 address).

When one node in one area is down, the ABR has also the ability to detect the missing neighbor from the neighbor list. It should then announce one new Summary LSA that includes the loopback addresses of this node, with the prefix originator set also to NULL(all 0 address).

For IS-IS, the above procedure is similar. The level-1/2 router will accomplish the above work when it judges that one prefix within the summary address range is missing.

These LSAs will be transported via the traditional flooding procedure.

When the routers in other area receives such LSA, they will generate automatically one black-hole route, with the prefix as the destination, and the next hop be set to Null. If there is other router advertise the summary prefix without carry unreachable information, it will prefer the other router to reach the specified prefix.

## 5. Intra-area prefix unreachable solution

In the intra-area scenario, like S1 illustrated in Figure 1, it will learn two types of prefixes, one is summary route, another is host route. When node S2 is down, S2 will withdraw the host route. But S1 can still match the summary route via the longest mask matching. For this scenario, when node S2 is down, S1 needs to keep the S2 host route for a period of time but updates S2 host route to black hole route. S1 will match the black hole route via the longest mask matching. Such mechanism can be used to trigger a SRv6 VPN for PE switching, or SRv6 TE mid-point protection.

The period for keeping the black hole route should be configured, to ensure the related protocols or services be converged.

## 6. Implementation Consideration

The above procedures will only be triggered under the following conditions:

1. The ABR or Level 1/2 router do the summary work.

2. The link prefix or loopback address of the node within the summary address range become unreachable.

The Summary LSA that includes the unreachable prefix, with the prefix originator set to NULL value, will be announced across the ABR router, reach the routers in other areas. It’s behavior is still the same as that defined in OSPFv2 [RFC2328] or OSPFv3 [RFC5340]

Considering the balances of reachable information and unreachable information announcement capabilities, the implementation of this mechanism should set one MAX_Address_Announcement (MAA) threshold value that can be configurable. Then, the ABR should make the following decisions to announce the prefixes:

1. If the number of unreachable prefixes is less than MAA, the ABR should advertise the summary address and the PUA.

2. If the number of reachable address is less than MAA, the ABR should advertise the detail reachable address only.

3. If the number of reachable prefixes and unreachable prefixes exceed MAA, then advertise the summary address with MAX metric.

When the receiver receives such LSA, it will do the following judgements and actions:

1. If all the source that announces the summary route announces the prefix unreachable information, the receiver should add the black hole route to this prefix.

2. If not, the receiver should prefer the router that does not include the prefix unreachable information to reach this prefix.

3. The receiver router should keep the black hole routes for PUA as one configurable time(MAX_T_PUA) to allow the services that depends on them converged. After the MAX_T_PUA time, such black hole routes can be deleted then.

## 6.1. Usages of Tunnel among ABRs

When in situation that all the ABRs reach the announcement limit, it is not viable to increase the cost of summary address, as described in above paragraph. In such situation, the operator should provide other solution to decrease the packet loss that due to the advertised summary route, which includes the failure prefix. Figure 2 illustrate such situation.

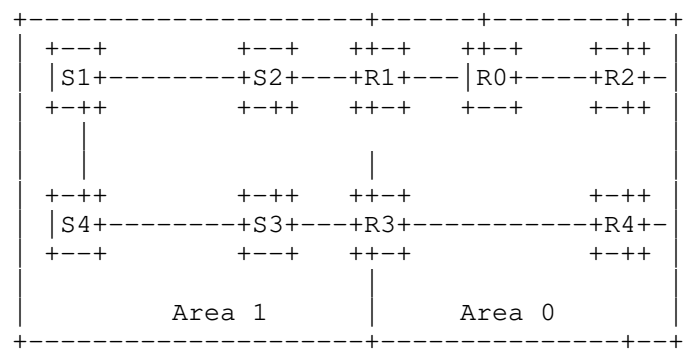



Figure 2: Usage of Tunnel among ABRs

In Figure 2, when R1 and R3 reach the PUA MAA state simultaneously, it is no use for these two ABRs increase the summary cost. For example, when the link between S1 and S4 is down, R1 can reach S1/S2

but not S3/S4, R3 can reach S3/S4 but not S1/S2. If the traffic destined to S3/S4 be sent via R1, it will be dropped by R1, but such traffic can be sent to the destination via R3. The traffic destined to S1/S2 that be sent via R3 will have the same fate.

In such situation, it is useful for R1 to send these traffic via some tunnel to R3 and vice versa. To achieve this, the ABR (R1/R3) should build the tunnel previously. When one of the ABRs receive the failure information, it should check whether the missed nodes can be reached via other ABRs. If such missed nodes can be reached, it then install the tunnel route as the next hop to these missed nodes. And when it receives the related traffic, it can transfer the traffic via other ABRs. Such implementation can mitigate the traffic loss in Figure 2.

