

OSPF Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 22, 2015

A. Atlas
S. Hegde
C. Bowers
Juniper Networks
J. Tantsura
Ericsson
Z. Li
Huawei Technologies
July 21, 2014

OSPF Extensions to Support Maximally Redundant Trees
draft-atlas-ospf-mrt-03

Abstract

This document specifies extensions to OSPF to support the distributed computation of Maximally Redundant Trees (MRT). Some example uses of the MRTs include IP/LDP Fast-Reroute and global protection or live-live for multicast traffic. The extensions indicate what MRT profile(s) each router supports. Different MRT profiles can be defined to support different uses and to allow transitioning of capabilities. An extension is introduced to flood MRT-Ineligible links, due to administrative policy.

The need for a mechanism to allow routers to advertise a worst-case FIB compute/install time is well understood for controlling convergence. This specification introduces the Controlled Convergence TLV to be carried in the Router Information LSA.

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 January 22, 2015.

Copyright Notice

Copyright (c) 2014 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
2. Requirements Language	3
3. Terminology	3
4. Overview of OSPF Extensions for MRT	4
4.1. Supporting MRT Profiles	4
4.2. GADAG Root Selection	5
4.3. Triggering an MRT Computation	5
5. MRT Capability Advertisement	6
5.1. Advertising MRT Capability in OSPFv2	6
5.2. Advertising MRT Capability in OSPFv3	7
5.3. MRT Profile TLV in Router Information LSA	8
6. Advertising MRT-ineligible links for MRT	9
6.1. MRT-Ineligible Link sub-TLV	10
7. Worst-Case Network Convergence Time	10
8. Backwards Compatibility	11
8.1. Handling MRT Capability Changes	12
9. Implementation Status	12
10. Security Considerations	12
11. IANA Considerations	12
12. References	13
12.1. Normative References	13
12.2. Informative References	13
Authors' Addresses	14

1. Introduction

This document describes the OSPF extensions necessary to support the architecture that defines how IP/LDP Fast-Reroute can use MRTs [I-D.ietf-rtgwg-mrt-frr-architecture]. At least one common standardized algorithm (such as the lowpoint algorithm explained and fully documented in [I-D.ietf-rtgwg-mrt-frr-algorithm]) is required

so that the routers supporting MRT computation consistently compute the same MRTs. MRT can also be used to protect multicast traffic via either global protection or local protection.[I-D.atlas-rtgwg-mrt-mc-arch]

IP/LDP Fast-Reroute using MRTs can provide 100% coverage for link and node failures in an arbitrary network topology where the failure doesn't split the network. It can also be deployed incrementally inside an OSPF area; an MRT Island is formed of connected supporting routers and the MRTs are computed inside that island.

In the default MRT profile, a supporting router both computes the MRTs and creates the necessary transit forwarding state necessary to provide the two additional forwarding topologies, known as MRT-Blue and MRT-Red.

2. 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 [RFC2119]

3. Terminology

For ease of reading, some of the terminology defined in [I-D.ietf-rtgwg-mrt-frr-architecture] is repeated here.

network graph: A graph that reflects the network topology where all links connect exactly two nodes and broadcast links have been transformed into the standard pseudo-node representation.

Redundant Trees (RT): A pair of trees where the path from any node X to the root R along the first tree is node-disjoint with the path from the same node X to the root along the second tree. These can be computed in 2-connected graphs.

Maximally Redundant Trees (MRT): A pair of trees where the path from any node X to the root R along the first tree and the path from the same node X to the root along the second tree share the minimum number of nodes and the minimum number of links. Each such shared node is a cut-vertex. Any shared links are cut-links. Any RT is an MRT but many MRTs are not RTs.

MRT Island: From the computing router, the set of routers that support a particular MRT profile and are connected via MRT-eligible links.

GADAG: Generalized Almost Directed Acyclic Graph - a graph that is the combination of the ADAGs of all blocks. Transforming a network graph into a GADAG is part of the MRT algorithm.

MRT-Red: MRT-Red is used to describe one of the two MRTs; it is used to describe the associated forwarding topology and MT-ID. Specifically, MRT-Red is the decreasing MRT where links in the GADAG are taken in the direction from a higher topologically ordered node to a lower one.

