Internet Engineering Task Force (IETF)
Request for Comments: 9855
Category: Standards Track
ISSN: 2070-1721

A. Bashandy
HPE
S. Litkowski
C. Filsfils
Cisco Systems
P. Francois
INSA Lyon
B. Decraene
Orange
D. Voyer
Bell Canada

October 2025

Topology Independent Fast Reroute Using Segment Routing

Abstract

This document presents Topology Independent Loop-Free Alternate (TI-LFA) Fast Reroute (FRR), which is aimed at providing protection of node and Adjacency segments within the Segment Routing (SR) framework. This FRR behavior builds on proven IP FRR concepts being LFAs, Remote LFAs (RLFAs), and Directed Loop-Free Alternates (DLFAs). It extends these concepts to provide guaranteed coverage in any two-connected networks using a link-state IGP. An important aspect of TI-LFA is the FRR path selection approach establishing protection over the expected post-convergence paths from the Point of Local Repair (PLR), reducing the operational need to control the tie-breaks among various FRR options.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

This document outlines a local repair mechanism that leverages Segment Routing (SR) to restore end-to-end connectivity in the event of a failure involving a directly connected network component. This mechanism is designed for standard link-state Interior Gateway Protocol (IGP) shortest path scenarios. Non-SR mechanisms for local repair are beyond the scope of this document. Non-local failures are addressed in a separate document [SR-LOOP].

The term Topology Independent (TI) describes the capability providing a loop-free backup path that is effective across all network topologies. This provides a major improvement compared to LFA [RFC5286] and RLFA [RFC7490], which cannot provide a complete protection coverage in some topologies as described in [RFC6571].

When the network reconverges after failure, micro-loops [RFC5715] can form due to transient inconsistencies in the forwarding tables of different routers. If it is determined that micro-loops are a significant issue in the deployment, then a suitable loop-free convergence method should be implemented, such as one of those described in [RFC5715], [RFC6976], [RFC8333], or [SR-LOOP].

TI-LFA operates locally at the Point of Local Repair (PLR) upon detecting a failure in one of its direct links. Consequently, this local operation does not influence:

* Micro-loops that may or may not form during the distributed IGP

convergence as delineated in [RFC5715]:

- These micro-loops occur on routes directed towards the destination that do not traverse paths configured for TI-LFA. According to [RFC5714], the formation of such micro-loops can prevent traffic from reaching the PLR, thereby bypassing the TI-LFA paths established for rerouting.

* Micro-loops that may or may not develop when the previously failed link is restored to functionality.

TI-LFA paths are activated from the instant the PLR detects a failure in a local link and remain in effect until the IGP convergence at the PLR is fully achieved. Consequently, they are not susceptible to micro-loops that may arise due to variations in the IGP convergence times across different nodes through which these paths traverse. This ensures a stable and predictable routing environment, minimizing disruptions typically associated with asynchronous network behavior. However, an early (relative to the other nodes) IGP convergence at the PLR and the consecutive "early" release of TI-LFA paths may cause micro-loops, especially if these paths have been computed using the methods described in Sections 5.2, 5.3, or 5.4 of this document. One of the possible ways to prevent such micro-loops is local convergence delay [RFC8333].

TI-LFA procedures are complementary to the application of any micro-loop avoidance procedures in the case of link or node failure:

* Link or node failure requires some urgent action to restore the traffic that passed through the failed resource. TI-LFA paths are pre-computed and pre-installed; therefore, they are suitable for urgent recovery.

* The paths used in the micro-loop avoidance procedures typically cannot be pre-computed.

For each destination (as specified by the IGP) in the network, TI-LFA pre-installs a backup forwarding entry for each protected destination ready to be activated upon detection of the failure of a link used to reach the destination. TI-LFA provides protection in the event of any one of the following: single link failure, single node failure, or single Shared Risk Link Group (SRLG) failure. In link failure mode, the destination is protected assuming the failure of the link. In node protection mode, the destination is protected assuming that the neighbor connected to the primary link (see Section 2) has failed. In SRLG protecting mode, the destination is protected assuming that a configured set of links sharing fate with the primary link has failed (e.g., a linecard or a set of links sharing a common transmission pipe).

Protection techniques outlined in this document are limited to protecting links, nodes, and SRLGs that are within a link-state IGP area. Protecting domain exit routers and/or links attached to another routing domain is beyond the scope of this document.

By utilizing SR, TI-LFA eliminates the need to establish Targeted Label Distribution Protocol sessions with remote nodes for leveraging the benefits of Remote Loop-Free Alternates (RLFAs) [RFC7490] [RFC7916] or Directed Loop-Free Alternates (DLFAs) [IPFRR-TUNNELS]. All the Segment Identifiers (SIDs) required are present within the Link State Database (LSDB) of the IGP. Consequently, there is no longer a necessity to prefer LFAs over RLFAs or DLFAs, nor is there a need to minimize the number of RLFA or DLFA repair nodes.

