Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: February 12, 2016

W. Cheng
L. Wang
H. Li
China Mobile
H. Helvoort
Hai Gaoming BV
K. Liu
J. Dong
J. He
Huawei Technologies
F. Li
China Academy of Telecommunication Research, MIIT., China
J. Yang
ZTE Corporation P.R.China
J. Wang
Fiberhome Telecommunication Technologies Co., LTD.

August 11, 2015

MPLS-TP Shared-Ring protection (MSRP) mechanism for ring topology
draft-cheng-mpls-tp-shared-ring-protection-06

Abstract

This document describes requirements, architecture and solutions for MPLS-TP Shared Ring Protection (MSRP) in the ring topology for point-to-point (P2P) services. The mechanism of MSRP is illustrated and how it satisfies the requirements for optimized ring protection in RFC 5654 is analyzed. This document also defines the Ring Protection Switch (RPS) Protocol which is used to coordinate the protection behavior of the nodes on MPLS ring.

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

Table of Contents

1. Introduction

As described in 2.5.6.1 of [RFC5654], Ring Protection of MPLS-TP requirements , several service providers have expressed much interest in operating MPLS-TP in ring topologies and require a high-level survivability function in these topologies. In operational transport network deployment, MPLS-TP networks are often constructed with ring topologies. It calls for an efficient and optimized ring protection mechanism to achieve simple operation and fast, sub 50 ms, recovery performance.

The requirements for MPLS-TP [RFC5654] state that recovery mechanisms which are optimized for ring topologies could be further developed if it can provide the following features:

a. Minimize the number of OAM entities for protection

b. Minimize the number of elements of recovery

c. Minimize the required label number

d. Minimize the amount of control and management-plane transactions during maintenance operation

e. Minimize the impact on information exchange during protection if a control plane is supported

This document specifies MPLS-TP Shared-Ring Protection mechanisms that can meet all those requirements on ring protection listed in [RFC5654].

The basic concepts and architecture of Shared-Ring protection mechanism are specified in this document. This document focuses on the solutions for point-to-point transport paths. While the basic concepts may also apply to point-to-multipoint transport paths, the solution for point-to-multipoint transport paths is under study and will be presented in a separate document.

2. Requirements for MPLS-TP Ring Protection

The requirements for MPLS-TP ring protection are specified in [RFC5654]. This document elaborates on the requirements in detail.

2.1. Recovery of Multiple Failures

MPLS-TP is expected to be used in carrier grade metro networks and backbone transport networks to provide mobile backhaul, business services etc., in which the network survivability is very important. According to R106 B in [RFC5654], MPLS-TP recovery mechanisms in a ring SHOULD protect against multiple failures. The following text provides some more detailed illustration about "multiple failures". In metro and backbone networks, a single risk factor often affects multiple links or nodes. Some examples of risk factors are given as follows:

o multiple links use fibers in one cable or pipeline o Several nodes share one power supply system

o Weather sensitive micro-wave system

Once one of the above risk factors happens, multiple links or nodes failures may occur simultaneously and those failed links or nodes may be located on a single ring as well as on interconnected rings. Ring protection against multiple failures should cover both multiple failures on a single ring and multiple failures on interconnected rings, as long as the connectivity between the ingress and egress node of the ring still exists.

2.2. Smooth Upgrade from Linear Protection to Ring Protection

It is beneficial for service providers to upgrade the protection scheme from linear protection to ring protection in their MPLS-TP network without service interruption. In-service insertion and removal of a node on the ring should also be supported. Therefore,

the MPLS-TP ring protection mechanism is supposed to be developed and optimized for compliance with this smooth upgrading principle.

2.3. Configuration Complexity

Ring protection can reduce the dependency of configuration on the quantity of services, thus will simplify the network protection configuration and operation effort. This is because the ring protection makes use of the characteristics of ring topology and mechanisms on the section layer. While in the application scenarios of deploying linear protection in ring topology MPLS-TP network, the configuration of protection has a close relationship with the quantities of services carried. Especially in some large metro networks with more than ten thousands of services in the access nodes, the LSP linear protection capabilities of the metro core nodes needs to be large enough to meet the network planning requirements, which also leads to the complexity of network protection configurations and operations.

3. Terminology and Notation

The following syntax will be used to describe the contents of the label stack:

1. The label stack will be enclosed in square brackets ("[]").

2. Each level in the stack will be separated by the '|' character. It should be noted that the label stack may contain additional layers. However, we only present the layers that are related to the protection mechanism.

3. If the Label is assigned by Node X, the Node Name is enclosed in bracket ("()")

4. Shared Ring Protection Architecture

4.1. Ring Tunnel

This document introduces a new logical layer of the ring for shared ring protection in MPLS-TP networks. As shown in Figure 1, the new logical layer consists of ring tunnels which provides a server layer for the LSPs traverse the ring. Once a ring tunnel is established, the configuration, management and protection of the ring are all performed at the ring tunnel level. One port can carry multiple ring tunnels, while one ring tunnel can carry multiple LSPs.


```
                                               +-------------
                                 +-------------|
                   +-------------|             |
     =====PW1======|             |             |
                   |             |    Ring     |   Physical
     =====PW2======|     LSP     |   Tunnel    |   Port
                   |             |             |
     =====PW3======|             |             |
                   +-------------|             |
                                 +-------------|
                                               +-------------
```

Figure 1. The logical layers of the ring

The label stack used in MPLS-TP Shared Ring Protection mechanism is shown as below:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Ring tunnel Label            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              LSP Label                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               PW Label                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Payload                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2. Label stack used in MPLS-TP Shared Ring Protection

4.1.1. Establishment of Ring Tunnel

The Ring tunnels are established based on the egress node. The egress node is the node where traffic leaves the ring. LSPs which have the same egress node on the ring share the same ring tunnels. In other words, all the LSPs that traverse the ring and exit from the same node share the same working ring tunnel and protection ring tunnel. For each egress node, four ring tunnels are established:

- o one clockwise working ring tunnel, which is protected by the anticlockwise protection ring tunnel
- o one anticlockwise protection ring tunnel
- o one anticlockwise working ring tunnel, which is protected by the clockwise protection ring tunnel
- o one clockwise protection ring tunnel

The structure of the protection tunnels are determined by the selected protection mechanism. This will be detailed in subsequent sections.

As shown in Figure 3, LSP1, LSP2 and LSP3 enter the ring from Node E, Node A and Node B, respectively, and all leave the ring at Node D. To protect these LSPs that traverse the ring, a clockwise working ring tunnel (RcW_D) via E->F->A->B->C->D, and its anticlockwise protection ring tunnel (RaP_D) via D->C->B->A->F->E->D are established, Also, an anti-clockwise working ring tunnel (RaW_D) via C->B->A->F->E->D, and its clockwise protection ring tunnel (RcP_D) via D->E->F->A->B->C->D are established. For simplicity Figure 3 only shows RcW_D and RaP_D. A similar provisioning should be applied for any other node on the ring. In summary, for each node in Figure 3 when acting as egress node, the ring tunnels are created as follows:

- o To Node A: RcW_A, RaW_A, RcP_A, RaP_A
- o To Node B: RcW_B, RaW_B, RcP_B, RaP_B
- o To Node C: RcW_C, RaW_C, RcP_C, RaP_C
- o To Node D: RcW_D, RaW_D, RcP_D, RaP_D
- o To Node E: RcW_E, RaW_E, RcP_E, RaP_E
- o To Node F: RcW_F, RaW_F, RcP_F, RaP_F


```
                    +---+############+---+
```

```
                       | F |-------------| A | +-- LSP2
                       +---+*************+---+
                       #/*                   *\#
                      #/*                     *\#
                     #/*                       *\#
                   +---+                     +---+
            LSP1-+ | E |                     | B |+-- LSP3
                   +---+                     +---+
                     #\                       */#
                      #\                     */#
                       #\                   */#
                       +---+*************+---+
             LSP1   +--| D |-------------| C |
             LSP2      +---+#############+---+
             LSP3

                      ---- physical links
                      **** RcW_D
                      #### RaP_D

                  Figure 3. Ring tunnels in MSRP
```

Through these working and protection ring tunnels, LSPs which enter the ring from any node can reach any egress nodes on the ring, and are protected from failures on the ring.

4.1.2. Label Assignment and Distribution

The ring tunnel labels are downstream-assigned labels as defined in [RFC3031]. The ring tunnel labels can be either configured statically, provisioned by a controller, or distributed dynamically via a control protocol.

4.1.3. Forwarding Operation

When an MPLS-TP transport path, such as an LSP, enters the ring, the ingress node on the ring pushes the working ring tunnel label according to the egress node and sends the traffic to the next hop. The transit nodes on the working ring tunnel swap the ring tunnel labels and forward the packets to the next hop. When the packet arrives at the egress node, the egress node pops the ring tunnel label and forwards the packets based on the inner LSP label and PW label. Figure 4 shows the label operation in the MPLS-TP shared ring protection mechanism. Assume that LSP1 enters the ring at Node A and exits from Node D, and the following label operations are executed.

