Network Working Group
Internet-Draft
Intended status: Historic
Expires: May 19, 2008

S. Bryant
C. Filsfils
S. Previdi
M. Shand
Cisco Systems
November 16, 2007

IP Fast Reroute using tunnels
draft-bryant-ipfrr-tunnels-03

Status of this Memo

Copyright Notice

Abstract

This draft describes an IP fast re-route mechanism that provides backup connectivity in the event of a link or router failure. In the absence of single points of failure and asymmetric costs, the mechanism provides complete protection against any single failure. If perfect repair is not possible, the identity of all the unprotected links and routers is known in advance.

This IP Fast Reroute advanced method was invented in 2002 and draft (draft-bryant-ipfrr-tunnels-00.txt) describing it was submitted to the IETF in May 2004. It was one of the first methods of achieving full repair coverage in an IP Network, and as such the draft has been widely referenced in the academic literature.

The authors DO NOT propose that this IPFRR method be implemented since better IPFRR advanced method capable of achieving full repair coverage have subsequently been invented.

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 [RFC2119].

Table of Contents

1. History

This IP Fast Reroute advanced method was invented in 2002 and draft (draft-bryant-ipfrr-tunnels-00.txt) describing it was submitted to the IETF in May 2004. It was one of the first methods of achieving full repair coverage in an IP Network, and as such the draft has been widely referenced in the academic literature. Since IETF drafts are ephemeral, the authors have requested the IETF Editor to publish this as a historic RFC so that it is available for reference.

The authors DO NOT propose that this IPFRR method be implemented. Better IPFRR advanced method capable of achieving full repair coverage have since been invented, and are the subject of work in progress in the IETF.

One final note, in some versions of the draft the abstract term P was renamed F, and the abstract term Q was renamed G. For reasons of personal preference this version of the document reverts to the terms P and Q.

2. Terminology

This draft uses the terms defined in [I-D.ietf-rtgwg-ipfrr-framework]. This section defines additional words, acronyms, and actions used in this draft.

Directed Forwarding

: The ability of the repairing router (S) to specify the next hop (Q) on exit from a tunnel end-point (P).

Extended P-space

: The union of the P-space of the neighbours of a specific router with respect to a common component.

  Extended P-space does not include the additional space reachable though directed forwarding.

P
: The router in P-space to which a packet is tunnelled for repair.

PQ
: A router that is in both P and Q-space and hence does not need directed forwarding.

P-space	P-space is the set of routers reachable from a specific router without any path (including equal cost path splits) transiting a specified component. For example, the P-space of S, is the set of routers that S can reach without using E (router failure case) or the S-E link failure case).
Q	The router in Q-space, to which the packet is directed by router P on exit from the repair tunnel. Q will always be adjacent to P, or P itself.
Q-space	Q-space is the set of routers from which a specific router can be reached without any path (including equal cost path splits) transiting a specified component.
Interference	The condition where the network costs are such that a repairing router cannot tunnel a packet sufficiently far from a failed node such that it is not attracted back to the failed node via another of that node's neighbours.

3. Introduction

When a link or node failure occurs in a routed network, there is inevitably a period of disruption to the delivery of traffic until the network re-converges on the new topology. Packets for destinations which were previously reached by traversing the failed component may be dropped or may suffer looping. Traditionally such disruptions have lasted for periods of at least several seconds, and most applications have been constructed to tolerate such a quality of service.

Recent advances in routers have reduced this interval to under a second for carefully configured networks using link state IGPs. However, new Internet services are emerging which may be sensitive to periods of traffic loss which are orders of magnitude shorter than this.

Addressing these issues is difficult because the distributed nature of the network imposes an intrinsic limit on the minimum convergence time which can be achieved.

However, there is an alternative approach, which is to compute backup routes that allow the failure to be repaired locally by the router(s) detecting the failure without the immediate need to inform other

routers of the failure. In this case, the disruption time can be limited to the small time taken to detect the adjacent failure and invoke the backup routes. This is analogous to the technique employed by MPLS Fast Reroute [RFC4090], but the mechanisms employed for the backup routes in pure IP networks are necessarily very different.

A framework for IP Fast Reroute [I-D.ietf-rtgwg-ipfrr-framework] provides a summary of the proposed IPFRR solutions, and a partial solution using equal cost multi-path and loop-free alternate case is described in [I-D.ietf-rtgwg-ipfrr-spec-base].

This draft describes extensions to the basic repair mechanism in which we propose the use of tunnels to provide additional logical downstream paths. These mechanisms provide almost 100% repair connectivity in practical networks.

4. Goals, non-goals, limitations and constraints

4.1. Goals

The following are the goals of IPFRR:

- Protect against any link or router failure in the network.
- No constraints on the network topology or link costs.
- Never worse than the existing routing convergence mechanism.
- Co-existence with non-IP fast-reroute capable routers in the network.

4.2. Non-Goals

The following are non-goals of IPFRR:

- Protection of a single point of failure.
- To provide protection in the presence of multiple concurrent failures other than those that occur due to the failure of a single router.
- Shared risk group protection.
- Complete fault coverage in networks that make use of asymmetric costs.

4.3. Limitations

The following limitations apply to IPFRR:

- Because the mechanisms described here rely on complete topological information from the link state routing protocol, they will only work within a single link state flooding domain.

