SPRING
Internet-Draft
Intended status: Standards Track
Expires: August 26, 2021

S. Hegde
W. Britto
R. Shetty
Juniper Networks Inc.
B. Decraene
Orange

February 22, 2021

Flexible Algorithms Bandwidth Constraints
draft-hegde-lsr-flex-algo-bw-con-00

Abstract

Many networks configure the link metric relative to the link capacity. High bandwidth traffic gets routed as per the link capacity. Flexible algorithms provides mechanisms to create constraint based paths in IGP. This draft documents a set of bandwidth related constraints to be used in Flexible Algorithms.

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 August 26, 2021.

Copyright Notice

Table of Contents

1. Introduction

High bandwidth traffic such as residential internet traffic and machine to machine elephant flows benefit from using high capacily links for the traffic. Many network operators define link metric relative to the link capacity. It may be useful to exclude the high bandwidth traffic from utilizing links below certain capacity. A flex-algo [I-D.ietf-lsr-flex-algo] is defined as a set of parameters consisting of calculation-type, metric-type and a set of constraints. It is very convenient to define a flex-algo that uses bandwidth based metric-type which can be used for carrying high bandwidth traffic.In this regard, it's useful to define additional metric-type and additional bandwidth related constraints to simplify the operations.

This document specifies a new metric-type to be used in flex-algo described in section Section 2. Additional Flexible Algorithm Definition (FAD) constraints defined in section Section 3. Section 4 defines mechanisms to automatically calculate metric based on reference bandwidth and actual link bandwidth.

2. Bandwidth Metric adevertisement

ISIS and OSPF advertise link metric in their respective link information. Multiple types of metric are supported, IGP cost, te-metric defined in [RFC5305] and [RFC3630] and delay metric defined in [RFC8570] and [RFC7471]. A brownfield network might have deployed legacy transport mechanisms using igp-cost and te-metric which continue to run during migration period. In this brownfield network if the operator wants to introduce two Flex-Algos, one for delay metric and another for bandwidth metric, a new metric-type to carry bandwidth related metric, would be needed. The IGP cost and te-metric may be already used by legacy applications and may not be available to carry link bandwidth based metric. This document defines a new metric called bandwidth metric. ISIS and OSPF will advertise this new type of metric in their link information. This document also defines a new metric-type called "bandwidth metric" in the FAD sub-TLV.

2.1. ISIS Bandwidth Metric sub-TLV

The ISIS Bandwidth Metric sub-TLV specifies the link metric based on link bandwidth. Typically, this metric is assigned by a network

administrator. The bandwidth metric sub-TLV is advertised in below TLVs

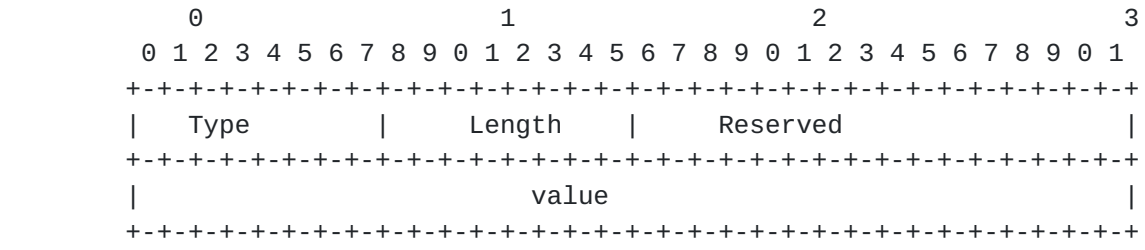
TLV-22 (Extended IS reachability) [RFC5305]

TLV-222 (MT-ISN) [RFC5120]

TLV-23 (IS Neighbor Attribute) [RFC5311]

TLV-223 (MT IS Neighbor Attribute) [RFC5311]

TLV-141 (inter-AS reachability information) [RFC5316]