1. Ingress node: Packets of LSP1 arrive at Node A with a label stack [LSP1] and is supposed to be forwarded in the clockwise direction of the ring. The clockwise working ring tunnel label RcW_D will be pushed at Node A, the label stack for the forwarded packet at Node A is changed to [RcW_D(B)|LSP1].

2. Transit nodes: In this case, Node B and Node C forward the packets by swapping the working ring tunnel labels. For example, the label [RcW_D(B)|LSP1] is swapped to [RcW_D(C)|LSP1] at Node B.

3. Egress node: When the packet arrives at Node D (i.e. the egress node) with label stack [RcW_D(D)|LSP1], Node D pops RcW_D(D), and subsequently deals with the inner labels of LSP1.

```
                      +---+#####[RaP_D(F)]######+---+
                      | F |---------------------| A | +-- LSP1
                      +---+*****[RcW_D(A)]******+---+
                       #/*                         *\#
            [RaP_D(E)]#/*[RcW_D(F)]      [RcW_D(B)]*\#[RaP_D(A)]
                     #/*                             *\#
                   +---+                           +---+
                   | E |                           | B |
                   +---+                           +---+
                     #\                             */#
            [RaP_D(D)]#\                [RxW_D(C)]*/#[RaP_D(B)]
                       #\                         */#
                       +---+*****[RcW_D(D)]****+---+
             LSP1  +-- | D |-------------------| C |
                       +---+#####[RaP_D(C)]####+---+

             -----physical links      ****** RcW_D    ###### RaP_D
```

Figure 4. Label operation of MSRP

4.2. Failure Detection

The MPLS-TP section layer OAM is used to monitor the connectivity between each two adjacent nodes on the ring using the mechanisms defined in [RFC6371]. Protection switching is triggered by the failure detected on the ring by the OAM mechanisms.

Two end ports of a link form a Maintenance Entity Group (MEG), and an MEG end point (MEP) function is installed in each ring port. CC OAM packets are periodically exchanged between each pair of MEPs to monitor the link health. Three consecutive CC packets losses will be interpreted as a link failure.

A node failure is regarded as the failure of two links attached to that node. The two nodes adjacent to the failed node detect the failure in the links that are connected to the failed node.

4.3. Ring Protection

This section specifies the ring protection mechanisms in detail. In general, the description uses the clockwise working ring tunnel and the corresponding anti-clockwise protection ring tunnel as example, but the mechanism is applicable in the same way to the anti-clockwise working and clockwise protection ring tunnels.

Taking the topology in Figure 4 as example, the LSP1 enters the ring at Node A and leaves the ring at Node D. In normal state, LSP1 is carried by clockwise working ring tunnel (RcW_D) through the path A->B->C->D, the label operation is:

[LSP1](original data traffic carried by LSP1) -> [RCW_D(B)|LSP1](NodeA) -> [RCW_D(C)|LSP1](NodeB) -> [RCW_D(D)| LSP1](NodeC) -> [LSP1](data traffic carried by LSP1). Then at node D the packet will be forwarded based on label stack of LSP1.

The following sections describes the protection mechanisms used in ring topology.

4.3.1. Wrapping

With the wrapping mechanism, the protection ring tunnel is a closed ring identified by the egress node. As shown in Figure 4, the RaP_D is the anticlockwise protection ring tunnel for the clockwise working ring tunnel RcW_D. As specified in the following sections, the closed ring protection tunnel can protect both the link failure and the node failure.

4.3.1.1. Wrapping for Link Failure

When a link failure between Node B and Node C occurs, if it is a bi-directional failure, both Node B and Node C can detect the failure via OAM mechanism; if it is a uni-directional failure, one of the two nodes would detect the failure and it would inform the other node via the Ring Protection Switch Protocol (RPS) which is specified in section 5. Then Node B switches the clockwise working ring tunnel (RcW_D) to the anticlockwise protection ring tunnel (RaP_D) and Node C switches anticlockwise protection ring tunnel(RaP_D) to the clockwise working ring tunnel(RcW_D). The data traffic which enters the ring at Node A and leaves the ring at Node D follows the path A->B->A->F->E->D->C->D. The label operation is:

[LSP1](Original data traffic) -> [RcW_D(B)|LSP1](Node A) -> [RaP_D(A)|LSP1](Node B) -> [RaP_D(F)|LSP1](Node A) -> [RaP_D(E)|LSP1] (Node F) -> [RaP_D(D)|LSP1] (Node E) -> [RaP_D(C)|LSP1] (Node D) -> [RcW_D(D)|LSP1](Node C) -> [LSP1](data traffic leaves the ring).

```
                         +----+#####[RaP_D(F)]######+----+
                         | F |--------------------| A | +-- LSP1
                         +----+*****[RcW_D(A)]******+----+
                         #/*                         *\#
              [RaP_D(E)]#/*[RcW_D(F)]      [RcW_D(B)]*\#RaP_D(A)
                       #/*                             *\#
                     +----+                         +----+
                     | E |                          | B |
                     +----+                         +----+
                       #\                            *x#
              [RaP_D(D)]#\                 [RcW_D(C)]*x#RaP_D(B)
                         #\                        *x#
                         +----+*****[RcW_D(D)]****+----+
               LSP1  +-- | D |-------------------| C |
                         +----+#####[RaP_D(C)]####+----+

                 -----physical links    xxxx Failure Link
                 ****** RcW_D           ###### RaP_D
```

Figure 5.Wrapping for link failure

4.3.1.2. Wrapping for Node Failure

As shown in Figure 6, when Node B fails, Node A detects the failure between A and B and switches the clockwise work ring tunnel (RcW_D) to the anticlockwise protection ring tunnel (RaP_D), Node C detects the failure between C and B and switches the anticlockwise protection ring tunnel (RaP_D) to the clockwise working ring tunnel (RcW_D). The data traffic which enters the ring at Node A and exits at Node D follows the path A->F->E->D->C->D. The label operation is:

[LSP1](original data traffic carried by LSP1) -> [RaP_D(F)|LSP1](NodeA) -> [RaP_D(E)|LSP1](NodeF) -> [RaP_D(D)|LSP1](NodeE) -> [RaP_D(C)|LSP1] (NodeD) -> [RcW_D(D)|LSP1] (NodeC) -> [LSP1](data traffic carried by LSP1).

In one special case where node D fails, all the ring tunnels with node D as egress will become unusable. However, before the failure location is propagated to all the ring nodes, the wrapping protection mechanism may cause temporary traffic loop: node C detects the failure and switches the traffic from the clockwise work ring tunnel (RcW_D) to the anticlockwise protection ring tunnel (RaP_D), node E also detects the failure and would switch the traffic from

anticlockwise protection ring tunnel (RaP_D) back to the clockwise work ring tunnel (RcW_D). A possible mechanism to mitigate the temporary loop problem is: the TTL of the ring tunnel label is set to 2*N by the ingress ring node of the traffic, where N is the number of nodes on the ring.

```
                     +---+#####[RaP_D(F)]######+---+
                     | F |---------------------| A | +-- LSP1
                     +---+*****[RcW_D(A)]******+---+
                     #/*                        *\#
          [RaP_D(E)]#/*[RcW_D(F)]      [RcW_D(B)]*\#RaP_D(A)
                   #/*                            *\#
                 +---+                          xxxxx
                 | E |                          x B x
                 +---+                          xxxxx
                   #\                            */#
          [RaP_D(D)]#\                [RcW_D(C)]*/#RaP_D(B)
                     #\                        */#
                     +---+*****[RcW_D(D)]****+---+
           LSP1  +-- | D |-------------------| C |
                     +---+#####[RaP_D(C)]####+---+

            -----physical links     xxxxxx  Failure Node
            *****RcW_D              ######  RaP_D
```

Figure 6. Wrapping for node failure

4.3.2. Short Wrapping

With the traditional wrapping protection scheme, Protection switching is executed at both nodes detecting the failure, consequently the traffic will be wrapped twice. This mechanism will cause additional latency and bandwidth consumption when traffic is switched to the protection path.

With short wrapping protection, data traffic switching is executed only at the upstream node detecting the failure, and data traffic leaves the ring in the protection ring tunnel at the egress node. This scheme can reduce the additional latency and bandwidth consumption when traffic is switched to the protection path.

In the traditional wrapping solution, the protection ring tunnel is a closed ring in normal state, while in the short wrapping solution, the protection ring tunnel is ended at the egress node, which is similar to the working ring tunnel. Short wrapping is easy to implement in shared ring protection because both the working and protection ring tunnels are terminated on the egress nodes. Figure 7

shows the clockwise working ring tunnel and the anticlockwise protection ring tunnel with node D as the egress node.

