

Network Working Group
Internet-Draft
Updates: 5340 (if approved)
Intended status: Standards Track
Expires: August 14, 2015

A. Lindem
Cisco Systems
J. Arkko
Ericsson
February 10, 2015

OSPFv3 Auto-Configuration
draft-ietf-ospf-ospfv3-autoconfig-15.txt

Abstract

OSPFv3 is a candidate for deployments in environments where auto-configuration is a requirement. One such environment is the IPv6 home network where users expect to simply plug in a router and have it automatically use OSPFv3 for intra-domain routing. This document describes the necessary mechanisms for OSPFv3 to be self-configuring. This document updates RFC 5340 by relaxing the HelloInterval/RouterDeadInterval checking during OSPFv3 adjacency formation and adding hysteresis to the update of self-originated Link State Advertisements (LSAs).

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 August 14, 2015.

Copyright Notice

Copyright (c) 2015 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	2
1.1. Requirements notation	3
2. OSPFv3 Default Configuration	3
3. OSPFv3 HelloInterval/RouterDeadInterval Flexibility	4
3.1. Wait Timer Reduction	4
4. OSPFv3 Minimal Authentication Configuration	5
5. OSPFv3 Router ID Selection	5
6. OSPFv3 Adjacency Formation	5
7. OSPFv3 Duplicate Router ID Detection and Resolution	6
7.1. Duplicate Router ID Detection for Neighbors	6
7.2. Duplicate Router ID Detection for Non-Neighbors	6
7.2.1. OSPFv3 Router Auto-Configuration LSA	7
7.2.2. Router-Hardware-Fingerprint TLV	8
7.3. Duplicate Router ID Resolution	9
7.4. Change to RFC 2328 Section 13.4, 'Receiving Self- Originated LSAs'	9
8. Security Considerations	10
9. Management Considerations	10
10. IANA Considerations	11
11. Acknowledgments	11
12. References	13
12.1. Normative References	13
12.2. Informative References	13
Authors' Addresses	14

1. Introduction

OSPFv3 [OSPFV3] is a candidate for deployments in environments where auto-configuration is a requirement. This document describes extensions to OSPFv3 to enable it to operate in these environments. In this mode of operation, the protocol is largely unchanged from the base OSPFv3 protocol specification [OSPFV3]. Since the goals of auto-configuration and security can be conflicting, operators and network administrators should carefully consider their security requirements before deploying the solution described in this document. Refer to Section 8 for more information.

The following aspects of OSPFv3 auto-configuration are described in this document:

1. Default OSPFv3 Configuration
2. HelloInterval/RouterDeadInterval Flexibility
3. Unique OSPFv3 Router ID generation
4. OSPFv3 Adjacency Formation
5. Duplicate OSPFv3 Router ID Resolution
6. Self-Originated LSA Processing

OSPFv3 [OSPFV3] is updated by allowing OSPFv3 adjacencies to be formed between OSPFv3 routers with differing HelloIntervals or RouterDeadIntervals (refer to Section 3). Additionally, hysteresis has been added to the processing of stale self-originated LSAs to mitigate the flooding overhead created by an OSPFv3 Router with a duplicate OSPFv3 Router ID in the OSPFv3 routing domain (refer to Section 7.4. Both updates are fully backward compatible.

1.1. Requirements notation

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-KEYWORDS].

2. OSPFv3 Default Configuration

For complete auto-configuration, OSPFv3 will need to choose suitable configuration defaults. These include:

1. Area 0 Only - All auto-configured OSPFv3 interfaces MUST be in area 0.
2. OSPFv3 SHOULD be auto-configured on all IPv6-capable interface on the router. An interface MAY be excluded if it is clear that running OSPFv3 on the interface is not required. For example, if manual configuration or another condition indicates that an interface is connected to an Internet Service Provider (ISP), there is typically no need to employ OSPFv3. In fact, [IPv6-CPE] specifically requires that IPv6 Customer Premise Equipment (CPE) routers do not initiate any dynamic routing protocol by default on the router's WAN, i.e., ISP-facing, interface. In home networking environments, an interface where no OSPFv3 neighbors are found but a DHCP IPv6 prefix can be acquired may be considered an ISP-facing interface and running OSPFv3 is unnecessary.

3. OSPFv3 interfaces will be auto-configured to an interface type corresponding to their layer-2 capability. For example, Ethernet interfaces and Wi-Fi interfaces will be auto-configured as OSPFv3 broadcast networks and Point-to-Point Protocol (PPP) interfaces will be auto-configured as OSPFv3 Point-to-Point interfaces. Most extant OSPFv3 implementations do this already. Auto-configured operation over wireless networks requiring a point-to-multipoint (P2MP) topology and dynamic metrics based on wireless feedback is not within the scope of this document. However, auto-configuration is not precluded in these environments.
4. OSPFv3 interfaces MAY use an arbitrary HelloInterval and RouterDeadInterval as specified in Section 3. Of course, an identical HelloInterval and RouterDeadInterval will still be required to form an adjacency with an OSPFv3 router not supporting auto-configuration [OSPFV3].
5. All OSPFv3 interfaces SHOULD be auto-configured to use an Interface Instance ID of 0 that corresponds to the base IPv6 unicast address family instance ID as defined in [OSPFV3-AF]. Similarly, if IPv4 unicast addresses are advertised in a separate auto-configured OSPFv3 instance, the base IPv4 unicast address family instance ID value, i.e., 64, SHOULD be auto-configured as the Interface Instance ID for all interfaces corresponding to the IPv4 unicast OSPFv3 instance [OSPFV3-AF].

