

Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2013

U. Chunduri
W. Lu
A. Tian
Ericsson Inc.
N. Shen
Cisco Systems, Inc.
October 22, 2012

IS-IS Extended Sequence number TLV
draft-chunduri-isis-extended-sequence-no-tlv-03

Abstract

This document defines Extended Sequence number TLV to protect Intermediate System to Intermediate System (IS-IS) PDUs from replay attacks.

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 April 25, 2013.

Copyright Notice

Copyright (c) 2012 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
1.1. Requirements Language	3
1.2. Acronyms	4
2. Replay attacks and Impact on IS-IS networks	4
2.1. Impact of Replays	4
3. Extended Sequence Number TLV	5
3.1. Sequence Number Wrap	6
4. Mechanism and Packet Encoding	6
4.1. IIHs	7
4.2. SNPs	7
4.2.1. CSNPs	7
4.2.2. PSNPs	7
4.3. LSPs	8
5. Backward Compatibility and Deployment	8
5.1. IIH and SNPs	8
5.2. LSPs	9
5.3. Operational Consideration	9
6. IANA Considerations	9
7. Security Considerations	9
8. Acknowledgements	10
9. Appendix A	10
9.1. Appendix A.1	10
9.2. Appendix A.2	10
10. Appendix B	11
10.1. Operational/Implementation consideration	11
11. References	11
11.1. Normative References	11
11.2. Informative References	11
Authors' Addresses	12

1. Introduction

With the rapid development of new data center infrastructures, due to its flexibility and scalability attributes, IS-IS has been adopted widely in various L2 and L3 routing deployment of the data centers for critical business operations. At the meantime the SDN-enabled networks even though put more power to Internet applications and also make network management easier, it does raise the security requirement of network routing infrastructure to another level.

This document defines Extended Sequence number (ESN) TLV to protect Intermediate System to Intermediate System (IS-IS) PDUs from replay attacks.

A replayed IS-IS PDU can potentially cause many problems in the IS-IS networks ranging from bouncing adjacencies to black hole or even some form of Denial of Service (DoS) attacks as explained in Section 2. This problem is also discussed in security consideration section, in the context of cryptographic authentication work as described in [RFC5304] and in [RFC5310].

Currently, there is no mechanism to protect IS-IS HELLO PDUs (IIHs) and Sequence number PDUs (SNPs) from the replay attacks. However, Link State PDUs (LSPs) have sequence number in the LSP header as defined in [RFC1195], with which it can effectively mitigate the intra-session replay attacks. But, LSPs are still susceptible to inter-session replay attacks.

The new ESN TLV defined here thwart these threats and can be deployed with authentication mechanism as specified in [RFC5304] and in [RFC5310] for a more secure network.

Replay attacks can be effectively mitigated by deploying a group key management protocol (being developed as defined in [I-D.yeung-g-ikev2] and [I-D.hartman-karp-mrkmpl]) with a frequent key change policy. Currently, there is no such mechanism defined for IS-IS. Even if such a mechanism is defined, usage of this TLV can be helpful to avoid replays before the keys are changed.

Also, it is believed, even when such key management system is deployed, there always will be some manual key based systems that co-exist with KMP (Key Management Protocol) based systems. The ESN TLV defined in this document is more helpful for such deployments.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this

document are to be interpreted as described in RFC 2119 [RFC2119].

1.2. Acronyms

CSNP	- Complete Sequence Number PDU
ESN	- Extended Sequence Number
IIH	- IS-IS Hello PDU
KMP	- Key Management Protocol (auto key management)
LSP	- IS-IS Link State PDU
MKM	- Manual Key management Protocols
PDU	- Protocol Data Unit
PSNP	- Partial Sequence Number PDU
SNP	- Sequence Number PDU

2. Replay attacks and Impact on IS-IS networks

This section explains the replay attacks and the applicability of the same for IS-IS networks. Replaying a captured protocol packet to cause damage is a common threat for any protocol. Securing the packet with cryptographic authentication information alone can not mitigate this threat completely. This has been described in detail in "Replay Attacks" section of KARP IS-IS gap analysis document [I-D.chunduri-karp-is-is-gap-analysis].

2.1. Impact of Replays

At the time of adjacency bring up an IS sends IIH packet with empty neighbor list (TLV 6) and with or without the authentication information as per provisioned authentication mechanism. If this packet is replayed later on the broadcast network all ISes in the broadcast network can bounce the adjacency to create a huge churn in the network.

Today Link State PDUs (LSPs) have intra-session replay protection as LSP header contains 32-bit sequence number which is verified for every received PDU against the local LSP database. But, if the key is not changed, an adversary can cause an inter-session replay attack by replaying a old LSP with higher sequence number and fewer prefixes or fewer adjacencies. This forces the receiver to accept and remove

the routes from the routing table, which eventually causes traffic disruption to those prefixes. The more common pre-conditions for inter-session replay attacks with LSPs and the current in-built recovery mechanism, have been discussed in details in KARP IS-IS gap analysis document [I-D.chunduri-karp-is-is-gap-analysis].

In broadcast networks a replayed Complete Sequence Number PDU (CSNP) can force the receiver to request Partial Sequence Number PDU (PSNP) in the network and similarly, a replayed PSNP can cause unnecessary LSP flood in the network.

Please refer KARP IS-IS gap analysis document for further details.

3. Extended Sequence Number TLV

The Extended Sequence Number (ESN) TLV is composed of 1 octet for the Type, 1 octet that specifies the number of bytes in the Value field and an 8 or 12 byte Value field.

x CODE - TBD.

x LENGTH - total length of the value field, which is 12 bytes for IIH, SNP PDUs and 8 bytes for LSPs.

x Value - 64-bit Extended Session Sequence Number (ESSN), which is present for all IS-IS PDUs followed 32 bit monotonically increasing per Packet Sequence Number (PSN). PSN is not required for LSPs.

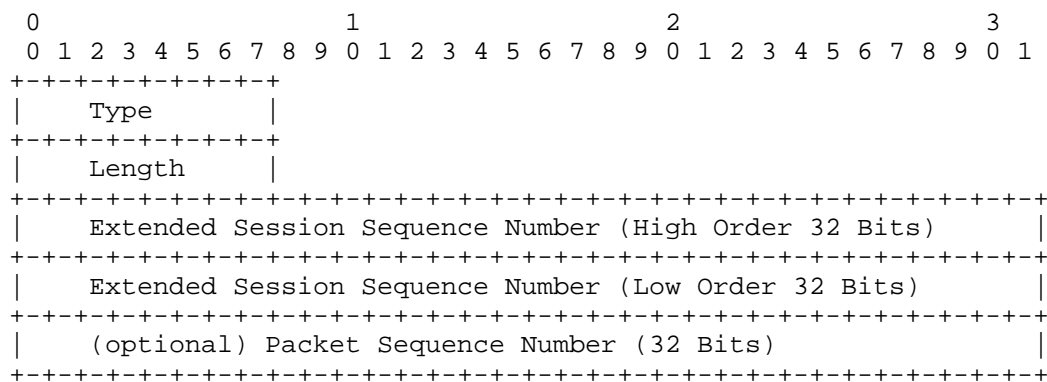


Figure 1: Extended Sequence Number (ESN) TLV

The Extended Sequence Number (ESN) TLV Type is TBD. Please refer to

IANA Considerations, in Section 6 for more details. Length indicates the length of the value field; which is 12 bytes for IIH and SNP PDUs and 8 bytes for LSPs.

In order to provide protection against both inter-session and intra-session replay attacks, the IS-IS Extended Session Sequence Number (ESSN) is defined a 64-bits value; the value MUST contain ever increasing number in all IS-IS PDUs including LSPs whenever it is changed due any situation as specified in Section 3.1.

The 32-bit Packet Sequence Number (PSN) MUST be set and increase monotonically for IIH or SNP PDUs sent by IS-IS node. Upon reception, the Packet Sequence number MUST be greater than the last sequence number in the IIH or SNP PDUs accepted from the sending IS-IS node. Otherwise, the IIH or SNP PDU is considered as replayed PDU and dropped.

As LSPs contain 32-bit sequence number field already in the LSP header, Packet Sequence Number in the ESN TLV MUST be omitted by setting the length field to 8 bytes and implementations continue to refer the header sequence number for all encoding and validation purposes.

The ESN TLV defined here is optional. The ESN TLV MAY present in any IS-IS PDU. If present and authentication is in use this TLV MUST be included as part of the authentication data to calculate the digest. A sender MUST only transmit a single ESN TLV in a IS-IS PDU.

3.1. Sequence Number Wrap

If the 32-bit Packet Sequence Number in ESN TLV and for LSPs the 32-bit header sequence number wraps; or session is refreshed; or even for the cold restarts the 64-bit ESSN value MUST be set higher than the previous value. IS-IS implementations MAY use guidelines provided in Section 9 for accomplishing this.