4.3.2.1. Short Wrapping for Link Failure

As shown in Figure 7, in normal state, LSP1 is carried by the clockwise working ring tunnel (RcW_D) through the path A->B->C->D. When a link failure between Node B and Node C occurs, Node B switches The working ring tunnel RcW_D to the protection ring tunnel RaP_D in the opposite direction. The difference occurs in the protection ring tunnel at egress node. In short wrapping protection, Rap_D ends in Node D and then traffic will be forwarded based on the LSP labels. Thus with short wrapping mechanism, LSP1 will follow the path A->B->A->F->E->D when link failure between Node B and Node C happens. For node failure, the protection with short wrapping is similar to the mechanism with link failure.

```
+----+#####[RaP_D(F)]######+----+
```

```
           | F |---------------------| A | +-- LSP1
           +---+*****[RcW_D(A)]******+---+
           #/*                          *\#
[RaP_D(E)]#/*[RcW_D(F)]      [RcW_D(B)]*\#RaP_D(A)
         #/*                              *\#
       +---+                           +---+
       | E |                           | B |
       +---+                           +---+
         #\                            *x#
[RaP_D(D)]#\                [RcW_D(C)]*x#RaP_D(B)
           #\                        *x#
           +---+*****[RcW_D(D)]****+---+
 LSP1  +-- | D |-------------------| C |
           +---+                   +---+

   ----- physical links    xxxxx Failure Link
   ****** RcW_D           ###### RaP_D
```

Figure 7. Short wrapping for link failure

4.3.2.2. Short Wrapping for Node Failure

For the failure scenarios which happen on a non-egress node, short wrapping protection switching is similar to the link failure as described in the previous section. This section specifies the scenario of egress node failure.

As shown in Figure 8, LSP1 enters the ring on node A, and leaves the ring on node D. in normal state, LSP1 is carried by the clockwise working ring tunnel (RcW_D) through the path A->B->C->D. When node D

fails, traffic of LSP1 cannot be protected by any ring tunnels which use node D as the egress node. However, before the failure location is propagated to all the ring nodes, node C switches all the traffic on the working ring tunnel RcW_D to the protection ring tunnel RaP_D in the opposite direction. When the traffic arrives at node E which also detects the failure of node D, the protection ring tunnel RaP_D cannot be used to forward traffic to node D. Since with short wrapping mechanism, protection switching can only be performed once from the working ring tunnel to the protection ring tunnel, thus node E MUST NOT switch the traffic which is already carried on the protection ring tunnel back to the working ring tunnel in the opposite direction. Instead, node E will discard the traffic received on RaP_D locally. This can avoid the temporary traffic loop when the faiulre happens on the egress node of the ring tunnel. This also illustrates one of the benefits of having separate working and protection ring tunnels in each ring direction.

```
                         +---+#####[RaP_D(F)]######+---+
                         | F |---------------------| A | +-- LSP1
                         +---+*****[RcW_D(A)]******+---+
                         #/*                        *\#
              [RaP_D(E)]#/*[RcW_D(F)]      [RcW_D(B)]*\#RaP_D(A)
                       #/*                            *\#
                     +---+                          +---+
                     | E |                          | B |
                     +---+                          +---+
                       #\                            */#
              [RaP_D(D)]#\                [RcW_D(C)]*/#RaP_D(B)
                         #\                       */#
                         xxxxx*****[RcW_D(D)]****+---+
               LSP1  +-- x D x-------------------| C |
                         xxxxx                   +---+

                 -----physical links     xxxxxx  Failure Node
                 *****RcW_D              ######  RaP_D
```

Figure 8. Short Wrapping for egress node failure

4.3.3. Steering

In ring topology, each working ring tunnel is associated with a protection ring tunnel in the opposite direction, and every node can obtain the ring topology either by configuration or via some topology discovery mechanism. The ring topology and the connectivity (Intact or Severed) between the adjacent ring nodes form the ring map. Every ring node maintains its ring map. When a failure occurs in the ring, the nodes that detect the failure via OAM mechanism will transmit the failure information in the opposite direction of the failure hop by

hop along the ring. When a node receives the message that identifies a failure, it can quickly determine the location of the fault by using the topology information that is maintained by the node and upate the ring map accordingly, then it can determine whether the

LSPs entering the ring locally need to switchover or not. For LSPs that needs to switchover, it will switch the LSPs from the working ring tunnels to its corresponding protection ring tunnels.

```
   Ring map of F                                             +--LSPl
  +-+-+-+-+-+-+-+     +---+ ###[RaP_D(F)]### +---/  +-+-+-+-+-+-+-+
  |F|A|B|C|D|E|F|     | F | ---------------- | A |  |A|B|C|D|E|F|A|
  +-+-+-+-+-+-+-+     +---+ ***[RcW_D(A)]*** +---+  +-+-+-+-+-+-+-+
   |I|I|I|S|I|I|                                     |I|I|S|I|I|I|
   +-+-+-+-+-+-+      #/*                   *\#     +-+-+-+-+-+-+
         [RaP_D(E)]  #/*          [RcW_D(B)]  *\# [RaP_D(A)]
                    #/* [RcW_D(F)]              *\#
 +-+-+-+-+-+-+-+   #/*                           *\#
 |E|F|A|B|C|D|E| +---+                            +---+ +-- LSP2
 +-+-+-+-+-+-+-+ | E |                            | B |  +-+-+-+-+-+-+-+
  |I|I|I|I|S|I|  +---+                            +---+  |B|C|D|E|F|A|B|
  +-+-+-+-+-+-+     #\*                             */#   +-+-+-+-+-+-+-+
                     #\* [RcW_D(E)]    [RcW_D(C)] */#     |I|S|I|I|I|I|
         [RaP_D(D)]   #\*                         */#      +-+-+-+-+-+-+
                       #\*                      */# [RaP_D(B)]
 +-+-+-+-+-+-+-+       +---+     [RcW_D(D)]    +---+    +-+-+-+-+-+-+-+
 |D|E|F|A|B|C|D|  +--  | D | xxxxxxxxxxxxxxxxx | C |    |C|D|E|F|A|B|C|
 +-+-+-+-+-+-+-+ LSP1  +---+     [RaP_D(C)]    +---+    +-+-+-+-+-+-+-+
  |I|I|I|I|I|S|  LSP2                                    |S|I|I|I|I|I|
  +-+-+-+-+-+-+                                          +-+-+-+-+-+-+

        ----- physical links  ***** RcW_D  ##### RaP_D
           I: Intact   S: Severed
```

Figure 9. Steering operation and protection switching

As shown in Figure 9, LSP1 enters the ring from Node A while LSP2 enters the ring from Node B, and both of them have the same destination node D.

In the normal state, LSP1 is carried by the clockwise working ring tunnel (RcW_D) through the path A->B->C->D, the label operation is: [LSP1] -> [RcW_D(B)|LSP1](NodeA) -> [RcW_D(C)| LSP1](NodeB) -> [RcW_D(D)|LSP1](NodeC) -> [LSP1] (data traffic carried by LSP1) .

LSP2 is carried by the clockwise working ring tunnel (RcW_D) throught the path B->C->D, the label operation is: [LSP2] -> [RcW_D(C)|LSP2](NodeB) -> [RcW_D(D)|LSP2](NodeC) -> [LSP2] (data traffic carried by LSP2) .

If the link between nodes C and D fails, according to the fault detection and distribution mechanisms, Node D will find out that there is a failure in the link between C and D, and it will update the link state of its ring topology, changing the link between C and D from normal to fault. In the direction that opposite to the failure position, Node D will send the state report message to Node E, informing Node E of the fault between C and D, and E will update the link state of its ring topology accordingly, changing the link between C and D from normal to fault. In this way, the state report message is sent hop by hop in the clockwise direction. Similar to Node D, Node C will send the failure information in the anti-clockwise direction.

When Node A receives the failure report message and updates the link state of its ring topology, it is aware that there is a fault on the clockwise working ring tunnel to node D (RcW_D), and LSP1 enters the ring locally and is carried by this ring tunnel, thus Node A will decide to switch the LSP1 onto the anticlockwise protection ring tunnel to node D (RaP_D). After the switchover, LSP1 will follow the path A->F->E->D, the label operation is: [LSP1] -> [RaP_D(F)|LSP1](NodeA) -> [RaP_D(E)|LSP1](NodeF) -> [RaP_D(D)|LSP1](NodeE) -> [LSP1] (data traffic carried by LSP1).

The same also apply to the operation of LSP2. When Node B updates the link state of its ring topology, and finds out that the working ring tunnel RcW_D has failed, it will switch the LSP2 to the anticlockwise protection tunnel RaP_D. After the switchover, LSP2 goes through the path B->A->F->E->D, and the label operation is: [LSP2] -> [RaP_D(A)|LSP2](NodeB) -> [RaP_D(F)|LSP2](NodeA) -> [RaP_D(E)|LSP2](NodeF) -> [RaP_D(D)|LSP2](NodeE) -> [LSP2](data traffic carried by LSP2).