- Reverse Path Forwarding (RPF) checks cannot be used in conjunction with IPFRR. This is because the use of tunnels may result in packets arriving over different interfaces than expected.

4.4. Constraints

The following constraints are assumed:

- Following a failure, only the routers adjacent to the failure have any knowledge of the failure.

- There is insufficient time following a failure to compute a repair strategy based on knowledge of the specific failure that has occurred.

- Multiple concurrent failures may not be protected.

5. Repair Paths

When a router detects an adjacent failure, it uses a set of repair paths in place of the failed component, and continues to use this until the completion of the routing transition. Only routers adjacent to the failed component are aware of the nature of the failure. Once the routing transition has been completed, the router will have no further use for the repair paths since all routers in the network will have revised their forwarding data and the failed link will have been eliminated from this computation.

Repair paths are pre-computed in anticipation of later failures so they can be promptly activated when a failure is detected.

Three types of repair path are used to achieve the repair:

1. Equal cost path-split.

2. Loop-free Alternate.

3. Tunnel.

The operation of equal cost path-split and loop-free alternate is described in [I-D.ietf-rtgwg-ipfrr-spec-base]. A tunnelled repair path tunnels traffic to some staging point from which it will travel to its destination using normal forwarding without looping back. The repair path can be thought of as providing a virtual link, originating at a router adjacent to a failure, and diverting traffic around the failure. This is equivalent to providing a virtual loop-free alternate to supplement the physical loop-free alternates.

5.1. Tunnels as Repair Paths

The repair strategies described in this draft operate on the basis that if a packet can somehow be sent to the other side of the failure, it will subsequently proceed towards its destination exactly as if it had traversed the failed component. See Figure 1.

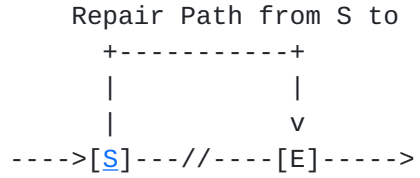

Figure 1: Simple Link Repair.

Creating a repair path from S to E may require a packet to traverse an unnatural route. If a suitable natural path starts at a neighbour (i.e. it is a loop-free alternate), then S can force the packet directly there. If this is not the case, then S may create one by using a tunnel to carry the packet to a point in the network where there is a real loop-free alternate. Note that the tunnel does not have to go from S to E. The tunnel can terminate at any router in the network, provided that S can be sure that the packet will proceed correctly to its destination from that router.

A repair path computed for a link failure may not however work satisfactorily when the neighbouring router has, itself, failed. This is illustrated in Figure 2

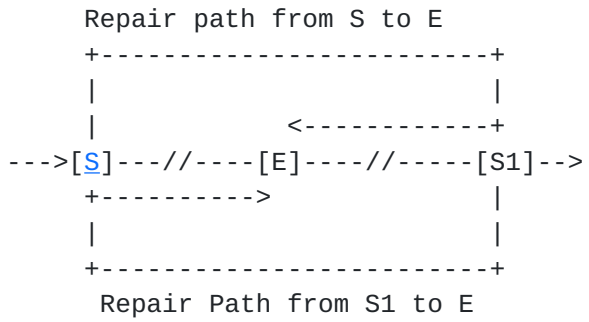

Figure 2: Looping Link Repair when Router Fails

Consider the case of a router E with just two neighbours S and S1. When router E fails, both S and S1 will observe the failure of their local link to E, but will have no immediate knowledge that E itself has failed. If they were both to attempt to repair traffic around their local link, they would invoke mutual repairs which would loop.

Since it is not easy for a router to immediately distinguish between a link failure and the failure of its neighbour, repair paths are calculated in anticipation of adjacent router failure. Thus, for each of its protected links, router S (Figure 3) pre-computes a set of tunnelled repair paths, one for each of the neighbours (S1, S2 and S3) of its neighbour E on the S-E link. The set of destinations that are normally assigned to link S-E will be assigned to a repair path based on the neighbour of E through which router E would have forwarded traffic to them.

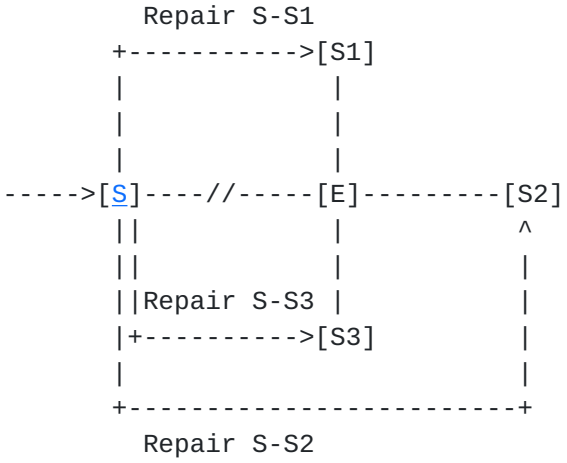

Figure 3: Repair paths in anticipation of a router failure.

The set of repair paths in Figure 3 will function correctly in the case of link and router failure. However, in some network topologies they may not provide a means for traffic to reach router E itself. This is important in cases where E is a single point of failure and E

is still functional (i.e. the failure was actually a failure of the S-E link). Hence, in addition to computing repair paths for the neighbours of its neighbour on a protected link, a router also calculates a repair path for the neighbour itself. This is illustrated in Figure 4.