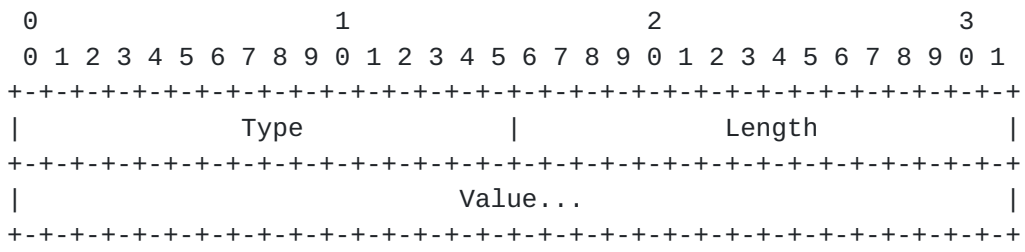
Type : TBD (To be assigned by IANA)
Length: 5 octets
Value : metric value range (1 - 4,261,412,864)

Figure 1: ISIS bandwidth metric sub-TLV

The bandwidth metric sub-TLV MUST be advertised only once. If there are multiple bandwidth metric sub-TLVs in one or more received LSPDUs, the first one MUST be used and the subsequesnt ones MUST be ignored.

2.2. OSPF Bandwidth Metric sub-TLV

The Bandwidth Metric sub-TLV specifies the link metric based on link bandwidth. Typically, this metric is assigned by a network administrator. The bandwidth metric sub-TLV is a sub-TLV of the OSPF Link TLV and advertised in OSPF extended Link LSA [RFC7684]. The Bandwidth Metric sub-TLV is TLV type TBD (IANA), and is four octets in length.



Type : TBD (To be assigned by IANA)
Length: 4 octets
Value : metric value (1- 4,294,967,296)

Figure 2: OSPF bandwidth metric sub-TLV

The bandwidth metric sub-TLV MUST be advertised only once. If there are multiple bandwidth metric sub-TLVs in a receieved Link TLV, the first one MUST be used and the subsequent ones MUST be ignored.

3. FAD constraint sub-TLVs

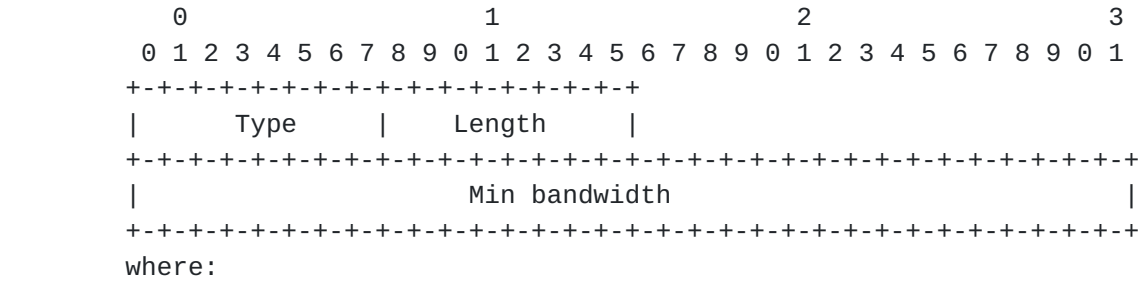
It is useful to exclude Links having capacity lower than a minimum value from the flex-algo topology that is designed to carry high bandwidth traffic. This can be achieved by associating link affinity to the lower capacity links and advertise exclude link constraint in the FAD for that link affinity. This works well where the link capacity is constant. When a L3 link is collection of L2 links (LAG/ L2 Bundle) , the link bandwidth varies based on the constituent link going up and down. The operator has to constantly monitor the link capacity and assign appropriate link affinity on link capacity changes beyond minimum value. In certain cases, the minimum link bandwidth required may change based on the applications that use the high bandwidth Flex-Algo. This document proposes a new exclude minimum bandwidth constraint. When this constraint is advertised in a FAD, based on the advertised link bandwidth, the link will be pruned from the flex-algo topology if its below the FAD advertised Minimum bandwidth value.

Similarly, exclude maximum link delay constraint is also defined in this document. Links may have the link delay measured dynamically and advertised in delay metric in IGP. For usecases that deploy low latency flex-algo, may want to exclude links that have delay more than a defined threshold.

3.1. ISIS FAD constraint sub-TLVs

3.1.1. ISIS Exclude Minimum Bandwidth sub-TLV

ISIS Flex-algo Exclude Minimum Bandwidth sub-TLV (FAEMB) is a sub-TLV of the ISIS FAD sub-TLV. It has the following format.



where:

Type: 1

Length: 4 octets.

