Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 5, 2015

Z. Li
N. Wu
Q. Zhao
Huawei Technologies
A. Atlas
C. Bowers
Juniper Networks
J. Tantsura
Ericsson

July 4, 2014

Intermediate System to Intermediate System (IS-IS) Extensions for Maximally Redundant Trees (MRT)

draft-li-isis-mrt-01

Abstract

This document describes necessary extensions to IS-IS 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 transition of capabilities. An extension is introduced to flood MRT-Ineligible links, due to administrative policy.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

The IS-IS protocol is specified in [ISO10589], with extensions for supporting IPv4 and IPv6 specified in [RFC1195] and [RFC5308]. Each Intermediate System (IS) (router) advertises one or more IS-IS Link State Protocol Data Units (LSPs) with routing information. Each LSP is composed of a fixed header and a number of tuples, each consisting of a Type, a Length, and a Value. Such tuples are commonly known as TLVs, and are a good way of encoding information in a flexible and extensible format.

[I-D.ietf-rtgwg-mrt-frr-architecture] gives a complete solution for IP/LDP fast-reroute using Maximally Redundant Trees (MRT) to provide alternates. This document describes the necessary signaling extensions for supporting MRT-FRR used in IS-IS routing domain.

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

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 R 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 R 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 which 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. Using MRT with Multi-Topology IGP Routing

Both IS-IS and OSPF have support for multi-topology routing (see [RFC5120] for ISIS and [RFC4915] for OSPF.) In addition to the standard topology (identified by MT-ID=0), these extensions allow the IGP to identify particular links and nodes as participating in additional topologies (identified by MT-ID!=0). A given link can belong to several topologies and be assigned different metrics in each topology. The IGP runs an independent SPF computation for each topology, finding independent shortest paths to prefixes in each topology.

It is straightforward to extend the MRT computations to multi-topology IGP routing. For each IGP topology identified by an IGP MT-ID, we need to identify the node and links belonging to an MRT Island for that IGP MT-ID. This process creates a graph for the MRT Island for that specific IGP MT-ID, which can then be used to compute the transit next-hops and alternate next-hops for MRT-Red and MRT-Blue for that specific IGP MT-ID.

We expect that initial implementation and deployments of MRT will be primarily concerned with computing MRT-Red and Blue trees for the standard topology (IGP MT-ID=0). However, we have chosen to specify the IS-IS MRT extensions to accommodate the computation of MRT-Red and MRT-Blue in a multi-topology IS-IS environment. This comes at the expense of 2-6 octets per TLV for MT-ID values, but it will allow for standards-based multi-topology aware MRT implementations for ISIS without any future standards work.

Using MRT in a multi-topology IGP environment does have one complication which should be discussed. Forwarding LDP traffic over MRT paths in the standard IGP topology requires the use of labels bound to topology-scoped FECs to identify traffic on MRT-Red and Blue trees. This is described in Section 6 of [I-D.ietf-rtgwg-mrt-frr-architecture]. To facilitate this, an MRT profile specifies IANA-assigned MRT-Red and MRT-Blue LDP MT-ID values, which are then used by LDP to advertise labels for the MRT-Red and Blue forwarding topologies. Note that the MRT-Red and MRT-Blue LDP MT-ID values assigned by IANA for a given MRT profile

correspond to the MRT-Red and Blue forwarding trees associated with the standard IGP topology with IGP MT-ID=0. For example, suppose that a future MRT profile X is assigned (hypothetical) MRT-Red and MRT-Blue LDP MT-ID values of 2001 and 2002. Then labels for shortest path forwarding trees associated with the standard IGP topology will be advertised using FECs with MT-ID=0, while the labels for the MRT-Red and Blue forwarding trees for profile X will be advertised using FECS with MT-ID=2001 and 2002, respectively. In the absence of multi-topology IGP routing, all MT-IDs used by LDP for MRT are assigned by IANA, so there are no potential conflicts in LDP MT-ID usage.

When MRT is used together with multi-topology IGP routing, additional LDP MT-IDs need to be specified for carrying traffic on the MRT-Red and Blue forwarding trees associated with the additional IGP routing topologies. Building on the previous example, suppose that a network is configured with an additional IGP routing topology using MT-ID=20, in addition to the standard topology with MT-ID=0. The router advertises support for MRT with respect to MT-ID=20 with profile X, as well as support for MRT with respect to MT-ID=0 with profile X. The MRT-Red and Blue LDP MT-IDs for MT-ID=0 with profile X are still inherited from profile X, as in the previous example. In order to

use LDP to create the MRT-Red and Blue forwarding trees for the IGP topology with MT-ID=20, the router could, for example, advertise MRT-Red and MRT-Blue LDP MT-ID values of 21 and 22 for IGP MT-ID=20 and profile X. This overrides the (hypothetical) IANA-assigned values MRT-Red and MRT-Blue LDP MT-ID values for profile X, but maintains all other properties of profile X. Care must be taken to avoid advertising LDP MT-ID values that conflict with implicitly advertised IANA-assigned values LDP MT-ID.

