

LSR Workgroup
Internet-Draft
Intended status: Standards Track
Expires: August 23, 2021

A. Lindem
Cisco Systems
Y. Qu
Futurewei
A. Roy
Arrcus, Inc.
S. Mirtorabi
Cisco Systems
February 19, 2021

OSPF Transport Instance Extensions
draft-acee-lsr-ospf-transport-instance-02

Abstract

OSPFv2 and OSPFv3 include a reliable flooding mechanism to disseminate routing topology and Traffic Engineering (TE) information within a routing domain. Given the effectiveness of these mechanisms, it is convenient to envision using the same mechanism for dissemination of other types of information within the domain. However, burdening OSPF with this additional information will impact intra-domain routing convergence and possibly jeopardize the stability of the OSPF routing domain. This document presents mechanism to relegate this ancillary information to a separate OSPF instance and minimize the impact.

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 August 23, 2021.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Requirements Language	3
3. Possible Use Cases	3
3.1. MEC Service Discovery	3
3.2. Application Data Dissemination	4
3.3. Intra-Area Topology for BGP-LS Distribution	4
4. OSPF Transport Instance	4
4.1. OSPFv2 Transport Instance Packet Differentiation	5
4.2. OSPFv3 Transport Instance Packet Differentiation	5
4.3. Instance Relationship to Normal OSPF Instances	5
4.4. Network Prioritization	5
4.5. OSPF Transport Instance Omission of Routing Calculation	6
4.6. Non-routing Instance Separation	6
4.7. Non-Routing Sparse Topologies	7
4.7.1. Remote OSPF Neighbor	7
4.8. Multiple Topologies	8
5. OSPF Transport Instance Information (TII) Encoding	8
5.1. OSPFv2 Transport Instance Information Encoding	8
5.2. OSPFv3 Transport Instance Information Encoding	9
5.3. Transport Instance Information (TII) TLV Encoding	10
5.3.1. Top-Level TII Application TLV	10
6. Manageability Considerations	11
7. Security Considerations	11
8. IANA Considerations	11
8.1. OSPFv2 Opaque LSA Type Assignment	11
8.2. OSPFv3 LSA Function Code Assignment	12
8.3. OSPF Transport Instance Information Top-Level TLV Registry	12
9. Acknowledgement	12
10. References	12
10.1. Normative References	12

10.2. Informative References	13
Authors' Addresses	14

1. Introduction

OSPFv2 [RFC2328] and OSPFv3 [RFC5340] include a reliable flooding mechanism to disseminate routing topology and Traffic Engineering (TE) information within a routing domain. Given the effectiveness of these mechanisms, it is convenient to envision using the same mechanism for dissemination of other types of information within the domain. However, burdening OSPF with this additional information will impact intra-domain routing convergence and possibly jeopardize the stability of the OSPF routing domain. This document presents mechanism to relegate this ancillary information to a separate OSPF instance and minimize the impact.

This OSPF protocol extension provides functionality similar to "Advertising Generic Information in IS-IS" [RFC6823]. Additionally, OSPF is extended to support sparse non-routing overlay topologies Section 4.7.

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. Possible Use Cases

3.1. MEC Service Discovery

Multi-Access Edge Computing (MEC) plays an important role in 5G architecture. MEC optimizes the performance for ultra-low latency and high bandwidth services by providing networking and computing at the edge of the network [ETSI-WP28-MEC]. To achieve this goal, it's important to expose the network capabilities and services of a MEC device to 5G User Equipment UE, i.e. UEs.

The followings are an incomplete list of the kind of information that OSPF transport instance can help to disseminate:

- o A network service is realized using one or more physical or virtualized hosts in MEC, and the locations of these service points might change. The auto-discovery of these service locations can be achieved using an OSPF transport instance.

- o UEs might be mobile, and MEC should support service continuity and application mobility. This may require service state transferring and synchronization. OSPF transport instance can be used to synchronize these states.
- o Network resources are limited, such as computing power, storage. The availability of such resources is dynamic, and OSPF transport instance can be used to populate such information, so applications can pick the right location of such resources, hence improve user experience and resource utilization.

3.2. Application Data Dissemination

Typically a network consists of routers from different vendors with different capabilities, and some applications may want to know whether a router supports certain functionality or where to find a router supports a functionality, so it will be ideal if such kind of information is known to all routers or a group of routers in the network. For example, an ingress router needs to find an egress router that supports In-situ Flow Information Telemetry (IFIT) [I-D.wang-lsr-igp-extensions-ifit] and obtain IFIT parameters.

OSPF transport instance can be used to populate such router capabilities/functionalities without impacting the performance or convergence of the base OSPF protocol.

3.3. Intra-Area Topology for BGP-LS Distribution

In some cases, it is desirable to limit the number of BGP-LS [RFC5572] sessions with a controller to the a one or two routers in an OSPF domain. However, many times those router(s) do not have full visibility to the complete topology of all the areas. To solve this problem without extended the BGP-LS domain, the OSPF LSAs for non-local area could be flooded over the OSPF transport instance topology using remote neighbors Section 4.7.1.

4. OSPF Transport Instance

In order to isolate the effects of flooding and processing of non-routing information, it will be relegated to a separate protocol instance. This instance should be given lower priority when contending for router resources including processing, backplane bandwidth, and line card bandwidth. How that is realized is an implementation issue and is outside the scope of this document.

Throughout the document, non-routing refers to routing information that is not used for IP or IPv6 routing calculations. The OSPF

transport instance is ideally suited for dissemination of routing information for other protocols and layers.

4.1. OSPFv2 Transport Instance Packet Differentiation

OSPFv2 currently does not offer a mechanism to differentiate Transport instance packets from normal instance packets sent and received on the same interface. However, the [RFC6549] provides the necessary packet encoding to support multiple OSPF protocol instances.

4.2. OSPFv3 Transport Instance Packet Differentiation

Fortunately, OSPFv3 already supports separate instances within the packet encodings. The existing OSPFv3 packet header instance ID field will be used to differentiate packets received on the same link (refer to section 2.4 in [RFC5340]).

4.3. Instance Relationship to Normal OSPF Instances

In OSPF transport instance, we must guarantee that any information we've received is treated as valid if and only if the router sending it is reachable. We'll refer to this as the "condition of reachability" in this document.

The OSPF transport instance is not dependent on any other OSPF instance. It does, however, have much of the same as topology information must be advertised to satisfy the "condition of reachability".

Further optimizations and coupling between an OSPF transport instance and a normal OSPF instance are beyond the scope of this document. This is an area for future study.

4.4. Network Prioritization

While OSPFv2 (section 4.3 in [RFC2328]) are normally sent with IP precedence Internetwork Control, any packets sent by an OSPF transport instance will be sent with IP precedence Flash (B'011'). This is only appropriate given that this is a pretty flashy mechanism.

Similarly, OSPFv3 transport instance packets will be sent with the traffic class mapped to flash (B'011') as specified in ([RFC5340]).

By setting the IP/IPv6 precedence differently for OSPF transport instance packets, normal OSPF routing instances can be given priority during both packet transmission and reception. In fact, some router

implementations map the IP precedence directly to their internal packet priority. However, internal router implementation decisions are beyond the scope of this document.

4.5. OSPF Transport Instance Omission of Routing Calculation

Since the whole point of the transport instance is to separate the routing and non-routing processing and fate sharing, a transport instance SHOULD NOT install any IP or IPv6 routes. OSPF routers SHOULD NOT advertise any transport instance LSAs containing IP or IPv6 prefixes and OSPF routers receiving LSAs advertising IP or IPv6 prefixes SHOULD ignore them. This implies that an OSPF transport instance Link State Database should not include any of the LSAs as shown in Table 1.

OSPFv2	summary-LSAs (type 3) AS-external-LSAs (type 5) NSSA-LSAs (type 7)
OSPFv3	inter-area-prefix-LSAs (type 2003) AS-external-LSAs (type 0x4005) NSSA-LSAs (type 0x2007) intra-area-prefix-LSAs (type 0x2009)
OSPFv3 Extended LSA	E-inter-area-prefix-LSAs (type 0xA023) E-as-external-LSAs (type 0xC025) E-Type-7-NSSA (type 0xA027) E-intra-area-prefix-LSA (type 0xA029)

LSAs not included in OSPF transport instance

If these LSAs are erroneously advertised, they will be flooded as per standard OSPF but MUST be ignored by OSPF routers supporting this specification.

4.6. Non-routing Instance Separation

It has been suggested that an implementation could obtain the same level of separation between IP routing information and non-routing information in a single instance with slight modifications to the OSPF protocol. The authors refute this contention for the following reasons:

- o Adding internal and external mechanisms to prioritize routing information over non-routing information are much more complex

than simply relegating the non-routing information to a separate instance as proposed in this specification.

- o The instance boundary offers much better separation for allocation of finite resources such as buffers, memory, processor cores, sockets, and bandwidth.
- o The instance boundary decreases the level of fate sharing for failures. Each instance may be implemented as a separate process or task.
- o With non-routing information, many times not every router in the OSPF routing domain requires knowledge of every piece of non-routing information. In these cases, groups of routers which need to share information can be segregated into sparse topologies greatly reducing the amount of non-routing information any single router needs to maintain.

4.7. Non-Routing Sparse Topologies

With non-routing information, many times not every router in the OSPF routing domain requires knowledge of every piece of non-routing information. In these cases, groups of routers which need to share information can be segregated into sparse topologies. This will greatly reduce the amount of information any single router needs to maintain with the core routers possibly not requiring any non-routing information at all.

With normal OSPF, every router in an OSPF area must have every piece of topological information and every intra-area IP or IPv6 prefix. With non-routing information, only the routers needing to share a set of information need be part of the corresponding sparse topology. For directly attached routers, one only needs to configure the desired topologies on the interfaces with routers requiring the non-routing information. When the routers making up the sparse topology are not part of a unconnected graph, two alternatives exist. The first alternative is configure tunnels to form a fully connected graph including only those routers in the sparse topology. The second alternative is use remote neighbors as described in Section 4.7.1.

4.7.1. Remote OSPF Neighbor

With sparse topologies, OSPF routers sharing non-routing information may not be directly connected. OSPF adjacencies with remote neighbors are formed exactly as they are with regular OSPF neighbors. The main difference is that a remote OSPF neighbor's address is configured and IP routing is used to deliver OSPF protocol packets to

the remote neighbor. Other salient feature of the remote neighbor include:

- o All OSPF packets have the remote neighbor's configured IP address as the IP destination address.
- o The adjacency is represented in the router Router-LSA as a router (type-1) link with the link data set to the remote neighbor's configured IP address.
- o Similar to NBMA networks, a poll-interval is configured to determine if the remote neighbor is reachable. This value is normally much higher than the hello interval with 40 seconds RECOMMENDED as the default.

4.8. Multiple Topologies

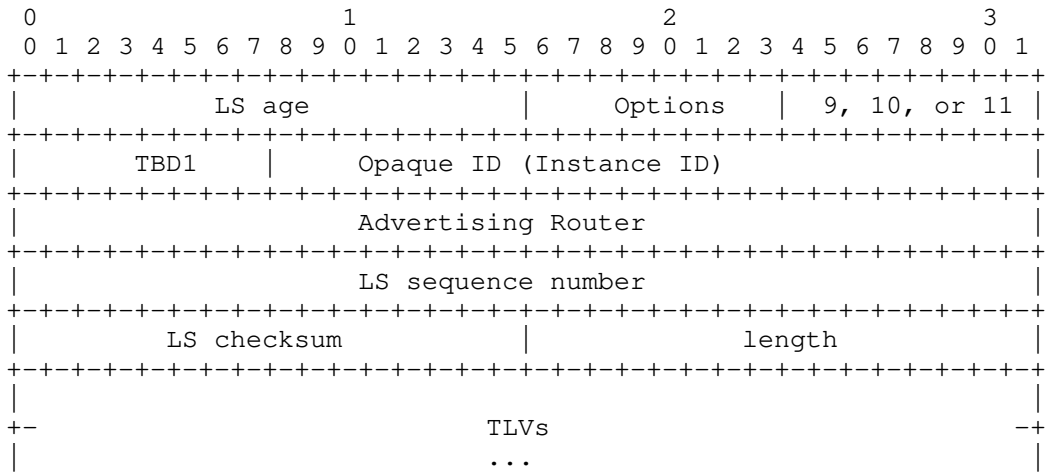
For some applications, the information need to be flooded only to a topology which is a subset of routers of the transport instance. This allows the application specific information only to be flooded to routers that support the application. A transport instance may support multiple topologies as defined in [RFC4915]. But as pointed out in Section 4.5, a transport instance or topology SHOULD NOT install any IP or IPv6 routes.

Each topology associated with the transport instance MUST be fully connected in order for the LSAs to be successfully flooded to all routers in the topology.

5. OSPF Transport Instance Information (TII) Encoding

5.1. OSPFv2 Transport Instance Information Encoding

Application specific information will be flooded in opaque LSAs as specified in [RFC5250]. An Opaque LSA option code will be reserved for Transport Instance Information (TII) as described in Section 8. The TII LSA can be advertised at any of the defined flooding scopes (link, area, or autonomous system (AS)).



