

Routing Area Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 22, 2013

S. Litkowski
B. Decraene
Orange
C. Filsfils
K. Raza
Cisco Systems
February 18, 2013

Operational management of Loop Free Alternates
draft-litkowski-rtgwg-lfa-manageability-01

Abstract

Loop Free Alternates (LFA), as defined in [RFC 5286](#) is an IP Fast ReRoute (IP FRR) mechanism enabling traffic protection for IP traffic (and MPLS LDP traffic by extension). Following first deployment experiences, this document provides operational feedback on LFA, highlights some limitations, and proposes a set of refinements to address those limitations. It also proposes required management specifications.

This proposal is also applicable to remote LFA solution.

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](#)].

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 22, 2013.

Copyright Notice

Internet-Draft

LFA manageability

February 2013

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](http://trustee.ietf.org/license-info) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
2.	Operational issues with default LFA tie breakers	3
2.1.	Case 1: Edge router protecting core failures	4
2.2.	Case 2: Edge router chosen to protect core failures while core LFA exists	5
2.3.	Case 3: suboptimal core alternate choice	6
2.4.	Case 4: ISIS overload bit on LFA computing node	7
3.	Configuration requirements	7
3.1.	LFA enabling/disabling scope	7
3.2.	Policy based LFA selection	8
3.2.1.	Mandatory criteria	8
3.2.2.	Enhanced criteria	9
4.	Operational aspects	13
4.1.	ISIS overload bit on LFA computing node	13
4.2.	Manual triggering of FRR	14
4.3.	Required local information	14
4.4.	Coverage monitoring	15
5.	Security Considerations	15
6.	Contributors	15
7.	Acknowledgements	15
8.	IANA Considerations	15
9.	References	16
9.1.	Normative References	16
9.2.	Informative References	16
	Authors' Addresses	17

Internet-Draft

LFA manageability

February 2013

1. Introduction

Following the first deployments of Loop Free Alternates (LFA), this document provides feedback to the community about the management of LFA.

[Section 2](#) provides real uses cases illustrating some limitations and suboptimal behavior.

[Section 3](#) proposes requirements for activation granularity and policy based selection of the alternate.

[Section 4](#) express requirements for the operational management of LFA.

2. Operational issues with default LFA tie breakers

[RFC5286] introduces the notion of tie breakers when selecting the LFA among multiple candidate alternate next-hops. When multiple LFA exist, [RFC 5286](#) has favored the selection of the LFA providing the best coverage of the failure cases. While this is indeed a goal, this is one among multiple and in some deployment this lead to the selection of a suboptimal LFA. The following sections details real use cases of such limitations.

Note that the use case of per-prefix LFA is assumed throughout this analysis.

[2.1.](#) Case 1: Edge router protecting core failures

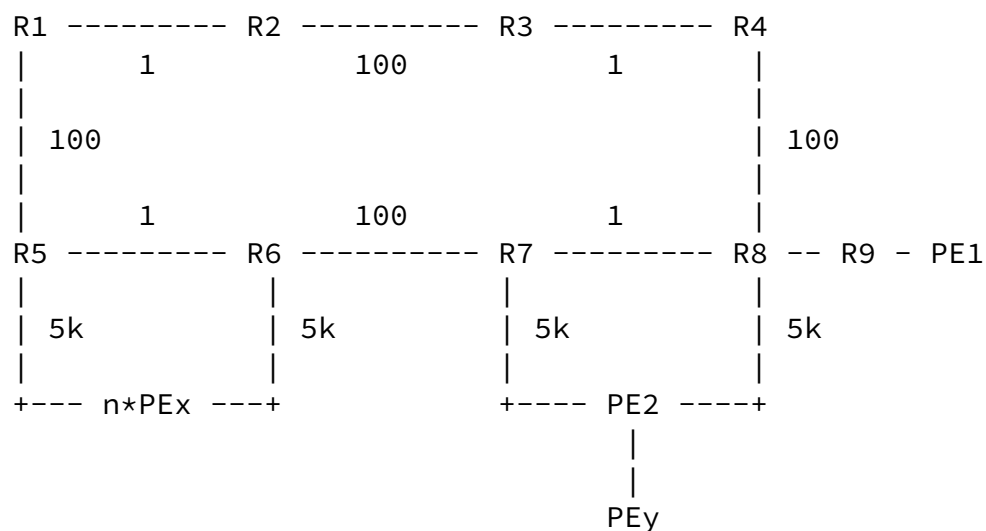


Figure 1

Rx routers are core routers using $n \times 10G$ links. PEs are connected using links with lower bandwidth.

