Routing Area Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 13, 2016

A. Atlas
C. Bowers
Juniper Networks
G. Enyedi
Ericsson

January 10, 2016

An Architecture for IP/LDP Fast-Reroute Using Maximally Redundant Trees

draft-ietf-rtgwg-mrt-frr-architecture-09

Abstract

This document defines the architecture for IP/LDP Fast-Reroute using Maximally Redundant Trees (MRT-FRR). MRT-FRR is a technology that gives link-protection and node-protection with 100% coverage in any network topology that is still connected after the failure.

Status of This Memo

This Internet-Draft will expire on July 13, 2016.

Copyright Notice

Table of Contents

1. Introduction

This document describes a solution for IP/LDP fast-reroute [RFC5714]. MRT-FRR creates two alternate trees separate from the primary next-hop forwarding used during stable operation. These two trees are maximally diverse from each other, providing link and node protection for 100% of paths and failures as long as the failure does not cut the network into multiple pieces. This document defines the architecture for IP/LDP fast-reroute with MRT.

[I-D.ietf-rtgwg-mrt-frr-algorithm] describes how to compute maximally redundant trees using a specific algorithm, the MRT Lowpoint algorithm. The MRT Lowpoint algorithm is used by a router that supports the Default MRT Profile, as specified in this document.

IP/LDP Fast-Reroute with MRT (MRT-FRR) uses two maximally diverse forwarding topologies to provide alternates. A primary next-hop should be on only one of the diverse forwarding topologies; thus, the other can be used to provide an alternate. Once traffic has been moved to one of the MRTs by one point of local repair (PLR), that traffic is not subject to further repair actions by another PLR, even in the event of multiple simultaneous failures. Therefore, traffic repaired by MRT-FRR will not loop between different PLRs responding to different simultaneous failures.

While MRT provides 100% protection for a single link or node failure, it may not protect traffic in the event of multiple simultaneous failures, nor does take into account Shared Risk Link Groups (SRLGs). Also, while the MRT Lowpoint algorithm is computationally efficient, it is also new. In order for MRT-FRR to function properly, all of the other nodes in the network that support MRT must correctly compute next-hops based on the same algorithm, and install the corresponding forwarding state. This is in contrast to other FRR methods where the calculation of backup paths generally involves repeated application of the simpler and widely-deployed SPF algorithm, and backup paths themselves re-use the forwarding state used for shortest path forwarding of normal traffic. Section 14 provides operational guidance related to verification of MRT forwarding paths.

In addition to supporting IP and LDP unicast fast-reroute, the diverse forwarding topologies and guarantee of 100% coverage permit fast-reroute technology to be applied to multicast traffic as described in [I-D.atlas-rtgwg-mrt-mc-arch]. However, the current document does not address the multicast applications of MRTs.

Figure 1 compares different methods of providing FRR for IP and LDP traffic, illustrating some of the trade-offs between the different approaches. For several methods, the methods are separated into link-protecting (LP) and node-protecting (NP) variants. The first column indicates whether the method provides 100% protection coverage (when topologically feasible). The second column indicates whether or not traffic traversing alternate path can loop when multiple failures occur.

The third column gives an estimate of the amount of computation required at each node to support the FRR method, in addition to the usual SPF computation rooted at the computing node itself. This metric of comparison is important for implementations that seek to quickly recompute repair paths following their initial use after a topology change, without reliance on an external system, in order to minimize the risk of a new failure occurring before the new repair paths are in place.

The fourth column indicates the maximum number of additional labels that need to be applied to packets to support the FRR method, while the fifth column gives the size of the MPLS label table needed to support the FRR method. These two metrics may be useful for evaluating requirements placed on hardware to support the different FRR methods.

The last column indicates the additional requirements placed on the control plane by the FRR method, beyond what is required for IGP shortest path forwarding with LDP.

Method	100% prot. cov.	Alt. can loop?	Additional computation at each node	Max. addit. labels	Label table size (relative to SPF labels)	Control plane reqs.

MRT-FRR LP and NP	Yes	No	equivalent of less than 3 SPFs	0(LDP)
				1(IP)
					3x	MRT-specific protocol extens.

LFA LP and NP	No	Yes	SPF per neighbor	0	1x	None
Remote LFA LP	No	Yes	SPF per neighbor	1	1x	T-LDP session for each selected PQ node
Remote LFA NP	No	Yes	SPF per nbr and SPF per per PQ node evaluated	1	1x	T-LDP session for each selected PQ node
Not-Via LP and NP	Yes	No	SPF per link and per node in topology	1	(average number of neighbors) x	T-LDP session for each nbr's nbr
TI-LFA LP with symm. metrics	Yes	Yes	SPF per neighbor	2	1x	uses SPRING for label dist.
TI-LFA NP or LP with asymm. metrics	Yes	Yes	# of SPFs dependent on topology	depth dependent on topo.	1x	uses SPRING for label dist.

Figure 1: Comparison of IP/LDP FRR Methods

MRT: MRT provides 100% coverage for link and node protection, and traffic on the alternate paths will not loop. The computation

required on each node is equivalent to less than 3 additional SPFs in terms of computational complexity. For IP traffic, MRT requires one additional label, while for LDP traffic, no additional labels are needed. However, the size of the MPLS label table is three times what would normally be required for shortest path LDP forwarding, due to the presence of additional red and blue labels for each FEC. MRT requires protocol extensions in order to allow a node to advertise support for MRT as well as a value for the GADAG Root Selection Priority. It also requires support for multi-topology LDP extensions.

Loop-Free Alternates (LFA): LFAs [RFC5286] provide limited topology-dependent coverage for link and node protection. Restrictions on choice of alternates can be relaxed to improve coverage, but this can cause forwarding loops if a worse failure is experienced than protected against. [RFC6571] discusses the applicability of LFA to different topologies with a focus on common PoP architectures. The computation required is one SPF per neighbor. LFA imposes no additional labels imposed, has no impact on the label table size, and has no additional control plane requirements.

Remote LFA: Remote LFA [RFC7490] improves coverage over LFA for both link and node protection, but it does not guarantee 100% coverage. The alternates can also loop with worse than expected failures. Computation for link protection is one SPF per neighbor, while computation for node protection requires an additional SPF per PQ node [I-D.ietf-rtgwg-rlfa-node-protection]. Remote LFA can impose up to one additional label on the packet, but does not increase the size of the label table. It requires a T-LDP session for each selected PQ node.

Not-Via: Not-Via [RFC6981] provides 100% coverage for link and node failures and does not have potential looping among alternates. The computation is high with an SPF per potential failure point (all links and nodes in the topology). When implemented with LDP, Not-Via adds one additional label to a tunnelled packet. It significantly increases the size of the label table, multiplying it by roughly the average number of neighbors. Not-Via also requires targeted LDP sessions to each neighbor's neighbor.

TI-LFA: Topology Independent Loop-free Alternate Fast Re-route (TI-LFA) [I-D.francois-rtgwg-segment-routing-ti-lfa] aims to provide link and node protection of node and adjacency segments within the Segment Routing (SR) framework. It guarantees complete coverage. The TI-LFA computation for link-protection is fairly straightforward, while the computation for node-protection is more complex. For link-protection with symmetric link costs, TI-LFA

can provide complete coverage by pushing up to two additional labels. For node protection on arbitrary topologies, the label stack size can grow significantly based on repair path. Note that TI-LFA requires shortest path forwarding based on SR Node-SIDs, as opposed to LDP labels, in order to construct label stacks for backups paths without relying on a large number of targeted LDP sessions to learn remote FEC-label bindings. It also requires the use of Adj-SIDs to achieve 100% coverage.

1.1. Importance of 100% Coverage

Fast-reroute is based upon the single failure assumption - that the time between single failures is long enough for a network to reconverge and start forwarding on the new shortest paths. That does not imply that the network will only experience one failure or change.

It is straightforward to analyze a particular network topology for coverage. However, a real network does not always have the same topology. For instance, maintenance events will take links or nodes out of use. Simply costing out a link can have a significant effect on what LFAs are available. Similarly, after a single failure has happened, the topology is changed and its associated coverage. Finally, many networks have new routers or links added and removed; each of those changes can have an effect on the coverage for topology-sensitive methods such as LFA and Remote LFA. If fast-reroute is important for the network services provided, then a method that guarantees 100% coverage is important to accomodate natural network topology changes.

Asymmetric link costs are also a common aspect of networks. There are at least three common causes for them. First, any broadcast interface is represented by a pseudo-node and has asymmetric link costs to and from that pseudo-node. Second, when routers come up or a link with LDP comes up, it is recommended in [RFC5443] and [RFC6987] that the link metric be raised to the maximum cost; this may not be symmetric and for [RFC6987] is not expected to be. Third, techniques such as IGP metric tuning for traffic-engineering can result in asymmetric link costs. A fast-reroute solution needs to handle network topologies with asymmetric link costs.

When a network needs to use a micro-loop prevention mechanism [RFC5715] such as Ordered FIB[RFC6976] or Nearside Tunneling[RFC5715], then the whole IGP area needs to have alternates available so that the micro-loop prevention mechanism, which requires slower network convergence, can take the necessary time without adversely impacting traffic. Without complete coverage, traffic to the unprotected destinations will be dropped for significantly longer

than with current convergence - where routers individually converge as fast as possible. See Section 12.1 for more discussion of micro-loop prevention and MRTs.

1.2. Partial Deployment and Backwards Compatibility

MRT-FRR supports partial deployment. Routers advertise their ability to support MRT. Inside the MRT-capable connected group of routers (referred to as an MRT Island), the MRTs are computed. Alternates to destinations outside the MRT Island are computed and depend upon the existence of a loop-free neighbor of the MRT Island for that destination. MRT Islands are discussed in detail in Section 7, and partial deployment is discussed in more detail in Section 14.5.

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

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.

cut-link: A link whose removal partitions the network. A cut-link by definition must be connected between two cut-vertices. If there are multiple parallel links, then they are referred to as cut-links in this document if removing the set of parallel links would partition the network graph.

cut-vertex: A vertex whose removal partitions the network graph.

