Routing area
Internet-Draft
Intended status: Informational
Expires: January 19, 2018

S. Hegde
C. Bowers
Juniper Networks, Inc.

July 18, 2017

Node Protection for SR-TE Paths
draft-hegde-spring-node-protection-for-sr-te-paths-01

Abstract

Segment routing supports the creation of explicit paths using adjacency-sids, node-sids, and binding-sids. It is important to provide fast reroute (FRR) mechanisms to respond to failures of links and nodes in the Segment-Routed Traffic-Engineered(SR-TE) path. A point of local repair (PLR) can provide FRR protection against the failure of a link in an SR-TE path by examining only the first (top) label in the SR label stack. In order to protect against the failure of a node, a PLR may need to examine the second label in the stack as well in order to determine SR-TE path beyond the failed node. This document specifies how a PLR can use the first and second label in the label stack describing an SR-TE path to provide protection against node failures.

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 RFC 2119 [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 January 19, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 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 2
2. Node Failures Along SR-TE Paths 3
2.1. Node protection for node-sid explicit paths 3
2.2. Node-protection for adj-sid explicit paths 4
2.3. Node-protection of binding-sid explicit paths 5
3. Detailed Solution using Context Tables 5
3.1. Building Context Tables 5
3.2. Building node protecting paths for node-sids 5
3.2.1. Building node protecting paths for adjacency-sids . . 7
3.3. Node protection for binding sids 8
3.4. Node protection for edge nodes 10
4. Security Considerations 11
5. IANA Considerations 11
6. Acknowledgments 11
7. References 11
7.1. Normative References 11
7.2. Informative References 11
Authors' Addresses 12

1. Introduction

It is possible for a routing device to completely go out of service abruptly due to power failure, hardware failure or software crashes. Node protection is an important property of the Fast Reroute mechanism. It provides protection against a node failure by rerouting traffic around the failed node. For example, the mechanisms described in Loop Free Alternates [RFC5286] and Remote loop free alternates [I-D.ietf-rtgwg-rlfa-node-protection] can be used to provide node protection to ensure minimal traffic loss after a node failure. The solutions to provide node protection in this draft use SPF based local protection mechanisms.

Section 2 describes problems with SR-TE paths and need for a specialized mechanism to provide node protection for the SR-TE paths. Section 3 describes the solution applied to paths built using adjacency-sids, node-sids and binding-sids. Section 3.4 describes the solution applied to egress node protection.

2. Node Failures Along SR-TE Paths

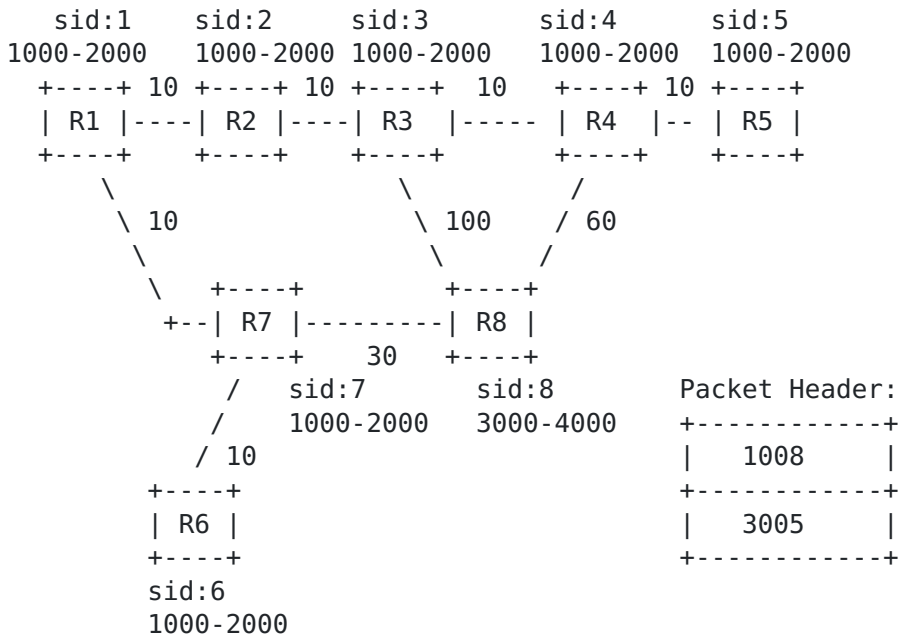

Figure 1: Sample Network

The topology shown in Figure 1. illustrates a sample network topology with SPRING enabled on each node. The SRGB and the segment index corresponding to each node is described in the topology diagram.