Utilizing SR also eliminates the need to establish an additional state within the network for enforcing explicit Fast Reroute (FRR) paths. This spares the nodes from maintaining a supplementary state and frees the operator from the necessity to implement additional protocols or protocol sessions solely to augment protection coverage.

TI-LFA also brings the benefit of the ability to provide a backup path that follows the expected post-convergence path considering a particular failure, which reduces the need of locally configured policies that influence the backup path selection [RFC7916]. The easiest way to express the expected post-convergence path in a loop-free manner is to encode it as a list of Adjacency segments. However, this may create a long segment list that some hardware may not be able to program. One of the challenges of TI-LFA is to encode the expected post-convergence path by combining Adjacency segments and node segments. Each implementation may independently develop its own algorithm for optimizing the ordered segment list. This document provides an outline of the fundamental concepts applicable to constructing the SR backup path, along with the related data plane procedures. Appendix A contains a more detailed description of some of the aspects of TI-LFA related to post-convergence path.

This document is structured as follows:

* Section 2 defines the main notations used in the document. They are in line with [RFC5714].

* Section 3 defines the main principles of TI-LFA backup path computation.

* Section 4 suggests to compute the P-space and Q-space properties defined in Section 2 for the specific case of nodes lying over the post-convergence paths towards the protected destinations.

* Section 5 describes how to compute protection lists that encode a loop-free post-convergence path towards the destination using the properties defined in Section 4.

* Section 6 defines the segment operations to be applied by the PLR to ensure consistency with the forwarding state of the repair node.

* Section 7 discusses aspects that are specific to the data plane.

* Section 8 discusses the relationship between TI-LFA and the SR algorithm.

* Section 9 provides an overview of the certain considerations that are needed when Adjacency segments are used in a Repair List (RL).

* Section 10 discusses security considerations.

* Appendix A highlights advantages of using the expected post-convergence path during FRR.

* Appendix B summarizes the measurements from implementing the algorithms detailed in this document within actual service provider and large enterprise network environments. Real-life measurements are presented regarding the number of SIDs utilized by repair paths.

2. Terminology

2.1. Abbreviations and Notations

DLFA: Directed Loop-Free Alternate

FRR: Fast Reroute

IGP: Interior Gateway Protocol

LFA: Loop-Free Alternate

LSDB: Link State Database

PLR: Point of Local Repair

RL: Repair List

RLFA: Remote Loop-Free Alternate

SID: Segment Identifier

SPF: Shortest Path First

SPT: Shortest Path Tree

SR: Segment Routing

SRLG: Shared Risk Link Group

TI-LFA: Topology Independent Loop-Free Alternate

The main notations used in this document are defined as follows:

* The terms "old" and "new" topologies refer to the LSDB state before and after the considered failure, respectively.

* SPT_old(R) is the SPT rooted at node R in the initial state of the network.

* SPT_new(R, X) is the SPT rooted at node R in the state of the network after the resource X has failed.

* The PLR is the router that applies fast traffic restoration after detecting failure in a directly attached link, set of links, and/or node.

* Similar to [RFC7490], the concept of P-space and Q-space is used for TI-LFA.

* The P-space P(R, X) of a router R with regard to a resource X (e.g., a link S-F, a node F, or an SRLG) is the set of routers reachable from R using the pre-convergence shortest paths without any of those paths (including equal-cost path splits) transiting through X. A P node is a node that belongs to the P-space.

* Consider the set of neighbors of a router R and a resource X. Exclude from that set the neighbors that are reachable from R using X. The extended P-space P'(R, X) of a node R with regard to a resource X is the union of the P-spaces of the neighbors in that reduced set of neighbors with regard to the resource X.

* The Q-space Q(R, X) of a router R with regard to a resource X is the set of routers from which R can be reached without any path (including equal-cost path splits) transiting through X. A Q node is a node that belongs to the Q-space.

* EP(P, Q) is an explicit SR path from a node P to a node Q.

* The primary interface and primary outgoing interface are one of the outgoing interfaces towards a destination according to the IGP link-state protocol.

* The primary link is a link connected to the primary interface.

* The Adj-SID(S-F) is the Adjacency segment from node S to node F.

2.2. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Base Principle

The basic algorithm to compute the repair path is to pre-compute SPT_new(R, X) and, for each destination, encode the repair path as a loop-free segment list. One way to provide a loop-free segment list is to use Adjacency SIDs only. However, this approach may create very long SID lists that hardware may not be able to handle due to Maximum SID Depth (MSD) limitations.