2-connected: A graph that has no cut-vertices. This is a graph that requires two nodes to be removed before the network is partitioned.

2-connected cluster: A maximal set of nodes that are 2-connected.

block: Either a 2-connected cluster, a cut-edge, or an isolated vertex.

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. Redundant trees can always 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. In graphs that are not 2-connected, it is not possible to compute RTs. However, it is possible to compute MRTs. MRTs are maximally redundant in the sense that they are as redundant as possible given the constraints of the network graph.

DAG: Directed Acyclic Graph - a graph where all links are directed and there are no cycles in it.

ADAG: Almost Directed Acyclic Graph - a graph that, if all links incoming to the root were removed, would be a DAG.

GADAG: Generalized ADAG - a graph that is the combination of the ADAGs of all blocks.

MRT-Red: MRT-Red is used to describe one of the two MRTs; it is used to describe the associated forwarding topology and MPLS multi-topology identifier (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 described the associated forwarding topology and MPLS 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.

Rainbow MRT: It is useful to have an MPLS MT-ID that refers to the multiple MRT forwarding topologies and to the default forwarding topology. This is referred to as the Rainbow MRT MPLS MT-ID and is used by LDP to reduce signaling and permit the same label to always be advertised to all peers for the same (MT-ID, Prefix).

MRT Island: The set of routers that support a particular MRT profile and the links connecting them that support MRT.

Island Border Router (IBR): A router in the MRT Island that is connected to a router not in the MRT Island and both routers are in a common area or level.

Island Neighbor (IN): A router that is not in the MRT Island but is adjacent to an IBR and in the same area/level as the IBR.

named proxy-node: A proxy-node can represent a destination prefix that can be attached to the MRT Island via at least two routers. It is named if there is a way that traffic can be encapsulated to reach specifically that proxy node; this could be because there is an LDP FEC for the associated prefix or because MRT-Red and MRT-Blue IP addresses are advertised in an undefined fashion for that proxy-node.

4. Maximally Redundant Trees (MRT)

A pair of Maximally Redundant Trees is a pair of directed spanning trees that provides maximally disjoint paths towards their common root. Only links or nodes whose failure would partition the network (i.e. cut-links and cut-vertices) are shared between the trees. The MRT Lowpoint algorithm is given in [I-D.ietf-rtgwg-mrt-frr-algorithm]. This algorithm can be computed in O(e + n log n); it is less than three SPFs. This document describes how the MRTs can be used and not how to compute them.

MRT provides destination-based trees for each destination. Each router stores its normal primary next-hop(s) as well as MRT-Blue next-hop(s) and MRT-Red next-hop(s) toward each destination. The alternate will be selected between the MRT-Blue and MRT-Red.

The most important thing to understand about MRTs is that for each pair of destination-routed MRTs, there is a path from every node X to the destination D on the Blue MRT that is as disjoint as possible from the path on the Red MRT.

For example, in Figure 2, there is a network graph that is 2-connected in (a) and associated MRTs in (b) and (c). One can consider the paths from B to R; on the Blue MRT, the paths are B->F->D->E->R or B->C->D->E->R. On the Red MRT, the path is B->A->R. These are clearly link and node-disjoint. These MRTs are redundant trees because the paths are disjoint.

```
[E]---[D]---|           [E]<--[D]<--|                [E]-->[D]---|
 |     |    |            |     ^    |                       |    |
 |     |    |            V     |    |                       V    V
[R]   [F]  [C]          [R]   [F]  [C]               [R]   [F]  [C]
 |     |    |                  ^    ^                 ^     |    |
 |     |    |                  |    |                 |     V    |
[A]---[B]---|           [A]-->[B]---|                [A]<--[B]<--|

      (a)                     (b)                         (c)
a 2-connected graph     Blue MRT towards R          Red MRT towards R
```

Figure 2: A 2-connected Network

By contrast, in Figure 3, the network in (a) is not 2-connected. If F, G or the link F<->G failed, then the network would be partitioned. It is clearly impossible to have two link-disjoint or node-disjoint paths from G, I or J to R. The MRTs given in (b) and (c) offer paths that are as disjoint as possible. For instance, the paths from B to R are the same as in Figure 2 and the path from G to R on the Blue MRT is G->F->D->E->R and on the Red MRT is G->F->B->A->R.

```
                      [E]---[D]---|
                       |     |    |     |----[I]
                       |     |    |     |     |
                      [R]---[C]  [F]---[G]    |
                       |     |    |     |     |
                       |     |    |     |----[J]
                      [A]---[B]---|

                                  (a)
                        a non-2-connected graph

       [E]<--[D]<--|                        [E]-->[D]
        |     ^    |          [I]                  |          |----[I]
        V     |    |           |                   V          V     ^
       [R]   [C]  [F]<--[G]    |            [R]<--[C]  [F]<--[G]    |
              ^    ^     ^     V             ^          |           |
              |    |     |----[J]            |          |          [J]
       [A]-->[B]---|                        [A]<--[B]<--|

                   (b)                                    (c)
            Blue MRT towards R                    Red MRT towards R
```

Figure 3: A non-2-connected network

5. Maximally Redundant Trees (MRT) and Fast-Reroute

In normal IGP routing, each router has its shortest path tree (SPT)to all destinations. From the perspective of a particular destination, D, this looks like a reverse SPT. To use maximally redundant trees, in addition, each destination D has two MRTs associated with it; by convention these will be called the MRT-Blue and MRT-Red. MRT-FRR is realized by using multi-topology forwarding. There is a MRT-Blue forwarding topology and a MRT-Red forwarding topology.

Any IP/LDP fast-reroute technique beyond LFA requires an additional dataplane procedure, such as an additional forwarding mechanism. The well-known options are multi-topology forwarding (used by MRT-FRR), tunneling (e.g. [RFC6981] or [RFC7490]), and per-interface

forwarding (e.g. Loop-Free Failure Insensitive Routing in [EnyediThesis]).

When there is a link or node failure affecting, but not partitioning, the network, each node will still have at least one path via one of the MRTs to reach the destination D. For example, in Figure 3, C would normally forward traffic to R across the C<->R link. If that C<->R link fails, then C could use the Blue MRT path C->D->E->R.

As is always the case with fast-reroute technologies, forwarding does not change until a local failure is detected. Packets are forwarded along the shortest path. The appropriate alternate to use is pre-computed. [I-D.ietf-rtgwg-mrt-frr-algorithm] describes exactly how to determine whether the MRT-Blue next-hops or the MRT-Red next-hops should be the MRT alternate next-hops for a particular primary next-hop to a particular destination.

MRT alternates are always available to use. It is a local decision whether to use an MRT alternate, a Loop-Free Alternate or some other type of alternate.

As described in [RFC5286], when a worse failure than is anticipated happens, using LFAs that are not downstream neighbors can cause looping among alternates. Section 1.1 of [RFC5286] gives an example of link-protecting alternates causing a loop on node failure. Even if a worse failure than anticipated happens, the use of MRT alternates will not cause looping.

6. Unicast Forwarding with MRT Fast-Reroute

There are three possible types of routers involved in forwarding a packet along an MRT path. At the MRT ingress router, the packet leaves the shortest path to the destination and follows an MRT path to the destination. In a FRR application, the MRT ingress router is the PLR. An MRT transit router takes a packet that arrives already associated with the particular MRT, and forwards it on that same MRT. In some situations (to be discussed later), the packet will need to leave the MRT path and return to the shortest path. This takes place at the MRT egress router. The MRT ingress and egress functionality may depend on the underlying type of packet being forwarded (LDP or IP). The MRT transit functionality is independent of the type of packet being forwarded. We first consider several MRT transit forwarding mechanisms. Then we look at how these forwarding mechanisms can be applied to carrying LDP and IP traffic.

6.1. MRT Forwarding Mechanisms

The following options for MRT forwarding mechanisms are considered.

1. MRT LDP Labels

    A. Topology-scoped FEC encoded using a single label

    B. Topology and FEC encoded using a two label stack

2. MRT IP Tunnels

    A. MRT IPv4 Tunnels

    B. MRT IPv6 Tunnels

6.1.1. MRT LDP labels

We consider two options for the MRT forwarding mechanisms using MRT LDP labels.

6.1.1.1. Topology-scoped FEC encoded using a single label (Option 1A)

[RFC7307] provides a mechanism to distribute FEC-Label bindings scoped to a given MPLS topology (represented by MPLS MT-ID). To use multi-topology LDP to create MRT forwarding topologies, we associate two MPLS MT-IDs with the MRT-Red and MRT-Blue forwarding topologies, in addition to the default shortest path forwarding topology with MT-ID=0.

With this forwarding mechanism, a single label is distributed for each topology-scoped FEC. For a given FEC in the default topology (call it default-FEC-A), two additional topology-scoped FECs would be created, corresponding to the Red and Blue MRT forwarding topologies (call them red-FEC-A and blue-FEC-A). A router supporting this MRT transit forwarding mechanism advertises a different FEC-label binding for each of the three topology-scoped FECs. When a packet is received with a label corresponding to red-FEC-A (for example), an MRT transit router will determine the next-hop for the MRT-Red forwarding topology for that FEC, swap the incoming label with the outgoing label corresponding to red-FEC-A learned from the MRT-Red next-hop router, and forward the packet.

This forwarding mechanism has the useful property that the FEC associated with the packet is maintained in the labels at each hop along the MRT. We will take advantage of this property when specifying how to carry LDP traffic on MRT paths using multi-topology LDP labels.

This approach is very simple for hardware to support. However, it reduces the label space for other uses, and it increases the memory needed to store the labels and the communication required by LDP to distribute FEC-label bindings.

This forwarding option uses the LDP signaling extensions described in [RFC7307]. The MRT-specific LDP extensions required to support this option will be described elsewhere.