The semantics of the IS-IS MRT extensions in this document are designed to handle the most common case (MRT in the absence of multi-topology IGP routing) in a simple manner. Setting the IGP MT-ID field as well as the MRT-Blue and MRT-Red LDP MT-ID fields to 0 in the TLV and sub-TLVs in this document results in the desired behavior for the standard IGP topology.

5. Overview of IS-IS Signaling Extensions for MRT

As stated in [I-D.ietf-rtgwg-mrt-frr-algorithm], it is necessary for each MRT-Capable router to compute MRT next hops in a consistent fashion. This is achieved by using same MRT profile and selecting the unique root in a MRT Island which is connected by MRT-Eligible links. Each of these issues will be discussed in following sections separately.

5.1. Supporting MRT Profiles

The contents and requirements of an MRT profile has been defined in [I-D.ietf-rtgwg-mrt-frr-architecture]. The parameters and behavioral rules contained in an MRT profile define one router's MRT capabilities. Based on common capabilities, one unified MRT Island is built.

The MRT-Capable router MUST advertise its corresponding MRT profiles by IS-IS protocol extension within IS-IS routing domain. The capabilities of advertiser MUST conform to the profile it claimed completely, especially the MT-IDs, the algorithm and the corresponding forwarding mechanism. This advertisement MUST have level scope. One router MAY support multiple MRT profiles and it MUST advertise these profiles in corresponding IS-IS level. The MT-

IDs used in one supported MRT Profile MUST NOT overlap with those MT-IDs used in a different supported MRT Profile.

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. MRT-Capable routers SHOULD support the default MRT profile.

5.2. Electing GADAG Root

As per [I-D.ietf-rtgwg-mrt-frr-algorithm], a GADAG root MUST be selected for one MRT Island. An unique GADAG root in common-sense among MRT Island routers is a necessity to do MRT computation. Since the selection of the GADAG root can affect the alternates and the traffic through it, the selection rules give network operator a knob to control the alternates and the traffic inside the MRT Island. Relevant discussion for the relationship between GADAG root role and MRT Island alternates is out of the scope of this document.

Each MRT-Capable router MUST advertise its priority for GADAG root selection. One router can only have one priority in the same MRT Island. It can have multiple priorities for different MRT Islands it supports. Routers that are marked as overloaded([RFC3787]) are not qualified as candidate for root selection.

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 in the IS-IS Router CAPABILITY TLV. 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.

When the current root is out of service or new router with higher priority joined into the MRT Island, the GADAG root MUST be re-selected. A new MRT computation will be triggered because of such a topology change.

5.3. Advertising MRT-Ineligible Links for MRT

For certain administrative or management reason, some links may not

be involved into MRT computation. In this scenario, MRT-Capable router MUST claim those MRT-Ineligible links are out of MRT Island scope. If such claim splits current MRT Island then MRT computation has to be done inside the modified MRT Island which the computing router belongs to.

5.4. Triggering an MRT Computation

A MRT Computation can be triggered through topology changes or MRT capability changes of any router in the MRT Island. It is always triggered for a given MRT Profile in the corresponding level. 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.

6. MRT Capability Advertisement

MRT-Capable router MUST identify its MRT capabilities through IS-IS Link State Packet(LSP) in level scope.

6.1. Advertising MRT Capability in IS-IS LSP

One new M-bit is introduced into TLV 229 to identify router is MRT-Capable. Structure of TLV 229 is stated in [RFC5120] as pictured below:

TYPE: 229

LENGTH: total length of the value field, it SHOULD be 2 times the number of MT components.

VALUE: one or more 2-byte MT components, structured as follows:

No. of Octets

```
+--------------------------------+
|O |A |M |R |        MT ID       |      2
+--------------------------------+
```

Bit M identifies the originator is of MRT-Capable. The MRT-Blue and the MRT-Red alternates will be calculated for the MT identified by MT-ID.

