Network Working Group   A. Atlas, Ed.
Internet-Draft   Avici Systems, Inc.
Expires: August 22, 2005   February 21, 2005

Basic Specification for IP Fast-Reroute: Loop-free Alternates
draft-ietf-rtgwg-ipfrr-spec-base-03

Abstract

This document describes the use of loop-free alternates to provide local protection for unicast traffic in pure IP and MPLS/LDP networks in the event of a single failure, whether link, node or shared risk link group (SRLG). The goal of this technology is to reduce the micro-looping and packet loss that happens while routers converge after a topology change due to a failure. Rapid failure repair is achieved through use of precalculated backup next-hops that are loop-free and safe to use until the distributed network convergence

process completes.

Table of Contents

1. Introduction

Applications for interactive multimedia services such as VoIP and pseudo-wires can be very sensitive to traffic loss, such as occurs when a link or router in the network fails. A router's convergence time is generally on the order of seconds; the application traffic may be sensitive to losses greater than 10s of milliseconds.

As discussed in [FRAMEWORK], minimizing traffic loss requires a mechanism for the router adjacent to a failure to rapidly invoke a repair path, which is minimally affected by any subsequent re-convergence. This specification describes such a mechanism which allows a router whose local link has failed to forward traffic to a pre-computed alternate until the router installs the new primary next-hops based upon the changed network topology. The terminology used in this specification is given in [FRAMEWORK]. The described mechanism assumes that routing in the network is performed using a link-state routing protocol-- OSPF[RFC2328] or ISIS [RFC1195][RFC2966].

When a local link fails, a router currently must signal the event to its neighbors via the IGP, recompute new primary next-hops for all affected prefixes, and only then install those new primary next-hops into the forwarding plane. Until the new primary next-hops are installed, traffic directed towards the affected prefixes is discarded. This process can take seconds.

```
                          <--
                               +-----+
                        /------|  S  |--\
                       /       +-----+   \
                      / 5               8 \
                     /                     \
                   +-----+                +-----+
                   |  E  |                | N_1 |
                   +-----+                +-----+
                      \                     /
                  \    \  4              3 /  /
                   \|   \                 / |/
                   -+    \    +-----+    /  +-
                          \---|  D  |---/
                              +-----+
```

Figure 1: Basic Topology

The goal of IP Fast-Reroute is to reduce failure reaction time to 10s of milliseconds by using a pre-computed alternate next-hop, in the event that the currently selected primary next-hop fails, so that the

alternate can be rapidly used when the failure is detected. A network with this feature experiences less traffic loss and less micro-looping of packets than a network without IPFRR. There are cases where micro-looping is still a possibility since IPFRR coverage varies but in the worst possible situation a network with IPFRR is equivalent with respect traffic convergence to a network without IPFRR.

To clarify the behavior of IP Fast-Reroute, consider the simple topology in Figure 1. When router S computes its shortest path to router D, router S determines to use the link to router E as its primary next-hop. Without IP Fast-Reroute, that link is the only next-hop that router S computes to reach D. With IP Fast-Reroute, S also looks for an alternate next-hop to use. In this example, S would determine that it could send traffic destined to D by using the link to router N_1 and therefore S would install the link to N_1 as its alternate next-hop. At some later time, the link between router S and router E could fail. When that link fails, S and E will be the first to detect it. On detecting the failure, S will stop sending traffic destined for D towards E via the failed link, and instead send the traffic to S's pre-computed alternate next-hop, which is the link to N_1, until a new SPF is run and its results are installed. As with the primary next-hop, an alternate next-hop is computed for each destination. The process of computing an alternate next-hop does not alter the primary next-hop computed via a standard SPF.

If in the example of Figure 1, the link cost from N_1 to D increased to 30 from 3, then N_1 would not be a loop-free alternate, because the cost of the path from N_1 to D via S would be 17 while the cost from N_1 directly to D would be 30. In real networks, we may often face this situation. The existence of a suitable loop-free alternate next-hop is topology dependent.

