IETF
Internet Draft

Ping Pan
Mohana Srinivas
Rajan Rao
Biao Lu

Expires: September 7, 2011

March 7, 2011

Supporting Shared Mesh Protection in MPLS-TP Networks

draft-pan-shared-mesh-protection-00.txt

Status of this Memo

Copyright Notice

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 [1]).

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.

Table of Contents

1. Introduction

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 [1]).Its operation has been further outlined in Section 4.7.6 of MPLS-TP Survivability Framework [2].

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.

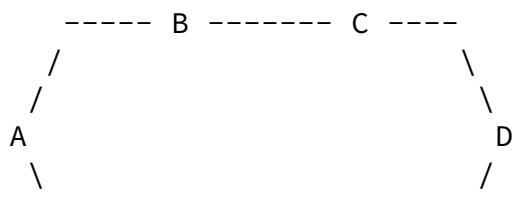
3. Modular: Depending on deployment scenarios, the signaling may need to support functions such as preemption, resource re-allocation and bi-directional activation in a modular fashion.

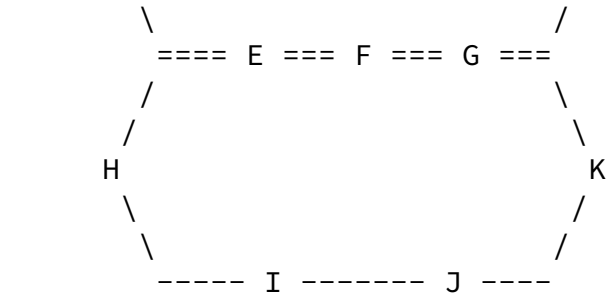
Here are some of the conventions used in this document. The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

2. Background and Problem Definition

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

We refer to the figure below for illustration:





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.

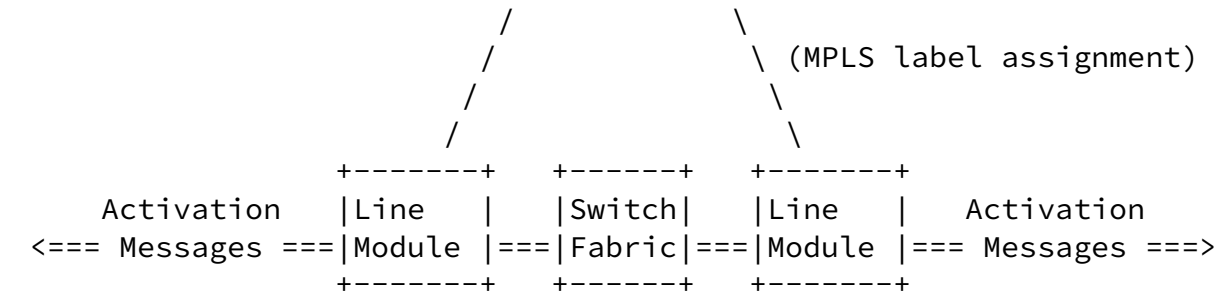
2. Over-subscription (1:N): A unit of network resource may be reserved by one or multiple protecting paths. In the example, the network resources on E-F and F-G are shared by two protecting paths, X' and Y'. In deployment, the over-subscription ratio is an important factor on network resource utilization.

3. Bi-direction Activation: The bi-directional paths can activate protection from both headend and tailend independently. In the example, upon the failure on link B-C, both A and D can trigger the activation of the protecting path X'. This procedure may improve the switch-over performance; however, it requires additional coordination between network nodes.

3. Protection Switching

The entire activation and switch-over operation need to be within the range of milliseconds to meet customer's expectation. In this section, we illustrate 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.



A typical MPLS-TP user flow (or, LSP) is bi-directional, 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.

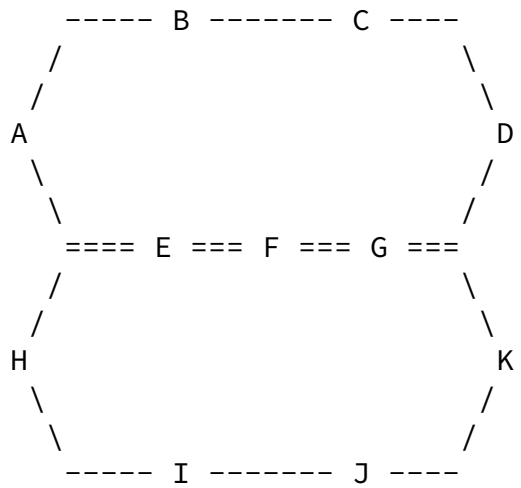
In summary, the activation procedure involves efficient path lookup and switch fabric re-programming.