2.1. Node protection for node-sid explicit paths

Consider an explicit path from R1->R5 via R1->R7->R8->R4->R5. This path can be built using R1->R8 and R8->R5 shortest paths. The label stack contains two node-sids 1008 and 3005. The 1008 label would take the packet to R8 and get popped. The next label in the stack 3005 would take the packet to the destination R5. If the node R8 goes down, it is not possible for R7 to perform FRR without examining the second label in the incoming label stack (3005). R7 does not need to understand the meaning of label 3005 in order to perform normal forwarding in the absence of a failure. However, in order to support node protection, R7 will need to understand the meaning of label 3005 in order to determine where the packet is headed after R8.

Anycast addresses are in general advertised by more than one node and if per-prefix LFA calculation [RFC5286] is used node protecting paths can be found for the anycast sids. If a node protecting path is available for the anycast sid then the context table lookup mechanism would not be required. Otherwise, the anycast label has to be popped and next label looked up to find where the packet should be forwarded.

2.2. Node-protection for adj-sid explicit paths

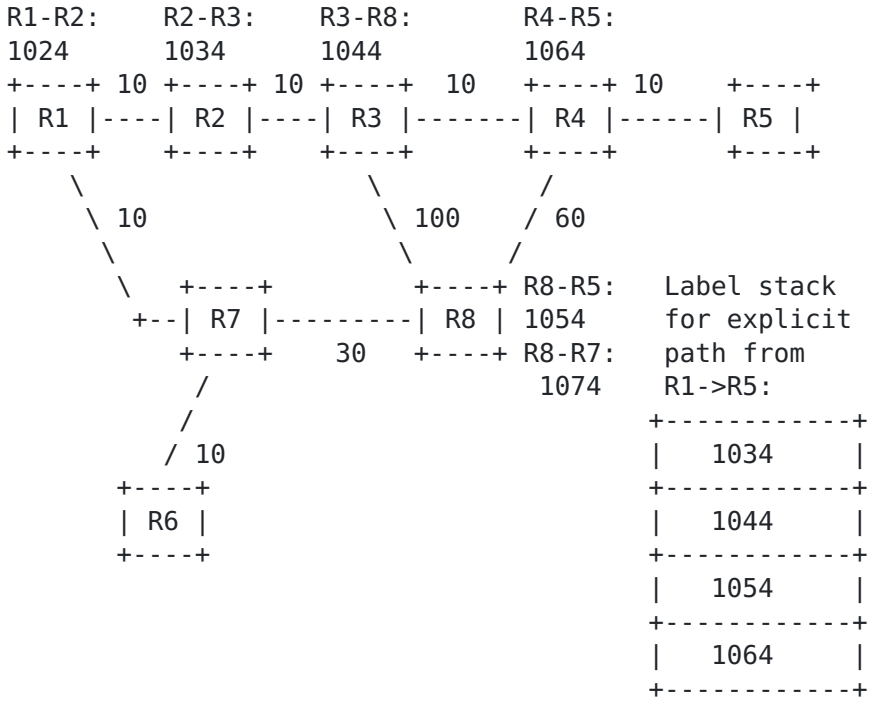

Figure 2: Explicit path using adjacency sids

Consider an explicit path from R1->R5 via R1->R2->R3->R8->R4->R5. This path can be built using adjacency sids, as shown in Figure 2. The diagram shows the adjacency sids advertised by each node required to realize this path, as well as the complete label stack. When a packet leaving R1 with this label stack reaches R3, the top of stack contains the label 1044 which will take the packet to R8. The next-next-hop in the path is R4. To provide protection for the failure of node R8, R3 would need to send the the packet to R4 without going through R8. However, the only way R3 can learn that the packet needs to go to the R4 is to examine the next label in the stack, label 1054.

2.3. Node-protection of binding-sid explicit paths

Binding sids (defined in SR architecture [I-D.ietf-spring-segment-routing]) allow the SR-TE path to be built using a hierarchy of sub-paths. The binding sid provides a single label to represent a set of nodes and links. If the node advertising the binding sid goes down, the traffic needs to be protected. The label stack involving the binding-sid contains next label in the stack which corresponds to the end point represented by the binding-sid. The penultimate node of the binding-sid advertiser cannot know the meaning of the next label in the stack.

3. Detailed Solution using Context Tables

3.1. Building Context Tables

[RFC5331] introduced the concept of Context Specific Label Spaces and there are various applications making use of this concept.A context label table on a router represents the Label Information Base (LIB) from the point of view of a particular neighbor . Context tables are built by constructing incoming label mappings advertised by the neighbor and the actions corresponding to those labels. The labels advertised by each node are local to the node and may not be unique across the segment routing domain. The context tables are separate tables built on a per-neighbor basis on every node to ensure they represent LIBs of a particular neighbor.