In figure 1, let us consider the traffic flowing from PE1 to PEx. The nominal path is R9-R8-R7-R6-PEx. Let us consider the failure of link R7-R8. For R8, R4 is not an LFA and the only available LFA is PE2.

When the core link R8-R7 fails, R8 switches all traffic destined to all the PEx towards the edge node PE2. Hence an edge node and edge

links are used to protect the failure of a core link. Typically, edge links have less capacity than core links and congestion may occur on PE2 links. Note that although PE2 was not directly affected by the failure, its links become congested and its traffic will suffer from the congestion.

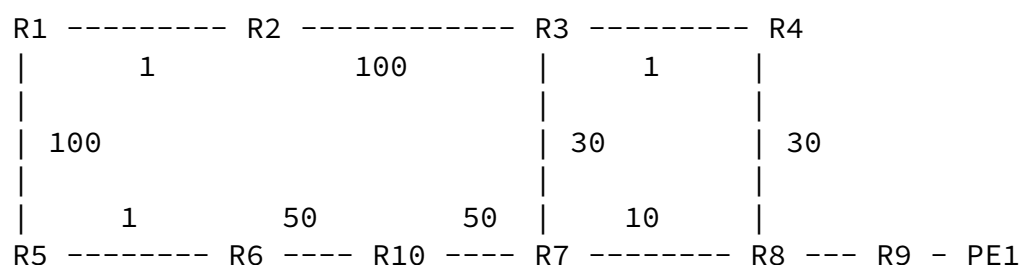
In summary, in case of failure, the impact on customer traffic is:

- o From PE2 point of view :
 - * without LFA: no impact
 - * with LFA: traffic is partially dropped (but possibly prioritized by a QoS mechanism). It must be highlighted that in such situation, traffic not affected by the failure may be affected by the congestion.

- o From R8 point of view:
 - * without LFA: traffic is totally dropped until convergence occurs.
 - * with LFA: traffic is partially dropped (but possibly prioritized by a QoS mechanism).

Besides the congestion aspects of using an Edge router as an alternate to protect a core failure, a service provider may consider this as a bad routing design and would like to prevent it.

[2.2.](#) Case 2: Edge router chosen to protect core failures while core LFA exists



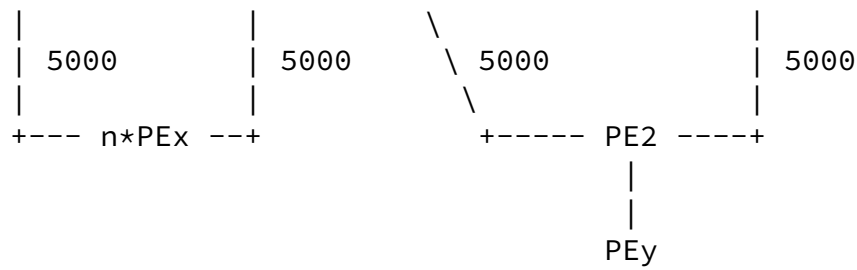
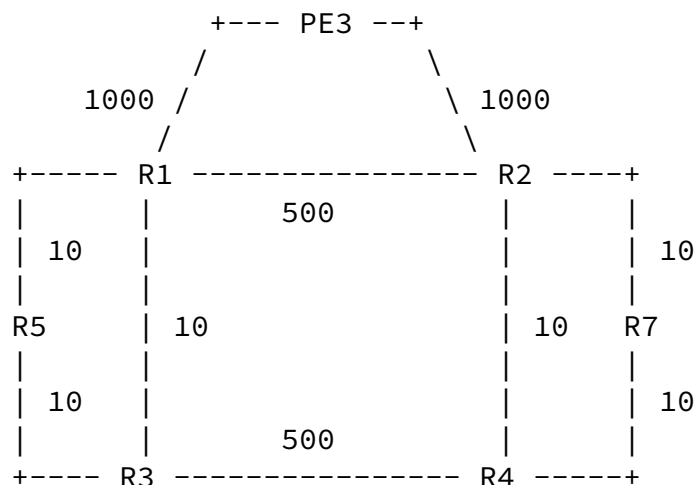


