Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 18, 2013

P. Pan
R. Rao
B. Lu
(Infinera)
L. Fang
(Cisco)
A. Malis
(Verizon)
F. Zhang
S. Aldrin
(Huawei)
F. Zhang
(ZTE)
S. Singamsetty
(Cisco)

October 15, 2012

Supporting Shared Mesh Protection in MPLS-TP Networks
draft-pan-shared-mesh-protection-05.txt

Abstract

Shared mesh protection is a common protection and recovery mechanism in transport networks, where multiple paths can share the same set of network resources for protection purposes.

In the context of MPLS-TP, it has been explicitly requested as a part of the overall solution (Req. 67, 68 and 69 in RFC5654 [RFC5654]).

It's important to note that each MPLS-TP LSP may be associated with transport network resources. In event of network failure, it may require explicit activation on the protecting paths before switching user traffic over.

In this memo, we define a lightweight signaling mechanism for protecting path activation in shared mesh protection-enabled MPLS-TP networks.

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

Copyright Notice

Table of Contents

1. Introduction

Shared mesh protection (SMP) is a common traffic protection mechanism in transport networks, where multiple paths can share the same set of network resources for protection purposes.

In the context of MPLS-TP, it has been explicitly requested as a part of the overall solution (Req. 67, 68 and 69 in RFC5654 [RFC5654]).Its operation has been further outlined in Section 4.7.6 of MPLS-TP Survivability Framework.

It's important to note that each MPLS-TP LSP may be associated with transport network resources. In event of network failure, it may require explicit activation on the protecting paths before switching user traffic over.

In this memo, we define a lightweight signaling mechanism for protecting path activation in shared mesh protection-enabled MPLS-TP networks.

Here are the key design goals:

1. Fast: The protocol is to activate the previously configured protecting paths in a timely fashion, with minimal transport and processing overhead. The goal is to support 50msec end-to-end traffic switch-over in large transport networks.

2. Reliable message delivery: Activation and deactivation operation have serious impact on user traffic. This requires the protocol to adapt a low-overhead reliable messaging mechanism. The activation messages may either traverse through a "trusted" transport channel, or require some level of built-in reliability mechanism.

3. Modular: Depending on deployment scenarios, the signaling may need to support functions such as preemption, resource re-allocation and bi-directional activation in a modular fashion.

2. Acronyms

This draft uses the following acronyms:

- SMP: Shared Mesh Protection

- LO: Lockout of protection

- o DNR: Do not revert

- o FS: Forced Switch

- o SF: Signal Fail

- o SD: Signal Degrade

- o MS: Manual Switch

- o NR: No Request

- o WTR: Wait-to-Restore

- o EXER: Exercise

- o RR: Reverse Request

- o ACK: Acknowledgement

- o NACK: Negative Acknowledgement

- o G-ACh: Generic Associated Channel

- o MPLS-TP Transport Profile for MPLS

3. Solution Overview

In this section, we describe the SMP operation in the context of MPLS-TP networks, and outline some of the relevant definitions.

We refer to the figure below for illustration:

```
            ----- B ------- C ----
           /                      \
          /                        \
         A                          D
          \                        /
           \                      /
            ==== E === F === G ===
           /                      \
          /                        \
         H                          K
          \                        /
           \                      /
            ----- I ------- J ----
```

Working paths: X = {A, B, C, D}, Y = {H, I, J, K}

Protecting paths: X' = {A, E, F, G, D}, Y' = {H, E, F, G, K}

The links between E, F and G are shared by both protecting paths. All paths are established via MPLS-TP control plane prior to network failure.

All paths are assumed to be bi-directional. An edge node is denoted as a headend or tailend for a particular path in accordance to the path setup direction.

Initially, the operators setup both working and protecting paths. During setup, the operators specify the network resources for each path. The working path X and Y will configure the appropriate resources on the intermediate nodes, however, the protecting paths, X' and Y', will reserve the resources on the nodes, but won't occupy them.

Depending on network planning requirements (such as SRLG), X' and Y' may share the same set of resources on node E, F and G. The resource assignment is a part of the control-plane CAC operation taking place on each node.