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) [3]. Detailed message encoding is explained in Section 5.

4. Activation Operation Overview

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



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.

To optimize traffic protection and resource management, each headend should 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.

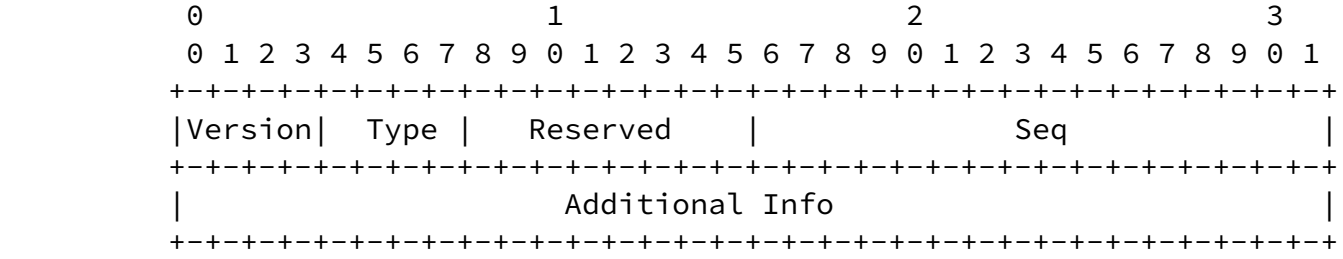
5. Protocol Definition

5.1. Activation Messages

The activation requires the following messages:

- ENABLE: this is initiated by the headend nodes to activate a protecting path
- DISABLE: this is initiated by the headend nodes to disable a protecting path and free the associated network resources
- GET: this is initiated by the headend to gather resource availability information on a particular protecting path
- NOTIFY: this is initiated by the intermediate nodes and terminate on the headend nodes to report preemption or protection failure conditions
- STATUS: this is the acknowledgement message for ENABLE, DISABLE, GET, and NOTIFY messages, and contains the relevant status information

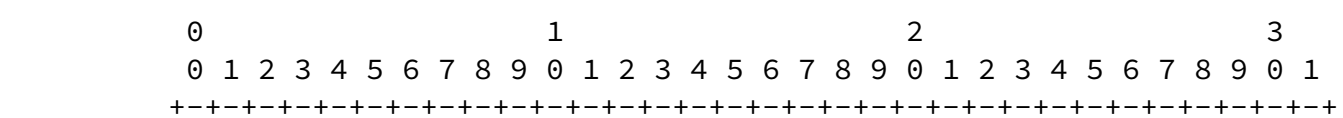
Each activation message has the following format:

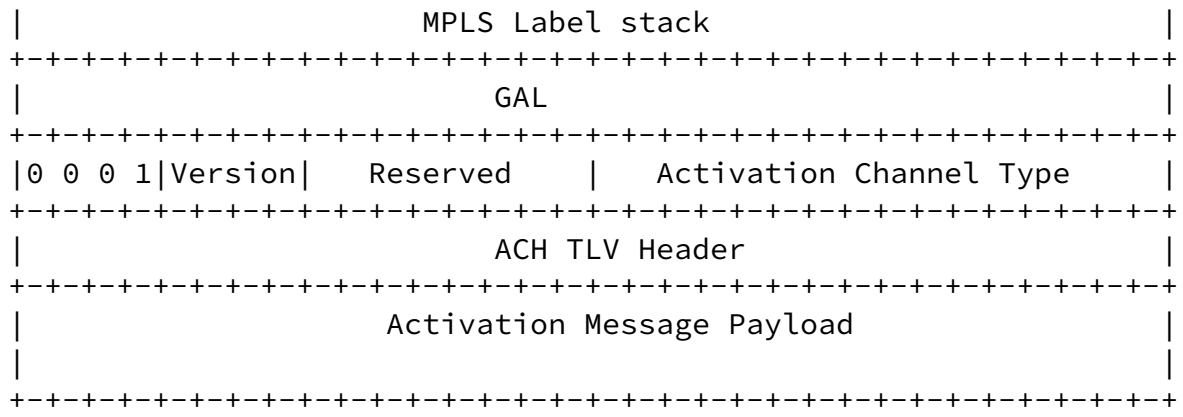


- Version: 1
- Type:
  - ENABLE 1
  - DISABLE 2
  - GET 3
  - STATUS 4
  - NOTIFY 5
- Reserved: This field is reserved for future use
- Seq: This uniquely identifies a particular message. This field is defined to support reliable message delivery
- Additional Info: the message-specific data