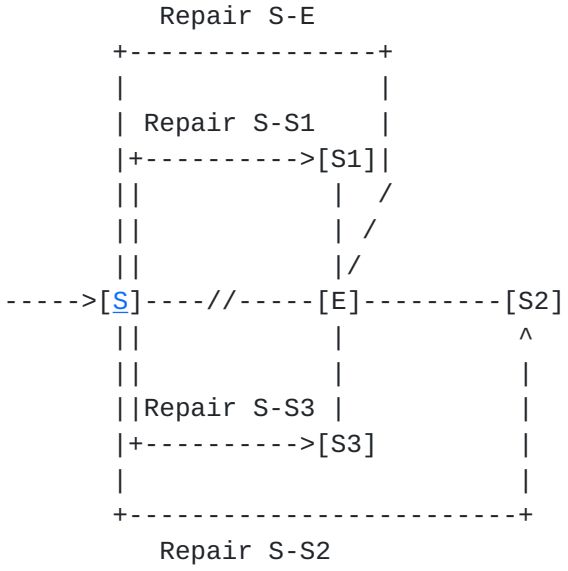

Figure 4: The full set of S-E repair paths.

In the event of a failure, the only traffic that is assigned to the link repair path (the S-E repair) is that traffic which has no other path to its destination except via E. As we have already seen, there is a danger that traffic assigned to this link repair path may loop if E has failed, therefore, when the repair paths are invoked, a loop detection mechanism is used which promptly detects the loop and, upon detection, withdraws the link (S-E) repair path from service.

5.2. Tunnel Requirements

There are a number of IP in IP tunnel mechanisms that may be used to fulfil the requirements of this design. Suitable candidates include IP-in-IP [RFC1853], GRE[RFC1701]] and L2TPv3 [RFC3931]. The selection of the specific tunnelling mechanism (and any necessary enhancements) used to provide a repair path is outside the scope of this document. However the following sections describe the requirements for the tunnelling mechanism.

5.2.1. Setup

When a failure is detected, it is necessary to immediately redirect traffic to the repair paths. Consequently, the tunnels used must be provisioned beforehand in anticipation of the failure. IP fast re-

route will determine which tunnels it requires. It must therefore be possible to establish tunnels automatically, without management action, and without the need to manually establish context at the tunnel endpoint.

5.2.2. Multipoint

To reduce the number of tunnel endpoints in the network the tunnels should be multi-point tunnels capable of receiving repair traffic from any IPFRR router in the network.

5.2.3. Directed forwarding

Directed forwarding must be supported such that the router at the tunnel endpoint (P can be directed by the router at the tunnel source (S) to forward the packet directly to a specific neighbour. Specification of the directed forwarding mechanism is outside the scope of this document. Directed forwarding might be provided using an enhancement to the IP tunnelling encapsulation, or it might be provided through the use of a single MPLS label stack entry [RFC3032] interposed between the IP tunnel header and the packet being repaired.

5.2.4. Security

A lightweight security mechanism should be supported to prevent the abuse of the repair tunnels by an attacker. This is discussed in more detail in Section 12.

6. Construction of Repair Paths

6.1. Identifying Repair Path Targets

To establish protection for a link or node it is necessary to determine which neighbours of the neighbouring node should be targets of repair paths. Normally all neighbours will be used as repair path targets. However, in some topologies, not all neighbours will be needed as targets because, prior to the failure, no traffic was being forwarded through them by the repairing router. This can be determined by examining the normal shortest path tree (SPT) computed by the repairing router.

In addition, the neighbouring router E will also be the target of a repair path for any destinations for which E is a single point of failure.

6.2. Determining Tunneled Repair Paths

The objective of each tunnelled repair path is to deliver traffic to a target router when a link is observed to have failed. However, it is seldom possible to use the target router itself as the tunnel endpoint because other routers on the repair path, that have not learned of the failure, will forward traffic addressed to it using their least cost path which may be via the failed link. This is illustrated in Figure 5 in which all link costs are one in both directions. Router S's intended repair path for traffic to D when link S-E fails is the path W-X-Y-Z-S1. However, if router S makes S1 be the tunnel endpoint and forwards the packet to router W, router W will immediately return it to S because its least cost path to S1 is S-E-S1 (cost 3 versus cost 4) and has no knowledge of the failure of link S-E.

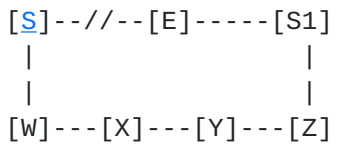

Figure 5: Repair path to target router S1.

Thus the tunnel endpoint needs to be somewhere on the repair path such that packets addressed to the tunnel end point will not loop back towards router S. In addition, the release point needs to be somewhere such that when packets are released from the tunnel they will flow towards the target router (or their actual destination) without being attracted back to the failed link. By inspection, in Figure 5, suitable tunnel endpoints are routers X, Y, and Z.

Note that it is not essential that traffic assigned to a repair path actually traverse the target router for which the repair path was created. If, for example, in Figure 5, if a packet's destination were normally reached via the path S-E-S1-Z-?-?-?, once it was released at any of the possible tunnel endpoints, it would arrive at its destination by the best available route without traversing S1.

