Network Working Group
Internet Draft
Intended status: Standard Track
Expires: June 2019

A. Bashandy
Arrcus
C. Filsfils
Cisco Systems
Bruno Decraene
Stephane Litkowski
Orange
Pierre Francois
INSA Lyon
D. Voyer
Bell Canada
Francois Clad
Pablo Camarillo
Cisco Systems

December 3, 2018

Topology Independent Fast Reroute using Segment Routing
draft-ietf-rtgwg-segment-routing-ti-lfa-00

Abstract

This document presents Topology Independent Loop-free Alternate Fast Re-route (TI-LFA), aimed at providing protection of node and adjacency segments within the Segment Routing (SR) framework. This Fast Re-route (FRR) behavior builds on proven IP-FRR concepts being LFAs, remote LFAs (RLFA), and remote LFAs with directed forwarding (DLFA). It extends these concepts to provide guaranteed coverage in any IGP network. A key aspect of TI-LFA is the FRR path selection approach establishing protection over post-convergence paths from the point of local repair, dramatically reducing the operational need to control the tie-breaks among various FRR options.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

This Internet-Draft will expire on June 3, 2019.

Copyright Notice

Table of Contents

1. Introduction

Segment Routing aims at supporting services with tight SLA guarantees [1]. By relying on segment routing this document provides a local repair mechanism for standard IGP shortest path capable of restoring end-to-end connectivity in the case of a sudden directly connected failure of a network component. Non-SR mechanisms for local repair are beyond the scope of this document. Non-local failures are addressed in a separate document [6].

The term topology independent (Ti) refers to the ability to provide a loop free backup path irrespective of the topologies prior the failure and after the failure.

For each destination in the network, TI-LFA prepares a data-plane switch-over 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 local 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 has failed. In local 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).

Protection techniques outlined in this document are limited to protecting links, nodes, and local SRLGs that are within a routing

domain. Protecting domain exit routers and/or links attached to another routing domains are beyond the scope of this document

Using segment routing, there is no need to establish TLDP sessions with remote nodes in order to take advantage of the applicability of remote LFAs (RLFA) [4][5] or remote LFAs with directed forwarding (DLFA)[2]. As a result, preferring LFAs over RLFAs or DLFAs, as well as minimizing the number of RLFA or DLFA repair nodes is not required

Using SR, there is no need to create state in the network in order to enforce an explicit FRR path thereby relieving the nodes from the extra state and the operator from having to deploy an extra protocol just to enhance FRR coverage.

The FRR behavior suggested in this document tailors the repair paths over the post-convergence path from the PLR to the protected destination, given the enabled protection mode for the interface. Using the post-convergence path in TI-LFA resolves some of operational issues with LFA selection that are mentioned in Section 3 of [5] (e.g. using PE routers to protect against core failures, or selecting links with low BW while links with high BW are available), because these issues presumably have been taken care of by the network operator as part of its original network engineering. Hence traffic that permanently uses the PLR after the failure achieves maximum benefits. Traffic that does not use the PLR prior to and after the failure remains unaffected. Traffic that temporarily continues to use the PLR after the failure benefits from the quick switching to the backup path by minimizing traffic loss until remote node(s) reacts.

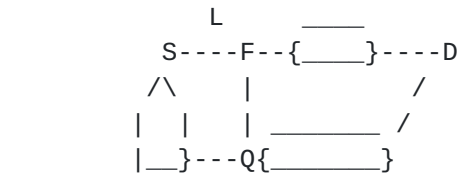

Figure 1 TI-LFA Protection

We use Figure 1 to illustrate the TI-LFA approach.

The Point of Local Repair (PLR), S, needs to find a node Q (a repair node) that is capable of safely forwarding the traffic to a destination D affected by the failure of the protected link L, a set of adjacent links including L (local SRLG), or the node F itself. The PLR also needs to find a way to reach Q without being affected by the convergence state of the nodes over the paths it wants to use to reach Q.

In Section 2 we define the main notations used in the document. They are in line with [2].

In Section 3, we suggest 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.

Using the properties defined in Section 3, Section 4 describes how to compute protection lists that encode a loopfree post-convergence path towards the destination.