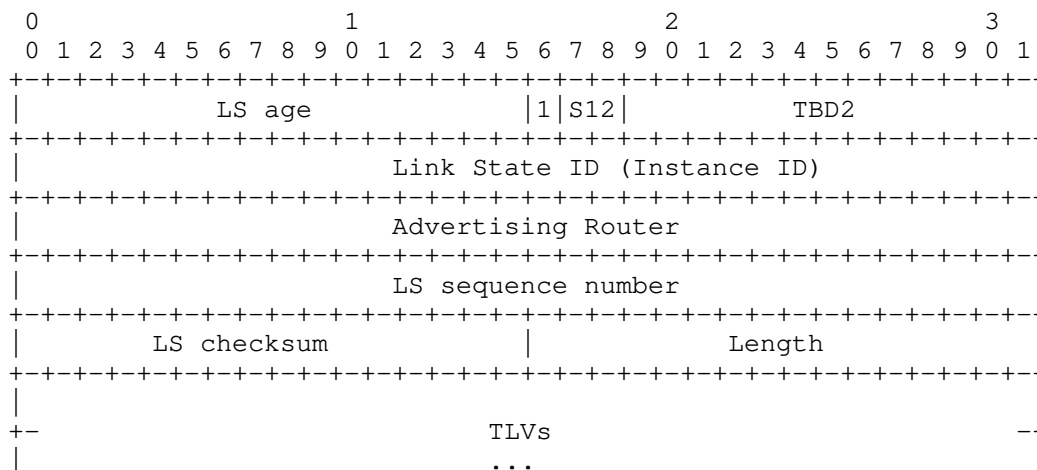
g

OSPFv2 Transport Instance Information Opaque LSA

The format of the TLVs within the body of an TII LSA is as defined in Section 5.3.

5.2. OSPFv3 Transport Instance Information Encoding

Application specific information will be flooded in separate LSAs with a separate function code. Refer to section A.4.2.1 of [RFC5340]. for information on the LS Type encoding in OSPFv3, and section 2 of [RFC8362] for OSPFv3 extended LSA types. An OSPFv3 function code will be reserved for Transport Instance Information (TII) as described in Section 8. Same as OSPFv2, the TII LSA can be advertised at any of the defined flooding scopes (link, area, or autonomous system (AS)). The U bit will be set indicating that OSPFv3 TTI LSAs should be flooded even if it is not understood.

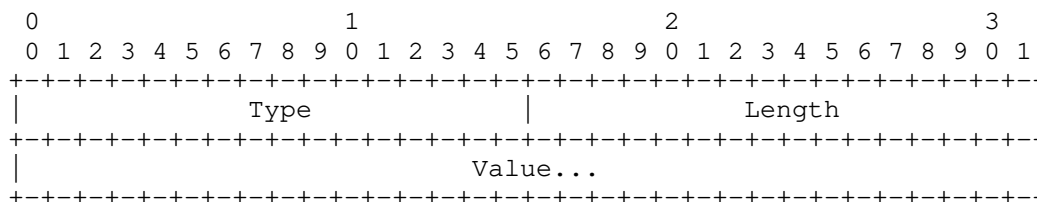


OSPFv3 Transport Instance Information LSA

The format of the TLVs within the body of an TII LSA is as defined in Section 5.3.

5.3. Transport Instance Information (TII) TLV Encoding

The format of the TLVs within the body of the LSAs containing non-routing information is the same as the format used by the Traffic Engineering Extensions to OSPF [RFC3630]. The LSA payload consists of one or more nested Type/Length/Value (TLV) triplets. The format of each TLV is:

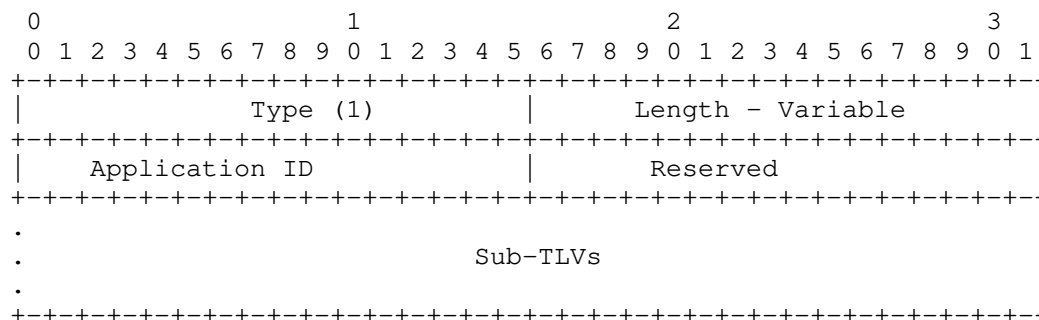


TLV Format

5.3.1. Top-Level TII Application TLV

An Application top-level TLV will be used to encapsulate application data advertised within TII LSAs. This top-level TLV may be used to handle the local publication/subscription for application specific

data. The details of such a publication/subscription mechanism are beyond the scope of this document. An Application ID is used in the top-level application TLV and shares the same code point with IS-IS as defined in [RFC6823].



Application ID:

An identifier assigned to this application via the IANA registry, as defined in RFC 6823. Each unique application will have a unique ID.

Additional Application-Specific Sub-TLVs:

Additional information defined by applications can be encoded as Sub-TLVs. Definition of such information is beyond the scope of this document.

Top-Level TLV

The specific TLVs and sub-TLVs relating to a given application and the corresponding IANA considerations MUST be specified in the document corresponding to that application.

6. Manageability Considerations

7. Security Considerations

The security considerations for the Transport Instance will not be different for those for OSPFv2 [RFC2328] and OSPFv3 [RFC5340].

8. IANA Considerations

8.1. OSPFv2 Opaque LSA Type Assignment

IANA is requested to assign an option type, TBD1, for Transport Instance Information (TII) LSA from the "Opaque Link-State Advertisements (LSA) Option Types" registry.

8.2. OSPFv3 LSA Function Code Assignment

IANA is requested to assign a function code, TBD2, for Transport Instance Information (TII) LSAs from the "OSPFv3 LSA Function Codes" registry.

8.3. OSPF Transport Instance Information Top-Level TLV Registry

IANA is requested to create a registry for OSPF Transport Instance Information (TII) Top-Level TLVs. The first available TLV (1) is assigned to the Application TLV Section 5.3.1. The allocation of the unsigned 16-bit TLV type are defined in the table below.

Range	Assignment Policy
0	Reserved (Not to be assigned)
1	Application TLV
2-16383	Unassigned (IETF Review)
16383-32767	Unassigned (FCFS)
32768-32777	Experimentation (No assignments)
32778-65535	Reserved (Not to be assigned)

TII Top-Level TLV Registry Assignments

9. Acknowledgement

The authors would like to thank Les Ginsberg for review and comments.

10. References

10.1. 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>>.

- [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>>.
- [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>>.
- [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, DOI 10.17487/RFC5250, July 2008, <<https://www.rfc-editor.org/info/rfc5250>>.
- [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>>.
- [RFC6549] Lindem, A., Roy, A., and S. Mirtorabi, "OSPFv2 Multi-Instance Extensions", RFC 6549, DOI 10.17487/RFC6549, March 2012, <<https://www.rfc-editor.org/info/rfc6549>>.
- [RFC6823] Ginsberg, L., Previdi, S., and M. Shand, "Advertising Generic Information in IS-IS", RFC 6823, DOI 10.17487/RFC6823, December 2012, <<https://www.rfc-editor.org/info/rfc6823>>.
- [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>>.

10.2. Informative References

- [ETSI-WP28-MEC]
Sami Kekki, etc., "MEC in 5G Networks", 2018, <https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf>.
- [I-D.wang-lsr-igp-extensions-ifit]
Wang, Y., Zhou, T., Qin, F., Chen, H., and R. Pang, "IGP Extensions for In-situ Flow Information Telemetry (IFIT) Capability Advertisement", draft-wang-lsr-igp-extensions-ifit-01 (work in progress), July 2020.

[RFC5572] Blanchet, M. and F. Parent, "IPv6 Tunnel Broker with the Tunnel Setup Protocol (TSP)", RFC 5572, DOI 10.17487/RFC5572, February 2010, <<https://www.rfc-editor.org/info/rfc5572>>.

Authors' Addresses

Acee Lindem
Cisco Systems
301 Midenhall Way
CARY, NC 27513
UNITED STATES

Email: acee@cisco.com

Yingzhen Qu
Futurewei
2330 Central Expressway
Santa Clara, CA 95050
USA

Email: yingzhen.qu@futurewei.com

Abhay Roy
Arrcus, Inc.

Email: abhay@arrcus.com

Sina Mirtorabi
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
USA

Email: smirtora@cisco.com

LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 19, 2021

W. Britto
S. Hegde
P. Kaneriya
R. Shetty
R. Bonica
Juniper Networks
P. Psenak
Cisco Systems
November 15, 2020

IGP Flexible Algorithms (Flex-Algorithm) In IP Networks
draft-bonica-lsr-ip-flexalgo-01

Abstract

An IGP Flexible Algorithm (Flex-Algorithm) allows IGP to compute constraint-based paths. As currently defined, IGP Flex-Algorithm is used with Segment Routing (SR) data planes - SR MPLS and SRv6. Therefore, Flex-Algorithm cannot be deployed in the absence of SR.

This document extends IGP Flex-Algorithm, so that it can be used for regular IPv4 and IPv6 prefixes. This allows Flex-Algorithm to be deployed in any IP network, even in the absence of SR.

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 May 19, 2021.

Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Requirements Language	3
3. Egress Node Procedures	3
4. Advertising Flex-Algorithm Definitions (FAD)	3
5. Advertising IP Flex-Algorithm Participation	3
5.1. The ISIS IP Algorithm Sub-TLV	4
5.2. The OSPF IP Algorithm TLV	5
6. Advertising IP Flex-Algorithm Reachability	6
6.1. The ISIS IPv4 Algorithm Prefix Reachability TLV	6
6.2. The ISIS IPv6 Algorithm Prefix Reachability TLV	8
6.3. The OSPFv2 Algorithm Prefix Reachability TLV	9
6.4. The OSPFv3 Flex-Algorithm IP Prefix Opaque LSA	11
7. Calculating of IP Flex-Algorithm Paths	11
8. IP Flex-Algorithm Forwarding	12
9. Deployment Considerations	12
10. IANA Considerations	13
11. Security Considerations	14
12. Acknowledgements	14
13. References	14
13.1. Normative References	14
13.2. Informative References	15
Authors' Addresses	16

1. Introduction

An IGP Flex-Algorithm as specified in [I-D.ietf-lsr-flex-algo] computes a constraint-based path to:

- o All Flex-Algorithm specific Prefix Segment Identifiers (SIDs) [RFC8402].
- o All Flex-Algorithm specific SRv6 Locators [I-D.ietf-spring-srv6-network-programming].

Therefore, Flex-Algorithm cannot be deployed in the absence of SR and SRv6.

This document extends Flex-Algorithm, allowing it to compute paths to:

- o An IPv4 [RFC0791] address.
- o An IPv6 [RFC8200] address.

This allows Flex-Algorithm to be deployed in any IP network, even in the absence of SR and SRv6.

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. Egress Node Procedures

Network operators configure multiple loopback interfaces on an egress node. They associate one or more IP addresses with each loopback interface and one Flex-Algorithm with each IP address.

If a packet is sent to a loopback address, and the loopback address is not associated with a Flex-Algorithm, the packet follows the IGP least-cost path to the egress node. If a packet is sent to a loopback address, and the loopback address is associated with a Flex-Algorithm, the packet follows the constraint-base path that the Flex-Algorithm calculated.

4. Advertising Flex-Algorithm Definitions (FAD)

To guarantee loop free forwarding, all routers that participate in a Flex-Algorithm MUST agree on the Flex-Algorithm Definition (FAD).

Selected nodes within the IGP domain MUST advertise FADs as described in Sections 5, 6 and 7 of [I-D.ietf-lsr-flex-algo].

5. Advertising IP Flex-Algorithm Participation

A node may use various algorithms when calculating paths to nodes and prefixes. Algorithm values are defined in the IGP Algorithm Type Registry [IANA-ALG].

A node MUST participate in a Flex-Algorithm to be:

- o able to compute path for such Flex-Algorithm

- o be part of the topology for such Flex-Algorithm

Flex-Algorithm participation MUST be advertised for each Flex-Algorithm application independently, as specified in Section 10.2 of [I-D.ietf-lsr-flex-algo]. Using Flex-Algorithm for regular IPv4 and IPv6 prefixes represents a new Flex-Algorithm application (IP Flex-Algorithm), and as such the Flex-Algorithm participation for the IP Flex-Algorithm application MUST be signalled independently of any other Flex-Algorithm applications (e.g. SR).

Following sections describe how the IP Flex-Algorithm participation is advertised in IGP protocols.

5.1. The ISIS IP Algorithm Sub-TLV

The ISIS IP Algorithm Sub-TLV is a sub-TLV of the ISIS Router Capability TLV [RFC7981] and has the following format:

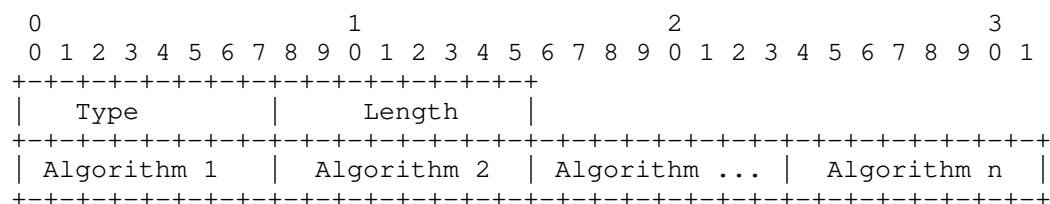


Figure 1: ISIS IP Algorithm Sub-TLV

- o Type: IP Algorithm Sub-TLV (Value TBD by IANA)
- o Length: Variable
- o Algorithm (1 octet): value from 1 to 255.

The IP Algorithm Sub-TLV MUST be propagated throughout the level and MUST NOT be advertised across level boundaries. Therefore, the S bit in the Router Capability TLV, in which the IP Algorithm Sub-TLV is advertised, MUST NOT be set.

The IP Algorithm Sub-TLV is optional. It MUST NOT be advertised more than once at a given level. A router receiving multiple IP Algorithm sub-TLVs from the same originator SHOULD select the first advertisement in the lowest-numbered LSP and subsequent instances of the IP Algorithm Sub-TLV MUST be ignored.

The IP Algorithm Sub-TLV advertises the participation in Flex-Algorithms, and MUST NOT impact the router participation in default

algorithm 0. The IP Algorithm Sub-TLV could be used to advertise support for non-zero standard algorithms, but that is outside the scope of this document.