6.1.1.2. Topology and FEC encoded using a two label stack (Option 1B)

With this forwarding mechanism, a two label stack is used to encode the topology and the FEC of the packet. The top label (topology-id label) identifies the MRT forwarding topology, while the second label (FEC label) identifies the FEC. The top label would be a new FEC type with two values corresponding to MRT Red and Blue topologies.

When an MRT transit router receives a packet with a topology-id label, the router pops the top label and uses that it to guide the next-hop selection in combination with the next label in the stack (the FEC label). The router then swaps the FEC label, using the FEC-label bindings learned through normal LDP mechanisms. The router then pushes the topology-id label for the next-hop.

As with Option 1A, this forwarding mechanism also has the useful property that the FEC associated with the packet is maintained in the labels at each hop along the MRT.

This forwarding mechanism has minimal usage of additional labels, memory and LDP communication. It does increase the size of packets and the complexity of the required label operations and look-ups.

This forwarding option is consistent with context-specific label spaces, as described in [RFC5331]. However, the precise LDP behavior required to support this option for MRT has not been specified.

6.1.1.3. Compatibility of Option 1A and 1B

In principle, MRT transit forwarding mechanisms 1A and 1B can coexist in the same network, with a packet being forwarding along a single MRT path using the single label of option 1A for some hops and the two label stack of option 1B for other hops.

6.1.1.4. Mandatory support for MRT LDP Label option 1A

If a router supports a profile that includes the MRT LDP Label option for MRT transit forwarding mechanism, then it MUST support option 1A, which encodes topology-scoped FECs using a single label.

6.1.2. MRT IP tunnels (Options 2A and 2B)

IP tunneling can also be used as an MRT transit forwarding mechanism. Each router supporting this MRT transit forwarding mechanism announces two additional loopback addresses and their associated MRT color. Those addresses are used as destination addresses for MRT-blue and MRT-red IP tunnels respectively. The special loopback addresses allow the transit nodes to identify the traffic as being forwarded along either the MRT-blue or MRT-red topology to reach the tunnel destination. For example, an MRT ingress router can cause a packet to be tunneled along the MRT-red path to router X by encapsulating the packet using the MRT-red loopback address advertised by router X. Upon receiving the packet, router X would remove the encapsulation header and forward the packet based on the original destination address.

Either IPv4 (option 2A) or IPv6 (option 2B) can be used as the tunneling mechanism.

Note that the two forwarding mechanisms using LDP Label options do not require additional loopbacks per router, as is required by the IP tunneling mechanism. This is because LDP labels are used on a hop-by-hop basis to identify MRT-blue and MRT-red forwarding topologies.

6.2. Forwarding LDP Unicast Traffic over MRT Paths

In the previous section, we examined several options for providing MRT transit forwarding functionality, which is independent of the type of traffic being carried. We now look at the MRT ingress functionality, which will depend on the type of traffic being carried (IP or LDP). We start by considering LDP traffic.

We also simplify the initial discussion by assuming that the network consists of a single IGP area, and that all routers in the network participate in MRT. Other deployment scenarios that require MRT egress functionality are considered later in this document.

In principle, it is possible to carry LDP traffic in MRT IP tunnels. However, for LDP traffic, it is desirable to avoid tunneling. Tunneling LDP traffic to a remote node requires knowledge of remote FEC-label bindings so that the LDP traffic can continue to be forwarded properly when it leaves the tunnel. This requires targeted LDP sessions which can add management complexity. As described below, the two MRT forwarding mechanisms that use LDP labels do not require targeted LDP sessions.

6.2.1. Forwarding LDP traffic using MRT LDP Labels (Option 1A)

The MRT LDP Label option 1A forwarding mechanism uses topology-scoped FECs encoded using a single label as described in section Section 6.1.1.1. When a PLR receives an LDP packet that needs to be forwarded on the Red MRT (for example), it does a label swap operation, replacing the usual LDP label for the FEC with the Red MRT label for that FEC received from the next-hop router in the Red MRT computed by the PLR. When the next-hop router in the Red MRT receives the packet with the Red MRT label for the FEC, the MRT transit forwarding functionality continues as described in Section 6.1.1.1. In this way the original FEC associated with the packet is maintained at each hop along the MRT.

6.2.2. Forwarding LDP traffic using MRT LDP Labels (Option 1B)

The MRT LDP Label option 1B forwarding mechanism encodes the topology and the FEC using a two label stack as described in Section 6.1.1.2. When a PLR receives an LDP packet that needs to be forwarded on the Red MRT, it first does a normal LDP label swap operation, replacing the incoming normal LDP label associated with a given FEC with the outgoing normal LDP label for that FEC learned from the next-hop on the Red MRT. In addition, the PLR pushes the topology-identification label associated with the Red MRT, and forward the packet to the appropriate next-hop on the Red MRT. When the next-hop router in the Red MRT receives the packet with the Red MRT label for the FEC, the MRT transit forwarding functionality continues as described in Section 6.1.1.2. As with option 1A, the original FEC associated with the packet is maintained at each hop along the MRT.

6.2.3. Other considerations for forwarding LDP traffic using MRT LDP Labels

Note that forwarding LDP traffic using MRT LDP Labels can be done without the use of targeted LDP sessions when an MRT path to the destination FEC is used. The alternates selected in [I-D.ietf-rtgwg-mrt-frr-algorithm] use the MRT path to the destination FEC, so targeted LDP sessions are not needed. If instead one found it desirable to have the PLR use an MRT to reach the primary next-next-hop for the FEC, and then continue forwarding the LDP packet along the shortest path tree from the primary next-next-hop, this would require tunneling to the primary next-next-hop and a targeted LDP session for the PLR to learn the FEC-label binding for primary next-next-hop to correctly forward the packet.

For greatest hardware compatibility, routers implementing MRT fast-reroute of LDP traffic MUST support Option 1A of encoding the MT-ID in the labels (See Section 9).

6.3. Forwarding IP Unicast Traffic over MRT Paths

For IPv4 traffic, there is no currently practical alternative except tunneling to gain the bits needed to indicate the MRT-Blue or MRT-Red forwarding topology. For IPv6 traffic, in principle one could define bits in the IPv6 options header to indicate the MRT-Blue or MRT-Red forwarding topology. However, in this document, we have chosen not to define a solution that would work for IPv6 traffic but not for IPv4 traffic.

The choice of tunnel egress is flexible since any router closer to the destination than the next-hop can work. This architecture assumes that the original destination in the area is selected (see Section 11 for handling of multi-homed prefixes); another possible choice is the next-next-hop towards the destination. As discussed in the previous section, for LDP traffic, using the MRT to the original destination simplifies MRT-FRR by avoiding the need for targeted LDP sessions to the next-next-hop. For IP, that consideration doesn't apply.

Some situations require tunneling IP traffic along an MRT to a tunnel endpoint that is not the destination of the IP traffic. These situations will be discussed in detail later. We note here that an IP packet with a destination in a different IGP area/level from the PLR should be tunneled on the MRT to the ABR/LBR on the shortest path to the destination. For a destination outside of the PLR's MRT Island, the packet should be tunneled on the MRT to a non-proxy-node immediately before the named proxy-node on that particular color MRT.

6.3.1. Tunneling IP traffic using MRT LDP Labels

An IP packet can be tunneled along an MRT path by pushing the appropriate MRT LDP label(s). Tunneling using LDP labels, as opposed to IP headers, has the the advantage that more installed routers can do line-rate encapsulation and decapsulation using LDP than using IP. Also, no additional IP addresses would need to be allocated or signaled.

6.3.1.1. Tunneling IP traffic using MRT LDP Labels (Option 1A)

The MRT LDP Label option 1A forwarding mechanism uses topology-scoped FECs encoded using a single label as described in section Section 6.1.1.1. When a PLR receives an IP packet that needs to be forwarded on the Red MRT to a particular tunnel endpoint, it does a label push operation. The label pushed is the Red MRT label for a FEC originated by the tunnel endpoint, learned from the next-hop on the Red MRT.

6.3.1.2. Tunneling IP traffic using MRT LDP Labels (Option 1B)

The MRT LDP Label option 1B forwarding mechanism encodes the topology and the FEC using a two label stack as described in Section 6.1.1.2. When a PLR receives an IP packet that needs to be forwarded on the Red MRT to a particular tunnel endpoint, the PLR pushes two labels on the IP packet. The first (inner) label is the normal LDP label learned from the next-hop on the Red MRT, associated with a FEC originated by the tunnel endpoint. The second (outer) label is the topology-identification label associated with the Red MRT.

For completeness, we note here a potential variation that uses a single label as opposed to two labels. In order to tunnel an IP packet over an MRT to the destination of the IP packet (as opposed to an arbitrary tunnel endpoint), then we could just push a topology-identification label directly onto the packet. An MRT transit router would need to pop the topology-id label, do an IP route lookup in the context of that topology-id , and push the topology-id label.

6.3.2. Tunneling IP traffic using MRT IP Tunnels

In order to tunnel over the MRT to a particular tunnel endpoint, the PLR encapsulates the original IP packet with an additional IP header using the MRT-Blue or MRT-Red loopack address of the tunnel endpoint.

6.3.3. Required support

For greatest hardware compatibility and ease in removing the MRT-topology marking at area/level boundaries, routers that support MPLS and implement IP MRT fast-reroute MUST support tunneling of IP traffic using MRT LDP Labels Option 1A (topology-scoped FEC encoded using a single label).

7. MRT Island Formation

The purpose of communicating support for MRT is to indicate that the MRT-Blue and MRT-Red forwarding topologies are created for transit traffic. The MRT architecture allows for different, potentially incompatible options. In order to create consistent MRT forwarding topologies, the routers participating in a particular MRT Island need to use the same set of options. These options are grouped into MRT profiles. In addition, the routers in an MRT Island all need to use the same set of nodes and links within the Island when computing the MRT forwarding topologies. This section describes the information used by a router to determine the nodes and links to include in a particular MRT Island. Some information already exists in the IGPs and can be used by MRT in Island formation, subject to the interpretation defined here.