5.2. Message Encapsulation

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



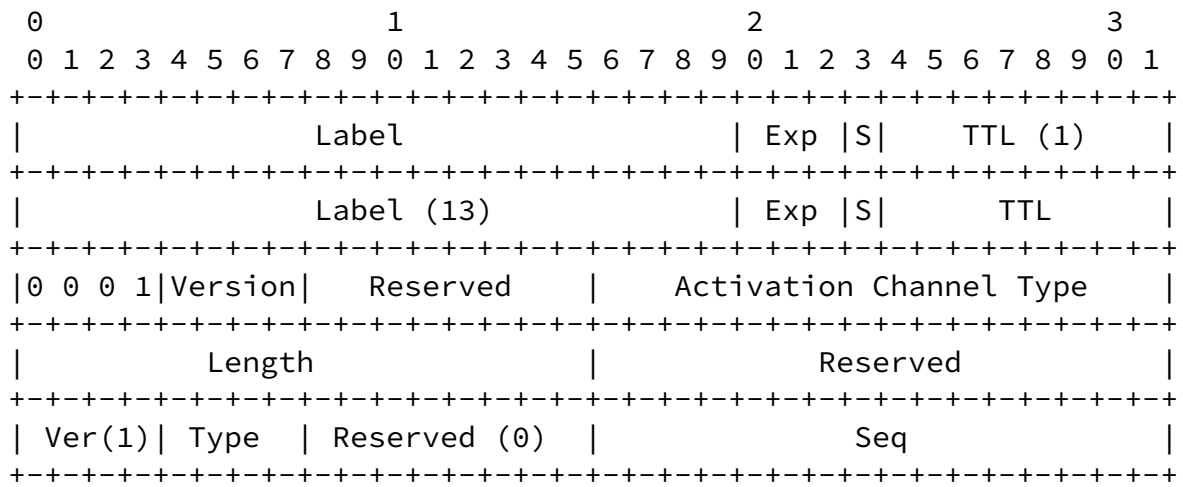


- GAL is described in [3]

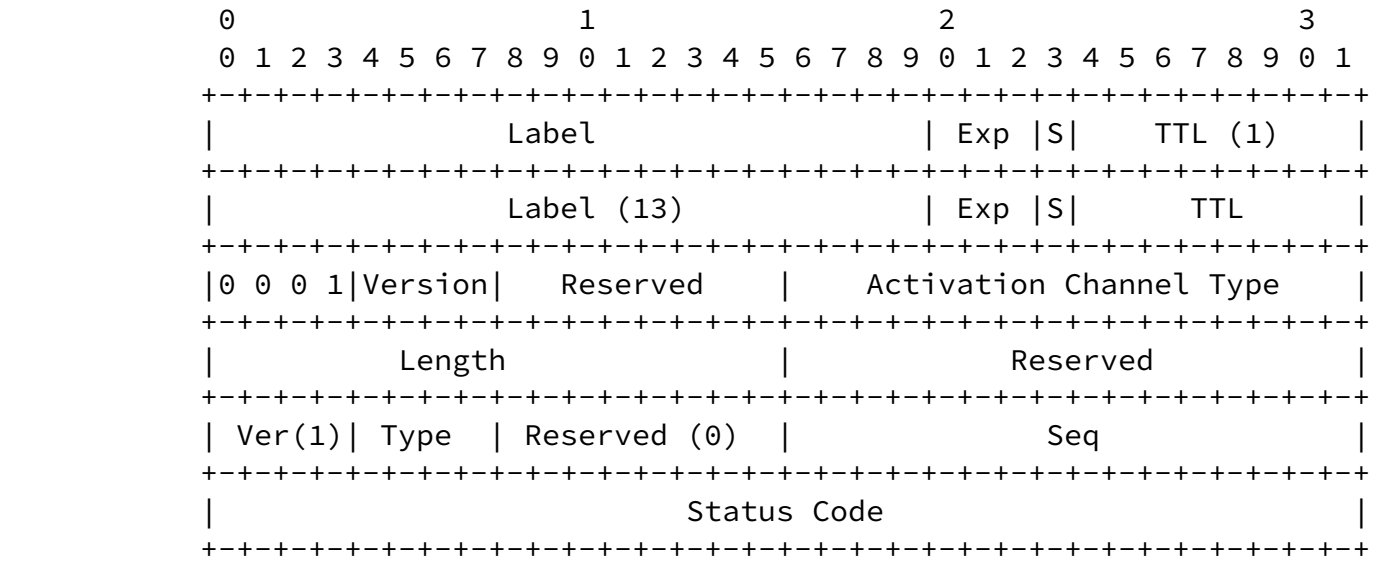
- Activation Channel Type is the GACH channel number assigned to the protocol. This uniquely identifies the activation messages.