In general, the properties that are required of tunnel endpoints are:

- The end point must be reachable from the tunnel source without traversing the failed link; and

- When released, tunnelled packets will proceed towards their destination without being attracted back over the failed link or node.

Provided both of these conditions are met, packets forwarded on the

repair path will not loop.

In some topologies it will not be possible to find a tunnel endpoint that exhibits both the required properties. For example, in Figure 5, if the cost of link X-Y were increased from one to four in both directions, there is no longer a viable endpoint within the fragment of the topology shown.

To solve this problem we introduce the concept of directed forwarding from the tunnel endpoint. Directed forwarding allows the originator of a tunnelled packet to instruct that, when it is decapsulated at the end of the tunnel, it be forwarded via a specific adjacency, and not be subjected to the normal forwarding decision process. This effectively allows the tunnel to be extended by one hop. So, for example, in Figure 5 with the cost of link X-Y set to four, it would be possible to select X as the tunnel endpoint with the directive that X always forward the packets it decapsulates via the adjacency to Y. Thus, router X is reached from S using normal forwarding, and directed forwarding is then used to force packets to router Y, from where S1 can be reached using normal forwarding.

Provided link costs are symmetrical, it can be proved that it is always possible to compute a tunnelled repair path (possibly using directed forwarding) around a link failure, and that the tunnel endpoint (P) and the release point (Q) will be coincident, or may be separated by at most one hop.

6.2.1. Computing Repair Paths

For a router S, determining tunnelled repair paths around a neighbouring router E, the set of potential tunnel end points includes all the routers that can be reached from S using normal forwarding without traversing the failed link S-E. This is termed the "P-space" of S with respect to the failure of E. Any router that is on an equal cost path split via the failed link is excluded from this set.

The resulting set defines all the possible tunnel end points that could be used in repair paths originating at router S for the failure of link S-E. This set can be obtained by computing an SPT rooted at S and excising the sub-tree reached via the S-E link. The set of possible release points can be determined by computing the set of routers that can reach the repair path target without traversing the failed link. This is termed the "Q-space" of the target with respect to the failure. The Q-space can be obtained by computing a reverse shortest path tree (rSPT) rooted at the repair path target, with the sub-tree which traverses the failed link (or node) excised. The rSPT uses the cost towards the root rather than from it and yields the

best paths towards the root from other nodes in the network.

The intersection of the target's Q-space with S's P-space includes all the possible release points for any repair path not employing directed forwarding. Where there is no intersection, but there exist a pair of routers, P in S's P-space and Q in the target's Q-space, router P can be used as the tunnel endpoint with directed forwarding to the release point Q.

6.2.2. Extended P-space

The description in Section 6.2.1 calculated router S's P-space rooted at S itself. However, since router S will only use a repair path when it has detected the failure of the link S-E, the initial hop of the repair path need not be subject to S's normal forwarding decision process. Thus we introduce the concept of extended P- space. Router S's extended P-space is the union of the P-spaces of each of S's neighbours. The use of extended P-space may allow router S to repair to targets that were otherwise unreachable.

6.2.3. Loop-free Alternates

When a loop-free alternate exists, no tunnelling is required.

6.2.4. Selecting Repair Paths

The mechanisms described above will identify all the possible release points that can be used to reach each particular target. (The circumstances when no release points exist are described in Section 6.4.) In a well-connected network there are likely to be multiple possible release points for each target, and all will work correctly. For simplicity, one release point per target is chosen. All will deliver the packets correctly so, arguably, it does not matter which is chosen. However, one release point may be preferred over the others on the basis of path cost or some other criteria.

It is an implementation matter as to how the release point is selected.

6.3. Assigning Traffic to Repair Paths

Once the repair path for each target has been selected, it is necessary to determine which of the destinations normally reached via the protected link should be assigned to which of the repair paths when the link fails.

This is achieved by recording which neighbour of E would be used to reach each destination reachable over S-E when running the original

SPF. Traffic assignment is then simply a matter of assigning the traffic which E would have forwarded via each neighbour to the repair path which has that neighbour as its target.

Although the repair paths are calculated based on traffic addressed to specific targets, it can be proved that the traffic assignment algorithm guarantees that the repair path can be used for any traffic assigned to it.

Where E would normally split the traffic to a particular destination via two or more of its neighbours, it is an implementation decision whether the repaired traffic should be split across the corresponding set of repair paths. The repair path to E itself is normally used just for traffic destined for E and any prefixes advertised by E. However, under some circumstances, it may be impossible to compute a repair path to one or more of E's neighbours, for example, because E is a single point of failure. In this case traffic for the destinations served by the otherwise irreparable targets is assigned to the repair path with E as its target, in the optimistic assumption that router E is still functioning. If router E is indeed still functioning, this will ensure delivery of the traffic. If, however, router E has failed, the traffic on this repair path will loop as previously shown in Section 5.1. The way this is detected, and the course of action when it is detected, is described in Section 7.3.

6.4. When no Repair Path is Possible

Under some circumstances, it will not be possible to identify a repair path to one or more of the targets. This can occur for the following reasons:

1. The neighbouring router that is presumed to have failed constitutes a single point of failure in the network.

2. Severely asymmetric link costs may cause an otherwise viable physical repair path to be unusable.

3. Interference may occur between the repair paths of individual targets

In practise, these cases are unlikely to be encountered frequently. Networks that will benefit from the mechanisms described here will usually exhibit considerable redundancy and are normally operated with largely symmetric link costs. Note that a router's inability to compute a full set of repair paths for one of its links does not necessarily affect its ability to do so for its other links.