3. OSPFv3 HelloInterval/RouterDeadInterval Flexibility

Auto-configured OSPFv3 routers will not require an identical HelloInterval and RouterDeadInterval to form adjacencies. Rather, the received HelloInterval will be ignored and the received RouterDeadInterval will be used to determine OSPFv3 liveness with the sending router. In other words, the Neighbor Inactivity Timer (Section 10 of [OSPFV2]) for each neighbor will reflect that neighbor's advertised RouterDeadInterval and MAY be different from other OSPFv3 routers on the link without impacting adjacency formation. A similar mechanism requiring additional signaling is proposed for all OSPFv2 and OSPFv3 routers [ASYNCH-HELLO].

3.1. Wait Timer Reduction

In many situations, auto-configured OSPFv3 routers will be deployed in environments where back-to-back ethernet connections are utilized. When this is the case, an OSPFv3 broadcast interface will not come up until the other OSPFv3 router is connected and the routers will wait RouterDeadInterval seconds before forming an adjacency [OSPFV2]. In order to reduce this delay, an auto-configured OSPFv3 router MAY reduce the wait interval to a value no less than (HelloInterval + 1).

Reducing the setting will slightly increase the likelihood of the Designated Router (DR) flapping but is preferable to the long adjacency formation delay. Note that this value is not included in OSPFv3 Hello packets and does not impact interoperability.

4. OSPFv3 Minimal Authentication Configuration

In many deployments, the requirement for OSPFv3 authentication overrides the goal of complete OSPFv3 autoconfiguration. Therefore, it is RECOMMENDED that OSPFv3 routers supporting this specification minimally offer an option to explicitly configure a single password for HMAC-SHA authentication as described in [OSPFV3-AUTH-TRAILER]. It is RECOMMENDED that the password entered as ASCII hexadecimal digits and that 32 or more digits to facilitate a password with a high degree of entropy. When configured, the password will be used on all auto-configured interfaces with the Security Association Identifier (SA ID) set to 1 and HMAC-SHA-256 used as the authentication algorithm.

5. OSPFv3 Router ID Selection

An OSPFv3 router requires a unique Router ID within the OSPFv3 routing domain for correct protocol operation. Existing Router ID selection algorithms (section C.1 in [OSPFV2] and [OSPFV3]) are not viable since they are dependent on a unique IPv4 interface address which is not likely to be available in autoconfigured deployments. An OSPFv3 router implementing this specification will select a router-id that has a high probability of uniqueness. A pseudo-random number SHOULD be used for the OSPFv3 Router ID. The generation SHOULD be seeded with a variable that is likely to be unique in the applicable OSPFv3 router deployment. A good choice of seed would be some portion or hash of the Router-Hardware-Fingerprint as described in Section 7.2.2.

Since there is a possibility of a Router ID collision, duplicate Router ID detection and resolution are required as described in Section 7 and Section 7.3. OSPFv3 routers SHOULD maintain the last successfully chosen Router ID in non-volatile storage to avoid collisions subsequent to when an autoconfigured OSPFv3 router is first added to the OSPFv3 routing domain.

6. OSPFv3 Adjacency Formation

Since OSPFv3 uses IPv6 link-local addresses for all protocol messages other than messages sent on virtual links (which are not applicable to auto-configuration), OSPFv3 adjacency formation can proceed as soon as a Router ID has been selected and the IPv6 link-local address has completed Duplicate Address Detection (DAD) as specified in IPv6

Stateless Address Autoconfiguration [SLAAC]. Otherwise, the only changes to the OSPFv3 base specification are supporting HelloInterval/RouterDeadInterval flexibility as described in Section 3 and duplicate Router ID detection and resolution as described in Section 7 and Section 7.3.

7. OSPFv3 Duplicate Router ID Detection and Resolution

There are two cases of duplicate OSPFv3 Router ID detection. One where the OSPFv3 router with the duplicate Router ID is directly connected and one where it is not. In both cases, the duplicate resolution is for one of the routers to select a new OSPFv3 Router ID.

7.1. Duplicate Router ID Detection for Neighbors

In this case, a duplicate Router ID is detected if any valid OSPFv3 packet is received with the same OSPFv3 Router ID but a different IPv6 link-local source address. Once this occurs, the OSPFv3 router with the numerically smaller IPv6 link-local address will need to select a new Router ID as described in Section 7.3. Note that the fact that the OSPFv3 router is a neighbor on a non-virtual interface implies that the router is directly connected. An OSPFv3 router implementing this specification should assure that the inadvertent connection of multiple router interfaces to the same physical link is not misconstrued as detection of an OSPFv3 neighbor with a duplicate Router ID.

