

Networking Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 23, 2016

S. Previdi, Ed.
Cisco Systems, Inc.
S. Giacalone
Unaffiliated
D. Ward
Cisco Systems, Inc.
J. Drake
Juniper Networks
Q. Wu
Huawei
January 20, 2016

IS-IS Traffic Engineering (TE) Metric Extensions
draft-ietf-isis-te-metric-extensions-09

Abstract

In certain networks, such as, but not limited to, financial information networks (e.g. stock market data providers), network performance criteria (e.g. latency) are becoming as critical to data path selection as other metrics.

This document describes extensions to IS-IS Traffic Engineering Extensions ([RFC5305](#)) such that network performance information can be distributed and collected in a scalable fashion. The information distributed using IS-IS TE Metric Extensions can then be used to make path selection decisions based on network performance.

Note that this document only covers the mechanisms with which network performance information is distributed. The mechanisms for measuring network performance or acting on that information, once distributed, are outside the scope of this document.

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](#)].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying [RFC-2119](#) significance.

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 <http://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 July 23, 2016.

Copyright Notice

Copyright (c) 2016 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 (<http://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.	TE Metric Extensions to IS-IS	4
3.	Interface and Neighbor Addresses	5
4.	Sub TLV Details	6
4.1.	Unidirectional Link Delay Sub-TLV	6
4.2.	Min/Max Unidirectional Link Delay Sub-TLV	7
4.3.	Unidirectional Delay Variation Sub-TLV	8
4.4.	Unidirectional Link Loss Sub-TLV	8
4.5.	Unidirectional Residual Bandwidth Sub-TLV	9
4.6.	Unidirectional Available Bandwidth Sub-TLV	10
4.7.	Unidirectional Utilized Bandwidth Sub-TLV	11
5.	Announcement Thresholds and Filters	12
6.	Announcement Suppression	13
7.	Network Stability and Announcement Periodicity	13

8.	Enabling and Disabling Sub-TLVs	14
9.	Static Metric Override	14
10.	Compatibility	14
11.	Security Considerations	14
12.	IANA Considerations	15
13.	Contributors	15
14.	Acknowledgements	16
15.	References	16
15.1.	Normative References	16
15.2.	Informative References	17
	Authors' Addresses	17

[1.](#) Introduction

In certain networks, such as, but not limited to, financial information networks (e.g. stock market data providers), network performance information (e.g. latency) is becoming as critical to data path selection as other metrics.

In these networks, extremely large amounts of money rest on the ability to access market data in "real time" and to predictably make trades faster than the competition. Because of this, using metrics such as hop count or cost as routing metrics is becoming only tangentially important. Rather, it would be beneficial to be able to make path selection decisions based on performance data (such as latency) in a cost-effective and scalable way.

This document describes extensions (hereafter called "IS-IS TE Metric Extensions") to IS-IS Extended Reachability TLV defined in [[RFC5305](#)], that can be used to distribute network performance information (such as link delay, delay variation, packet loss, residual bandwidth, and available bandwidth).

The data distributed by the IS-IS TE Metric Extensions proposed in this document is meant to be used as part of the operation of the routing protocol (e.g. by replacing cost with latency or considering bandwidth as well as cost), by enhancing Constrained-SPF (CSPF), or for other uses such as supplementing the data used by an ALTO server [[RFC7285](#)]. With respect to CSPF, the data distributed by IS-IS TE Metric Extensions can be used to setup, fail over, and fail back data paths using protocols such as RSVP-TE [[RFC3209](#)].

Note that the mechanisms described in this document only disseminate performance information. The methods for initially gathering that performance information, such as [[RFC6375](#)], or acting on it once it is distributed are outside the scope of this document. Example mechanisms to measure latency, delay variation, and loss in an MPLS network are given in [[RFC6374](#)]. While this document does not specify

how the performance information should be obtained, the measurement of delay SHOULD NOT vary significantly based upon the offered traffic load. Thus, queuing delays SHOULD NOT be included in the delay measurement. For links, such as Forwarding Adjacencies, care must be taken that measurement of the associated delay avoids significant queuing delay; that could be accomplished in a variety of ways, including either by measuring with a traffic class that experiences minimal queuing or by summing the measured link delays of the components of the link's path.