In order to prevent the traffic loop, when one ABR receives such traffic via the tunnel interface but can’t find the next hop for these traffic, it should drop such traffic and can’t send again via tunnel to other ABRs.

If ABR receive the link/node failure information, and can’t find other ABRs to reach the missed nodes, it should send some notify messages to the operator because some nodes are out of the network and the ABRs can’t notify the nodes in other area via the PUA mechanism.

### 6.2. Fast Rerouting to Avoid Routing Backhole

Fast rerouting is a mechanism that allows a router whose local link has failed to forward traffic to a pre-computed alternate path until the router installs the new primary next-hops based upon the changed network topology. If the area border router R2/R4 does the summary action, both R2 and R4 should pre-install one path to the summary address, with the nexthop address pointed to each other. When the ABR R2 becomes unreachable to a node in one area, R2 will withdraw the detailed route of the node. The pre-install summary route will be the longest match route for the summary address. The traffic destined to the failed node arrived on R2 will be forwarded to another ABR R4 then. If R4 have the detailed route of the node, R4 will forward the traffic to the corresponding node along the correct path. When both R2 and R4 becomes unreachable, how to avoid the traffic loops between R2 and R4 is beyond the scope of this document.
.

### 6.3. PUA Capabilities Announcement

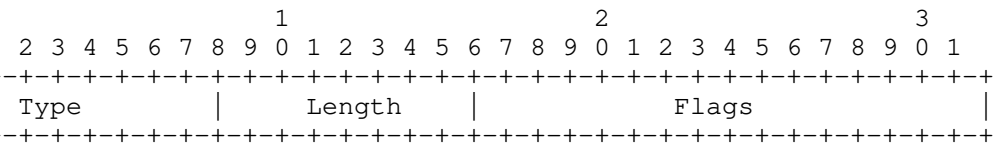
When not all of the nodes in one area support the PUA information, there are possibilities to form traffic loop. To avoid this happen, the ABR should not send PUA information to one area until it ensures that all of nodes in this area can parse the PUA information. In the situation that not all of nodes support PUA information, the ABR should use the mechanism that described section Section 6.1 and Section 6.2 to forward the received traffic that bound to the unreachable prefixes.

To accomplish this, this draft defines the capabilities sub-TLV as the followings:

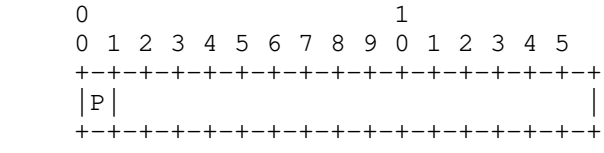
For OSPFv2, this bit (Bit number TBD, suggest bit 6, 0x20) should be carried in "OSPF Router-LSA Option", as that described in RFC2328 [RFC2328]

For OSPFv3, one bit (Bit number TBD, suggest bit 8) should be defined to indicate the router's capabilities to support PUA that described in this draft, the defined bit should be carried in "OSPF Router Informational Capabilities" TLV, which is described in [RFC7770]

For ISIS, one new sub-TLV(Type TBD, suggest 29), PUA Capabilities sub-TLV, which is included in the "IS-IS Router CAPABILITY TLV" [RFC7981] is defined in the followings:




Type: TBD, Suggested value 29, to be assigned by IANA
Length: 2
Flags: 2 octets
The following flags are defined:




where:

P-flag: If set, the router supports PUA information.

Figure 3: PUA Capabilities sub-TLV format

## 7. Security Considerations

Security concerns for OSPF are addressed in [RFC5709]

Advertisement of the additional information defined in this document may raise some compatible issues when the node does not recognize it or consider such information is illegal. During deployment, the operator should make sure all the routers within its domain have supported such features.

## 8. IANA Considerations

IANA is requested to register the following in the "OSPF Router Properties Registry" and "OSPF Router Informational Capability Bits Registry" respectively.

| Bit Number | Capability Name | Reference |
|---|---|---|
| TBD(0x20) | OSPF PUA Support | this document |

Table 1: P-Bit in OSPF Router-LSA Option

| Bit Number | Capability Name | Reference |
|---|---|---|
| TBD(bit 8) | OSPF PUA Support | this document |

Table 2: OSPF Router PUA Capability Support Bit

IANA is requested to register the following in "Sub-TLVs for TLV242(IS-IS Router CAPABILITY TLV)

Type: 29 (Suggested - to be assigned by IANA)

Description: PUA Support Capabilities

## 9. Acknowledgement

Thanks Peter Psenak, Les Ginsberg and Acee Lindem for their suggestions and comments on this draft.