An implementation is free to use any local optimization to provide smaller segment lists by combining Node-SIDs and Adjacency SIDs. In addition, the usage of Node-SIDs allow for maximizing ECMPs over the backup path. These optimizations are out of scope of this document; however, the subsequent sections provide some guidance on how to leverage P-spaces and Q-spaces to optimize the size of the segment list.

4. Intersecting P-space and Q-space with Post-Convergence Paths

One of the challenges of defining an SR path following the expected post-convergence path is to reduce the size of the segment list. In order to reduce this segment list, an implementation MAY determine the P-space / extended P-space and Q-space properties (defined in [RFC7490]) of the nodes along the expected post-convergence path from the PLR to the protected destination and compute an SR explicit path from P to Q when they are not adjacent. Such properties will be used in Section 5 to compute the TI-LFA RL.

4.1. Extended P-space Property Computation for a Resource X over Post-Convergence Paths

The objective is to determine which nodes on the post-convergence path from the PLR R to the destination D are in the extended P-space of R with regard to resource X (where X can be a link or a set of links adjacent to the PLR or a neighbor node of the PLR).

This can be found by:

* excluding neighbors that are not on the post-convergence path when computing P’(R, X), then

* intersecting the set of nodes belonging to the post-convergence path from R to D, assuming the failure of X, with P’(R, X).

4.2. Q-space Property Computation for a Resource X over Post-Convergence Paths

The goal is to determine which nodes on the post-convergence path from the PLR R to the destination D are in the Q-space of destination D with regard to resource X (where X can be a link or a set of links adjacent to the PLR, or a neighbor node of the PLR).

This can be found by intersecting the set of nodes belonging to the post-convergence path from R to D, assuming the failure of X, with

Q(D, X).

4.3. Scaling Considerations When Computing Q-space

[RFC7490] raises scaling concerns about computing a Q-space per destination. Similar concerns may affect TI-LFA computation if an implementation tries to compute a reverse Shortest Path Tree (SPT) [RFC7490] for every destination in the network to determine the Q-space. It will be up to each implementation to determine the good tradeoff between scaling and accuracy of the optimization.

5. TI-LFA Repair Path

The TI-LFA repair path consists of an outgoing interface and a list of segments (a Repair List (RL)) to insert on the SR header in accordance with the data plane used. The RL encodes the explicit (and possibly post-convergence) path to the destination, which avoids the protected resource X. At the same time, the RL is guaranteed to be loop-free, irrespective of the state of FIBs along the nodes belonging to the explicit path, as long as the states of the FIBs are programmed according to a link-state IGP. Thus, there is no need for any coordination or message exchange between the PLR and any other router in the network.

The TI-LFA repair path is found by intersecting P(S, X) and Q(D, X) with the post-convergence path to D and computing the explicit SR-based path EP(P, Q) from a node P in P(S, X) to a node Q in Q(D, X) when these nodes are not adjacent along the post-convergence path. The TI-LFA RL is expressed generally as (Node-SID(P), EP(P, Q)).

```
S --------- N1 --------- D
*\          | \          |
* \         |  \         |
*  \        |   \        |
*   N2----- R1***R2 *** R3
*           *
N3 **********

  ***** : link with high metric (1k)
  ----- : link with metric 1
```

Figure 1: Sample Topology with TI-LFA

As an example, in Figure 1, the focus is on the TI-LFA backup from S to D, considering the failure of node N1.

* First, P(S, N1) is computed and results in [N3, N2, R1].

* Then, Q(D, N1) is computed and results in [R3].

* The expected post-convergence path from S to D considering the failure of N1 is <N2 -> R1 -> R2 -> R3 -> D> (we are naming it "PCPath" in this example).

* P(S, N1) intersection with PCPath is [N2, R1]. With R1 being the deeper downstream node in PCPath, it can be assumed to be used as a P node (this is an example, and an implementation could use a different strategy to choose the P node).

* Q(D, N1) intersection with PCPath is [R3], so R3 is picked as a Q node. An SR-explicit path is then computed from R1 (P node) to R3 (Q node) following PCPath (R1 -> R2 -> R3): <Adj-SID(R1-R2), Adj-SID(R2-R3)>.

As a result, the TI-LFA RL of S for destination D considering the failure of node N1 is: <Node-SID(R1), Adj-SID(R1-R2), Adj-

SID(R2-R3)>.

Most often, the TI-LFA RL has a simpler form, as described in the following sections. Appendix B provides statistics for the number of SIDs in the explicit path to protect against various failures.

5.1. FRR Path Using a Direct Neighbor

When a direct neighbor is in P(S, X) and Q(D, x), and the link to that direct neighbor is on the post-convergence path, the outgoing interface is set to that neighbor and the repair segment list is empty.

This is comparable to a post-convergence LFA FRR repair.