Section 5 defines the segment operations to be applied by the PLR to ensure consistency with the forwarding state of the repair node.

By applying the algorithms specified in this document to actual service providers and large enterprise networks, we provide real life measurements for the number of SIDs used by repair paths. Section 6 summarizes these measurements.

1.1. Conventions used in this document

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 RFC-2119

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

2. Terminology

We define the main notations used in this document as the following.

We refer to "old" and "new" topologies as the LSDB state before and after the considered failure.

SPT_old(R) is the Shortest Path Tree rooted at node R in the initial state of the network.

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

Dist_old(A,B) is the shortest distance from node A to node B in SPT_old(A).

Dist_new(A,B, X) is the shortest distance from node A to node B in SPT_new(A,X).

PLR stands for "Point of Local Repair". It is the router that applies fast traffic restoration after detecting failure in a directly attached link, set of links, and/or node.

Similar to [4], we use the concept of P-Space and Q-Space for TI-LFA.

The P-Space P(R,X) of a node R w.r.t. a resource X (e.g. a link S-F, a node F, or a local SRLG) is the set of nodes that are reachable from R without passing through X. It is the set of nodes that are not downstream of X in SPT_old(R).

The Extended P-Space P'(R,X) of a node R w.r.t. a resource X is the set of nodes that are reachable from R or a neighbor of R, without passing through X.

The Q-Space Q(D,X) of a destination node D w.r.t. a resource X is the set of nodes which do not use X to reach D in the initial state of the network. In other words, it is the set of nodes which have D in their P-Space w.r.t. S-F, F, or a set of links adjacent to S).

A symmetric network is a network such that the IGP metric of each link is the same in both directions of the link.

3. Intersecting P-Space and Q-Space with post-convergence paths

In this section, we suggest to determine the P-Space and Q-Space properties of the nodes along the post-convergence paths from the PLR to the protected destination and compute an SR-based explicit path from P to Q when they are not adjacent. Such properties will be used in Section 4 to compute the TI-LFA repair list.

3.1. P-Space property computation for a resource X

A node N is in P(R, X) if it is not downstream of X in SPT_old(R). X can be a link, a node, or a set of links adjacent to the PLR. A node N is in P'(R,X) if it is not downstream of X in SPT_old(N), for at least one neighbor N of R.

3.2. Q-Space property computation for a link S-F, over post-convergence paths

We want to determine which nodes on the post-convergence path from the PLR to the destination D are in the Q-Space of destination D w.r.t. link S-F.

This can be found by intersecting the post-convergence path to D, assuming the failure of S-F, with Q(D, S-F).

3.3. Q-Space property computation for a set of links adjacent to S, over post-convergence paths

We want to determine which nodes on the post-convergence path from the PLR to the destination D are in the Q-Space of destination D w.r.t. a set of links adjacent to S (S being the PLR). That is, we aim to find the set of nodes on the post-convergence path that use none of the members of the protected set of links, to reach D.

This can be found by intersecting the post-convergence path to D, assuming the failure of the set of links, with the intersection among Q(D, S->X) for all S->X belonging to the set of links.

3.4. Q-Space property computation for a node F, over post-convergence paths

We want to determine which nodes on the post-convergence from the PLR to the destination D are in the Q-Space of destination D w.r.t. node F.

This can be found by intersecting the post-convergence path to D, assuming the failure of F, with Q(D, F).

4. TI-LFA Repair Tunnel

The TI-LFA repair tunnel consists of an outgoing interface and a list of segments (repair list) to insert on the SR header. The repair list encodes the explicit post-convergence path to the destination, which avoids the protected resource X and, at the same time, is guaranteed to be loop free irrespective of the state of FIBs along the nodes belonging to the explicit path. Thus there is no need for any co-ordination or message exchange between the PLR and any other router in the network.

The TI-LFA repair tunnel 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 P to Q when these nodes are not adjacent along the post convergence path. The TI-LFA repair list is expressed generally as (Node_SID(P), EP(P, Q)).

Most often, the TI-LFA repair list has a simpler form, as described in the following sections. Section 6 provides statistics for the number of SIDs in the explicit path to protect against various failures.

4.1. The repair node is a direct neighbor

When the repair node is a direct neighbor, 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.