The IP Flex-Algorithm participation advertised in ISIS IP Algorithm Sub-TLV is topology independent. When a router advertises participation in ISIS IP Algorithm Sub-TLV, the participation applies to all topologies in which the advertising node participates.

5.2. The OSPF IP Algorithm TLV

The OSPF IP Algorithm TLV is a top-level TLV of the Router Information Opaque LSA [RFC7770] and has the following format:

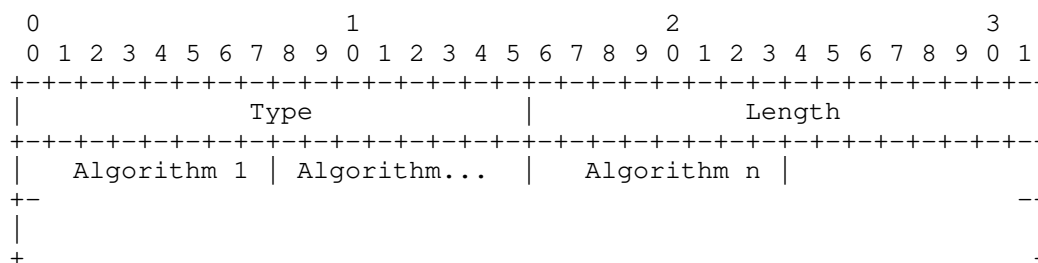


Figure 2: OSPF IP Algorithm TLV

- o Type: IP Algorithm TLV (Value TBD by IANA)
- o Length: Variable
- o Algorithm (1 octet): value from 1 to 255.

The IP Algorithm TLV is optional. It SHOULD only be advertised once in the Router Information Opaque LSA.

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

The IP Algorithm TLV advertises the participation in Flex-Algorithms, and MUST NOT impact the router participation in default algorithm 0. The IP Algorithm TLV could be used to advertise support for non-zero standard algorithms, but that is outside the scope of this document.

The IP Flex-Algorithm participation advertised in OSPF IP Algorithm TLV is topology independent. When a router advertises participation in OSPF IP Algorithm TLV, the participation applies to all topologies in which the advertising node participates.

6. Advertising IP Flex-Algorithm Reachability

To be able to associate the prefix with the Flex-Algorithm, the existing prefix reachability advertisements can not be used, because they advertise the prefix reachability in default algorithm 0. Instead, a new IP Flex-Algorithm reachability advertisements are defined in ISIS and OSPF.

Two new top-level TLVs are defined in ISIS [ISO10589] to advertise prefix reachability associated with a Flex-Algorithm.

- o The IPv4 Algorithm Prefix Reachability TLV
- o The IPv6 Algorithm Prefix Reachability TLV

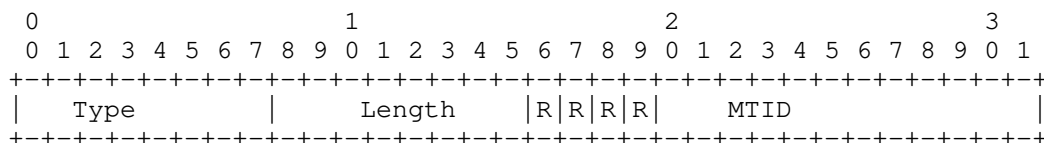
New top-level TLV of OSPFv2 Extended Prefix Opaque LSA [RFC7684] is defined to advertise prefix reachability associated with a Flex-Algorithm in OSPFv2.

6.1. The ISIS IPv4 Algorithm Prefix Reachability TLV

A new top level TLV is defined for advertising IPv4 Flex-Algorithm Prefix Reachability in ISIS - IPv4 Algorithm Prefix Reachability TLV.

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

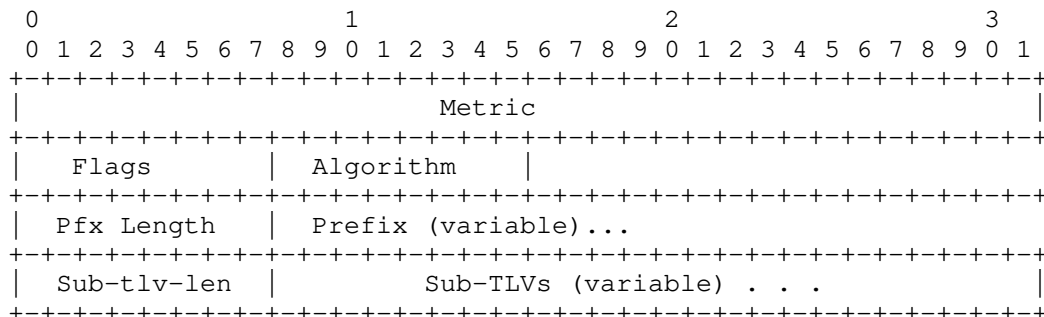
The ISIS IPv4 Algorithm Prefix Reachability TLV has the following format:



ISIS IPv4 Algorithm Prefix Reachability TLV

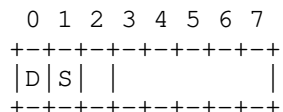
- o Type: IPv4 Algorithm Prefix Reachability TLV (Value TBD by IANA).
- o Length: variable.
- o R bits (4 bits): reserved for future use. They MUST be set to zero on transmission and MUST be ignored on receipt.
- o MTID (12 bits): Multitopology Identifier as defined in [RFC5120]. Note that the value 0 is legal.

Followed by one or more prefix entries of the form:



ISIS IPv4 Algorithm Prefix Reachability TLV

- o Metric (4 octets): Metric information.
- o Flags (1 octet):



D-flag: When the Prefix is leaked from level-2 to level-1, the D bit MUST be set. Otherwise, this bit MUST be clear. Prefixes with the D bit set MUST NOT be leaked from level-1 to level-2. This is to prevent looping.

S-flag: Set when Sub-TLVs are present for the prefix entry.

- o Algorithm (1 octet): Associated Algorithm from 1 to 255.
- o Prefix Len (1 octet): Prefix length measured in bits.
- o Prefix (variable length): Prefix mapped to Flex-Algorithm.
- o Optional Sub-TLV-length (1 octet): Number of octets used by sub-TLVs
- o Optional sub-TLVs (variable length).

A router receiving multiple IPv4 Algorithm Prefix Reachability advertisements for the same prefix, from the same originator, each with a different Algorithm, MUST select the first advertisement in the lowest-numbered LSP and ignore any subsequent IPv4 Algorithm Prefix Reachability advertisements for the same prefix for any other Algorithm.

A router receiving multiple IPv4 Algorithm Prefix Reachability advertisements for the same prefix, from different originators, each with a different Algorithm, MUST ignore all of them and MUST NOT install any forwarding entries based on these advertisements.

In cases where a prefix advertisement is received in both a IPv4 Prefix Reachability TLV and an IPv4 Algorithm Prefix Reachability TLV, the IPv4 Prefix Reachability advertisement MUST be preferred when installing entries in the forwarding plane.

6.2. The ISIS IPv6 Algorithm Prefix Reachability TLV

The ISIS IPv6 Algorithm Prefix Reachability TLV is identical to the ISIS IPv4 Algorithm Prefix Reachability TLV, except that it has a unique type. The type is TBD by IANA.

A router receiving multiple IPv6 Algorithm Prefix Reachability advertisements for the same prefix, from the same originator, each with a different Algorithm, MUST select the first advertisement in the lowest-numbered LSP and ignore any subsequent IPv6 Algorithm Prefix Reachability advertisements for the same prefix for any other Algorithm.

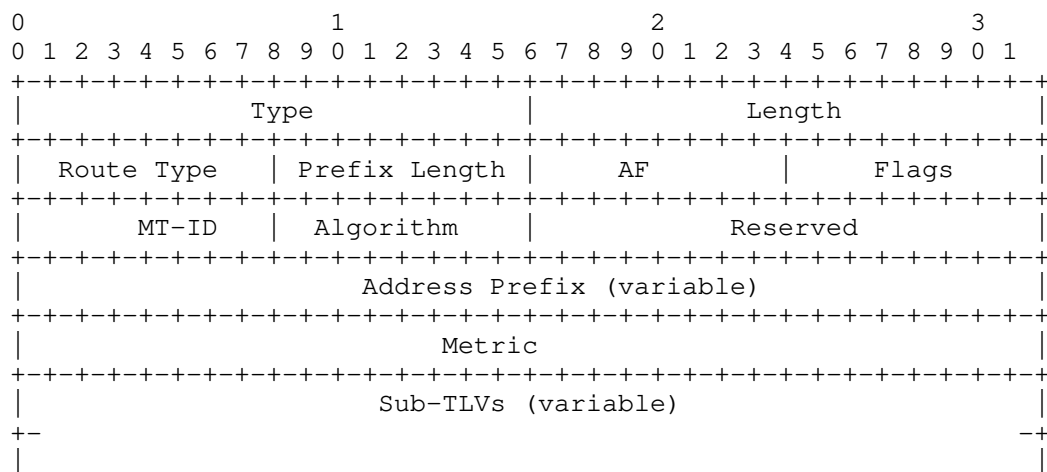
A router receiving multiple IPv6 Algorithm Prefix Reachability advertisements for the same prefix, from different originators, each with a different Algorithm, MUST ignore all of them and MUST NOT install any forwarding entries based on these advertisements.

In cases where a prefix advertisement is received in both a IPv6 Prefix Reachability TLV and an IPv6 Algorithm Prefix Reachability TLV, the IPv6 Prefix Reachability advertisement MUST be preferred when installing entries in the forwarding plane.

6.3. The OSPFv2 Algorithm Prefix Reachability TLV

A new top level TLV of OSPFv2 Extended Prefix Opaque LSA is defined for advertising IPv4 Algorithm Prefix Reachability in OSPFv2 - OSPF Algorithm Prefix Reachability TLV

Multiple Algorithm Prefix Reachability TLV MAY be advertised in each OSPFv2 Extended Prefix Opaque LSA. However, since the opaque LSA type defines the flooding scope, the LSA flooding scope MUST satisfy the application specific requirements for all the prefixes included in a single OSPFv2 Extended Prefix Opaque LSA. The Algorithm Prefix Reachability TLV has the following format:



OSPFv2 Algorithm Prefix Reachability TLV

Type: Algorithm Prefix Reachability TLV (Value TBD by IANA).

Length: Variable dependent on sub-TLVs.

Route Type (1 octet): type of the OSPF route. Supported types are:

1 - Intra-Area

- 2 - Inter-Area
- 3 - AS External with Type-1 Metric
- 4 - AS External with Type-2 Metric
- 5 - NSSA External with Type-1 Metric
- 6 - NSSA External with Type-2 Metric

Prefix Length (1 octet): Length of prefix in bits.

AF (1 octet): Address family for the prefix. Currently, the only supported value is 0 for IPv4 unicast. The inclusion of address family in this TLV allows for future extension.

Flags (1 octet): Flags applicable to the prefix. Supported Flags include:

0x80 - A-Flag (Attach flag): An Area Border Router (ABR) generating an Extended Prefix TLV for inter-area prefix that is locally connected or attached in other connected area SHOULD set this flag.

0x40 - N-Flag (Node Flag): Set when the prefix identifies the advertising router i.e., the prefix is a host prefix advertising a globally reachable address typically associated with a loopback address. The advertising router MAY choose to not set this flag even when the above conditions are met. If the flag is set and the prefix length is not a host prefix then the flag MUST be ignored. The flag is preserved when the OSPFv2 Extended Prefix Opaque LSA is propagated between areas.

MT-ID (1 octet): Multi-Topology ID as defined in [RFC8402]

Algorithm: (1 octet). Associated Algorithm from 1 to 255.

Address Prefix: For the address family IPv4 unicast, the prefix itself encoded as a 32-bit value. The default route is represented by a prefix of length 0. Prefix encoding for other address families is beyond the scope of this specification.

Metric (4 octets): Metric information.

If this TLV is advertised multiple times for the same prefix in the same OSPFv2 Extended Prefix Opaque LSA, only the first instance of the TLV is used by receiving OSPFv2 Routers. This situation SHOULD be logged as an error.

If this TLV is advertised multiple times for the same prefix in different OSPFv2 Extended Prefix Opaque LSAs originated by the same OSPF router, the OSPF advertising router is re-originating Extended Prefix Opaque LSAs for multiple prefixes and is most likely repacking Algorithm Prefix Reachability TLVs in Extended Prefix Opaque LSAs. In this case, the Algorithm Prefix Reachability TLV in the Extended Prefix Opaque LSA with the smallest Opaque ID is used by receiving OSPFv2 Routers. This situation may be logged as a warning.

It is RECOMMENDED that OSPF routers advertising Algorithm Prefix Reachability TLVs in different Extended Prefix Opaque LSAs re-originate these LSAs in ascending order of Opaque ID to minimize the disruption.

A router receiving multiple Algorithm Prefix Reachability TLVs for the same prefix, from different originators, each with a different Algorithm, MUST ignore all of them and MUST NOT install any forwarding entries based on these advertisements.

In cases where a prefix advertisement is received in any of the LSAs advertising the prefix reachability for algorithm 0 (Router-LSA, Summary-LSA, AS-external-LSA or NSSA AS-external LSA) and in an IPv4 Algorithm Prefix Reachability TLV, the prefix reachability advertisement for algorithm 0 MUST be preferred when installing entries in the forwarding plane, regardless of the Route Type advertised in IPv4 Algorithm Prefix Reachability TLV.

6.4. The OSPFv3 Flex-Algorithm IP Prefix Opaque LSA

TBD.

7. Calculating of IP Flex-Algorithm Paths

IP Flex-Algorithm is considered as yet another application of the Flex-Algorithm as described in Section 10 and Section 12 of the [I-D.ietf-lsr-flex-algo].

Participation for the IP Flex-Algorithm is signalled as described in Section 5 and is specific to the IP Flex-Algorithm application.