5.2. FRR Path Using a PQ Node

When a remote node R is in P(S, X) and Q(D, x) and on the post-convergence path, the RL is made of a single node segment to R, and the outgoing interface is set to the outgoing interface used to reach R.

This is comparable to a post-convergence RLFA repair tunnel.

5.3. FRR Path Using a P Node and Q Node That Are Adjacent

When a node P is in P(S, X) and a node Q is in Q(D, x), and both are on the post-convergence path and are adjacent to each other, the RL is made of two segments: a node segment to P (to be processed first), followed by an Adjacency segment from P to Q.

This is comparable to a post-convergence Directed Loop-Free Alternate (DLFA) repair tunnel.

5.4. Connecting Distant P and Q Nodes Along Post-Convergence Paths

In some cases, there is no adjacent P and Q node along the post-convergence path. As mentioned in Section 3, a list of Adjacency SIDs can be used to encode the path between P and Q. However, the PLR can perform additional computations to compute a shorter list of segments that represent a loop-free path from P to Q. How these computations are done is out of scope of this document and is left to implementation.

6. Building TI-LFA Repair Lists for SR Segments

The following sections describe how to build the RLs using the terminology defined in [RFC8402]. The procedures described in this section are equally applicable to both the Segment Routing over MPLS (SR-MPLS) and the Segment Routing over IPv6 (SRv6) data plane, while the data plane-specific considerations are described in Section 7.

This section explains the process by which a protecting router S handles the active segment of a packet upon the failure of its primary outgoing interface for the packet S-F. The failure of the primary outgoing interface may occur due to various triggers, such as link failure, neighbor node failure, and others.

6.1. The Active Segment Is a Node Segment

The active segment MUST be kept on the SR header unchanged and the RL MUST be added. The active segment becomes the first segment after the RL. The way the RL is added depends on the data plane used (see Section 7).

6.2. The Active Segment Is an Adjacency Segment

This section defines the FRR behavior applied by S for any packet received with an active Adjacency segment S-F for which protection was enabled. Since protection has been enabled for the segment S-F and signaled in the IGP (for instance, using protocol extensions from [RFC8667] and [RFC8665]), a calculator of any SR policy utilizing this segment is aware that it may be transiently rerouted out of S-F in the event of an S-F failure.

The simplest approach for link protection of an Adjacency segment S-F is to create an RL that will carry the traffic to F. To do so, one or more "PUSH" operations are performed. If the RL, while avoiding S-F, terminates on F, S only pushes segments of the RL. Otherwise, S pushes a node segment of F, followed by the segments of the RL. For details on the "NEXT" and "PUSH" operations, refer to [RFC8402].

This method, which merges back the traffic at the remote end of the Adjacency segment, has the advantage of keeping as much traffic as possible on the pre-failure path. When SR policies are involved and strict compliance with the policy is required, an end-to-end protection (beyond the scope of this document) should be preferred over the local repair mechanism described above.

Note, however, that when the SR source node is using Traffic Engineering (TE), it will generally not be possible for the PLR to know what post-convergence path will be selected by the source node once it detects the failure, since computation of the TE path is a local matter that depends on constraints that may not be known at the PLR. Therefore, no method applied at the PLR can guarantee protection will follow the post-convergence path.

The case where the active segment is followed by another Adjacency segment is distinguished from the case where it is followed by a node segment. Repair techniques for the respective cases are provided in the following subsections.

6.2.1. Protecting [Adjacency, Adjacency] Segment Lists

If the next segment in the list is an Adjacency segment, then the packet has to be conveyed to F.

To do so, S MUST apply a "NEXT" operation on Adj-SID(S-F) and then one or more "PUSH" operations. If the RL, while avoiding S-F, terminates on F, S only pushes the segments of the RL. Otherwise, S pushes a node segment of F, followed by the segments of the RL. For details on the "NEXT" and "PUSH" operations, refer to [RFC8402].

Upon failure of S-F, a packet reaching S with a segment list matching [adj-sid(S-F), adj-sid(F-M), ...] will thus leave S with a segment list matching [RL(F), node(F), adj-sid(F-M), ...], where RL(F) is the RL for destination F.

6.2.2. Protecting [Adjacency, Node] Segment Lists

If the next segment in the stack is a node segment, say for node T, the segment list on the packet matches [adj-sid(S-F), node(T), ...].

In this case, S MUST apply a "NEXT" operation on the Adjacency segment related to S-F, followed by a "PUSH" of an RL redirecting the traffic to a node Q, whose path to node segment T is not affected by the failure.

Upon failure of S-F, packets reaching S with a segment list matching [adj-sid(S-F), node(T), ...] would leave S with a segment list matching [RL(Q), node(T), ...].

7. Data Plane-Specific Considerations

7.1. MPLS Data Plane Considerations

The MPLS data plane for SR is described in [RFC8660].