4. Mechanism and Packet Encoding

The ESN TLV defined in this document is optional and the encoding and decoding of this TLV in each IS-IS PDU is as detailed below. Also refer, when to ignore processing of the ESN TLV as described in Section 5 for appropriate operation in the face of legacy node(s) in the network with out having this capability.

4.1. IIHs

The IIH ESN TLV information is maintained per IS-IS interface and per level. For a broadcast interface, it can have two sets of ESN TLV information, if the circuit belongs to both level-1 and level-2. For point-to-point (P2P) interface, only one ESN TLV information is needed. This TLV information can be maintained as part of the adjacency state.

While transmitting, the 64-bit ESSN MUST always be started with a non zero number and MAY use the guidelines as specified in Section 9 to encode this 64-bit value. The 32-bit PSN starts from 1 and increases monotonically for every subsequent PDU.

While receiving, the 64-bit ESSN MUST always be either same or higher than the stored value in the adjacency state. Similarly, the 32-bit PSN MUST be higher than the stored value in the adjacency state. If the PDU is accepted then the adjacency state should be updated with the last received IIH PDU's ESN TLV information.

For an adjacency refresh or the 32-bit PSN wrap the associated higher order 64-bit ESSN MUST always be higher than the previous value and the lower order 32-bit packet sequence number starts all over again.

4.2. SNPs

4.2.1. CSNPs

In broadcast networks, only Designated Intermediate System (DIS) CSNP ESN TLV information is maintained per adjacency (per level) similar to IIH ESN TLV information. The procedure for encoding, verification and sequence number wrap scenarios are similar as explained in Section 4.1, except separate DIS ESN TLV information should be used. In case of DIS change all adjacencies in the broadcast network MUST reflect new DIS's CSNP ESN TLV information in the adjacency and should be used for encoding/verification.

In P2P networks, CSNP ESN TLV information is maintained per adjacency similar to IIH ESN TLV information. The procedure for encoding, verification and sequence number wrap scenarios are similar as explained in Section 4.1, except separate CSNP ESN TLV information should be used.

4.2.2. PSNPs

In both broadcast and P2P networks, PSNP ESN TLV information is maintained per adjacency (per level) similar to IIH ESN TLV information. The procedure for encoding, verification and sequence

number wrap scenarios are similar as explained in Section 4.1, except separate PSNP ESN TLV information should be used.

4.3. LSPs

For LSPs, while originating, the 64-bit ESSN MUST always be started with a non zero number and MAY use the guidelines as specified in Section 9 for encoding this value.

While receiving, the 64-bit Extended Sequence Number MUST always be either same or higher than the stored value in the LSP database. This document does not specify any changes for the existing LSP header 32-bit sequence number validation mechanism.

5. Backward Compatibility and Deployment

The implementation and deployment of the ESN TLV can be done to support backward compatibility and gradual deployment in the network without requiring a flag day. The deployment can be done for IS-IS links only, or for both IS-IS links and nodes in the networks. This feature can also be deployed for the links in a certain area of the network where the maximum security mechanism is needed, or it can be deployed for the entire network.

The implementation SHOULD allow the configuration of ESN TLV feature on each IS-IS link level and on IS-IS node level with area/level scope. The implementation SHOULD also allow operators to control the configuration of 'send' and/or 'verify' the feature of IS-IS PDUs for the links and for the node. In this document, the 'send' operation is to include the ESN TLV in it's own IS-IS PDUs; and the 'verify' operation is to process the ESN TLV in the receiving IS-IS PDUs from neighbors.

5.1. IIH and SNPs

On the link level, ESN TLV involves the IIH PDUs and SNPs (both CSNP and PSNP). When the router software is upgraded to include this feature, the network operators can configure the IS-IS to 'send' the ESN TLV in it's IIH PDUs and SNPs for those IS-IS interfaces on the IS-IS area or level. When all the routers attached to the link or links have been upgraded with this feature, network operators can start to configure 'verify' on the IS-IS interfaces for all the routers sharing the same link(s). This way deployment can be done in per link basis in the network. Please further refer Section 5.3 for note on operational considerations at the time of 'verify' operation in the network. The operators may decide to only apply ESN TLV feature on some of the links in the network, or only on their multi-

access media links.

5.2. LSPs

On the node level with an area or level scope, ESN TLV involves the IS-IS LSPs. This feature has to be done for the entire IS-IS area or levels with the same flooding domain. The deployment and upgrade of software to support ESN TLV can be gradual and from node to node. When a node is upgraded to support this feature, the operators can configure the node level 'send' in the desired area/level(s) to include the ESN TLV in it's own LSPs. No 'verify' is enabled until all the routers in the entire IS-IS area/level or entire network is upgraded. Then the operators can configure the 'verify' for the IS-IS node level from node to node. Please further refer Section 5.3 for note on operational considerations at the time of 'verify' operation in the network. When all the nodes performs the 'verify' of ESN TLVs, the node level ESN TLV feature is supported fully in the network.

5.3. Operational Consideration

In the face of an adversary doing an active attack, it is possible to have inconsistent data view in the network, if there is a considerable delay in enabling ESN TLV 'verify' operation from first node to the last node in the network. This can happen primarily because, replay PDUs can potentially be accepted by the nodes where 'verify' operation is still not provisioned at the time of the attack. To minimize such a window it is recommended that provisioning of 'verify' SHOULD be done in a timely fashion by the network operators.

6. IANA Considerations

This document requests that IANA allocate from the IS-IS TLV Codepoints Registry a new TLV, referred to as the "Extended Sequence Number" TLV, with the following attributes: IIH = y, LSP = y, SNP = y, Purge = y.

7. Security Considerations

This document describes a mechanism to the replay attack threat as discussed in the Security Considerations section of [RFC5304] and in [RFC5310]. This document does not introduce any new security concerns to IS-IS or any other specifications referenced in this document.

8. Acknowledgements

The authors of this document do not make any claims on the originality of the ideas described. Authors are thankful for the review and the valuable feedback provided by Acee Lindem, Joel Halpern and Les Ginsberg.

9. Appendix A

IS-IS nodes implementing this specification SHOULD use available mechanisms to preserve the 64-bit Extended Session Sequence Number's strictly increasing property, when ever it is changed for the deployed life of the IS-IS node (including cold restarts).

This Appendix provides only guidelines for achieving the same and implementations can resort to any similar method as far as strictly increasing property of the 64-bit ESSN in ESN TLV is maintained.

9.1. Appendix A.1

One mechanism for accomplishing this is by encoding 64-bit ESSN as system time represented in 64-bit unsigned integer value. This MAY be similar to the system timestamp encoding for NTP long format as defined in Appendix A.4 of [RFC5905]. New current time MAY be used when the IS-IS node loses its sequence number state including in Packet Sequence Number wrap scenarios.

Implementations MUST make sure while encoding the 64-bit ESN value with current system time, it should not default to any previous value or some default node time of the system; especially after cold restarts or any other similar events. In general system time must be preserved across cold restarts in order for this mechanism to be feasible. One example of such implementation is to use a battery backed real-time clock (RTC).

9.2. Appendix A.2

One other mechanism for accomplishing this would be similar to the one as specified in [I-D.ietf-ospf-security-extension-manual-keying], to use the 64-bit ESSN as a wrap/boot count stored in non-volatile storage. This value is incremented anytime the IS-IS node loses its sequence number state including in Packet Sequence Number wrap scenarios.

The drawback of this approach per Section 6 of [I-D.ietf-ospf-security-extension-manual-keying], if used is applicable here. The only drawback is, it requires the IS-IS implementation to be able to

save its boot count in non-volatile storage. If the non-volatile storage is ever repaired or upgraded such that the contents are lost, keys MUST be changed to prevent replay attacks.

10. Appendix B

10.1. Operational/Implementation consideration

Since the ESN is maintained per interface, per level and per PDU type, this scheme can be useful for monitoring the health of the IS-IS adjacency. A Packet Sequence Number skip on IIH can be recorded by the neighbors which can be used later to correlate with adjacency state changes over the interface. For instance in a multi-access media, all the neighbors have the skips from the same IIH sender or only one neighbor has the Packet Sequence Number skips can indicate completely different issues on the network.

11. References

11.1. Normative References

- [RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, December 1990.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC5905] Mills, D., Martin, J., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, June 2010.

11.2. Informative References

- [I-D.chunduri-karp-is-is-gap-analysis]
Chunduri, U., Tian, A., and W. Lu, "KARP IS-IS security gap analysis", draft-chunduri-karp-is-is-gap-analysis-01 (work in progress), March 2012.
- [I-D.hartman-karp-mrkmp]
Hartman, S., Zhang, D., and G. Lebovitz, "Multicast Router Key Management Protocol (MaRK)", draft-hartman-karp-mrkmp-05 (work in progress), September 2012.
- [I-D.ietf-karp-threats-reqs]
Lebovitz, G. and M. Bhatia, "Keying and Authentication for

Routing Protocols (KARP) Overview, Threats, and Requirements", draft-ietf-karp-threats-reqs-05 (work in progress), May 2012.