Calculation of IP Flex-Algorithm paths follows the Section 12 of [I-D.ietf-lsr-flex-algo]. This computation uses the IP Flex-Algorithm participation and is independent of the Flex-Algorithm calculation done for any other Flex-Algorithm applications (e.g. SR, SRv6).

IP Flex-Algorithm application only considers participating nodes during the Flex-Algorithm calculation. When computing paths for a

given Flex-Algorithm, all nodes that do not advertise participation for IP Flex-Algorithm, as described in Section 5, MUST be pruned from the topology.

8. IP Flex-Algorithm Forwarding

IP Algorithm Prefix Reachability advertisement as described in Section 5 includes the MTID value that associates the prefix with a specific topology. Algorithm Prefix Reachability advertisement also includes an Algorithm value that explicitly associates the prefix with a specific Flex-Algorithm. The paths to the prefix MUST be calculated using the specified Flex-Algorithm in the associated topology.

Forwarding entries for the IP Flex-Algorithm prefixes advertised in IGP MUST be installed in the forwarding plane of the receiving IP Flex-Algorithm prefix capable routers when they participate in the associated topology and algorithm. Forwarding entries for IP Flex-Algorithm prefixes associated with Flex-Algorithms in which the node is not participating MUST NOT be installed in the forwarding plane.

When the IP Flex-Algorithm prefix is associated with a Flex-Algorithm, LFA paths to the 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.

9. Deployment Considerations

IGP Flex-Algorithm can be used by many applications. Original specification was done for SR and SRv6, this specification adds IP as another application that can use IGP Flex-Algorithm. Other applications may be defined in the future. This section provides some details about the coexistence of the various applications of the IGP Flex-Algorithm.

Flex-Algorithm definition (FAD), as described in [I-D.ietf-lsr-flex-algo], is application independent and is used by all Flex-Algorithm applications.

Participation in the Flex-Algorithm, as described in [I-D.ietf-lsr-flex-algo], is application specific.

Calculation of the flex-algo paths is application specific and uses application specific participation advertisements.

Application specific participation and calculation guarantee that the forwarding of the traffic over the Flex-Algorithm application

specific paths is consistent between all nodes over which the traffic is being forwarded.

Multiple application can use the same Flex-Algorithm value at the same time and as such share the FAD for it. For example SR-MPLS and IP can both use such common Flex-Algorithm. Traffic for SR-MPLS will be forwarded based on Flex-algorithm specific SR SIDs. Traffic for IP Flex-Algorithm will be forwarded based on Flex-Algorithm specific prefix reachability announcements.

10. IANA Considerations

This specification updates the OSPF Router Information (RI) TLVs Registry as follows:

Value	TLV Name	Reference
TBD	IP Algorithm TLV	This Document Section 5.2

This document also updates the "Sub-TLVs for TLV 242" registry as follows:

Value	TLV Name	Reference
TBD	IP Algorithm Sub-TLV	This Document Section 5.1

This document also updates the "ISIS TLV Codepoints Registry" registry as follows:

Value	TLV Name	Reference
TBD	IPv4 Algorithm Prefix Reachability TLV	This document, Section 6.1
TBD	IPv6 Algorithm Prefix Reachability TLV	This document, Section 6.2
TBD		

This document updates the "OSPFv2 Extended Prefix Opaque LSA TLVs" registry as follows::

Value	TLV Name	Reference
TBD	OSPFv2 Algorithm Prefix Reachability TLV	This Document, Section 6.1

11. Security Considerations

TBD

12. Acknowledgements

TBD.

13. References

13.1. 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-13 (work in progress), October 2020.

[ISO10589]

IANA, "Intermediate system to Intermediate system routing information exchange protocol for use in conjunction with the Protocol for providing the Connectionless-mode Network Service (ISO 8473)", August 1987, <ISO/IEC 10589:2002>.

[RFC0791]

Postel, J., "Internet Protocol", STD 5, RFC 791, DOI 10.17487/RFC0791, September 1981, <<https://www.rfc-editor.org/info/rfc791>>.

[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>>.

[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>>.

- [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>>.
- [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>>.
- [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>>.
- [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>>.
- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.

13.2. Informative References

[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-24 (work in progress), October 2020.

[IANA-ALG]

IANA, "Sub-TLVs for TLV 242 (IS-IS Router CAPABILITY TLV)", August 1987, <<https://www.iana.org/assignments/igp-parameters/igp-parameters.xhtml#igp-algorithm-types>>.

[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

William Britto
Juniper Networks
Elnath-Exora Business Park Survey
Bangalore, Karnataka 560103
India

Email: bwilliam@juniper.net

Shraddha Hegde
Juniper Networks
Elnath-Exora Business Park Survey
Bangalore, Karnataka 560103
India

Email: shraddha@juniper.net

Parag Kaneriya
Juniper Networks
Elnath-Exora Business Park Survey
Bangalore, Karnataka 560103
India

Email: pkaneria@juniper.net

Rejesh Shetty
Juniper Networks
Elnath-Exora Business Park Survey
Bangalore, Karnataka 560103
India

Email: mrjesh@juniper.net

Ron Bonica
Juniper Networks
2251 Corporate Park Drive
Herndon, Virginia 20171
USA

Email: rbonica@juniper.net

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

Email: ppsenak@cisco.com

Link State Routing Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 2, 2021

Y. Wang
T. Zhou
Z. Hu
Huawei
October 29, 2020

IGP Extensions for Advertising Hop-by-Hop Options Header Processing
Action
draft-wang-lsr-hbh-process-00

Abstract

This document extends Node and Link attribute TLVs to Interior Gateway Protocols (IGP) to advertise the Hop-by-Hop Options header processing action and supported services (e.g. IOAM Trace Option and Alternate Marking) at node and link granularity. Such advertisements allow entities (e.g., centralized controllers) to determine whether the Hop-by-Hop Options header and specific services can be supported in a given network.

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 May 2, 2021.

Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Terminology	3
3. Hop-by-Hop Options Header Processing Action	4
4. Signaling Processing Action Using IS-IS	5
4.1. IS-IS Node Processing-Action Sub-TLV	5
4.2. IS-IS Link Processing-Action Sub-TLV	6
5. Signaling Processing Action Using OSPF	6
5.1. OSPF Node Processing-Action TLV	7
5.2. OSPFv2 Link Processing-Action sub-TLV	7
5.3. OSPFv3 Link Processing-Action Sub-TLV	8
6. IANA Considerations	8
7. Security Considerations	9
8. Acknowledgements	9
9. References	9
9.1. Normative References	9
9.2. Informative References	10
Authors' Addresses	10

1. Introduction

[RFC8200] specifies IPv6 extension headers, including Hop-by-Hop Options header, Destination Options header, Routing header, etc. An IPv6 packet may carry zero, one, or more extension headers that must be processed strictly in the order they appear in the packet. Except for the Hop-by-Hop Options header, other extension headers are not processed, inserted, or deleted by any transit node along a packet's delivery path, until the packet arrives at the Destination node.

As specified in [RFC8200], although the Hop-by-Hop Options header is not inserted or deleted by any transit node along a packet's delivery path, it is only examined and processed by nodes along a packet's

delivery path if they are explicitly configured to process. Besides, nodes may be configured to ignore the Hop-by-Hop Options header, drop packets containing a Hop-by-Hop Options header, or assign packets containing a Hop-by-Hop Options header to a slow processing path. In the results, devices of different vendors can be configured to process the Hop-by-Hop Options header in different ways.

Until now, the Hop-by-Hop Options header has been widely used. In-situ Operations, Administration, and Maintenance (IOAM) data fields are encapsulated in two types of extension headers in IPv6 packets, either Hop-by-Hop Options header or Destination Options header, depending on IOAM usage [I-D.ietf-ippm-ioam-ipv6-options]. For example, IOAM-tracing options are represented as an IPv6 options in Hop-by-Hop extension header. Similarly, the Alternate Marking technique can be carried by the Hop-by-Hop Options header and the Destination Options header [I-D.ietf-6man-ipv6-alt-mark]. If nodes are not explicitly configured to process the Hop-by-Hop Option header, they should ignore them. In this case, the performance measurement does not account for all links and nodes along a path. Therefore, such advertisement can be useful for entities (e.g., the centralized controller) to determine whether a specific service can be implemented in IPv6 network by encoding in the Hop-by-Hop Options header.

BGP-LS defines a way to advertise topology and associated attributes and capabilities of the nodes in that topology to a centralized controller [RFC7752]. Typically, BGP-LS is configured on a small number of nodes that do not necessarily act as head-ends. In order for BGP-LS to signal the processing action of the Hop-by-Hop Options header for all the devices in the network, the processing action SHOULD be advertised by every IGP router in the network.

This document defines a mechanism to signal the configured processing action of the Hop-by-Hop Options header and supported services at node and/or link granularity using IS-IS, OSPFv2 and OSPFv3.

2. Terminology

Following are abbreviations used in this document:

- o IGP: Interior Gateway Protocols
- o IS-IS: Intermediate System to Intermediate System
- o OSPF: Open Shortest Path First
- o BGP-LS: Border Gateway Protocol - Link State

- o NLRI: Network Layer Reachability Information

3. Hop-by-Hop Options Header Processing Action

This document defines the information of processing action formed of a tuple of a 1-octet Extension Header Options identifier and 8-bit Processing Action Flag.

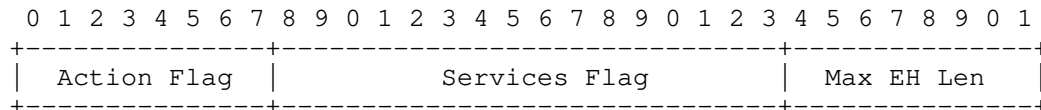


Fig. 1 Processing Action Format

Where:

- o Action Flag: A 8-bit field. The highest-order 3-bit indicates the processing action, i.e., 000 - drop packets; 001 - dispatch to control plane; 010 - forward, skip to Next Header; 011 - forward, ignoring all extension Options header; 100 - examine and process.
- o Max EH Len: A one octet field. The maximum length of the Extension Header in 8-octet units can be examined and processed at node or link granularity. The definition is same as the Next Header Length in [RFC8200].
- o Services Flag: A 16-bit bitmap. The format is defined as follows.

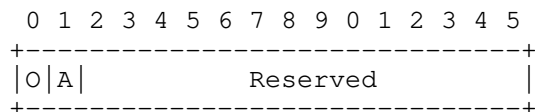


Fig. 2 Services Flag Format

where fields are defined as the following:

- o O (IOAM Trace Option) is a one-bit flag. The O flag is set to 1 if the IOAM Trace Option is supported at node or link granularity.
- o A (Alternate Marking) is a one-bit flag. The A flag is set to 1 if the Alternate Marking method is supported at node or link granularity.
- o R - reserved bits for future use. These flags MUST be zeroed on transmit and ignored on receipt.

In this document, the processing action at link granularity is defined as the supported the Hop-by-Hop Options header processing action of the interface associated with the link. When all interfaces associated with links support the same processing action, the processing action at node granularity SHOULD represent the Link processing action. Both of Node and Link processing action information are formed of a tuple of a 1-octet Extension Header Options identifier and 8-bit Processing Action Flag.

When both of Node and Link processing action are advertised, the Link processing action information MUST take precedence over the Node processing action. Besides, when a Link processing action is not signaled, then the Node processing action SHOULD be considered to be the processing action for this link.

4. Signaling Processing Action Using IS-IS

The IS-IS Extensions for advertising Router Information TLV named IS-IS Router CAPABILITY TLV [RFC7981], which allows a router to announce its capabilities within an IS-IS level or the entire routing domain.

4.1. IS-IS Node Processing-Action Sub-TLV

According to the format of IS-IS Router CAPABILITY TLV [RFC7981], the Node Processing-Action sub-TLV within the body of the IS-IS router CAPABILITY TLV is composed of three fields, a one-octet Type field, a one-octet Length field, and a 4-octet Value field. The following format is used:

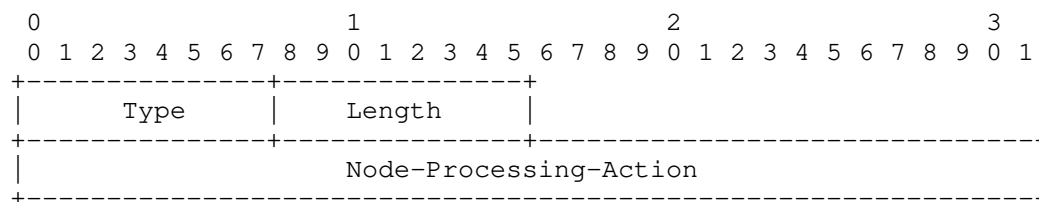


Fig. 3 IS-IS Node Processing-Action Sub-TLV Format

Where:

- o Type: To be assigned by IANA
- o Length: A one-octet field that indicates the length of the value portion in octets.
- o Node-Processing-Action: A 4-octet field, which is same as defined in Section 3.

4.2. IS-IS Link Processing-Action Sub-TLV

The Link Processing-Action sub-TLV is defined for TLVs 22, 23, 25, 141, 222, and 223 to carry the Processing-Action information of the interface associated with the link. The Link Processing-Action sub-TLV is formed of three fields, a one-octet Type field, a one-octet Length field, and a 4-octet Value field. The following format is used:

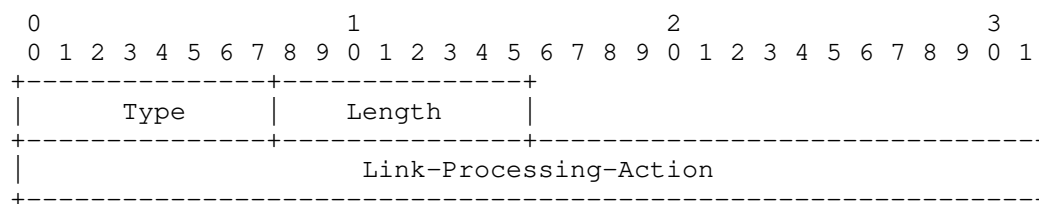


Fig. 4 IS-IS Link Processing-Action Sub-TLV Format

Where:

- o Type: To be assigned by IANA
- o Length: A one-octet field that indicates the length of the value portion in octets.
- o Link-Processing-Action: A 4-octet field, which is same as defined in Section 3.