2. TE Metric Extensions to IS-IS

This document proposes new IS-IS TE sub-TLVs that can be announced in TLVs 22, 141, 222, and 223 in order to distribute network performance information. The extensions in this document build on the ones provided in IS-IS TE [[RFC5305](#)] and GMPLS [[RFC4203](#)].

IS-IS Extended Reachability TLV 22 (defined in [[RFC5305](#)]), Inter-AS reachability information TLV 141 (defined in [[RFC5316](#)]) and MT-ISIS TLV 222 (defined in [[RFC5120](#)]) have nested sub-TLVs which permit the TLVs to be readily extended. This document proposes several additional sub-TLVs:

Type	Value

33 (Suggested)	Unidirectional Link Delay
34 (Suggested)	Min/Max Unidirectional Link Delay
35 (Suggested)	Unidirectional Delay Variation
36 (Suggested)	Unidirectional Packet Loss
37 (Suggested)	Unidirectional Residual Bandwidth
38 (Suggested)	Unidirectional Available Bandwidth
39 (Suggested)	Unidirectional Bandwidth Utilization

As can be seen in the list above, the sub-TLVs described in this document carry different types of network performance information. The new sub-TLVs include a bit called the Anomalous (or "A") bit. When the A bit is clear (or when the sub-TLV does not include an A bit), the sub-TLV describes steady state link performance. This information could conceivably be used to construct a steady state performance topology for initial tunnel path computation, or to verify alternative failover paths.

When network performance violates configurable link-local thresholds a sub-TLV with the A bit set is advertised. These sub-TLVs could be used by the receiving node to determine whether to fail traffic to a backup path, or whether to calculate an entirely new path. From an MPLS perspective, the intent of the A bit is to permit LSP ingress nodes to:

- A) Determine whether the link referenced in the sub-TLV affects any of the LSPs for which it is ingress. If there are, then:
- B) Determine whether those LSPs still meet end-to-end performance objectives. If not, then:
- C) The node could then conceivably move affected traffic to a pre-established protection LSP or establish a new LSP and place the traffic in it.

If link performance then improves beyond a configurable minimum value (reuse threshold), that sub-TLV can be re-advertised with the Anomalous bit cleared. In this case, a receiving node can conceivably do whatever re-optimization (or fallback) it wishes to do (including nothing).

Note that when a sub-TLV does not include the A bit, that sub-TLV cannot be used for failover purposes. The A bit was intentionally omitted from some sub-TLVs to help mitigate oscillations. See [Section 5](#) for more information.

Consistent with existing IS-IS TE specification [[RFC5305](#)], the bandwidth advertisements defined in this draft MUST be encoded as IEEE floating point values. The delay and delay variation advertisements defined in this draft MUST be encoded as integer values. Delay values MUST be quantified in units of microseconds, packet loss MUST be quantified as a percentage of packets sent, and bandwidth MUST be sent as bytes per second. All values (except residual bandwidth) MUST be calculated as rolling averages where the averaging period MUST be a configurable period of time. See [Section 5](#) for more information.

3. Interface and Neighbor Addresses

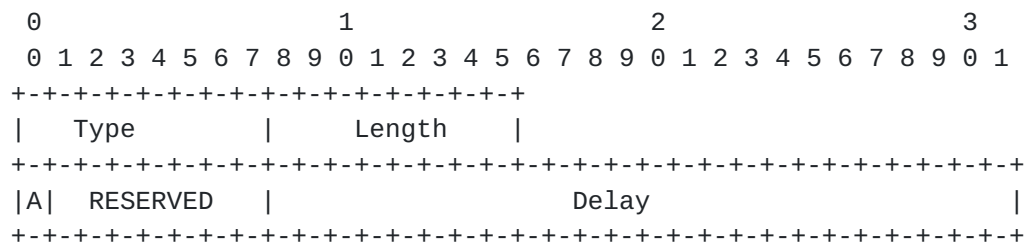
The use of IS-IS TE Metric Extensions sub-TLVs is not confined to the TE context. In other words, IS-IS TE Metric Extensions sub-TLVs defined in this document can also be used for computing paths in the absence of a TE subsystem.