A neighbor N can provide a loop-free alternate (LFA) if and only if

$$\text{Distance\_opt}(N, D) < \text{Distance\_opt}(N, S) + \text{Distance\_opt}(S, D)$$

Equation 1: Loop-Free Criterion

A sub-set of loop-free alternate are downstream paths which must meet the more restrictive condition of

Distance_opt(N, D) < Distance_opt(S, D)

Equation 2: Downstream Path Criterion

1.1 Failure Scenarios

The alternate next-hop can protect against a single link failure, a single node failure, one or more shared risk link group failure, or a combination of these. Whenever a failure occurs that is more extensive than what the alternate was intended to protect, there is the possibility of looping traffic. The example where a node fails when the alternate provided only link protection is illustrated below. If unexpected simultaneous failures occur, then micro-looping may occur since the alternates are not pre-computed to avoid the set of failed links.

If only link protection is provided and the node fails, it is possible for traffic using the alternates to experience micro-looping. This issue is illustrated in Figure 2. If Link(S->E) fails, then the link-protecting alternate via N will work correctly. However, if router E fails, then both S and N will detect a failure and switch to their alternates. In this example, that would cause S to redirect the traffic to N and N to redirect the traffic to S and thus causing a forwarding loop. Such a scenario can arise because the key assumption, that all other routers in the network are forwarding based upon the shortest path, is violated because of a second simultaneous correlated failure - another link connected to the same primary neighbor. If there are not other protection mechanisms a node failure is still a concern when only using link protection.

```
                              <@@@
                        @@@>
                 +-----+       +-----+
                 |  S  |-------|  N  |
                 +-+---+   5   +-----+
                   |              |
                   | 5          4 |  |
                |  |              | \|/
               \|/ |              |
                   |    +-----+   |
                   +----|  E  |---+
                        +--+--+
                           |
                           |
                           | 10
                           |
                        +--+--+
                        |  D  |
                        +-----+
```

Figure 2: Link-Protecting Alternates Causing Loop on Node Failure

Micro-looping of traffic via the alternates caused when a more extensive failure than planned for can be prevented via selection of only downstream paths as alternates. In Figure 2, S would be able to use N as an alternate, but N could not use S; therefore N would have no alternate and would discard the traffic, thus avoiding the micro-loop. A micro-loop due to the use of alternates can be avoided by using downstream paths because each router in the path to the destination must be closer to the destination (according to the topology prior to the failures). Although use of downstream paths ensures that the micro-looping via alternates does not occur, such a restriction can severely limit the coverage of alternates.

It may be desirable to find an alternate that can protect against other correlated failures (of which node failure is a specific instance). In the general case, these are handled by shared risk link groups (SRLGs) where any links in the network can belong to the SRLG. General SRLGs may add unacceptably to the computational complexity of finding a loop-free alternate.

However, a sub-category of SRLGs is of interest and can be applied only during the selection of an acceptable alternate. This sub-category is to express correlated failures of links that are connected to the same router. For example, if there are multiple logical sub-interfaces on the same physical interface, such as VLANs on an Ethernet interface, if multiple interfaces use the same physical port because of channelization, or if multiple interfaces

share a correlated failure because they are on the same line-card. This sub-category of SRLGs will be referred to as local-SRLGs. A local-SRLG has all of its member links with one end connected to the same router. Thus, router S could select a loop-free alternate which does not use a link in the same local-SRLG as the primary next-hop. The local-SRLGs belonging to E can be protected against via node-protection; i.e. picking a loop-free node-protecting alternate.

2. Applicability of Described Mechanisms

IP Fast Reroute mechanisms described in this memo cover intra-domain routing only, with OSPF[RFC2328] or ISIS [RFC1195][RFC2966] as the IGP. Specifically, Fast Reroute for BGP inter-domain routing is not part of this specification.

3. Alternate Next-Hop Calculation

In addition to the set of primary next-hops obtained through a shortest path tree (SPT) computation that is part of standard link-state routing functionality, routers supporting IP Fast Reroute also calculate a set of backup next hops that are engaged when a local failure occurs. These backup next hops are calculated to provide required type of protection (i.e. link-protecting and/or node-protecting) and to guarantee that when the expected failure occurs, forwarding traffic through them will not result in a loop. Such next hops are called loop-free alternates or LFAs throughout this specification.