At some time, link B-C is cut. Node A will detect the outage, and initiate activation messages to bring up the protecting path X'. The intermediate nodes, E, F and G will program the switch fabric and configure the appropriate resources. Upon the completion of the activation, A will switch the user traffic to X'.

The operation may have extra caveat:

1. Preemption: Protecting paths X' and Y' may share the same resources on node E, F or G due to resource constraints. Y' has higher priority than that of X'. In the previous example, X' is up and running. When there is a link outage on I-J, H can activate its protecting path Y'. On E, F or G, Y' can take over the resources from X' for its own traffic. The behavior is acceptable with the condition that A should be notified about the preemption action.

2. Over-subscription (1:N): A unit of network resource may be reserved by one or multiple protecting paths. In the example, the network resources on E-F and F-G are shared by two protecting paths, X' and Y'. In deployment, the over-subscription ratio is an important factor on network resource utilization.

3.1. Protection Switching

The entire activation and switch-over operation need to be within the range of milliseconds to meet customer's expectation. This section illustrates how this may be achieved on MPLS-TP-enabled transport switches. Note that this is for illustration of protection switching operation, not mandating the implementation itself.

The diagram below illustrates the operation:

```
                         +---------------+
            Control      |    MPLS-TP    |     Control
      <=== Signaling ====| Control Plane |=== Signaling ===>
                         +---------------+
                           /           \
                          /             \ (MPLS label assignment)
                         /               \
                        /                 \
                  +--------+   +-------+   +--------+
     Activation   |Line    |   |Switch|   |Line    |   Activation
 <=== Messages ===|Module  |===|Fabric|===|Module  |=== Messages ===>
                  +--------+   +-------+   +--------+
```

Typical MPLS-TP user flows (or, LSP's) are bi-directional, and setup as co-routed or associated tunnels, with a MPLS label for each of the upstream and downstream traffic. On this particular type of transport switch, the control-plane can download the labels to the line modules. Subsequently, the line module will maintain a label lookup table on all working and protecting paths.

Upon the detection of network failure, the headend nodes will transmit activation messages along the MPLS LSP's. When receiving the messages, the line modules can locate the associated protecting path from the label lookup table, and perform activation procedure by programming the switching fabric directly. Upon its success, the line module will swap the label, and forward the activation messages to the next hop.

In summary, the activation procedure involves efficient path lookup and switch fabric re-programming.

To achieve the tight end-to-end switch-over budget, it's possible to implement the entire activation procedure with hardware-assistance (such as in FPGA or ASIC). The activation messages are encapsulated with a MPLS-TP Generic Associated Channel Header (GACH) [RFC5586]. Detailed message encoding is explained in later sections.

3.2. Operation Overview

To achieve high performance, the activation procedure is designed to be simple and straightforward on the network nodes.

In this section, we describe the activation procedure using the same figure shown before:

```
          ----- B ------- C ----
         /                      \
        /                        \
       A                          D
        \                        /
         \                      /
          ==== E === F === G ===
         /                      \
        /                        \
       H                          K
        \                        /
         \                      /
          ----- I ------- J ----
```

Working paths: X = {A, B, C, D}, Y = {H, I, J, K}

Protecting paths: X' = {A, E, F, G, D}, Y' = {H, E, F, G, K}

Upon the detection of working path failure, the edge nodes, A, D, H and K may trigger the activation messages to activate the protecting paths, and redirect user traffic immediately after.

We assume that there is a consistent definition of priority levels among the paths throughout the network. At activation time, each node may rely on the priority levels to potentially preempt other paths.

When the nodes detect path preemption on a particular node, they should inform all relevant nodes to free the resources by sending out notification messages. Upon the reception of notification messages, the relevant nodes will send out de-activation messages.

To optimize traffic protection and resource management, each headend may periodically poll the protecting paths about resource availability. The intermediate nodes have the option to inform the current resource utilization.

Note that, upon the detection of a working path failure, both headend

and tailend may initiate the activation simultaneously (known as bi-directional activation). This may expedite the activation time. However, both headend and tailend nodes need to coordinate the order of protecting paths for activation, since there may be multiple protecting paths for each working path (i.e., 1:N protection). For clarity, we will describe the operation from headend in the memo. The tailend operation will be available in the subsequent revisions.

