Workgroup: SPRING
Internet-Draft:
draft-ietf-lsr-flex-algo-bw-con-02
Published: 23 March 2022
Intended Status: Standards Track
Expires: 24 September 2022
Authors: S. Hegde, Juniper Networks Inc.
W. Britto, Juniper Networks Inc.
R. Shetty, Juniper Networks Inc.
B. Decraene, Orange
P. Psenak, Cisco Systems
T. Li, Arista Networks

Flexible Algorithms: Bandwidth, Delay, Metrics and Constraints

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 generic metric type and 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

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 capacity links. Accordingly, many network operators define a link's metric relative to its capacity to help direct traffic to higher bandwidth links, but this is no guarantee that lower bandwidth links will be avoided, especially in failure scenarios. To ensure that elephant flows are only placed on high capacity links, it would be useful to explicitly exclude the high bandwidth traffic from utilizing links below a certain capacity. Flex-Algorithm [I-D.ietf-lsr-flex-algo] has already defined as a set of parameters consisting of calculation-type, metric-type and a set of constraints for allowing operators to have more control over network path computation. In this document, we define further extensions to Flex-Algorithm that will allow operators additional control over their traffic, especially with respect to constraints about bandwidth.

Historically, IGPs have done path computation by minimizing the sum of the link metrics along the path from source to destination. While the metric has been administratively defined, implementations have defaulted to a metric that is inversely proportional to link bandwidth. This has driven traffic to higher bandwidth links and has required manual metric manipulation to achieve the desired loading of the network.

Over time, with the addition of different traffic types, the need for alternate types of metrics has become clear. Flex-Algorithm already supports using the minimum link delay and the administratively assigned traffic-engineering metrics in path computation. However, it is clear that additional metrics may be of interest in different situations. A network operator may seek to minimize their operational costs and thus may want a metric that reflects the actual fiscal costs of using a link. Other traffic may require low jitter, leading to an entirely different set of metrics. With Flex-Algorithm, all of these different metrics, and more, could be used concurrently on the same network.

In some circumstances, path computation constraints, such as administrative groups, can be used to ensure that traffic avoids particular portions of the network. These strict constraints are appropriate when there is an absolute requirement to avoid parts of the topology, even in failure conditions. If, however, the requirement is less strict, then using a high metric in a portion of the topology may be more appropriate.

This document defines a family of generic metrics that can carry various types of administratively assigned metrics. This document proposes standard metric-types which require specific standard document. This document also proposes user defined metric-types where specifics are not defined, so that adminstrators are free to assign semantics as they fit. This document also specifies a new bandwidth based metric type to be used with Flex-Algorithm and other applications in Section Section 4. Additional Flexible Algorithm Definition (FAD) constraints are defined in Section Section 3 that allow the network adminstrator to preclude the use of low bandwidth links or high delay links. Section Section 4.1 defines mechanisms to automatically calculate link metrics based on parameters defined in the FAD and the advertised Maximum Link Bandwidth of each link. This is advantageous because administrators can change their criteria for metric assignment centrally, without individual modification of each link metric throughout the network.

2. Generic Metric Advertisement

ISIS and OSPF advertise a metric for each link in their respective link state advertisements. Multiple metric types are are already supported. Administratively assigned metrics are described in the original OSPF and ISIS specifications. The Traffic Engineering Default Metric is defined in [RFC5305] and [RFC3630] and the Min Unidirectional delay metric is defined in [RFC8570] and [RFC7471]. Other metrics, such as jitter, reliability, and fiscal cost may be helpful, depending on the traffic class. Rather than attempt to enumerate all possible metrics of interest, this document specifies a generic mechanism for advertising metrics.

Each generic metric advertisement is on a per-link and per metric type basis. The metric advertisement consists of a metric type field and a value for the metric. The metric type field is assigned by the "IGP metric type" IANA registry. Metric types 0-127 are standard metric types as assigned by IANA. This document further specifies a user defined metric type space of metric types 128-255. These are user defined and can be assigned by an operator for local use.