When a node learns the node-sid, SRGB, and adjacency-sids or binding-sids from a neighbor, the label mapping is added to the context table corresponding to that neighbor. The output actions for the label mapping are derived based on the actions that the neighbor would perform on receipt of the label.

The following section illustrates how the context table is constructed to allow the PLR to provide node-protecting paths for the next-next hops in the previous examples

3.2. Building node protecting paths for node-sids

R7's Transit Routing table

in-label	Out label
1001	Fwd to R1,
1002	swap 1002, Fwd to R1
1003	swap 1003, Fwd to R1
1004	swap 1004, Fwd to R1
1005,	swap 1005, Fwd to R1
1008,	pop, fwd to r8 *pop,lookup context.r8

* - Indicates backup path.

R7's Context Table for R8

in-label	Out label
3001	Fwd to R1,
3002	swap 1002, Fwd to R1
3003	swap 1003, Fwd to R1
3004	swap 1004, Fwd to R1
3005,	swap 1005, Fwd to R1

Figure 3: Transit routing table and Context Table at R7

The above Figure 3 shows the transit routing table and the context table of neighbor R8 built at R7 for the example network shown in

Figure 1. When the adjacency with R8 comes up, R7 builds the context table for R8 and adds the label mappings to the context table by adding the node-sid index of all the nodes to the SRGB advertised by R8. The output action is constructed by looking into the R7's SPF and backup SPF computations for the next-nexthop. The backup SPF computations as defined in LFA [RFC5286] are applicable here. The R7's SPF and backup SPF computations for the next-nexthop may provide multiple loop free primary or backup paths. A loop free path that does not include the failure node (R8 in this example) is chosen and downloaded to the context table.

R7's routing table entry for R8's sid i.e label 1008 will have a pop and forward action and the backup path SHOULD have action pop and lookup into the context table of R8. When the node R7 detects R8 goes down, R7's forwarding plane does a local repair and points to the backup path. When a packet whose top label is 1008 arrives at R7, the top label is popped, and the next label is looked up in the context table for R8. As shown in Figure 3, if the next label is 3005, the packet will be directed to R5 along a path that avoids R8.

3.2.1. Building node protecting paths for adjacency-sids

R3's Transit Routing table (partial)

in-label	Out label
1044	pop,Fwd to R8, *pop, lookup context.r8
1004	pop, Fwd to R4 *push 3004, fwd to R8

* - Indicates backup path.

R3's Context Table for R8 (partial)

in-label	Out label
1054	pop,Fwd to R4,
1074	swap 1007, Fwd to R2

Figure 4: Context Table at R3

The processing for the packet is similar to mechanism explained for node sids in section Section 3.2.

Figure 4 shows the context table constructed at R3 corresponding to R8 for the sample network shown in Figure 2. Adjacency sids are attached to the link advertisements in IGPs and the link advertisements contain the node information of the remote end. When R3 learns adjacency sids from R8, it builds context table for R8 which contains the adjacency labels advertised by R8 and the output action is built by looking at R3's own SPF and backup SPF computations for the remote end point of the link. Among the multiple primary/backup paths to the remote end of the link, a loop free path that does not pass through R8 is chosen.