5. Signaling Processing Action Using OSPF

Given that OSPF uses the options field in LSAs and hello packets to advertise optional router capabilities [RFC7770], this document defines a new Node Processing-Action TLV within the body of the OSPF RI Opaque LSA [RFC7770] to carry the Processing Action of the router originating the RI LSA.

This document defines the Link Processing-Action sub-TLV to carry the Processing-Action information of the interface associated with the link. For OSPFv2, the link-level Processing-Action information is advertised as a sub-TLV of the OSPFv2 Extended Link TLV as defined in [RFC7684]. For OSPFv3, the link-level Processing-Action information is advertised a sub-TLV of the E-Router-LSA TLV as defined in [RFC8362].

5.1. OSPF Node Processing-Action TLV

The Node Processing-Action TLV is composed of three fields, a 2-octet Type field, a 2-octet Length field, and a 4-octet Value field. The following format is used:

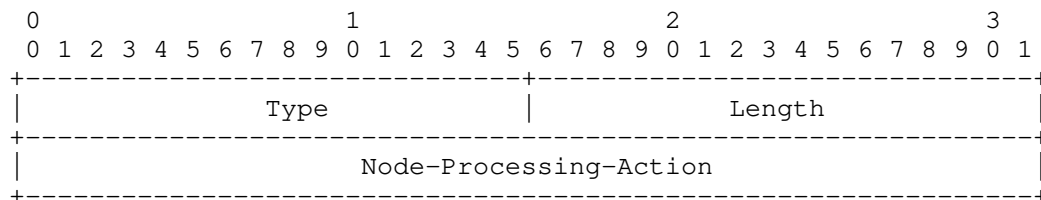


Fig. 5 OSPF Node Processing-Action TLV

Where:

- o Type: To be assigned by IANA
- o Length: A 2-octet field that indicates the length of the value field.
- o Node-Processing-Action: A 4-octet field, which is as defined in Section 3.

5.2. OSPFv2 Link Processing-Action sub-TLV

The Link Processing-Action sub-TLV encoded in the OSPFv2 Extended Link TLV as defined in [RFC7684], which is constructed of three fields, a 2-octet Type field, a 2-octet Length field, and a 4-octet Value field. The following format is used:

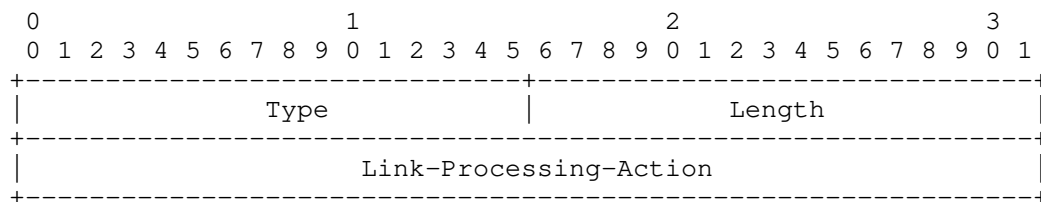


Fig. 6 OSPFv2 Link Processing-Action sub-TLV

Where:

- o Type: To be assigned by IANA

- o Length: A 2-octet field that indicates the length of the value field.
- o Link-Processing-Action: A 4-octet field, which is as defined in Section 3.

5.3. OSPFv3 Link Processing-Action Sub-TLV

The Link Processing-Action sub-TLV encoded in the OSPFv3 E-Router-LSA TLV as defined in [RFC8362], which is constructed of three fields, a 2-octet Type field, a 2-octet Length field, and a 4-octet Value field. The following format is used:

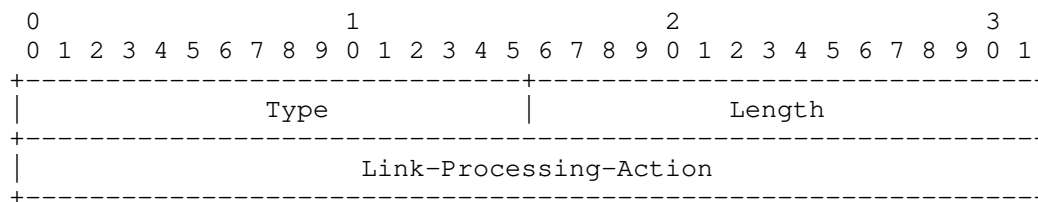


Fig. 7 OSPFv3 Link Processing-Action sub-TLV

Where:

- o Type: To be assigned by IANA
- o Length: A 2-octet field that indicates the length of the value field.
- o Link-Processing-Action: A 4-octet field, which is as defined in Section 3.

6. IANA Considerations

IANA is requested to allocate values for the following new TLV and sub-TLVs.

Type	Description
TBA	IS-IS Node Processing-Action Sub-TLV
TBA	IS-IS Link Processing-Action Sub-TLV

Type	Description
TBA	OSPF Node Processing-Action TLV
TBA	OSPFv2 Link Processing-Action sub-TLV
TBA	OSPFv3 Link Processing-Action sub-TLV

7. Security Considerations

This document introduces new IGP Node and Link Attribute TLVs and sub-TLVs for advertising processing actions of the Hop-by-Hop Options header at node and/or link granularity. It does not introduce any new security risks to IS-IS, OSPFv2 and OSPFv3.

8. Acknowledgements

TBD

9. References

9.1. 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>>.
- [RFC7684] "OSPFv2 Prefix/Link Attribute Advertisement", <<https://www.rfc-editor.org/info/rfc7684>>.
- [RFC7752] "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", <<https://datatracker.ietf.org/doc/rfc7752/>>.
- [RFC7770] "Extensions to OSPF for Advertising Optional Router Capabilities", <<https://www.rfc-editor.org/info/rfc7770>>.
- [RFC7981] "IS-IS Extensions for Advertising Router Information", <<https://www.rfc-editor.org/info/rfc7981>>.
- [RFC8200] "Internet Protocol, Version 6 (IPv6) Specification", <<https://datatracker.ietf.org/doc/rfc8200/>>.
- [RFC8362] "OSPFv3 Link State Advertisement (LSA) Extensibility", <<https://www.rfc-editor.org/info/rfc8362>>.

9.2. Informative References

- [I-D.ietf-6man-ipv6-alt-mark]
"IPv6 Application of the Alternate Marking Method",
<<https://datatracker.ietf.org/doc/draft-ietf-6man-ipv6-alt-mark/>>.
- [I-D.ietf-ippm-ioam-ipv6-options]
"In-situ OAM IPv6 Options",
<<https://datatracker.ietf.org/doc/draft-ietf-ippm-ioam-ipv6-options/>>.

Authors' Addresses

Yali Wang
Huawei
156 Beiqing Rd., Haidian District
Beijing
China

Email: wangyalil1@huawei.com

Tianran Zhou
Huawei
156 Beiqing Rd., Haidian District
Beijing
China

Email: zhoutianran@huawei.com

Zhibo Hu
Huawei
156 Beiqing Rd., Haidian District
Beijing
China

Email: huzhibo@huawei.com

LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 13, 2022

A. Wang
China Telecom
Z. Hu
Huawei Technologies
G. Mishra
Verizon Inc.
J. Sun
ZTE Corporation
July 12, 2021

Passive Interface Attribute
draft-wang-lsr-passive-interface-attribute-08

Abstract

This document describes the mechanism that can be used to differentiate the passive interfaces from the normal interfaces within ISIS or OSPF domain.

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 13, 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

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

include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Conventions used in this document	3
3. Consideration for flagging passive interface	3
4. Passive Interface Attribute	4
4.1. OSPFv2 Extended Stub-Link TLV	4
4.2. OSPFv3 Router-Stub-Link TLV	5
4.3. ISIS Stub-link TLV	6
4.4. Stub-Link Prefix Sub-TLV	7
5. Security Considerations	8
6. IANA Considerations	8
7. Acknowledgement	9
8. References	9
8.1. Normative References	9
8.2. Informative References	10
Authors' Addresses	11

1. Introduction

Passive interfaces are used commonly within an operators enterprise or service provider networks. One of the most common use cases for passive interface is in a data center Layer 2 and Layer 3 Top of Rack(TOR) switch where the inter connected links between the TOR switches and uplinks to the Core switch are only a few links and a majority of the links are Layer 3 VLAN switched virtual interface trunked between the TOR switches serving Layer 2 broadcast domains. In this scenario all the VLANs are made passive as it is recommended to limit the number of network LSAs between routers and switches to avoid unnecessary hello processing overhead.

Another common use case is an inter-as routing scenario where the same routing protocol but different IGP instance is running between the adjacent BGP domains. Using passive interface on the inter-as connections can ensure that prefixes contained within a domain are only reachable within the domain itself and not allow the link state database to be merged between domain which could result in undesirable consequences.

For operator which runs different IGP domains that interconnect with each other via the passive interfaces, there is desire to obtain the inter-as topology information as described in [I-D.ietf-idr-bgpls-inter-as-topology-ext]. If the router that runs BGP-LS within one IGP domain can distinguish passive interfaces from

other normal interfaces, it is then easy for the router to report these passive links using BGP-LS to a centralized PCE controller.

Draft [I-D.dunbar-lsr-5g-edge-compute-ospf-ext] describes the case that edge compute server attach the network and needs to flood some performance index information to the network to facilitate the network select the optimized application resource. The edge compute server will also not run IGP protocol.

And, passive interfaces are normally the boundary of one IGP domain, knowing them can facilitate the operators to apply various policies on such interfaces, for example, to secure their networks, or filtering the incoming traffic with scrutiny.

But OSPF and ISIS have no position to flag such passive interface and their associated attributes now.

This document defines the protocol extension for OSPF and ISIS to indicate the passive interfaces and their associated attributes.

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. Consideration for flagging passive interface

ISIS [RFC5029] defines the Link-Attributes Sub-TLV to carry the link attribute information, but this Sub-TLV can only be carried within the TLV 22, which is used to described the attached neighbor. For passive interface, there is no ISIS neighbor, then it is not appropriate to use this Sub-TLV to indicate the passive attribute of the interface.

OSPFv2[RFC2328] defines link type field within Router LSA, the type 3 for connections to a stub network can be used to identified the passive interface. But in OSPFv3 [RFC5340], type 3 within the Router-LSA has been reserved. The information that associated with stub network has been put in the Intra-Area-Prefix-LSAs.

It is necessary to define one general solution for ISIS and OSPF to flag the passive interface and transfer the associated attributes then.

4. Passive Interface Attribute

The following sections define the protocol extension to indicate the passive interface and associated attributes in OSPFv2/v3 and ISIS.

4.1. OSPFv2 Extended Stub-Link TLV

[RFC7684] defines the OSPFv2 Extended Link Opaque LSA to contain the additional link attribute TLV. Currently, only OSPFv2 Extended Link TLV is defined to contain the link related sub-TLV. Because passive interface is not the normal link that participate in the OSPFv2 process, we select to define one new top TLV within the OSPFv2 Extended Link Opaque LSA to contain the passive interface related attribute information.

The OSPFv2 Extended Stub-Link TLV has the following format:

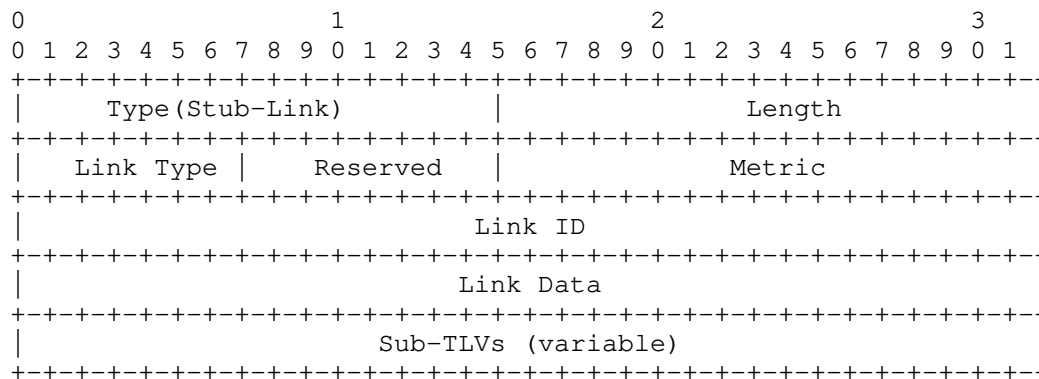


Figure 1: OSPFv2 Extended Stub-Link TLV

Type: The TLV type. The value is 2(TBD) for this stub-link type

Length: Variable, dependent on sub-TLVs

Link Type: Define the type of the stub-link. This document defines the followings type:

- o 0: Reserved
- o 1: AS boundary link
- o 2: Loopback link
- o 3: Vlan interface link
- o 4-255: For future extension

Metric: Link metric used for inter-AS traffic engineering.

Link ID: Link ID is defined in Section A.4.2 of [RFC2328]

Link Data: Link Data is defined in Section A.4.2 of [RFC2328]

Sub-TLVs: Existing sub-TLV that defined within "OSPFv2 Extended Link TLV Sub-TLV" can be included if necessary, the definition of new sub-TLV can refer to Section 4.4

If this TLV is advertised multiple times in the same OSPFv2 Extended Link Opaque LSA, only the first instance of the TLV is used by receiving OSPFv2 routers. This situation SHOULD be logged as an error.

If this TLV is advertised multiple times for the same link in different OSPFv2 Extended Link Opaque LSAs originated by the same OSPFv2 router, the OSPFv2 Extended Stub-Link TLV in the OSPFv2 Extended Link Opaque LSA with the smallest Opaque ID is used by receiving OSPFv2 routers. This situation may be logged as a warning.