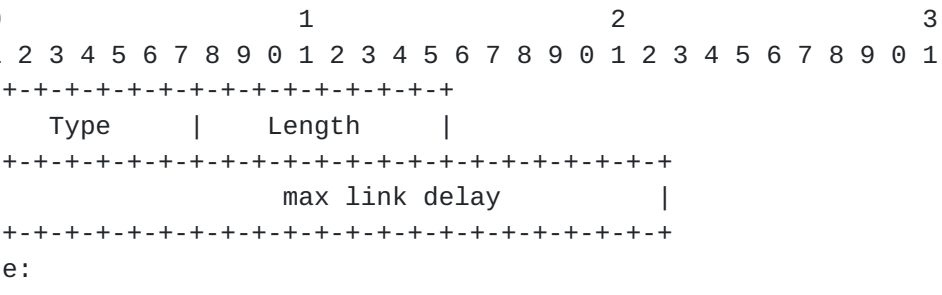
Min Bandwidth: link bandwidth is encoded in 32 bits in IEEE floating point format. The units are bytes per second.

Figure 3: ISIS FAEMB sub-TLV

The FAEMB sub-TLV MUST appear only once in the FAD sub-TLV. If it appears more than once, the ISIS FAD Sub-TLV MUST be ignored by the receiver. The total link bandwidth as advertised by the sub-TLV 9 of the TLV 22/222/23/223/141 is compared against the Min bandwidth advertised in FAEMB sub-TLV. If the link bandwidth is lower, the link is excluded from the Flex-algo topology of the corresponding Flex-algo which advertised the FAEMB sub-TLV. If a link does not have the link bandwidth advertised but the FAD contains this sub-TLV, then that link MUST be excluded from the topology.

3.1.2. ISIS Exclude Maximum Delay sub-TLV

ISIS Flex-algo Exclude Maximum Delay sub-TLV (FAEMD) is a sub-TLV of the ISIS FAD sub-TLV. It has the following format.



where:

Type: TBD

Length: 3 octets

Max link delay: Maximum link delay in microseconds

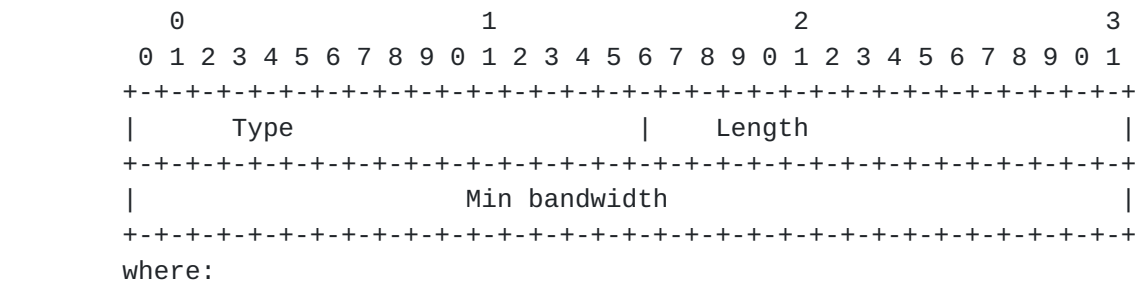
Figure 4: ISIS FAEMD sub-TLV

The FAEMD sub-TLV MUST appear only once in the FAD sub-TLV. If it appears more than once, the ISIS FAD Sub-TLV MUST be ignored by the receiver. The link delay [RFC8570].as advertised by the sub-TLV 33 of the TLV 22/222/23/223/141 is compared against the Max link delay advertised in FAEMD sub-TLV. If the link delay value is higher, the link is excluded from the Flex-algo topology of the corresponding Flex-algo which advertised the FAEMD sub-TLV. If a link does not have the link delay advertised but the FAD contains this sub-TLV, then that link MUST be excluded from the topology.

3.2. OSPF FAD constraint sub-TLVs

3.2.1. OSPF Exclude Minimum Bandwidth sub-TLV

OSPF Flex-algo Bandwidth Exclusion sub-TLV (FAEMB) is a sub-TLV of the OSPF FAD TLV. It has the following format.



where:

Type: TBD

Length: 4 octets.