4. SMP Message and Action Definition

4.1. Protection Switching Control (PSC) Logic

Protection switching processes the local triggers described in requirements 74-79 of [RFC5654] together with inputs received from the tailend node. Based on these inputs the headend will take SMP actions, and transmit different protocol messages.

Here, we reuse the switching control logic described in MPLS Linear Protection, with the following logical decomposition at headend node:

```
        Server Indication     Control Plane Indication
        -----------------+  +-------------
      Operator Command   |  |   OAM Indication
      ----------------+  |  |  +---------------
                      |  |  |  |
                      V  V  V  V
                   +---------------+         +-------+
                   | Local Request |<--------|  WTR  |
                   |    logic      |WTR Exps | Timer |
                   +---------------+         +-------+
                          |                      ^
             Highest local|request               |
                          V                      | Start/Stop
                  +-----------------+            |
      Remote PSC  |  PSC  Control   |------------+
     ------------>|      logic      |
        Request   +-----------------+
                          |
                          |  Action         +------------+
                          +---------------->|  Message   |
                                            | Generator  |
                                            +------------+
                                                  |
                                     Output PSC | Message
                                                  V
```

The Local Request logic unit accepts the triggers from the OAM, external operator commands, from the local control plane (when present), and the Wait-to-Restore timer. By considering all of these local request sources it determines the highest priority local request. This high-priority request is passed to the PSC Control logic, that will cross-check this local request with the information received from the tailend node. The PSC Control logic uses this input to determine what actions need to be taken, e.g. local actions at the headend, or what message should be sent to the tailend node.

Specifically, the signals could be the following:

- o Clear - if the operator cancels an active local administrative command, i.e. LO/FS/MS.

- o Lockout of Protection (LO) - if the operator requested to prevent switching data traffic to the protection path, for any purpose.

- o Signal Fail (SF) - if any of the Server Layer, Control plane, or OAM indications signaled a failure condition on either the protection path or one of the working paths.

- o Signal Degrade (SD) - if any of the Server Layer, Control plane, or OAM indications signaled a degraded transmission condition on either the protection path or one of the working paths.

- o Forced Switch (FS) - if the operator requested that traffic be switched from one of the working paths to the protection path,

- o Manual Switch (MS) - if the operator requested that traffic is switched from the working path to the protection path. This is only relevant if there is no currently active fault condition or Operator command.

- o WTR Expires - generated by the WTR timer completing its period. If none of the input sources have generated any input then the request logic should generate a No Request (NR) request.

In addition to the local requests, the PSC Control Logic SHALL accept PSC messages from the tailend node of the transport path. Remote messages indicate the status of the transport path from the viewpoint of the tailend nodes. The remote requests may include remote LO, SF, SD, FS, MS, WTR and NR.

Much of the signal definition is further described in ITU G.709 and G.873.1.

4.2. SMP Action Types

As shown in the previous section, SMP requires four actions types throughout the operation:

- ACTIVATION: This action is triggered by the head-end (or tail-end) to activate a protecting connection. The intermediate nodes need to propagate this towards the other end of the protecting connection.

- DE-ACTIVATION: This action is used to de-activate a particular protecting connection. This can be originated by one end of a protecting connection (i.e. head-end, or tail-end). The intermediate nodes need to propagate this towards the other end of the protecting connection.

- QUERY: This action is used when an operator decides to query a particular protecting connection.

- NOTIFICATION: SMP operation requires the coordination between nodes. The coordination takes places in two occasions:

  1. The activation/de-activation is initiated at the headend (tailend) nodes. To avoid potential mis-connection, the user traffic cannot be switched on to the protecting connection until the reception of an acknowledgement from the tailend (headend) nodes.

  2. If an intermediate node cannot process the activation requests, due to lack of resources or preemption levels, it needs to report the failure to the request originators.

It is conceivable that this message can be used to report the location of the fault, with respect to a protecting connection so that the head-end may use this information as one of the criteria for restoring the working transport entity. The fault location could be used by the head-end node to select among a list of possible protecting connections associated with the working connection (i.e. avoid the faulty location), or to determine that none of the provisioned protecting connections is usable at the time the failure is reported and then fallback to restoring the working connection.