[I-D.ietf-ospf-security-extension-manual-keying]
Bhatia, M., Hartman, S., Zhang, D., and A. Lindem,
"Security Extension for OSPFv2 when using Manual Key
Management",
draft-ietf-ospf-security-extension-manual-keying-02 (work
in progress), April 2012.

[I-D.weis-gdoi-mac-tek]
Weis, B. and S. Rowles, "GDOI Generic Message
Authentication Code Policy", draft-weis-gdoi-mac-tek-03
(work in progress), September 2011.

[RFC5304] Li, T. and R. Atkinson, "IS-IS Cryptographic
Authentication", RFC 5304, October 2008.

[RFC5310] Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R.,
and M. Fanto, "IS-IS Generic Cryptographic
Authentication", RFC 5310, February 2009.

[RFC6518] Lebovitz, G. and M. Bhatia, "Keying and Authentication for
Routing Protocols (KARP) Design Guidelines", RFC 6518,
February 2012.

Authors' Addresses

Uma Chunduri
Ericsson Inc.
300 Holger Way,
San Jose, California 95134
USA

Phone: 408 750-5678
Email: uma.chunduri@ericsson.com

Wenhu Lu
Ericsson Inc.
300 Holger Way,
San Jose, California 95134
USA

Email: wenhu.lu@ericsson.com

Albert Tian
Ericsson Inc.
300 Holger Way,
San Jose, California 95134
USA

Phone: 408 750-5210
Email: albert.tian@ericsson.com

Naiming Shen
Cisco Systems, Inc.
225 West Tasman Drive,
San Jose, California 95134
USA

Email: naiming@cisco.com

Network Working Group
INTERNET-DRAFT
Intended status: Proposed Standard
Expires: April 21, 2013

Donald Eastlake
Yizhou Li
Huawei
October 22, 2012

Interface Addresses TLV
<draft-eastlake-isis-ia-tlv-01.txt>

Abstract

This document specifies an IS-IS (Intermediate System to Intermediate System) TLV that enables the reporting by an Intermediate System of sets of addresses of different types such that all of the addresses in each set designate the same interface (port). For example, an EUI-48 MAC (Extended Unique Identifier 48-bit, Media Access Control) address, IPv4 address, and IPv6 address can be reported as all corresponding to the same interface. Such information could be used, for example, to synthesize responses to or by-pass the need for the Address Resolution Protocol (ARP) or the IPv6 Neighbor Discovery (ND) protocol in some cases.

Status of This Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Distribution of this document is unlimited. Comments should be sent to the TRILL working group mailing list.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/lid-abstracts.html>. The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

Table of Contents

1. Introduction.....	3
1.1 Conventions Used in This Document.....	3
2. The Interface Addresses TLV.....	4
3 IA-TLV sub-TLVs.....	7
3.1 AFN Size sub-TLV.....	7
3.2 Data Label sub-TLV.....	8
3.3 Nickname sub-TLV.....	9
3.4 Fixed Address sub-TLV.....	9
4. Example.....	11
5. IANA Considerations.....	12
6. Security Considerations.....	13
7. Normative References.....	14
8. Informative References.....	14
Change History.....	16
00 to 01.....	16
Acknowledgements.....	17

1. Introduction

This document specifies an IS-IS (Intermediate System to Intermediate System [ISO-10589] [RFC1195]) TLV that enables the reporting by an Intermediate System in its LSP (Link State PDU) of sets of addresses of different types such that all of the addresses in each set designate the same interface (port). For example, an EUI-48 MAC (Extended Unique Identifier 48-bit, Media Access Control [RFC5342]) address, IPv4 address, and IPv6 address can be reported as all three corresponding to the same interface. Such information could be used, for example, to synthesize responses to or by-pass the need for the Address Resolution Protocol (ARP, [RFC826]), Reverse Address Resolution Protocol (RARP, [RFC903]) or the IPv6 Neighbor Discovery (ND [RFC4861]) protocols in some cases.

Although, in some IETF protocols, address field types are represented by EtherType [802] or Hardware Type [RFC5494] only Address Family Number is used in this TLV.

1.1 Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. The Interface Addresses TLV

The Interface Addresses TLV is used to indicate that a set of addresses indicate the same interface. These addresses can be in different address families. For example, it can be used to declare that an interface has a particular IPv4 address, IPv6 address, and EUI-48 MAC address.

The Template Size gives the number of AFNs present below. The set of AFNs present give a template for the type and order of addresses in each Address Set.

```

+---+---+---+---+---+
| Type = TBD | (1 byte)
+---+---+---+---+---+
| Length | (1 byte)
+---+---+---+---+---+---+---+---+---+---+---+---+
| E|RESV | Topology-ID | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+
| Addr Set End | (1 byte)
+---+---+---+---+---+
| Confidence | (1 byte)
+---+---+---+---+---+
| Template Size | (1 byte)
+---+---+---+---+---+---+---+---+---+---+---+---+
| AFN 1 | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+
| AFN 2 | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+
| ...
+---+---+---+---+---+---+---+---+---+---+---+---+
| AFN K | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+...-+
| Address Set 1 (size determined by Template) |
+---+---+---+---+---+---+---+---+---+---+---+---+...-+
| Address Set 2 (size determined by Template) |
+---+---+---+---+---+---+---+---+---+---+---+---+...-+
| ...
+---+---+---+---+---+---+---+---+---+---+---+---+...-+
| Address Set N (size determined by Template) |
+---+---+---+---+---+---+---+---+---+---+---+---+...-+
| optional sub-TLVs ...
+---+---+---+---+---+---+---+---+---+---+---+---+...

```

Figure 1. The Interface Addresses TLV

- o Type: Interface Addresses TLV type, set to TBD[#247 suggested] (IA-TLV).
- o Length: Variable, minimum 7. If length is 6 or less, the TLV MUST

be ignored.

- o E: The E (rEachability) flag is set to one to indicate that the interfaces whose addresses are being given in the TLV are reachable through the Intermediate System that is advertising the TLV.
- o RESV: Reserved. MUST be sent as zero and ignored on receipt.
- o Topology-ID: This field carries a topology ID [RFC5120] or zero if topologies are not in use.
- o Addr Set End: The unsigned offset of the byte, within the TLV value part, of the last byte of the last Address Set. This will be the byte just before the first sub-TLV if any sub-TLVs are present. [RFC5305]
- o Confidence: This 8-bit quantity indicates the confidence level in the addresses being transported. The semantics of the Confidence are further defined by the technology using the IA-TLV.
- o Template Size: A byte containing the unsigned integer number K of AFNs (Address Family Numbers) in the template specifying the structure of each Address Set occurring later in the TLV. The minimum valid value is 1 and there must be room in the TLV for the AFNs. If Template Size is zero or too big, the TLV MUST be ignored.
- o AFN: A two-byte Address Family Number. The number of AFNs present is given in the Template Size field. This sequence specifies the structure of the Address Sets occurring later in the TLV. For example, if Template Size is 2 and the two AFNs present are the AFNs for IPv4 and EUI-48, in that order, then each Address set present will consist of a 4-byte IPv4 address followed by a 6-byte MAC address. If any AFNs are present that are unknown to the receiving IS and the length of the corresponding address is not provided by a sub-TLV as specified below, the receiving IS will be unable to parse the Address Sets and MUST ignore the enclosing TLV.
- o Address Set: Each address set consists of a sequence of Template Size number of addresses of the types given by the AFN sequence template earlier in the TLV in the same order as those AFNs. No alignment, other than to a byte boundary, is guaranteed. The addresses in each Address Set are contiguous with no unused bytes between them and the Address Sets are contiguous with no unused bytes between Address Sets. The Address Sets must fit within the TLV. If the product of the size of an Address Set and the number of Address Sets is so large that this is not true, the TLV is ignored.

- o sub-TLVs: If the Address Sets indicated by Addr Sets End do not completely fill the Length of the TLV, the remaining bytes are parsed as sub-TLVs [RFC5305]. These sub-TLVs are from a new sequence of sub-TLVs. Any such sub-TLVs that are not known to the receiving IS are ignored. Should this not be possible, for example there is only one remaining byte or an apparent sub-TLV extends beyond the end of the TLV, the containing IA-TLV is considered corrupt and is ignored. Several sub-TLV types are specified in Section 3.

This Reachable Interface Addresses TLV occurs only within LSP PDUs and may occur multiple times in the same or different LSP PDUs with the same or different templates.