4.2. The repair node is a PQ node

When the repair node is in P(S,X), the repair list is made of a single node segment to the repair node.

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

4.3. The repair is a Q node, neighbor of the last P node

When the repair node is adjacent to P(S,X), the repair list is made of two segments: A node segment to the adjacent P node, and an adjacency segment from that node to the repair node.

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

4.4. 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. However, the PLR can perform additional computations to compute a list of segments that represent a loopfree path from P to Q.

5. Protecting segments

In this section, we explain how a protecting router S processes the active segment of a packet upon the failure of its primary outgoing interface for the packet, S-F.

The behavior depends on the type of active segment to be protected.

5.1. The active segment is a node segment

The active segment is kept on the SR header, unchanged (1). The repair list is inserted at the head of the list. The active segment becomes the first segment of the inserted repair list.

Note (1): If SR-MPLS is being used and the SRGB at the repair node is different from the SRGB at the PLR, then the active segment MUST be updated to fit the SRGB of the repair node.

In Section 5.3, we describe the node protection behavior of PLR S, for the specific case where the active segment is a prefix segment for the neighbor F itself.

5.2. The active segment is an adjacency segment

We define hereafter the FRR behavior applied by S for any packet received with an active adjacency segment S-F for which protection was enabled. We distinguish the case where this active segment is followed by another adjacency segment from the case where it is followed by a node segment.

5.2.1. 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 applies a "NEXT" operation on Adj(S-F) and then two consecutive "PUSH" operations: first it pushes a node segment for F, and then it pushes a protection list allowing to reach F while bypassing S-F. For details on the "NEXT" and "PUSH" operations, refer to [7].

Upon failure of S-F, a packet reaching S with a segment list matching [adj(S-F),adj(M),...] will thus leave S with a segment list matching [RT(F),node(F),adj(M)], where RT(F) is the repair tunnel for destination F. If MPLS forwarding plane is used, then Note(1) from Section 5.1 applies here. Hence MPLS label representing Node(F) MUST be calculated according to the exit point of the repair tunnel "RT(F)"

In Section 5.3.2, we describe the TI-LFA behavior of PLR S when node protection is applied and the two first segments are Adjacency Segments.

5.2.2. 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(S-F),node(T),...].

A first solution would consist in steering the packet back to F while avoiding S-F. To do so, S applies a "NEXT" operation on Adj(S-F) and then two consecutive "PUSH" operations: first it pushes a node segment for F, and then it pushes a repair list allowing to reach F while bypassing S-F.

Upon failure of S-F, a packet reaching S with a segment list matching [adj(S-F),node(T),...] will thus leave S with a segment list matching [RT(F),node(F),node(T)]. Again if MPLS forwarding plane is used, then Note(1) from Section 5.1 applies and the label representing the node(F) MUST be calculated according to the SRGB of the last node in the repair tunnel RT(F).

Another solution is to not steer the packet back via F but rather follow the new shortest path to T. In this case, S just needs to apply a "NEXT" operation on the Adjacency segment related to S-F, and push a repair list 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(L), node(T), ...], would leave S with a segment list matching [RT(Q),node(T), ...]. Note that this second behavior is the one followed for node protection, as described in Section 5.3.1.

Just like the first solution above, if MPLS forwarding plane is used, then Note(1) from Section 5.1 applies. Hence the label corresponding to Node(T) MUST be calculated according to the SRGB of node Q.

5.3. Protecting SR policy midpoints against node failure

In this section, we describe the behavior of a node S configured to interpret the failure of link S->F as the node failure of F, in the specific case where the active segment of the packet received by S is a Prefix SID of F represented as "F"), or an Adjacency SID for the link S-F (represented as "S->F").

5.3.1. Protecting {F, T, D} or {S->F, T, D}

This section describes the protection behavior of S when all of the following conditions are true:

1. the active segment is a prefix SID for a neighbor F, or an adjacency segment S->F

2. the primary interface used to forward the packet failed

3. the segment following the active segment is a prefix SID (for node T)

4. node protection is active for that interface.