2.1. ISIS Generic Metric sub-TLV

The ISIS Generic Metric sub-TLV specifies the link metric for a given metric type. Typically, this metric is assigned by a network administrator. Generic metric is application-independent attribute similar to igp-metric. The Generic Metric sub-TLV is advertised in the TLVs/sub-TLVs below:

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]

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type        |     Length    |  metric-type  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Value                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

   Type  :   TBD (To be assigned by IANA)
   Length: 4 octets
   metric-type:  A value from the IGP metric-type registry
   Value : metric value range (1 - 16,777,215)
```

Figure 1: ISIS Generic Metric sub-TLV

The Generic Metric sub-TLV MAY be advertised multiple times. For a particular metric type, the Generic Metric sub-TLV MUST be advertised only once for a link when advertised in TLV 22,222,23,223 and 141. If there are multiple Generic Metric sub-TLVs advertised for a link for same metric type in one or more received LSPDUs, advertisement in the lowest numbered fragment MUST be used and the subsequent ones MUST be ignored.If the metric type indicates a standard metric type for which there are other advertisement mechanisms (e.g., the IGP metric, the Min Unidirectional Link Delay, or the Traffic Engineering Default Metric, as of this writing), the Generic Metric advertisement MUST be ignored.

2.2. OSPF Generic Metric sub-TLV

The OSPF Generic Metric sub-TLV specifies the link metric for a given metric type. Typically, this metric is assigned by a network administrator.Generic metric is application-independent attribute similar to igp-metric. The Generic Metric sub-TLV is advertised in the TLVs below:

sub-TLV of the OSPF Link TLV of OSPF extended Link LSA [RFC7684].

sub-TLV of TE Link TLV (2) of OSPF TE LSA [RFC3630].

sub-TLV of the Router-Link TLV in the E-Router-LSA in OSPFv3 [RFC8362].

The Generic Metric sub-TLV is TLV type TBD (IANA), and is eight octets in length.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Type             |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| metric-type   |         Reserved                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Value                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Type  :   TBD (To be assigned by IANA)
Length: 8 octets
metric-type = A value from the IGP metric type registry

Value : metric value (1- 4,294,967,295)
```

Figure 2: OSPF Generic Metric sub-TLV

The Generic Metric sub-TLV MAY be advertised multiple times. For a particular metric type, the Genreric Metric sub-TLV MUST be advertised only once for a link when advertised in OSPF Link TLV of Extended Link LSA, Link TLV of TE LSA and sub-TLV of the Router-Link TLV in the E-Router-LSA Router-Link TLV in OSPFv3. If there are multiple Genreric Metric sub-TLVs advertised for a link for the same metric type in a received LSA, the first one MUST be used and the subsequent ones MUST be ignored.If the metric type indicates a standard metric type for which there are other advertisement mechanisms (e.g., the IGP metric, the Min Unidirectional Link Delay, or the Traffic Engineering Default Metric, as of this writing), the Generic Metric advertisement MUST be ignored.

2.3. Generic Metric applicability to Flexible Algorithms Multi-domain/ Multi-area networks

Generic Metric can be used by Flex-Algorithms by specifying the metric type in the Flexible Algorithm Definitions. When Flex-Algorithms is used in a multi-area network, [I-D.ietf-lsr-flex-algo] defines FAPM sub-TLV that carries the Flexible Algorithm specific metric. Metric carried in FAPM will be equal to the metric to reach the prefix for that Flex-Algorithm in its source area or domain. When Flex-Algorithm uses Generic metric, the same procedures as described in section 13 of [I-D.ietf-lsr-flex-algo] are used to send and process FAPM sub-TLV.

3. FAD constraint sub-TLVs

In networks that carry elephant flows, directing an elephant flow down a low-bandwidth link would be catastrophic. Thus, in the context of Flex-Algorithm, it would be useful to be able to constrain the topology to only those links capable of supporting a minimum amount of bandwidth.

If the capacity of a link is constant, this can already be achived through the use of administrative groups. However, when a Layer 3 link is actually a collection of Layer 2 links (LAG/Layer 2 Bundle), the link bandwidth will vary based on the set of active constituent links. This could be automated by having an implementation vary the advertised administrative groups based on bandwidth, but this seems unnecessarily complex and expressing this requirement as a direct constraint on the topology seems simpler. This is also advantageous if the minimum required bandwidth changes, as this constraint would provide a single centralized, coordinated point of control.

To implement this idea, this document defines a new Exclude Minimum Bandwidth constraint. When this constraint is advertised in a FAD, a link will be pruned from the Flex-Algorithm topology if the link's advertised Maximum Link Bandwidth is below the advertised Minimum Bandwidth value.