Then assume the link between nodes A and B breaks down, as shown in Figure 10. Similar to the above failure case, Node B will detect a fault in the link between A and B, and it will update the link state of its ring topology, changing the link state between A and B from normal to fault. The state report message is sent hop by hop in the clockwise direction, notifying every node that there is a fault between node A and B, and every node updates the link state of its ring topology. As a result, Node A will detect a fault in the working ring tunnel to node D, and switch LSP1 to the protection ring tunnel, while Node B determine that the working ring tunnel for LSP2 still works fine, and will not perform the switchover.

```
                                                    /-- LSPl
 +-+-+-+-+-+-+-+      +----+ ###[RaP_D(F)]####  +---/  +-+-+-+-+-+-+-+
 |F|A|B|C|D|E|F|      | F | -----------------  | A |  |A|B|C|D|E|F|A|
 +-+-+-+-+-+-+-+      +----+ ***[RcW_D(A)]****  +----+  +-+-+-+-+-+-+-+
  |I|S|I|I|I|I|      #/*                          x      |S|I|I|I|I|I|
  +-+-+-+-+-+-+     #/*                            x     +-+-+-+-+-+-+
        [RaP_D(E)]  #/*[RcW_D(F)]       [RcW_D(B)]x [RaP_D(A)]
                   #/*                                x    +-- LSP2
 +-+-+-+-+-+-+-+  +----+                             +----++-+-+-+-+-+-+-+
 |E|F|A|B|C|D|E|  | E |                              | B ||B|C|D|E|F|A|B|
 +-+-+-+-+-+-+-+  +----+                             +----++-+-+-+-+-+-+-+
  |I|I|S|I|I|I|     #\*                              */#    |I|I|I|I|I|S|
  +-+-+-+-+-+-+      #\*[RcW_D(E)]    [RcW_D(C)]  */#     +-+-+-+-+-+-+
          [RaP_D(D)]  #\*                        */# [RaP_D(B)]
 +-+-+-+-+-+-+-+       #\*                      */#     +-+-+-+-+-+-+-+
 |D|E|F|A|B|C|D|       +----+ ***[RcW_D(D)]*** +----+     |C|D|E|F|A|B|C|
 +-+-+-+-+-+-+-+  +--  | D | ---------------- | C |     +-+-+-+-+-+-+-+
  |I|I|I|S|I|I|   LSP1 +----+ ###[RaP_D(C)]### +----+      |I|I|I|I|S|I|
  +-+-+-+-+-+-+   LSP2                                     +-+-+-+-+-+-+

           ----- physical links  ***** RcW_D  ##### RaP_D
```

Figure 10. Steering operation and protection switching (2)

4.4. Interconnected Ring Protection

4.4.1. Interconnected Ring Topology

Interconnected ring topology is often used in MPLS-TP networks. This document will discuss two typical interconnected ring topologies:

1. Single-node interconnected rings

   In single-node interconnected rings, the connection between the two rings is through a single node. Because the interconnection node is in fact a single point of failure, this topology should be avoided in real transport networks. Figure 10 shows the topology of single-node interconnected rings. Node C is the interconnection node between Ring1 and Ring2.

2. Dual-node interconnected rings

In dual-node interconnected rings, the connection between the two rings is through two nodes. The two interconnection nodes belong to both interconnected rings. This topology can recover from one interconnection node failure.

Figure 11 shows the topology of single-node interconnected rings. Node C is the interconnection node between Ring1 and Ring2.

```
          +---+      +---+                        +---+      +---+
          | A |------| B |-----              -----| G |------| H |
          +---+      +---+      \           /     +---+      +---+
            |                    \         /                   |
            |                     \ +---+ /                    |
            |        Ring1          | C |         Ring2        |
            |                     / +---+ \                    |
            |                    /         \                   |
          +---+      +---+      /           \     +---+      +---+
          | F |------| E |-----              -----| J |------| I |
          +---+      +---+                        +---+      +---+
```

Figure 11. Single-node interconnected rings

Figure 12 shows the topology of dual-node interconnected rings. Nodes C and Node D are the interconnection nodes between Ring1 and Ring2.

```
             +---+      +---+      +---+      +---+      +---+
             | A |------| B |------| C |------| G |------| H |
             +---+      +---+      +---+      +---+      +---+
               |                    | |                    |
               |                    | |                    |
               |        Ring1       | |       Ring2        |
               |                    | |                    |
               |                    | |                    |
             +---+      +---+      +---+      +---+      +---+
             | F |------| E |------| D |------| J |------| I |
             +---+      +---+      +---+      +---+      +---+
```

Figure 12. Dual-node interconnected rings

4.4.2. Interconnected Ring Protection Mechanisms

Interconnected rings can be treated as two independent rings. Ring protection switching (RPS) protocol operates on each ring independently. Failure in one ring only triggers protection switching on the ring itself and does not affect the other ring. This way, protection switching on each ring is the same as the mechanisms described in section 4.3.

The service LSPs that traverse the interconnected rings via the interconnection nodes MUST use different ring tunnels in different rings, and the service LSPs traversing the interconnected rings are stitched by the interconnection node. On the interconnection node,

the ring tunnel label used in the source ring will be popped, the service LSP label will be swapped, and the ring tunnel label of the destination ring will be pushed.

In the dual-node interconnected ring scenario, the two interconnection nodes can be managed as a virtual interconnection node group. Each ring should assign working and protection ring tunnels for the virtual interconnection node group. Both the interconnection nodes in the virtual interconnection node group can terminate the working ring tunnel of each ring. The protection ring tunnel is used to protect the working ring tunnel of each ring and can be terminated by any node in the virtual interconnection node group.

On the nodes in the virtual interconnection node group of the dual-node interconnected ring, the same label is allocated for each service LSP. This way any interconnection node in the virtual node group can stitch the service LSPs between the source ring tunnel and the destination ring tunnel.

When the service traffic passes through the interconnection node, the direction of the working ring tunnels in each ring for this service traffic should be the same. For example, if the working ring tunnel follows the clockwise direction in Ring1, the working ring tunnel for the same service traffic in Ring2 SHOULD also follow the clockwise direction when the service leaves Ring1 and enters Ring2.

4.4.3. Ring Tunnels in Interconnected Rings

The same ring tunnels as described in section 4.1 are used in each ring of the interconnected rings. Note that ring tunnels to the virtual interconnection node group will be established by each ring of the interconnected rings, i.e.:

- one clockwise working ring tunnel to the virtual interconnection node group
- one anticlockwise protection ring tunnel to the virtual interconnection node group
- one anticlockwise working ring tunnel to the virtual interconnection node group
- one clockwise protection ring tunnel to the virtual interconnection node group

These ring tunnels will terminated at all nodes in the virtual interconnection node group.

For example, all the ring tunnels on Ring1 of Figure 12 are established as follows:

- To Node A: R1cW_A, R1aW_A, R1cP_A, R1aP_A
- To Node B: R1cW_B, R1aW_B, R1cP_B, R1aP_B
- To Node C: R1cW_C, R1aW_C, R1cP_C, R1aP_C
- To Node D: R1cW_D, R1aW_D, R1cP_D, R1aP_D
- To Node E: R1cW_E, R1aW_E, R1cP_E, R1aP_E
- To Node F: R1cW_F, R1aW_F, R1cP_F, R1aP_F
- To the virtual interconnection node group (including Node F and Node A): R1cW_F&A, R1aW_F&A, R1cP_F&A, R1aP_F&A

All the ring tunnels established in Ring2 in Figure 13 are provisioned as follows:

o To Node A: R2cW_A, R2aW_A, R2cP_A, R2aP_A

o To Node F: R2cW_F, R2aW_F, R2cP_F, R2aP_F

o To Node G: R2cW_G, R2aW_G, R2cP_G, R2aP_G

o To Node H: R2cW_H, R2aW_H, R2cP_H, R2aP_H

o To Node I: R2cW_I, R2aW_I, R2cP_I, R2aP_I

o To Node J: R2cW_J, R2aW_J, R2cP_J, R2aP_J

o To the virtual interconnection node group(including Node F and Node A): R2cW_FandA, R2aW_FandA, R2cP_FandA, R2aP_FandA


```
                          +---+cccccccccccc +---+
                          | H |-------------| I |--->LSP1
                          +---+             +---+
                          c/a                   a\
                         c/a                     a\
                        c/a                       a\
                      +---+                     +---+
                      | G |      Ring2          | J |
                      +---+                     +---+
                        c\a                      a/c
                         c\a                    a/c
                          c\a  aaaaaaaaaaaaa   a/c
```

```
                          +---+ccccccccccccc+---+
                          | F |-------------| A |
                          +---+ccccccccccccc+---+
                          c/aaaaaaaaaaaaaaaaaaa a\
                         c/                      a\
                        c/                        a\
                      +---+                     +---+
                      | E |        Ring1        | B |
                      +---+                     +---+
                        c\a                      a/c
                         c\a                    a/c
                          c\a                  a/c
                          +---+aaaaaaaaaaaa +---+
                  LSP1--->| D |-------------| C |
                          +---+ccccccccccccc+---+

                           ccccccccccc  R1cW_F&A
                           aaaaaaaaaaa  R1aP_F&A
                           ccccccccccc  R2cW_I
                           aaaaaaaaaaa  R2aP_I
            Figure 13. Ring tunnels for the interconnected rings
```

4.4.4. Interconnected Ring Switching Procedure

As shown in Figure 13, for the service traffic LSP1 which enters Ring1 at Node D and leaves Ring1 at Node F and continues to enter Ring2 at Node F and leaves Ring2 at Node I, the protection scheme is described as below.