4.3. PSC Signal to SMP Action Mapping

In SMP operation, there is the action-signal mapping:

- Activation action: FS, SF, SD, MS

o De-activation action: NR

o Query action: EXER

o Notification action: ACK, NACK (see next section)

5. Protocol Definition

Each SMP message has the following format:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Ver|Request|Rsv|R|  Reserved    |     Status    |     Seq       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o Version: 1

o Request:

   * 1111b: Lockout of Protection (LO)

   * 1110b: Forced Switch (FS). This triggers activation

   * 1100b: Signal Fail (SF). This triggers activation

   * 1011b: Acknowledgement (ACK). This is to acknowledge a successful activation/de-activation request

   * 1010b: Signal Degrade (SD). This triggers activation

   * 1001b: Negative Acknowledgement (NACK). This is to report failure in activation/de-activation process.

   * 1000b: Manual Switch (MS). This triggers activation

   * 0110b: Wait-to-Restore (WTR). Used for revertive switching

   * 0100b: Exercise (EXER). Triggers SMP query

   * 0001b: Do Not Revert (DNR). Used for revertive switching

   * 0000b: No Request (NR). This triggers de-activation

- o R: Revertive field
  - * 0: non-revertive mode
  - * 1: revertive mode
- o Rsv/Reserved: This field is reserved for future use
- o Status: this informs the status of the AMP activation. This field is only relevant with ACK and NACK requests. Specifically, the Status Code has the following encoding value and definition:
  - * 1: end-to-end ack
  - * 2: hop-to-hop ack
  - * 3: no such path
  - * 4: no more resource for the path
  - * 5: preempted by another path
  - * 6: system failure
  - * 7: shared resource has been taken by other paths
- o Seq: This uniquely identifies a particular message. This field is defined to support reliable message delivery

Note that the message format and naming convention are very similar to that of MPLS linear protection [6] and ITU G.873.1.

5.1. Message Encapsulation

SMP messages use MPLS labels to identify the paths. Further, the messages are encapsulated in GAL/GACH:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      MPLS Label stack                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          GAL                                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 0 0 1|Version|   Reserved    |   Activation Channel Type     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Activation Message Payload                 |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

GAL is described in [RFC5586]. Activation Channel Type is the GACH channel number assigned to the protocol. This uniquely identifies the activation messages.

Specifically, the messages have the following message format:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Label                  | Exp |S|     TTL       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Label (13)             | Exp |S|       TTL     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 0 0 1|Version|   Reserved    |   Activation Channel Type     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Ver|Request|Rsv|R|  Reserved   |     Status    |     Seq       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

5.2. Reliable Messaging

The activation procedure adapts a simple two-way handshake reliable messaging. Each node maintains a sequence number generator. Each new sending message will have a new sequence number. After sending a message, the node will wait for a response with the same sequence number.

Specifically, upon the generation of activation, de-activation, query and notification messages, the message sender expects to receive acknowledgement in reply with same sequence number.

If a sender is not getting the reply within a time interval, it will retransmit the same message with a new sequence number, and starts to wait again. After multiple retries (by default, 3), the sender will declare activation failure, and alarm the operators for further service.

5.3. Message Scoping

Activation signaling uses MPLS label TTL to control how far the message would traverse. Here are the processing rules on each intermediate node:

o On receive, if the message has label TTL = 0, the node must drop the packet without further processing

- o The receiving node must always decrement the label TTL value by one. If TTL = 0 after the decrement, the node must process the message. Otherwise, the node must forward the message without further processing (unless, of course, the node is headend or tailend)

- o On transmission, the node will adjust the TTL value. For hop-by-hop messages, TTL = 1. Otherwise, TTL = 0xFF, by default.

6. Processing Rules

6.1. Enable a Protecting Path

Upon the detection of network failure (SF/SD/FS) on a working path, the headend node identifies the corresponding MPLS-TP label and initiates the protection switching by sending an activation message.

The activation messages always use MPLS label TTL = 1 to force hop-by-hop process. Upon reception, a next-hop node will locate the corresponding path and activate the path.

If the activation message is received on an intermediate node, due to label TTL expiry, the message is processed and then propagated to the next hop of the MPLS TP LSP, by setting the MPLS TP label TTL = 1.