Example topologies illustrating each of the three cases above are

described in the following subsections.

6.4.1. Unreachable Target

If the failure of a neighbouring router makes one or more of its neighbours genuinely unreachable, clearly it will not be possible to establish a repair path to such targets. Such single points of failure are not expected to be encountered frequently in properly designed networks, and will probably occur only when the network has previously suffered other failures that have reduced its connectivity.

6.4.2. Asymmetric Link Costs

When link costs have been set asymmetrically, it is possible that a repair path cannot be constructed even using directed forwarding.

Although it is trivial to construct a network fragment with this property, this should not be regarded as a major problem. Firstly, asymmetric link costs are seldom used deliberately. And, secondly, even when an asymmetric link cost prevents one potential repair path being used, there will normally be other ones available.

6.4.3. Interference Between Potential Node Repair Paths

Under some circumstances the existence of one neighbour may interfere with a potential repair path to another. Consider the topology shown in Figure 6, in which all links have a symmetrical cost of one, with the exception of that between H and I, which has a cost of 3. In this example, the fact that router J is a neighbour of E prevents the discovery of a repair path from router S to router S1 despite the existence of an apparently suitable path.

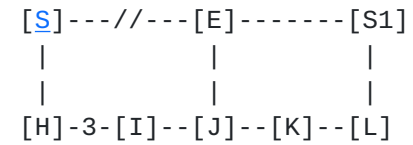

Figure 6: Interference between repair paths

A repair path from router S to J can use J itself as the release point by employing directed forwarding from I. However, it is not possible to identify a suitable release point for a repair path to router S1 within the topology shown since there is nowhere that router S can reach that will subsequently forward traffic to router S1 except via the forbidden link E-S1 (J's least cost path to S1 is J-E-S1). This is because the extended P-space of router S is

separated by more than one hop from the Q-space of router S1.

Since the topology shown in Figure 6 will typically form part of a much larger topology, a different, and possibly more circuitous repair path from S to S1, that does not go via J, may be discovered. This is illustrated in Figure 7. In this enhanced topology, a repair path to S1 using Y as the release point can be used.

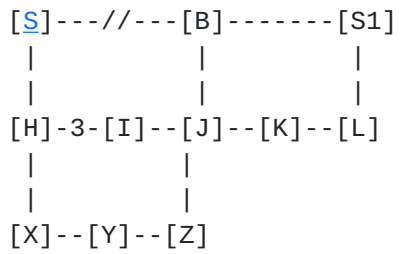

Figure 7: Resolving interference in a larger network

Note that, in Figure 6, if the traffic for S1 were assigned to the repair path for J, it would correctly reach S1 because J would assign it to its repair path to S1. That is, packets from S to S1 would travel via two successive tunnels. Consequently, this is referred to as a "secondary repair path". However, it is not always the case that interference can be handled in this fashion and it is possible to create looping repair paths.

One possibility of looping repair paths is illustrated in Figure 8. All links have a symmetrical cost of one with the exception of H-I, which is cost 3 in both directions, and K-L and L-S1 which are cost 5 in the indicated direction and cost 1 in the other.

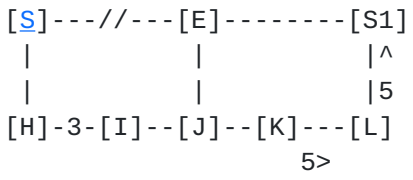

Figure 8: Looping secondary repair paths

In this topology, S can establish a repair path to J, but cannot establish a repair path to S1 because of interference. Router S might assign traffic intended for S1 onto its repair path to J expecting it to undergo a secondary repair towards S1. However, because of the asymmetrical link costs, J is unable to establish a repair path to S1. It is only able to establish a repair path to S. If J, like S, elected to forward repaired traffic to S1 using its (only) repair path to S, similarly expecting a secondary repair to get it to its destination, traffic for S1 would loop between S and J.

Thus when interference occurs, the possibility of a secondary repair path cannot be relied upon to ensure that traffic reaches its destination.

In order to determine the viability of secondary repair paths, it is necessary for each router to take into account the repair paths which the other neighbours of router E can achieve. These can be computed locally by running the repair path computation algorithms rooted at each of those neighbours. It is only necessary to compute the repair paths from the routers to which router S can establish repair paths, with targets of those routers to which repair paths have not yet been established.

It is then possible to determine whether all routers can now be reached by invoking secondary (or if necessary tertiary, etc.) repair paths, and if so, to which primary repair path traffic for each target should be assigned.

There is another, more subtle, possibility of loops arising when secondary repair paths are used. This is illustrated in Figure 9, where all links are cost 1 with the exception of L-K which has a cost 5 in that direction and cost 1 in the direction K-L.

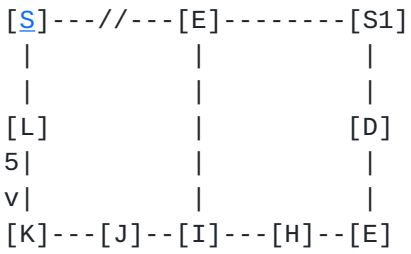

Figure 9: Example of an apparently non-looping secondary repair path which results in a loop.