However, as for the TE case, Interface Address and Neighbor Address sub-TLVs (IPv4 or IPv6) MUST be present. The encoding is defined in [RFC5305] for IPv4 and in [RFC6119] for IPv6.

4. Sub TLV Details

4.1. Unidirectional Link Delay Sub-TLV

This sub-TLV advertises the average link delay between two directly connected IS-IS neighbors. The delay advertised by this sub-TLV MUST be the delay from the local neighbor to the remote one (i.e. the forward path latency). The format of this sub-TLV is shown in the following diagram:



where:

Figure 1

Type: TBA (suggested value: 33).

Length: 4.

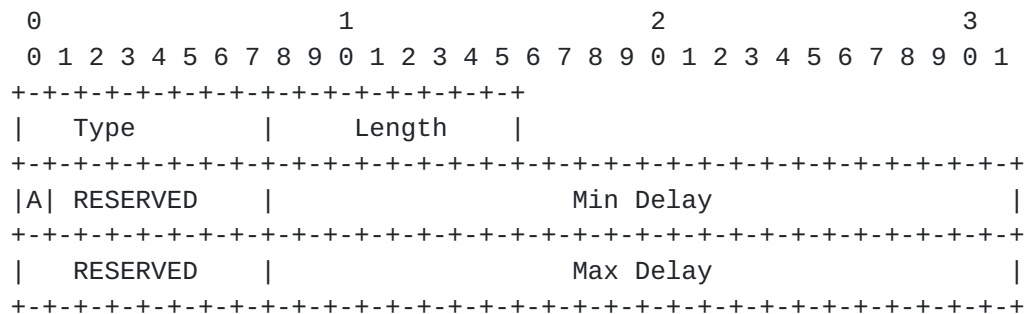
A-bit. The A-bit represents the Anomalous (A) bit. The A-bit is set when the measured value of this parameter exceeds its configured maximum threshold. The A bit is cleared when the measured value falls below its configured reuse threshold. If the A-bit is clear, the sub-TLV represents steady state link performance.

RESERVED. This field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

Delay. This 24-bit field carries the average link delay over a configurable interval in micro-seconds, encoded as an integer value. When set to the maximum value 16,777,215 (16.777215 sec), then the delay is at least that value and may be larger.

4.2. Min/Max Unidirectional Link Delay Sub-TLV

This sub-TLV advertises the minimum and maximum delay values between two directly connected IS-IS neighbors. The delay advertised by this sub-TLV MUST be the delay from the local neighbor to the remote one (i.e. the forward path latency). The format of this sub-TLV is shown in the following diagram:



where:

Figure 2

Type: TBA (suggested value: 34).

Length: 8.

A-bit. The A-bit represents the Anomalous (A) bit. The A-bit is set when the measured value of this parameter exceeds its configured maximum threshold. The A bit is cleared when the measured value falls below its configured reuse threshold. If the A-bit is clear, the sub-TLV represents steady state link performance.

RESERVED. This field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

Min Delay. This 24-bit field carries minimum measured link delay value (in microseconds) over a configurable interval, encoded as an integer value.

Max Delay. This 24-bit field carries the maximum measured link delay value (in microseconds) over a configurable interval, encoded as an integer value.

Implementations MAY also permit the configuration of an offset value (in microseconds) to be added to the measured delay value, to facilitate the communication of operator specific delay constraints.

It is possible for the Min and Max delay to be the same value.