Other information needs to be communicated between routers for which there do not currently exist protocol extensions. This new information needs to be shared among all routers in an IGP area, so defining extensions to existing IGPs to carry this information makes sense. These new protocol extensions will be defined elsewhere.

Deployment scenarios using multi-topology OSPF or IS-IS, or running both ISIS and OSPF on the same routers is out of scope for this specification. As with LFA, it is expected that OSPF Virtual Links will not be supported.

7.1. IGP Area or Level

All links in an MRT Island MUST be bidirectional and belong to the same IGP area or level. For ISIS, a link belonging to both level 1 and level 2 would qualify to be in multiple MRT Islands. A given ABR or LBR can belong to multiple MRT Islands, corresponding to the areas or levels in which it participates. Inter-area forwarding behavior is discussed in Section 10.

7.2. Support for a specific MRT profile

All routers in an MRT Island MUST support the same MRT profile. A router advertises support for a given MRT profile using an 8-bit MRT Profile ID value. The registry for the MRT Profile ID is defined in this document. The protocol extensions for advertising the MRT Profile ID value will be defined elsewhere. A given router can support multiple MRT profiles and participate in multiple MRT Islands. The options that make up an MRT profile, as well as the default MRT profile, are defined in Section 8.

Note that a router may advertise support for multiple different MRT profiles. The process of MRT Island formation takes place independently for each MRT profile advertised by a given router. For example, consider a network with 40 connected routers in the same area advertising support for MRT Profile A and MRT Profile B. Two distinct MRT Islands will be formed corresponding to Profile A and Profile B, with each island containing all 40 routers. A complete set of maximally redundant trees will be computed for each island following the rules defined for each profile. If we add a third MRT Profile to this example, with Profile C being advertised by a connected subset of 30 routers, there will be a third MRT Island formed corresponding to those 30 routers, and a third set of maximally redundant trees will be computed. In this example, 40 routers would compute and install two sets of MRT transit forwarding entries corresponding to Profiles A and B, while 30 routers would compute and install three sets of MRT transit forwarding entries corresponding to Profiles A, B, and C.

7.3. Excluding additional routers and interfaces from the MRT Island

MRT takes into account existing IGP mechanisms for discouraging traffic from using particular links and routers, and it introduces an MRT-specific exclusion mechanism for links.

7.3.1. Existing IGP exclusion mechanisms

Mechanisms for discouraging traffic from using particular links already exist in ISIS and OSPF. In ISIS, an interface configured with a metric of 2^24-2 (0xFFFFFE) will only be used as a last resort. (An interface configured with a metric of 2^24-1 (0xFFFFFF) will not be advertised into the topology.) In OSPF, an interface configured with a metric of 2^16-1 (0xFFFF) will only be used as a last resort. These metrics can be configured manually to enforce administrative policy, or they can be set in an automated manner as with LDP IGP synchronization [RFC5443].

Mechanisms also already exist in ISIS and OSPF to discourage or prevent transit traffic from using a particular router. In ISIS, the overload bit is prevents transit traffic from using a router.

For OSPFv2 and OSPFv3, [RFC6987] specifies setting all outgoing interface metrics to 0xFFFF to discourage transit traffic from using a router.([RFC6987] defines the metric value 0xFFFF as MaxLinkMetric, a fixed architectural value for OSPF.) For OSPFv3, [RFC5340] specifies that a router be excluded from the intra-area shortest path tree computation if the V6-bit or R-bit of the LSA options is not set in the Router LSA.

The following rules for MRT Island formation ensure that MRT FRR protection traffic does not use a link or router that is discouraged or prevented from carrying traffic by existing IGP mechanisms.

1. A bidirectional link MUST be excluded from an MRT Island if either the forward or reverse cost on the link is 0xFFFFFE (for ISIS) or 0xFFFF for OSPF.

2. A router MUST be excluded from an MRT Island if it is advertised with the overload bit set (for ISIS), or it is advertised with metric values of 0xFFFF on all of its outgoing interfaces (for OSPFv2 and OSPFv3).

3. A router MUST be excluded from an MRT Island if it is advertised with either the V6-bit or R-bit of the LSA options not set in the Router LSA.

7.3.2. MRT-specific exclusion mechanism

This architecture also defines a means of excluding an otherwise usable link from MRT Islands. The protocol extensions for advertising that a link is MRT-Ineligible will be defined elsewhere. A link with either interface advertised as MRT-Ineligible MUST be excluded from an MRT Island. Note that an interface advertised as MRT-Ineligible by a router is ineligible with respect to all profiles advertised by that router.

7.4. Connectivity

All of the routers in an MRT Island MUST be connected by bidirectional links with other routers in the MRT Island. Disconnected MRT Islands will operate independently of one another.

7.5. Algorithm for MRT Island Identification

An algorithm that allows a computing router to identify the routers and links in the local MRT Island satisfying the above rules is given in section 5.2 of [I-D.ietf-rtgwg-mrt-frr-algorithm].

8. MRT Profile

An MRT Profile is a set of values and options related to MRT behavior. The complete set of options is designated by the corresponding 8-bit Profile ID value.

This document specifies the values and options that correspond to the Default MRT Profile (Profile ID = 0). Future documents may define other MRT Profiles by specifying the MRT Profile Options below.

8.1. MRT Profile Options

Below is a description of the values and options that define an MRT Profile.

MRT Algorithm: This identifies the particular algorithm for computing maximally redundant trees used by the router for this profile.

MRT-Red MT-ID: This specifies the MPLS MT-ID to be associated with the MRT-Red forwarding topology. It is allocated from the MPLS Multi-Topology Identifiers Registry.

MRT-Blue MT-ID: This specifies the MPLS MT-ID to be associated with the MRT-Blue forwarding topology. It is allocated from the MPLS Multi-Topology Identifiers Registry.

GADAG Root Selection Policy: This specifies the manner in which the GADAG root is selected. All routers in the MRT island need to use the same GADAG root in the calculations used construct the MRTs. A valid GADAG Root Selection Policy MUST be such that each router in the MRT island chooses the same GADAG root based on information available to all routers in the MRT island. GADAG Root Selection Priority values, advertised as router-specific MRT parameters, MAY be used in a GADAG Root Selection Policy.

MRT Forwarding Mechanism: This specifies which forwarding mechanism the router uses to carry transit traffic along MRT paths. A router which supports a specific MRT forwarding mechanism must program appropriate next-hops into the forwarding plane. The current options are MRT LDP Labels, IPv4 Tunneling, IPv6 Tunneling, and None. If the MRT LDP Labels option is supported, then option 1A and the appropriate signaling extensions MUST be supported. If IPv4 is supported, then both MRT-Red and MRT-Blue IPv4 Loopback Addresses SHOULD be specified. If IPv6 is supported, both MRT-Red and MRT-Blue IPv6 Loopback Addresses SHOULD be specified.

Recalculation: Recalculation specifies the process and timing by which new MRTs are computed after the topology has been modified.

Area/Level Border Behavior: This specifies how traffic traveling on the MRT-Blue or MRT-Red in one area should be treated when it passes into another area.

Other Profile-Specific Behavior: Depending upon the use-case for the profile, there may be additional profile-specific behavior.

When a new MRT Profile is defined, new and unique values should be allocated from the MPLS Multi-Topology Identifiers Registry, corresponding to the MRT-Red and MRT-Blue MT-ID values for the new MRT Profile .

If a router advertises support for multiple MRT profiles, then it MUST create the transit forwarding topologies for each of those, unless the profile specifies the None option for MRT Forwarding Mechanism.

The ability of MRT-FRR to support transit forwarding entries for multiple profiles can be used to facilitate a smooth transition from an existing deployed MRT Profile to a new MRT Profile. The new profile can be activated in parallel with the existing profile, installing the transit forwarding entries for the new profile without affecting the transit forwarding entries for the existing profile. Once the new transit forwarding state has been verified, the router

can be configured to use the alternates computed by the new profile in the event of a failure.

8.2. Router-specific MRT paramaters

For some profiles, additional router-specific MRT parameters may need to be advertised. While the set of options indicated by the MRT Profile ID must be identical for all routers in an MRT Island, these router-specific MRT parameters may differ between routers in the same MRT island. Several such parameters are described below.

GADAG Root Selection Priority: A GADAG Root Selection Policy MAY rely on the GADAG Root Selection Priority values advertised by each router in the MRT island. A GADAG Root Selection Policy may use the GADAG Root Selection Priority to allow network operators to configure a parameter to ensure that the GADAG root is selected from a particular subset of routers. An example of this use of the GADAG Root Selection Priority value by the GADAG Root Selection Policy is given in the Default MRT profile below.

MRT-Red Loopback Address: This provides the router's loopback address to reach the router via the MRT-Red forwarding topology. It can be specified for either IPv4 or IPv6. Note that this parameter is not needed to support the Default MRT profile.

MRT-Blue Loopback Address: This provides the router's loopback address to reach the router via the MRT-Blue forwarding topology. It can be specified for either IPv4 and IPv6. Note that this parameter is not needed to support the Default MRT profile.

Protocol extensions for advertising a router's GADAG Root Selection Priority value will be defined in other documents. Protocol extensions for the advertising a router's MRT-Red and MRT-Blue Loopback Addresses will be defined elsewhere.

8.3. Default MRT profile

The following set of options defines the default MRT Profile. The default MRT profile is indicated by the MRT Profile ID value of 0.

MRT Algorithm: MRT Lowpoint algorithm defined in [I-D.ietf-rtgwg-mrt-frr-algorithm].

MRT-Red MPLS MT-ID: This value will be allocated from the MPLS Multi-Topology Identifiers Registry. The IANA request for this allocation will be in another document.

MRT-Blue MPLS MT-ID: This value will be allocated from the MPLS Multi-Topology Identifiers Registry. The IANA request for this allocation will be in another document.

GADAG Root Selection Policy: Among the routers in the MRT Island and with the most preferred GADAG Root Selection Priority advertised (corresponding to the lowest numerical value of GADAG Root Selection Priority), an implementation MUST pick the router with the highest Router ID to be the GADAG root.