Different IA-TLVs within the same or different LSP PDUs from the same IS may have different templates. The same AFN may occur more than once in a template and the same address may occur in more than one address set. For example, an EUI-48 MAC address interface might have three IPv6 addresses. This could be represented by an IA-TLV whose template specifically provided for one EUI-48 address and three IPv6 addresses, which might be an efficient format if there were multiple interfaces with that pattern. Alternatively, a template with one EUI-48 and one IPv6 address could be used in an IA-TLV with three address sets each having the same EUI-48 address but different IPv6 addresses, which might be the most efficient format if only one interface had multiple IPv6 addresses and other interfaces had only one IPv6 address.

In order to be able to parse the Address Sets, a receiving IS must know at least the size of the address each AFN in the Template specifies; however, the presence of the Addr Set End field means that the sub-TLVs, if any, can always be located by a receiving IS. An IS can be assumed to know the size of IPv4 and IPv6 addresses (AFNs 1 and 2) and the size of the additional AFNs allocated by the IANA Considerations below. Should an IS wish to include an AFN that some receiving IS in the campus may not know, it SHOULD include an AFN-Size sub-TLV as described below. If an IA-TLV is received with one or more AFNs in its template for which the receiving IS does not know the length and for which an AFN-Size sub-TLV is not present, that IA-TLV will be ignored.

3 IA-TLV sub-TLVs

IA-TLVs may have trailing sub-TLVs [RFC5305] as specified below. These sub-TLVs occur after the Address Sets and the amount of space available for sub-TLVs is determined from the overall IA-TLV length and the value of the Addr Set End byte.

There is no ordering restriction on IA-TLV sub-TLVs. Unless otherwise specified each sub-TLV type can occur zero, one, or many times in an IA-TLV.

3.1 AFN Size sub-TLV

Using this sub-TLV, the originating IS can specify the size of an address type. This is useful under two circumstances:

1. One or more AFNs that are unknown to the receiving IS appears in the template. If an AFN Size sub-TLV is present for each such AFN, the at least the IA-TLV can be parses the Address Sets and make use of any address types present that it does understand.
2. If an AFN occurs in the template that represents a variable length address, this sub-TLV gives its size for all occurrences in that IA-TLV.

```

+---+---+---+---+---+
|subType = AFNsz|           (1 byte)
+---+---+---+---+---+
| Length          |           (1 byte)
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| AFN Size Record(s) |           (3 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Where each AFN Size Record is structured as follows:

```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| AFN |           (2 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| AddrSize |           (1 byte)
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

- o Type: AFN-Size sub-TLV type, set to 1 (AFNsz).
- o Length: 3*n where n is the number of AFN Size Records present. If n is not a multiple of 3, the sub-TLV MUST be ignored.
- o AFN Size Record(s): Zero or more 3-byte records, each giving the size of an address type identified by an AFN,

- o AFN: The AFN whose length is being specified by the AFN Size Record.
- o AdrSize: The length of the address specified by the AFN field.

This sub-TLV may occur multiple times in an enclosing IA-TLV.

The declaration of a size for an AFN address type controls for the occurrences of the AFN in the enclosing IA-TLV and for and subsequent IA-TLVs in the enclosing LSP PDU until the next occurrence in the LSP PDU of an AFN Size sub-TLV for that AFN. Thus, an AFN Size sub-TLV for a fixed size AFN need only be included in the first IA-TLV in a PDU. But one must be included in or before first IA-TLV using an AFN in each PDU where that AFN appears in the template if needed. Otherwise Address Sets will not be parseable by a receiving IS. If multiple AFN Size Records occur for the same AFN in the same AFN Size sub-TLV the last size given controls.

An AFN Size sub-TLV for any AFN known to the receiving IS (which always includes AFN 1 and 2 and the AFNs specified in Section 5) is compared with the size known to the IS and if they differ, the enclosing IA-TLV is ignored.

3.2 Data Label sub-TLV

```

+-----+
|Type==DATA-LABEL|          (1 byte)
+-----+
| Length          |          (1 byte)
+-----+-----+
| Data Label      |          (variable)
+-----+-----+

```

- o Type: Data Label sub-TLV type, set to 2 (DATA-LABEL).
- o Length: variable, minimum 1. If Length is zero, the sub-TLV MUST be ignored.
- o Data Label: A data label under which all the interfaces listed in the TLV can be reached. Exact meaning for different lengths depends on the technology in use. If absent, no label is specified. If more than one different Data Label sub-TLV occurs in an Interface Addresses TLV, it indicates that the interfaces listed can be reach via all the labels given.

For TRILL use, if Length=2, the Data Label is a VLAN-ID which appears right justified in two bytes with four leading bits that MUST be sent as zero and ignored on receipt.

3.3 Nickname sub-TLV

```

+---+---+---+---+---+
|Type=NICKNAME|          (1 byte)
+---+---+---+---+---+
| Length      |          (1 byte)
+---+---+---+---+---+---+---+---+---+...
| Nickname    |          (variable)
+---+---+---+---+---+---+---+---+---+...

```

- o Type: Data Label sub-TLV type, set to 3 (NICKNAME).
- o Length: variable, minimum 1. If Length is zero, the sub-TLV MUST be ignored.
- o Nickname: The nickname of an Intermediate System through which all the interfaces listed in the TLV can be reached. Exact meaning for different lengths depends on the technology in use. If absent, no nickname is specified. If more than one different Nickname sub-TLV occurs in an Interface Addresses TLV, it indicates that the interfaces listed can be reach via all the nicknames given. Occurrence of one or more Nickname sub-TLVs in an Interface Addresses TLV does not change the effect of the E flag bit in the TLV initial fixed fields; the E flag having the value one still indicates that the interfaces are reachable through the Intermediate System advertising the TLV.

3.4 Fixed Address sub-TLV

There may be cases where, in an Interface Addresses TLV, the same address would appear across every address set in the TLV. To avoid having a larger template and wasted space in all Address Sets, this sub-TLV can be used to indicate such a fixed address

```

+---+---+---+---+---+
|Type=FIXEDADR|          (1 byte)
+---+---+---+---+---+
| Length      |          (1 byte)
+---+---+---+---+---+
| AFN         |          (2 bytes)
+---+---+---+---+---+---+---+---+---+...
| Fixed Address|          (variable)
+---+---+---+---+---+---+---+---+---+...

```

- o Type: Data Label sub-TLV type, set to 4 (FIXEDADR).
- o Length: variable, minimum 3. If Length is 2 or less, the sub-TLV MUST be ignored.

- o AFN: Address Family Number of the Fixed Address.
- o Fixed Address: The address of the type indicated by the preceeding AFN field that is considered to be part of every Address Set in the TLV.

4. Example

TBD

5. IANA Considerations

IANA is requested to allocate a new TLV number for IA-TLV [#247 suggested because #147 is the MAC Reachability (MAC-RI) TLV].

IANA is requested to allocate three new AFN numbers as follows:

Number	Description	References
TBD(26)	EUI-48	RFC 5342, this document
TBD(27)	EUI-64	RFC 5342, this document
TBD(28)	IPv6/64	This document.

[[[Curiously enough, the AFN RFC references seem to always use "Address Family Identifier" although IANA uses "Address Family Number". Furthermore, neither of the two RFCs pointed to by the IANA Address Family Number Registry actually has IANA Considerations for Address Family Numbers although the IANA Registry has shows policies for different ranges of AFNs. Conceivably, such IANA Considerations should appear in this document.]]]

IPv6/64 is an 8-byte quantity that is the first 64 bits of an IPv6 address. If present, there will normally be an EUI-64 or EUI-48 address in the address set to provide the lower 64 bits of the IPv6 address. For this purpose, an EUI-48 is expanded to 64 bits as described in [RFC5342].

IANA is requested to establish a new subregistry for sub-TLVs of the IA-TLV with initial contents as shown below.

Name: Sub-TLVs for TLV TBD[#247]
 Procedure: Expert Review
 Reference: This document

Type	Description	Reference
----	-----	-----
0	Reserved	
1	AFN Size	This document
2	Nickname	This document
3	Data Label	This document
4	Fixed Address	This document
5-254	Available	This document
255	Reserved	

[[[Should administrative tag sub-TLVs (see RFC 5130) be allowed?]]]

6. Security Considerations

This document raises no new security issues for IS-IS. IS-IS security may be used to secure IS-IS PDUs containing the IA-TLV. See [RFC5304] and [RFC5310].

7. Normative References

- [ISO-10589] - ISO/IEC 10589:2002, Second Edition, "Intermediate System to Intermediate System Intra-Domain Routing Exchange Protocol for use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service (ISO 8473)", 2002.
- [RFC1195] - Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments", 1990.
- [RFC2119] - Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC5120] - Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, February 2008.
- [RFC5304] - Li, T. and R. Atkinson, "IS-IS Cryptographic Authentication", RFC 5304, October 2008.
- [RFC5305] - Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", 2008.
- [RFC5310] - Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", RFC 5310, February 2009.