This M-bit MUST be set and checked in LSP fragment 0. A MRT-Capable router MUST advertise this TLV with M-bit set for corresponding MT. For instance, if M-bit is set for MT-ID #0, MRT alternates will be calculated for standard topology.

If only M-bit is advertised for MRT-Capabilities without any other MRT information then the router is regarded as supporting default MRT profile with default GADAG root selection priority.

6.2. MRT Profile sub-TLV in IS-IS Router CAPABILITY TLV

A new MRT Profile sub-TLV is introduced into IS-IS Router CAPABILITY TLV[RFC4971] to advertise MRT capabilities. Since MRT is per level scope, the S-bit and D-bit of IS-IS Router CAPABILITY TLV MUST be set to zero. The structure of the MRT Profile sub-TLV is pictured as below:

TYPE: TBA-MRT-ISIS-1 (To Be Allocated by IANA)

LENGTH: 8

VALUE:

MT ID (2 octet with 4 bits reserved)

Profile ID (1 octet)

MRT-Red LDP MT-ID (2 octet)

MRT-Blue LDP MT-ID (2 octet)

```
+--------------------------------+
|R |R |R |R |        MT ID       |      2
+----------------+---------------+
|   Profile ID   |                      1
+----------------+
| GADAG Priority |                      1
+----------------+---------------+
|      MRT-Red LDP MT-ID         |      2
+--------------------------------+
|      MRT-Blue LDP MT-ID        |      2
+--------------------------------+
```

12-bit MT ID represents the base MT topology which MRT computation is based on. Profile ID represents the MRT profile this router supports and GADAG Root Selection Priority is the priority for root selection. The range of this priority is [0, 255] with 128 as the default value. The GADAG Root Selection Policy defined as part of a given MRT profile determine how the GADAG Root Selection Priority value is used.

If the MRT-Blue LDP MT-ID is 0, then the value specified in the associated MRT Profile is assumed. If the MRT-Red LDP MT-ID is 0, then the value specified in the associated MRT profile is assumed. The MRT-Blue LDP MT-ID and MRT-Red LDP MT-ID MUST NOT be the reserved values for LDP MT-IDs ([I-D.ietf-mpls-ldp-multi-topology]). The value for MRT-Blue LDP MT-ID and MRT-Red LDP MT-ID MUST be different except for 0. As stated above, the MRT-Blue LDP MT-ID and MRT-Red LDP MT-ID MUST NOT overlap among profiles if multiple MRT-Profile sub-TLVs are advertised.

This sub-TLV can occur multiple times if this router support multiple MRT profiles. This can happen during transition or to support multiple uses of MRT which prefer different profiles.

6.3. MRT-Ineligible Links sub-TLV in IS-IS Router CAPABILITY TLV

As a matter of policy, some links may not be available for the MRT computation, which can prevent alternates or traffic using these links. For instance, policy can be made to prevent fast-rerouted traffic from taking those links.

For a link to be excluded from the MRT computation, it MUST be advertised as sub-TLV in IS-IS Router CAPABILITY TLV which is in level scope with S-bit and D-bit unset. The MRT-Ineligible Link sub-TLV is structured as below:

TYPE: TBA-MRT-ISIS-2 (To Be Allocated by IANA)

LENGTH: from 9 to 255 octets

VALUE:

MT ID (2 octet with 4 bits reserved)

System ID and pseudo-node number (7 octet for each MRT-Ineligible Link)

```
                                                No. of Octets
+--------------------------------+
|R |R |R |R |        MT ID       |      2
+--------------------------------+
|System ID and pseudonode number |      7
+--------------------------------+
|         Default metric         |      3
+--------------------------------+
.                                .
.                                .
+--------------------------------+
|System ID and pseudonode number |      7
+--------------------------------+
|         Default metric         |      3
+--------------------------------+
```

Each MRT-Ineligible Link is identified by neighbor's System ID and pseudo-node number and Default metric, same as IS Reachability TLV. This sub-TLV MAY occur multiple times if multiple links are ineligible.

7. Controlled Convergence sub-TLV in IS-IS Router CAPABILITY TLV

Section 12.2 of [I-D.ietf-rtgwg-mrt-frr-architecture] describes the need to wait for a configured or advertised period after a network failure to insure that all routers are using their new SPTs. Similarly, 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 this

value are given in [I-D.atlas-bryant-shand-lf-timers].