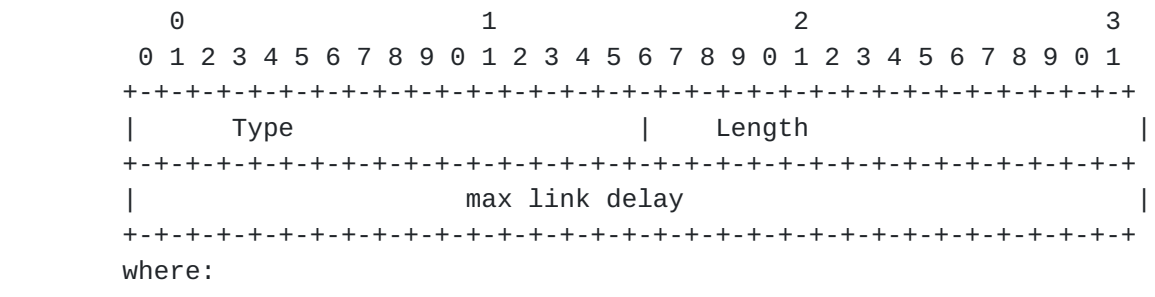
Min Bandwidth: link bandwidth is encoded in 32 bits in IEEE floating point format. The units are bytes per second.

Figure 5: OSPF FAEMB sub-TLV

The FAEMB sub-TLV MUST appear only once in the FAD sub-TLV. If it appears more than once, the OSPF FAD TLV MUST be ignored by the receiver. The total link bandwidth as advertised by the sub-TLV 6 of the Extended Link TLV of Extended Link Opaque LSA [RFC 7684] is compared against the Min bandwidth advertised in FAEMB sub-TLV. If the link bandwidth is lesser, the link is excluded from the Flex-algo topology of the corresponding Flex-algo which advertised the FAEMB sub-TLV. If a link does not have the link bandwidth advertised but the FAD contains this sub-TLV, then that link MUST be excluded from the topology.

3.2.2. OSPF Exclude Maximum Delay sub-TLV

OSPF Flex-algo Exclude Maximum Delay sub-TLV (FAEMD) is a sub-TLV of the OSPF FAD TLV. It has the following format.



where:

Type: TBD

Length: 4 octets

Max link delay: Maximum link delay in microseconds

Figure 6: OSPF FAEMD sub-TLV

The FAEMD sub-TLV MUST appear only once in the OSPF FAD TLV. If it appears more than once, the OSPF FAD TLV MUST be ignored by the receiver. The link delay as advertised by the sub-TLV 27 of the Extended Link TLV of Extended link opaque LSA [RFC 7684] is compared against the Max delay advertised in FAEMD sub-TLV. If the link delay value is higher, the link is excluded from the Flex-algo topology of the corresponding Flex-algo which advertised the FAEMD sub-TLV. If a link does not have the link delay advertised but the FAD contains this sub-TLV, then that link MUST be excluded from the topology.

4. Automatic metric calculation

Networks which are designed to be highly regular and follow uniform metric assignment may want to further simplify the operations by automatically calculating the metric based on the reference bandwidth or a staircase metric assignment based on bandwidth thresholds. Based on the advertised rules, every node automatically calculates the link metric of the links in the network before running SPF algorithm. Based on the delay in receiving the link bandwidth changes, there may be possibility of micro-loops which is no different from IGP susceptibility to micro-loops during metric change. The micro-loop avoidance procedures described in [I-D.bashandy-rtgwg-segment-routing-uloop] can be used to avoid micro-loops when the automatic metric calculation is deployed.

4.1. Simple mode

In simple mode, the link bandwidth of a single Layer 3 link is used. Two ways of automatic metric calculation is supported.

1. Based on reference bandwidth

2. Staircase metric values based on bandwidth thresholds

4.2. Interface group mode

Simple mode of metric calculation may not work well when there are multiple parallel layer 3 interfaces between two nodes.