MRT-Blue: MRT-Blue is used to describe one of the two MRTs; it is used to describe the associated forwarding topology and MT-ID. Specifically, MRT-Blue is the increasing MRT where links in the GADAG are taken in the direction from a lower topologically ordered node to a higher one.

4. Overview of OSPF Extensions for MRT

There are two separate aspects that need to be advertised in OSPF. Both derive from the need for all routers supporting an MRT profile to compute the same pair of MRTs to each destination. By executing the same algorithm on the same network graph, distributed routers will compute the same MRTs. Convergence considerations are discussed in [I-D.ietf-rtgwg-mrt-frr-architecture].

The first aspect that must be advertised is which MRT profile(s) are supported and the associated GADAG Root Selection Priority. The second aspect that must be advertised is any links that are not eligible, due to administrative policy, to be part of the MRTs. This must be advertised consistently across the area so that all routers in the MRT Island use the same network graph.

4.1. Supporting MRT Profiles

An MRT Profile defines the exact MRT Algorithm, the MRT-Red LDP MT-ID, the MRT-Blue LDP MT-ID, and the forwarding mechanisms supported for the transit MRT-Red and MRT-Blue forwarding topologies. Finally, the MRT Profile defines exact behavioral rules such as:

- o how reconvergence is handled,
- o inter-area forwarding behavior,

A router that advertises support for an MRT Profile MUST provide the specified forwarding mechanism for its MRT-Red and MRT-Blue forwarding topologies. A router that advertises support for an MRT Profile MUST implement an algorithm that produces the same set of MRT-Red and MRT-Blue next-hops for its MRT-Red and MRT-Blue

topologies as is provided by the algorithm specified in the MRT Profile.

A router MAY indicate support for multiple MRT Profiles. A router computes its local MRT Island for each separate MRT Profile that the router supports. Supporting multiple MRT Profiles also provides a mechanism for transitioning from one profile to another. Different uses of MRT forwarding topologies may behave better on different MRT profiles.

The default MRT Profile is defined in [I-D.ietf-rtgwg-mrt-frr-architecture]. Its behavior is intended to support IP/LDP unicast and multicast fast-reroute.

4.2. GADAG Root Selection

One aspect of the MRT algorithms is that the selection of the GADAG root can affect the alternates and the traffic through that GADAG root. Therefore, it is important to provide an operator with control over which router will play the role of GADAG root. A measure of the centrality of a node may help determine how good a choice a particular node is.

The GADAG Root Selection Policy (defined as part of an MRT profile) may make use of the GADAG Root Selection Priority value advertised in the MRT Profile TLV of the Router Information LSA. For example, the GADAG Root Selection Policy for the default MRT profile is the following: Among the routers in the MRT Island and with the highest priority advertised, an implementation MUST pick the router with the highest Router ID to be the GADAG root.

4.3. Triggering an MRT Computation

When an MRT Computation is triggered, it is triggered for a given MRT Profile in a given area. First, the associated MRT Island is determined. Then, the GADAG Root is selected. Finally, the actual MRT algorithm is run to compute the transit MRT-Red and MRT-Blue topologies. Additionally, the router MAY choose to compute MRT-FRR alternates or make other use of the MRT computation results.

Prefixes can be attached and detached and have their associated MRT-Red and MRT-Blue next-hops computed without requiring a new MRT computation.