8. Informative References

- [802] - IEEE, "Standard for Local and Metropolitan Area Networks: Overview and Architecture", IEEE Std. 802-2001, 8 March 2002.
- [RFC826] - Plummer, D., "Ethernet Address Resolution Protocol: Or Converting Network Protocol Addresses to 48-bit Ethernet Address for Transmission on Ethernet Hardware", STD 37, RFC 826, November 1982.
- [RFC903] - Finlayson, R., Mann, T., Mogul, J., and M. Theimer, "A Reverse Address Resolution Protocol", STD 38, RFC 903, June 1984.
- [RFC4861] - Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.
- [RFC5342] - Eastlake 3rd, D., "IANA Considerations and IETF Protocol Usage for IEEE 802 Parameters", BCP 141, RFC 5342, September

2008.

[RFC5494] - Arkko, J. and C. Pignataro, "IANA Allocation Guidelines
for the Address Resolution Protocol (ARP)", RFC 5494, April
2009.

Change History

RFC Editor Note: Please delete this section before publication.

00 to 01

Add the Fixed Address sub-TLV.

Minor editorial changes.

Acknowledgements

The authors gratefully acknowledge the contributions and review by the following:

Linda Dunbar

This document was produced with raw nroff. All macros used were defined in the source files.

Authors' Addresses

Donald Eastlake
Huawei R&D USA
155 Beaver Street
Milford, MA 01757 USA

Phone: +1-508-333-2270
Email: d3e3e3@gmail.com

Yizhou Li
Huawei Technologies
101 Software Avenue,
Nanjing 210012, China

Phone: +86-25-56624558
Email: liyizhou@huawei.com

Copyright, Disclaimer, and Additional IPR Provisions

Copyright (c) 2012 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. The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions. For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of RFC 5378. No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under RFC 5378, shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

Networking Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 20, 2013

S. Previdi, Ed.
L. Ginsberg
Cisco Systems
M. Shand

A. Roy
D. Ward
Cisco Systems
October 17, 2012

IS-IS Multi-Instance
draft-ietf-isis-mi-08.txt

Abstract

This draft describes a mechanism that allows a single router to share one or more circuits among multiple Intermediate System To Intermediate System (IS-IS) routing protocol instances.

Multiple instances allow the isolation of resources associated with each instance. Routers will form instance specific adjacencies. Each instance can support multiple topologies. Each topology has a unique Link State Database (LSDB). Each Protocol Data Unit (PDU) will contain a new Type Length Value (TLV) identifying the instance and the topology(ies) to which the PDU belongs.

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 <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 April 20, 2013.

Copyright Notice

Copyright (c) 2012 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.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1. Introduction	4
2. Elements Of Procedure	4
2.1. Instance Identifier TLV	5
2.2. Instance Membership	6
2.3. Use of Authentication	7
2.4. Adjacency Establishment	7
2.4.1. Point-to-Point Adjacencies	7
2.4.2. Multi-Access Adjacencies	7
2.5. Update Process Operation	8
2.5.1. Update Process Operation on Point-to-Point Circuits	8
2.5.2. Update Process Operation on Broadcast Circuits	8
2.6. Interoperability Considerations	8
2.6.1. Interoperability Issues on Broadcast Circuits	8
2.6.2. Interoperability using point-to-point circuits	9
3. Usage Guidelines	10
3.1. One-One Mapping Between Topologies and Instances	10
3.2. Many-to-one Mapping Between Topologies and Instances	10
3.3. Considerations for the Number of Instances	11
4. Relationship to M-ISIS	11
5. Graceful Restart Interactions	12
6. IANA Considerations	12
7. Security Considerations	12
8. Acknowledgements	12
9. References	13
9.1. Normative References	13
9.2. Informative References	14
Authors' Addresses	14

1. Introduction

An existing limitation of the protocol defined by [IS-IS] is that only one instance of the protocol can operate on a given circuit. This document defines an extension to IS-IS to remove this restriction. The extension is referred to as "Multi-instance IS-IS" (MI-IS-IS).

Routers which support this extension are referred to as "Multi-instance capable routers" (MI-RTR).

The use of multiple instances enhances the ability to isolate the resources associated with a given instance both within a router and across the network. Instance specific prioritization for processing PDUs and performing routing calculations within a router may be specified. Instance specific flooding parameters may also be defined so as to allow different instances to consume network wide resources at different rates.

Another existing protocol limitation is that a given instance supports a single Update Process operating on a single Link State Database (LSDB). This document defines an extension to IS-IS to allow non-zero instances of the protocol to support multiple Update Processes. Each Update Process is associated with a topology and a unique topology specific LSDB. Non-zero instances of the protocol are only supported by MI-RTRs. Legacy routers support the standard or zero instance of the protocol. The behavior of the standard instance is not changed in any way by the extensions defined in this document.

MI-IS-IS might be used to support topology specific routing. When used for this purpose it is an alternative to [RFC5120].

MI-IS-IS might also be used to support advertisement of information on behalf of applications [I-D.ietf-isis-genapp]. The advertisement of information not directly related to the operation of the IS-IS protocol can therefore be done in a manner which minimizes its impact on the operation of routing.

The above are examples of how MI-IS-IS might be used. The specification of uses of MI-IS-IS is outside the scope of this document.

2. Elements Of Procedure

An Instance Identifier (IID) is introduced to uniquely identify an IS-IS instance. The protocol extension includes a new TLV (IID-TLV)

in each IS-IS PDU originated by an MI-RTR except as noted in this document. The IID-TLV identifies the unique instance as well as the topology/topologies to which the PDU applies. Each IS-IS PDU is associated with only one IS-IS instance.

MI-RTRs form instance specific adjacencies. The IID-TLV included in Intermediate System-Intermediate System Hellos (IIH) includes the IID and the set of Instance Specific Topology Identifiers (ITID) that the sending IS supports. When multiple instances share the same circuit each instance will have a separate set of adjacencies.

MI-RTRs support the exchange of topology specific Link State PDUs for the IID/ITID pairs that each neighbor supports. A unique IS-IS Update process [see IS-IS] operates for each IID/ITID pair. This MAY also imply IID/ITID specific routing calculations and IID/ITID specific routing and forwarding tables. However, this aspect is outside the scope of this specification.

The mechanisms used to implement support of the separation of IS-IS instances and topology specific Update processes within a router are outside the scope of this specification.

2.1. Instance Identifier TLV

A new TLV is defined in order to convey the IID and ITIDs supported. The IID-TLV associates PDUs with each IS-IS instance using a unique 16-bit number. The IID-TLV is carried in all IS-IS PDUs (Intermediate System to Intermediate System Hellos (IIH), Sequence Number PDUs (SNP) and Link State PDUs (LSP)) which are associated with a non-zero instance.

Multiple instances of IS-IS may co-exist on the same circuit and on the same physical router. IIDs MUST be unique within the same routing domain.

Instance identifier #0 is reserved for the standard instance supported by legacy systems. IS-IS PDUs associated with the standard instance MUST NOT include an IID-TLV except where noted in this document.

The IID-TLV MAY include one or more ITIDs. An ITID is a 16 bit identifier where all values (0 - 65535) are valid.

The following format is used for the IID-TLV:

Type: 7
 Length: 2 - 254
 Value:

No. of octets	
IID (0 - 65535)	2
Supported ITID	2
:	:
Supported ITID	2

When the IID = 0, the list of supported ITIDs MUST NOT be present.

An IID-TLV with IID = 0 MUST NOT appear in an SNP or LSP. When the TLV appears (with a non-zero IID) in an SNP or LSP, exactly one ITID MUST be present indicating the topology with which the PDU is associated. If no ITIDs or multiple ITIDs are present or the IID is zero then the PDU MUST be ignored.

When the IID is non-zero and the TLV appears in an IIH, the set of ITIDs supported on the circuit over which the IIH is sent is included. There MUST BE at least one ITID present.

Multiple IID-TLVs MAY appear in IIHs. If multiple IID-TLVs are present and the IID value in all IID-TLVs is not the same then the PDU MUST BE ignored.

A single IID-TLV will support advertisement of up to 126 ITIDs. If multiple IID-TLVs are present in an IIH PDU the supported set of ITIDs is the union of all ITIDs present in all IID-TLVs.

When an LSP purge is initiated, the IID-TLV MUST be retained but the remainder of the body of the LSP SHOULD be removed. Purge procedure is described in [RFC6233] and [RFC6232].

A PDU without an IID-TLV belongs to the standard instance (#0).