Forwarding Mechanisms: MRT LDP Labels

Recalculation: Recalculation of MRTs SHOULD occur as described in Section 12.2. This allows the MRT forwarding topologies to support IP/LDP fast-reroute traffic.

Area/Level Border Behavior: As described in Section 10, ABRs/LBRs SHOULD ensure that traffic leaving the area also exits the MRT-Red or MRT-Blue forwarding topology.

9. LDP signaling extensions and considerations

The protocol extensions for LDP will be defined in another document. A router must indicate that it has the ability to support MRT; having this explicit allows the use of MRT-specific processing, such as special handling of FECs sent with the Rainbow MRT MT-ID.

A FEC sent with the Rainbow MRT MT-ID indicates that the FEC applies to all the MRT-Blue and MRT-Red MT-IDs in supported MRT profiles. The FEC-label bindings for the default shortest-path based MT-ID 0 MUST still be sent (even though it could be inferred from the Rainbow FEC-label bindings) to ensure continuous operation of normal LDP forwarding. The Rainbow MRT MT-ID is defined to provide an easy way to handle the special signaling that is needed at ABRs or LBRs. It avoids the problem of needing to signal different MPLS labels to different LDP neighbors for the same FEC. Because the Rainbow MRT MT-ID is used only by ABRs/LBRs or an LDP egress router, it is not MRT profile specific.

The value of the Rainbow MRT MPLS MT-ID will be allocated from the MPLS Multi-Topology Identifiers Registry. The IANA request for this allocation will be in another document.

10. Inter-area Forwarding Behavior

An ABR/LBR has two forwarding roles. First, it forwards traffic within areas. Second, it forwards traffic from one area into another. These same two roles apply for MRT transit traffic.

Traffic on MRT-Red or MRT-Blue destined inside the area needs to stay on MRT-Red or MRT-Blue in that area. However, it is desirable for traffic leaving the area to also exit MRT-Red or MRT-Blue and return to shortest path forwarding.

For unicast MRT-FRR, the need to stay on an MRT forwarding topology terminates at the ABR/LBR whose best route is via a different area/ level. It is highly desirable to go back to the default forwarding topology when leaving an area/level. There are three basic reasons for this. First, the default topology uses shortest paths; the packet will thus take the shortest possible route to the destination. Second, this allows failures that might appear in multiple areas (e.g. ABR/LBR failures) to be separately identified and repaired around. Third, the packet can be fast-rerouted again, if necessary, due to a failure in a different area.

An ABR/LBR that receives a packet on MRT-Red or MRT-Blue towards destination Z should continue to forward the packet along MRT-Red or MRT-Blue only if the best route to Z is in the same area as the interface that the packet was received on. Otherwise, the packet should be removed from MRT-Red or MRT-Blue and forwarded on the shortest-path default forwarding topology.

To avoid per-interface forwarding state for MRT-Red and MRT-Blue, the ABR/LBR needs to arrange that packets destined to a different area arrive at the ABR/LBR already not marked as MRT-Red or MRT-Blue.

10.1. ABR Forwarding Behavior with MRT LDP Label Option 1A

For LDP forwarding where a single label specifies (MT-ID, FEC), the ABR/LBR is responsible for advertising the proper label to each neighbor. Assume that an ABR/LBR has allocated three labels for a particular destination; those labels are L_primary, L_blue, and L_red. To those routers in the same area as the best route to the destination, the ABR/LBR advertises the following FEC-label bindings: L_primary for the default topology, L_blue for the MRT-Blue MT-ID and L_red for the MRT-Red MT-ID, as expected. However, to routers in other areas, the ABR/LBR advertises the following FEC-label bindings: L_primary for the default topology, and L_primary for the Rainbow MRT MT-ID. Associating L_primary with the Rainbow MRT MT-ID causes the receiving routers to use L_primary for the MRT-Blue MT-ID and for the MRT-Red MT-ID.

The ABR/LBR installs all next-hops for the best area: primary next-hops for L_primary, MRT-Blue next-hops for L_blue, and MRT-Red next-hops for L_red. Because the ABR/LBR advertised (Rainbow MRT MT-ID, FEC) with L_primary to neighbors not in the best area, packets from those neighbors will arrive at the ABR/LBR with a label L_primary and

will be forwarded into the best area along the default topology. By controlling what labels are advertised, the ABR/LBR can thus enforce that packets exiting the area do so on the shortest-path default topology.

10.1.1. Motivation for Creating the Rainbow-FEC

The desired forwarding behavior could be achieved in the above example without using the Rainbow-FEC. This could be done by having the ABR/LBR advertise the following FEC-label bindings to neighbors not in the best area: L1_primary for the default topology, L1_primary for the MRT-Blue MT-ID, and L1_primary for the MRT-Red MT-ID. Doing this would require machinery to spoof the labels used in FEC-label binding advertisements on a per-neighbor basis. Such label-spoofing machinery does not currently exist in most LDP implmentations and doesn't have other obvious uses.

Many existing LDP implmentations do however have the ability to filter FEC-label binding advertisements on a per-neighbor basis. The Rainbow-FEC allows us to re-use the existing per-neighbor FEC filtering machinery to achieve the desired result. By introducing the Rainbow FEC, we can use per-neighbor FEC-filtering machinery to advertise the FEC-label binding for the Rainbow-FEC (and filter those for MRT-Blue and MRT-Red) to non-best-area neighbors of the ABR.

The use of the Rainbow-FEC by the ABR for non-best-area advertisements is RECOMMENDED. An ABR MAY advertise the label for the default topology in separate MRT-Blue and MRT-Red advertisements. However, a router that supports the LDP Label MRT Forwarding Mechanism MUST be able to receive and correctly interpret the Rainbow-FEC.

10.2. ABR Forwarding Behavior with IP Tunneling (option 2)

If IP tunneling is used, then the ABR/LBR behavior is dependent upon the outermost IP address. If the outermost IP address is an MRT loopback address of the ABR/LBR, then the packet is decapsulated and forwarded based upon the inner IP address, which should go on the default SPT topology. If the outermost IP address is not an MRT loopback address of the ABR/LBR, then the packet is simply forwarded along the associated forwarding topology. A PLR sending traffic to a destination outside its local area/level will pick the MRT and use the associated MRT loopback address of the selected ABR/LBR advertising the lowest cost to the external destination.

Thus, for these two MRT Forwarding Mechanisms (MRT LDP Label option 1A and IP tunneling option 2), there is no need for additional computation or per-area forwarding state.

10.3. ABR Forwarding Behavior with LDP Label option 1B

The other MRT forwarding mechanism described in Section 6 uses two labels, a topology-id label, and a FEC-label. This mechanism would require that any router whose MRT-Red or MRT-Blue next-hop is an ABR/ LBR would need to determine whether the ABR/LBR would forward the packet out of the area/level. If so, then that router should pop off the topology-identification label before forwarding the packet to the ABR/LBR.

For example, in Figure 4, if node H fails, node E has to put traffic towards prefix p onto MRT-Red. But since node D knows that ABR1 will use a best route from another area, it is safe for D to pop the Topology-Identification Label and just forward the packet to ABR1 along the MRT-Red next-hop. ABR1 will use the shortest path in Area 10.

In all cases for ISIS and most cases for OSPF, the penultimate router can determine what decision the adjacent ABR will make. The one case where it can't be determined is when two ASBRs are in different non-backbone areas attached to the same ABR, then the ASBR's Area ID may be needed for tie-breaking (prefer the route with the largest OPSF area ID) and the Area ID isn't announced as part of the ASBR link-state advertisement (LSA). In this one case, suboptimal forwarding along the MRT in the other area would happen. If that becomes a realistic deployment scenario, protocol extensions could be developed to address this issue.

```
       +----[C]----     --[D]--[E]                --[D]--[E]
       |           \   /         \               /         \
   p--[A] Area 10 [ABR1]  Area 0 [H]--p   +-[ABR1]  Area 0 [H]-+
       |           /   \         /        |       \         /   |
       +----[B]----     --[F]--[G]        |        --[F]--[G]   |
                                          |                     |
                                          | other               |
                                          +----------[p]-------+
                                            area

         (a) Example topology        (b) Proxy node view in Area 0 nodes


                   +----[C]<---       [D]->[E]
                   V           \             \
                +-[A] Area 10 [ABR1]  Area 0 [H]-+
                |  ^           /             /   |
                |  +----[B]<---       [F]->[G]   V
                |                                |
                +------------->[p]<--------------+

                  (c) rSPT towards destination p



             ->[D]->[E]                         -<[D]<-[E]
            /          \                       /         \
       [ABR1]  Area 0 [H]-+             +-[ABR1]         [H]
                      /   |             |      \
               [F]->[G]   V             V       -<[F]<-[G]
                          |             |
                          |             |
                [p]<------+             +--------->[p]

     (d) Blue MRT in Area 0           (e) Red MRT in Area 0
```

Figure 4: ABR Forwarding Behavior and MRTs

11. Prefixes Multiply Attached to the MRT Island

How a computing router S determines its local MRT Island for each supported MRT profile is already discussed in Section 7.

There are two types of prefixes or FECs that may be multiply attached to an MRT Island. The first type are multi-homed prefixes that usually connect at a domain or protocol boundary. The second type represent routers that do not support the profile for the MRT Island.

The key difference is whether the traffic, once out of the MRT Island, might re-enter the MRT Island if a loop-free exit point is not selected.

FRR using LFA has the useful property that it is able to protect multi-homed prefixes against ABR failure. For instance, if a prefix from the backbone is available via both ABR A and ABR B, if A fails, then the traffic should be redirected to B. This can be accomplished with MRT FRR as well.

If ASBR protection is desired, this has additional complexities if the ASBRs are in different areas. Similarly, protecting labeled BGP traffic in the event of an ASBR failure has additional complexities due to the per-ASBR label spaces involved.