- ACH TLV Header contains the message length, and is described in [3]

Specifically, ENABLE, DISABLE and GET messages have the following message format:



Both STATUS and NOTIFY messages have the following message format:



Currently, the status code has the following definition:

- o 1xx: OK
  - . 101: end-to-end ack
- o 2xx: message processing errors
  - . 201: no such path
- o 3xx: processing issues:
  - . 301: no more resource for the path
  - . 302: preempted by another path
  - . 303: system failure
- o 4xx: informative data:
  - . 401: shared resource has been taken by other paths

5.3. 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 ENABLE, DISABLE, GET and NOTIFY messages, the message sender expects to receive a STATUS in reply with same sequence number.

If a sender is not getting the reply (STATUS) 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.

5.4. Message Scoping

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

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

- The receiving node must always decrement the 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)

- 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 on a working path, the headend node initiates the protection switching by sending an ENABLE message.

ENABLE messages always use MPLS TTL one to force hop-by-hop process. Upon reception, a next-hop node will locate the corresponding path and activate the path.

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 STATUS messages to the headend nodes in all cases.

If the headend node is not receiving the acknowledgement within a time internal, it will retransmit another ENABLE 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 an NOTIFY message to report failure. When the headend node receives a NOTIFY message, it must initiate DISABLE 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 DISABLE 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 DISABLE message must traverse through and be processed on all the nodes of the corresponding path. When the DISABLE message reaches to the tailend node, the tailend is required to reply with a STATUS message to the headend.

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

6.3. Get protecting path status

The operators have the option to trigger GET 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 ENABLE messages on the intermediate nodes, except the GET messages should

not trigger any path re-programming.

Upon reception, the node will check the availability of resources.

If the resource is no longer available, the node will reply a NOTIFY with error conditions.

6.4. Acknowledgement with STATUS

The STATUS message is the acknowledgement packet to all messages, and may be generated by any node in the network.

Each STATUS message must use the same sequence number as the corresponding message (ENABLE, DISABLE, GET and NOTIFY).

When replying to headend, the tailend nodes must originate STATUS messages will a large MPLS TTL value (0xff, by default).

6.5. Preemption

The preemption operation typically takes place when processing an ENABLE 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 ENABLE message by sending a STATUS message 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 NOTIFY message 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. Normative References

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

[2] IETF draft, Multiprotocol Label Switching Transport Profile Survivability Framework (draft-ietf-mpls-tp-survive-fwk-06.txt), N. Sprecher, A. Farrel, June 2010

[3] RFC5586 - Vigoureux,, M., Bocci, M., Swallow, G., Aggarwal, R., and D. Ward, "MPLS Generic Associated Channel", May 2009.

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

[5] Crocker, D. and Overell, P.(Editors), "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, Internet Mail Consortium and Demon Internet Ltd., November 1997.

10. Acknowledgments

Authors like to thank Maneesh Jain, Mohit Misra, Yalin Wang, Ted Sprague, Ann Gui and Tony Jorgenson for review and feedback.

Authors' Addresses

Ping Pan
Email: ppan@infinera.com

Sri Mohana Satya Srinivas Singamsetty
Email: ssingamsetty@infinera.com

Rajan Rao

Email: rrao@infinera.com

Biao Lu
Email: blu@infinera.com