7.2. Duplicate Router ID Detection for Non-Neighbors

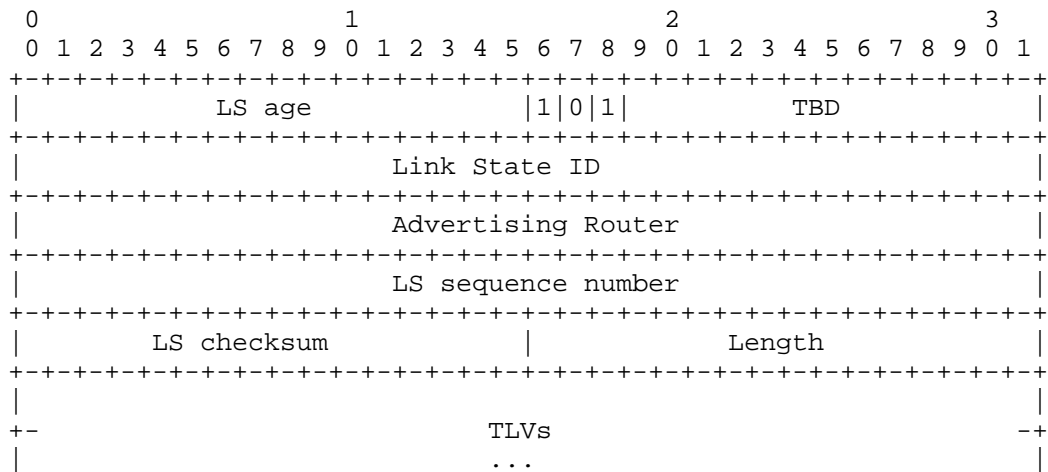
OSPFv3 routers implementing auto-configuration, as specified herein, MUST originate an Auto-Configuration (AC) Link State Advertisement (LSA) including the Router-Hardware-Fingerprint Type-Length-Value (TLV). The Router-Hardware-Fingerprint TLV contains a variable length value that has a very high probability of uniquely identifying the advertising OSPFv3 router. An OSPFv3 router implementing this specification MUST detect received Auto-Configuration LSAs with its Router ID specified in the LSA header. LSAs received with the local OSPFv3 Router's Router ID in the LSA header are perceived as self-originated (see section 4.6 of [OSPFV3]). In these received Auto-Configuration LSAs, the Router-Hardware-Fingerprint TLV is compared against the OSPFv3 Router's own router hardware fingerprint. If the fingerprints are not equal, there is a duplicate Router ID conflict and the OSPFv3 router with the numerically smaller router hardware fingerprint MUST select a new Router ID as described in Section 7.3.

This new LSA is designated for information related to OSPFv3 Auto-configuration and, in the future, could be used for other auto-

configuration information, e.g., global IPv6 prefixes. However, this is beyond the scope of this document.

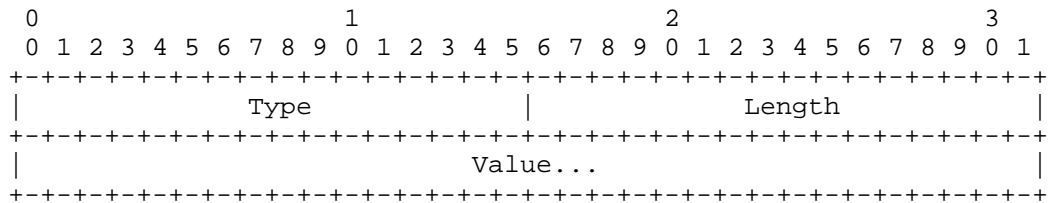
7.2.1. OSPFv3 Router Auto-Configuration LSA

The OSPFv3 Auto-Configuration (AC) LSA has a function code of TBD and the S2/S1 bits set to 01 indicating Area Flooding Scope. The U bit will be set indicating that the OSPFv3 AC LSA should be flooded even if it is not understood. The Link State ID (LSID) value will be a integer index used to discriminate between multiple AC LSAs originated by the same OSPFv3 router. This specification only describes the contents of an AC LSA with a Link State ID (LSID) of 0.



OSPFv3 Auto-Configuration (AC) LSA

The format of the TLVs within the body of an AC LSA is the same as the format used by the Traffic Engineering Extensions to OSPF [TE]. The LSA payload consists of one or more nested Type/Length/Value (TLV) triplets. The format of each TLV is:



TLV Format

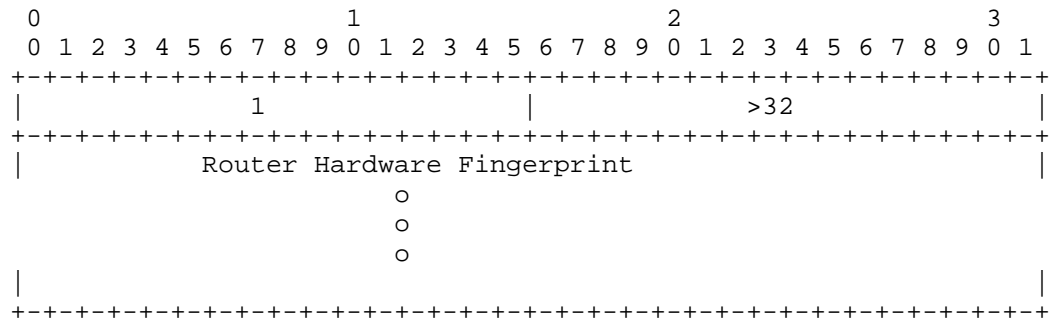
The Length field defines the length of the value portion in octets (thus a TLV with no value portion would have a length of 0). The TLV is padded to 4-octet alignment; padding is not included in the length field (so a 3-octet value would have a length of 3, but the total size of the TLV would be 8 octets). Nested TLVs are also 32-bit aligned. For example, a 1-byte value would have the length field set to 1, and 3 octets of padding would be added to the end of the value portion of the TLV. Unrecognized types are ignored.

The new LSA is designated for information related to OSPFv3 Auto-configuration and, in the future, can be used other auto-configuration information.

7.2.2. Router-Hardware-Fingerprint TLV

The Router-Hardware-Fingerprint TLV is the first TLV defined for the OSPFv3 Auto-Configuration (AC) LSA. It will have type 1 and MUST be advertised in the LSID OSPFv3 AC LSA with an LSID of 0. It SHOULD occur, at most, once and the first instance of the TLV will take precedence over subsequent TLV instances. The length of the Router-Hardware-Fingerprint is variable but must be 32 octets or greater. If the Router-Hardware-Fingerprint TLV is not present as the first TLV, the AC-LSA is considered malformed and is ignored for the purposes of duplicate Router ID detection. Additionally, the event SHOULD be logged.