It is RECOMMENDED that OSPFv2 routers advertising OSPFv2 Extended Stub-Link TLVs in different OSPFv2 Extended Link Opaque LSAs re-originate these LSAs in ascending order of Opaque ID to minimize the disruption.

This document creates a registry for Stub-Link attribute in Section 6.

4.2. OSPFv3 Router-Stub-Link TLV

[RFC8362] extend the LSA format by encoding the existing OSPFv3 LSA [RFC5340] in TLV tuples and allowing advertisement of additional information with additional TLV.

This document defines the Router-Stub-Link TLV to describes a single router passive interface. The Router-Stub-Link TLV is only applicable to the E-Router-LSA. Inclusion in other Extended LSA MUST be ignored.

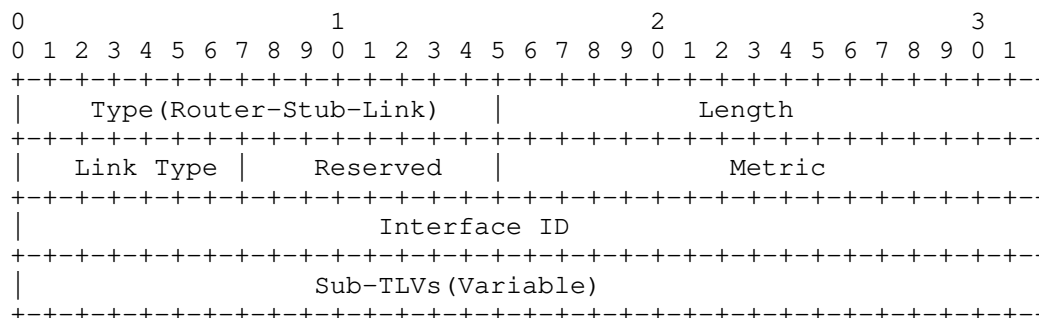


Figure 2: OSPFv3 Router-Stub-Link TLV

Type: OSPFv3 Extended-LSA TLV Type. Value is 10(TBD) for Router-Stub-Link TLV.

Length: Variable, dependent on sub-TLVs

Link Type: Define the type of the stub-link. This document defines the followings type:

- o 0: Reserved
- o 1: AS boundary link
- o 2: Loopback link
- o 3: Vlan interface link
- o 4-255: For future extension

Metric: Link metric used for inter-AS traffic engineering.

Interface ID: 32-bit number uniquely identifying this interface among the collection of this router's interfaces. For example, in some implementations it may be possible to use the MIB-II IfIndex [RFC2863].

Sub-TLVs: Existing sub-TLV that defined within "OSPFv3 Extended-LSA Sub-TLV" can be included if necessary. The definition of new sub-TLV can refer to Section 4.4.

4.3. ISIS Stub-link TLV

This document defines one new top TLV to contain the passive interface attributes, which is shown in Figure 4:

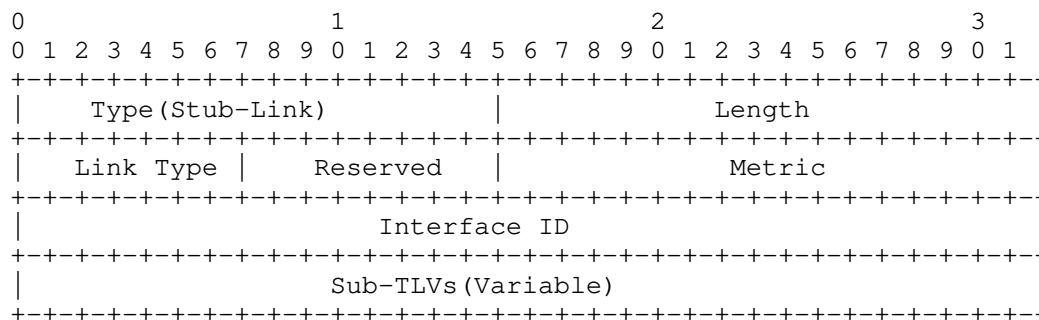


Figure 3: ISIS Stub-Link TLV

Type: ISIS TLV Codepoint. Value is 28(TBD) for stub-link TLV.

Length: Variable, dependent on sub-TLVs

Link Type: Define the type of the stub-link. This document defines the followings type:

- o 0: Reserved
- o 1: AS boundary link
- o 2: Loopback link
- o 3: Vlan interface link
- o 4-255: For future extension

Metric: Link metric used for inter-AS traffic engineering.

Interface ID: 32-bit number uniquely identifying this interface among the collection of this router's interfaces. For example, in some implementations it may be possible to use the MIB-II IfIndex [RFC2863].

Sub-TLVs: Existing sub-TLV that defined within "Sub-TLVs for TLVs 22, 23, 25, 141, 222, and 223" can be included if necessary. The definition of new sub-TLV can refer to Section 4.4.

4.4. Stub-Link Prefix Sub-TLV

This document defines one new sub-TLV that can be contained within the OSPFv2 Extended Stub-Link TLV, OSPFv3 Router-Stub-Link TLV or ISIS Stub-Link TLV, to describe the prefix information associated with the passive interface.

The format of the sub-TLV is the followings:

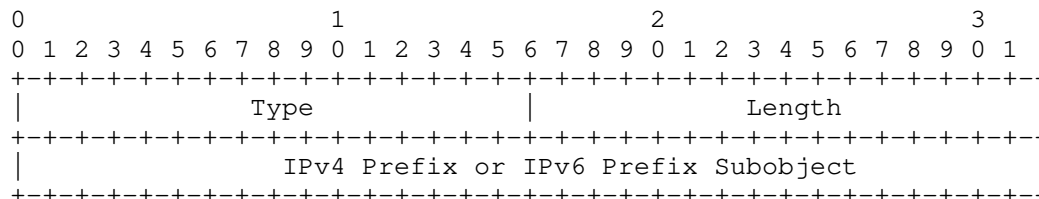


Figure 4: Stub-Link Prefix Sub-TLV

Type: The TLV type. The value is 01(TBD) for this Stub-Link Prefix type

Length: Variable, dependent on associated subobjects

Subobject: IPv4 prefix subobject or IPv6 prefix subobject, as that defined in [RFC3209]

If the passive interface has multiple address, then multiple subobjects will be included within this sub-TLV.

5. Security Considerations

Security concerns for ISIS are addressed in [RFC5304] and[RFC5310]

Security concern for OSPFv3 is addressed in [RFC4552]

Advertisement of the additional information defined in this document introduces no new security concerns.

6. IANA Considerations

IANA is requested to the allocation in following registries:

Registry	Type	Meaning
OSPFv2 Extended Link Opaque LSA TLV	2	Stub-Link TLV
OSPFv3 Extended-LSA TLV	10	Router-Stub-Link TLV
IS-IS TLV Codepoint	28	Stub-Link TLV

Figure 5: Newly defined TLV in existing IETF registry

IANA is requested to allocate one new registry that can be referred by OSPFv2, OSPFv3 and ISIS respectively.

New Registry	Meaning
Stub-Link Attribute	Attributes for stub-link

Figure 6: Newly defined Registry for stub-link attributes

One new sub-TLV is defined in this document under this registry codepoint:

Registry	Type	Meaning
Stub-Link Attribute	0	Reserved
	1	Stub-Link Prefix sub-TLV
	2-65535	Reserved

Figure 7: Stub-Link Prefix Sub-TLV

7. Acknowledgement

Thanks Shunwan Zhang, Tony Li, Les Ginsberg, Acee Lindem, Dhruv Dhody, Jeff Tantsura and Robert Raszuk for their suggestions and comments on this idea.

8. References

8.1. 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>>.
- [RFC2863] McCloghrie, K. and F. Kastenholtz, "The Interfaces Group MIB", RFC 2863, DOI 10.17487/RFC2863, June 2000, <<https://www.rfc-editor.org/info/rfc2863>>.

- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <<https://www.rfc-editor.org/info/rfc3209>>.
- [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>>.
- [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>>.
- [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>>.
- [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>>.
- [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>>.
- [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>>.

8.2. Informative References

- [I-D.dunbar-lsr-5g-edge-compute-ospf-ext]
Dunbar, L., Chen, H., and A. Wang, "OSPF extension for 5G Edge Computing Service", draft-dunbar-lsr-5g-edge-compute-ospf-ext-04 (work in progress), March 2021.

[I-D.ietf-idr-bgppls-inter-as-topology-ext]

Wang, A., Chen, H., Talaulikar, K., and S. Zhuang, "BGP-LS
Extension for Inter-AS Topology Retrieval", draft-ietf-
idr-bgppls-inter-as-topology-ext-09 (work in progress),
September 2020.

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

Gyan S. Mishra
Verizon Inc.
13101 Columbia Pike
Silver Spring MD 20904
United States of America

Email: gyan.s.mishra@verizon.com

Jinsong Sun
ZTE Corporation
No. 68, Ziiijnhua Road
Nan Jing 210012
China

Email: sun.jinsong@zte.com.cn

LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 18, 2022

A. Wang
China Telecom
G. Mishra
Verizon Inc.
Z. Hu
Y. Xiao
Huawei Technologies
October 15, 2021

Prefix Unreachable Announcement
draft-wang-lsr-prefix-unreachable-announcement-08

Abstract

This document describes a mechanism to solve an existing issue with Longest Prefix Match (LPM), that exists where an operator domain is divided into multiple areas or levels where summarization is utilized. This draft addresses a fail-over issue related to a multi areas or levels domain, where a link or node down event occurs resulting in an LPM component prefix being omitted from the FIB resulting in black hole sink of routing and connectivity loss. This draft introduces a new control plane convergence signaling mechanism using a negative prefix called Prefix Unreachable Announcement Mechanism(PUAM), utilized to detect a link or node down event and signal the RIB that the event has occurred to force immediate control plane 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 April 18, 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Conventions used in this document	3
3. Scenario Description	3
3.1. Inter-Area Node Failure Scenario	4
3.2. Inter-Area Links Failure Scenario	4
4. PUA (Prefix Unreachable Advertisement) Procedures	5
5. MPLS and SRv6 LPM based BGP Next-hop Failure Application	5
6. PUAM Capabilities Announcement	6
7. Implementation Consideration	7
8. Deployment Considerations	7
9. Security Considerations	8
10. IANA Considerations	8
11. Acknowledgement	9
12. Normative References	9
Authors' Addresses	10

1. Introduction

As part of an operator optimized design criteria, a critical requirement is to limit Shortest Path First (SPF) churn which occurs within a single OSPF area or ISIS level. This is accomplished by sub-dividing the IGP domain into multiple areas for flood reduction of intra area prefixes so they are contained within each discrete area to avoid domain wide flooding.

OSPF and ISIS have a default and summary route mechanism which is performed on the OSPF area border router or ISIS L1-L2 node. The OSPF summary route is triggered to be advertised conditionally when at least one component prefix exists within the non-zero area. ISIS Level-L1-L2 node as well generate a summary prefix into the level-2 backbone area for Level 1 area prefixes that is triggered to be

advertised conditionally when at least a single component prefix exists within the Level-1 area. ISIS L1-L2 node with attach bit set also generates a default route into each Level-1 area along with summary prefixes generated for other Level-1 areas.

Operators have historically relied on MPLS architecture which is based on exact match host route FEC binding for single area. [RFC5283] LDP inter-area extension provides the ability to LPM, so now the RIB match can now be a summary match and not an exact match of a host route of the egress PE for an inter-area LSP to be instantiated. SRV6 routing framework utilizes the IPv6 data plane standard IGP LPM. When operators start to migrate from MPLS LSP based host route bootstrapped FEC binding, to SRV6 routing framework, the IGP LPM now comes into play with summarization which will influence the forwarding of traffic when a link or node event occurs for a component prefix within the summary range resulting in black hole routing of traffic.

The motivation behind this draft is based on either MPLS LPM FEC binding, or SRv6 BGP service overlay using traditional unicast routing (uRIB) LPM forwarding plane where the IGP domain has been carved up into OSPF or ISIS areas and summarization is utilized. In this scenario where a failure conditions result in a black hole of traffic where multiple ABRs exist and either the area is partitioned or other link or node failures occur resulting in the component prefix host route missing within the summary range. Summarization of inter-area types routes propagated into the backbone area for flood reduction are made up of component prefixes. It is these component prefixes that the PUAM tracks to ensure traffic is not black hole sink routed due to a PE or ABR failure. The PUA mechanism ensures immediate control plane convergence with ABR or PE node switchover when area is partitioned or ABR has services down to avoid black hole of traffic.

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

Figure 1 illustrates the topology scenario when OSPF or ISIS is running in multi areas or multi levels domain. 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 or ISIS Level 1-2 border nodes between area 0 and area 1. R2 and R4 are area border routers between area 0 and area 2.

S1/S4 and T2/T4 PEs peer to customer CEs for overlay VPNs. Ps1/Ps4 is the loopback0 address of S1/S4 and Pt2/Pt4 is the loopback0 address of T2/T4.