As discussed in [RFC5286], a multi-homed prefix could be:

- An out-of-area prefix announced by more than one ABR,

- An AS-External route announced by 2 or more ASBRs,

- A prefix with iBGP multipath to different ASBRs,

- etc.

See Appendix A for a discussion of a general issue with multi-homed prefixes connected in two different areas.

There are also two different approaches to protection. The first is tunnel endpoint selection where the PLR picks a router to tunnel to where that router is loop-free with respect to the failure-point. Conceptually, the set of candidate routers to provide LFAs expands to all routers that can be reached via an MRT alternate, attached to the prefix.

The second is to use a proxy-node, that can be named via MPLS label or IP address, and pick the appropriate label or IP address to reach it on either MRT-Blue or MRT-Red as appropriate to avoid the failure point. A proxy-node can represent a destination prefix that can be attached to the MRT Island via at least two routers. It is termed a named proxy-node if there is a way that traffic can be encapsulated to reach specifically that proxy-node; this could be because there is an LDP FEC for the associated prefix or because MRT-Red and MRT-Blue IP addresses are advertised (in an as-yet undefined fashion) for that proxy-node. Traffic to a named proxy-node may take a different path than traffic to the attaching router; traffic is also explicitly forwarded from the attaching router along a predetermined interface towards the relevant prefixes.

For IP traffic, multi-homed prefixes can use tunnel endpoint selection. For IP traffic that is destined to a router outside the MRT Island, if that router is the egress for a FEC advertised into the MRT Island, then the named proxy-node approach can be used.

For LDP traffic, there is always a FEC advertised into the MRT Island. The named proxy-node approach should be used, unless the computing router S knows the label for the FEC at the selected tunnel endpoint.

If a FEC is advertised from outside the MRT Island into the MRT Island and the forwarding mechanism specified in the profile includes LDP, then the routers learning that FEC MUST also advertise labels for (MRT-Red, FEC) and (MRT-Blue, FEC) to neighbors inside the MRT Island. Any router receiving a FEC corresponding to a router outside the MRT Island or to a multi-homed prefix MUST compute and install the transit MRT-Blue and MRT-Red next-hops for that FEC. The FEC-label bindings for the topology-scoped FECs ((MT-ID 0, FEC), (MRT-Red, FEC), and (MRT-Blue, FEC)) MUST also be provided via LDP to neighbors inside the MRT Island.

11.1. Protecting Multi-Homed Prefixes using Tunnel Endpoint Selection

Tunnel endpoint selection is a local matter for a router in the MRT Island since it pertains to selecting and using an alternate and does not affect the transit MRT-Red and MRT-Blue forwarding topologies.

Let the computing router be S and the next-hop F be the node whose failure is to be avoided. Let the destination be prefix p. Have A be the router to which the prefix p is attached for S's shortest path to p.

The candidates for tunnel endpoint selection are those to which the destination prefix is attached in the area/level. For a particular candidate B, it is necessary to determine if B is loop-free to reach p with respect to S and F for node-protection or at least with respect to S and the link (S, F) for link-protection. If B will always prefer to send traffic to p via a different area/level, then this is definitional. Otherwise, distance-based computations are necessary and an SPF from B's perspective may be necessary. The following equations give the checks needed; the rationale is similar to that given in [RFC5286].

Loop-Free for S: D_opt(B, p) < D_opt(B, S) + D_opt(S, p)

Loop-Free for F: D_opt(B, p) < D_opt(B, F) + D_opt(F, p)

The latter is equivalent to the following, which avoids the need to compute the shortest path from F to p.

Loop-Free for F: D_opt(B, p) < D_opt(B, F) + D_opt(S, p) - D_opt(S, F)

Finally, the rules for Endpoint selection are given below. The basic idea is to repair to the prefix-advertising router selected for the shortest-path and only to select and tunnel to a different endpoint if necessary (e.g. A=F or F is a cut-vertex or the link (S,F) is a cut-link).

1. Does S have a node-protecting alternate to A? If so, select that. Tunnel the packet to A along that alternate. For example, if LDP is the forwarding mechanism, then push the label (MRT-Red, A) or (MRT-Blue, A) onto the packet.

2. If not, then is there a router B that is loop-free to reach p while avoiding both F and S? If so, select B as the end-point. Determine the MRT alternate to reach B while avoiding F. Tunnel the packet to B along that alternate. For example, with LDP, push the label (MRT-Red, B) or (MRT-Blue, B) onto the packet.

3. If not, then does S have a link-protecting alternate to A? If so, select that.

4. If not, then is there a router B that is loop-free to reach p while avoiding S and the link from S to F? If so, select B as the endpoint and the MRT alternate for reaching B from S that avoid the link (S,F).

The tunnel endpoint selected will receive a packet destined to itself and, being the egress, will pop that MPLS label (or have signaled Implicit Null) and forward based on what is underneath. This suffices for IP traffic since the tunnel endpoint can use the IP header of the original packet to continue forwarding the packet. However, tunnelling of LDP traffic requires targeted LDP sessions for learning the FEC-label binding at the tunnel endpoint.

11.2. Protecting Multi-Homed Prefixes using Named Proxy-Nodes

Instead, the named proxy-node method works with LDP traffic without the need for targeted LDP sessions. It also has a clear advantage over tunnel endpoint selection, in that it is possible to explicitly forward from the MRT Island along an interface to a loop-free island neighbor when that interface may not be a primary next-hop.

A named proxy-node represents one or more destinations and, for LDP forwarding, has a FEC associated with it that is signalled into the MRT Island. Therefore, it is possible to explicitly label packets to go to (MRT-Red, FEC) or (MRT-Blue, FEC); at the border of the MRT Island, the label will swap to meaning (MT-ID 0, FEC). It would be possible to have named proxy-nodes for IP forwarding, but this would require extensions to signal two IP addresses to be associated with MRT-Red and MRT-Blue for the proxy-node. A named proxy-node can be uniquely represented by the two routers in the MRT Island to which it is connected. The extensions to signal such IP addresses will be defined elsewhere. The details of what label-bindings must be originated will be described in another document.

Computing the MRT next-hops to a named proxy-node and the MRT alternate for the computing router S to avoid a particular failure node F is straightforward. The details of the simple constant-time functions, Select_Proxy_Node_NHs() and Select_Alternates_Proxy_Node(), are given in [I-D.ietf-rtgwg-mrt-frr-algorithm]. A key point is that computing these MRT next-hops and alternates can be done as new named proxy-nodes are added or removed without requiring a new MRT computation or impacting other existing MRT paths. This maps very well to, for example, how OSPFv2 (see [RFC2328] Section 16.5) does incremental updates for new summary-LSAs.

The remaining question is how to attach the named proxy-node to the MRT Island; all the routers in the MRT Island MUST do this consistently. No more than 2 routers in the MRT Island can be selected; one should only be selected if there are no others that meet the necessary criteria. The named proxy-node is logically part of the area/level.

There are two sources for candidate routers in the MRT Island to connect to the named proxy-node. The first set are those routers in the MRT Island that are advertising the prefix; the named-proxy-cost assigned to each prefix-advertising router is the announced cost to the prefix. The second set are those routers in the MRT Island that are connected to routers not in the MRT Island but in the same area/level; such routers will be defined as Island Border Routers (IBRs). The routers connected to the IBRs that are not in the MRT Island and are in the same area/level as the MRT island are Island Neighbors(INs).

Since packets sent to the named proxy-node along MRT-Red or MRT-Blue may come from any router inside the MRT Island, it is necessary that whatever router to which an IBR forwards the packet be loop-free with respect to the whole MRT Island for the destination. Thus, an IBR is a candidate router only if it possesses at least one IN whose

shortest path to the prefix does not enter the MRT Island. A method for identifying loop-free Island Neighbors(LFINs) is given in [I-D.ietf-rtgwg-mrt-frr-algorithm]. The named-proxy-cost assigned to each (IBR, IN) pair is cost(IBR, IN) + D_opt(IN, prefix).

From the set of prefix-advertising routers and the set of IBRs with at least one LFIN, the two routers with the lowest named-proxy-cost are selected. Ties are broken based upon the lowest Router ID. For ease of discussion, the two selected routers will be referred to as proxy-node attachment routers.

A proxy-node attachment router has a special forwarding role. When a packet is received destined to (MRT-Red, prefix) or (MRT-Blue, prefix), if the proxy-node attachment router is an IBR, it MUST swap to the shortest path forwarding topology (e.g. swap to the label for (MT-ID 0, prefix) or remove the outer IP encapsulation) and forward the packet to the IN whose cost was used in the selection. If the proxy-node attachment router is not an IBR, then the packet MUST be removed from the MRT forwarding topology and sent along the interface(s) that caused the router to advertise the prefix; this interface might be out of the area/level/AS.

11.3. MRT Alternates for Destinations Outside the MRT Island

A natural concern with new functionality is how to have it be useful when it is not deployed across an entire IGP area. In the case of MRT FRR, where it provides alternates when appropriate LFAs aren't available, there are also deployment scenarios where it may make sense to only enable some routers in an area with MRT FRR. A simple example of such a scenario would be a ring of 6 or more routers that is connected via two routers to the rest of the area.

Destinations inside the local island can obviously use MRT alternates. Destinations outside the local island can be treated like a multi-homed prefix and either Endpoint Selection or Named Proxy-Nodes can be used. Named Proxy-Nodes MUST be supported when LDP forwarding is supported and a label-binding for the destination is sent to an IBR.

Naturally, there are more complicated options to improve coverage, such as connecting multiple MRT islands across tunnels, but the need for the additional complexity has not been justified.

12. Network Convergence and Preparing for the Next Failure

After a failure, MRT detours ensure that packets reach their intended destination while the IGP has not reconverged onto the new topology. As link-state updates reach the routers, the IGP process calculates

the new shortest paths. Two things need attention: micro-loop prevention and MRT re-calculation.

12.1. Micro-loop prevention and MRTs

A micro-loop is a transient packet forwarding loop among two or more routers that can occur during convergence of IGP forwarding state. [RFC5715] discusses several techniques for preventing micro-loops. This section discusses how MRT-FRR relates to two of the micro-loop prevention techniques discussed in [RFC5715], Nearside Tunneling and Farside Tunneling.

