TEAS WG
Internet-Draft
Intended status: Standards Track
Expires: January 9, 2017

A. Deshmukh
K. Kompella
Juniper Networks, Inc.

July 8, 2016

RSVP Extensions for RMR
draft-deshmukh-rsvp-rmr-extension-00

Abstract

Rings are the most common topology in access and aggregation networks. However, the use of MPLS as the transport protocol for rings is very limited today. draft-ietf-mpls-rmr-02 describes a mechanism to handle rings efficiently using MPLS. This document describes the extensions to the RSVP protocol for signaling MPLS label switched paths in rings.

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].

Table of Contents

1. Introduction

This document extends RSVP-TE [RFC3209] to establish label-switched path (LSP) tunnels in the ring topology. Rings are auto-discovered using the mechanisms mentioned in the [draft-ietf-mpls-rmr-02]. Either IS-IS [RFC5305] or OSPF[RFC3630] can be used as the IGP for auto-discovering the rings.

After the rings are auto-discovered, each ring node knows its clockwise (CW) and anticlockwise (AC) ring neighbors and its ring links. All of the express links in the ring also get identified as part of the auto-discovery process. At this point, every node in the ring informs the RSVP protocol to begin the signaling of the ring LSPs.

Section 2 covers the terminology used in this document. Section 3 presents the RSVP protocol extensions needed to support MPLS rings. Section 4 describes the procedures of RSVP LSP signaling in detail.

2. Terminology

A ring consists of a subset of n nodes {R_i, 0 <= i < n}. We define the direction from node R_i to R_i+1 as "clockwise" (CW) and the reverse direction as "anti-clockwise" (AC). As there may be several rings in a graph, we number each ring with a distinct ring ID RID.

```
                          R0 . . . R1
                        .             .
                     R7                  R2
       Anti-      |  .        Ring        .  |
       Clockwise  |  .                    .  | Clockwise
                  v  .      RID = 17      .  v
                     R6                  R3
                        .             .
                          R5 . . . R4
```

Figure 1: Ring with 8 nodes

The following terminology is used for ring LSPs:

Ring ID (RID): A non-zero number that identifies a ring; this is unique in some scope of a Service Provider’s network. A node may belong to multiple rings.

Ring node: A member of a ring. Note that a device may belong to several rings.

Node index: A logical numbering of nodes in a ring, from zero upto one less than the ring size. Used purely for exposition in this document.

Ring neighbors: Nodes whose indices differ by one (modulo ring size).

Ring links: Links that connect ring neighbors.

Express links: Links that connect non-neighboring ring nodes.

MP2P LSP: Each LSP in the ring is a multipoint to point LSP such that LSP can have multiple ingress nodes and one egress node.

3. RSVP Extensions

Since the procedures of signaling ring LSPs will be different from the signaling of regular RSVP LSPs, a new C-Type is defined here for the SESSION object. This new C-Type will help to clearly differentiate ring LSPs from regular LSPs. In addition, new flags are introduced in the SESSION object to represent the ring direction of the corresponding Path message.

3.1. Session Object

Class = SESSION, LSP_TUNNEL_IPv4 C-Type = TBD

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Ring anchor node address                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Ring Flags             |           MBB ID              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Ring ID                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

SESSION Object

Ring anchor node address: IPv4 address of the anchor node. Each anchor node creates a LSP addressed to itself.

MBB ID: A 16-bit identifier used in the SESSION. This "Make-before-break" (MBB) ID is useful for graceful ring changes. If a new node is being added to the ring or some existing node goes down and we have to signal a smaller ring, in those cases, anchor node creates a new tunnel with a different "MBB ID".

Ring ID: A 32-bit number that identifies a ring; this is unique in some scope of a Service Provider’s network. This number remains constant throughout the existence of ring.

Ring Flags: For each ring, the anchor node starts signaling of a ring LSP. Ring LSP RL_i, anchored on node R_i, consists of two counter-rotating unicast LSPs that start and end at R_i. One LSP will be in the clockwise direction and other LSP will be in the anti-clockwise direction. A ring LSP is "multipoint": any node R_j can use RL_i to send traffic to R_i; this can be in either the CW or AC directions, or both (i.e., load balanced). Two new flags are defined in the SESSION object which define the ring direction of the corresponding Path message.

ClockWise(CW) Direction 0x01: This flag indicates that the corresponding Path message is traveling in the ClockWise(CW) direction along the ring.

Anti-ClockWise(AC) Direction 0x02: This flag indicates that the corresponding Path message is traveling in the Anti-ClockWise(AC) direction along the ring.

3.2. SENDER_TEMPLATE,FILTER_SPEC Objects

There will be no changes to the SENDER_TEMPLATE and FILTER_SPEC objects. The format of the above 2 objects will be similar to the definitions in RFC 3209. [RFC3209] Only the semantics of these objects will slightly change. This will be explained in section Section 4.5 below.