In normal state, LSP1 follows R1cW_F&A in Ring1 and R2CW_I in Ring2. The label used for the working ring tunnel R1cW_F&A in Ring1 is popped and the label used for the working ring tunnel R2cW_I will be pushed based the inner label lookup at the interconnection node F. The working path that the service traffic LSP1 follows is: LSP1->R1cW_F&A (D->E->F)->R2cW_I(F->G->H->I)->LSP1.

In case of link failure, for example, when a failure occurs on the link between Node F and Node E, Nodes F and E will detect the failure and execute protection switching as described in 4.3.1.1. The path that the service traffic LSP1 follows after switching change to LSP1->R1cW_F&A(D->E)->R1aP_F&A(E->D->C->B->A->F)->R1cW_F(F)

->R2cW_I(F->G->H->I)->LSP1.

In case of a non interconnection node failure, for example, when the failure occurs at Node E in Ring1, Nodes F and D will detect the failure and execute protection switching as described in 4.3.1.2. The path that the service traffic LSP1 follows after switching becomes: LSP1->R1cW_F&A(D)->R1aP_F&A(D->C->B->A->F)-> R1cW_F(F)->R2cW_I(F->G->H->I).

In case of an interconnection node failure, for example, when the failure occurs at the interconnection Node F. Nodes E and A in Ring1 will detect the failure, and execute protection switching as described in 4.3.1.2. Nodes G and A in Ring2 will also detects the failure, and execute protection switching. The path that the service traffic LSP1 follows after switching is:
LSP1->R1cW_F&A(D->E)->R1aP_F&A(E->D->C->B->A)->R1cW_A(A)
->R2aP_I(A->J->I)->LSP1.

4.4.5. Interconnected Ring Detection Mechanism

As show in Figure 14, the service traffic LSP1 traverses A->B->C in Ring1 and C->G->H->I in Ring2. Node C and Node D are the interconnection nodes. When both the link between Node C and Node G and the link between Node C and Node D fail, the ring tunnel from Node C to Node I in Ring2 becomes unreachable. However, Node D is still available, and LSP1 can still reach Node I.

```
      +---+ *********+---+**********+---+          +---+**********+---+
LSP1->| A |----------| B |----------| C |XXXXXXXXXX| G |----------| H |
      +---+##########+---+##########+---+          +---+##########+---+
        |#                            X                            #|*
        |#                            X                            #|*
        |#           Ring1            X           Ring2            #|*
        |#                            X                            #|*
        |#                            X                            #|*
      +---+##########+---+##########+---+######### +---+##########+---+
      | F |----------| E |----------| D |----------| J |----------| I | ->LSP1
      +---+          +---+          +---+          +---+          +---+

                        ***********  R1cW_C&D
                        ###########  R1aP_C&D
                        ***********  R2cW_I
                        ###########  R2aP_I
```

Figure 14. Interconnected ring

In order to achieve this, the interconnection nodes need to know the ring topology of each ring so that they can judge whether a node is reachable. This judgment is based on the knowledge of each ring topology and the fault location as described in section 3.4. The ring topology can be obtained from the NMS or topology discovery mechanisms. The fault location can be obtained by transmitting the fault information around the ring. The nodes that detect the failure will transmit the fault information in the opposite direction node by node in the ring. When the interconnection node receives the message that informs the failure, it will quickly calculate the location of the fault by the topology information that is maintained by itself and determines whether the LSPs entering the ring at itself can reach the destination. If the destination node is reachable, the LSP will leave the source ring and enter the destination ring. If the destination node is not reachable, the LSP will switch to the anticlockwise protection ring tunnel.

In Figure 14, Node C determines that the ring tunnel to Node I is unreachable, the service traffic LSP1 for which the destination node on the ring tunnel is Node I should switch to the protection LSP (R1aP_C&D) and consequently the service traffic LSP1 traverses the interconnected rings at Node D. Node D will remove the ring tunnel label of Ring1 and add the ring tunnel label of Ring2.

5. Ring Protection Coordination Protocol

5.1. RPS Protocol

The MSRP protection operation MUST be controlled with the help of the Ring Protection Switch Protocol (RPS). The RPS processes in each of the individual ring nodes that form the ring SHOULD communicate using the G-ACh channel.

The RPS protocol MUST carry the ring status information and RPS requests, i.e., automatically initiated and externally initiated, between the ring nodes.

Each node on the ring MUST be uniquely identified by assigning it a node ID. The node ID MUST be unique on each ring. The maximum number of nodes on the ring supported by the RPS protocol is 127. The node ID SHOULD be independent of the order in which the nodes appear on the ring. The node ID is used to identity the source and destination nodes of each RPS request.

Every node obtains the ring topology either by configuration or via some topology discovery mechanism. The ring map consists of the ring topology information, and connectivity status (Intact or Severed) between the adjacent ring nodes, which is determined via the OAM message exchange between the adjacent nodes. The ring map is used by every ring node to determine the switchover behavoir of the ring tunnels.

When no protection switching is active on the ring, each node MUST dispatch periodically RPS requests to the two adjacent nodes, indicating No Request (NR). When a node determines that a protection switching is required, it MUST send the appropriate RPS request in both directions.

```
                 +---+ A->B(NR)    +---+ B->C(NR)    +---+ C->D(NR)
          -------| A |-------------| B |-------------| C |-------
        (NR)F<-A +---+    (NR)A<-B +---+    (NR)B<-C +---+
```

Figure 15. RPS communication between the ring nodes in case of no failures in the ring

A destination node is a node that is adjacent to a node that identified a failed span. When a node that is not the destination node receives an RPS request and it has no higher priority local request, it MUST transfer in the same direction the RPS request as received. In this way, the switching nodes can maintain direct RPS protocol communication in the ring.


```
                +---+ C->B(SF)    +---+ B->C(SF)    +---+ C->B(SF)
         -------| A |-------------| B |----- X -----| C |-------
       (SF)C<-B +---+    (SF)C<-B +---+    (SF)B<-C +---+
```

Figure 16. RPS communication between the ring nodes in case of failure between nodes B and C

Note that in the case of a bidirectional failure such as a cable cut, the two adjacent nodes detect the failure and send each other an RPS request in opposite directions.

o In rings utilizing the wrapping protection. When the destination node receives the RPS request it MUST perform the switch from/to the working ring tunnels to/from the protection ring tunnels if it has no higher priority active RPS request.

o In rings utilizing the steering protection. When a ring switch is required, any node MUST perform the switches if its added/dropped traffic is affected by the failure. Determination of the affected traffic SHOULD be performed by examining the RPS requests (indicating the nodes adjacent to the failure or failures) and the stored ring maps (indicating the relative position of the failure and the added traffic destined towards that failure).

When the failure has cleared and the Wait-to-Restore (WTR) timer has expired, the nodes sourcing RPS requests MUST drop their respective switches (tail end) and MUST source an RPS request carrying the NR code. The node receiving from both directions such RPS request (head end) MUST drop its protection switches.

A protection switch MUST be initiated by one of the criteria specified in Section 3.2. A failure of the RPS protocol or controller MUST NOT trigger a protection switch.

Ring switches MUST be preempted by higher priority RPS requests. For example, consider a protection switch that is active due to a manual switch request on the given span, and another protection switch is required due to a failure on another span. Then an RPS request MUST be generated, the former protection switch MUST be dropped, and the latter protection switch established.

MSRP mechanism SHOULD support multiple protection switches in the ring, resulting the ring being segmented into two or more separate segments. This may happen when several RPS requests of the same priority exist in the ring due to multiple failures or external switch commands.

Proper operation of the MSRP mechanism relies on all nodes having knowledge of the state of the ring (nodes and spans) so that nodes do not preempt existing RPS request unless they have a higher-priority RPS request. In order to accommodate ring state knowledge, during a protection switch the RPS requests MUST be sent in both directions.

5.1.1. Transmission and Acceptance of RPS Requests

A new RPS request MUST be transmitted immediately when a change in the transmitted status occurs.