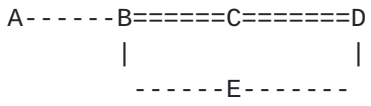


Figure 7: Parallel interfaces

In the above diagram, there are two parallel links between B->C and C->D. Lets assume the link bandwidth is uniform 10Gbps on all links. When Simple mode of metric derivation is used,the metric is derived as 10 on all links. Traffic will be load balanced beween B-> C->D and B->E->D. Since the bandwidth is higher B->C->D path, the requirement is t be able to assign smaller metric based on cumulative metric for the parallel links.

In the interface group mode, every node identifies the set of parallel links betweena pair of nodes based on IGP link advertisements and considers cumulative bandwidth of the parallel links while arriving at the metric for the link. Two ways of automatic metric calculation is supported for interface group mode as well.

1. Based on reference bandwidth

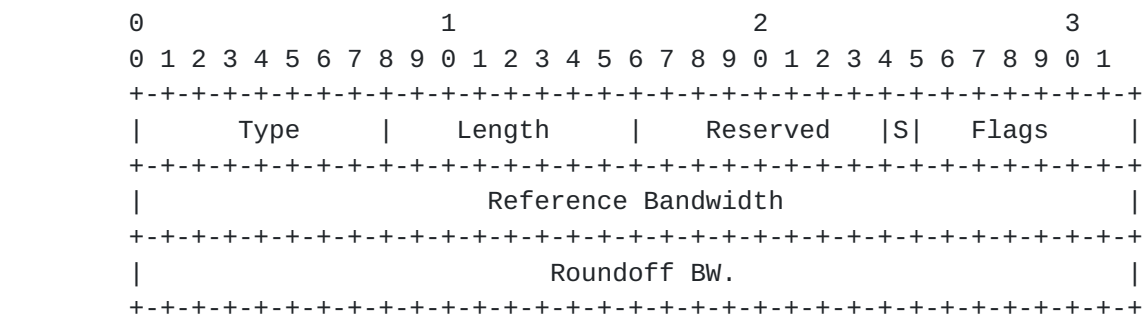
2. Staircase metric values based on bandwidth thresholds

The cumulative metric of the parallel link is used and based on reference bandwidth or staircase metric assignment method, metric value is derived. The derived metric is assigned to each of the parallel link between nodes. Interface group mode is useful for deployments that do not use L2 bundles.

4.3. ISIS FAD constraint sub-TLVs for automatic metric calculation

4.3.1. Reference Bandwidth sub-TLV

The Flexible Algorithm Definition Reference Bandwidth Sub-TLV (FADRB Sub-TLV) is a Sub-TLV of the ISIS FAD sub-TLV. It has the following format:



where:

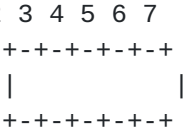
Type: TBD

Length: 10 octets.

Reference Bandwidth: link bandwidth is encoded in 32 bits in IEEE floating point format. The units are bytes per second.

Roundoff BW: link bandwidth is encoded in 32 bits in IEEE floating point format. The units are bytes per second.

Flags:



S-flag: when set, interface group Mode MUST be used to derive total link bandwidth.

Metric calculation: (Reference_bandwidth) /
(Total_link_bandwidth -
(Mod of(Total_link_bandwidth,roundoff_bw)))