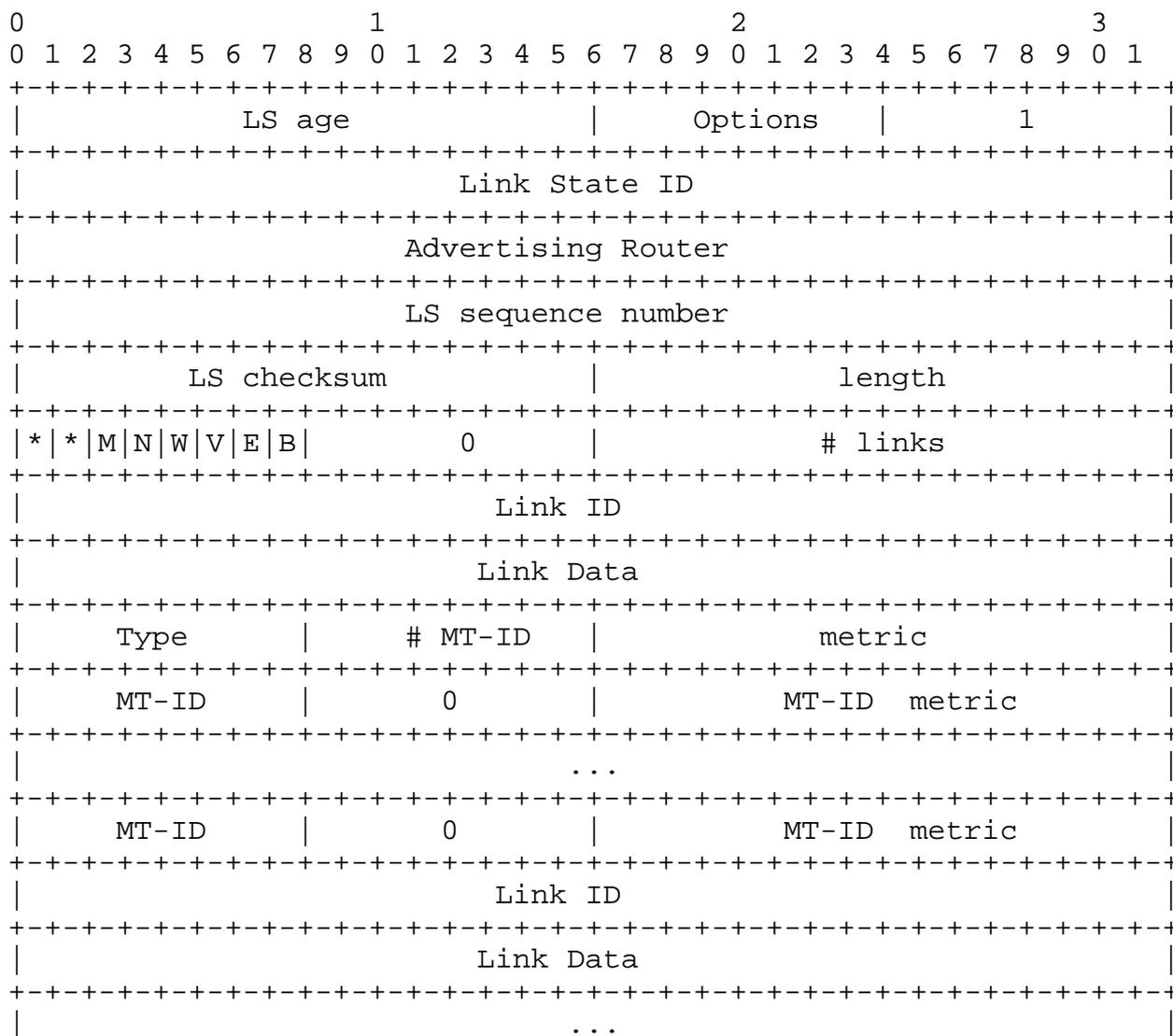
5. MRT Capability Advertisement

A router that supports MRT indicates this by setting a newly defined M bit in the Router LSA. If the router provides no other information via a separate MRT Profile TLV, then the router supports the default MRT Profile with a GADAG Root Selection Priority of 128.

In addition, a router can advertise a newly-defined MRT Profile TLV within the scope of the OSPF router information LSA [RFC4970]. This TLV also includes the GADAG Root Selection Priority.

5.1. Advertising MRT Capability in OSPFv2

A new M-bit is defined in the Router-LSA (defined in [RFC2328] and updated in [RFC4915]), as pictured below.