Figure 2

Rx routers are core routers meshed with $n \times 10G$ links. PEs are meshed using links with lower bandwidth.

In the figure 2, let us consider the traffic coming from PE1 to PEx. Nominal path is R9-R8-R7-R6-PEx. Let us consider the failure of the link R7-R8. For R8, R4 is a link-protecting LFA and PE2 is a node-protecting LFA. PE2 is chosen as best LFA due to its better protection type. Just like in case 1, this may lead to congestion on PE2 links upon LFA activation.

[2.3.](#) Case 3: suboptimal core alternate choice



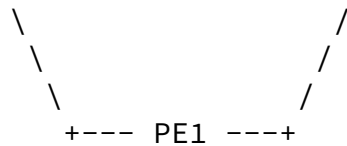


Figure 3

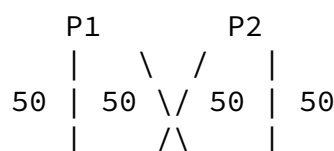
Rx routers are core routers. R1-R2 and R3-R4 links are 1G links. All others inter Rx links are 10G links.

In the figure above, let us consider the failure of link R1-R3. For destination PE3, R3 has two possible alternates:

- o R4, which is node-protecting
- o R5, which is link-protecting

R4 is chosen as best LFA due to its better protection type. However, it may not be desirable to use R4 for bandwidth capacity reason. A service provider may prefer to use high bandwidth links as preferred LFA. In this example, preferring shortest path over protection type may achieve the expected behavior, but in cases where metric are not reflecting bandwidth, it would not work and some other criteria would need to be involved when selecting the best LFA.

[2.4.](#) Case 4: ISIS overload bit on LFA computing node



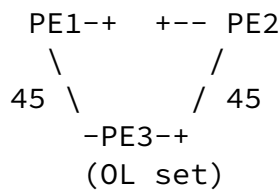


Figure 4

In the figure above, PE3 has its overload bit set (permanently, for design reason) and wants to protect traffic using LFA for destination PE2.

On PE3, the loopfree condition is not satisfied : $100 \nless 45 + 45$. PE1 is thus not considered as an LFA. However thanks to the overload bit set on PE3, we know that PE1 is loopfree so PE1 is an LFA to reach PE2.

In case of overload condition set on a node, LFA behavior must be clarified.

[3.](#) Configuration requirements

Controlling best alternate and LFA activation granularity is a requirement for Service Providers. This section defines configuration requirements for LFA.

[3.1.](#) LFA enabling/disabling scope

The granularity of LFA activation should be controlled (as alternate nexthop consume memory in forwarding plane).

An implementation of LFA SHOULD allow its activation with the following criteria:

- o Per address-family : ipv4 unicast, ipv6 unicast, LDP IPv4 unicast, LDP IPv6 unicast ...
- o Per routing context : VRF, virtual/logical router, global routing table, ...

- o Per interface

- o Per protocol instance, topology, area
- o Per prefixes: prefix protection SHOULD have a better priority compared to interface protection. This means that if a specific prefix must be protected due to a configuration request, LFA must be computed and installed for this prefix even if the primary outgoing interface is not configured for protection.