4. Ring Signaling Procedures

A ring node indicates in its IGP updates the ring LSP signaling protocols that it supports. This can be LDP and/or RSVP-TE. Ideally, each node should support both. If the ring is configured with RSVP as the signaling protocol, then once a ring node R_i knows the RID, its ring links and directions, it kicks off ring RSVP LSP signaling automatically.

4.1. Differences from regular RSVP-TE LSPs

Ring LSPs differ from regular RSVP-TE LSPs in several ways:

1. Ring LSPs (by construction) form a loop.

2. Ring LSPs are multipoint-to-point. Any ring node can inject traffic into a ring LSP.

3. The bandwidth of a ring LSP can change hop-to-hop.

4. Ring LSPs are protected without the use of bypass or detour LSPs. Ring LSP protection is akin to SONET/SDH ring protection.

4.2. LSP signaling

After the ring auto-discovery process, each anchor node creates a LSP addressed to itself. This ring LSP contains of a pair of counter-rotating unicast LSPs. So, for a ring containing N nodes, there will be 2N total LSPs signaled.

There is no need for ERO object in the Path message. The Path message for ring LSPs has the following format:

```
        <Path Message> ::=  <Common Header> [ <INTEGRITY> ]
                                <SESSION> <RSVP_HOP>
                                <TIME_VALUES>
                                <LABEL_REQUEST>
                                [ <SESSION_ATTRIBUTE> ]
                                <sender descriptor list>

        <sender descriptor list> ::= <sender descriptor>|
                                         <sender descriptor list> <sender descri
ptor>
        <sender descriptor> ::= <SENDER_TEMPLATE> <SENDER_TSPEC>
```

The anchor node creates 2 Path messages traveling in opposite directions. The SESSION format MUST be as per the description in Section 3.1. The anchor node which creates the LSP will insert it’s own address in the "Ring node anchor address" field of the SESSION object. So effectively, the Path messages are addressed to the originating node itself.

The SESSION flags of these 2 Path messages are different. The Path message sent to the CW neighbor MUST have the CW flag set in the SESSION object to signal the LSP going in the clockwise direction. The Path message sent to the AC neighbor MUST have the AC flag set to signal the LSP in the anti-clockwise direction. The details for signaling over express links will be given in a future version.

When an incoming Path message is received at the ring node R_i, it consults the results of auto-discovery to find the appropriate ring neighbor. If the incoming Path message has CW direction flag set, then R_i sends a Path message to its CW ring neighbor (and vice versa). Thus, there is no need of ERO in the Path message. The Path message is routed locally at each ring based on the ring auto-discovery calculations.

The RESV message for ring LSPs also uses the new RING_IPv4 SESSION object. When the Path message originated from the anchor node R_i reaches back to R_i, R_i generates a Resv message. Note that this means that anchor node is both Ingress and Egress for the Path message. The Resv message copies the same ring flags as received in the corresponding Path message. So, a Resv message for a CW LSP goes in the AC direction (unlike the Path message, which goes CW). This is done to correctly match Path and corresponding Resv messages at transit ring nodes. Upon receiving Resv message with CW flag set, the ring node will forward the Resv message to its AC neighbor.

Each ring node R_i allocates CW and AC labels for each ring LSP RL_k. As the signaling propagates around the ring, CW and AC labels are exchanged. When R_i receives CW and AC labels for RL_k from its ring neighbors, primary and fast reroute (FRR) paths for RL_k are installed at R_i.

Consider the following three nodes of the ring, and their signaling interactions for LSP RL_5 originating from anchor node R5:

```
                        P5_CW ->     P5_CW ->
                        Q5_CW <-     Q5_CW <-
        ... ------ R7 --------- R8 --------- R9 ------ ...
                        P5_AC <-     P5_AC <-
                        Q5_AC ->     Q5_AC ->
```

P corresponds to the Path message and Q corresponds to the Resv message.

Also, since ring LSPs are MP2P in nature, each ring node SHOULD also signal a Path message towards anchor node. The procedure for that is as follows:

When a ring node R5 receives a Path message initiated by anchor node R1(for anchor lsp "lsp1"), R5 SHOULD make a copy of the received Path message for "lsp1". R5 then modifies the sender-template object from the copied Path message for "lsp1". In the sender-template object, R5 uses the sender address as the loopback address of node R5 and lsp-id = X. R5 then forwards this new Path message to it’s ring neighbor. The original anchor Path message has sender address as loopback address of R1 and lsp-id = X.

So at this point, there will be 2 different path messages existing for lsp1 First Path message will be for the anchor LSP with sender address = node R1. Second Path message will be for the ring LSP with sender address = node R5.