The contents of the hardware fingerprint MUST have an extremely high probability of uniqueness. It SHOULD be constructed from the concatenation of a number of local values that themselves have a high likelihood of uniqueness, such as MAC addresses, CPU ID, or serial numbers. It is RECOMMENDED that one or more available universal tokens (e.g., IEEE 802 48-bit MAC addresses or IEEE EUI-64 Identifiers [EUI64]) associated with the OSPFv3 router be included in the hardware fingerprint. It MUST be based on hardware attributes that will not change across hard and soft restarts.



Router-Hardware-Fingerprint TLV Format

7.3. Duplicate Router ID Resolution

The OSPFv3 router selected to resolve the duplicate OSPFv3 Router ID condition must select a new OSPFv3 Router ID. The OSPFv3 router SHOULD reduce the possibility of a subsequent Router ID collision by checking the Link State Database for an OSPFv3 Auto-Configuration LSA with the newly selected Router ID and a different Router-Hardware-Fingerprint. If one is detected, a new Router ID should be selected without going through the resolution process Section 7. After selecting a new Router ID, all self-originated LSAs MUST be reoriginated, and any OSPFv3 neighbor adjacencies MUST be reestablished. The OSPFv3 router retaining the Router ID causing the conflict will reoriginate or purge stale any LSAs as described in Section 13.4 [OSPFV2].

7.4. Change to RFC 2328 Section 13.4, 'Receiving Self-Originated LSAs'

RFC 2328 [OSPFV2], Section 13.4, describes the processing of received self-originated LSAs. If the received LSA doesn't exist, the receiving router will purge it from the OSPF routing domain. If the LSA is newer than the version in the Link State Database (LSDB), the receiving router will originate a newer version by advancing the LSA sequence number and reoriginating. Since it is possible for an auto-configured OSPFv3 router to choose a duplicate OSPFv3 Router ID, OSPFv3 routers implementing this specification should detect when multiple instances of the same self-originated LSA are purged or reoriginated since this is indicative of an OSPFv3 router with a duplicate Router ID in the OSPFv3 routing domain. When this condition is detected, the OSPFv3 router SHOULD delay self-originated LSA processing for LSAs that have recently been purged or reoriginated. This specification recommends 10 seconds as the interval defining recent self-originated LSA processing and an exponential back off of 1 to 8 seconds for the processing delay.

This additional delay should allow for the mechanisms described in Section 7 to resolve the duplicate OSPFv3 Router ID conflict.

Since this mechanism is useful in mitigating the flooding overhead associated with the inadvertent or malicious introduction of an OSPFv3 router with a duplicate Router ID into an OSPFv3 routing domain, it MAY be deployed outside of autoconfigured deployments. The detection of a self-originated LSA that is being repeated reoriginated or purged SHOULD be logged.

8. Security Considerations

A unique OSPFv3 Interface Instance ID is used for auto-configuration to prevent inadvertent OSPFv3 adjacency formation, see Section 2

The goals of security and complete OSPFv3 auto-configuration are somewhat contradictory. When no explicit security configuration takes place, auto-configuration implies that additional devices placed in the network are automatically adopted as a part of the network. However, auto-configuration can also be combined with password configuration (see Section 4) or future extensions for automatic pairing between devices. These mechanisms can help provide an automatically configured, securely routed network.

In deployments where different authentication algorithm, per-interface keys, or encryption is required, OSPFv3 IPsec [OSPFV3-IPSEC] or alternate OSPFv3 Authentication trailer [OSPFV3-AUTH-TRAILER] algorithms MAY be used at the expense of additional configuration. The configuration and operational description of such deployments is beyond the scope of this document. However, a deployment could always revert to explicit configuration as described in Section 9 for features such as IPsec, per-interface keys, or alternate authentication algorithms.

The introduction, either malicious or accidental, of an OSPFv3 router with a duplicate Router ID is an attack point for OSPFv3 routing domains. This is due to the fact that OSPFv3 routers will interpret LSAs advertised by the router with the same Router ID as self-originated LSAs and attempt to purge them from the routing domain. The mechanisms in Section 7.4 will mitigate the effects of duplication.

9. Management Considerations

It is RECOMMENDED that OSPFv3 routers supporting this specification also support explicit configuration of OSPFv3 parameters as specified in Appendix C of [OSPFV3]. This would allow explicit override of autoconfigured parameters in situations where it is required (e.g.,

if the deployment requires multiple OSPFv3 areas). This is in addition to the authentication key configuration recommended in Section 4. Additionally, it is RECOMMENDED that OSPFv3 routers supporting this specification allow autoconfiguration to be completely disabled.

Since there is a small possibility of OSPFv3 Router ID collisions, manual configuration of OSPFv3 Router IDs is RECOMMENDED in OSPFv3 routing domains where route convergence due to a router ID change is intolerable.

OSPFv3 Routers supporting this specification MUST augment mechanisms for displaying or otherwise conveying OSPFv3 operational state to indicate whether or not the OSPFv3 router was autoconfigured and whether or not its OSPFv3 interfaces have been auto-configured.

10. IANA Considerations

This specification defines an OSPFv3 LSA Type for the OSPFv3 Auto-Configuration (AC) LSA, as described in Section 7.2.1. The value TBD will be allocated from the existing "OSPFv3 LSA Function Code" registry for the OSPFv3 Auto-Configuration LSA.

This specification also creates a registry for OSPFv3 Auto-Configuration (AC) LSA TLVs. This registry should be placed in the existing OSPFv3 IANA registry, and new values can be allocated via IETF Review or, under exceptional circumstances, IESG Approval. [IANA-GUIDELINES]

Three initial values are allocated:

- o 0 is marked as reserved.
- o 1 is Router-Hardware-Fingerprint TLV (Section 7.2.2).
- o 65535 is an Auto-configuration-Experiment-TLV, a common value that can be used for experimental purposes.