The following data plane behaviors apply when creating an RL using an MPLS data plane:

1. If the active segment is a node segment that has been signaled with penultimate hop popping, and the RL ends with an Adjacency segment terminating on the penultimate node of the active segment, then the active segment MUST be popped before pushing the RL.

2. If the active segment is a node segment, but the other conditions in 1. are not met, the active segment MUST be popped and then pushed again with a label value computed according to the Segment Routing Global Block (SRGB) of Q, where Q is the endpoint of the RL. Finally, the RL MUST be pushed.

7.2. SRv6 Data Plane Considerations

SRv6 data plane and programming instructions are described respectively in [RFC8754] and [RFC8986].

The TI-LFA path computation algorithm is the same as in the SR-MPLS data plane. Note, however, that the Adjacency SIDs are typically globally routed. In such a case, there is no need for preceding an Adjacency SID with a Prefix-SID [RFC8402], and the resulting RL is likely shorter.

If the traffic is protected at a Transit Node, then an SRv6 SID list is added on the packet to apply the RL. The addition of the RL follows the head-end behaviors as specified in Section 5 of [RFC8986].

If the traffic is protected at an SR Segment Endpoint Node, first the Segment Endpoint packet processing is executed. Then, the packet is protected as if it were a transit packet.

8. TI-LFA and SR Algorithms

SR allows an operator to bind an algorithm to a Prefix-SID (as defined in [RFC8402]). The algorithm value dictates how the path to the prefix is computed. The SR default algorithm is known as the "Shortest Path" algorithm. The SR default algorithm allows an operator to override the IGP shortest path by using local policies. When TI-LFA uses Node-SIDs associated with the default algorithm, there is no guarantee that the path will be loop-free, as a local policy may have overridden the expected IGP path. As the local policies are defined by the operator, it becomes the responsibility of this operator to ensure that the deployed policies do not affect the TI-LFA deployment. It should be noted that such a situation can already happen today with existing mechanisms such as RLFA.

[RFC9350] defines a Flexible Algorithm framework to be associated with Prefix-SIDs. A Flexible Algorithm allows a user to associate a constrained path to a Prefix-SID rather than using the regular IGP shortest path. An implementation MAY support TI-LFA to protect Node-SIDs associated with a Flexible Algorithm. In such a case, rather than computing the expected post-convergence path based on the regular SPF, an implementation SHOULD use the constrained SPF algorithm bound to the Flexible Algorithm (using the Flexible Algorithm Definition) instead of the regular Dijkstra in all the SPF/reverse SPF computations that are occurring during the TI-LFA computation. This includes the computation of the P-space and Q-space as well as the post-convergence path. Furthermore, the implementation SHOULD only use Node-SIDs/Adj-SIDs bound to the Flexible Algorithm and/or unprotected Adj-SIDs of the regular SPF to build the RL. The use of regular Dijkstra for the TI-LFA computation or for building the repair path using SIDs other than those recommended does not ensure that the traffic going over the TI-LFA repair path during the FRR period is honoring the Flexible Algorithm constraints.

9. Usage of Adjacency Segments in the Repair List

The RL of segments computed by TI-LFA may contain one or more Adjacency segments. An Adjacency segment may be protected or not protected.

```
        S --- R2 --- R3 ---- R4 --- R5 --- D
                 *   |  \   *
                   * |   \ *
                    R7 ** R8
                     *    |
                     *    |
                    R9 -- R10
```

Figure 2

In Figure 2, all the metrics are equal to 1 except R2-R7,R7-R8,R8-R4,R7-R9, which have a metric of 1000. Considering R2 as a PLR to protect against the failure of node R3 for the traffic S->D, the RL computed by R2 will be [adj-sid(R7-R8), adj-sid(R8-R4)], and the outgoing interface will be to R7. If R3 fails, R2 pushes the RL onto the incoming packet to D. During the FRR, if R7-R8 fails and if TI-LFA has picked a protected Adjacency segment for Adj-SID(R7-R8), R7 will push an additional RL onto the packet following the procedures defined in Section 6.

To avoid the possibility of this double FRR activation, an implementation of TI-LFA MAY pick only unprotected Adjacency segments when building the RL. However, it is important to note that FRR in general is intended to protect for a single pre-planned failure. If the failure that happens is worse than expected or multiple failures happen, FRR is not guaranteed to work. In such a case, fast IGP convergence remains important to restore traffic as quickly as possible.

10. Security Considerations

The techniques described in this document are internal functionalities to a router that can guarantee an upper bound on the time taken to restore traffic flow upon the failure of a directly connected link or node. As these techniques steer traffic to the post-convergence path as quickly as possible, this serves to minimize the disruption associated with a local failure, which can be seen as a modest security enhancement. The protection mechanism does not protect external destinations, but rather provides quick restoration for destinations that are internal to a routing domain.