3.2. Policy based LFA selection

When multiple alternates exist, LFA selection algorithm is based on tie breakers. Current tie breakers do not provide sufficient control on how the best alternate is chosen. This document proposes an enhanced tie breaker allowing service providers to manage all specific cases:

1. An implementation of LFA SHOULD support policy-based decision for determining the best LFA.
2. Policy based decision SHOULD be based on multiple criterions, with each criteria having a level of preference.
3. If the defined policy does not permit to determine a unique best LFA, an implementation SHOULD pick only one based on its own decision, as a default behavior. An implementation SHOULD also support election of multiple LFAs, for loadbalancing purposes.
4. Policy SHOULD be applicable to a protected interface or to a specific set of destinations. In case of application on the protected interface, all destinations primarily routed on this interface SHOULD use the interface policy.
5. It is an implementation choice to reevaluate policy dynamically or not (in case of policy change). If a dynamic approach is chosen, the implementation SHOULD recompute the best LFAs and reinstall them in FIB, without service disruption. If a non-dynamic approach is chosen, the policy would be taken into account upon the next IGP event. In this case, the implementation SHOULD support a command to manually force the recomputation/reinstallation of LFAs.

3.2.1. Mandatory criteria

An implementation of LFA MUST support the following criteria:

- o Non candidate link: A link marked as "non candidate" will never be used as LFA.
- o A primary nexthop being protected by another primary nexthop of the same prefix (ECMP case).
- o Type of protection provided by the alternate: link protection, node protection. In case of node protection preference, an implementation SHOULD support fallback to link protection if node protection is not available.
- o Shortest path: lowest IGP metric used to reach the destination.
- o SRLG (as defined in [\[RFC5286\] Section 3](#)).

[3.2.2](#). Enhanced criteria

An implementation of LFA SHOULD support the following enhanced criteria:

- o Downstreamness of a neighbor : preference of a downstream path over a non downstream path SHOULD be configurable.
- o Link coloring with : include, exclude and preference based system.
- o Link Bandwidth.
- o Neighbor preference.
- o Neighbor type: link or tunnel alternate. This means that user may change preference between link alternate or tunnel alternate (link preferred over tunnel, or considered as equal).

[3.2.2.1](#). Link coloring

Link coloring is a powerful system to control the choice of alternates. Protecting interfaces are tagged with colors. Protected interfaces are configured to include some colors with a preference level, and exclude others.

Link color information SHOULD be signalled in the IGP. How signalling is done is out of scope of the document but it may be useful to reuse existing admin-groups from traffic-engineering extensions.

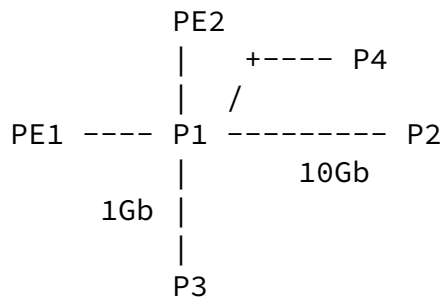


Figure 5

Example : P1 router is connected to three P routers and two PEs.

P1 is configured to protect the P1-P4 link. We assume that given the topology, all neighbors are candidate LFA. We would like to enforce a policy in the network where only a core router may protect against the failure of a core link, and where high capacity links are preferred.

In this example, we can use the proposed link coloring by:

- o Marking PEs links with color RED
- o Marking 10Gb CORE link with color BLUE
- o Marking 1Gb CORE link with color YELLOW
- o Configured the protected interface P1->P4 with :
 - * Include BLUE, preference 200
 - * Include YELLOW, preference 100
 - * Exclude RED

Using this, PE links will never be used to protect against P1-P4 link failure and 10Gb link will be preferred.

The main advantage of this solution is that it can easily be

duplicated on other interfaces and other nodes without change. A Service Provider has only to define the color system (associate color with a significance), as it is done already for TE affinities or BGP communities.

An implementation of link coloring:

- o SHOULD support multiple include and exclude colors on a single protected interface.

- o SHOULD provide a level of preference between included colors.
- o SHOULD support multiple colors configuration on a single protecting interface.