11. Acknowledgments

This specification was inspired by the work presented in the Homenet working group meeting in October 2011 in Philadelphia, Pennsylvania. In particular, we would like to thank Fred Baker, Lorenzo Colitti, Ole Troan, Mark Townsley, and Michael Richardson.

Arthur Dimitrelis and Aidan Williams did prior work in OSPFv3 auto-configuration in the expired "Autoconfiguration of routers using a

link state routing protocol" IETF Draft. There are many similarities between the concepts and techniques in this document.

Thanks for Abhay Roy and Manav Bhatia for comments regarding duplicate router-id processing.

Thanks for Alvaro Retana and Michael Barnes for comments regarding OSPFv3 Instance ID auto-configuration.

Thanks to Faraz Shamim for review and comments.

Thanks to Mark Smith for the requirement to reduce the adjacency formation delay in the back-to-back ethernet topologies that are prevalent in home networks.

Thanks to Les Ginsberg for document review and recommendations on OSPFv3 hardware fingerprint content.

Thanks to Curtis Villamizar for document review and analysis of duplicate router-id resolution nuances.

Thanks to Uma Chunduri for comments during OSPF WG last call.

Thanks to Martin Vigoureux for Routing Area Directorate review and comments.

Thanks to Adam Montville for Security Area Directorate review and comments.

Thanks to Qin Wu for Operations & Management Area Directorate review and comments.

Thanks to Robert Sparks for General Area (GEN-ART) review and comments.

Thanks to Rama Darbha for review and comments.

Special thanks to Adrian Farrel for his in-depth review, copious comments, and suggested text.

Special thanks go to Markus Stenberg for his implementation of this specification in Bird.

Special thanks also go to David Lamparter for his implementation of this specification in Quagga.

The RFC text was produced using Marshall Rose's xml2rfc tool.

12. References

12.1. Normative References

- [OSPFV2] Moy, J., "OSPF Version 2", RFC 2328, April 1998.
- [OSPFV3] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.
- [OSPFV3-AF] Lindem, A., Mirtorabi, S., Roy, A., Barnes, M., and R. Aggarwal, "Support of Address Families in OSPFv3", RFC 5838, April 2010.
- [OSPFV3-AUTH-TRAILER] Bhatia, M., Manral, V., and A. Lindem, "Supporting Authentication Trailer for OSPFv3", RFC 7166, February 2012.
- [RFC-KEYWORDS] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.
- [SLAAC] Thomson, S., Narten, T., and J. Tatuya, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.
- [TE] Katz, D., Yeung, D., and K. Kompella, "Traffic Engineering Extensions to OSPF", RFC 3630, September 2003.

12.2. Informative References

- [ASYNCH-HELLO] Anand, M., Grover, H., and A. Roy, "Asymmetric OSPF Hold Timer", draft-madhukar-ospf-agr-asymmetric-01.txt (work in progress), June 2013.
- [EUI64] IEEE, "Guidelines for 64-bit Global Identifier (EUI-64) Registration Authority", IEEE Tutorial <http://standards.ieee.org/regauth/oui/tutorials/EUI64.html>, March 1997.
- [IANA-GUIDELINES] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 5226, May 2008.

[IPv6-CPE]

Singh, H., Beebee, W., Donley, C., and B. Stark, "Basic Requirements for IPv6 Customer Edge Routers", RFC 7084, November 2013.

[OSPFV3-IPSEC]

Gupta, M. and S. Melam, "Authentication/Confidentiality for OSPFv3", RFC 4552, June 2006.

Authors' Addresses

Acee Lindem
Cisco Systems
301 Midenhall Way
Cary, NC 27513
USA

Email: acee@cisco.com

Jari Arkko
Ericsson
Jorvas, 02420
Finland

Email: jari.arkko@piuha.net

Network Working Group
Internet Draft
Intended status: Proposed Standard
Expires: July 2015

S. Giacalone
Unaffiliated

D. Ward
Cisco Systems

J. Drake
Juniper Networks

A. Atlas
Juniper Networks

S. Previdi
Cisco Systems

January 09, 2015

OSPF Traffic Engineering (TE) Metric Extensions
draft-ietf-ospf-te-metric-extensions-11.txt

Abstract

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

This document describes common extensions to RFC 3630 "Traffic Engineering (TE) Extensions to OSPF Version 2" and RFC 5329 "Traffic Engineering Extensions to OSPF Version 3" to enable network performance information to be distributed in a scalable fashion. The information distributed using OSPF TE Metric Extensions can then be used to make path selection decisions based on network performance.

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

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), 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/ietf/lid-abstracts.txt>

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

This Internet-Draft will expire on July 9, 2015.

Copyright Notice

Copyright (c) 2015 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. Conventions used in this document.....	5
3. TE Metric Extensions to OSPF TE.....	5
4. Sub-TLV Details.....	7
4.1. Unidirectional Link Delay Sub-TLV.....	7
4.1.1. Type.....	7
4.1.2. Length.....	7