Router S has a primary repair path to I (with a release point of K), and I has a primary repair path to S1 (with a release point of E). It would appear that these form a non-looping secondary repair path from S to S1. As usual, the primary repair path from S to I has been computed on the basis of destinations normally reachable through E-I. However, when making use of the secondary repair path, the traffic inserted in the repair path from S to I will be destined not for one of the routers normally reachable via E-I, but for S1. Hence this repair path is not necessary valid for such traffic and in this example it will have a 50% probability of being forwarded back along the path K-L-S-E-S1, and hence looping.

This problem can in general be avoided by choosing a release point for the initial primary repair with the property that traffic for the

secondary target (S1) is guaranteed to traverse the primary target (I). This can be achieved by computing the rSTF rooted at the secondary target (S1) and examining the sub-tree which traverses the primary target. It can be proved that in the absence of asymmetric link costs, such a release point will always exist. Where asymmetric link costs prevent this, the traffic can be encapsulated to the intermediate router (I), which may require the use of double encapsulation. On reaching router I, the traffic for S1 is decapsulated and then forwarded in I's primary repair path to S1 (via router E, in the example).

6.5. Multi-homed Prefixes

Up to this point, it has been assumed that any particular prefix is "attached" to exactly one router in the network, and consequently only the routers in the network need be considered when constructing repair paths, etc. However, in many cases the same prefix will be attached to two or more routers. Common cases are:

- The subnet present on a link is advertised from both ends of the link.
- Prefixes are propagated from one routing domain to another by multiple routers.
- Prefixes are advertised from multiple routers to provide resilience in the event of the failure of one of the routers.

In general, this causes no particular problems, and the shortest route to each prefix (and hence which of the routers to which it is attached should be used to reach it) is resolved by the normal SPF process. However, in the particular case where one of the instances of a prefix is attached to router E, or to a router for which router E is a single point of failure, the situation is more complicated.

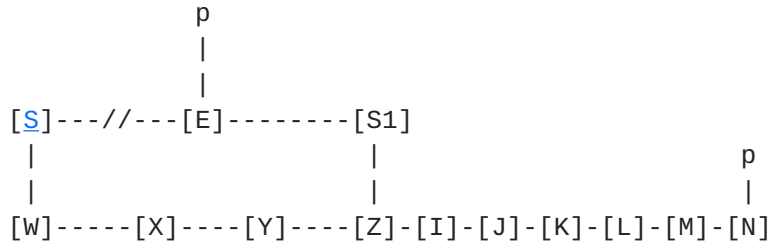

Figure 10: A multi-homed prefix p

Consider a prefix p, which is attached to router E and some other router N as illustrated in Figure 10. Before the failure of the link S-E, p is reachable from S via S-E. After the failure it cannot be

assumed that E is still reachable. If traffic to p is assigned to a link repair path to E (as it would be if p were attached only to E), and router E has failed, then it would loop and subsequently be dropped. Traffic for p cannot simply be assigned to whatever repair path would be used for traffic to N, because other routers, which are not yet aware of any failure, may direct the traffic back towards E, since the instance of p attached to E is closer.

A solution is to treat p itself as a neighbour of E, and compute a repair path with p as a target. However, although correct, this solution may be infeasible where there are a very large number of such prefixes, which would result in an unacceptably large computational overhead.

Some simplification is possible where there exist a large number of multi-homed prefixes which all share the same connectivity and metrics. These may be treated as a single router and a single repair path computed for the entire set of prefixes.

An alternative solution is to tunnel the traffic for a multi-homed prefix to the router N where it is also attached (see Figure 10). If this involves a repair path that was already tunnelled, then this requires double encapsulation.

6.6. LANs and pseudo-nodes

In link state protocols a LAN is represented by a construct known as a pseudo-node in IS-IS and a network LSA in OSPF.

In order to deal correctly with this representation of LANs, the algorithms described in this draft require certain modifications. There are four cases which require consideration. These are described in the following subsections.

6.6.1. The Link between Routers S and E is a LAN

In this case, the link which is being protected is a LAN, and the router E which has potentially failed is reachable over the LAN. This is illustrated in Figure 11.

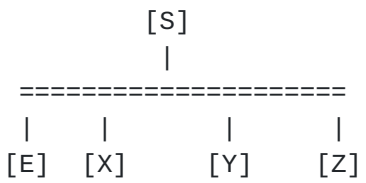

Figure 11: The link between routers S and E is a LAN

There are two possible failure modes in this case.

6.6.1.1. Case 1

Router E or its interface to the LAN may have failed independently of the rest of the LAN. In this case the remaining routers on the LAN (routers X, Y and Z) will remain reachable from router S. These routers do not appear as direct neighbours of router E in the link state database and are not treated as neighbours of router E for the purposes of this specification because no traffic from router S would be directed through router E to any of these routers. However, each of these neighbouring routers will have router E as a neighbour and they will initiate their own repair paths in the event of the failure of router E or its LAN interface.

Repair paths are computed with the non-LAN neighbours of E as targets, and also E itself (the "link-failure" repair path). Note that since the remaining neighbours of S on the LAN are assumed to be still reachable when the link to E has failed, these repair paths may traverse the LAN.