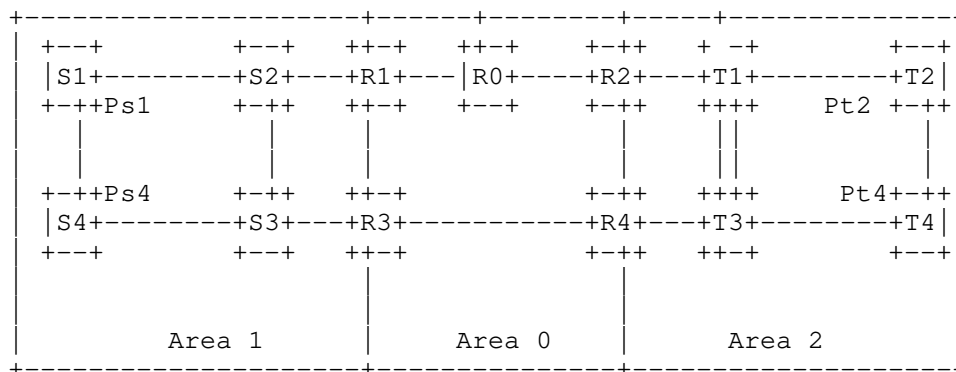


Figure 1: OSPF Inter-Area Prefix Unreachable Announcement Scenario

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 bgp next hop becomes unreachable while the LPM summary prefix continues to be advertised into the backbone area. Except the border router R2/R4, the other routers within area 0 and area 1 do not know the unreachable status of the Pt2 bgp next hop prefix. Traffic will continue to forward LPM match to prefix Pt2 and will be dropped on the ABR or Level 1-2 border node resulting in black hole routing and connectivity loss. Customer overlay VPN dual homed to both S1/S4 and T2/R4, traffic will not be able to fail-over to alternate egress PE T4 bgp next hop Pt4 due to the summarization.

3.2. Inter-Area Links Failure Scenario

In a link failure scenario, if the link between T1/T2 and T1/T3 are down, R2 will not be able to reach node T2. But as R2 and R4 do the summary announcement, and the summary address covers the bgp next hop prefix of Pt2, other nodes in area 0 area 1 will still send traffic to T2 bgp next hop prefix Pt2 via the border router R2, thus black hole sink routing the traffic.

In such a situation, the border router R2 should notify other routers that it can't reach the prefix Pt2, and lets the other ABRs(R4) that can reach prefix Pt2 advertise one specific route to Pt2, then the

internal routers will select R4 as the bypass router to reach prefix Pt2.

4. PUA (Prefix Unreachable Advertisement) Procedures

[RFC7794] and [I-D.ietf-lsr-ospf-prefix-originator] draft both define one sub-tlv to announce the originator information of the one prefix from a specified node. This draft utilizes such TLV for both OSPF and ISIS to signal the negative prefix in the perspective PUAM when a link or node goes down.

ABR detects link or node down and floods PUAM negative prefix advertisement along with the summary advertisement according to the prefix-originator specification. The ABR or ISIS L1-L2 border node has the responsibility to add the prefix originator information when it receives the Router LSA from other routers in the same area or level.

When the ABR or ISIS L1-L2 border node generates the summary advertisement based on component prefixes, the ABR will announce one new summary LSA or LSP which includes the information about this down prefix, with the prefix originator set to NULL. The number of PUAMs is equivalent to the number of links down or nodes down. The LSA or LSP will be propagated with standard flooding procedures.

If the nodes in the area receive the PUAM flood from all of its ABR routers, they will start BGP convergence process if there exist BGP session on this PUAM prefix. The PUAM creates a forced fail over action to initiate immediate control plane convergence switchover to alternate egress PE. Without the PUAM forced convergence the down prefix will yield black hole routing resulting in loss of connectivity.

When only some of the ABRs can't reach the failure node/link, as that described in Section 3.2, the ABR that can reach the PUAM prefix should advertise one specific route to this PUAM prefix. The internal routers within another area can then bypass the ABRs that can't reach the PUAM prefix, to reach the PUAM prefix.

5. MPLS and SRv6 LPM based BGP Next-hop Failure Application

In an MPLS or SR-MPLS service provider core, scalability has been a concern for operators which have split up the IGP domain into multiple areas to avoid SPF churn. Normally, MPLS FEC binding for LSP instantiation is based on egress PE exact match of a host route Looback0. [RFC5283] LDP inter-area extension provides the ability to LPM, so now the RIB match can now be a summary match and not an exact match of host route of the egress PE for an inter-area LSP to be

instantiated. The caveat related to this feature that has prevented operators from using the [RFC5283] LDP inter-area extension concept is that when the component prefixes are now hidden in the summary prefix, and thus the visibility of the BGP next-hop attribute is lost.

In a case where a PE is down, and the [RFC5283] LDP inter-area extension LPM summary is used to build the LSP inter-area, the LSP remains partially established black hole on the ABR performing the summarization. This major gap with [RFC5283] inter-area extension forces operators into a workaround of having to flood the BGP next-hop domain wide. In a small network this is fine, however if you have 1000s PEs and many areas, the domain wide flooding can be painful for operators as far as resource usage memory consumption and computational requirements for RIB / FIB / LFIB label binding control plane state. The ramifications of domain wide flooding of host routes is described in detail in [RFC5302] domain wide prefix distribution with 2 level ISIS Section 1.2 - Scalability. As SRv6 utilizes LPM, this problem exists as well with SRv6 when IGP domain is broken up into areas and summarization is utilized.

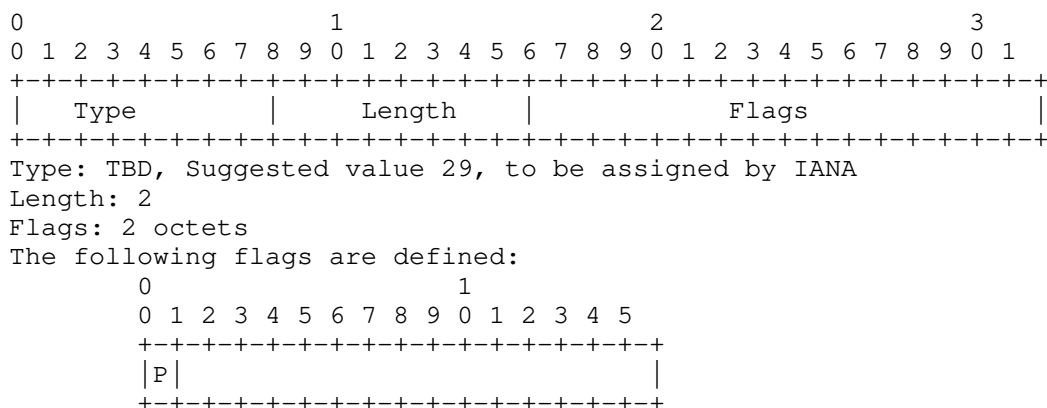
PUAM is now able to provide the negative prefix component flooded across the backbone to the other areas along with the summary prefix, which is now immediately programmed into the RIB control plane. MPLS LSP exact match or SRv6 LPM match over fail over path can now be established to the alternate egress PE. No disruption in traffic or loss of connectivity results from PUAM. Further optimizations such as LFA and BFD can be done to make the data plane convergence hitless. The PUAM solution applies to MPLS or SR-MPLS where LDP inter-area extension is utilized for LPM aggregate FEC, as well a SRv6 IPv6 control plane LPM match summarization of BGP next hop.

6. PUAM Capabilities Announcement

When not all of the nodes in one area support the PUAM information, there are possibilities to form traffic loop. To avoid this happen, the ABR should not send PUAM information to one area until it ensures that all of nodes in this area can parse the PUAM information. 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]. For OSPFv3, one bit (Bit number TBD, suggest bit 8) should be defined to indicate the router's capabilities to support PUAM 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), PUAM Capabilities

sub-TLV, which is included in the "IS-IS Router CAPABILITY TLV" [RFC7981] is defined in the followings:



where:

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

Figure 2: PUA Capabilities sub-TLV format

7. Implementation Consideration

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 PUAM.
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.

8. Deployment Considerations

To support the PUAM advertisement, the ABRs should be upgraded according to the procedures described in Section 4. The PEs that want to accomplish the BGP switchover that described in Section 3.1 and Section 5 should also be upgraded to act upon the receive of the PUAM message. Other nodes within the network can ignore such PUAM message if they don't care or don't support.

As described in Section 4, the ABR will advertise the PUAM message once it detects there is link or node down within the summary address. In order to reduce the unnecessary advertisements of PUAM messages on ABRs, the ABRs should support the configuration of the protected prefixes. Based on such information, the ABR will only advertise the PUAM message when the protected prefixes (for example, the loopback addresses of PEs that run BGP) that within the summary address is missing.

The advertisement of PUAM message should only last one configurable period to allow the services that run on the failure prefixes are converged or switchover. If one prefix is missed before the PUAM takes effect, the ABR will not declare its absence via the PUAM.

9. Security Considerations

Advertisement of PUAM information follow the same procedure of traditional LSA. The action based on the PUAM is clearly defined in this document for ABR or Level1/2 router and the receiver that run BGP.

There is no changes to the forward behavior of other internal routers.

10. 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

11. Acknowledgement

Thanks Peter Psenak, Les Ginsberg, Acee Lindem, Shraddha Hegde, Robert Raszuk, Tonly Li, Jeff Tantsura, Tony Przygienda and Bruno Decraene for their suggestions and comments on this draft.

12. 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-12 (work in progress), April 2021.
- [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>>.
- [RFC5283] Decraene, B., Le Roux, J.L., and I. Minei, "LDP Extension for Inter-Area Label Switched Paths (LSPs)", RFC 5283, DOI 10.17487/RFC5283, July 2008, <<https://www.rfc-editor.org/info/rfc5283>>.

- [RFC5302] Li, T., Smit, H., and T. Przygienda, "Domain-Wide Prefix Distribution with Two-Level IS-IS", RFC 5302, DOI 10.17487/RFC5302, October 2008, <<https://www.rfc-editor.org/info/rfc5302>>.
- [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

Gyan Mishra
Verizon Inc.

Email: gyan.s.mishra@verizon.com

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: Informational
Expires: August 26, 2021

C. Xie
C. Ma
China Telecom
J. Dong
Z. Li
Huawei Technologies
February 22, 2021

Using IS-IS Multi-Topology (MT) for Segment Routing based Virtual
Transport Network
draft-xie-lsr-isis-sr-vtn-mt-03

Abstract

Enhanced VPN (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 underlay network which consists of a subset of the network topology and network resources allocated from the physical network. A VTN could be used as the underlay for one or a group of VPN+ services.

In some network scenarios, each VTN can be associated with a unique logical network topology. This document describes a mechanism to build the SR based VTNs using IS-IS Multi-Topology together with other well-defined IS-IS extensions.

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 August 26, 2021.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Advertisement of SR VTN Topology Attribute	3
3. Advertisement of SR VTN Resource Attribute	4
3.1. Advertising Topology-specific TE attributes	4
4. Forwarding Plane Operations	5
5. Scalability Considerations	5
6. Security Considerations	5
7. IANA Considerations	6
8. Acknowledgments	6
9. References	6
9.1. Normative References	6
9.2. Informative References	7
Authors' Addresses	7

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. Thus these properties 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) can be created, each with a subset of the underlay network topology and a subset of network resources allocated

from the underlay network to meet the requirement of one or a group of VPN+ services. Another possible approach is to create a set of point-to-point paths, each with a set of network resource reserved along the path, such paths are called Virtual Transport Path (VTP). Although using a set of dedicated VTPs can provide similar characteristics as a VTN, it has some scalability issues due to the per-path state in the network.

[I-D.ietf-spring-resource-aware-segments] introduces resource awareness to Segment Routing (SR) [RFC8402]. The resource-aware SIDs have additional semantics to identify the set of network resources available for the packet processing action associated with the SIDs. As described in [I-D.ietf-spring-sr-for-enhanced-vpn], the resource-aware SIDs 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 both 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 can have an independent topology and a set of dedicated network resources. This document describes a simplified mechanism to build 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 attribute of network resources allocated to a VTN can be advertised by using IS-IS MT with the Traffic Engineering (TE) extensions defined in [RFC5305] and [RFC8570].

2. Advertisement of SR VTN Topology Attribute

IS-IS Multi-Topology Routing (MTR) [RFC5120] has been defined to create independent topologies in one network. In [RFC5120], MT-based TLVs are introduced to carry topology-specific link-state information. The MT-specific Link or Prefix TLVs are defined by adding additional two bytes, which includes 12-bit MT-ID field in front of the ISN TLV and IP or IPv6 Reachability TLVs. This provides the capability of specifying the customized attributes of each topology. When each VTN is associated with an independent network

topology, MT-ID could be used as the identifier of VTN in control plane.

MTR can be used with segment routing based data plane. Thus the topology attribute of an SR based VTN could be advertised using MTR with segment routing. The IS-IS extensions to support the advertisement of topology-specific MPLS SIDs are specified in [RFC8667]. Topology-specific Prefix-SIDs can be 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 can be 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).

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 and other attributes associated with each VTN need to be advertised.

3.1. Advertising Topology-specific TE attributes

On each network link, the information of the network resources and other attributes associated with a VTN can be specified by carrying the TE attributes sub-TLVs [RFC5305] and [RFC8570] 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 of a topology, 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 notel: 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 in [RFC5120].

4. Forwarding Plane Operations

For SR-MPLS data plane, a Prefix-SID is associated with the paths calculated in the corresponding topology of a VTN. An outgoing interface is determined for each path. In addition, the prefix-SID also steers the traffic to use the subset of network resources allocated to the VTN on the outgoing interface for packet forwarding. An Adj-SID is associated with a subset of network resources allocated to a VTN on the link. The Adj-SIDs and Prefix-SIDs associated with the same VTN can be used together to build SR-MPLS paths with the topological and resource constraints of the VTN.

For SRv6 data plane, an SRv6 Locator is a prefix which is associated with the paths calculated in the corresponding topology of a VTN. An outgoing interface is determined for each path. In addition, the SRv6 Locator prefix also steers the traffic to use the subset of network resources which are allocated to the VTN on the outgoing interface for packet forwarding. An End.X SID is associated with a subset of network resources allocated to a VTN on the link. The End.X SIDs and the SRv6 Locator prefixes associated with the same VTN can be used together to build SRv6 paths with the topological and resource constraints of the VTN.

5. Scalability Considerations

The mechanism described in this document assumes that each VTN is associated with a unique topology, so that the MT-IDs can be reused to identify the VTNs in the control plane. While this brings the benefit of simplicity, it also has some limitations. For example, it means that even if multiple VTNs have the same topology, they would still need to be identified using different MT-IDs in the control plane, then independent path computation needs to be executed for each VTN. Thus the number of VTNs supported in a network may be dependent on the number of topologies supported, which is related to the control plane computation overhead.