4.1.3. A bit.....	7
4.1.4. Reserved.....	7
4.1.5. Delay Value.....	7
4.2. Min/Max Unidirectional Link Delay Sub-TLV.....	8
4.2.1. Type.....	8
4.2.2. Length.....	8
4.2.3. A bit.....	8
4.2.4. Reserved.....	8
4.2.5. Min Delay.....	9
4.2.6. Reserved.....	9
4.2.7 Max Delay.....	9
4.3. Unidirectional Delay Variation Sub-TLV.....	9
4.3.1. Type.....	10
4.3.2. Length.....	10
4.3.3. Reserved.....	10
4.3.4. Delay Variation.....	10
4.4. Unidirectional Link Loss Sub-TLV.....	10
4.4.1. Type.....	11
4.4.2. Length.....	11
4.4.3. A bit.....	11
4.4.4. Reserved.....	11
4.4.5. Link Loss.....	11
4.5. Unidirectional Residual Bandwidth Sub-TLV.....	11
4.5.1. Type.....	12
4.5.2. Length.....	12
4.5.3. Residual Bandwidth.....	12
4.6. Unidirectional Available Bandwidth Sub-TLV.....	12
4.6.1. Type.....	13
4.6.2. Length.....	13
4.6.3. Available Bandwidth.....	13
4.7. Unidirectional Utilized Bandwidth Sub-TLV.....	13
4.7.1. Type.....	14
4.7.2. Length.....	14
4.7.3. Utilized Bandwidth.....	14
5. Announcement Thresholds and Filters.....	14
6. Announcement Suppression.....	15
7. Network Stability and Announcement Periodicity.....	15
8. Enabling and Disabling Sub-TLVs.....	16
9. Static Metric Override.....	16
10. Compatibility.....	16
11. Security Considerations.....	16
12. IANA Considerations.....	17
13. References.....	17
13.1. Normative References.....	17
13.2. Informative References.....	18
14. Acknowledgments.....	19
15. Author's Addresses.....	19

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., link propagation delay) is becoming as critical to data path selection as other metrics.

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 network performance information (such as link propagation delay) in a cost-effective and scalable way.

This document describes extensions to OSPFv2 and OSPFv3 TE (hereafter called "OSPF TE Metric Extensions"), that can be used to distribute network performance information (viz link propagation delay, delay variation, link loss, residual bandwidth, available bandwidth, and utilized bandwidth).

The data distributed by OSPF TE Metric Extensions is meant to be used as part of the operation of the routing protocol (e.g., by replacing cost with link propagation delay or considering bandwidth as well as cost), by enhancing CSPF, or for use by a PCE [RFC4655] or an Alto server [RFC7285]. With respect to CSPF, the data distributed by OSPF 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 distribute network performance information. The methods for measuring that information or acting on it once it is distributed are outside the scope of this document. A method for measuring loss and delay in an MPLS network is described in [RFC6374].

While this document does not specify the method for measuring network performance information, any measurement of link propagation delay SHOULD NOT vary significantly based upon the offered traffic load and hence SHOULD NOT include queuing delays. For a forwarding adjacency (FA) [RFC4206], care must be taken that measurement of the link propagation delay avoids significant queuing delay; this can be accomplished in a variety of ways, e.g., measuring with a traffic class that experiences minimal queuing or summing the measured link propagation delay of the links on the FA's path.

2. Conventions used in this document

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

3. TE Metric Extensions to OSPF TE

This document defines new OSPF TE sub-TLVs that are used to distribute network performance information. The extensions in this document build on the ones provided in OSPFv2 TE [RFC3630] and OSPFv3 TE [RFC5329].

OSPFv2 TE LSAs [RFC3630] are opaque LSAs [RFC5250] with area flooding scope while OSPFv3 Intra-Area-TE-LSAs have their own LSA type, also with area flooding scope; both consist of a single TLV with one or more nested sub-TLVs. The Link TLV is common to both and describes the characteristics of a link between OSPF neighbors.

This document defines several additional sub-TLVs for the Link TLV:

Type	Length	Value
TBD1	4	Unidirectional Link Delay
TBD2	8	Min/Max Unidirectional Link Delay
TBD3	4	Unidirectional Delay Variation
TBD4	4	Unidirectional Link Loss
TBD5	4	Unidirectional Residual Bandwidth
TBD6	4	Unidirectional Available Bandwidth
TBD7	4	Unidirectional Utilized Bandwidth

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 move 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) The node determines 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 failback) it wishes (including nothing).

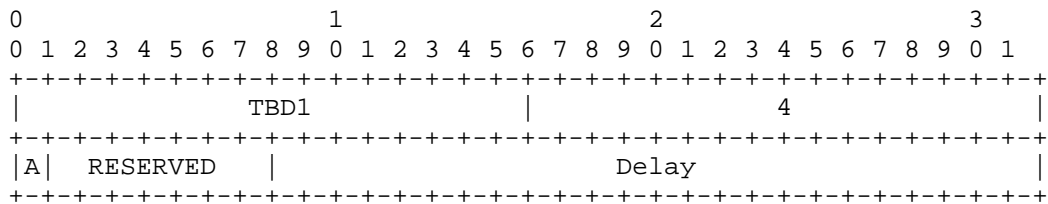
The A bit was intentionally omitted from some sub-TLVs to help mitigate oscillations. See section 7. 1. for more information.

Link delay, delay variation, and link loss MUST be encoded as integers. Consistent with existing OSPF TE specifications [RFC3630], residual, available, and utilized bandwidth MUST be encoded in IEEE single precision floating point [IEEE754]. Link delay and delay variation MUST be in units of microseconds, link loss MUST be a percentage, and bandwidth MUST be in units of 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.

4. Sub-TLV Details

4.1. Unidirectional Link Delay Sub-TLV

This sub-TLV advertises the average link delay between two directly connected OSPF neighbors. The delay advertised by this sub-TLV **MUST** be the delay from the advertising node to its neighbor (i.e., the forward path delay). The format of this sub-TLV is shown in the following diagram:



4.1.1. Type

This sub-TLV has a type of TBD1.

4.1.2. Length

The length is 4.

4.1.3. A bit

This field 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.