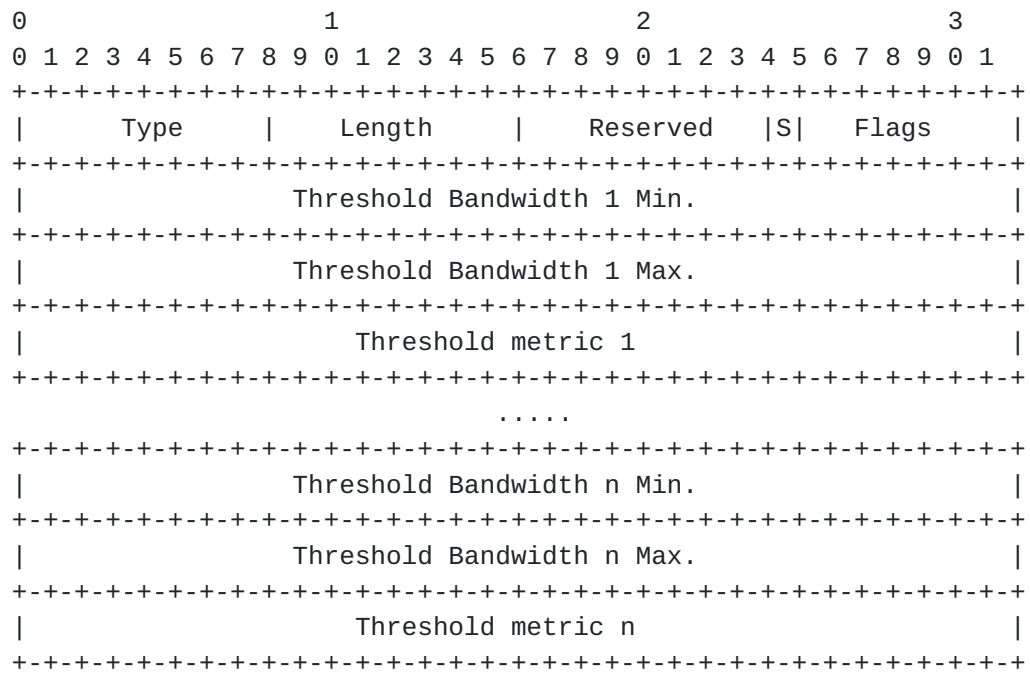
Round-off BW value is used to make sure the the metric does not change when there is smaller change in the link bandwidth.

Figure 8: ISIS FADRB sub-TLV

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

4.3.2. Threshold metric sub-TLV

The Flexible Algorithm Definition Threshold Bandwidth Sub-TLV (FADTB Sub-TLV) is a Sub-TLV of the ISIS FAD sub-TLV. It has the following format:

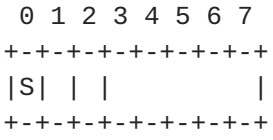


where:

Type: TBD

Length: 2 + n*12 octets.Here n is equal to Number of Threshold Bandwidth sets specified.

Flags:



S-flag: when set, interface group Mode MUST be used to derive total link bandwidth.

Staircase bandwidth threshold and associated metric values.
Threshold Bandwidth 1 Min.: Minimum link bandwidth is encoded in 32 bits in IEEE
floating point format. The units are bytes per second.
Threshold Bandwidth 1 Max.:Maximum link bandwidth is encoded in 32 bits in IEEE
floating point format. The units are bytes per second.
Threshold metric 1 : metric value range (1 - 4,294,967,296)

When the computed link bandwidth is in the range specified by
Min and Max threshold bandwidth values,
the corresponding metric value is assigned to the link during SPF calculation.

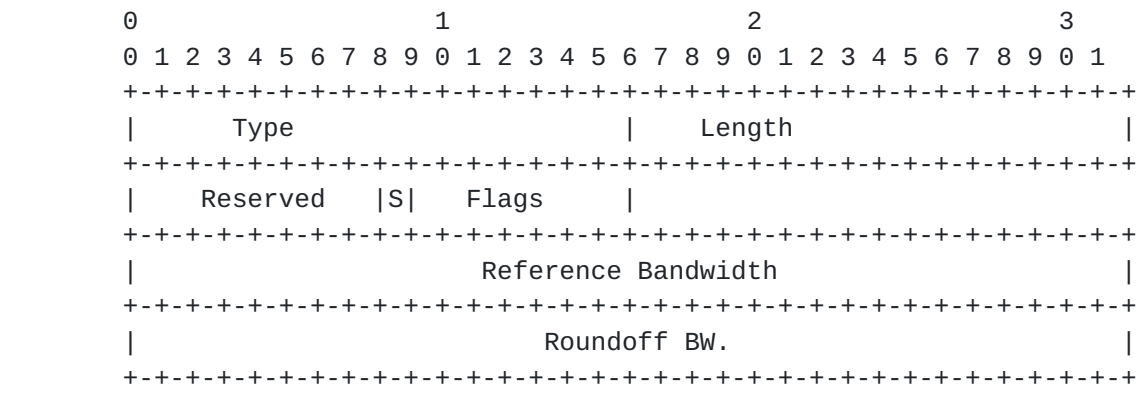
Figure 9: ISIS FADTB sub-TLV

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

4.4. OSPF FAD constraint sub-TLVs for automatic metric calculation

4.4.1. Reference Bandwidth sub-TLV

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



where:

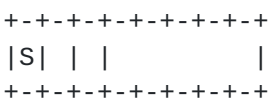
Type: TBD

Length: 10 octets.