In general, to be able to calculate the set of LFAs for a specific destination D, a router needs to know the following basic pieces of information:

- o Shortest-path distance from the calculating router to the destination (Distance_opt(S, D))

- o Shortest-path distance from the routerÆs IGP neighbors to the destination (Distance_opt(N, D))

- o Shortest path distance from the routerÆs IGP neighbors to itself (Distance_opt(N, S))

- o Distance_opt(S, D) is normally available from the regular SPF calculation performed by the link-state routing protocols. Distance_opt(N, D) and Distance_opt(N, S) can be obtained by performing additional SPF calculations from the perspective of each IGP neighbor (i.e. considering the neighbor's vertex as the root of the SPT--called SPT(N) hereafter--rather than the calculating router's one, called SPT(S)).

This specification defines a form of SRLG protection limited to those SRLGs that include a link that the calculating router is directly connected to. Information about local link SRLG membership is manually configured. Information about remote link SRLG membership is dynamically obtained using [ISIS-SRLG] or [OSPF-SRLG]. In order to choose among all available LFAs those that provide required SRLG protection for a given destination, the calculating router needs to track the set of SRLGs that the path through a specific IGP neighbor involves. To do so, each node D in the network topology is associated with SRLG_set(N, D), which is the set of SRLGs that would be crossed if traffic to D was forwarded through N. To calculate this set, the router initializes SRLG_set(N, N) for each of its IGP neighbors to be empty. During the SPT(N) calculation, when a new vertex V is added to the SPT, its SRLG_set(N, V) is set to the union of SRLG sets associated with its parents, and the SRLG sets associated with the links from V's parents to V. The union of the set of SRLG associated with a candidate alternate next-hop and the SRLG_set(N, D) for the neighbor reached via that candidate next-hop is used to determine SRLG protection.

The following sections provide information required for calculation of LFAs. Sections Section 3.1 through Section 3.5 define different types of LFA conditions. Section 3.6 describes constrains imposed by the IS-IS overload and OSPF stub router functionality. Section 3.7 defines the summarized algorithm for LFA calculation using the definitions in the previous sections.

3.1 Basic Loop-free Condition

Alternate next hops used by implementations following this specification MUST conform to at least the loop-freeness condition stated above in Equation 1. This condition guarantees that forwarding traffic to an LFA will not result in a loop after a link failure.

Further conditions may be applied when determining link-protecting and/or node-protecting alternate next-hops as described in Sections Section 3.2 and Section 3.3.

3.2 Node-Protecting Alternate Next-Hops

For an alternate next-hop N to protect against node failure of a primary neighbor E for destination D, N must be loop-free with respect to both E and D. In other words, N's path to D must not go through E. This is the case if Equation 3 is true, where N is the neighbor providing a loop-free alternate.

Distance_opt(N, D) < Distance_opt(N, E) + Distance_opt(E, D)

Equation 3: Criteria for a Node-Protecting Loop-Free Alternate

If Distance_opt(N,D) = Distance_opt(N, E) + Distance_opt(E, D), it is possible that N has equal-cost paths and one of those could provide protection against E's node failure. However, it is equally possible that one of N's paths goes through E, and the calculating router has no way to influence N's decision to use it. Therefore, it must be assumed that an alternate next-hop does not offer node protection if Equation 3 is not met.

3.3 Broadcast and NBMA Links

Verification of the link-protection property of a next hop in the case of a broadcast link is more elaborate than for a point-to-point link. This is because of the fact that a broadcast link is represented as a pseudo-node with zero-cost links connecting it to other nodes.

Because failure of an interface attached to a broadcast segment may mean loss of connectivity of the whole segment, the condition for broadcast link protection is pessimistic and requires that the alternate is loop- free with regard to the pseudo-node. Consider the example in Figure 3.

```
                      +-----+    15
                      |  S  |--------
                      +-----+       |
                         | 5        |
                         |          |
                         | 0        |
                       /----\ 0 5 +-----+
                       | PN |-----|  N  |
                       \----/     +-----+
                          | 0        |
                          |          | 8
                          | 5        |
                       +-----+ 5  +-----+
                       |  E  |----|  D  |
                       +-----+    +-----+
```

Figure 3: Loop-Free Alternate that is Link-Protecting

In Figure 3, N offers a loop-free alternate which is link-protecting. If the primary next-hop uses a broadcast link, then an alternate must be loop-free with respect to that link's pseudo-node to provide link protection. This requirement is described in Equation 4 below.