Similarly, this document defines a Exclude Maximum Link Delay constraint. Delay is an important consideration in High Frequency Trading applications, networks with transparent L2 link recovery, or in satellite networks, where link delay may fluctuate. Mechanisms already exist to measure the link delay dynamically and advertised it in the IGP. Networks that employ dynamic link delay measurement, may want to exclude links that have a delay over a given threshold.

3.1. ISIS FAD constraint sub-TLVs

3.1.1. ISIS Exclude Minimum Bandwidth sub-TLV

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

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |    Length     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Min Bandwidth                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

Type: TBA

Length: 4 octets.

Min Bandwidth: The 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 at most once in the FAD sub-TLV. If it appears more than once, the ISIS FAD Sub-TLV MUST be ignored by the receiver.

The Minimum bandwidth advertised in FAEMB sub-TLV MUST be compared with Maximum Link Bandwidth advertised in sub-sub-TLV 9 of ASLA sub-TLV [RFC 8919]. If L-Flag is set in the ASLA sub-TLV, the Minimum bandwidth advertised in FAEMB sub-TLV MUST be compared with Maximum Link Bandwidth as advertised by the sub-TLV 9 of the TLV 22/222/23/223/141 [RFC 5305] as defined in [RFC8919] Section 4.2.

If the Maximum Link Bandwidth is lower than the Minimum link bandwidth advertised in FAEMB sub-TLV, the link MUST be excluded from the Flex-Algorithm topology. If a link does not have the Maximum Link Bandwidth advertised but the FAD contains this sub-TLV, then that link MUST NOT be excluded from the topology based on the Minimum Bandwidth constraint.

3.1.2. ISIS Exclude Maximum Delay sub-TLV

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

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |    Length     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     max link delay          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
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 Maximum link delay advertised in FAEMD sub-TLV MUST be compared with Min Unidirectional Link Delay advertised in sub-sub-TLV 34 of ASLA sub-TLV [RFC 8919]. If L-Flag is set in the ASLA sub-TLV, the Maximum link delay advertised in FAEMD sub-TLV MUST be compared with Min Unidirectional Link Delay as advertised by the sub-TLV 34 of the TLV 22/222/23/223/141 [RFC 8570] as defined in [RFC8919] Section 4.2.

If the Min Unidirectional Link Delay value is higher than the Maximum link delay advertised in FAEMD sub-TLV, the link MUST be excluded from the Flex-Algorithm topology. If a link does not have the Min Unidirectional Link Delay advertised but the FAD contains this sub-TLV, then that link MUST NOT be excluded from the topology based on the Maximum Delay constraint.

3.2. OSPF FAD constraint sub-TLVs

3.2.1. OSPF Exclude Minimum Bandwidth sub-TLV

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

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type                     |    Length                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Min Bandwidth                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
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 Maximum Link Bandwidth as advertised in Extended Link TLV in the Extended Link Opaque LSA in OSPFv2 [RFC7684] or as a sub-TLV of the Router-Link TLV in the E-Router-LSA Router-Link TLV in OSPFv3 [RFC8362] MUST be compared against the Minimum bandwidth advertised in FAEMB sub-TLV. If the link bandwidth is lower than the Minimum bandwidth advertised in FAEMB sub-TLV, the link MUST be excluded from the Flex-Algorithm topology. If a link does not have the Maximum Link Bandwidth advertised but the FAD contains this sub-TLV, then that link MUST be included in the topology and proceed to apply further pruning rules for the link.

3.2.2. OSPF Exclude Maximum Delay sub-TLV

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

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type                     |    Length                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Max Delay                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
where:

   Type: TBD

   Length:  3 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 Min Unidirectional Link Delay as advertised by sub-sub-TLV 12 of ASLA sub-TLV [RFC 8920], MUST be compared against the Maximum delay advertised in FAEMD sub-TLV. If the Min Unidirectional Link Delay is higher than the Maximum delay advertised in FAEMD sub-TLV, the link MUST be excluded from the Flex-Algorithm topology If a link does not have the Min Unidirectional Link Delay advertised but the FAD contains this sub-TLV, then then that link MUST NOT be excluded from the topology based on the Maximum Delay constraint.

4. Bandwidth Metric Advertisement

Historically, IGP implementations have made default metric assignments based on link bandwidth. This has proven to be useful, but has suffered from having different defaults across implementations and from the rapid growth of link bandwidths. With Flex-Algorithm, the network administrator can define a function that will produce a metric for each link have each node automatically compute each link's metric based its bandwidth.