This sub-TLV advertises the loss (as a packet percentage) between two directly connected IS-IS neighbors. The link loss advertised by this sub-TLV MUST be the packet loss from the local neighbor to the remote one (i.e. the forward path loss). The format of this sub-TLV is shown in the following diagram:

This sub-TLV advertises the residual bandwidth between two directly connected IS-IS neighbors. The residual bandwidth advertised by this sub-TLV MUST be the residual bandwidth from the system originating the LSA to its neighbor.

This sub-TLV advertises the available bandwidth between two directly connected IS-IS neighbors. The available bandwidth advertised by this sub-TLV MUST be the available bandwidth from the system originating this sub-TLV. The format of this sub-TLV is shown in the following diagram:

1. If the measured parameter falls outside a configured upper bound for all but the min delay metric (or lower bound for min delay metric only) and the advertised sub-TLV is not already outside that bound or,
2. If the difference between the last advertised value and current measured value exceed a configured threshold then,
3. The advertisement is made immediately.
4. For sub-TLVs which include an A-bit (except min/max delay), an additional threshold SHOULD be included corresponding to the threshold for which the performance is considered anomalous (and sub-TLVs with the A-bit are sent). The A-bit is cleared when the sub-TLV's performance has been below (or re-crosses) this threshold for an advertisement interval(s) to permit fail back.

To prevent oscillations, only the high threshold or the low threshold (but not both) may be used to trigger any given sub-TLV that supports both.

Additionally, once outside of the bounds of the threshold, any readvertisement of a measurement within the bounds would remain governed solely by the measurement interval for that sub-TLV.

6. Announcement Suppression

When link performance values change by small amounts that fall under thresholds that would cause the announcement of a sub-TLV, implementations SHOULD suppress sub-TLV readvertisement and/or lengthen the period within which they are refreshed.

Only the accelerated advertisement threshold mechanism described in [Section 5](#) may shorten the re-advertisement interval. All suppression and re-advertisement interval backoff timer features SHOULD be configurable.

7. Network Stability and Announcement Periodicity

[Section 5](#) and [Section 6](#) provide configurable mechanisms to bound the number of re-advertisements. Instability might occur in very large networks if measurement intervals are set low enough to overwhelm the processing of flooded information at some of the routers in the topology. Therefore care should be taken in setting these values.

Additionally, the default measurement interval for all sub-TLVs SHOULD be 30 seconds.

Announcements MUST also be able to be throttled using configurable inter-update throttle timers. The minimum announcement periodicity is 1 announcement per second. The default value SHOULD be set to 120 seconds.

Implementations SHOULD NOT permit the inter-update timer to be lower than the measurement interval.

Furthermore, it is RECOMMENDED that any underlying performance measurement mechanisms not include any significant buffer delay, any significant buffer induced delay variation, or any significant loss due to buffer overflow or due to active queue management.

8. Enabling and Disabling Sub-TLVs

Implementations MUST make it possible to individually enable or disable each sub-TLV based on configuration.

9. Static Metric Override

Implementations SHOULD permit the static configuration and/or manual override of dynamic measurements for each sub-TLV in order to simplify migration and to mitigate scenarios where dynamic measurements are not possible.

10. Compatibility

As per [[RFC5305](#)], unrecognized sub-TLVs should be silently ignored.

11. Security Considerations

This document does not introduce security issues beyond those discussed in [[RFC5305](#)].

[Section 7](#) describe a mechanism in order to ensure network stability when the new sub-TLVs defined in this document are advertised. Implementation SHOULD follow the described guidelines in order to mitigate the instability risk.

[RFC5304] describes an authentication method for IS-IS LSP that allows cryptographic authentication of IS-IS LSPs.

It is anticipated that in most deployments, IS-IS protocol is used within an infrastructure entirely under control of the same operator. However, it is worth to consider that the effect of sending IS-IS Traffic Engineering sub-TLVs over insecure links could result in a man-in-the-middle attacker delaying real time data to a given site (or destination), which could negatively affect the value of the data

for that site/destination. The use of LSP cryptographic authentication allows to mitigate the risk of man-in-the-middle attack.