D_opt(N, D) < D_opt(N, pseudo) + D_opt(pseudo, D)

Equation 4: Loop-Free Link-Protecting Criterion for Broadcast Links

Because the shortest path from the pseudo-node goes through E, if a loop-free alternate from a neighbor N is node-protecting, the alternate will also be link-protecting unless the router S can only reach the neighbor N via the same pseudo-node. This can occur because S will direct traffic away from the shortest path to use an alternate. Therefore link protection must be considered during the alternate selection.

3.4 Downstream Alternate Next-Hops

In certain situations, described later, alternate next-hops must comply with the stricter condition provided in Equation 2 that defines a downstream path. The main property of the downstream paths is that traffic is always forwarded to a node that is closer to the destination, i.e. a node with a smaller metric. This property guarantees that no looping occurs regardless of the type of failure or the network architecture.

To ensure node-protection in certain scenarios, it is not sufficient to satisfy Equation 3. Instead the stricter downstream condition given in Equation 5 must be satisfied.

Distance_opt(N, D) < Distance_opt(E, D)

Equation 5: Criteria for a Node-Protecting Downstream Alternate

Similarly, to ensure link-protection in certain scenarios, the stricter downstream condition given in Equation 6 must be satisfied instead of merely Equation 4.

D_opt(N, D) < D_opt(pseudo, D)

Equation 6: Link-Protecting Downstream Criterion for Broadcast Links

The following types of alternate next-hops are defined. These describe increasingly contrained subsets of alternates; all strict downstream alternates are downstream alternates and all downstream alternates are loop-free alternates.

Loop-Free Alternate (LFA): Satisfies Equation 1. Link protection determined via Equation 4. Node protection determined via Equation 3.

Downstream Alternate: Satisfies Equation 2. Link protection determined via Equation 4. Node protection determined via Equation 3.

Strict Downstream Alternate (SDA): Satisfies Equation 2. Link protection determined via Equation 6. Node protection determined via Equation 5.

A downstream alternate is sufficient to guarantee that no looping occurs regardless of the type of failure. An SDA is necessary to guarantee protection in certain scenarios described in Section 6.2.

3.5 ECMP and Alternates

With equal-cost multi-path, a prefix may have multiple primary next-hops that are used to forward traffic. When a particular primary next-hop fails, alternate next-hops should be used to preserve the traffic. These alternate next-hops may themselves also be primary next-hops, but need not be. Other primary next-hops are not guaranteed to provide protection against the failure scenarios of concern.

```
       20 L1      L3   3
    [N]-----[ S ]--------[E3]
     |        |            |
     |      5 | L2         |
  20 |        |            |
     |    ---------        | 2
     |  5 |       | 5      |
     |   [E1]    [E2]------|
     |    |       |
     | 10 |    10 |
     |---[A]     [B]
          |       |
        2 |--[D]--| 2
```

Figure 4: ECMP where Primary Next-Hops Provide Limited Protection

In Figure 4 S has three primary next-hops to reach D; these are L2 to E1, L2 to E2 and L3 to E3. The primary next-hop L1 to E1 can obtain link and node protection from L3 to E3, which is one of the other primary next-hops; L1 to E1 cannot obtain link protection from the other primary next-hop L2 to E2. Similarly, the primary next-hop L2 to E2 can only get node protection from L2 to E1 and can only get link protection from L3 to E3. The third primary next-hop E3 can obtain link and node protection from L2 to E1, but can only get link protection from L2 to E2. It is possible for both the primary next-hop L2 to E2 and the primary next-hop L2 to E1 to obtain an

alternate next-hop that provides both link and node protection by using L1.

Alternate next-hops are determined for each primary next-hop separately. As with alternate selection in the non-ECMP case, these alternate next-hops should maximize the coverage of the failure cases.