In Nearside Tunneling, a router (PLR) adjacent to a failure perform local repair and inform remote routers of the failure. The remote routers initially tunnel affected traffic to the nearest PLR, using tunnels which are unaffected by the failure. Once the forwarding state for normal shortest path routing has converged, the remote routers return the traffic to shortest path forwarding. MRT-FRR is relevant for Nearside Tunneling for the following reason. The process of tunneling traffic to the PLRs and waiting a sufficient amount of time for IGP forwarding state convergence with Nearside Tunneling means that traffic will generally be relying on the local repair at the PLR for longer than it would in the absence of Nearside Tunneling. Since MRT-FRR provides 100% coverage for single link and node failure, it may be an attractive option to provide the local repair paths when Nearside Tunneling is deployed.

MRT-FRR is also relevant for the Farside Tunneling micro-loop prevention technique. In Farside Tunneling, remote routers tunnel traffic affected by a failure to a node downstream of the failure with respect to traffic destination. This node can be viewed as being on the farside of the failure with respect to the node initiating the tunnel. Note that the discussion of Farside Tunneling in [RFC5715] focuses on the case where the farside node is immediately adjacent to a failed link or node. However, the farside node may be any node downstream of the failure with respect to traffic destination, including the destination itself. The tunneling mechanism used to reach the farside node must be unaffected by the failure. The alternative forwarding paths created by MRT-FRR have the potential to be used to forward traffic from the remote routers upstream of the failure all the way to the destination. In the event of failure, either the MRT-Red or MRT-Blue path from the remote upstream router to the destination is guaranteed to avoid a link failure or inferred node failure. The MRT forwarding paths are also guaranteed to not be subject to micro-loops because they are locked to the topology before the failure.

We note that the computations in [I-D.ietf-rtgwg-mrt-frr-algorithm] address the case of a PLR adjacent to a failure determining which choice of MRT-Red or MRT-Blue will avoid a failed link or node. More computation may be required for an arbitrary remote upstream router to determine whether to choose MRT-Red or MRT-Blue for a given destination and failure.

12.2. MRT Recalculation for the Default MRT Profile

This section describes how the MRT recalculation SHOULD be performed for the Default MRT Profile. This is intended to support FRR applications. Other approaches are possible, but they are not specified in this document.

When a failure event happens, traffic is put by the PLRs onto the MRT topologies. After that, each router recomputes its shortest path tree (SPT) and moves traffic over to that. Only after all the PLRs have switched to using their SPTs and traffic has drained from the MRT topologies should each router install the recomputed MRTs into the FIBs.

At each router, therefore, the sequence is as follows:

1. Receive failure notification

2. Recompute SPT.

3. Install the new SPT in the FIB.

4. If the network was stable before the failure occured, wait a configured (or advertised) period for all routers to be using their SPTs and traffic to drain from the MRTs.

5. Recompute MRTs.

6. Install new MRTs in the FIB.

While the recomputed MRTs are not installed in the FIB, protection coverage is lowered. Therefore, it is important to recalculate the MRTs and install them quickly.

New protocol extensions for advertising the time needed to recompute shortest path routes and install them in the FIB will be defined elsewhere.

13. Implementation Status

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

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC6982]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC6982], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

Juniper Networks Implementation

- Organization responsible for the implementation: Juniper Networks

- Implementation name: MRT-FRR

- Implementation description: MRT-FRR using OSPF as the IGP has been implemented and verified.

- The implementation's level of maturity: prototype

- Protocol coverage: This implementation of the MRT-FRR includes Island identification, GADAG root selection, MRT Lowpoint algorithm, augmentation of GADAG with additional links, and calculation of MRT transit next-hops alternate next-hops based on draft "draft-ietf-rtgwg-mrt-frr-algorithm-00". This implementation also includes the M-bit in OSPF based on "draft-atlas-ospf-mrt-01" as well as LDP MRT Capability based on "draft-atlas-mpls-ldp-mrt-00".

- Licensing: proprietary

- Implementation experience: Implementation was useful for verifying functionality and lack of gaps. It has also been useful for improving aspects of the algorithm.

- Contact information: akatlas@juniper.net, shraddha@juniper.net, kishoret@juniper.net

Huawei Technology Implementation

- Organization responsible for the implementation: Huawei Technology Co., Ltd.

- Implementation name: MRT-FRR and IS-IS extensions for MRT.

- Implementation description: The MRT-FRR using IS-IS extensions for MRT and LDP multi-topology have been implemented and verified.

- The implementation's level of maturity: prototype

- Protocol coverage: This implementation of the MRT algorithm includes Island identification, GADAG root selection, MRT Lowpoint algorithm, augmentation of GADAG with additional links, and calculation of MRT transit next-hops alternate next-hops based on draft "draft-enyedi-rtgwg-mrt-frr-algorithm-03". This implementation also includes IS-IS extension for MRT based on "draft-li-mrt-00".

- Licensing: proprietary

- Implementation experience: It is important produce a second implementation to verify the algorithm is implemented correctly without looping. It is important to verify the ISIS extensions work for MRT-FRR.

- Contact information: lizhenbin@huawei.com, eric.wu@huawei.com

14. Operational Considerations

The following aspects of MRT-FRR are useful to consider when deploying the technology in different operational environments and network topologies.

14.1. Verifying Forwarding on MRT Paths

The forwarding paths created by MRT-FRR are not used by normal (non-FRR) traffic. They are only used to carry FRR traffic for a short period of time after a failure has been detected. It is RECOMMENDED that an operator proactively monitor the MRT forwarding paths in

order to be certain that the paths will be able to carry FRR traffic when needed. Therefore, an implementation SHOULD provide an operator with the ability to test MRT paths with OAM traffic. For example, when MRT paths are realized using LDP labels distributed for topology-scoped FECs, an implementation can use the MPLS ping and traceroute as defined in [RFC4379] and extended in [RFC7307] for topology-scoped FECs.

14.2. Traffic Capacity on Backup Paths

During a fast-reroute event initiated by a PLR in response to a network failure, the flow of traffic in the network will generally not be identical to the flow of traffic after the IGP forwarding state has converged, taking the failure into account. Therefore, even if a network has been engineered to have enough capacity on the appropriate links to carry all traffic after the IGP has converged after the failure, the network may still not have enough capacity on the appropriate links to carry the flow of traffic during a fast-reroute event. This can result in more traffic loss during the fast-reroute event than might otherwise be expected.

Note that there are two somewhat distinct aspects to this phenomenon. The first is that the path from the PLR to the destination during the fast-reroute event may be different from the path after the IGP converges. In this case, any traffic for the destination that reaches the PLR during the fast-reroute event will follow a different path from the PLR to the destination than will be followed after IGP convergence.

The second aspect is that the amount of traffic arriving at the PLR for affected destinations during the fast-reroute event may be larger than the amount of traffic arriving at the PLR for affected destinations after IGP convergence. Immediately after a failure, any non-PLR routers that were sending traffic to the PLR before the failure will continue sending traffic to the PLR, and that traffic will be carried over backup paths from the PLR to the destinations. After IGP convergence, upstream non-PLR routers may direct some traffic away from the PLR.

In order to reduce or eliminate the potential for transient traffic loss due to inadequate capacity during fast-reroute events, an operator can model the amount of traffic taking different paths during a fast-reroute event. If it is determined that there is not enough capacity to support a given fast-reroute event, the operator can address the issue either by augmenting capacity on certain links or modifying the backup paths themselves.

The MRT Lowpoint algorithm produces a pair of diverse paths to each destination. These paths are generated by following the directed links on a common GADAG. MRT-FRR allows an operator to exclude a link from the MRT Island, and thus the GADAG, by advertising it as MRT-Ineligible. Such a link will not be used on the MRT forwarding path for any destination. Advertising links as MRT-Ineligible is the main tool provided by MRT-FRR for keeping backup traffic off of lower bandwidth links during fast-reroute events.

Note that all of the backup paths produced by the MRT Lowpoint algorithm are closely tied to the common GADAG computed as part of that algorithm. Therefore, it is generally not possible to modify a subset of paths without affecting other paths. This precludes more fine-grained modification of individual backup paths when using only paths computed by the MRT Lowpoint algorithm.

However, it may be desirable to allow an operator to use MRT-FRR alternates together with alternates provided by other FRR technologies. A policy-based alternate selection process can allow an operator to select the best alternate from those provided by MRT and other FRR technologies. As a concrete example, it may be desirable to implement a policy where a downstream LFA (if it exists for a given failure mode and destination) is preferred over a given MRT alternate. This combination gives the operator the ability to affect where traffic flows during a fast-reroute event, while still producing backup paths that use no additional labels for LDP traffic and will not loop under multiple failures. This and other choices of alternate selection policy can be evaluated in the context of their effect on fast-reroute traffic flow and available capacity, as well as other deployment considerations.

Note that future documents may define MRT profiles in addition to the default profile defined here. Different MRT profiles will generally produce alternate paths with different properties. An implementation may allow an operator to use different MRT profiles instead of or in addition to the default profile.

14.3. MRT IP Tunnel Loopback Address Management

As described in Section 6.1.2, if an implementation uses IP tunneling as the mechanism to realize MRT forwarding paths, each node must advertise an MRT-Red and an MRT-Blue loopback address. These IP addresses must be unique within the routing domain to the extent that they do not overlap with each other or with any other routing table entries. It is expected that operators will use existing tools and processes for managing infrastructure IP addresses to manage these additional MRT-related loopback addresses.

14.4. MRT-FRR in a Network with Degraded Connectivity

Ideally, routers is a service provider network using MRT-FRR will be initially deployed in a 2-connected topology, allowing MRT-FRR to find completely diverse paths to all destinations. However, a network can differ from an ideal 2-connected topology for many possible reasons, including network failures and planned maintenance events.