The security considerations described in [RFC5286] and [RFC7490] apply to this document. Similarly, as the solution described in this document is based on SR technology, the reader should be aware of the security considerations related to this technology (see [RFC8402]) and its data plane instantiations (see [RFC8660], [RFC8754], and [RFC8986]). However, this document does not introduce additional security concerns.

11. IANA Considerations

This document has no IANA actions.

12. References

12.1. Normative References

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

[RFC7916] Litkowski, S., Ed., Decraene, B., Filsfils, C., Raza, K., Horneffer, M., and P. Sarkar, "Operational Management of Loop-Free Alternates", RFC 7916, DOI 10.17487/RFC7916, July 2016, <https://www.rfc-editor.org/info/rfc7916>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <https://www.rfc-editor.org/info/rfc8660>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

[RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <https://www.rfc-editor.org/info/rfc8986>.

12.2. Informative References

[IPFRR-TUNNELS] Bryant, S., Filsfils, C., Previdi, S., and M. Shand, "IP Fast Reroute using tunnels", Work in Progress, Internet-Draft, draft-bryant-ipfrr-tunnels-03, 16 November 2007, <https://datatracker.ietf.org/doc/html/draft-bryant-ipfrr-tunnels-03>.

[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, <https://www.rfc-editor.org/info/rfc5286>.

[RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, DOI 10.17487/RFC5714, January 2010, <https://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, <https://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, <https://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, <https://www.rfc-editor.org/info/rfc6976>.

[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, <https://www.rfc-editor.org/info/rfc7490>.

[RFC8333] Litkowski, S., Decraene, B., Filsfils, C., and P. Francois, "Micro-loop Prevention by Introducing a Local Convergence Delay", RFC 8333, DOI 10.17487/RFC8333, March 2018, <https://www.rfc-editor.org/info/rfc8333>.

[RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <https://www.rfc-editor.org/info/rfc8665>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

[RFC9350] Psenak, P., Ed., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", RFC 9350, DOI 10.17487/RFC9350, February 2023, <https://www.rfc-editor.org/info/rfc9350>.

[SR-LOOP] Bashandy, A., Filsfils, C., Litkowski, S., Decraene, B., Francois, P., and P. Psenak, "Loop avoidance using Segment Routing", Work in Progress, Internet-Draft, draft-bashandy-rtgwg-segment-routing-uloop-17, 29 June 2024, <https://datatracker.ietf.org/doc/html/draft-bashandy-rtgwg-segment-routing-uloop-17>.

Appendix A. Advantages of Using the Expected Post-Convergence Path During FRR

[RFC7916] raises several operational considerations when using LFA or RLFA. Section 3 of [RFC7916] presents a case where a high bandwidth link between two core routers is protected through a Provider Edge (PE) router connected with low bandwidth links. In such a case, congestion may happen when the FRR backup path is activated. [RFC7916] introduces a local policy framework to let the operator tuning manually the best alternate election based on its own requirements.

From a network capacity planning point of view, it is often assumed for simplicity that if a link L fails on a particular node X, the bandwidth consumed on L will be spread over some of the remaining links of X. The remaining links to be used are determined by the IGP routing considering that the link L has failed (we assume that the traffic uses the post-convergence path starting from the node X). In Figure 3, we consider a network with all metrics equal to 1 except the metrics on links used by PE1, PE2, and PE3, which are 1000. An easy network capacity planning method is to consider that if the link L (X-B) fails, the traffic actually flowing through L will be spread over the remaining links of X (X-H, X-D, X-A). Considering the IGP metrics, only X-H and X-D can be used in reality to carry the traffic flowing through the link L. As a consequence, the bandwidth of links X-H and X-D is sized according to this rule. We should observe that this capacity planning policy works; however, it is not fully accurate.

In Figure 3, considering that the source of traffic is only from PE1 and PE4, when the link L fails, depending on the convergence speed of the nodes, X may reroute its forwarding entries to the remote PEs onto X-H or X-D; however, in a similar timeframe, PE1 will also reroute a subset of its traffic (the subset destined to PE2) out of its nominal path, reducing the quantity of traffic received by X. The capacity planning rule presented previously has the drawback of oversizing the network; however, it allows for preventing any transient congestion (for example, when X reroutes traffic before PE1 does).

```
              H --- I --- J *
              |           |  *
   PE4        |           |   PE3
      \       | (L)       |  *
      * A --- X --- B --- G *
     *        |           |  *
   PE1        |           |   PE2
     *        |           |  *
      * C --- D --- E --- F *
```

Figure 3

Based on this assumption, in order to facilitate the operation of FRR and limit the implementation of local FRR policies, traffic can be steered by the PLR onto its expected post-convergence path during the FRR phase. In our example, when link L fails, X switches the traffic destined to PE3 and PE2 on the post-convergence paths. This is perfectly in line with the capacity planning rule that was presented before and also in line with the fact that X may converge before PE1 (or any other upstream router) and may spread the X-B traffic onto the post-convergence paths rooted at X.