3.6 Interactions with ISIS Overload, RFC 3137 and Costed Out Links

As described in [RFC3137], there are cases where it is desirable not to have a router used as a transit node. For those cases, it is also desirable not to have the router used on an alternate path.

For computing an alternate, a router MUST NOT consider diverting from the SPF tree along a link whose cost or reverse cost is LSInfinity (for OSPF) or the maximum cost (for ISIS) or whose next-hop router has the overload bit set (for ISIS).

In the case of OSPF, if all links from router S to a neighbor N_i have a reverse cost of LSInfinity, then router S MUST NOT consider using N_i as an alternate.

Similarly in the case of ISIS, if N_i has the overload bit set, then S MUST NOT consider using N_i as an alternate.

This preserves the desired behavior of diverting traffic away from a router which is following [RFC3137] and it also preserves the desired behavior when an operator sets the cost of a link to LSInfinity for maintenance which is not permitting traffic across that link unless there is no other path.

If a link or router which is costed out was the only possible alternate to protect traffic from a particular router S to a particular destination, then there will be no alternate provided for protection.

3.7 Selection Procedure

A router supporting this specification SHOULD select at least one loop-free alternate next-hop for each primary next-hop used for a given prefix. A router MAY decide to not use an available loop-free alternate next-hop. A reason for such a decision might be that the loop-free alternate next-hop does not provide protection for the failure scenario of interest.

The alternate selection should maximize the coverage of the failure cases.

S SHOULD select a loop-free node-protecting alternate next-hop, if one is available. If S has a choice between a loop-free link-protecting node-protecting alternate and a loop-free node-protecting alternate which is not link-protecting, S SHOULD select a loop-free node-protecting alternate which is also link-protecting. This can occur as explained in Section 3.3. If S has multiple primary next-hops, then S SHOULD select as a loop-free alternate either one of the other primary next-hops or a loop-free node-protecting alternate. If no loop-free node-protecting alternate is available, then S MAY select a loop-free link-protecting alternate.

Each next-hop can be categorized as to the type of alternate it can provide to a particular destination D from router S for a particular primary next-hop which goes to a neighbor E. A next-hop may provide one of the following types of paths:

Primary Path - This is the primary next-hop.

Loop-Free Node-Protecting Alternate - This next-hop satisfies Equation 1 and Equation 3. The path avoids S, S's primary neighbor E, and the link from S to E.

Loop-Free Link-Protecting Alternate - This next-hop satisfies Equation 1 but not Equation 3. If the primary next-hop uses a broadcast link, then this next-hop satisfies Equation 4.

Unavailable - This may be because the path goes through S to reach D, because the link is costed out, etc.

An alternate path may also provide none, some or complete SRLG protection as well as node and link or link protection. For instance, a link may belong to two SRLGs G1 and G2. The alternate path might avoid other links in G1 but not G2, in which case the alternate would only provide partial SRLG protection.

4. Using an Alternate

If an alternate next-hop is available, the router SHOULD redirect traffic to the alternate next-hop when the primary next-hop has failed.

When a local interface failure is detected, traffic that was destined to go out the failed interface must be redirected to the appropriate alternate next-hops. Other failure detection mechanisms which detect the loss of a link or a node may also be used to trigger redirection of traffic to the appropriate alternate next-hops. The mechanisms available for failure detection are discussed in [FRAMEWORK] and are

outside the scope of this specification.

The alternate next-hop MUST be used only for traffic types which are routed according to the shortest path. Multicast traffic is specifically out of scope for this specification.

4.1 Terminating Use of Alternate

A router MUST limit the amount of time an alternate next-hop is used after the primary next-hop has become unavailable. This ensures that the router will start using the new primary next-hops. It ensures that all possible transient conditions are removed and the network converges according to the deployed routing protocol.