4.1.4. Reserved

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

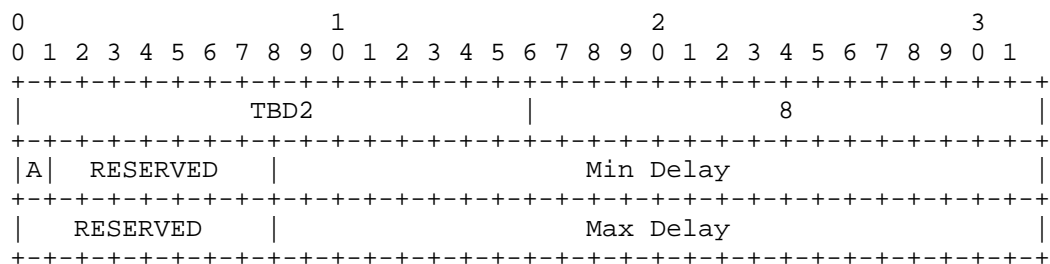
4.1.5. Delay Value

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. If there is no value to send

(unmeasured and not statically specified), then the sub-TLV should not be sent or be withdrawn.

4.2. Min/Max Unidirectional Link Delay Sub-TLV

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



4.2.1. Type

This sub-TLV has a type of TBD2.

4.2.2. Length

The length is 8.

4.2.3. A bit

This field represents the Anomalous (A) bit. The A bit is set when one or more measured values exceed a 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.

4.2.4. Reserved

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


```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

4.3.1. Type

This sub-TLV has a type of TBD3.

4.3.2. Length

The length is 4.

4.3.3. Reserved

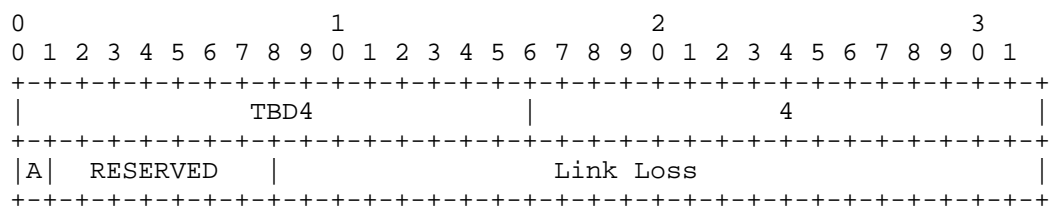
This field is reserved for future use. It MUST be set to 0 when sent and MUST be ignored when received.

4.3.4. Delay Variation

This 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.

4.4. Unidirectional Link Loss Sub-TLV

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




```

|-----Residual Bandwidth-----|
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

4.5.1. Type

This sub-TLV has a type of TBD5.

4.5.2. Length

The length is 4.

4.5.3. Residual Bandwidth

This field carries the residual bandwidth on a link, forwarding adjacency [RFC4206], or bundled link in IEEE floating point format with units of bytes per second. For a link or forwarding adjacency, residual bandwidth is defined to be Maximum Bandwidth [RFC3630] 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.

The calculation of Residual Bandwidth is different than that of Unreserved Bandwidth [RFC3630]. Residual Bandwidth subtracts tunnel reservations from Maximum Bandwidth (i.e., the link capacity) [RFC3630] and provides an aggregated remainder across QoS classes. Unreserved Bandwidth [RFC3630], on the other hand, is subtracted from the Maximum Reservable Bandwidth (the bandwidth that can theoretically be reserved) [RFC3630] and provides per-QoS-class remainders. Residual Bandwidth and Unreserved Bandwidth [RFC3630] can be used concurrently, and each has a separate use case (e.g., the former can be used for applications like Weighted ECMP while the latter can be used for call admission control).

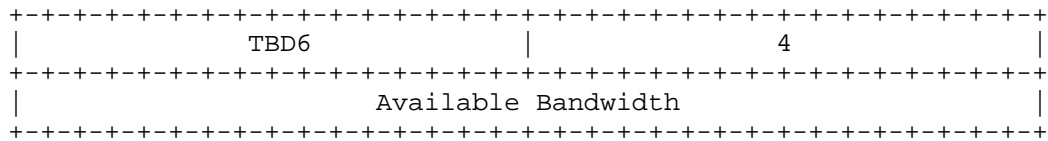
4.6. Unidirectional Available Bandwidth Sub-TLV

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

```

0           1           2           3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

```



4.6.1. Type

This sub-TLV has a type of TBD6.

4.6.2. Length

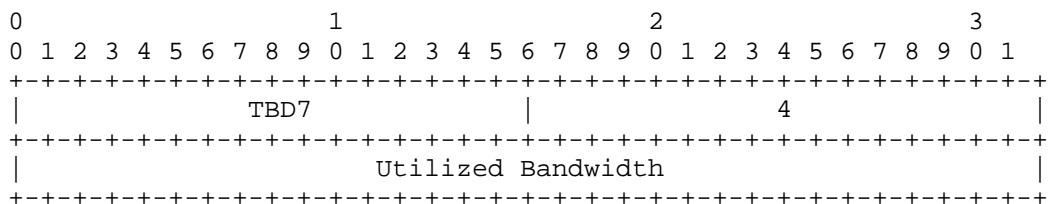
The length is 4.

4.6.3. Available Bandwidth

This field 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 4.5.) minus the measured bandwidth used for the actual forwarding of non-RSVP-TE LSP packets. For a bundled link, available bandwidth is defined to be the sum of the component link available bandwidths.

4.7. Unidirectional Utilized Bandwidth Sub-TLV

This Sub-TLV advertises the bandwidth utilization between two directly connected OSPF neighbors. The bandwidth utilization advertised by this sub-TLV MUST be the bandwidth from the advertising node to its neighbor. The format of this Sub-TLV is shown in the following diagram:



4.7.1. Type

This sub-TLV has a type of TBD7.