A new Controlled Convergence sub-TLV is introduced into the IS-IS Router CAPABILITY TLV [RFC4971] to advertise the worst-case time for a router to compute and install all IS-IS routes in the level after a change to a stable network. This advertisement has per level scope, so the S-bit and D-bit of IS-IS Router CAPABILITY TLV MUST be set to zero. The advertisement is scoped by IGP MT-ID, allowing a router supporting multi-topology IGP routing to advertise a different worst-

case compute and install time for each IGP topology. This make sense as the SPF computations for each IGP topology are independent of one another, and may have different worst-case compute and install times.

The structure of the Controlled Convergence sub-TLV is shown below:

TYPE: TBA-MRT-ISIS-3 (To Be Allocated by IANA)

LENGTH: 3

VALUE:

MT ID (2 octet with 4 bits reserved)

FIB compute/install time (1 octet)

```
+--------------------------------+
|R |R |R |R |        MT ID       |      2
+----------------+---------------+
|FIB comp/in time|                      1
+----------------+
```

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

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 a level, 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. Handling MRT Capability Sending and Receiving

The M-bit which identifies router's MRT capability MUST be advertised in LSP fragment 0. Those MRT related sub-TLVs SHOULD be ignored when MRT Capability bit is unset. When changes in MRT capabilities are received, a MRT computation SHOULD be triggered but MAY be delayed for a while to allow reception of all MRT-related information.

8.1. Advertising MRT extension

MRT sub-TLVs are encapsulated in the Router Capability TLV and advertised through LSP PDU for the level-wide. MRT sub-TLVs are optional. If one router does not support MRT, it MUST NOT advertise those sub-TLVs.

Since the advertisement scope of the MRT sub-TLV is level-wide, the D-Bit and S-Bit of the Router Capability TLV MUST be set as 0 when it is advertised. If other sub-TLVs in the Router Capability TLV need different values for those two bits, there MUST be an independent Router Capability TLV for MRT sub-TLVs.

When MRT related information is changed for the router or existing IS-IS LSP mechanisms are triggered for refreshing or updating, MRT sub-TLVs MUST be advertised if the router is MRT-Capable.

For administrative policies or reasons, it may be desirable to exclude certain links from the MRT computation. MRT-Ineligible sub-TLV is used to advertise which links should be excluded. Note that an interface advertised as MRT-Ineligigle by a router is ineligible with respect to all profiles advertised by that router.

8.2. Parsing MRT extension

MRT extension MUST NOT affect the peer setup and the routing calculation of the standard topology.

MRT sub-TLVs SHOULD be validated like other sub-TLVs when received. MRT sub-TLVs SHOULD also be taken for the checksum calculation and authentication.

If MT-ID conflict is found for MRT-Red or MRT-blue from multiple sub-TLVs then those associated sub-TLVs MUST be ignored.

Links advertised in MRT-Ineligible sub-TLV MUST be precluded from MRT Computation. The removal of those links may change the computing router's MRT Island significantly.

9. Backwards Compatibility

The M-bit for MRT capability, the MRT Profile sub-TLV and the MRT-Ineligible Link sub-TLV defined in this document SHOULD NOT introduce any interoperability issues. Routers that do not support these MRT extensions SHOULD silently ignore them. Alternates or traffic MUST NOT be affected in current IS-IS routing domain.

10. 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.

11. Security Considerations

This IS-IS extension is not believed to introduce new security concerns.

12. IANA Considerations

Please allocate values for the following IS-IS Router CAPABILITY TLV Types [RFC4971]: MRT Profile sub-TLV (TBA-MRT-ISIS-1), MRT-Ineligible Link sub-TLV (TBA-MRT-ISIS-2), and Controlled Convergence sub-TLV (TBA-MRT-ISIS-3).

13. References

13.1. Normative References

[I-D.ietf-mpls-ldp-multi-topology]
Zhao, Q., Raza, K., Zhou, C., Fang, L., Li, L., and D. King, "LDP Extensions for Multi Topology", draft-ietf-mpls-ldp-multi-topology-12 (work in progress), April 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.

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

[RFC3787] Parker, J., "Recommendations for Interoperable IP Networks using Intermediate System to Intermediate System (IS-IS)", RFC 3787, May 2004.

13.2. Infomative 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.

[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.

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

[RFC4971] Vasseur, JP., Shen, N., and R. Aggarwal, "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information", RFC 4971, July 2007.

[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.

[RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, October 2008.

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

Authors' Addresses

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

Email: lizhenbin@huawei.com

Nan Wu
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095

China

Email: eric.wu@huawei.com

Quintin Zhao
Huawei Technologies
125 Nagog Technology Park
Acton, MA 01719
USA

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

Email: akatlas@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