It is desirable to avoid micro-forwarding loops involving S. An example illustrating the problem is given in Figure 5. If the link from S to E fails, S will use N1 as an alternate and S will compute N2 as the new primary next-hop to reach D. If S starts using N2 as soon as S can compute and install its new primary, it is probable that N2 will not have yet installed its new primary next-hop. This would cause traffic to loop and be dropped until N2 has installed the new topology. This can be avoided by S delaying its installation and leaving traffic on the alternate next-hop.

```
      +-----+
      |  N2 |--------   |
      +-----+   1   |  \|/
          |         |
          |     +-----+    @@>  +-----+
          |     |  S  |---------|  N1 |
       10 |     +-----+   10    +-----+
          |        |               |
          |      1 |    |          |
          |        |   \|/    10   |
          |     +-----+            |  |
          |     |  E  |            | \|/
          |     +-----+            |
          |        |               |
          |      1 |  |            |
          |        | \|/           |
          |    +-----+             |
          |----|  D  |--------------
               +-----+
```

Figure 5: Example where Continued Use of Alternate is Desirable

This is an example of a case where the new primary is not a loop-free alternate before the failure and therefore may have been forwarding traffic through S. This will occur when the path via a previously upstream node is shorter than the the path via a loop-free alternate neighbor. In these cases, it is useful to give sufficient time to ensure that the new primary neighbor and other nodes on the new primary path have switched to the new route.

If the newly selected primary was loop-free before the failure, then it is safe to switch to that new primary immediately; the new primary wasn't dependent on the failure and therefore its path will not have changed.

Given that there is an alternate providing appropriate protection and while the assumption of a single failure holds, it is safe to delay the installation of the new primaries; this will not create forwarding loops because the alternate's path to the destination is known to not go via S or the failed element and will therefore not be affected by the failure.

An implementation SHOULD continue to use the alternate next-hops for packet forwarding even after the new routing information is available based on the new network topology. The use of the alternate next-hops for packet forwarding SHOULD terminate:

a. if the new primary next-hop was loop-free prior to the topology change, or

b. if a configured hold-down, which represents a worst-case bound on the length of the network convergence transition, has expired, or

c. if notification of an unrelated topological change in the network is received.

5. Requirements on LDP Mode

Since LDP traffic will follow the path specified by the IGP, it is also possible for the LDP traffic to follow the loop-free alternates indicated by the IGP. To do so, it is necessary for LDP to have the appropriate labels available for the alternate so that the appropriate out-segments can be installed in the forwarding plane before the failure occurs.

This means that a Label Switched Router (LSR) running LDP must distribute its labels for the FECs it can provide to all its neighbors, regardless of whether or not they are upstream. Additionally, LDP must be acting in liberal label retention mode so that the labels which correspond to neighbors that aren't currently the primary neighbor are stored. Similarly, LDP should be in downstream unsolicited mode, so that the labels for the FEC are distributed other than along the SPT.

If these requirements are met, then LDP can use the loop-free alternates without requiring any targeted sessions or signaling extensions for this purpose.

6. Routing Aspects

6.1 Multi-Homed Prefixes

An SPF-like computation is run for each topology, which corresponds to a particular OSPF area or ISIS level. The IGP needs to determine loop-free alternates to multi-homed routes. Multi-homed routes occur for routes obtained from outside the routing domain by multiple routers, for subnets on links where the subnet is announced from multiple ends of the link, and for routes advertised by multiple routers to provide resiliency.

Figure 6 demonstrates such a topology. In this example, the shortest path to reach the prefix p is via E. The prefix p will have the link to E as its primary next-hop. If the alternate next-hop for the prefix p is simply inherited from the router advertising it on the

shortest path to p, then the prefix p's alternate next-hop would be the link to C. This would provide link protection, but not the node protection that is possible via A.

```
                   5   +---+  4   +---+  5  +---+
                 ------| S |------| A |-----| B |
                 |     +---+      +---+     +---+
                 |       |                    |
                 |     5 |                  5 |
                 |       |                    |
               +---+ 5 +---+   5       7    +---+
               | C |---| E |------ p -------| F |
               +---+   +---+                +---+
```

Figure 6: Multi-homed prefix

To determine the best protection possible, the prefix p can be treated in the SPF computations as a node with uni-directional links to it from those routers that have advertised the prefix. Such a node need never have its links explored, as it has no out-going links.