This document defines a new standard metric type for this purpose called the "Bandwidth Metric". The Bandwidth Metric MAY be advertised in the Generic Metric sub-TLV with the metric type set to "Bandwidth Metric". ISIS and OSPF will advertise this new type of metric in their link advertisements. Bandwidth metric is a link attribute and for advertisement and processing of this attribute for Flex-algorithm purposes, MUST follow the the section 12 of [I-D.ietf-lsr-flex-algo]

Flex-Algorithm uses this metric type by specifying the bandwidth metric as the metric type in a FAD TLV. A FAD TLV may also specify an automatic computation of the bandwidth metric based on a links advertised bandwidth. An explicit advertisement of a link's bandwidth metric using the Generic Metric sub-TLV overrides this automatic computation. The automatic bandwidth metric calculation sub-TLVs are advertised in FAD TLV and these parameters are applicable to applications such as Flex-algorithm that make use of the FAD TLV.

4.1. Automatic Metric Calculation

Networks which are designed to be highly regular and follow uniform metric assignment may want to simplify their operations by automatically calculating the bandwidth metric. When a FAD advertises the metric type as Bandwidth Metric and the link does not have the Bandwidth Metric advertised, automatic metric derivation can be used with additional FAD constraint advertisements as described in this section.

If a link's bandwidth changes, then the delay in learning about the change may create the possibility of micro-loops in the topology. This is no different from the IGP's susceptibility to micro-loops during a 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.

Computing the metric between adjacent systems based on bandwidth becomes more complex in the face of parallel adjacencies. If there are parallel adjacnecies between systems, then the bandwidth between the systems is the sum of the bandwidth of the parallel links. This is somewhat more complex to deal with, so there is an optional mode for computing the aggregate bandwidth.

4.1.1. Automatic Metric Calculation Modes

4.1.1.1. Simple Mode

In simple mode, the Maximum Link Bandwidth of a single Layer 3 link is used to derive the metric. This mode is suitable for deployments that do not use parallel Layer 3 links. In this case, the computation of the metric is straightforward. If a layer 3 link is composed of a layer 2 bundle, then the link bandwidth is the sum of the bandwidths of the working components and may vary with layer 2 link failures.

4.1.1.2. Interface Group Mode

The simple mode of metric calculation may not work well when there are multiple parallel layer 3 interfaces between two nodes. Ideally, the metric between two systems should be the same given the same bandwidth, whether the bandwidth is provided by parallel layer 2 links or parallel layer 3 links. To address this, in Interface Group Mode, nodes MUST compute the aggregate bandwidth of all parallel adjacencies, MUST derive the metric based on the aggregate bandwidth, and MUST apply the resulting metric to each of the parallel adjacencies.

```
     A------B====C====F====D
            |              |
             ------E-------
```

Figure 7: Parallel interfaces

For exmple, in the above diagram, there are two parallel links between B->C, C->F, F->D. Let us assume the link bandwidth is uniform 10Gbps on all links and the metric for each link will be the same. Traffic from B to D will be forwarded B->E->D. Since the bandwidth is higher on the B->C->F->D path, the metric for that path should be lower, and that path should be selected. Interface Group Mode is preferred in cases where there are parallel layer 3 links.

In the interface group mode, every node MUST identify the set of parallel links between a pair of nodes based on IGP link advertisements and MUST consider cumulative bandwidth of the parallel links while arriving at the metric of each link.

4.1.2. Automatic Metric Calculation Methods

In automatic metric calculation for simple and interface group mode, Maximum Link Bandwidth of the links is used to derive the metric. There are two types of automatic metric derivation methods.

1. Reference bandwidth method

2. Bandwidth thresholds method

4.1.2.1. Reference Bandwidth method

In many networks, the metric is inversely proportional to the link bandwidth. The administrator or implementation selects a reference bandwdith and the metric is derived by dividing the reference bandwidth by the advertised Maximum Link Bandwidth. Advertising the reference bandwidth in the FAD constraints allows the metric computation to be done automatically. Centralized control of this reference bandwidth simplifies management in the case that the reference bandwidth changes. In order to ensure that small bandwidth changes do not change the link metric, it is useful to define the granularity of the bandwidth that is of interest. The link bandwidth will be truncated to this granularity before deriving the metric.