The first three RPS protocol messages carrying new RPS request SHOULD be transmitted as fast as possible. For fast protection switching within 50 ms, the interval of the first three RPS protocol messages SHOULD be 3.3 ms. The successive RPS requests SHOULD be transmitted with the interval of 5 seconds.

5.1.2. RPS PDU Format

Figure 17 depicts the format of an RPS packet that is sent on the G-ACh. The Channel Type field is set to indicate that the message is an RPS message. The ACH MUST NOT include the ACH TLV Header [RFC5586] meaning that no ACH TLVs can be included in the message.

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0 0 0 1|0 0 0 0|0 0 0 0 0 0 0 0|    RPS Channel Type (TBD)     |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Dest Node ID  | Src Node ID   |    Request     |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 17. G-ACh RPS Packet Format

The following fields MUST be provided:

- o Destination Node ID: The destination node ID MUST always be set to value of the node ID of the adjacent node. The Node ID MUST be unique on each ring. Valid destination node ID values are 1-127.

- o Source node ID: The source node ID MUST always be set to the ID value of the node generating the RPS request. The Node ID MUST be unique on each ring. Valid source node ID values are 1-127.

- o RPS request code: A code consisting of eight bits as specified below:

Bits (MSB - LSB)	Condition, State or external Request	Priority
0 0 0 0 1 1 1 1	Lockout of Protection (LP)	highest
0 0 0 0 1 1 0 1	Forced Switch (FS)	
0 0 0 0 1 0 1 1	Signal Fail (SF)	
0 0 0 0 0 1 1 0	Manual Switch (MS)	
0 0 0 0 0 1 0 1	Wait-To-Restore (WTR)	
0 0 0 0 0 0 1 1	Exercise (EXER)	
0 0 0 0 0 0 0 1	Reverse Request (RR)	
0 0 0 0 0 0 0 0	No Request (NR)	lowest

5.1.3. Ring Node RPS States

Idle state: A node is in the idle state when it has no RPS request and is sourcing and receiving NR code to/from both directions.

Switching state: A node not in the idle or pass-through states is in the switching state.

Pass-through state: A node is in the pass-through state when its highest priority RPS request is a request not destined to it or sourced by it. The pass-through is bidirectional.

5.1.3.1. Idle State

A node in the idle state MUST source the NR request in both directions.

A node in the idle state MUST terminate RPS requests flow in both directions.

A node in the idle state MUST block the traffic flow on protection LSPs/tunnels in both directions.

5.1.3.2. Switching State

A node in the switching state MUST source RPS request to adjacent node with its highest RPS request code in both directions when it detects a failure or receives an external command.

A node in the switching state MUST terminate RPS requests flow in both directions.

As soon as it receives an RPS request from the short path, the node to which it is addressed MUST acknowledge the RPS request by replying with the RR code on the short path, and with the received RPS request code on the long path. Here the short path refers to the shorter span on the ring between the source and destination node of the RPS request, and the long path refers to the longer span on the ring between the source and destination node of the RPS request.

This rule refers to the unidirectional failure detection: the RR SHOULD be issued only when the node does not detect the failure condition (i.e., the node is a head end), that is, it is not applicable when a bidirectional failure is detected, because, in this case, both nodes adjacent to the failure will send an RPS request for the failure on both paths (short and long).

The following switches MUST be allowed to coexist:

o LP and LP

o FS and FS

o SF and SF

o FS and SF

When multiple MS RPS requests over different spans exist at the same time, no switch SHOULD be executed and existing switches MUST be dropped. The nodes MUST signal, anyway, the MS RPS request code.

Multiple EXER requests MUST be allowed to coexist in the ring.

A node in a ring switching state that receives the external command LP for the affected span MUST drop its switch and MUST signal NR for the locked span if there is no other RPS request on another span. Node still SHOULD signal relevant RPS request for another span.

5.1.3.3. Pass-through State

When a node is in a pass-through state, it MUST transfer the received RPS Request in the same direction.

When a node is in a pass-through state, it MUST enable the traffic flow on protection ring tunnels in both directions.

5.1.4. RPS State Transitions

All state transitions are triggered by an incoming RPS request change, a WTR expiration, an externally initiated command, or locally detected MPLS-TP section failure conditions.

RPS requests due to a locally detected failure, an externally initiated command, or received RPS request shall pre-empt existing RPS requests in the prioritized order given in Section 3.1.2, unless the requests are allowed to coexist.

5.1.4.1. Transitions Between Idle and Pass-through States

The transition from the idle state to pass-through state MUST be triggered by a valid RPS request change, in any direction, from the NR code to any other code, as long as the new request is not destined to the node itself. Both directions move then into a pass-through state, so that, traffic entering the node through the protection Ring tunnels are transferred transparently through the node.

A node MUST revert from pass-through state to the idle state when it detects NR codes incoming from both directions. Both directions revert simultaneously from the pass-through state to the idle state.

5.1.4.2. Transitions Between Idle and Switching States

Transition of a node from the idle state to the switching state MUST be triggered by one of the following conditions:

- o A valid RPS request change from the NR code to any code received on either the long or the short path and destined to this node
- o An externally initiated command for this node
- o The detection of an MPLS-TP section layer failure at this node

Actions taken at a node in the idle state upon transition to switching state are:

- o For all protection switch requests, except EXER and LP, the node MUST execute the switch
- o For EXER, and LP, the node MUST signal appropriate request but not execute the switch

A node MUST revert from the switching state to the idle state when it detects NR codes received from both directions.

o At the tail end: When a WTR time expires or an externally initiated command is cleared at a node, the node MUST drop its switch, transit to the Idle State and signal the NR code in both directions.

o At the head end: Upon reception of the NR code, from both directions, the head-end node MUST drop its switch, transition to Idle State and signal the NR code in both directions.

5.1.4.3. Transitions Between Switching States

When a node that is currently executing any protection switch receives a higher priority RPS request (due to a locally detected failure, an externally initiated command, or a ring protection switch request destined to it) for the same span, it MUST update the priority of the switch it is executing to the priority of the received RPS request.

When a failure condition clears at a node, the node MUST enter WTR condition and remain in it for the appropriate time-out interval, unless:

o A different RPS request with a higher priority than WTR is received

o Another failure is detected

o An externally initiated command becomes active

The node MUST send out a WTR code on both the long and short paths.

When a node that is executing a switch in response to incoming SF RPS request (not due to a locally detected failure) receives a WTR code (unidirectional failure case), it MUST send out RR code on the short path and the WTR on the long path.

5.1.4.4. Transitions Between Switching and Pass-through States

When a node that is currently executing a switch receives an RPS request for a non-adjacent span of higher priority than the switch it is executing, it MUST drop its switch immediately and enter the pass-through state.

The transition of a node from pass-through to switching state MUST be triggered by:

o An equal priority, a higher priority, or an allowed coexisting externally initiated command

- o The detection of an equal priority, a higher priority, or an allowed coexisting automatic initiated command

- o The receipt of an equal, a higher priority, or an allowed coexisting RPS request destined to this node

5.2. RPS State Machine

5.2.1. Initial States

	State	Signaled RPS
A	Idle Working: no switch Protection: no switch	NR
B	Pass-trough Working: no switch Protection: pass through	N/A
C	Switching - LP Working: no switch Protection: no switch	LP
D	Idle - LW Working: no switch Protection: no switch	NR
E	Switching - FS Working: switched Protection: switched	FS
F	Switching - SF Working: switched Protection: switched	SF
G	Switching - MS Working: switched Protection: switched	MS
H	Switching - WTR Working: switched Protection: switched	WTR
I	Switching - EXER	EXER

```
   |      |  Working: no switch          |                |
   |      |  Protection: no switch       |                |
   +------+------------------------------+----------------+
```