The TILFA Node FRR behavior becomes equivalent to:

1. Pop; the segment F or S->F is removed

2. Confirm that the next segment is in the SRGB of F, meaning that the next segment is a prefix segment, e.g. for node T

3. Identify T (as per the SRGB of F)

4. Pop the next segment and push T's segment based on the SRGB of node "S".

5. forward the packet according to T.

5.3.2. Protecting {F, F->T, D} or {S->F, F->T, D}

This section describes the protection behavior of S when all of the following conditions are true:

1. the active segment is a prefix SID for a neighbor F, or an adjacency segment S->F

2. the primary interface used to forward the packet failed

3. the segment following the active segment is an adjacency SID (F->T)

4. node protection is active for that interface.

The TILFA Node FRR behavior becomes equivalent to:

1. Pop; the segment F or S->F is removed

2. Confirm that the next segment is an adjacency SID of F, say F->T

3. Identify T (as per the set of Adjacency Segments of F)

4. Pop the next segment and push T's segment based on the SRGB of the node "S"

5. forward the packet according to T.

It is noteworthy to mention that node "S" in the procedures described in Sections 5.3.1 and 5.3.2 can always determine whether the segment after popping the top segment is an adjacency SID or a prefix-SID of the next-hop "F" as follows:

1. In a link state environment, the node "S" knows the SRGB and the adj-SIDs of the neighboring node "F"

2. If the new segment after popping the top segment is within the SRGB or the adj-SIDs of "F", then node "S" is certain that the failure of node "F" is a midpoint failure and hence node "S" applies the procedures specified in Sections 5.3.1 or 5.3.2, respectively.

3. Otherwise the failure is not a midpoint failure and hence the node "S" may apply other protection techniques that are beyond the scope of this document or simply drop the packet and wait for normal protocol conversion.

6. Measurements on Real Networks

This section presents measurements performed on real service provider and large enterprise networks. The objective of the measurements is to assess the number of SIDs required in an explicit path when the mechanism 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 segment routing over the MPLS forwarding plane.

The measurements below indicate that for link and local SRLG protection, a 1 SID repair path delivers more than 99% coverage. For node protection a 2 SIDs repair path yields 99% coverage.

Table 1 below lists the characteristics of the networks used in our measurements. 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 recored

- The percentage of number of SIDs are listed in Tables 2A/B, 3A/B, and 4A/B

The measurements listed in the tables indicate that for link and local SRLG protection, 1 SID repair paths are sufficient to protect more than 99% of the prefix in almost all cases. For node protection 2 SIDs repair paths yield 99% coverage.

Network	Nodes	Circuits	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 convention that we use is 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 4.1.

- 1 SIDs: the repair node is a PQ node, in which case only 1 SID is needed to guarantee loop-freeness. This scenario is covered in Section 4.2.

- 2 or more SIDs: The repair path consists of 2 or more SIDs as described in Sections 4.3 and 4.4. We do not cover the case for 2 SIDs (Section 4.3) 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.

Table 2A and 2B below summarize the measurements on the number of SIDs needed for link protection

Network	0 SIDs	1 SID	2 SIDs	3 SIDs
T1	74.227%	25.256%	0.517%	0.001%
T2	81.097%	18.738%	0.165%	0.0%
T3	95.878%	4.067%	0.056%	0.0%
T4	62.547%	35.666%	1.788%	0.0%
T5	85.733%	14.267%	0.0%	0.0%
T6	81.252%	18.714%	0.033%	0.0%
T7	98,857%	1.143%	0.0%	0.0%
T8	94,118%	5.882%	0.0%	0.0%
T9	98.950%	1.050%	0.0%	0.0%

Table 2A: Link protection (repair size distribution)

Network	0 SIDs	1 SID	2 SIDs	3 SIDs
T1	74.227%	99.482%	99.999%	100.0%
T2	81.097%	99.835%	100.0%	100.0%
T3	95.878%	99.944%	100.0%	100.0%
T4	62.547%	98.212%	100.0%	100.0%
T5	85.733%	100.000%	100.0%	100.0%
T6	81.252%	99.967%	100.0%	100.0%
T7	98,857%	100.000%	100.0%	100.0%
T8	94,118%	100.000%	100.0%	100.0%
T9	98,950%	100.000%	100.0%	100.0%

Table 2B: Link protection repair size cumulative distribution

Table 3A and 3B summarize the measurements on the number of SIDs needed for local SRLG protection.