12. IANA Considerations

IANA maintains the registry for the sub-TLVs. IS-IS TE Metric Extensions will require one new type code per sub-TLV defined in this document in the following sub-TLV registry: TLVs 22, 23, 141, 222, and 223:

Type	Value

33 (Suggested)	Unidirectional Link Delay
34 (Suggested)	Min/Max Unidirectional Link Delay
35 (Suggested)	Unidirectional Delay Variation
36 (Suggested)	Unidirectional Packet Loss
37 (Suggested)	Unidirectional Residual Bandwidth
38 (Suggested)	Unidirectional Available Bandwidth
39 (Suggested)	Unidirectional Bandwidth Utilization

13. Contributors

The following people gave a substantial contribution to the content of this document and should be considered as co-authors:

Alia Atlas
Juniper Networks
US

akatlas@juniper.net

Clarence Filsfils
Cisco Systems Inc.
Belgium

Email: cfilsfil@cisco.com

14. Acknowledgements

The authors would like to recognize Ayman Soliman, Nabil Bitar, David McDysan, Les Ginsberg, Edward Crabbe, Don Fedyk, Hannes Gredler, Uma Chunduri, Alvaro Retana and Brian Weis for their contribution and review of this document.

The authors also recognize Curtis Villamizar for significant comments and direct content collaboration.

15. References

15.1. 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, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", [RFC 4206](#), DOI 10.17487/RFC4206, October 2005, <<http://www.rfc-editor.org/info/rfc4206>>.
- [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, <<http://www.rfc-editor.org/info/rfc5120>>.
- [RFC5304] Li, T. and R. Atkinson, "IS-IS Cryptographic Authentication", [RFC 5304](#), DOI 10.17487/RFC5304, October 2008, <<http://www.rfc-editor.org/info/rfc5304>>.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", [RFC 5305](#), DOI 10.17487/RFC5305, October 2008, <<http://www.rfc-editor.org/info/rfc5305>>.
- [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, <<http://www.rfc-editor.org/info/rfc5316>>.
- [RFC6119] Harrison, J., Berger, J., and M. Bartlett, "IPv6 Traffic Engineering in IS-IS", [RFC 6119](#), DOI 10.17487/RFC6119, February 2011, <<http://www.rfc-editor.org/info/rfc6119>>.

15.2. Informative References

- [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, <<http://www.rfc-editor.org/info/rfc3209>>.
- [RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", [RFC 4203](#), DOI 10.17487/RFC4203, October 2005, <<http://www.rfc-editor.org/info/rfc4203>>.
- [RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", [RFC 6374](#), DOI 10.17487/RFC6374, September 2011, <<http://www.rfc-editor.org/info/rfc6374>>.
- [RFC6375] Frost, D., Ed. and S. Bryant, Ed., "A Packet Loss and Delay Measurement Profile for MPLS-Based Transport Networks", [RFC 6375](#), DOI 10.17487/RFC6375, September 2011, <<http://www.rfc-editor.org/info/rfc6375>>.
- [RFC7285] Alimi, R., Ed., Penno, R., Ed., Yang, Y., Ed., Kiesel, S., Previdi, S., Roome, W., Shalunov, S., and R. Woundy, "Application-Layer Traffic Optimization (ALTO) Protocol", [RFC 7285](#), DOI 10.17487/RFC7285, September 2014, <<http://www.rfc-editor.org/info/rfc7285>>.

Authors' Addresses

Stefano Previdi (editor)
Cisco Systems, Inc.
Via Del Serafico 200
Rome 00191
IT

Email: sprevidi@cisco.com

Spencer Giacalone
Unaffiliated

Email: spencer.giacalone@gmail.com

Dave Ward
Cisco Systems, Inc.
3700 Cisco Way
SAN JOSE, CA 95134
US

Email: wardd@cisco.com

John Drake
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
USA

Email: jdrake@juniper.net

Qin Wu
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: sunseawq@huawei.com