4.7.2. Length

The length is 4.

4.7.3. Utilized Bandwidth

This field carries the bandwidth utilization 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, bandwidth utilization represents the actual utilization of the link (i.e., as measured by the advertising node). For a bundled link, bandwidth utilization is defined to be the sum of the component link bandwidth utilizations.

5. Announcement Thresholds and Filters

The values advertised in all sub-TLVs (except min/max delay and residual bandwidth) MUST represent an average over a period or be obtained by a filter that is reasonably representative of an average. For example, a rolling average is one such filter.

Min and max delay MAY be the lowest and/or highest measured value over a measurement interval or MAY make use of a filter, or other technique to obtain a reasonable representation of a min and max value representative of the interval with compensation for outliers.

The measurement interval, any filter coefficients, and any advertisement intervals MUST be configurable for each sub-TLV.

In addition to the measurement intervals governing re-advertisement, implementations SHOULD provide for each sub-TLV configurable accelerated advertisement thresholds, such that:

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 re-advertisement 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 re-advertisement 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 back-off timer features SHOULD be configurable.

7. Network Stability and Announcement Periodicity

Sections 5 and 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 the advertisement of each sub-TLV.

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 [RFC3630], an unrecognized TLV should be silently ignored. I.e., it should not be processed but it should be included in LSAs sent to OSPF neighbors.

11. Security Considerations

This document does not introduce security issues beyond those discussed in [RFC3630]. OSPFv2 HMAC-SHA [RFC5709] provides additional protection for OSPFv2. OSPFv3 IPsec [RFC4552] and OSPFv3 Authentication Trailer [RFC7166] provide additional protection for OSPFv3.

OSPF KARP [RFC6863] provides an analysis of OSPFv2 and OSPFv3 routing security and OSPFv2 Security Extensions [OSPFSEC] provides extensions designed to address the identified gaps in OSPFv2.

12. IANA Considerations

IANA maintains the registry for the Link TLV sub-TLVs. OSPF TE Metric Extensions will require one new type code for each sub-TLV defined in this document, as follows:

Type	Description
------	-------------

TBD1	Unidirectional Link Delay
------	---------------------------

TBD2	Min/Max Unidirectional Link Delay
------	-----------------------------------

TBD3	Unidirectional Delay Variation
------	--------------------------------

TBD4	Unidirectional Link Loss
------	--------------------------

TBD5	Unidirectional Residual Bandwidth
------	-----------------------------------

TBD6	Unidirectional Available Bandwidth
------	------------------------------------

TBD7	Unidirectional Utilized Bandwidth
------	-----------------------------------

13. References

13.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3630] Katz, D., Kompella, K., Yeung, D., "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC5329] Ishiguro, K., Manral, V., Davey, A., Lindem, A., "Traffic Engineering Extensions to OSPF Version 3", RFC 5329, September 2009.

- [IEEE754] Institute of Electrical and Electronics Engineers,
"Standard for Floating-Point Arithmetic", IEEE Standard
754, August 2008.

13.2. Informative References

- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan,
V., Swallow, G., "RSVP-TE: Extensions to RSVP for LSP
Tunnels", RFC 3209, December 2001.
- [RFC4206] Kompella, K., Rekhter, Y., "Label Switched Paths (LSP)
Hierarchy with Generalized Multi-Protocol Label Switching
(GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.
- [RFC4552] Gupta, M., Melam, M., "Authentication/Confidentiality for
OSPFv3", RFC 4552, June 2006.
- [RFC4655] Farrel, A., Vasseur, J.-P., Ash, J., "A Path Computation
Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5250] Berger, L., Bryskin I., Zinin, A., Coltun, R., "The OSPF
Opaque LSA Option", RFC 5250, July 2008.
- [RFC5709] Bhatia, M., Manral, V., Fanto, M., White, R., Barnes, M.,
Li, T., Atkinson, R., "OSPFv2 HMAC-SHA Cryptographic
Authentication", RFC 5709, October 2009.
- [RFC6374] Frost, D., Bryant, S., "Packet Loss and Delay
Measurement for MPLS Networks", RFC 6374, September 2011.
- [RFC6863] Hartman, S., Zhang, D., "Analysis of OSPF Security
According to the Keying and Authentication for Routing
Protocols (KARP) Design Guide", RFC 6863, March 2013.
- [RFC7166] Bhatia, M., Manral, V., Lindem, A., "Supporting
Authentication Trailer for OSPFv3", RFC 7166, March 2014.
- [RFC7285] Almi, R., Penno, R., Yang, Y., Kiesel, S., Previdi, S.,
Roome, W., Shalunov, S., Woundy, R., "Application-Layer
Traffic Optimization (ALTO) Protocol", RFC 7285, September
2014.
- [OSPFSEC] Bhatia, M., Hartman, S., Zhang, D., Lindem, A., "Security
Extensions for OSPFv2 when using Manual Key Management",

draft-ietf-ospf-security-extension-manual-keying, Work in Progress.

14. Acknowledgments

The authors would like to recognize Nabil Bitar, Edward Crabbe, Don Fedyk, Acee Lindem, David McDysan, and Ayman Soliman for their contributions to this document.

The authors would also like to acknowledge Curtis Villamizar for his significant comments and direct content collaboration.

This document was prepared using 2-Word-v2.0.template.dot.

15. Author's Addresses

Spencer Giacalone
Unaffiliated

Email: spencer.giacalone@gmail.com

Dave Ward
Cisco Systems
170 West Tasman Dr.
San Jose, CA 95134, USA

Email: dward@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

Stefano Previdi
Cisco Systems
Via Del Serafico 200
00142 Rome
Italy

Email: sprevidi@cisco.com