Network	0 SIDs	1 SID	2 SIDs	3 SIDs
T1	74.177%	25.306%	0.517%	0.001%
T2	No SRLG Information			
T3	93.650%	6.301%	0.049%	0.0%
T4	62,547%	35.666%	1.788%	0.0%
T5	83.139%	16.861%	0.0%	0.0%
T6	No SRLG Information			
T7	No SRLG Information			
T8	85.185%	14.815%	0.0%	0.0%
T9	98,940%	1.060%	0.0%	0.0%

Table 3A: Local SRLG protection repair size distribution

Network	0 SIDs	1 SID	2 SIDs	3 SIDs
T1	74.177%	99.482%	99.999%	100.001%
T2	No SRLG Information			
T3	93.650%	99.951%	100.000%	0.0%
T4	62,547%	98.212%	100.000%	100.0%
T5	83.139%	100.000%	100.0%	100.0%
T6	No SRLG Information			
T7	No SRLG Information			
T8	85.185%	100,000%	100.000%	100.0%
T9	98,940%	100,000%	100.000%	100.0%

Table 3B: 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	49.771%	47.902%	2.156%	0.148%	0.023%
T2	36,528%	59.625%	3.628%	0.194%	0.025%
T3	73,287%	25,574%	1,128%	0.010%	0%
T4	36.112%	57.350%	6.329%	0.199%	0.010%
T5	73.185%	26.815%	0%	0%	0%
T6	78.362%	21.320%	0.318%	0%	0%
T7	66.106%	32.813%	1.082%	0%	0%
T8	59.712%	40.288%	0%	0%	0%
T9	98.950%	1.050%	0%	0%	0%

Table 4A: Node protection (repair size distribution)

Network	0 SIDs	1 SID	2 SIDs	3 SIDs	4 SIDs
T1	49.771%	97.673%	99.829%	99.977%	100%
T2	36,528%	96.153%	99.781%	99.975%	100%
T3	73,287%	98.862%	99.990%	100.0%	100%
T4	36.112%	93.461%	99.791%	99.990%	100%
T5	73.185%	100.0%	100.0%	100.0%	100%
T6	78.362%	99.682%	100.0%	100.0%	100%
T7	66.106%	98,918%	100.0%	100.0%	100%
T8	59.712%	100.0%	100.0%	100.0%	100%
T9	98.950%	100.0%	100.0%	100.0%	100%

Table 4B: Node protection (repair size cumulative distribution)

7. Security Considerations

The techniques described in this document is internal functionality to a router that result in the ability to 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 mechanisms does not protect external destinations, but rather provides quick restoration for destination that are internal to a routing domain.

8. IANA Considerations

No requirements for IANA

9. Conclusions

This document proposes a mechanism that is able to pre-calculate a backup path for every primary path so as to be able to protect against the failure of a directly connected link, node, or SRLG. The mechanism is able to calculate the backup path irrespective of the topology as long as the topology is sufficiently redundant.

10. References

10.1. Normative References

10.2. Informative References

[1] Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-08 (work in progress), May 2016.

[2] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, January 2010.

[3] 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.

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

[5] 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>.

[6] Bashandy, A., Filsfils, C., and Litkowski, S., " Loop avoidance using Segment Routing", draft-bashandy-rtgwg-segment-routing-uloop-00, (work in progress), May 2017

[7] Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and Shakir, R, "Segment Routing Architecture", draft-ietf-spring-segment-routing-11 (work in progress), February 2017

11. Acknowledgments

We would like to give Les Ginsberg special thanks for the valuable comments and contribution

This document was prepared using 2-Word-v2.0.template.dot.

Authors' Addresses

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

Ahmed Bashandy
Arrcus
Email: abashandy.ietf@gmail.com

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

Bruno Decraene
Orange
Issy-les-Moulineaux
FR
Email: bruno.decraene@orange.com

Stephane Litkowski
Orange
FR
Email: stephane.litkowski@orange.com

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

Pablo Camarillo
Cisco Systems
Email: pcamaril@cisco.com

Francois Clad
Cisco Systems
Email: fclad@cisco.com