When node R1 receives this modified Path message, it replies with the Resv message containing the same label it advertised for the original anchor lsp "lsp1". The SESSION object of the Resv message will also exactly match with the received Path message. Only the FILTER_SPEC object in the Resv message will have the sender address as loopback of node R5. As this Resv message propagates back towards R5, all the transit nodes also send the same label that they have allocated for the original anchor lsp "lsp1". So no new label routes get installed as part of signaling for this ring lsp. The anchor LSP and all of their associated ring LSPs share label routes. The label actions are described below in Section 4.3.

4.3. Protection

In the rings, there are no protection LSPs -- no node or link bypass LSPs, no standby LSPs and no detours. Protection is via the "other" direction around the ring, which is why ring LSPs are in counter-rotating pairs. Protection works in the same way for link, node and ring LSP failures.

Since each ring LSP is a MP2P LSP, any ring node can inject traffic onto a LSP whose anchor might be a different ring node. To achieve the above, an ingress route will be installed as follows at every ring node J, for a given ring-LSP with anchor Rk (say 1.2.3.4).

```
1.2.3.4  ->  (Push CL_J+1,K, NH: R_J+1)       # CW
         ->  (Push AL_J-1,K, NH: R_J-1)       # AC

         CL = Clockwise label
         AL = Anti-Clockwise label
```

Traffic will either be load balanced in the CW and AC directions or the traffic will be sent on just CW or AC lsp based on parameters such as hop-count, policy etc.

Also, 2 transit routes will be installed for the anchor LSP transiting from node Rj as follows:

```
CL_J,K ->  SWAP(CL_J+1,K,  NH: R_J+1)             #CW
       ->  SWAP(AL_J-1,K , NH: R_J-1)             #AC

           CL = Clockwise label
           AL = Anti-Clockwise label
           CW NH has weight 1, AC NH has higher-weight.


AL_J,K -> SWAP(AL_J-1,K , NH: R_J-1)  #AC
       -> SWAP(CL_J+1,K,  NH: R_J+1)  #CW

           CL = Clockwise label
           AL = Anti-Clockwise label
           AC NH has weight 1, CW NH has higher weight.
```

Suppose a packet headed in anti-clockwise direction towards R5 and it arrives at node R8. Lets say that now R8 learns there is a link failure in the AC direction. R8 reroutes this packet back onto the clockwise direction. This reroute action is pre-programmed in the LFIB, to minimize the time between detection of a fault and the corresponding recovery action.

At this time, R8 also sends a notification to R7 that the AC direction is not working, so that R7 can similarly switch traffic to the CW direction. These notification SHOULD propagate CW until each traffic source on the ring CW of the failure uses the CW direction.For RSVP-TE, this notification is sent in the form of PathErr message.

To provide this notification, the ring node detecting failure SHOULD send a Path Error message with error code of "Notify" and an error value field of ("Tunnel locally repaired"). This Path Error code and value is same as defined in RFC 4090[RFC4090] for the notification of local repair.

Note that the failure of a node or a link will not necessarily affect all ring LSPs. Thus, it is important to identify the affected LSPs and only switch the affected LSPs.

4.4. Ring changes

A ring node can go down resulting in a smaller ring or a new node can be added to the ring which will increase the ring size. In both of the above cases, the ring auto-discovery process SHOULD kick in and it SHOULD calculate a new ring with the changed ring nodes.

When the ring auto-discovery process is complete, IGP will signal RSVP to begin the MBB process for the existing ring LSPs. For this MBB process, the anchor node will create a new Path message with a different "MBB ID" in the SESSION object. All other fields in the SESSION Object will remain same as the existing Path message(before the ring change).

This new Path message will then propagate along the ring neighbors in the same way as the original Path message. Each ring neighbor SHOULD forward the Path message to it’s appropriate neighbor based on the new auto-discovery calculations.

For the ring links which are common between the old and new LSPs, the LSPs will share resources(SE style reservation) on those ring links. Note that here we are using MBB_ID in the SESSION object to share resources instead of the LSP_ID in the SENDER_TEMPLATE Object(which is used in RSVP-TE for sharing resources as described in RFC 3209 [RFC4090]). The LSP_ID use is reserved for a different functionality as described in section Section 4.5.