5.2.2. State transitions When Local Request is Applied

In the state description below 'O' means that new local request will be rejected because of exiting request.

```
   =====================================================================
   Initial state        New request        New state
   -------------        -----------        ---------
```



```
   A (Idle)             LP                 C (Switching - LP)
                        LW                 D (Idle - LW)
                        FS                 E (Switching - FS)
                        SF                 F (Switching - SF)
                        Recover from SF    N/A
                        MS                 G (Switching - MS)
                        Clear              N/A
                        WTR expires        N/A
                        EXER               I (Switching - EXER)
   =====================================================================
   Initial state        New request        New state
   -------------        -----------        ---------
   B (Pass-trough)      LP                 C (Switching - LP)
                        LW                 B (Pass-trough)
                        FS                 O - if current state is due to
                                               LP sent by another node
                                           E (Switching - FS) - otherwise
                        SF                 O - if current state is due to
                                               LP sent by another node
                                           F (Switching - SF) - otherwise
                        Recover from SF    N/A
                        MS                 O - if current state is due to
                                               LP, SF or FS sent by
                                               another node
                                           G (Switching - MS) - otherwise
                        Clear              N/A
                        WTR expires        N/A
                        EXER               O
   =====================================================================
```

```
   Initial state         New request        New state
   -------------         -----------        ---------
   C (Switching - LP)    LP                 N/A
                         LW                 O
                         FS                 O
                         SF                 O
                         Recover from SF    N/A
                         MS                 O
                         Clear              A (Idle) - if there is no
                                               failure in the ring
                                            F (Switching - SF) - if there
                                               is a failure at this node
                                            B (Pass-trough) - if there is
                                               a failure at another node
                         WTR expires        N/A
                         EXER               O
   ======================================================================
   Initial state         New request        New state
   -------------         -----------        ---------
```



```
   D (Idle - LW)         LP                 C (Switching - LP)
                         LW                 N/A - if on the same span
                                            D (Idle - LW) - if on another
                                               span
                         FS                 O - if on the same span
                                            E (Switching - FS) - if on
                                               another span
                         SF                 O - if on the addressed span
                                            F (Switching - SF) - if on
                                               another span
                         Recover from SF    N/A
                         MS                 O - if on the same span
                                            G (Switching - MS) - if on
                                               another span
                         Clear              A (Idle) - if there is no
                                               failure on addressed span
                                            F (Switching - SF) - if there
                                               is a failure on this span
                         WTR expires        N/A
                         EXER               O
   ======================================================================
   Initial state         New request        New state
```

```
   -------------        -----------       ---------
   E (Switching - FS)   LP                C (Switching - LP)
                        LW                O - if on another span
                                          D (Idle - LW) - if on the same
                                             span
                        FS                N/A - if on the same span
                                          E (Switching - FS) - if on
                                             another span
                        SF                O - if on the addressed span
                                          E (Switching - FS) - if on
                                             another span
                        Recover from SF   N/A
                        MS                O
                        Clear             A (Idle) - if there is no
                                             failure in the ring
                                          F (Switching - SF) - if there
                                             is a failure at this node
                                          B (Pass-trough) - if there is
                                             a failure at another node
                        WTR expires       N/A
                        EXER              O
   ==================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   F (Switching - SF)   LP                C (Switching - LP)
                        LW                O - if on another span
```



```
                                          D (Idle - LW) - if on the same
                                             span
                        FS                E (Switching - FS)
                        SF                N/A - if on the same span
                                          F (Switching - SF) - if on
                                             another span
                        Recover from SF   H (Switching - WTR)
                        MS                O
                        Clear             N/A
                        WTR expires       N/A
                        EXER              O
   ==================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   G (Switching - MS)   LP                C (Switching - LP)
```

```
                       LW                 O - if on another span
                                          D (Idle - LW) - if on the same
                                             span
                       FS                 E (Switching - FS)
                       SF                 F (Switching - SF)
                       Recover from SF    N/A
                       MS                 N/A - if on the same span
                                          G (Switching - MS) - if on
                                             another span release the
                                             switches but signal MS
                       Clear              A
                       WTR expires        N/A
                       EXER               O
   ====================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   H (Switching - WTR)  LP                C (Switching - LP)
                        LW                D (Idle - W)
                        FS                E (Switching - FS)
                        SF                F (Switching - SF)
                        Recover from SF   N/A
                        MS                G (Switching - MS)
                        Clear             A
                        WTR expires       A
                        EXER              O
   ====================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   I (Switching - EXER) LP                C (Switching - LP)
                        LW                D (idle - W)
                        FS                E (Switching - FS)
                        SF                F (Switching - SF)
                        Recover from SF   N/A
```



```
                        MS                G (Switching - MS)
                        Clear             A
                        WTR expires       N/A
                        EXER              N/A - if on the same span
                                          I (Switching - EXER)
   ====================================================================
```

5.2.3. State Transitions When Remote Request is Applied

The priority of a remote request does not depend on the side from which the request is received.

Initial state	New request	New state
A (Idle)	LP	C (Switching - LP)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	MS	G (Switching - MS)
	WTR	N/A
	EXER	I (Switching - EXER)
	RR	N/A
	NR	A (Idle)

Initial state	New request	New state
B (Pass-trough)	LP	C (Switching - LP)
	FS	N/A - cannot happen when there is LP request in the ring E (Switching - FS) - otherwise
	SF	N/A - cannot happen when there is LP request in the ring F (Switching - SF) - otherwise
	MS	N/A - cannot happen when there is LP, FS or SF request in the ring G (Switching - MS) - otherwise
	WTR	N/A - cannot happen when there is LP, FS, SF or MS request in the ring
	EXER	N/A - cannot happen when there is LP, FS, SF, MS or WTR request in the ring I (Switching - EXER) - otherwise
	RR	N/A
	NR	A (Idle) - if received from

both sides

Initial state	New request	New state
C (Switching - LP)	LP	C (Switching - LP)
	FS	N/A - cannot happen when there is LP request in the ring
	SF	N/A - cannot happen when there is LP request in the ring
	MS	N/A - cannot happen when there is LP request in the ring
	WTR	N/A
	EXER	N/A - cannot happen when there is LP request in the ring
	RR	C (Switching - LP)
	NR	N/A

Initial state	New request	New state
D (Idle - LW)	LP	C (Switching - LP)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	MS	G (Switching - MS)
	WTR	N/A
	EXER	I (Switching - EXER)
	RR	N/A
	NR	D (Idle - LW)

Initial state	New request	New state
E (Switching - FS)	LP	C (Switching - LP)
	FS	E (Switching - FS)
	SF	E (Switching - FS)
	MS	N/A - cannot happen when there is FS request in the ring
	WTR	N/A
	EXER	N/A - cannot happen when there is FS request in the ring
	RR	E (Switching - FS)
	NR	N/A

Initial state	New request	New state
F (Switching - SF)	LP	C (Switching - LP)
	FS	F (Switching - SF)
	SF	F (Switching - SF)
	MS	N/A - cannot happen when there is SF request in the ring

```
                      WTR                 N/A
                      EXER                N/A - cannot happen when there
                                               is SF request in the ring
                      RR                  F (Switching - SF)
                      NR                  N/A
   ======================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   G (Switching - MS)   LP                C (Switching - LP)
                        FS                E (Switching - FS)
                        SF                F (Switching - SF)
                        MS                G (Switching - MS) - release
                                             the switches but signal MS
                        WTR               N/A
                        EXER              N/A - cannot happen when there
                                               is MS request in the ring
                        RR                G (Switching - MS)
                        NR                N/A
   ======================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   H (Switching - WTR)  LP                C (Switching - LP)
                        FS                E (Switching - FS)
                        SF                F (Switching - SF)
                        MS                G (Switching - MS)
                        WTR               H (Switching - WTR)
                        EXER              N/A - cannot happen when there
                                              is WTR request in the ring
                        RR                H (Switching - WTR)
                        NR                N/A
   ======================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   I (Switching - EXER) LP                C (Switching - LP)
                        FS                E (Switching - FS)
                        SF                F (Switching - SF)
                        MS                G (Switching - MS)
                        WTR               N/A
                        EXER              I (Switching - EXER)
                        RR                I (Switching - EXER)
                        NR                N/A
   ======================================================================
```