3.3. Node protection for binding sids

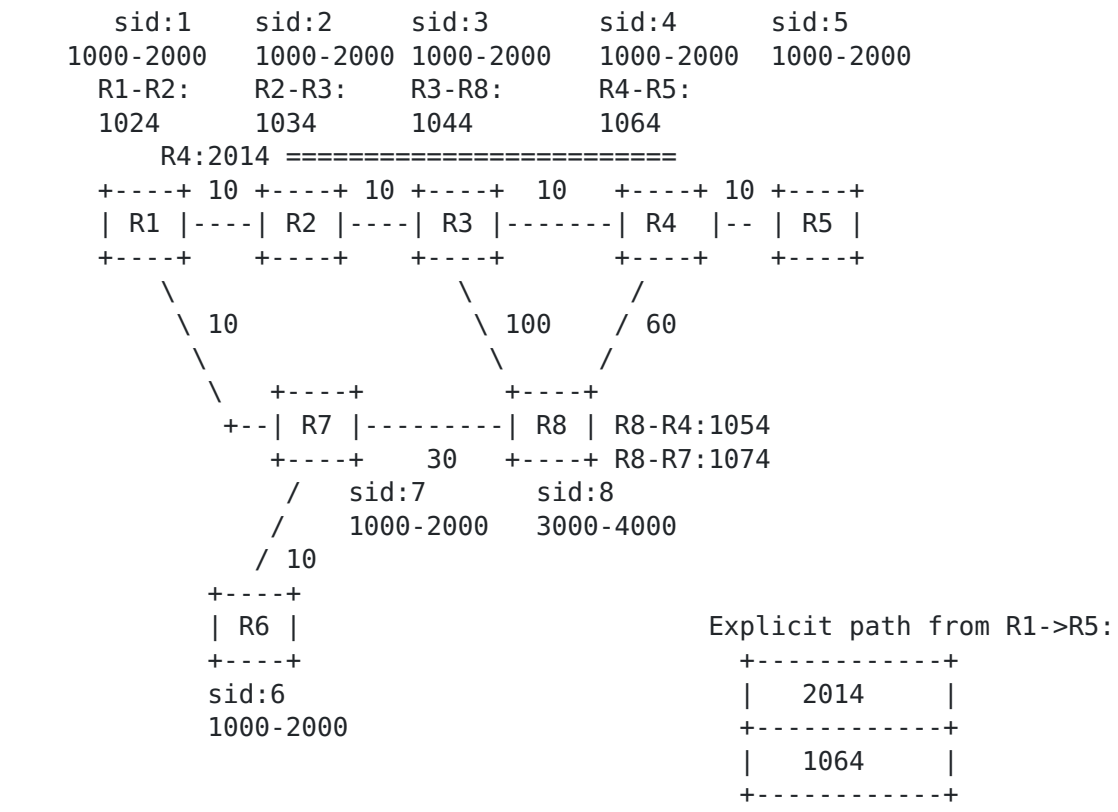

Figure 5: Node Protection for Binding SID

Figure Section 3.3 describes a sample network where R2 advertises a binding sid 2014 for the path R2->R3->R4. This mechanism is very useful in compressing the label stack depth as a sub-path can be represented using a single label. The explicit path R1->R2->R3->R4->R5 can be represented by 2 label stack as shown in above figure. If the node that advertises the binding-sid goes down, protection mechanisms are needed for the binding sid that the node advertised. A receiving node that programs a forwarding path for the binding sid should find a node protecting path to the last node of the path represented by the binding sid. In the above sample network, R1 programs a backup path for binding sid 2014 with the node protecting R-LFA path to R4 which consists of two labels [1008, 1004]. When the packet reached R4, it has the label 1064 in the label stack and can recognize this label and forward to R5. The node protecting path could be computed using various FRR technologies like LFA [RFC5286], Remote-LFA [RFC7490] , TI-LFA [I-D.francois-rtgwg-segment-routing-ti-lfa] etc.