[3.2.2.2](#). Bandwidth

As mentionned in previous sections, not taking into account bandwidth of an alternate could lead to congestion during FRR activation. We propose to base the bandwidth criteria on the link speed information for the following reason :

- o if a router S has a set of X destinations primarily forwarded to N, using per prefix LFA may lead to have a subset of X protected by a neighbor N1, another subset by N2, another subset by Nx ...
- o S is not aware about traffic flows to each destination and is not able to evaluate how much traffic will be sent to N1,N2, ... Nx in case of FRR activation.

Based on this, it is not useful to gather available bandwidth on alternate paths, as the router does not know how much bandwidth it requires for protection. The proposed link speed approach provides a good approximation with a small cost as information is easily available.

The bandwidth criteria of the policy framework SHOULD work in two ways :

- o PRUNE : exclude a LFA if link speed to reach it is lower than the link speed of the primary nexthop interface.

- o PREFER : prefer a LFA based on his bandwidth to reach it compared to the link speed of the primary nexthop interface.

3.2.2.3. Neighbor preference

Rather than tagging interface on each node (using link color) to identify neighbor node type (as example), it would be helpful if routers could be identified in the IGP. This would permit a grouped processing on multiple nodes. Some existing IGP extension like SUB-TLV 1 of TLV 135 may be useful for this purpose. As an implementation must be able to exclude some specific neighbors (see mandatory criterions), an implementation :

- o SHOULD be able to give a preference to specific neighbor.

- o SHOULD be able to give a preference to a group of neighbor.
- o SHOULD be able to exclude a group of neighbor.

A specific neighbor may be identified by its interface or IP address and group of neighbors may be identified by a marker like SUB-TLV1 in TLV135. As multiple prefixes may be present in TLVs 135, an heuristic is required to choose the appropriate one that will identify the neighbor and will transport the tag associated with the neighbor preference.

We propose the following algorithm to select the prefix :

1. Select the prefix in TLV#135 that is equal to TLV#134 value (Router ID) and prefix length is 32.
2. Select the prefix in TLV#135 that is equal to TLV#132 value (IP Addresses) and prefix length is 32, it must be noted that TLV#132 may transport multiple addresses and so multiple matches may happen.
3. If multiple prefixes are matching TLV#132 values, choose the highest one.

Consider the following network:

alternates may not always provide an optimal routing path and it may be preferable to select a remote alternate over a link alternate. The usage of tunnels to extend LFA coverage is described in [\[I-D.ietf-rtgwg-remote-lfa\]](#) and [\[I-D.litkowski-rtgwg-lfa-rsvpte-cooperation\]](#).

In figure 1, there is no core alternate for R8 to reach PEs located behind R6, so R8 is using PE2 as alternate, which may generate congestion when FRR is activated. Instead, we could have a remote core alternate for R8 to protect PEs destinations. For example, a tunnel from R8 to R3 would ensure a LFA protection without any impact.

There is a requirement to be able to compare remote alternates (reachable through a tunnel) to link alternates (a remote alternate may provide a better protection than a link alternate based on service provider's criteria). Policy will associate a preference to each alternate whatever their type (link or remote) and will elect the best one.

[4.](#) Operational aspects

[4.1.](#) ISIS overload bit on LFA computing node

In [\[RFC5286\], Section 3.5](#), the setting of the overload bit condition in LFA computation is only taken into account for the case where a neighbor has the overload bit set.

In addition to [RFC 5286](#) inequality 1 Loop-Free Criterion

(Distance_opt(N, D) < Distance_opt(N, S) + Distance_opt(S, D)), the IS-IS overload bit of the LFA calculating neighbor (S) SHOULD be taken into account. Indeed, if it has the overload bit set, no neighbor will loop back to traffic to itself.

[4.2.](#) Manual triggering of FRR

Service providers often use using manual link shutdown (using router CLI) to perform some network changes/tests. Especially testing or troubleshooting FRR requires to perform the manual shutdown on the remote end of the link as generally a local shutdown would not

trigger FRR. To enhance such situation, an implementation SHOULD support triggering/activating LFA Fast Reroute for a given link when a manual shutdown is done.