A separate set of repair paths is required in anticipation of the potential failure of each router on the LAN.

6.6.1.2. Case 2

Router S's interface to the LAN may have failed (or the entire LAN may have failed). In either event, simultaneous failures will be observed from router S to all the remaining routers on the LAN (routers E, X, Y and Z). In this case, the pseudo-node itself can be treated as the "adjacent" router (i.e. the router normally referred to as "router E"), and repairs constructed using the normal mechanisms with all the neighbours of the pseudo-node (routers E, X, Y and Z) as repair path targets. If one or more of the routers had failed in addition to the LAN connectivity, treating it as a repair path target would not be viable, but this would be a case of multiple simultaneous failures which is out of scope of this specification.

The entire sub-tree over S's LAN interface is the failed component and is excised from the SPT when computing S's extended P-space. For

the Q-spaces of the targets, the sub-tree over the LAN interface of the target is excised.

6.6.1.3. Simplified LAN repair

A simpler alternative strategy is to always consider the LAN and all routers attached to it as failing as a single unit. In this case, a single set of repair paths is computed with targets being the entire set of non-LAN neighbours of all the routers on the LAN, together with "link-repair" paths with all the routers on the LAN as targets. Any failure of one or more LAN adjacencies results in these repair paths being invoked for all neighbours on the LAN. These repair paths must not traverse the LAN, and so must be computed by excising the entire sub-tree reachable over S's LAN interface from S's SPT (i.e. the entire LAN is the failed component). The Q-spaces are computed as normal, with the LAN neighbours or their interface to the LAN being excised as appropriate. This is simpler than the approach proposed above, but will fail to make use of possible repair paths (or even path splits) over the LAN. In particular, if the only viable repair paths involve the LAN, it will prevent any repair being possible.

6.6.2. A LAN exists at the release point

When computing the viable release points, it may be that one or more of the leaf nodes are actually pseudo-nodes. In this case, the release point is deemed to be any of the parent nodes on the LAN by which the pseudo-node had been reached, and when computing the extended set of release points (reachable by directed forwarding), all the remaining routers on the LAN may be included.

6.6.3. A LAN between E and its neighbors

If there is a LAN between router E and one or more of E's neighbours (other than router S), then rather than treating each of those neighbours as a separate target to which a repair path must be computed, the pseudo-node itself can be treated as a single target for which a repair path can be computed. If there are other neighbours of E which are directly attached to E, including those which may also be attached to the LAN, they must still be treated as an individual repair path target.

Normally a repair path with the pseudo-node as its target will have a release point before the pseudo-node. However it is possible that the release point would be computed as the pseudo-node itself. This will occur if the rSPT rooted at the pseudo-node includes no routers other than itself. In this case a single repair with the pseudo-node as target is not possible, and it is necessary to compute individual

repair paths whose target are each of the neighbours of E on the LAN.

6.6.4. The LAN is a Transit Subnet

This is the most common case, where a LAN is traversed by a repair path, but is not in any of the special positions described above. In this case no special treatment is required, and the normal SPF mechanisms are applicable.

7. Failure Detection and Repair Path Activation

The details of repair path activation are inherently implementation-dependent and must be addressed by individual design specifications. This section describes the implementation independent aspects of the fail-over to the repair path.

7.1. Failure Detection

The failure detection mechanism must provide timely detection of the failure and activation of the repair paths. The failure detection mechanisms may be media specific (for example loss of light), or may be generic (for example BFD [I-D.ietf-bfd-base]). Multiple detection mechanisms may be used in order to improve detection latency. Note that in the case of a LAN it may be necessary to monitor connectivity to all of the adjacent routers on the LAN.

7.2. Repair Path Activation

The mechanism used by the router to activate the repair path following failure will be implementation specific.

An implementation that is capable of withdrawing the repair may delay the start of network convergence in order to minimise network disruption in the event that the failure was a transient.

7.3. Node Failure Detection Mechanism

When router S detects a failure of the S-E link, it will invoke the link repair path from itself to router S. This S-E link repair is always invoked because even if all other traffic can be re-routed, E is always a single point of failure to itself. If router E has failed, the S-E link repair can result in a forwarding loop. A node failure detection mechanism is therefore needed. A suitable mechanism might be to run BFD ([I-D.ietf-bfd-base]) between S and E, over the S-E link repair path.

When the node failure detection mechanism has determined that router

E has failed it withdraws the S-E link repair path. The node failure detection and revocation of the S-E link repair needs to be expedited, in order to minimise the duration of collateral damage to the network cause by packets looping around the S-E link repair path.

If E is a single point of failure to some destinations, then withdrawing the S-E link repair has no impact on network connectivity, because those destinations will have been rendered unreachable by the failure of router E.

If E is not a single point of failure, but traffic to some destinations is being repaired via the S-E link because of the inability to provide suitable repair paths, then there are destinations that are rendered temporarily unreachable by IPFRR. The IPFRR loop free convergence mechanism delays normal convergence of the network. Consideration therefore has to be given to the relative importance of the traffic being protected and the traffic being black-holed. Depending on the outcome of that consideration, the IPFRR loop-free strategy may need to be abandoned.

8. Loop Free Transition

Once the repair paths have been activated, data will again be forwarded correctly. At this stage only the routers directly adjacent to the failure will be aware of the failure because no routing information concerning the failure has yet been propagated to other routers. The network now has to be transitioned to normal operation using the available components.