For example,

reference bandwidth = 1000G

Granularity = 20G

The derived metric is 10 for link bandwidth in the range 100G to 119G

4.1.2.2. Bandwidth Thresholds method

The reference bandwidth approach described above provides a uniform metric value for a range of link bandwidths. In certain cases there may be a need to define non-proportional metric values for the varying ranges of link bandwidth. For example, bandwidths from 10G to 30G are assigned metric value 100, bandwidth from 30G to 70G get a metric value of 50, and bandwidths greater than 70G have a metric of 10. In order to support this, a staircase mapping based on bandwidth thresholds is supported in the FAD. This advertisement contains a set of threshold values and associated metrics.

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

4.1.3.1. Reference Bandwidth sub-TLV

This section provides FAD constraint advertisement details for the reference bandwidth method of metric calculation as described in Section 4.1.2.1. 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:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |    Length     |     Flags     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Reference Bandwidth                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Granularity Bandwidth                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+


where:

   Type: TBD

   Length: 9 octets.
   Reference Bandwidth: Bandwidth encoded in 32 bits in IEEE floating
                        format. The units are in bytes per second.
   Granularity Bandwidth: Bandwidth encoded in 32 bits in IEEE floati
                        format. The units are in bytes per second.

Flags:

              0 1 2 3 4 5 6 7
             +-+-+-+-+-+-+-+-+
             |G| | |         |
             +-+-+-+-+-+-+-+-+

      G-flag: when set, interface group Mode MUST be used to derive t

      Metric calculation: (Reference_bandwidth) /
                           (Total_link_bandwidth -
                           (Modulus of(Total_link_bandwidth,granulari
```

Figure 8: ISIS FADRB sub-TLV

Granularity Bandwidth value ensures that the metric does not change when there is a small change in the link bandwidth. The ISIS FADRB Sub-TLV MUST NOT appear more than once in an ISIS FAD sub-TLV. If it appears more than once, the ISIS FAD sub-TLV MUST be ignored by the receiver. If a Generic Metric sub-TLV with Bandwidth metric type is advertised for a link, the Flex-Algorithm calculation MUST use the advertised Bandwidth Metric, and MUST NOT use the automatically derived metric for that link.

4.1.3.2. Bandwidth Thresholds sub-TLV

This section provides FAD constraint advertisement details for the Bandwidth Thresholds method of metric calculation as described in Section 4.1.2.2. The Flexible Algorithm Definition Bandwidth Threshold Sub-TLV (FADBT Sub-TLV) is a Sub-TLV of the ISIS FAD sub-TLV. It has the following format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |    Length     |       Flags   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Bandwidth Threshold 1                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Threshold Metric 1        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Bandwidth Threshold 1                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Threshold Metric 2        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Bandwidth Threshold 2                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                               .....
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Threshold Metric n-1      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Bandwidth Threshold n-1                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Threshold Metric n        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

Type: TBD

Length: 1 + n*7 octets. Here n is equal to number of Threshold Met
n MUST be greater than or equal to 1.

Flags:

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|G| | |         |
+-+-+-+-+-+-+-+-+
```

G-flag: when set, interface group Mode MUST be used to derive t

Staircase bandwidth threshold and associated metric values.
Bandwidth Threshold 1: Minimum Link Bandwidth is encoded in 32
floating point format. The units are bytes p
Bandwidth Threshold 2: Maximum Link Bandwidth is encoded in 32
floating point format. The units are bytes p
Threshold Metric 1 : metric value range (1 - 4,261,412,864)

Figure 9: ISIS FADBT sub-TLV

When G-flag is set, the cumulative bandwidth of the parallel links is computed as described in section Section 4.1.1.2. If G-flag is not set, the advertised Maximum Link Bandwidth is used.

When the computed link bandwidth is less than Bandwidth Threshold 1, the MAX_METRIC value of 4,261,412,864 MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

When the computed link bandwidth is greater than or equal to Bandwidth Threshold 1 and less than Bandwidth Threshold 1, Threshold Metric 1 MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

Similarly, when the computed link bandwidth is greater than or equal to Bandwidth Threshold 1 and less than Bandwidth Threshold 2, Threshold Metric 2 MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

In general, when the computed link bandwidth is greater than or equal to Bandwidth Threshold X AND less than Bandwidth Threshold X+1, Threshold Metric X MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

Finally, when the computed link bandwidth is greater than or equal to Bandwidth Threshold n, then Threshold Metric n MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

The ISIS FADBT Sub-TLV MUST NOT appear more than once in an ISIS FAD sub-TLV. If it appears more than once, the ISIS FAD sub-TLV MUST MUST stop participating in such flex-algorithm.