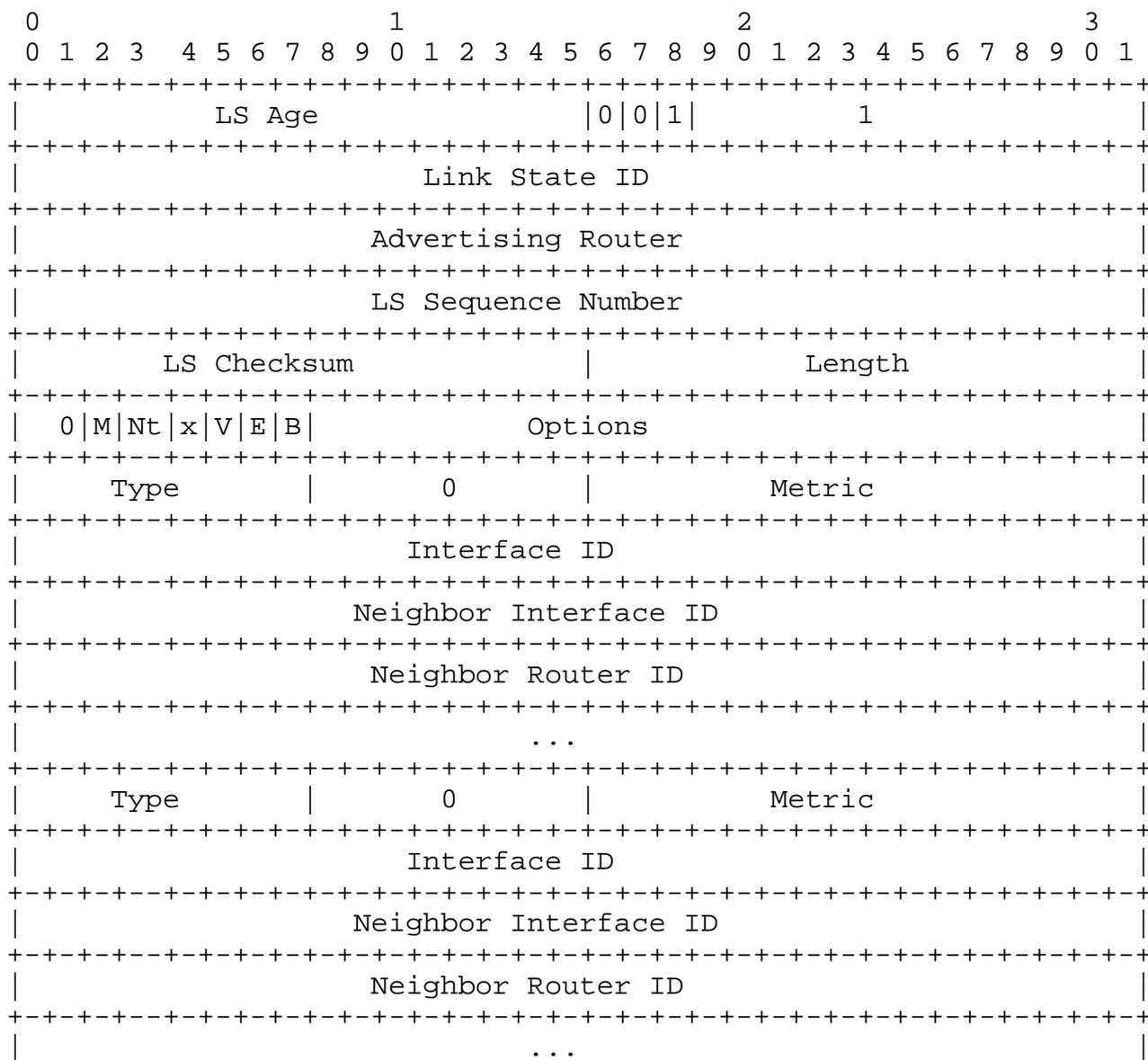


M-bit in OSPFv2 Router LSA

M bit: When set, the router supports MRT. If no separate MRT Profile TLV is advertised in the associated Router Information LSA, then the router supports the default MRT Profile and has a GADAG Root Selection Priority of 128.

5.2. Advertising MRT Capability in OSPFv3

Similarly, the M-bit is defined in the OSPFv3 Router LSA as shown below. Since there can be multiple router LSAs, the M-bit needs to be set on all of them.