MRT-FRR is designed to continue to function properly when network connectivity is degraded. When a network contains cut-vertices or cut-links dividing the network into different 2-connected blocks, MRT-FRR will continue to provide completely diverse paths for destinations within the same block as the PLR. For a destination in a different block from the PLR, the redundant paths created by MRT-FRR will be link and node diverse within each block, and the paths will only share links and nodes that are cut-links or cut-vertices in the topology.

If a network becomes partitioned with one set of routers having no connectivity to another set of routers, MRT-FRR will function independently in each set of connected routers, providing redundant paths to destinations in same set of connected routers as a given PLR.

14.5. Partial Deployment of MRT-FRR in a Network

A network operator may choose to deploy MRT-FRR only on a subset of routers in an IGP area. MRT-FRR is designed to accommodate this partial deployment scenario. Only routers that advertise support for a given MRT profile will be included in a given MRT Island. For a PLR within the MRT Island, MRT-FRR will create redundant forwarding paths to all destinations with the MRT Island using maximally redundant trees all the way to those destinations. For destinations outside of the MRT Island, MRT-FRR creates paths to the destination which use forwarding state created by MRT-FRR within the MRT Island and shortest path forwarding state outside of the MRT Island. The paths created by MRT-FRR to non-Island destinations are guaranteed to be diverse within the MRT Island (if topologically possible). However, the part of the paths outside of the MRT Island may not be diverse.

15. Acknowledgements

The authors would like to thank Mike Shand for his valuable review and contributions.

The authors would like to thank Joel Halpern, Hannes Gredler, Ted Qian, Kishore Tiruveedhula, Shraddha Hegde, Santosh Esale, Nitin Bahadur, Harish Sitaraman, Raveendra Torvi, Anil Kumar SN, Bruno Decraene, Eric Wu, Janos Farkas, Rob Shakir, and Stewart Bryant for their suggestions and review.

16. IANA Considerations

IANA is requested to create a registry entitled "MRT Profile Identifier Registry". The range is 0 to 255. The Default MRT Profile defined in this document has value 0. Values 1-200 are allocated by Standards Action. Values 201-220 are for experimentation. Values 221-255 are for vendor private use.

17. Security Considerations

In general, MRT forwarding paths do not follow shortest paths. The transit forwarding state corresponding to the MRT paths is created during normal operations (before a failure occurs). Therefore, a malicious packet with an appropriate header injected into the network from a compromised location would be forwarded to a destination along a non-shortest path. When this technology is deployed, a network security design should not rely on assumptions about potentially malicious traffic only following shortest paths.

It should be noted that the creation of non-shortest forwarding paths is not unique to MRT.

MRT-FRR requires that routers advertise information used in the formation of MRT backup paths. While this document does not specify the protocol extensions used to advertise this information, we discuss security considerations related to the information itself. Injecting false MRT-related information could be used to direct some MRT backup paths over compromised transmission links. Combined with the ability to generate network failures, this could be used to send traffic over compromised transmission links during a fast-reroute event. In order to prevent this potential exploit, a receiving router needs to be able to authenticate MRT-related information that claims to have been advertised by another router.

18. Contributors

Robert Kebler
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
USA
Email: rkebler@juniper.net

Andras Csaszar
Ericsson
Konyves Kalman krt 11
Budapest 1097
Hungary
Email: Andras.Csaszar@ericsson.com

Jeff Tantsura
Ericsson
300 Holger Way
San Jose, CA 95134
USA
Email: jeff.tantsura@ericsson.com

Russ White
VCE
Email: russw@riw.us

19. References

19.1. Normative References

[I-D.ietf-rtgwg-mrt-frr-algorithm]
Envedi, G., Csaszar, A., Atlas, A., Bowers, C., and A. Gopalan, "Algorithms for computing Maximally Redundant Trees for IP/LDP Fast- Reroute", draft-ietf-rtgwg-mrt-frr-algorithm-06 (work in progress), October 2015.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC5286] Atlas, A., Ed. and A. Zinin, Ed., "Basic Specification for IP Fast Reroute: Loop-Free Alternates", RFC 5286, DOI 10.17487/RFC5286, September 2008, <http://www.rfc-editor.org/info/rfc5286>.

19.2. Informative References

[EnyediThesis]
Enyedi, G., "Novel Algorithms for IP Fast Reroute", Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics Ph.D. Thesis, February 2011, <http://timon.tmit.bme.hu/theses/thesis_book.pdf>.

[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.

[I-D.francois-rtgwg-segment-routing-ti-lfa]
Francois, P., Filsfils, C., Bashandy, A., and B. Decraene, "Topology Independent Fast Reroute using Segment Routing", draft-francois-rtgwg-segment-routing-ti-lfa-00 (work in progress), August 2015.

[I-D.ietf-rtgwg-lfa-manageability]
Litkowski, S., Decraene, B., Filsfils, C., Raza, K., Horneffer, M., and P. Sarkar, "Operational management of Loop Free Alternates", draft-ietf-rtgwg-lfa-manageability-11 (work in progress), June 2015.

[I-D.ietf-rtgwg-rlfa-node-protection]
Sarkar, P., Hegde, S., Bowers, C., Gredler, H., and S. Litkowski, "Remote-LFA Node Protection and Manageability", draft-ietf-rtgwg-rlfa-node-protection-05 (work in progress), December 2015.

[LightweightNotVia]
Enyedi, G., Retvari, G., Szilagyi, P., and A. Csaszar, "IP Fast ReRoute: Lightweight Not-Via without Additional Addresses", Proceedings of IEEE INFOCOM , 2009, <http://mycite.omikk.bme.hu/doc/71691.pdf>.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <http://www.rfc-editor.org/info/rfc2328>.

[RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, DOI 10.17487/RFC4379, February 2006, <http://www.rfc-editor.org/info/rfc4379>.

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, DOI 10.17487/RFC5331, August 2008, <http://www.rfc-editor.org/info/rfc5331>.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, DOI 10.17487/RFC5340, July 2008, <http://www.rfc-editor.org/info/rfc5340>.

[RFC5443] Jork, M., Atlas, A., and L. Fang, "LDP IGP Synchronization", RFC 5443, DOI 10.17487/RFC5443, March 2009, <http://www.rfc-editor.org/info/rfc5443>.

[RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, DOI 10.17487/RFC5714, January 2010, <http://www.rfc-editor.org/info/rfc5714>.

[RFC5715] Shand, M. and S. Bryant, "A Framework for Loop-Free Convergence", RFC 5715, DOI 10.17487/RFC5715, January 2010, <http://www.rfc-editor.org/info/rfc5715>.

[RFC6571] Filsfils, C., Ed., Francois, P., Ed., Shand, M., Decraene, B., Uttaro, J., Leymann, N., and M. Horneffer, "Loop-Free Alternate (LFA) Applicability in Service Provider (SP) Networks", RFC 6571, DOI 10.17487/RFC6571, June 2012, <http://www.rfc-editor.org/info/rfc6571>.

[RFC6976] Shand, M., Bryant, S., Previdi, S., Filsfils, C., Francois, P., and O. Bonaventure, "Framework for Loop-Free Convergence Using the Ordered Forwarding Information Base (oFIB) Approach", RFC 6976, DOI 10.17487/RFC6976, July 2013, <http://www.rfc-editor.org/info/rfc6976>.

[RFC6981] Bryant, S., Previdi, S., and M. Shand, "A Framework for IP and MPLS Fast Reroute Using Not-Via Addresses", RFC 6981, DOI 10.17487/RFC6981, August 2013, <http://www.rfc-editor.org/info/rfc6981>.

[RFC6982] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", RFC 6982, DOI 10.17487/RFC6982, July 2013, <http://www.rfc-editor.org/info/rfc6982>.

[RFC6987] Retana, A., Nguyen, L., Zinin, A., White, R., and D. McPherson, "OSPF Stub Router Advertisement", RFC 6987, DOI 10.17487/RFC6987, September 2013, <http://www.rfc-editor.org/info/rfc6987>.

[RFC7307] Zhao, Q., Raza, K., Zhou, C., Fang, L., Li, L., and D. King, "LDP Extensions for Multi-Topology", RFC 7307, DOI 10.17487/RFC7307, July 2014, <http://www.rfc-editor.org/info/rfc7307>.

[RFC7490] Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, "Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)", RFC 7490, DOI 10.17487/RFC7490, April 2015, <http://www.rfc-editor.org/info/rfc7490>.

Appendix A. General Issues with Area Abstraction

When a multi-homed prefix is connected in two different areas, it may be impractical to protect them without adding the complexity of explicit tunneling. This is also a problem for LFA and Remote-LFA.

```
          50
        |----[ASBR Y]---[B]---[ABR 2]---[C]      Backbone Area 0:
        |                                |           ABR 1, ABR 2, C, D
        |                                |
        |                                |       Area 20:  A, ASBR X
        |                                |
        p ---[ASBR X]---[A]---[ABR 1]---[D]      Area 10: B, ASBR Y
           5                                  p is a Type 1 AS-external
```

Figure 5: AS external prefixes in different areas

Consider the network in Figure 5 and assume there is a richer connective topology that isn't shown, where the same prefix is announced by ASBR X and ASBR Y which are in different non-backbone areas. If the link from A to ASBR X fails, then an MRT alternate could forward the packet to ABR 1 and ABR 1 could forward it to D, but then D would find the shortest route is back via ABR 1 to Area 20. This problem occurs because the routers, including the ABR, in one area are not yet aware of the failure in a different area.

The only way to get it from A to ASBR Y is to explicitly tunnel it to ASBR Y. If the traffic is unlabeled or the appropriate MPLS labels are known, then explicit tunneling MAY be used as long as the shortest-path of the tunnel avoids the failure point. In that case, A must determine that it should use an explicit tunnel instead of an MRT alternate.

Authors' Addresses

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

Gabor Sandor Enyedi
Ericsson
Konyves Kalman krt 11.
Budapest 1097
Hungary

Email: Gabor.Sandor.Enyedi@ericsson.com