2.2. Instance Membership

Each MI-RTR is configured to be participating in one or more instances of IS-IS. For each non-zero instance in which it participates, an MI-RTR marks IS-IS PDUs (IIHs, LSPs or SNPs) generated pertaining to that instance by including the IID-TLV with the appropriate instance identifier.

2.3. Use of Authentication

When authentication is in use, the IID, if present, is first used to select the authentication configuration which is applicable. The authentication check is then performed as normal. When multiple ITIDs are supported, ITID specific authentication MAY be used in SNPs and LSPs.

2.4. Adjacency Establishment

In order to establish adjacencies, IS-IS routers exchange IIH PDUs. Two types of adjacencies exist in IS-IS: point-to-point and broadcast. The following sub-sections describe the additional rules an MI-RTR MUST follow when establishing adjacencies.

2.4.1. Point-to-Point Adjacencies

MI-RTRs include the IID-TLV in the point-to-point hello PDUs they originate. Upon reception of an IIH, an MI-RTR inspects the received IID-TLV and if the IID matches any of the IIDs which the router supports on that circuit, normal adjacency establishment procedures are used to establish an instance specific adjacency. Note that the absence of the IID TLV implies instance ID #0. For instances other than IID #0, an adjacency SHOULD NOT be established unless there is at least one ITID in common.

This extension allows an MI-RTR to establish multiple adjacencies to the same physical neighbor over a point-to-point circuit. However, as the instances are logically independent, the normal expectation of at most one neighbor on a given point-to-point circuit still applies.

2.4.2. Multi-Access Adjacencies

Multi-Access (broadcast) circuits behave differently than point-to-point in that PDUs sent by one router are visible to all routers and all routers must agree on the election of a Designated Intermediate System (DIS) independent of the set of ITIDs supported.

MI-RTRs will establish adjacencies and elect a DIS per IS-IS instance. Each MI-RTR will form adjacencies only with routers which advertise support for the instances which the local router has been configured to support on that circuit. Since an MI-RTR is not required to support all possible instances on a LAN, it's possible to elect a different DIS for different instances.

2.5. Update Process Operation

For non-zero instances, a unique Update Process exists for each supported ITID.

2.5.1. Update Process Operation on Point-to-Point Circuits

On Point-to-Point circuits - including Point-to-Point Operation over LAN [RFC5309] - the ITID specific Update Process only operates on that circuit for those ITIDs which are supported by both ISs operating on the circuit.

2.5.2. Update Process Operation on Broadcast Circuits

On Broadcast circuits, a single DIS is elected for each supported IID independent of the set of ITIDs advertised in LAN IIHs. This requires that the DIS generate pseudo-node LSPs for all supported ITIDs and that the Update Process for all supported ITIDs operate on the Broadcast Circuit. In cases where the set of supported ITIDs for a given non-zero IID is inconsistent among the MI-RTRs operating on a broadcast circuit, connectivity for the topology(ies) associated with ITIDs not supported by some MI-RTRs operating on the circuit can be compromised.

2.6. Interoperability Considerations

[IS-IS] requires that any TLV that is not understood is silently ignored without compromising the processing of the whole IS-IS PDU (IIH, LSP, SNP).

To a router not implementing this extension, all IS-IS PDUs received will appear to be associated with the standard instance regardless of whether an IID TLV is present in those PDUs. This can cause interoperability issues unless the mechanisms and procedures discussed below are followed.

2.6.1. Interoperability Issues on Broadcast Circuits

In order for routers to correctly interoperate with routers not implementing this extension and in order not to cause disruption, a specific and dedicated Media Access Control (MAC) address is used for multicasting IS-IS PDUs with any non-zero IID. Each level will use a specific layer 2 multicast address. Such an address allows MI-RTRs to exchange IS-IS PDUs with non-zero IIDs without these PDUs being processed by legacy routers and therefore no disruption is caused.

An MI-RTR will use the AllL1IS and AllL2IS ISIS MAC layer addresses (as defined in [IS-IS]) as the destination address when sending ISIS

PDUs for the standard instance (IID #0). An MI-RTR will use two new (TBD) dedicated layer 2 multicast addresses (one for each level) as the destination address when sending IS-IS PDUs for any non-zero IID. If operating in point-to-point mode on a broadcast circuit [RFC5309] an MI-RTR MUST use one of the two new multicast addresses as the destination address when sending point-to-point IIHs associated with a non-zero instance. (Either address will do.)

MI-RTRs MUST discard IS-IS PDUs received if either of the following is true:

- o The destination multicast address is AllL1IS or AllL2IS and the PDU contains an IID-TLV
- o The destination multicast address is one of the two new addresses and the PDU contains an IID-TLV with a zero value for the IID or has no IID-TLV.

NOTE: If the multicast addresses AllL1IS and/or AllL2IS are improperly used to send IS-IS PDUs for non-zero IIDs, legacy systems will interpret these PDUs as being associated with IID #0. This will cause inconsistencies in the LSDB in those routers, may incorrectly maintain adjacencies, and may lead to inconsistent DIS election.

2.6.2. Interoperability using point-to-point circuits

In order for an MI-RTR to interoperate over a point-to-point circuit with a router which does NOT support this extension, the MI-RTR MUST NOT send IS-IS PDUs for instances other than IID #0 over the point-to-point circuit as these PDUs may affect the state of IID #0 in the neighbor.

The presence/absence of the IID-TLV in an IIH indicates that the neighbor does/does not support this extension. Therefore, all IIHs sent on a point-to-point circuit by an MI-RTR MUST include an IID-TLV. This includes IIHs associated with IID #0. Once it is determined that the neighbor does not support this extension, an MI-RTR MUST NOT send PDUs (including IIHs) for instances other than IID #0.

Until an IIH is received from a neighbor, an MI-RTR MAY send IIHs for a non-zero instance. However, once an IIH with no IID TLV has been received - indicating that the neighbor is not an MI-RTR - the MI-RTR MUST NOT send IIHs for a non-zero instance. The temporary relaxation of the restriction on sending IIHs for non-zero instances allows a non-zero instance adjacency to be established on an interface on which an MI-RTR does NOT support instance #0.

Point-to-point adjacency setup MUST be done through the use of three-way handshaking procedure as defined in [RFC5303] in order to prevent a non-MI capable neighbor from bringing up an adjacency prematurely based on reception of an IIH w an IID-TLV for a non-zero instance.

3. Usage Guidelines

As discussed above, MI-IS-IS extends IS-IS to support multiple instances on a given circuit. Each instance is uniquely identified by the IID and forms instance specific adjacencies. Each instance supports one or more topologies as represented by the ITIDs. All topologies associated with a given instance share the instance specific adjacencies. The set of topologies supported by a given IID MAY differ from circuit to circuit. Each topology has its own set of LSPs and runs a topology specific Update process. Flooding of topology specific LSPs is only performed on circuits on which both the local router and the neighbor(s) support a given topology (i.e. advertise the same ITID in the set of supported ITIDs sent in the IID-TLV included in IIHs).

The following sub-sections provide some guidelines for usage of instances and topologies within each instance. While this represents examples based on the intent of the authors, implementors are not constrained by the examples.

3.1. One-One Mapping Between Topologies and Instances

When the set of information to be flooded in LSPs is intended to be flooded to all MI-RTRs supporting a given IID a single topology MAY be used. The information contained in the single LSDB MAY still contain information associated with multiple applications as the GENINFO TLV for each application has an application specific ID which identifies the application to which the TLV applies [I-D.ietf-isis-genapp].

3.2. Many-to-one Mapping Between Topologies and Instances

When the set of information to be flooded in LSPs includes subsets which are of interest to a subset of the MI-RTRs supporting a given IID, support of multiple ITIDs allows each subset to be flooded only to those MI-RTRs which are interested in that subset. In the simplest case, a one-one mapping between a given application and an ITID allows the information associated with that application to be flooded only to MI-RTRs which support that application - but a many-to-one mapping between applications and a given ITID is also possible. When the set of application specific information is large, the use of multiple ITIDs provides significantly greater efficiencies

as MI-RTRs only need to maintain the LSDB for applications of interest and that information only needs to be flooded over a topology defined by the MI-RTRs who support a given ITID.

The use of multiple ITIDs also allows the dedication of a full LSP set (256 LSPs at each level) for the use of a given (set of) applications, thereby minimizing the possibility of exceeding the carrying capacity of an LSP set which might arise if information for all applications were to be included in a single LSP set.

Note that the topology associated with each ITID MUST be fully connected in order for ITID specific LSPs to be successfully flooded to all MI-RTRs who support that ITID.

3.3. Considerations for the Number of Instances

The support of multiple topologies within the context of a single instance provides better scalability in support of multiple applications both in terms of the number of adjacencies which are required and in the flooding of topology specific LSDB. In many cases the use of a single non-zero instance would be sufficient and optimal. However, in cases where the set of topologies desired in support of a set of applications is largely disjoint from the set of topologies desired in support of a second set of applications, it could make sense to use multiple instances.