The headend node will declare the success of the activation only when it gets a positive reply from the tailend node. This requires that the tailend nodes must reply the messages with ACK to the headend nodes in all cases.

If the headend node is not receiving the acknowledgement within a time internal, it will retransmit another activation message with a different Seq number.

If the headend node is not receiving a positive reply within a longer time interval, it will declare activation failure.

If an intermediate node cannot activate a protecting path, it will reply a message with NACK to report failure. When the headend node receives the message for failure, it must initiate the de-activation messages to clean up networks resources on all the relevant nodes on the path.

6.2. Disable a Protecting Path

The headend removes the network resources on a path by sending the de-activation messages.

In the message, the MPLS label represents the path to be de-activated. The MPLS TTL is one to force hop-by-hop processing.

Upon reception, a node will de-activate the path, by freeing the resources from the data-plane.

As a part of the clean-up procedure, each de-activation message must traverse through and be processed on all the nodes of the corresponding path. When the de-activation message reaches to the tailend node, the tailend is required to reply with an acknowledgement message to the headend.

The de-activation process is complete when the headend receives the corresponding acknowledgement message from the tailend.

6.3. Get Protecting Path Status

The operators have the option to trigger the query messages from the headend to check on the protecting path periodically or on-demand. The process procedure on each node is very similar to that of the activation messages on the intermediate nodes, except the query messages should not trigger any network resource re-programming. Upon reception, the node will check the availability of resources.

If the resource is no longer available, the node will reply an NACK with error conditions.

6.4. Preemption

The preemption operation typically takes place when processing an activation message. If the activating network resources have been used by another path and carrying user traffic, the node needs to compare the priority levels.

If the existing path has higher priority, the node needs to reject the activation request by sending an ACK to the corresponding headend to inform the unavailability of network resources.

If the new path has higher priority, the node will reallocate the resource to the new path, and send an ACK to old path's headend node to inform about the preemption.

7. Security Consideration

The protection activation takes place in a controlled networking environment. Nevertheless, it is expected that the edge nodes will encapsulate and transport external traffic into separated tunnels,

and the intermediate nodes will never have to process them.

8. IANA Considerations

Activation messages are encapsulated in MPLS-TP with a specific GACH channel type that needs to be assigned by IANA.

9. Acknowledgments

Authors like to thank Eric Osborne, Lou Berger, Nabil Bitar and Deborah Brungard for detailed feedback on the earlier work, and the guidance and recommendation for this proposal.

We also thank Maneesh Jain, Mohit Misra, Yalin Wang, Ted Sprague, Ann Gui and Tony Jorgenson for discussion on network operation, feasibility and implementation methodology.

During ITU-T SG15 Interim meeting in May 2011, we have had long discussion with the G.SMP contributors, in particular Fatai Zhang, Bin Lu, Maarten Vissers and Jeong-dong Ryoo. We thank their feedback and corrections.

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4447] Martini, L., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, April 2006.

[RFC6370] Bocci, M., Swallow, G., and E. Gray, "MPLS Transport Profile (MPLS-TP) Identifiers", RFC 6370, September 2011.

10.2. Informative References

[RFC5586] Bocci, M., Vigoureux, M., and S. Bryant, "MPLS Generic Associated Channel", RFC 5586, June 2009.

[RFC5654] Niven-Jenkins, B., Brungard, D., Betts, M., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, September 2009.

Authors' Addresses

Ping Pan
(Infinera)

Email: ppan@infinera.com

Rajan Rao
(Infinera)

Email: rrao@infinera.com

Biao Lu
(Infinera)

Email: blu@infinera.com

Luyuan Fang
(Cisco)

Email: lufang@cisco.com

Andrew G. Malis
(Verizon)

Email: andrew.g.malis@verizon.com

Fatai Zhang
(Huawei)

Email: zhangfatai@huawei.com

Sam Aldrin
(Huawei)

Email: sam.aldrin@huawei.com

Fei Zhang
(ZTE)

Email: zhang.fei3@zte.com.cn

Sri Mohana Satya Srinivas Singamsetty
(Cisco)

Email: srsingam@cisco.com