[4.3.](#) Required local information

LFA introduction requires some enhancement in standard routing information provided by implementations. Moreover, due to the non 100% coverage, coverage informations is also required.

Hence an implementation :

- o MUST be able to display, for every prefixes, the primary nexthop as well as the alternate nexthop information.
- o MUST provide coverage information per activation domain of LFA (area, level, topology, instance, virtual router, address family ...).
- o MUST provide number of protected prefixes as well as non protected prefixes globally.
- o SHOULD provide number of protected prefixes as well as non protected prefixes per link.
- o MAY provide number of protected prefixes as well as non protected prefixes per priority if implementation supports prefix-priority insertion in RIB/FIB.
- o SHOULD provide a reason for chosing an alternate (policy and criteria) and for excluding an alternate.
- o SHOULD provide the list of non protected prefixes and the reason why they are not protected (no protection required or no alternate available).

[4.4.](#) Coverage monitoring

It is pretty easy to evaluate the coverage of a network in a nominal situation, but topology changes may change the coverage. In some

situations, the network may no longer be able to provide the required level of protection. Hence, it becomes very important for service providers to get alerted about changes of coverage.

An implementation SHOULD :

- o provide an alert system if total coverage (for a node) is below a defined threshold or comes back to a normal situation.
- o provide an alert system if coverage of a specific link is below a defined threshold or comes back to a normal situation.

An implementation MAY :

- o provide an alert system if a specific destination is not protected anymore or when protection comes back up for this destination

Although the procedures for providing alerts are beyond the scope of this document, we recommend that implementations consider standard and well used mechanisms like syslog or SNMP traps.

[5.](#) Security Considerations

This document does not introduce any change in security consideration compared to [[RFC5286](#)].

[6.](#) Contributors

Significant contributions were made by Pierre Francois, Hannes Gredler and Mustapha Aissaoui which the authors would like to acknowledge.

[7.](#) Acknowledgements

[8.](#) IANA Considerations

This document has no action for IANA.

[9.](#) References

[9.1.](#) Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC5286] Atlas, A. and A. Zinin, "Basic Specification for IP Fast Reroute: Loop-Free Alternates", [RFC 5286](#), September 2008.

[9.2.](#) Informative References

- [I-D.ietf-rtgwg-remote-lfa]
Bryant, S., Filsfils, C., Previdi, S., Shand, M., and S. Ning, "Remote LFA FRR", [draft-ietf-rtgwg-remote-lfa-01](#) (work in progress), December 2012.
- [I-D.litkowski-rtgwg-lfa-rsvp-te-cooperation]
Litkowski, S., Decraene, B., Filsfils, C., and K. Raza, "Interactions between LFA and RSVP-TE", [draft-litkowski-rtgwg-lfa-rsvp-te-cooperation-01](#) (work in progress), February 2013.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", [RFC 3630](#), September 2003.
- [RFC3906] Shen, N. and H. Smit, "Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels", [RFC 3906](#), October 2004.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", [RFC 4090](#), May 2005.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", [RFC 5305](#), October 2008.
- [RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", [RFC 5714](#), January 2010.
- [RFC5715] Shand, M. and S. Bryant, "A Framework for Loop-Free Convergence", [RFC 5715](#), January 2010.
- [RFC6571] Filsfils, C., Francois, P., Shand, M., Decraene, B., Uttaro, J., Leymann, N., and M. Horneffer, "Loop-Free Alternate (LFA) Applicability in Service Provider (SP) Networks", [RFC 6571](#), June 2012.

Internet-Draft

LFA manageability

February 2013

Authors' Addresses

Stephane Litkowski
Orange

Email: stephane.litkowski@orange.com

Bruno Decraene
Orange

Email: bruno.decraene@orange.com

Clarence Filsfils
Cisco Systems

Email: cfilsfil@cisco.com

Kamran Raza
Cisco Systems

Email: skraza@cisco.com

Litkowski, et al.

Expires August 22, 2013

[Page 17]