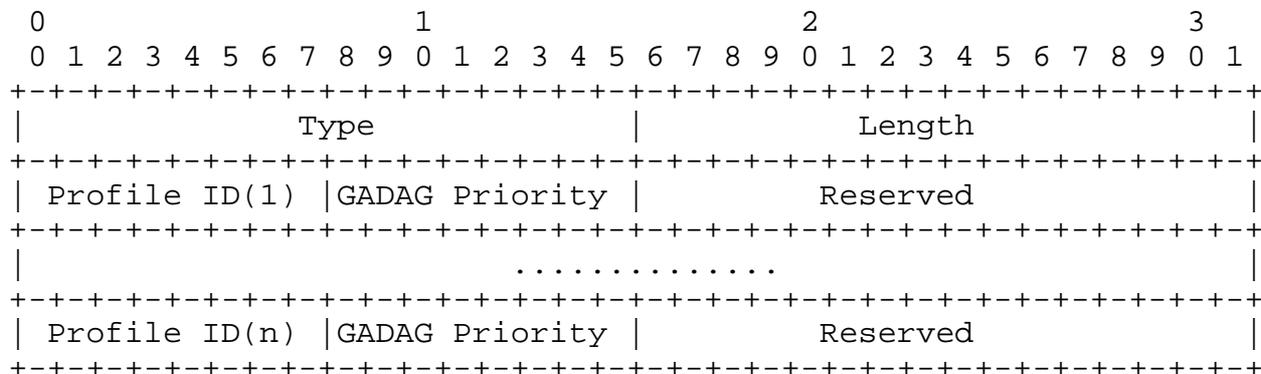
M-bit in OSPFv3 Router LSA

M bit: When set, the router supports MRT. If no separate MRT Profile TLV is advertised in the associated Router Information LSA, then the router supports the default MRT Profile and has a GADAG Root Selection Priority of 128.

5.3. MRT Profile TLV in Router Information LSA

A router may advertise an MRT Profile TLV to indicate support for multiple MRT Profiles, for a non-default MRT Profile, and/or to indicate a non-default GADAG Root Selection Priority. The MRT Profile TLV is advertised within the OSPF router information LSA

which is defined for both OSPFv2 and OSPFv3 in [RFC4970]. The RI LSA MUST have area scope.



TYPE: TBA-MRT-OSPF-1 (To Be Allocated by IANA)
 LENGTH: 4 * (number of Profiles)
 Profile ID : 1 byte
 GADAG Priority: 1 byte

MRT Profile TLV in Router Information LSA

Each Profile ID listed indicates support for a given MRT Profile, as defined in [I-D.ietf-rtgwg-mrt-frr-architecture]. A Profile ID value of 0 corresponds to the default MRT profile.

The GADAG Priority is the GADAG Root Selection Priority associated with the advertising router in the MRT Island for the associated MRT Profile, as indicated by the Profile ID. If multiple MRT Profiles are supported, then the length of this TLV varies. The ordering of the profiles inside the TLV is not significant. Multiple appearances of this TLV is not an error.

Lack of support for the default MRT profile is indicated by the presence of an MRT Profile TLV with a non-zero Profile ID value, and the absence of an MRT Profile TLV with a zero Profile ID value.

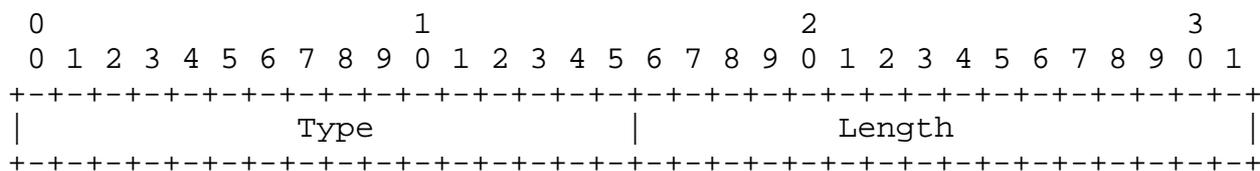
6. Advertising MRT-ineligible links for MRT

Due to administrative policy, some otherwise eligible links in the network topology may need to be excluded from the network graph upon which the MRT algorithm is run. Since the same network graph must be used across the area, it is critical for OSPF to flood which links to exclude from the MRT calculation. This is done by introducing a new MRT-Ineligible Link sub-TLV. For OSPFv2, this sub-TLV is carried in

the Extended Link TLV defined in [I-D.ietf-ospf-segment-routing-extensions]. For OSPFv3, this sub-TLV is carried in the Router-Link TLV defined in [I-D.ietf-ospf-ospfv3-lsa-extend].