3.4. Node protection for edge nodes

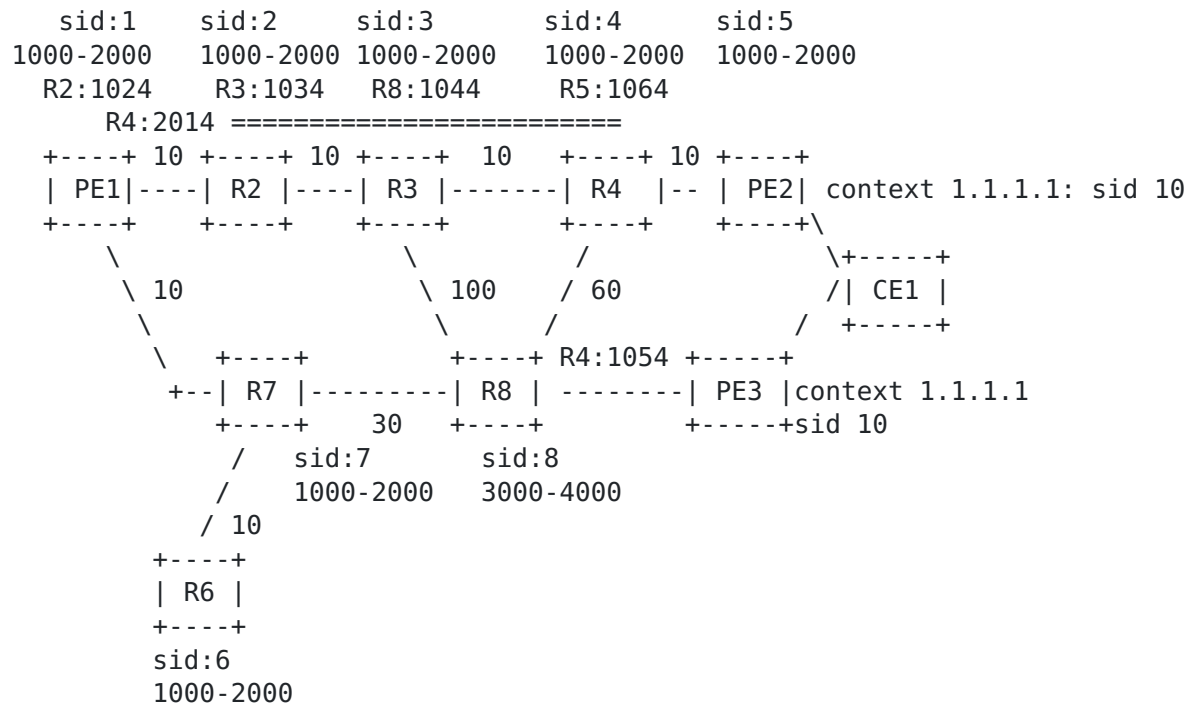

Figure 6: Node Protection for edge nodes

The node protection mechanisms that are described in previous sections depend on the assumption that the label below the top label in the label stack are understood in the IGP domain. If the edge node goes down, the label below the top label representing the edge node could be BGP service label or labels representing other applications. Service mirroring use case is described in [I-D.filsfils-spring-segment-routing-use-cases] The Customer edges are multi-homed to provider edges and one of the PE's acts in primary role and the other in protector role. The two PEs advertise a context ip address for each customer site and attaches a prefix-sid to the context. The protector PE advertises a binding sid with M bit set which implies mirroring capability for the context. Protector PE builds the context table for the BGP service labels advertised by the primary PE for the same context. The BGP service is built using stack of labels with context-sid at the bottom of the label stack.when the label ranges advertised by the PE2 and the penultimate node, Penultimate node does not understand the bottom label which is advertised by the node PE2. Any penultimate node of PE2 builds a context table for PE2 as explained in the section Section 3.1. This context table contains the sid for the context-id and output action is to pop the top label and replace with the binding sid that the

protector PE advertised for the context 1.1.1.1. The binding sid directs the protector PE to lookup the context table of Primary PE for the BGP service labels. The node protection mechanisms described in this document also ensure the edge node protection when uniform label range is not assigned across the entire IGP domain.

4. Security Considerations

TBD

5. IANA Considerations

6. Acknowledgments

7. References

7.1. Normative References

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

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, DOI 10.17487/RFC5331, August 2008, <http://www.rfc-editor.org/info/rfc5331>.

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

7.2. Informative References

[I-D.filsfils-spring-segment-routing-use-cases] Filsfils, C., Francois, P., Previdi, S., Decraene, B., Litkowski, S., Horneffer, M., Milojevic, I., Shakir, R., Ytti, S., Henderickx, W., Tantsura, J., Kini, S., and E. Crabbe, "Segment Routing Use Cases", draft-filsfils-spring-segment-routing-use-cases-01 (work in progress), October 2014.

[I-D.francois-rtgwg-segment-routing-ti-lfa] Francois, P., Bashandy, A., Filsfils, C., Decraene, B., and S. Litkowski, "Abstract", draft-francois-rtgwg-segment-routing-ti-lfa-04 (work in progress), December 2016.

[I-D.ietf-rtgwg-rlfa-node-protection]
Sarkar, P., Hegde, S., Bowers, C., Gredler, H., and S. Litkowski, "Remote-LFA Node Protection and Manageability", draft-ietf-rtgwg-rlfa-node-protection-13 (work in progress), January 2017.

[I-D.ietf-spring-segment-routing]
Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-12 (work in progress), June 2017.

[I-D.minto-rsvp-lsp-egress-fast-protection]
Jeganathan, J., Gredler, H., and Y. Shen, "RSVP-TE LSP egress fast-protection", draft-minto-rsvp-lsp-egress-fast-protection-03 (work in progress), November 2013.

[ISO10589]
"Intermediate system to Intermediate system intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473), ISO/IEC 10589:2002, Second Edition.", Nov 2002.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, DOI 10.17487/RFC1195, December 1990, <http://www.rfc-editor.org/info/rfc1195>.

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

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <http://www.rfc-editor.org/info/rfc2328>.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, DOI 10.17487/RFC5340, July 2008, <http://www.rfc-editor.org/info/rfc5340>.

Authors' Addresses

Shraddha Hegde
Juniper Networks, Inc.
Exora Business Park
Bangalore, KA 560103
India

Email: shraddha@juniper.net

Chris Bowers
Juniper Networks, Inc.

Email: cbowers@juniper.net