If there exist multiple multi-homed prefixes exist that share the same connectivity and the difference in metrics to those routers, then a single node can be used to represent the set. For instance, if in Figure 6 there were another prefix X that was connected to E with a metric of 1 and to F with a metric of 3, then that prefix X could use the same alternate next-hop as was computed for prefix p.

A router SHOULD compute the alternate next-hop for an IGP multi-homed prefix by considering alternate paths via all routers that have announced that prefix.

6.2 OSPF

OSPF introduces certain complications because it is possible for the traffic path to exit an area and then re-enter that area. This can occur whenever the same route is considered from multiple areas. There are several cases where issues such as this can occur. They happen when another area permits a shorter path to connect two ABRs than is available in the area where the LFA has been computed. To clarify, an example topology is given in Appendix A.

a. Virtual Links: These allow paths to leave the backbone area and traverse the transit area. The path provided via the transit area can exit via any ABR. The path taken is not the shortest path determined by doing an SPF in the backbone area.

b. Alternate ABR[RFC3509]: When an ABR is not connected to the backbone, it considers the inter-area summaries from multiple areas. The ABR A may determine to use area 2 but that path could traverse another alternate ABR B that determines to use area 1. This can lead to scenarios similar to that illustrated in Figure 7.

c. ASBR Summaries: An ASBR may itself be an ABR and can be announced into multiple areas. This presents other ABRs with a decision as to which area to use. This is the example illustrated in Figure 7.

d. AS External Prefixes: A prefix may be advertised by multiple ASBRs in different areas and/or with multiple forwarding addresses that are in different areas, which are connected via at least one common ABR. This presents such ABRs with a decision as to which area to use to reach the prefix.

This issue does not exist for non-backbone intra-area routes. A candidate alternate next-hop must be an LFA. For intra-area backbone, inter-area, and AS External routes, a candidate alternate next-hop must be an SDA to be used.

If no virtual links exist, backbone intra-area routes can use candidate alternate next-hops that are LFAs and not SDAs. If no Alternate ABRs exist, then inter-area routes can use candidate alternate next-hops that are LFAs and not SDAs.

If no ASBR exists simultaneously in multiple non-backbone areas and no prefix is included in announcements either by two or more ASBRs that are in different areas or in announcements associated with multiple forwarding addresses that are in different areas, then AS External routes can use candidate alternate next-hops that are LFAs and not SDAs.

The inappropriate use of an LFA that isn't an SDA can cause forwarding loops or lack of protection.

In all cases where an SDA is required, this is because the path taken cannot be determined via the SPT in the local area. The use of an SDA relies on the fact that any path taken will use hops with monotonically decreasing distance to the destination. This does not allow knowledge of the actual path the traffic will traverse. Therefore, it is not possible, based on the computations described in this specification, to determine whether an SDA will provide protection against an SRLG failure.

6.2.1 OSPF External Routing

An additional complication comes from forwarding addresses, where an ASBR uses a forwarding address to indicate to all routers in the Autonomous System to use the specified address instead of going through the ASBR. When a forwarding address has been indicated, all routers in the topology calculate the shortest path to the link specified in the external LSA. In this case, the alternate next-hop should be computed by selecting among the alternate paths to the forwarding link(s) instead of among alternate paths to the ASBR.

6.3 BGP Next-Hop Synchronization

Typically BGP prefixes are advertised with AS exit routers router-id, and AS exit routers are reached by means of IGP routes. BGP resolves its advertised next-hop to the immediate next-hop by potential recursive lookups in the routing database. IP Fast-Reroute computes the alternate next-hops to all IGP destinations, which include alternate next-hops to the AS exit router's router-id. BGP simply inherits the alternate next-hop from IGP. The BGP decision process is unaltered; BGP continues to use the IGP optimal distance to find the nearest exit router. MBGP routes do not need to copy the alternate next hops.

It is possible to provide ASBR protection if BGP selected a set of IGP next-hops and allowed the IGP to determine the primary and alternate next-hops as if the BGP route were a multi-homed prefix. This is for future study.

6.4 Multicast Considerations

Multicast traffic is out of scope for this specification of IP Fast-Reroute. The alternate next-hops SHOULD not used for multi-cast RPF checks.