Reference Bandwidth: link bandwidth is encoded in 32 bits in IEEE floating point format. The units are bytes per second.

Roundoff BW: link bandwidth is encoded in 32 bits in IEEE floating point format. The units are bytes per second.

Flags:



S-flag: when set, interface group Mode MUST be used to derive total link bandwidth.

Metric calculation: (Reference_bandwidth) / (Total_link_bandwidth - (Mod of(Total_link_bandwidth,roundoff_bw)))

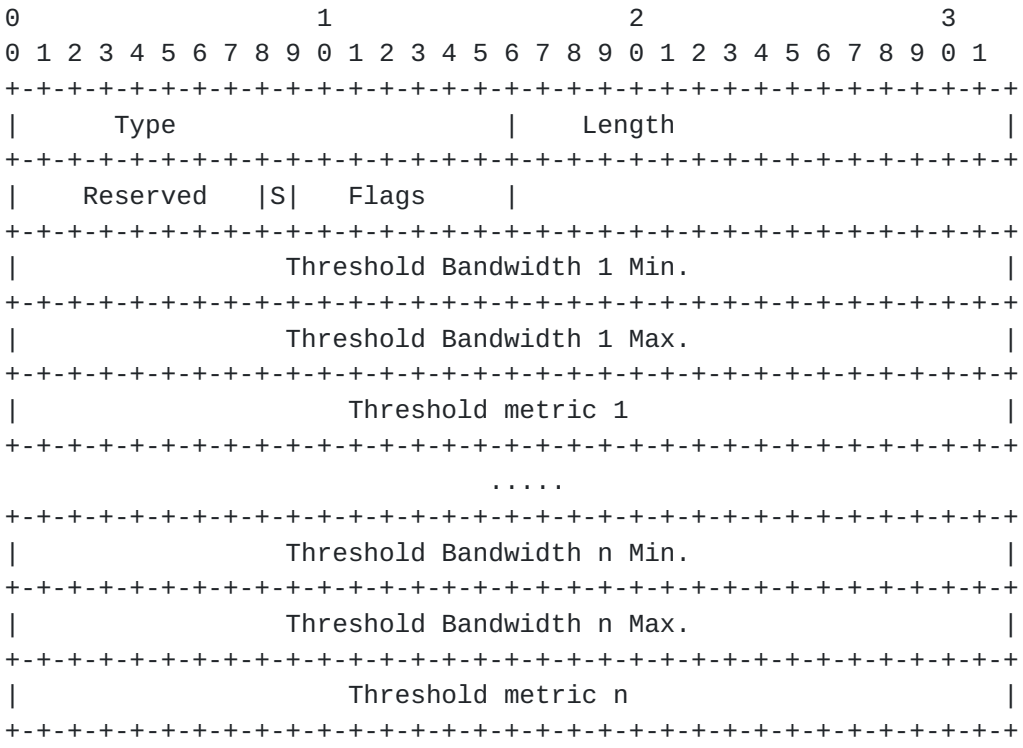
Round-off BW value is used to make sure the the metric does not change when there is smaller change in the link bandwidth.

Figure 10: OSPF FADRB sub-TLV

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

4.4.2. Threshold metric sub-TLV

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

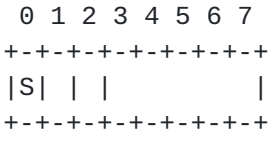


where:

Type: TBD

Length: 2 + n*12 octets.Here n is equal to Number of Threshold Bandwidth specified.

Flags:



S-flag: when set, interface group Mode MUST be used to derive total link bandwidth.

Stircase bandwidth threshold and associated metric values.
Threshold Bandwidth 1 Min.: Minimum link bandwidth is encoded in 32

bits in IEEE
floating point format. The units are bytes per
second.