If a link is marked by administrative policy as MRT-Ineligible, then a router MUST flood the OSPFv2 Extended Link TLV (or OSPFv3 Router-Link TLV) for that link, including the MRT-Ineligible Link sub-TLV. The OSPFv2 Extended Link TLV and OSPFv3 Router-Link TLV have area wide scope.

6.1. MRT-Ineligible Link sub-TLV



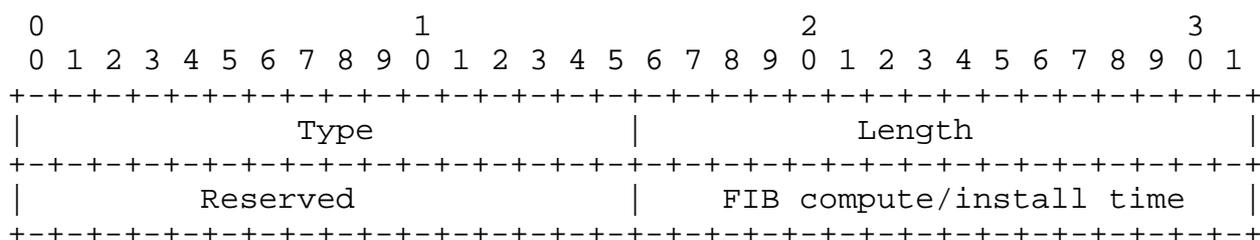
TYPE: TBA-MRT-OSPF-2 in OSPFv2 Extended Link TLV
 TBA-MRT-OSPF-3 in OSPFv3 Router-Link TLV
 (To Be Allocated by IANA)
 LENGTH: 0

MRT-Ineligible Link sub-TLV

This zero-length sub-TLV can appear in the OSPFv2 Extended Link TLV or the OSPFv3 Router-Link TLV. Its presence indicates that the associated link MUST NOT be used in the MRT calculation for all profiles.

7. Worst-Case Network Convergence Time

As part of converging the network after a single failure, Section 12.2 of [I-D.ietf-rtgwg-mrt-frr-architecture] describes the need to wait for a configured or advertised period for all routers to be using their new SPTs. Similarly, any work on avoiding micro-forwarding loops during convergence[RFC5715] requires determining the maximum among all routers in the area of the worst-case route computation and FIB installation time. More details on the specific reasoning and need for flooding it are given in [I-D.atlas-bryant-shand-lf-timers].



TYPE: TBA-MRT-OSPF-4 (To Be Allocated by IANA)

LENGTH: 4

FIB compute/install time: This is the worst-case time the router may take to compute and install all OSPF routes in the area after a change to a stable network. The value is in milliseconds.

Controlled Convergence TLV in Router Information LSA

The Controlled Convergence TLV is carried in the Router Information LSA and flooded with area-wide scope. The FIB compute/install time value sent by a router SHOULD be an estimate taking into account network scale or real-time measurements, or both. Advertisements SHOULD be dampened to avoid frequent communication of small changes in the FIB compute/install time.

A router receiving the Controlled Convergence sub-TLV SHOULD estimate the network convergence time as the maximum of the FIB compute/install times advertised by the routers in an area, including itself. In order to account for routers that do not advertise the Controlled Convergence sub-TLV, a router MAY use a locally configured minimum network convergence time as a lower bound on the computed network convergence time. A router MAY use a locally configured maximum network convergence time as an upper bound on the computed network convergence time.

8. Backwards Compatibility

The MRT capability bit, the MRT Profile, the MRT-Ineligible Link, and the OSPFv3 MRT-Ineligible Link TLVs are defined in this document. They should not introduce any interoperability issues. Routers that do not support the MRT capability bit in the router LSA SHOULD silently ignore it. Routers that do not support the new MRT-related TLVs SHOULD silently ignore them.

8.1. Handling MRT Capability Changes

When a router changes from supporting MRT to not supporting MRT, it is possible that Router Information LSAs with MRT-related TLVs remain in the neighbors' database briefly. Such MRT-related TLVs SHOULD be ignored when the associated Router LSA from that router does not have the MRT capability set in its Router LSA.