4.5. Bandwidth management

For RSVP-TE LSPs, bandwidths may be signaled in both directions. However, these are not provisioned either; rather, one does "reverse call admission control". When a service needs to use an LSP, the ring node where the traffic enters the ring attempts to increase the bandwidth on the LSP to the egress. If successful, the service is admitted to the ring.

```
                               . R0 . . . R1
                              . __________|| .
                             . /   ________|  .
                           R7 /  /            R2
              Anti-      |  . /  /              .  |
              Clockwise |  . | /               .  | Clockwise
                        v  . | \               .  v
                           R6   \             R3
                              .  \           .
                                R5 . . . R4
```

Figure 2: BW Management in Ring with 8 nodes

Let’s say that Ring node R5 wants to increase the BW for the LSP whose egress is at node R1. To achieve this BW increase, Ring node R5 has to increase BW along the LSP anchored at node R1(say lsp1).

R5 makes a copy of the existing ring Path message for lsp1. R5 then modifies the sender-template object from the copied Path message for "lsp1". In the sender-template object, R5 uses the sender address as the loopback address of node R5 and lsp-id = X+1. R5 also modifies the TSPEC object which represents the BW increase/decrease in this new Path message. R5 then forwards this new Path message to it’s ring neighbor. Note that R5 MUST also continue signaling the original anchor Path message received from ring node R1 for lsp1. The original anchor Path message has sender address as loopback address of R1.

Now, let’s say, node 5 wants to increase BW again for lsp1, then R5 adds a new SENDER_TEMPLATE object in the existing Path message for "lsp1" with sender address as loopback of node 5 and lsp-id = X+2. So at this point, there will be 2 different path messages existing for lsp1 First Path message will be for the anchor LSP with sender address = node 1. Second Path message will contain 2 SENDER_TEMPLATE objects as [node5, lsp-id = X+1] and [node5, lsp-id =X+2].

Similarly, if node R6 wants to increase the BW for "lsp1", it SHOULD create a new Path message containing SENDER_TEMPLATE object with sender address = loopback of node 6 and lsp-id = Y+1. Thus, the LSP-ID field is local to each sender node along the ring.

If sufficient BW is available all the way towards ring node R1, then this new Path message reaches node R1. R1 generates a Resv message with the correct FILTER_SPEC object corresponding to the received SENDER_TEMPLATE object. This Resv message will also have the correct FLOWSPEC object as per the requested bandwidth.

If sufficient BW is not available at some downstream (say node R9), then ring node R9 SHOULD generate a PathErr message with the corresponding Sender Template Object. When node R5 receives this PathErr message, R5 understands that the BW increase was not successful. Note that the existing established bandwidths for lsp1 are not affected by this new PathErr message.

When ring node R5 no longer needs the BW reservation, then ring node R5 SHOULD originate a PathTear message with the appropriate Sender Template Object as described above. Every downstream node SHOULD then remove bandwidth allocated on the corresponding link on receipt of this PathTear message.

Also, note that as part of this BW increase or decrease process, any ring node does not actually change any label associated with the LSP. So, the label remains same as it was signaled initially when the anchor LSP came up.

5. Security Considerations

It is not anticipated that either the notion of MPLS rings or the extensions to various protocols to support them will cause new security loopholes. As this document is updated, this section will also be updated.

6. Contributors

Ravi Singh
Juniper Networks, Inc.
1194 N Mathilda Ave
Sunnyvale, CA 94089
USA

Email: ravis@juniper.net

Santosh Esale
Juniper Networks, Inc.
1194 N Mathilda Ave
Sunnyvale, CA 94089
USA

Email: sesale@juniper.net

Raveendra Torvi
Juniper Networks, Inc.
10 Technology Park Dr
Westford, MA 01886
USA

Email: rtorvi@juniper.net

7. IANA Considerations

Requests to IANA will be made in a future version of this document.

8. References

8.1. Normative References

[I-D.ietf-mpls-rmr]
Kompella, K. and L. Contreras, "Resilient MPLS Rings", draft-ietf-mpls-rmr-02 (work in progress), July 2016.

[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>.

8.2. Informative References

[I-D.dai-mpls-rsvp-te-mbb-label-reuse]
Dai, M. and M. Chaudhry, "MPLS RSVP-TE MBB Label Reuse", draft-dai-mpls-rsvp-te-mbb-label-reuse-01 (work in progress), September 2015.

[RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, DOI 10.17487/RFC2205, September 1997, <http://www.rfc-editor.org/info/rfc2205>.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <http://www.rfc-editor.org/info/rfc3209>.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <http://www.rfc-editor.org/info/rfc3630>.

[RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, DOI 10.17487/RFC4090, May 2005, <http://www.rfc-editor.org/info/rfc4090>.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <http://www.rfc-editor.org/info/rfc5305>.

Authors' Addresses

Abhishek Deshmukh
Juniper Networks, Inc.
10 Technology Park Dr
Westford, MA 01886
USA

Email: adeshmukh@juniper.net

Kireeti Kompella
Juniper Networks, Inc.
1194 N Mathilda Ave
Sunnyvale, CA 94089
USA

Email: kireeti@juniper.net