It should be noted that some networks may have a different capacity planning rule, leading to an allocation of less bandwidth on X-H and X-D links. In such a case, using the post-convergence paths rooted at X during FRR may introduce some congestion on X-H and X-D links. However, it is important to note that a transient congestion may possibly happen even without FRR activated, for instance, when X converges before the upstream routers. Operators are still free to use the policy framework defined in [RFC7916] if the usage of the post-convergence paths rooted at the PLR is not suitable.

Readers should be aware that FRR protection is pre-computing a backup path to protect against a particular type of failure (link, node, or SRLG). When using the post-convergence path as an FRR backup path, the computed post-convergence path is the one considering the failure we are protecting against. This means that FRR is using an expected post-convergence path, and this expected post-convergence path may be actually different from the post-convergence path used if the failure that happened is different from the failure FRR was protecting against. As an example, if the operator has implemented a protection against a node failure, the expected post-convergence path used during FRR will be the one considering that the node has failed. However, even if a single link is failing or a set of links is failing (instead of the full node), the node-protecting post-convergence path will be used. The consequence is that the path used during FRR is not optimal with respect to the failure that has actually occurred.

Another consideration to take into account is as follows: while using the expected post-convergence path for SR traffic using node segments only (for instance, PE to PE traffic using the shortest path) has some advantages, these advantages reduce when SR policies [RFC9256] are involved. A segment list used in an SR policy is computed to obey a set of path constraints defined locally at the head-end or centrally in a controller. TI-LFA cannot be aware of such path constraints, and there is no reason to expect the TI-LFA backup path protecting one segment in that segment list to obey those constraints. When SR policies are used and the operator wants to have a backup path that still follows the policy requirements, this backup path should be computed as part of the SR policy in the ingress node (or central controller), and the SR policy should not rely on local protection. Another option could be to use a Flexible Algorithm [RFC9350] to express the set of constraints and use a single node segment associated with a Flexible Algorithm to reach the destination. When using a node segment associated with a Flexible Algorithm, TI-LFA keeps providing an optimal backup by applying the appropriate set of constraints. The relationship between TI-LFA and the SR algorithm is detailed in Section 8.

Appendix B. Analysis Based on Real Network Topologies

This section presents an analysis performed on real service provider and large enterprise network topologies. The objective of the analysis is to assess the number of SIDs required in an explicit path when the mechanisms described in this document are used to protect against the failure scenarios within the scope of this document. The number of segments described in this section are applicable to instantiating SR over the MPLS forwarding plane.

The measurement below indicates that, for link and local SRLG protection, a repair path of 1 SID or less delivers more than 99% coverage. For node protection, a repair path of 2 SIDs or less yields 99% coverage.

Table 1 below lists the characteristics of the networks used in our measurements. The number of links refers to the number of "bidirectional" links (not directed edges of the graph). The measurements are carried out as follows:

* For each network, the algorithms described in this document are applied to protect all prefixes against link, node, and local SRLG failure.

* For each prefix, the number of SIDs used by the repair path is recorded.

* The percentage of number of SIDs are listed in Tables 2, 3, 4, 5, 6, and 7.

Network	Nodes	Links	Node-to-Link Ratio	SRLG Info?
T1	408	665	1.63	Yes
T2	587	1083	1.84	No
T3	93	401	4.31	Yes

T4	247	393	1.59	Yes
T5	34	96	2.82	Yes
T6	50	78	1.56	No
T7	82	293	3.57	No
T8	35	41	1.17	Yes
T9	177	1371	7.74	Yes

Table 1: Data Set Definition

The rest of this section presents the measurements done on the actual topologies. The conventions that we use are as follows:

* 0 SIDs: The calculated repair path starts with a directly connected neighbor that is also a loop-free alternate; in which case, there is no need to explicitly route the traffic using additional SIDs. This scenario is described in Section 5.1.

* 1 SID: The repair node is a PQ node; in which case, only 1 SID is needed to guarantee a loop-free path. This scenario is covered in Section 5.2.

* 2 or more SIDs: The repair path consists of 2 or more SIDs as described in Sections 5.3 and 5.4. We do not cover the case for 2 SIDs (Section 5.3) separately because there was no granularity in the result. Also, we treat the Node-SID + Adj-SID and Node-SID + Node-SID the same because they do not differ from the data plane point of view.

Tables 2 and 3 below summarize the measurements on the number of SIDs needed for link protection.