4. Relationship to M-ISIS

[RFC5120] defines support for Multi-Topology Routing. In that document 12 bit Multi-topology IDs are defined to identify the topologies which an IS-IS instance (a "standard instance" as defined by this document) supports. There is no relationship between the Multi-topology IDs defined in [RFC5120] and the ITIDs defined in this document.

If an MI-RTR uses the extensions in support of the BFD Enabled TLV [RFC6213], the ITID SHOULD be used in place of the MTID in which case all 16 bits of the identifier field are useable.

An MI-RTR MAY use the extensions defined in this document to support multiple topologies in the context of an instance with a non-zero IID. Each MI topology is associated with a unique LSDB identified by an ITID. An ITID specific IS-IS Update Process operates on each topology. This differs from [RFC5120] where a single LSDB/single IS-IS Update Process is used in support of all topologies.

An MI-RTR MUST NOT support [RFC5120] multi-topology within a non-zero

instance. The following TLVs MUST NOT be sent in an LSP associated with a non-zero instance and MUST be ignored when received:

TLV 222 - MT IS Neighbors
TLV 235 - MT IP Reachability
TLV 237 - MT IPv6 Reachability

5. Graceful Restart Interactions

[RFC5306] defines protocol extensions in support of graceful restart of a routing instance. The extensions defined there apply to MI-RTRs with the notable addition that as there are topology specific LSP databases each of these must be synchronized following restart in order for database synchronization to be complete. This involves the use of additional T2 timers. See [RFC5306] for further details.

6. IANA Considerations

This document requires the definition of a new ISIS TLV that needs to be reflected in the ISIS TLV code-point registry:

Type	Description	IIH	LSP	SNP	Purge
---	-----	---	---	---	-----
7	Instance Identifier	y	y	y	y

This document requires that two EUI-48 multicast addresses from the IANA managed EUI address space be allocated as specified in [RFC5342]. Requested block size is 2.

The two addresses will be identified as AllL1MI-ISs and AllL2MI-ISs destination addresses.

7. Security Considerations

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

8. Acknowledgements

The authors would like to acknowledge contributions made by Dino Farinacci and Tony Li.

9. References

9.1. Normative References

- [I-D.ietf-isis-genapp]
Ginsberg, L., Previdi, S., and M. Shand, "Advertising Generic Information in IS-IS", draft-ietf-isis-genapp-04 (work in progress), November 2010.
- [IS-IS] "Intermediate system to Intermediate system intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473), ISO/IEC 10589:2002, Second Edition.", Nov 2002.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, February 2008.
- [RFC5303] Katz, D., Saluja, R., and D. Eastlake, "Three-Way Handshake for IS-IS Point-to-Point Adjacencies", RFC 5303, October 2008.
- [RFC5304] Li, T. and R. Atkinson, "IS-IS Cryptographic Authentication", RFC 5304, October 2008.
- [RFC5306] Shand, M. and L. Ginsberg, "Restart Signaling for IS-IS", RFC 5306, October 2008.
- [RFC5310] Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", RFC 5310, February 2009.
- [RFC6213] Hopps, C. and L. Ginsberg, "IS-IS BFD-Enabled TLV", RFC 6213, April 2011.
- [RFC6232] Wei, F., Qin, Y., Li, Z., Li, T., and J. Dong, "Purge Originator Identification TLV for IS-IS", RFC 6232, May 2011.
- [RFC6233] Li, T. and L. Ginsberg, "IS-IS Registry Extension for Purges", RFC 6233, May 2011.

9.2. Informative References

- [RFC5309] Shen, N. and A. Zinin, "Point-to-Point Operation over LAN in Link State Routing Protocols", RFC 5309, October 2008.
- [RFC5342] Eastlake, D., "IANA Considerations and IETF Protocol Usage for IEEE 802 Parameters", BCP 141, RFC 5342, September 2008.

Authors' Addresses

Stefano Previdi (editor)
Cisco Systems
Via Del Serafico 200
Rome 0144
Italy

Email: sprevidi@cisco.com

Les Ginsberg
Cisco Systems
510 McCarthy Blvd.
Milpitas, CA 95035
USA

Email: ginsberg@cisco.com

Mike Shand

Email: imc.shand@gmail.com

Abhay Roy
Cisco Systems
170 W. Tasman Dr.
San Jose, CA 95134
USA

Email: akr@cisco.com

Dave Ward
Cisco Systems
3700 Cisco Way
San Jose, CA 95134
USA

Email: wardd@cisco.com

Networking Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 11, 2013

S. Previdi, Ed.
Cisco Systems, Inc.
S. Giacalone
Thomson Reuters
D. Ward
Cisco Systems, Inc.
J. Drake
A. Atlas
Juniper Networks
C. Filsfils
Cisco Systems, Inc.
October 08, 2012

IS-IS Traffic Engineering (TE) Metric Extensions
draft-previdi-isis-te-metric-extensions-02

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 TE [RFC5305] such that network performance information can be distributed and collected in a scalable fashion. The information distributed using ISIS 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 April 11, 2013.

Copyright Notice

Copyright (c) 2012 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	4
2. TE Metric Extensions to IS-IS	4
3. Interface and Neighbor Addresses	6
4. Sub TLV Details	6
4.1. Unidirectional Link Delay Sub-TLV	6
4.2. Unidirectional Delay Variation Sub-TLV	7
4.3. Unidirectional Link Loss Sub-TLV	8
4.4. Unidirectional Residual Bandwidth Sub-TLV	8
4.5. Unidirectional Available Bandwidth Sub-TLV	9
5. Announcement Thresholds and Filters	10
6. Announcement Suppression	11
7. Network Stability and Announcement Periodicity	11
8. Compatibility	11
9. Security Considerations	11
10. IANA Considerations	11
11. Acknowledgements	11
12. References	12
12.1. Normative References	12
12.2. Informative References	12
Authors' Addresses	13

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 to IS-IS Extended Reachability TLV defined in [RFC5305] (hereafter called "IS-IS TE Metric Extensions"), 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 IS-IS TE Metric Extensions 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 CSPF, or for other uses such as supplementing the data used by an Alto server [I-D.ietf-alto-protocol]. 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.

2. TE Metric Extensions to IS-IS

This document proposes new IS-IS TE sub-TLVs that can be announced in ISIS Extended Reachability TLV (TLV-22) 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-ISN 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
TBD1	Unidirectional Link Delay
TBD2	Unidirectional Delay Variation
TBD3	Unidirectional Packet Loss
TBD4	Unidirectional Residual Bandwidth Sub TLV
TBD5	Unidirectional Available Bandwidth Sub TLV

As can be seen in the list above, the sub-TLVs described in this document carry different types of network performance information. Many (but not all) of the 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 specifications [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

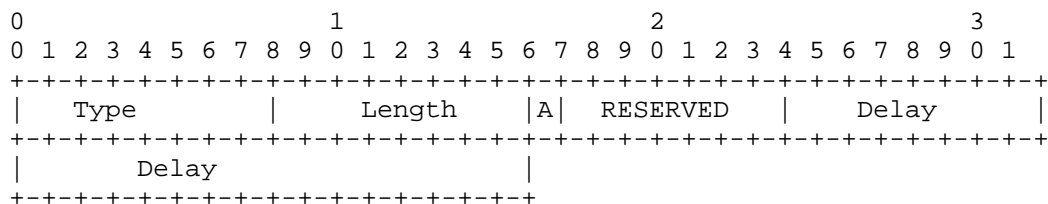
The use of TE Metric Extensions SubTLVs is not confined to the TE context. In other words, IS-IS TE Metric Extensions SubTLVs defined in this document can also be used for computing paths in the absence of a TE subsystem.

However, as for the TE case, Interface Address and Neighbor Address SubTLVs (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:



This sub-TLV has a type of TBD1.
The length is 4.

Where:

"A" represents the Anomalous (A) bit. The A bit is set when the

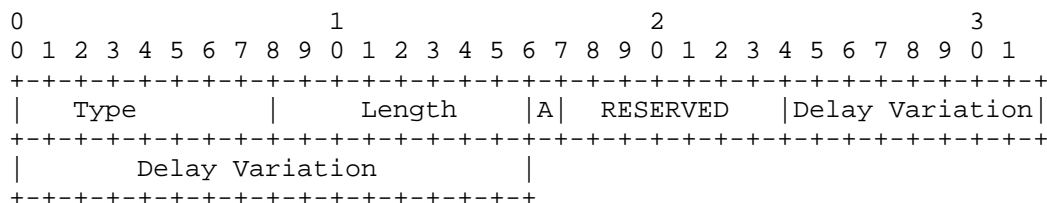
measured value of this parameter exceeds its configured maximum threshold. The A bit is cleared when the measured value falls below its configured reuse threshold. If the A bit is clear, the sub-TLV represents steady state link performance.