## 10. Normative References

[I-D.ietf-lsr-ospf-prefix-originator]
Wang, A., Lindem, A., Dong, J., Psenak, P., and K. Talaulikar, "OSPF Prefix Originator Extensions", draft-ietf-lsr-ospf-prefix-originator-06 (work in progress), June 2020.

[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>.

[RFC5709] Bhatia, M., Manral, V., Fanto, M., White, R., Barnes, M., Li, T., and R. Atkinson, "OSPFv2 HMAC-SHA Cryptographic Authentication", RFC 5709, DOI 10.17487/RFC5709, October 2009, <https://www.rfc-editor.org/info/rfc5709>.

[RFC7770] Lindem, A., Ed., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 7770, DOI 10.17487/RFC7770, February 2016, <https://www.rfc-editor.org/info/rfc7770>.

[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>.

[RFC7981] Ginsberg, L., Previdi, S., and M. Chen, "IS-IS Extensions for Advertising Router Information", RFC 7981, DOI 10.17487/RFC7981, October 2016, <https://www.rfc-editor.org/info/rfc7981>.

## Authors' Addresses

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

Email: wangaj3@chinatelecom.cn

Zhibo Hu
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: huzhibo@huawei.com

Yaqun Xiao
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: xiaoyaqun@huawei.com

LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 14, 2021

C. Xie
C. Ma
China Telecom
J. Dong
Z. Li
Huawei Technologies

July 13, 2020

# Using IS-IS Multi-Topology (MT) for Segment Routing based Virtual Transport Network
## draft-xie-lsr-isis-sr-vtn-mt-01

## Abstract

Enhanced VPN (VPN+) as defined in I-D.ietf-teas-enhanced-vpn aims to provide enhanced VPN service to support some application's needs of enhanced isolation and stringent performance requirements. VPN+ requires integration between the overlay VPN and the underlay network. A Virtual Transport Network (VTN) is a virtual network which consists of a subset of the network topology and network resources allocated from the underlay network. A VTN could be used as the underlay for one or a group of VPN+ services.

I-D.dong-lsr-sr-enhanced-vpn defines the IGP extensions to build a set of Segment Routing (SR) based VTNs. This document describes a simplified mechanism to build the SR based VTNs using IGP multi-topology together with other well-defined IS-IS extensions.

## 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].

## 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 14, 2021.



Table of Contents



## 1. Introduction

Enhanced VPN (VPN+) is an enhancement to VPN services to support the needs of new applications, particularly including the applications that are associated with 5G services. These applications require enhanced isolation and have more stringent performance requirements than that can be provided with traditional overlay VPNs. These properties cannot be met with pure overlay networks, as they require integration between the underlay and the overlay networks. [I-D.ietf-teas-enhanced-vpn] specifies the framework of enhanced VPN and describes the candidate component technologies in different network planes and layers. An enhanced VPN may be used for 5G

transport network slicing, and will also be of use in other generic scenarios.

To meet the requirement of enhanced VPN services, a number of virtual transport networks (VTN) need to be created, each with a subset of the underlay network topology and a set of network resources allocated to meet the requirement of a specific VPN+ service or a group of VPN+ services. Another existing approach is to build a set of point-to-point paths, each with a set of network resource reserved along the path, such paths is called Virtual Transport Path (VTP). Although using a set of dedicated VTPs can provide similar characteristics, it has some scalability issues in large networks.

[I-D.dong-spring-sr-for-enhanced-vpn] specifies how segment routing (SR) [RFC8402] can be used to build virtual transport networks (VTNs) with the required network topology and network resource attributes to support enhanced VPN services. With segment routing based data plane, Segment Identifiers (SIDs) can be used to represent the topology and the set of network resources allocated by network nodes to a virtual network. The SIDs of each VTN and the associated topology and resource attributes need to be distributed using control plane.

[I-D.dong-lsr-sr-enhanced-vpn] defines the IGP mechanisms with necessary extensions to build a set of Segment Routing (SR) based VTNs. The VTNs could be used as the underlay of the enhanced VPN service. The mechanism described in [I-D.dong-lsr-sr-enhanced-vpn] allows flexible combination of the topology and resource attribute to build customized VTNs. In some network scenarios, it is assumed that each VTN has an independent topology and a set of dedicated network resources. This document describes a simplified mechanism to build the SR based VTNs in those scenarios.

The approach is to use IS-IS Multi-Topology [RFC5120] with segment routing [RFC8667] to define the independent network topologies of each VTN. The information of network resources allocated to a VTN can be advertised by using IS-IS MT with the Traffic Engineering (TE) extensions defined in [RFC5305].