During network transition inconsistent state may lead to the formation of micro-loops. During this period, packets may be prevented from reaching the repair path, may expire due to transiting an excessive number of hops, may be subject to excessive delay, and the resultant congestion may disrupt the passage of other packets through the network. A loop free transition technique which allows the network to re-converge without packet loss or disruption is therefore required.

A number of suitable loop-free convergence techniques are described in [I-D.ietf-rtgwg-lf-conv-frmwk].

9. IPFRR Capability

In the previous sections it has been assumed that all routers in the network are capable of acting as IPFRR routers, performing such tasks as tunnel termination and directed forwarding. In practise this is

unlikely to be the case, partially because of the heterogeneous nature of a practical network, and partially because of the need to progressively deploy such capability. IPFRR therefore needs to support some form of capability announcement, and the algorithms need to take these capabilities into account when calculating their path repair strategies. For example, the ability of routers to function as tunnel end points and perform directed forwarding will influence the choice of repair path. However, routers which are simply traversed by repair paths (tunnelled or not) do not need to be IPFRR capable in order to guarantee correct operation of the repair paths.

10. Enhancements to routing protocols

It will be seen from the above that a number of enhancements to the appropriate routing protocols are needed to support IPFRR. The following possible enhancements have been identified:

- The ability to advertise IPFRR capability
- The ability to advertise tunnel endpoint capability
- The ability to advertise directed forwarding identifiers
- The ability to announce the start of a loop-free transition, and to abort a loop-free transition.
- The ability to signal transition completion status to neighbours.
- The ability to advertise that a link is protected.

Capability advertisement should make use of existing capability mechanisms in the routing protocols. The exact set of enhancements will depend on specific IPFRR design choices.

11. IANA Considerations

There are no IANA considerations that arise from this architectural description of IPFRR. However there will be changes to the IGPs to support IPFRR in which there will be IANA considerations.

12. Security Considerations

Changes to the IGPs to support IPFRR do not introduce any additional security risks.

The security implications of the increased convergence time due to the loop avoidance strategy depend on the approach to multiple failures. If the presence of multiple failures results in the network aborting the loop free strategy, then the convergence time will be similar to that of a conventional network. On the other hand, an attacker in a position to disrupt part of a network might use this to disrupt the repair of a critical path.

The tunnel endpoints need to be secured to prevent their use as a facility by an attacker. Performance considerations indicate that tunnels cannot be secured by IPsec [RFC4301]. A system of packet address policing, both at the tunnel endpoints and at the edges of the network would prevent an attacker's packet arriving at a tunnel endpoint and would seem to be the best strategy.

When a fast re-route is in progress, there may be an unacceptable increase in traffic load over the repair path. Network operators need to examine the computed repair paths and ensure that they have sufficient capacity.

13. Acknowledgments

The authors acknowledge the significant technical contributions made to this work by their colleagues: John Harper and Kevin Miles.

14. Security Considerations

All micro-loop control mechanisms raise significant security issues which must be addressed in their detailed technical description.

15. Informative References

[I-D.ietf-bfd-base]
Katz, D. and D. Ward, "Bidirectional Forwarding Detection", draft-ietf-bfd-base-06 (work in progress), March 2007.

[I-D.ietf-rtgwg-ipfrr-framework]
Shand, M. and S. Bryant, "IP Fast Reroute Framework", draft-ietf-rtgwg-ipfrr-framework-07 (work in progress), July 2007.

[I-D.ietf-rtgwg-ipfrr-notvia-addresses]
Bryant, S., "IP Fast Reroute Using Not-via Addresses", draft-ietf-rtgwg-ipfrr-notvia-addresses-01 (work in

progress), July 2007.

[I-D.ietf-rtgwg-ipfrr-spec-base]
Atlas, A. and A. Zinin, "Basic Specification for IP Fast-Reroute: Loop-free Alternates", draft-ietf-rtgwg-ipfrr-spec-base-09 (work in progress), September 2007.

[I-D.ietf-rtgwg-lf-conv-frmwk]
Bryant, S. and M. Shand, "A Framework for Loop-free Convergence", draft-ietf-rtgwg-lf-conv-frmwk-01 (work in progress), July 2007.

[RFC1701] Hanks, S., Li, T., Farinacci, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 1701, October 1994.

[RFC1853] Simpson, W., "IP in IP Tunneling", RFC 1853, October 1995.

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

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, January 2001.

[RFC3931] Lau, J., Townsley, M., and I. Goyret, "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, March 2005.

[RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

Authors' Addresses

Stewart Bryant
Cisco Systems
250, Longwater, Green Park,
Reading RG2 6GB, UK
UK

Email: stbryant@cisco.com

Clarence Filsfils
Cisco Systems
De Kleetlaan 6a
1831 Diegem,
Belgium

Phone:
Fax:
Email: cfilsfil@cisco.com
URI:

Stefano Previdi
Cisco Systems
Via Del Serafico 200
00142 Roma,
Italy

Phone:
Fax:
Email: sprevidi@cisco.com
URI:

Mike Shand
Cisco Systems
250, Longwater, Green Park,
Reading RG2 6GB, UK
UK

Email: mshand@cisco.com

Full Copyright Statement

Intellectual Property

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).