7. Security Considerations

This document does not introduce any new security issues. The mechanisms described in this document depend upon the network topology distributed via an IGP, such as OSPF or ISIS. It is dependent upon the security associated with those protocols.

8 References

[FRAMEWORK]
Shand, M., "IP Fast Reroute Framework", draft-ietf-rtgwg-ipfrr-framework-02.txt (work in progress), October 2004.

[ISIS-SRLG]
Kompella, K. and Y. Rekhter, "IS-IS Extensions in Support of Generalized MPLS", draft-ietf-isis-gmpls-extensions-19 (work in progress), October 2003.

[OSPF-SRLG]
Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching", draft-ietf-ccamp-ospf-gmpls-extensions-12 (work in progress), October 2003.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, December 1990.

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

[RFC2966] Li, T., Przygienda, T. and H. Smit, "Domain-wide Prefix Distribution with Two-Level IS-IS", RFC 2966, October 2000.

[RFC3036] Andersson, L., Doolan, P., Feldman, N., Fredette, A. and B. Thomas, "LDP Specification", RFC 3036, January 2001.

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

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V. and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RFC3509] Zinin, A., Lindem, A. and D. Yeung, "Alternative Implementations of OSPF Area Border Routers", RFC 3509, April 2003.

Authors' Addresses

Alia K. Atlas (editor)
Avici Systems, Inc.
101 Billerica Avenue
N. Billerica, MA 01862
USA

Phone: +1 978 964 2070
EMail: aatlas@avici.com

Raveendra Torvi
Avici Systems, Inc.
101 Billerica Avenue
N. Billerica, MA 01862
USA

Phone: +1 978 964 2026
EMail: rtorvi@avici.com

Gagan Choudhury
AT&T
200 Laurel Avenue, Room D5-3C21
Middletown, NJ 07748
USA

Phone: +1 732 420-3721
EMail: gchoudhury@att.com

Christian Martin
Verizon
1880 Campus Commons Drive
Reston, VA 20191
USA

Brent Imhoff
LightCore
14567 North Outer Forty Rd.
Chesterfield, MO 63017
USA

Phone: +1 314 880 1851
EMail: brent@lightcore.net

Don Fedyk
Nortel Networks
600 Technology Park
Billerica, MA 01821
USA

Phone: +1 978 288 3041
EMail: dwfedyk@nortelnetworks.com

Appendix A. OSPF Example Where LFA Based on Local Area Topology is Insufficient

This appendix provides an example scenario where the local area topology does not suffice to determine that an LFA is available. As described in Section 6.2, one problem scenario is for ASBR summaries where the ASBR is available in two areas via intra-area routes and there is at least one ABR or alternate ABR that is in both areas. The following Figure 7 illustrates this case.

```
                  5
        [ F ]-----------[ C ]
          |               |
          |               | 5
       20 |          5    |     1
          |   [ N ]-----[ A ]*****[ F ]
          |     |         #         *
          |  40 |         # 50      *  2
          |     |    5    #    2    *
          |   [ S ]-----[ B ]*****[ G ]
          |     |         *
          |   5 |         * 15
          |     |         *
          |   [ E ]     [ H ]
          |     |         *
          |   5 |         * 10**
          |     |         *
          |---[ X ]-----[ASBR]
                     5

          ----  Link in Area 1
          ****  Link in Area 2
          ####  Link in Backbone Area 0
```

Figure 7: Topology with Multi-area ASBR Causing Area Transiting

In Figure 7, the ASBR is also an ABR and is announced into both area 1 and area 2. A and B are both ABRs that are also connected to the backbone area. S determines that N can provide a loop-free alternate to reach the ASBR. N's path goes via A. A also sees an intra-area route to ASBR via Area 2; the cost of the path in area 2 is 30, which is less than 35, the cost of the path in area 1. Therefore, A uses the path from area 2 and directs traffic to F. The path from F in area 2 goes to B. B is also an ABR and learns the ASBR from both areas 1 and area 2; B's path via area 1 is shorter (cost 20) than B's path via area 2 (cost 25). Therefore, B uses the path from area 1 that connects to S.

Intellectual Property Statement

Disclaimer of Validity

Copyright Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.