## 2. Advertisement of SR VTN Topology Attribute

Multi-Topology Routing (MTR) [RFC5120] has been defined to create independent topologies in one network. It also has the capability of specifying the customized attributes of each topology. MTR can be used with segment routing based data plane. The IS-IS extensions to support the advertisement of topology-specific MPLS SIDs are specified in [RFC8667]. Topology-specific Prefix-SIDs are advertised by carrying the Prefix-SID sub-TLVs in the IS-IS TLV 235 (MT IP

Reachability) and TLV 237 (MT IPv6 IP Reachability). Topology-specific Adj-SIDs are advertised by carrying the Adj-SID sub-TLVs in IS-IS TLV 222 (MT-ISN) and TLV 223 (MT IS Neighbor Attribute).

The IS-IS extensions to support the advertisement of topology-specific SRv6 Locators and SIDs are specified in [I-D.ietf-lsr-isis-srv6-extensions]. The topology-specific SRv6 locators are advertised using SRv6 Locator TLV, and SRv6 End SIDs inherit the MT-ID from the parent locator. The topology-specific End.X SID are advertised by carrying SRv6 End.X SID sub-TLVs in the IS-IS TLV 222 (MT-ISN) and TLV 223 (MT IS Neighbor Attribute).

When each VTN has an independent network topology, the MT-ID could be used as the identifier of VTN in control plane. Thus the topology attribute of a VTN could be advertised using MTR with segment routing.

## 3. Advertisement of SR VTN Resource Attribute

In order to perform constraint based path computation for each VTN on the network controller or on the ingress nodes, the network resource attribute associated with each VTN needs to be advertised.

### 3.1. Advertising Topology-specific TE attributes

On each network link, the information of the network resources associated with a VTN can be specified by carrying the TE attributes sub-TLVs [RFC5305] in the IS-IS TLV 222 (MT-ISN) and TLV 223 (MT IS Neighbor Attribute) of the corresponding topology.

When Maximum Link Bandwidth sub-TLV is carried in the MT-ISN TLV, it indicates the amount of link bandwidth allocated to the corresponding VTN. The bandwidth allocated to a VTN can be exclusive for services carried in the corresponding VTN. The usage of other TE attributes in topology-specific TLVs is for further study.

Editor’s note1: It is noted that carrying per-topology TE attributes was considered as a possible feature in future when the encoding of IS-IS multi-topology was defined [RFC5120].

## 4. Scalability Considerations

The mechanism described in this document requires that each VTN has an independent topology. Even if multiple VTNs may have the same topology attribute, they would still need to be identified using different MT-IDs in the control plane. This requires that for each VTN, independent path computation would be executed. The number of

VTNs supported in a network may be dependent on the control plane computation overhead.

## 5. Security Considerations

This document introduces no additional security vulnerabilities to IS-IS.

The mechanism proposed in this document is subject to the same vulnerabilities as any other protocol that relies on IGPs.

## 6. IANA Considerations

This document does not request any IANA actions.

## 7. Acknowledgments

The authors would like to thank Zhibo Hu, Dean Cheng, Les Ginsberg and Peter Psenak for the review and discussion of this document.

## 8. References

### 8.1. Normative References

[I-D.dong-spring-sr-for-enhanced-vpn]
Dong, J., Bryant, S., Miyasaka, T., Zhu, Y., Qin, F., and Z. Li, "Segment Routing for Resource Guaranteed Virtual Networks", draft-dong-spring-sr-for-enhanced-vpn-08 (work in progress), June 2020.

[I-D.ietf-lsr-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extension to Support Segment Routing over IPv6 Dataplane", draft-ietf-lsr-isis-srv6-extensions-08 (work in progress), April 2020.

[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>.

[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>.

[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>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

## 8.2. Informative References

[I-D.dong-lsr-sr-enhanced-vpn]
Dong, J., Hu, Z., Li, Z., Tang, X., Pang, R., JooHeon, L., and S. Bryant, "IGP Extensions for Segment Routing based Enhanced VPN", draft-dong-lsr-sr-enhanced-vpn-04 (work in progress), June 2020.

[I-D.ietf-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-ietf-spring-srv6-network-programming-16 (work in progress), June 2020.

[I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Enhanced Virtual Private Networks (VPN+) Services", draft-ietf-teas-enhanced-vpn-05 (work in progress), February 2020.

## Authors' Addresses

Chongfeng Xie
China Telecom
China Telecom Beijing Information Science & Technology, Beiqijia
Beijing 102209
China

Email: xiechf@chinatelecom.cn

Chenhao Ma
China Telecom
China Telecom Beijing Information Science & Technology, Beiqijia
Beijing 102209
China

Email: machh@chinatelecom.cn

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Road
Beijing 100095
China

Email: jie.dong@huawei.com

Zhenbin Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Road
Beijing 100095
China

Email: lizhenbin@huawei.com