When a router changes from not supporting MRT to supporting MRT, it will flood its Router LSA(s) with the M-bit set and may send an updated Router Information LSA. If a Router LSA is received with the M-bit newly set, an MRT computation SHOULD be scheduled but MAY be delayed up to 60 seconds to allow reception of updated related Router Information LSAs. In general, when changes in MRT-related information is received, an MRT computation SHOULD be triggered.

The rationale behind using the M bit in router LSA is to handle the MRT capability changes gracefully in case of version upgrade/downgrade. The M bit in router LSA ensures the latest "MRT capability" information is available for computation when there is a downgrade to the version that doesn't support MRT and RI LSA.

9. Implementation Status

[RFC Editor: please remove this section prior to publication.]

Please see [I-D.ietf-rtgwg-mrt-frr-architecture] for details on implementation status.

10. Security Considerations

This OSPF extension is not believed to introduce new security concerns. It relies upon the security architecture already provided for Router LSAs and Router Information LSAs.

11. IANA Considerations

IANA is requested to allocate values for the following OSPF Router Information TLV Types [RFC4970]: MRT Profile TLV (TBA-MRT-OSPF-1), and Controlled Convergence TLV (TBA-MRT-OSPF-4).

IANA is requested to allocate a value from the OSPF Extended Link LSA sub-TLV registry [I-D.ietf-ospf-segment-routing-extensions] for the MRT-Ineligible Link sub-TLV (TBA-MRT-OSPF-2).

IANA is requested to allocate a value from the OSPFv3 Extended-LSA sub-TLV registry [I-D.ietf-ospf-ospfv3-lsa-extend] for the MRT-Ineligible Link sub-TLV (TBA-MRT-OSPF-3).

12. References

12.1. Normative References

[I-D.ietf-ospf-ospfv3-lsa-extend]

Lindem, A., Mirtorabi, S., Roy, A., and F. Baker, "OSPFv3 LSA Extendibility", draft-ietf-ospf-ospfv3-lsa-extend-03 (work in progress), May 2014.

[I-D.ietf-ospf-segment-routing-extensions]

Psenak, P., Previdi, S., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", draft-ietf-ospf-segment-routing-extensions-01 (work in progress), July 2014.

[I-D.ietf-rtgwg-mrt-frr-algorithm]

Enyedi, G., Csaszar, A., Atlas, A., Bowers, C., and A. Gopalan, "Algorithms for computing Maximally Redundant Trees for IP/LDP Fast-Reroute", draft-rtgwg-mrt-frr-algorithm-01 (work in progress), July 2014.

[I-D.ietf-rtgwg-mrt-frr-architecture]

Atlas, A., Kebler, R., Bowers, C., Enyedi, G., Csaszar, A., Tantsura, J., Konstantynowicz, M., and R. White, "An Architecture for IP/LDP Fast-Reroute Using Maximally Redundant Trees", draft-rtgwg-mrt-frr-architecture-04 (work in progress), July 2014.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC4970] Lindem, A., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 4970, July 2007.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.

12.2. Informative References

[I-D.atlas-bryant-shand-lf-timers]

K, A. and S. Bryant, "Synchronisation of Loop Free Timer Values", draft-atlas-bryant-shand-lf-timers-04 (work in progress), February 2008.

[I-D.atlas-rtgwg-mrt-mc-arch]

Atlas, A., Kebler, R., Wijnands, I., Csaszar, A., and G. Envedi, "An Architecture for Multicast Protection Using Maximally Redundant Trees", draft-atlas-rtgwg-mrt-mc-arch-02 (work in progress), July 2013.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3137] Retana, A., Nguyen, L., White, R., Zinin, A., and D. McPherson, "OSPF Stub Router Advertisement", RFC 3137, June 2001.

[RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P. Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF", RFC 4915, June 2007.

[RFC5715] Shand, M. and S. Bryant, "A Framework for Loop-Free Convergence", RFC 5715, January 2010.

Authors' Addresses

Alia Atlas
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
USA

Email: akatlas@juniper.net

Shraddha Hegde
Juniper Networks
Embassy Business Park
Bangalore, KA 560093
India

Email: shraddha@juniper.net

Chris Bowers
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
USA

Email: cbowers@juniper.net

Jeff Tantsura
Ericsson
300 Holger Way
San Jose, CA 95134
USA

Email: jeff.tantsura@ericsson.com

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: lizhenbin@huawei.com