Network	0 SIDs	1 SID	2 SIDs	3 SIDs
T1	74.3%	25.3%	0.5%	0.0%
T2	81.1%	18.7%	0.2%	0.0%
T3	95.9%	4.1%	0.1%	0.0%
T4	62.5%	35.7%	1.8%	0.0%
T5	85.7%	14.3%	0.0%	0.0%
T6	81.2%	18.7%	0.0%	0.0%
T7	98.9%	1.1%	0.0%	0.0%
T8	94.1%	5.9%	0.0%	0.0%
T9	98.9%	1.0%	0.0%	0.0%

Table 2: Link Protection (Repair Size Distribution)

Network	0 SIDs	1 SID	2 SIDs	3 SIDs

T1	74.2%	99.5%	99.9%	100%
T2	81.1%	99.8%	100%	100%
T3	95.9%	99.9%	100%	100%
T4	62.5%	98.2%	100%	100%
T5	85.7%	100%	100%	100%
T6	81.2%	99.9%	100%	100%
T7	98.8%	100%	100%	100%
T8	94.1%	100%	100%	100%
T9	98.9%	100%	100%	100%

Table 3: Link Protection (Repair Size Cumulative Distribution)

Tables 4 and 5 summarize the measurements on the number of SIDs needed for local SRLG protection.

Network	0 SIDs	1 SID	2 SIDs	3 SIDs
T1	74.2%	25.3%	0.5%	0.0%
T2	No SRLG information			
T3	93.6%	6.3%	0.0%	0.0%
T4	62.5%	35.6%	1.8%	0.0%
T5	83.1%	16.8%	0.0%	0.0%
T6	No SRLG information			
T7	No SRLG information			
T8	85.2%	14.8%	0.0%	0.0%
T9	98.9%	1.1%	0.0%	0.0%

Table 4: Local SRLG Protection (Repair Size Distribution)

Network	0 SIDs	1 SID	2 SIDs	3 SIDs
T1	74.2%	99.5%	99.9%	100%
T2	No SRLG information			
T3	93.6%	99.9%	100%	0.0%
T4	62.5%	98.2%	100%	100%
T5	83.1%	100%	100%	100%
T6	No SRLG information			
T7	No SRLG information			

T8	85.2%	100%	100%	100%
T9	98.9%	100%	100%	100%

Table 5: Local SRLG Protection (Repair Size Cumulative Distribution)

The remaining two tables summarize the measurements on the number of SIDs needed for node protection.

Network	0 SIDs	1 SID	2 SIDs	3 SIDs	4 SIDs
T1	49.8%	47.9%	2.1%	0.1%	0.0%
T2	36.5%	59.6%	3.6%	0.2%	0.0%
T3	73.3%	25.6%	1.1%	0.0%	0.0%
T4	36.1%	57.3%	6.3%	0.2%	0.0%
T5	73.2%	26.8%	0.0%	0.0%	0.0%
T6	78.3%	21.3%	0.3%	0.0%	0.0%
T7	66.1%	32.8%	1.1%	0.0%	0.0%
T8	59.7%	40.2%	0.0%	0.0%	0.0%
T9	98.9%	1.0%	0.0%	0.0%	0.0%

Table 6: Node Protection (Repair Size Distribution)

Network	0 SIDs	1 SID	2 SIDs	3 SIDs	4 SIDs
T1	49.7%	97.6%	99.8%	99.9%	100%
T2	36.5%	96.1%	99.7%	99.9%	100%
T3	73.3%	98.9%	99.9%	100%	100%
T4	36.1%	93.4%	99.8%	99.9%	100%
T5	73.2%	100%	100%	100%	100%
T6	78.4%	99.7%	100%	100%	100%
T7	66.1%	98.9%	100%	100%	100%
T8	59.7%	100%	100%	100%	100%
T9	98.9%	100%	100%	100%	100%

Table 7: Node Protection (Repair Size Cumulative Distribution)

Acknowledgments

The authors would like to thank Les Ginsberg, Stewart Bryant, Alexander Vainsthein, Chris Bowers, Shraddha Hedge, Wes Hardaker, Gunter Van de Velde, and John Scudder for their valuable comments.

Contributors

In addition to the authors listed on the front page, the following co-authors have also contributed to this document:

Francois Clad
Cisco Systems

Pablo Camarillo
Cisco Systems

Authors' Addresses

Ahmed Bashandy
HPE
United States of America
Email: abashandy.ietf@gmail.com

Stephane Litkowski
Cisco Systems
France
Email: slitkows@cisco.com

Clarence Filsfils
Cisco Systems
Brussels
Belgium
Email: cfilsfil@cisco.com

Pierre Francois
INSA Lyon
Email: pierre.francois@insa-lyon.fr

Bruno Decraene
Orange
France
Email: bruno.decraene@orange.com

Daniel Voyer
Bell Canada
Canada
Email: daniel.voyer@bell.ca