Threshold Bandwidth 1 Max.:Maximum link bandwidth is encoded in 32 bits in IEEE
floating point format. The units are bytes per second.
Threshold metric 1 : metric value range (1 - 4,294,967,296)

When the computed link bandwidth is in the range specified by Min and Max threshold bandwidth values,
the corresponding metric value is assigned to the link during SPF calculation.

Figure 11: OSPF FADTB sub-TLV

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

5. Backward Compatibility

6. Security Considerations

TBD

7. IANA Considerations

7.1. IGP Metric-Type Registry

Type: Suggested 3 (TBA)

Description: Bandwidth metric

Reference: This document

7.2. ISIS Sub-Sub-TLVs for Flexible Algorithm Definition Sub-TLV

Type: Suggested 6 (TBA)

Description: ISIS Exclude Minimum Bandwidth sub-TLV

Reference: This document Section 3.1.1

Type: Suggested 7 (TBA)

Description: ISIS Exclude Maximum Delay sub-TLV

Reference: This document Section 3.1.2

Type: Suggested 8 (TBA)

Description: ISIS Reference Bandwidth sub-TLV

Reference: This document Section 4.3.1

Type: Suggested 9 (TBA)

Description: ISIS Threshold metric sub-TLV

Reference: This document Section 4.3.2

7.3. OSPF Sub-TLVs for Flexible Algorithm Definition Sub-TLV

Type: Suggested 6 (TBA)

Description: OSPF Exclude Minimum Bandwidth sub-TLV

Reference: This document Section 3.2.1

Type: Suggested 7 (TBA)

Description: OSPF Exclude Maximum Delay sub-TLV

Reference: This document Section 3.2.2

Type: Suggested 8 (TBA)

Description: OSPF Reference Bandwidth sub-TLV

Reference: This document Section 4.4.1

Type: Suggested 9 (TBA)

Description: OSPF Threshold metric sub-TLV

Reference: This document Section 4.4.2

7.4. Sub-TLVs for TLVs 22, 23, 25, 141, 222, and 223

Type: Suggested 45 (TBA)

Description: Bandwidth metric

Reference: This document Section 2.1

7.5. Sub-sub-TLV Codepoints for Application-Specific Link Attributes

Type: Suggested 45 (TBA)

Description: Bandwidth metric

Reference: This document Section 2.1

7.6. OSPFv2 Extended Link TLV Sub-TLVs

Type: Suggested 45 (TBA)

Description: Bandwidth metric

Reference: This document Section 2.2

7.7. Types for sub-TLVs of TE Link TLV (Value 2)

Type: Suggested 45 (TBA)

Description: Bandwidth metric

Reference: This document Section 2.2

8. Acknowledgements

Many thanks to Chris Bowers, Krzysztof Szarcowitz, Julian Lucek, Ram Santhanakrishnan for discussions and inputs.

9. Contributors

1. Salih K A

Juniper Networks

salih@juniper.net

10. References

10.1. 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.

[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>.

[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>.

[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>.

[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>.

10.2. Informative References

[I-D.bashandy-rtgwg-segment-routing-uloop]
Bashandy, A., Filsfils, C., Litkowski, S., Decraene, B., Francois, P., and P. Psenak, "Loop avoidance using Segment Routing", draft-bashandy-rtgwg-segment-routing-uloop-10 (work in progress), December 2020.

[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>.

[RFC5311] McPherson, D., Ed., Ginsberg, L., Previdi, S., and M. Shand, "Simplified Extension of Link State PDU (LSP) Space for IS-IS", RFC 5311, DOI 10.17487/RFC5311, February 2009, <https://www.rfc-editor.org/info/rfc5311>.

[RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, DOI 10.17487/RFC5316, December 2008, <https://www.rfc-editor.org/info/rfc5316>.

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

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

Authors' Addresses

Shraddha Hegde
Juniper Networks Inc.
Exora Business Park
Bangalore, KA 560103
India

Email: shraddha@juniper.net

William Britto A J
Juniper Networks Inc.

Email: bwilliam@juniper.net

Rajesh Shetty
Juniper Networks Inc.

Email: mrajesh@juniper.net

Bruno Decraene
Orange

Email: bruno.decraene@orange.com