A FAD MUST NOT contain both FADBT sub-TLV and FADRB sub-TLV. If both these sub-TLVs are advertised in the same FAD for a Flexible Algorithm, the FAD MUST be ignored by the receiver.

If a Generic Metric sub-TLV with Bandwidth metric type is advertised for a link, the Flex-Algorithm calculation MUST use the Bandwidth Metric advertised on the link, and MUST NOT use the automatically derived metric for that link.

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

4.1.4.1. 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:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type                     |    Length                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Flags   |     Reserved                                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Reference Bandwidth                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Granularity Bandwidth                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+


where:

   Type: TBD

   Length: 14 octets.
   Reference Bandwidth: Bandwidth encoded in 32 bits in IEEE floating
                        format. The units are in bytes per second.
   Granularity Bandwidth: Bandwidth encoded in 32 bits in IEEE floati
                        format. The units are in bytes per second.

Flags:

              0 1 2 3 4 5 6 7
             +-+-+-+-+-+-+-+-+
             |G| | |         |
             +-+-+-+-+-+-+-+-+

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

      Metric calculation: (Reference_bandwidth) /
                           (Total_link_bandwidth -
                           (Modulus of(Total_link_bandwidth, Granular
```

Figure 10: OSPF FADRB sub-TLV

Granularity Bandwidth value is used to ensure that the metric does not change when there is a small change in the link bandwidth. The OSPF FADRB Sub-TLV MUST NOT appear more than once in an OSPF FAD TLV. If it appears more than once, the OSPF FAD TLV MUST be ignored by the receiver. If a Generic Metric sub-TLV with Bandwidth metric type is advertised for a link, the Flex-Algorithm calculation MUST use the advertised Bandwidth Metric on the link, and MUST NOT use the automatically derived metric for that link.

4.1.4.2. Bandwidth Threshold sub-TLV

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

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type                     |    Length                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Flags    | Reserved                                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Bandwidth Threshold 1                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Threshold Metric 1                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Bandwidth Threshold 2                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Threshold Metric 2                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Bandwidth Threshold 3                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                              .....
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Threshold Metric n-1                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Bandwidth Threshold n                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Threshold Metric n                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

Type: TBD

Length: 2 + n*8 octets. Here n is equal to number of Threshold Metrics specified. n MUST be greater than or equal to 1.

Flags:

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|G| | |         |
+-+-+-+-+-+-+-+-+
```

G-flag: when set, interface group Mode MUST be used to derive t

Staircase bandwidth threshold and associated metric values.
Bandwidth Threshold 1: Minimum Link Bandwidth is encoded in 32 floating point format. The units are bytes p

```
Bandwidth Threshold 2: Maximum Link Bandwidth is encoded in 32
                  floating point format.  The units are bytes p
Threshold Metric 1 : metric value range (1 - 4,294,967,296)
```

Figure 11: OSPF FADBT sub-TLV

When G-flag is set, the cumulative bandwidth of the parallel links is computed as described in section Section 4.1.1.2. If G-flag is not set, the advertised Maximum Link Bandwidth is used.

When the computed link bandwidth is less than Bandwidth Threshold 1 , the MAX_METRIC value of 4,294,967,296 MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

When the computed link bandwidth is greater than or equal to Bandwidth Threshold 1 and less than Bandwidth Threshold 1, Threshold Metric 1 MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

Similarly, when the computed link bandwidth is greater than or equal to Bandwidth Threshold 1 and less than Bandwidth Threshold 2, Threshold Metric 2 MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

In general, when the computed link bandwidth is greater than or equal to Bandwidth Threshold X AND less than Bandwidth Threshold X+1, Threshold Metric X MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

Finally, when the computed link bandwidth is greater than or equal to Bandwidth Threshold n, then Threshold Metric n MUST be assigned as the Bandwidth Metric on the link during Flex-Algorithm SPF calculation.

The ISIS FADBT Sub-TLV MUST NOT appear more than once in an ISIS FAD sub-TLV. If it appears more than once, the ISIS FAD sub-TLV MUST stop participating in such flex-algorithm.