The "Reserved" field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

"Delay Value" is a 24-bit field carries the average link delay over a configurable interval in micro-seconds, encoded as an integer value. When set to 0, it has not been measured. 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. Unidirectional Delay Variation Sub-TLV

This sub-TLV advertises the average link delay variation between two directly connected IS-IS neighbors. The delay variation 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:



This sub-TLV has a type of TBD2.

The length is 4.

Where:

"A" 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.

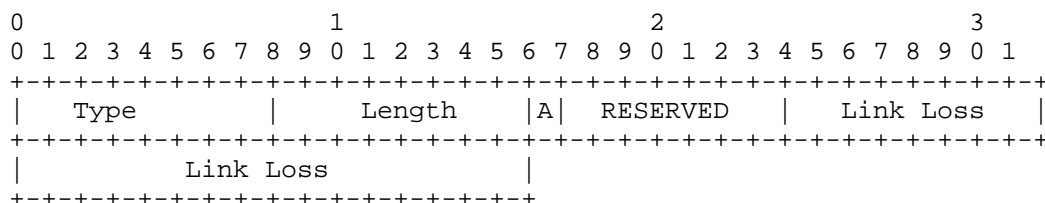
The "Reserved" field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

"Delay Variation" is a 24-bit field carries the average link delay variation over a configurable interval in micro-seconds, encoded as an integer value. When set to 0, it has not been measured. When set to the maximum value 16,777,215 (16.777215 sec), then the delay is at

least that value and may be larger.

4.3. Unidirectional Link Loss Sub-TLV

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 has a type of TBD3.
The length is 4.

Where:

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.

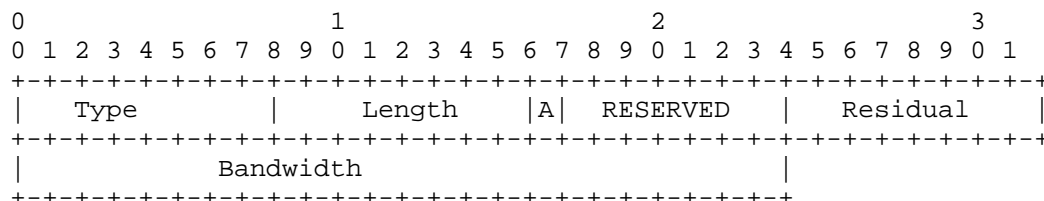
"Reserved" field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

"Link Loss" is a 24-bit field carries link packet loss as a percentage of the total traffic sent over a configurable interval. The basic unit is 0.000003%, where $(2^{24} - 2)$ is 50.331642%. This value is the highest packet loss percentage that can be expressed (the assumption being that precision is more important on high speed links than the ability to advertise loss rates greater than this, and that high speed links with over 50% loss are unusable). Therefore, measured values that are larger than the field maximum SHOULD be encoded as the maximum value. When set to a value of all 1s ($2^{24} - 1$), the link packet loss has not been measured.

4.4. Unidirectional Residual Bandwidth Sub-TLV

This 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 sub-TLV to its neighbor. The format of this sub-TLV is shown in the following diagram:



This sub-TLV has a type of TBD4.
The length is 5.

Where:

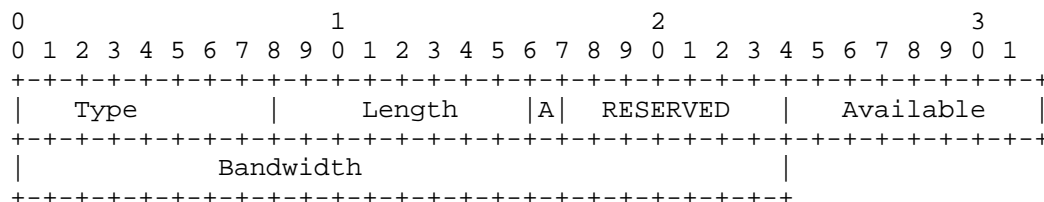
The "A" bit represents the Anomalous (A) bit. The A bit is set when the measured value of this parameter exceeds its configured maximum threshold. The A bit is cleared when the measured value falls below its configured reuse threshold. If the A bit is clear, the sub-TLV represents steady state link performance.

"Residual Bandwidth" is the residual bandwidth in IEEE floating point format in units of bytes per second. The link may be a single link, forwarding adjacency [RFC4206], or bundled link. For a link or forwarding adjacency, residual bandwidth is defined to be Maximum Link Bandwidth [RFC5305] minus the bandwidth currently allocated to RSVP-TE LSPs. For a bundled link, residual bandwidth is defined to be the sum of the component link residual bandwidths.

Note that although it may seem possible to calculate Residual Bandwidth using the existing sub-TLVs in [RFC5305], this is not a consistently reliable approach and hence the Residual Bandwidth sub-TLV has been added here. For example, because the Maximum Reservable Bandwidth [RFC5305] can be larger than the capacity of the link, using it as part of an algorithm to determine the value of the Maximum Link Bandwidth [RFC5305] minus the bandwidth currently allocated to RSVP-TE Label Switched Paths cannot be considered reliably accurate.

4.5. Unidirectional Available Bandwidth Sub-TLV

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



This sub-TLV has a type of TBD5.
The length is 5.

Where:

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.

"Available Bandwidth" is a field that carries the available bandwidth on a link, forwarding adjacency, or bundled link in IEEE floating point format with units of bytes per second. For a link or forwarding adjacency, available bandwidth is defined to be residual bandwidth (see Section 4.4) minus the measured bandwidth used for the actual forwarding of non-RSVP-TE Label Switched Paths packets. For a bundled link, available bandwidth is defined to be the sum of the component link available bandwidths.

5. Announcement Thresholds and Filters

The values advertised in all sub-TLVs MUST be controlled using an exponential filter (i.e. a rolling average) with a configurable measurement interval and filter coefficient.

Implementations are expected to provide separately configurable advertisement thresholds. All thresholds MUST be configurable on a per sub-TLV basis.

The announcement of all sub-TLVs that do not include the A bit SHOULD be controlled by variation thresholds that govern when they are sent.

Sub-TLVs that include the A bit are governed by several thresholds. Firstly, a threshold SHOULD be implemented to govern the announcement of sub-TLVs that advertise a change in performance, but not an SLA violation (i.e. when the A bit is not set). Secondly, implementations MUST provide configurable thresholds that govern the

announcement of sub-TLVs with the A bit set (for the indication of a performance violation). Thirdly, implementations SHOULD provide reuse thresholds. These thresholds govern sub-TLV re-announcement with the A bit cleared to permit fail back.

6. Announcement Suppression

When link performance average values change, but fall under the threshold that would cause the announcement of a sub-TLV with the A bit set, implementations MAY suppress or throttle sub-TLV announcements. All suppression features and thresholds SHOULD be configurable.

7. Network Stability and Announcement Periodicity

To mitigate concerns about stability, all values (except residual bandwidth) MUST be calculated as rolling averages where the averaging period MUST be a configurable period of time, rather than instantaneous measurements.

Announcements MUST also be able to be throttled using configurable inter-update throttle timers. The minimum announcement periodicity is 1 announcement per second.

8. Compatibility

As per [RFC5305], unrecognized Sub-TLVs should be silently ignored

9. Security Considerations

This document does not introduce security issues beyond those discussed in [RFC3630] and [RFC5329].

10. 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.

11. Acknowledgements

The authors would like to recognize Ayman Soliman and Les Ginsberg

for their contributions.

12. References

12.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.
- [RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, February 2008.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.
- [RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, December 2008.
- [RFC5329] Ishiguro, K., Manral, V., Davey, A., and A. Lindem, "Traffic Engineering Extensions to OSPF Version 3", RFC 5329, September 2008.
- [RFC6119] Harrison, J., Berger, J., and M. Bartlett, "IPv6 Traffic Engineering in IS-IS", RFC 6119, February 2011.

12.2. Informative References

- [I-D.ietf-alto-protocol]
Alimi, R., Penno, R., and Y. Yang, "ALTO Protocol",

draft-ietf-alto-protocol-13 (work in progress),
September 2012.

[RFC6375] Frost, D. and S. Bryant, "A Packet Loss and Delay
Measurement Profile for MPLS-Based Transport Networks",
RFC 6375, September 2011.

Authors' Addresses

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

Email: sprevidi@cisco.com

Spencer Giacalone
Thomson Reuters
195 Broadway
New York, NY 10007
USA

Email: Spencer.giacalone@thomsonreuters.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

Alia Atlas
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
USA

Email: akatlas@juniper.net

Clarence Filsfils
Cisco Systems, Inc.
Brussels
Belgium

Email: cfilsfil@cisco.com