6. 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.

7. IANA Considerations

This document does not request any IANA actions.

8. Acknowledgments

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

9. References

9.1. 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-11 (work in progress), October 2020.
- [I-D.ietf-spring-resource-aware-segments]
Dong, J., Bryant, S., Miyasaka, T., Zhu, Y., Qin, F., Li, Z., and F. Clad, "Introducing Resource Awareness to SR Segments", draft-ietf-spring-resource-aware-segments-01 (work in progress), January 2021.
- [I-D.ietf-spring-sr-for-enhanced-vpn]
Dong, J., Bryant, S., Miyasaka, T., Zhu, Y., Qin, F., Li, Z., and F. Clad, "Segment Routing based Virtual Transport Network (VTN) for Enhanced VPN", February 2021, <<https://tools.ietf.org/html/draft-ietf-spring-sr-for-enhanced-vpn>>.
- [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>>.

- [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>>.
- [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>>.

9.2. 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-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Enhanced Virtual Private Networks (VPN+) Service", draft-ietf-teas-enhanced-vpn-06 (work in progress), July 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

LSR Working Group
Internet-Draft
Intended status: Standards Track
Expires: 8 September 2022

Y. Zhu
China Telecom
J. Dong
Z. Hu
Huawei Technologies
7 March 2022

Using Flex-Algo for Segment Routing based VTN
draft-zhu-lsr-isis-sr-vtn-flexalgo-04

Abstract

Enhanced VPN (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 connectivity and the characteristics provided by the underlay network. A Virtual Transport Network (VTN) is a virtual underlay network which has a customized network topology and a set of network resources allocated from the physical network. A VTN could be used as the underlay for one or a group of VPN+ services.

The topological constraints of a VTN can be defined using Flex-Algo. In some network scenarios, each VTN can be associated with a unique Flex-Algo, and the set of network resources allocated to a VTN can be instantiated as layer-2 sub-interfaces or member links of the layer-3 interfaces. This document describes the mechanisms to build the SR based VTNs using SR Flex-Algo and IGP L2 bundle with minor extensions.

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 8 September 2022.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	2
1.1. Requirements Language	3
2. Advertisement of SR VTN Topology Attributes	3
3. Advertisement of SR VTN Resource Attributes	4
4. Forwarding Plane Operations	5
5. Scalability Considerations	6
6. Security Considerations	6
7. IANA Considerations	6
8. Acknowledgments	6
9. References	6
9.1. Normative References	7
9.2. Informative References	8
Authors' Addresses	9

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. Thus these properties 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) can be created, each with a subset of the underlay network topology and a set of network resources allocated from the underlay network to meet the requirement of a specific VPN+

service or a group of VPN+ services. Another possible approach is to create a set of point-to-point paths, each with a set of network resource reserved along the path, such paths are called Virtual Transport Paths (VTPs). Although using a set of dedicated VTPs can provide similar characteristics as VTN, it has some scalability issues due to the per-path state in the network.

[I-D.ietf-spring-resource-aware-segments] introduces resource awareness to Segment Routing (SR) [RFC8402]. As described in [I-D.ietf-spring-sr-for-enhanced-vpn], the resource-aware SIDs can be used to build VTNs with the required network topology and network resource attributes to support VPN+ services. With segment routing based data plane, Segment Identifiers (SIDs) can be used to represent both the topology and the set of network resources allocated by network nodes to a VTN. The SIDs of each VTN together with its associated topology and resource attributes need to be distributed using control plane.

[I-D.dong-lsr-sr-enhanced-vpn] defines the IGP mechanisms and extensions to provide scalable Segment Routing (SR) based VTNs. The mechanism in [I-D.dong-lsr-sr-enhanced-vpn] allows flexible combination of the topology and resource attribute to provide a relatively large number of VTNs. In some network scenarios, each VTN can be associated with a unique Flex-Algo, and the set of network resources allocated to the VTN can be instantiated using layer-2 sub-interfaces or member links of the L3 interfaces. This document describes a mechanism to build the SR based VTNs using SR Flex-Algo and IGP L2 bundle with minor extensions.

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 BCP14 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Advertisement of SR VTN Topology Attributes

[I-D.ietf-lsr-flex-algo] specifies the mechanism to provide distributed constraint-path computation, and the usage of SR-MPLS prefix-SIDs and SRv6 locators for steering traffic along the constrained paths.

The Flex-Algo Definition (FAD) is the combination of calculation-type, metric-type and the topological constraints used for path computation. According to the network nodes' participation of a

Flex-Algo, and the rules of including or excluding Admin Groups (i.e. colors) and Shared Risk Link Groups (SRLGs), the topology of a VTN can be described using the associated Flex-Algo. If each VTN is associated with a unique Flex-Algo, the Flex-Algo identifier could be reused as the identifier of the VTN in the control plane.

With the mechanisms defined in[RFC8667] [I-D.ietf-lsr-flex-algo], SR-MPLS prefix-SID advertisement can be associated with a specific topology and a specific algorithm, which can be a Flex-Algo. This allows the nodes to use the prefix-SIDs to steer traffic along distributed computed constraint paths according to the associated Flex-Algo in a particular topology.

[I-D.ietf-lsr-isis-srv6-extensions] specifies the IS-IS extensions to support SRv6 data plane, in which the SRv6 locators advertisement is associated with a topology and a specific algorithm, which can be a Flex-Algo. This allows the nodes to use the SRv6 locators to steer traffic along distributed computed constraint paths according to the associated Flex-Algo in a particular topology. In addition, topology/algorithm specific SRv6 End SIDs and End.X SIDs can be used to enforce traffic over the Loop-Free Alternatives (LFA) computed backup paths.

3. Advertisement of SR VTN Resource Attributes

Each VTN can be allocated with a set of dedicated network resources on different network nodes and links. In order to perform constraint based path computation for each VTN on network controller and the ingress nodes, the resource attribute of each VTN also needs to be advertised. This way, the network controller or the ingress node can compute an SR TE path in a VTN by taking both the Flex-Algo constraints and the resource attribute of the VTN into consideration.

IS-IS L2 Bundle [RFC8668] was defined to advertise the link attributes of the layer-2 bundle member links. In this section, it is extended to advertise the set of network resource attributes associated with different VTNs on a layer-3 link.

The layer-3 link may or may not be a bundle of layer-2 links, as long as it has the capability of partitioning the link resources into different subsets for different VTNs it participates in. One partition of the link resources can be instantiated as a layer-2 sub-interface, which can be seen as a virtual layer-2 member link of the layer-3 link. If the layer-3 link is a layer-2 link bundle, it is possible that the set of link resource allocated to a specific VTN is provided by one or multiple physical layer-2 member links.

A new flag "E" (Exclusive) is defined in the flag field of the Parent L3 Neighbor Descriptor in the L2 Bundle Member Attributes TLV (25).

```

      0 1 2 3 4 5 6 7
    +---+---+---+---+
    |P|E|         |
    +---+---+---+---+

```

E flag: When the E flag is set, it indicates each member link under the Parent L3 link are used exclusively for one VTN, and load sharing among the member links is not allowed. When the E flag is clear, it indicates load balancing and sharing among the member links are allowed.

For each virtual or physical layer-2 member link, the TE attributes defined in [RFC5305] such as the Maximum Link Bandwidth and Admin Groups SHOULD be advertised using the mechanism as defined in [RFC8668]. The SR-MPLS Adj-SIDs or SRv6 End.X SIDs associated with each of the virtual or physical Layer-2 member links SHOULD also be advertised according to [RFC8668] and [I-D.dong-lsr-l2bundle-srv6].

In order to correlate the virtual or physical layer-2 member links with the Flex-Algo ID which is used to identify the VTN, each VTN SHOULD be assigned with a unique Admin Group (AG) or Extended Admin Group (EAG), and the virtual or physical layer-2 member links associated with this VTN SHOULD be configured with the AG or EAG assigned to the VTN. The AG or EAG of the parent layer-3 link SHOULD be set to the union of all the AGs or EAGs of its virtual or physical layer-2 member links. In the definition of the Flex-Algo corresponding to the VTN, It MUST use the Include-Any Admin Group rule with only the AG or EAG assigned to the VTN as the link constraints, the Include-All Admin Group rule or the Exclude Admin Group rule MUST NOT be used. This is to ensure that the layer-3 link is included in the Flex-Algo constraint based path computation for each VTN it participates in.

4. Forwarding Plane Operations

For SR-MPLS data plane, a prefix SID is associated with the paths calculated using the Flex-Algo corresponding to a VTN. An outgoing layer-3 interface is determined for each path. In addition, the prefix-SID also steers the traffic to use the virtual or physical layer-2 member link which is associated with the VTN on the outgoing layer-3 interface for packet forwarding. The Adj-SIDs associated with the virtual or physical member links of a VTN MAY be used with the prefix-SIDs of the same VTN together to build SR-MPLS TE paths with the topological and resource constraints of the VTN.

For SRv6 data plane, an SRv6 Locator is a prefix which is associated with the paths calculated using the Flex-Algo corresponding to a VTN. An outgoing Layer-3 interface is determined for each path. In addition, the SRv6 Locator prefix also steers the traffic to use the virtual or physical layer-2 member link which is associated with the VTN on the outgoing layer-3 interface for packet forwarding. The End.XU SIDs associated with the virtual or physical member links of a VTN MAY be used with the SRv6 Locator prefix of the same VTN together to build SRv6 paths with the topological and resource constraints of the VTN.

5. Scalability Considerations

The mechanism described in this document assumes that each VTN is associated with a unique Flex-Algo, so that the Flex-Algo IDs can be reused to identify the VTNs in the control plane. While this brings the benefit of simplicity, it also has some limitations. For example, it means that even if multiple VTNs share the same topological constraints, they still need to be identified using different Flex-Algo IDs in the control plane, then independent path computation needs to be executed for each VTN. The number of VTNs supported in a network may be dependent on the number of Flex-Algos supported, which is related to the number of Flex-Algos defined in the protocol (which is 128) and the control plane overhead on network nodes. The mechanism described in this document is applicable to network scenarios where the number of required VTN is relatively small. A detailed analysis about the VTN scalability and the possible optimizations for supporting a large number of VTNs is described in [I-D.dong-teas-nrp-scalability].

6. 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.

7. IANA Considerations

This document does not request any IANA actions.

8. Acknowledgments

The authors would like to thank Zhenbin Li and Peter Psenak for the review and discussion of this document.

9. References

9.1. Normative References

[I-D.dong-lsr-l2bundle-srv6]

Dong, J. and Z. Hu, "Advertising SRv6 SIDs for Layer 2 Bundle Member Links in IGP", Work in Progress, Internet-Draft, draft-dong-lsr-l2bundle-srv6-01, 24 October 2021, <<https://www.ietf.org/archive/id/draft-dong-lsr-l2bundle-srv6-01.txt>>.

[I-D.ietf-lsr-flex-algo]

Psenak, P., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", Work in Progress, Internet-Draft, draft-ietf-lsr-flex-algo-18, 25 October 2021, <<https://www.ietf.org/archive/id/draft-ietf-lsr-flex-algo-18.txt>>.

[I-D.ietf-lsr-isis-srv6-extensions]

Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extensions to Support Segment Routing over IPv6 Dataplane", Work in Progress, Internet-Draft, draft-ietf-lsr-isis-srv6-extensions-18, 20 October 2021, <<https://www.ietf.org/archive/id/draft-ietf-lsr-isis-srv6-extensions-18.txt>>.

[I-D.ietf-spring-resource-aware-segments]

Dong, J., Bryant, S., Miyasaka, T., Zhu, Y., Qin, F., Li, Z., and F. Clad, "Introducing Resource Awareness to SR Segments", Work in Progress, Internet-Draft, draft-ietf-spring-resource-aware-segments-03, 12 July 2021, <<https://www.ietf.org/archive/id/draft-ietf-spring-resource-aware-segments-03.txt>>.

[I-D.ietf-spring-sr-for-enhanced-vpn]

Dong, J., Bryant, S., Miyasaka, T., Zhu, Y., Qin, F., Li, Z., and F. Clad, "Segment Routing based Virtual Transport Network (VTN) for Enhanced VPN", Work in Progress, Internet-Draft, draft-ietf-spring-sr-for-enhanced-vpn-01, 12 July 2021, <<https://www.ietf.org/archive/id/draft-ietf-spring-sr-for-enhanced-vpn-01.txt>>.

[I-D.ietf-teas-enhanced-vpn]

Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Enhanced Virtual Private Network (VPN+) Services", Work in Progress, Internet-Draft, draft-ietf-teas-enhanced-vpn-09, 25 October 2021, <<https://www.ietf.org/archive/id/draft-ietf-teas-enhanced-vpn-09.txt>>.

- [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>>.
- [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>>.
- [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>>.
- [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>>.
- [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>>.

9.2. 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 Scalable Segment Routing based Enhanced VPN", Work in Progress, Internet-Draft, draft-dong-lsr-sr-enhanced-vpn-07, 29 January 2022, <<https://www.ietf.org/archive/id/draft-dong-lsr-sr-enhanced-vpn-07.txt>>.
- [I-D.dong-teas-nrp-scalability]
Dong, J., Li, Z., Gong, L., Yang, G., Guichard, J. N., Mishra, G., Qin, F., Saad, T., and V. P. Beeram, "Scalability Considerations for Network Resource Partition", Work in Progress, Internet-Draft, draft-dong-teas-nrp-scalability-01, 7 February 2022, <<https://www.ietf.org/archive/id/draft-dong-teas-nrp-scalability-01.txt>>.

Authors' Addresses

Yongqing Zhu
China Telecom
Email: zhuyq8@chinatelecom.cn

Jie Dong
Huawei Technologies
Email: jie.dong@huawei.com

Zhibo Hu
Huawei Technologies
Email: huzhibo@huawei.com