A FAD MUST NOT contain both FADBT sub-TLV and FADRB sub-TLV. If both these sub-TLVs are advertised in the same FAD for a Flexible Algorithm, the FAD MUST be ignored by the receiver.

If a Generic Metric sub-TLV with Bandwidth metric type is advertised for a link, the Flex-Algorithm calculation MUST use the Bandwidth Metric advertised on the link, and MUST NOT use the automatically derived metric for that link.

5. Bandwidth metric considerations

This section specifies the rules of deriving the Bandwidth Metric if and only if the winning FAD for the Flex-Algorithm specifies the metric-type as "Bandwidth Metric".

1. If the the Generic Metric sub-TLV with Bandwidth metric type is advertised for the link as described in Section 4, it MUST be used during the Flex-Algorithm calculation.

2. If the Generic Metric sub-TLV with Bandwidth metric type is not advertised for the link and the winning FAD for the Flex-Algorithm does not specify the automatic bandwidth metric calculation (as defined in Section 4.1), the Bandwidth Metric is considered as not being advertised for the link.

3. If the Generic Metric sub-TLV with Bandwidth metric type is not advertised for the link and the winning FAD for the Flex-Algorithm specifies the automatic bandwidth metric calculation (as defined in Section 4.1), the Bandwidth Metric metric MUST be automatically calculated as per the procedures defined in Section 4.1. If the Bandwidth Metric can not be calculated due to lack of Flex-Algorithm specific ASLA advertisement of sub-sub-TLV 9 [RFC 8919], or in case of IS-IS, in presence of the L-Flag in the Flex-Algorithm specific ASLA advertisement the lack of sub-TLV 9 in the TLV 22/222/23/223/141 [RFC 5305], the Bandwidth Metric is considered as not being advertised for the link.

6. Calculation of Flex-Algorithm paths

Two new additional rules are added to the existing rules in the Flex-rules specified in sec 13 of [I-D.ietf-lsr-flex-algo].

6. Check if any exclude FAEMB rule is part of the Flex-Algorithm definition. If such exclude rule exists and the link has Maximum Link Bandwidth advertised, check if the link bandwidth satisfies the FAEMB rule. If the link does not satisfy the FAEMB rule, the link MUST be pruned from the computation.

7. Check if any exclude FAEMD rule is part of the Flex-Algorithm definition. If such exclude rule exists and the link has Min Unidirectional link delay advertised, check if the link delay satisfies the FAEMD rule. If the link does not satisfy the FAEMD rule, the link MUST be pruned from the computation.

7. Backward Compatibility

8. Security Considerations

TBD

9. IANA Considerations

9.1. IGP Metric-Type Registry

Type: Suggested 3 (TBA)

Description: Bandwidth metric

Reference: This document

Type: 128 to 255(TBA)

Description: User defned metric

Reference: This document

9.2. 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 4.1.3.1

Type: Suggested 9 (TBA)

Description: ISIS Threshold Metric sub-TLV

Reference: This document Section 4.1.3.2

9.3. 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 4.1.4.1

Type: Suggested 9 (TBA)

Description: OSPF Threshold Metric sub-TLV

Reference: This document Section 4.1.4.2

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

Type: Suggested 45 (TBA)

Description: Generic metric

Reference: This document Section 2.1

9.5. OSPFv2 Extended Link TLV Sub-TLVs

Type: Suggested 45 (TBA)

Description: Generic metric

Reference: This document Section 2.2

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

Type: Suggested 45 (TBA)

Description: Generic metric

Reference: This document Section 2.2

9.7. OSPFv3 Extended-LSA Sub-TLVs

Type: Suggested 45 (TBA)

Description: Generic metric

Reference: This document Section 2.2

10. Acknowledgements

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

11. Contributors

1. Salih K A

Juniper Networks

salih@juniper.net

12. References

12.1. Normative References

[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>.

[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>.

12.2. 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", Work in Progress, Internet-Draft, draft-bashandy-rtgwg-segment-routing-uloop-12, 22

December 2021, <https://www.ietf.org/archive/id/draft-bashandy-rtgwg-segment-routing-uloop-12.txt>.

[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 560103
KA
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

Peter Psenak
Cisco Systems

Email: ppsenak@cisco.com

Tony Li
Arista Networks

Email: tony.li@tony.li