5.2.4. State Transitions When Request Addresses to Another Node is Received

The priority of a remote request does not depend on the side from which the request is received.

```
=====================================================================
Initial state         New request        New state
-------------         -----------        ---------
A (Idle)              LP                 B (Pass-trough)
                      FS                 B (Pass-trough)
                      SF                 B (Pass-trough)
                      MS                 B (Pass-trough)
                      WTR                B (Pass-trough)
                      EXER               B (Pass-trough)
                      RR                 N/A
                      NR                 N/A
=====================================================================
Initial state         New request        New state
-------------         -----------        ---------
B (Pass-trough)       LP                 B (Pass-trough)
                      FS                 N/A - cannot happen when there
                                              is LP request in the ring
                                         B (Pass-trough) - otherwise
                      SF                 N/A - cannot happen when there
                                              is LP request in the ring
                                         B (Pass-trough) - otherwise
                      MS                 N/A - cannot happen when there
                                               is LP, FS or SF request
                                               in the ring
                                         B (Pass-trough) - otherwise
                      WTR                N/A - cannot happen when there
                                               is LP, FS, SF or MS
                                               request in the ring
                                         B (Pass-trough) - otherwise
                      EXER               N/A - cannot happen when there
                                               is LP, FS, SF, MS or WTR
                                               request in the ring
                                         B (Pass-trough) - otherwise
```

```
                       RR                 N/A
                       NR                 B (Pass-trough)
   ===================================================================
   Initial state       New request        New state
   -------------       -----------        ---------
   C (Switching - LP)  LP                 C (Switching - LP)
                       FS                 N/A - cannot happen when there
                                               is LP request in the ring
                       SF                 N/A - cannot happen when there
```



```
                                               is LP request in the ring
                       MS                 N/A - cannot happen when there
                                               is LP request in the ring
                       WTR                N/A - cannot happen when there
                                               is LP in the ring
                       EXER               N/A - cannot happen when there
                                               is LP request in the ring
                       RR                 N/A
                       NR                 N/A
   ===================================================================
   Initial state       New request        New state
   -------------       -----------        ---------
   D (Idle - LW)       LP                 B (Pass-trough)
                       FS                 B (Pass-trough)
                       SF                 B (Pass-trough)
                       MS                 B (Pass-trough)
                       WTR                B (Pass-trough)
                       EXER               B (Pass-trough)
                       RR                 N/A
                       NR                 N/A
   ===================================================================
   Initial state       New request        New state
   -------------       -----------        ---------
   E (Switching - FS)  LP                 B (Pass-trough)
                       FS                 E (Switching - FS)
                       SF                 E (Switching - FS)
                       MS                 N/A - cannot happen when there
                                               is FS request in the ring
                       WTR                N/A - cannot happen when there
                                               is FS request in the ring
                       EXER               N/A - cannot happen when there
                                               is FS request in the ring
```

```
                    RR                N/A
                    NR                N/A
   ====================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   F (Switching - SF)   LP                B (Pass-trough)
                        FS                F (Switching - SF)
                        SF                F (Switching - SF)
                        MS                N/A - cannot happen when there
                                               is SF request in the ring
                        WTR               N/A - cannot happen when there
                                               is SF request in the ring
                        EXER              N/A - cannot happen when there
                                               is SF request in the ring
                        RR                N/A
                        NR                N/A
```



```
   ====================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   G (Switching - MS)   LP                B (Pass-trough)
                        FS                B (Pass-trough)
                        SF                B (Pass-trough)
                        MS                G (Switching - MS) - release
                                             the switches but signal MS
                        WTR               N/A - cannot happen when there
                                               is MS request in the ring
                        EXER              N/A - cannot happen when there
                                               is MS request in the ring
                        RR                N/A
                        NR                N/A
   ====================================================================
   Initial state        New request       New state
   -------------        -----------       ---------
   H (Switching - WTR)  LP                B (Pass-trough)
                        FS                B (Pass-trough)
                        SF                B (Pass-trough)
                        MS                B (Pass-trough)
                        WTR               N/A
                        EXER              N/A - cannot happen when there
                                              is WTR request in the ring
                        RR                N/A
```

Initial state	New request	New state
	NR	N/A
Initial state	New request	New state
I (Switching - EXER)	LP	B (Pass-trough)
	FS	B (Pass-trough)
	SF	B (Pass-trough)
	MS	B (Pass-trough)
	WTR	N/A
	EXER	I (Switching - EXER)
	RR	N/A
	NR	N/A

5.3. RPS and PSC Comparison on Ring Topology

This section provides comparison between RPS and PSC [RFC6378] [RFC6974] on ring topologies. This can be helpful to explain the reason of defining a new protocol for ring protection switching.

The PSC protocol [RFC6378] is designed for point-to-point LSPs, on which the protection switching can only be performed on one or both of the end points of the LSP. While RPS is designed for ring

tunnels, which consist of multiple ring nodes, and the failure could happen on any segment of the ring, thus RPS SHOULD be capable of identifying and handling the different failures on the ring, and coordinating the protection switching behavior of all the nodes on the ring. As specified in section 5, this is achieved with the introduction of the "Pass-Through" state for the ring nodes, and the location of the protection request is identified via the Node IDs in the RPS Request message.

Taking a ring topology with N nodes as example:

With the mechanism specified in RFC6974, on every ring-node, a linear protection configuration has to be provisioned with every other node in the ring, i.e. with (N-1) other nodes. This means that on every ring node there will be (N-1) instances of the PSC protocol. And in order to detect faults and to transport the PSC message, each instance shall have a MEP on the working path and a MEP on the protection path respectively. This means that every node on the ring needs to be configured with (N-1) * 2 MEPs.

With the mechanism defined in this document, on every ring node there will only be a single instance of the RPS protocol. In order to detect faults and to transport the RPS message, each node only needs to have a MEP on the section to its adjacent nodes respectively. In this way, every ring-node only needs to be configured with 2 MEPs.

As shown in the above example, RPS is designed for ring topologies and can achieve ring protection efficiently with minimum protection instances and OAM entities, which meets the requirements on topology specific recovery mechanisms as specified in [RFC5654].

6. IANA Considerations

IANA is requested to administer the assignment of new values defined in this document and summarized in this section.

6.1. G-ACh Channel Type

The Channel Types for the Generic Associated Channel (GACH) are allocated from the IANA PW Associated Channel Type registry defined in [RFC4446] and updated by [RFC5586].

IANA is requested to allocate a new GACH Channel Type as follows:

Value	Description	Reference
TBD	Ring Protection Switching Protocol (RPS)	this document

6.2. RSP Request Codes

IANA is requested to create a new sub-registry under the "Multiprotocol Label Switching (MPLS) Operations, Administration, and Management (OAM) Parameters" registry called the "MPLS RPS Request Code Registry". All code points within this registry shall be allocated according to the "Standards Action" procedure as specified in [RFC5226].

The RPS Request Field is 8 bits, the allocated values are as follows:

Value	Description	Reference
0	No Request (NR)	this document
1	Reverse Request (RR)	this document
2	not assigned	
3	Exercise (EXER)	this document
4	not assigned	
5	Wait-To-Restore (WTR)	this document
6	Manual Switch (MS)	this document
7-10	not assigned	
11	Signal Fail (SF)	this document
12	not assigned	
13	Forced Switch (FS)	this document
14	not assigned	
15	Lockout of Protection (LP)	this document
16-255	not assigned	

7. Security Considerations

The RPS protocol defined in this document is carried in the G-ACh [RFC5586], which is a generalization of the Associated Channel defined in [RFC4385]. The security considerations specified in these documents apply to the proposed RPS mechanism.

8. Contributing Authors

Wen Ye, Minxue Wang, Sheng Liu (China Mobile)

Guanghui Sun (Huawei)

9. Acknowledgements

The authors would like to thank Gregory Mirsky, Yimin Shen, Eric Osborne and Spencer Jackson for their valuable comments and suggestions.

10. References

10.1. 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, <http://www.rfc-editor.org/info/rfc2119>.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <http://www.rfc-editor.org/info/rfc3031>.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, DOI 10.17487/RFC4385, February 2006, <http://www.rfc-editor.org/info/rfc4385>.

[RFC4446] Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)", BCP 116, RFC 4446, DOI 10.17487/RFC4446, April 2006, <http://www.rfc-editor.org/info/rfc4446>.

[RFC5586] Bocci, M., Ed., Vigoureux, M., Ed., and S. Bryant, Ed., "MPLS Generic Associated Channel", RFC 5586, DOI 10.17487/RFC5586, June 2009, <http://www.rfc-editor.org/info/rfc5586>.

[RFC5654] Niven-Jenkins, B., Ed., Brungard, D., Ed., Betts, M., Ed., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, DOI 10.17487/RFC5654, September 2009, <http://www.rfc-editor.org/info/rfc5654>.

[RFC6371] Busi, I., Ed. and D. Allan, Ed., "Operations, Administration, and Maintenance Framework for MPLS-Based Transport Networks", RFC 6371, DOI 10.17487/RFC6371, September 2011, <http://www.rfc-editor.org/info/rfc6371>.

10.2. Informative References

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, DOI 10.17487/RFC5226, May 2008, <http://www.rfc-editor.org/info/rfc5226>.

[RFC6378] Weingarten, Y., Ed., Bryant, S., Osborne, E., Sprecher, N., and A. Fulignoli, Ed., "MPLS Transport Profile (MPLS-TP) Linear Protection", RFC 6378, DOI 10.17487/RFC6378, October 2011, <http://www.rfc-editor.org/info/rfc6378>.

[RFC6974] Weingarten, Y., Bryant, S., Ceccarelli, D., Caviglia, D., Fondelli, F., Corsi, M., Wu, B., and X. Dai, "Applicability of MPLS Transport Profile for Ring Topologies", RFC 6974, DOI 10.17487/RFC6974, July 2013, <http://www.rfc-editor.org/info/rfc6974>.

Authors' Addresses

Weiqiang Cheng
China Mobile

Email: chengweiqiang@chinamobile.com

Lei Wang
China Mobile

Email: wangleiyj@chinamobile.com

Han Li
China Mobile

Email: lihan@chinamobile.com

Huub van Helvoort
Hai Gaoming BV

Email: huubatwork@gmail.com

Kai Liu
Huawei Technologies

Email: alex.liukai@huawei.com

Jie Dong
Huawei Technologies

Email: jie.dong@huawei.com

Jia He
Huawei Technologies

Email: hejia@huawei.com

Fang Li
China Academy of Telecommunication Research, MIIT., China

Email: lifang@ritt.cn

Jian Yang
ZTE Corporation P.R.China

Email: yang.jian90@zte.com.cn

Junfang Wang
Fiberhome Telecommunication Technologies Co., LTD.

Email: wjf@fiberhome.com.cn