

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

Zafar Ali  
Rakesh Gandhi  
Tarek Saad  
Cisco Systems, Inc.  
Robert H. Venator  
Defense Information Systems Agency  
Yuji Kamite  
NTT Communications Corporation  
October 15, 2012

Signaling RSVP-TE P2MP LSPs in an Inter-domain Environment  
draft-ali-mpls-inter-domain-p2mp-rsvp-te-lsp-09

## Abstract

Point-to-MultiPoint (P2MP) Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs) are established using signaling procedures defined in [RFC4875]. However, [RFC4875] does not address several issues that arise when a P2MP-TE LSP is signaled in inter-domain networks. One such issue is the computation of a loosely routed inter-domain P2MP-TE LSP paths that are re-merge free. Another issue is the reoptimization of the inter-domain P2MP-TE LSP tree vs. an individual destination(s), since the loosely routing domain ingress border node is not aware of the reoptimization scope. This document defines the required protocol extensions needed for establishing and reoptimizing P2MP MPLS and GMPLS TE LSPs in inter-domain networks.

## Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 18, 2013.

## Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
1.1. Summary of Solutions . . . . .	4
1.2. Path Computation Techniques . . . . .	5
1.3. Use cases . . . . .	5
2. Conventions used in this document . . . . .	6
3. Control Plane Solution For Re-merge Handling . . . . .	6
3.1. Single Border Node For All S2Ls . . . . .	6
3.2. Crankback and PathErr Signaling Procedure . . . . .	6
4. Data Plane Solution For Re-merge Handling . . . . .	8
4.1. P2MP-TE Re-merge Recording Request Flag . . . . .	8
4.2. P2MP-TE Re-merge Present Flag . . . . .	8
4.3. Signaling Procedure . . . . .	9
5. Intra-domain P2MP-TE LSP Re-merge Handling . . . . .	10
6. Reoptimization Handling . . . . .	11
6.1. P2MP-TE Tree Re-evaluation Request Flag . . . . .	11
6.2. Preferable P2MP-TE Tree Exists Flag . . . . .	11
6.3. Signaling Procedure . . . . .	11
7. Compatibility . . . . .	13
8. Security Considerations . . . . .	13
9. IANA Considerations . . . . .	13
10. Acknowledgments . . . . .	14
11. References . . . . .	15
11.1. Normative References . . . . .	15
11.2. Informative References . . . . .	15
Author's Addresses . . . . .	16

## 1. Introduction

[RFC4875] describes procedures to set up Point-to-Multipoint (P2MP) Traffic Engineering Label Switched Paths (TE LSPs) for use in MultiProtocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks.

As with Point-to-Point (P2P) TE LSPs, P2MP TE LSP state is managed using RSVP messages. While the use of RSVP messages is mostly similar to their P2P counterpart, P2MP LSP state differs from P2P LSP in a number of ways. In particular, the P2MP LSP must also handle the "re-merge" problem described in [RFC4875] section 18.

The term "re-merge" refers to the situation when two source-to-leaf (S2L) sub-LSPs branch at some point in the P2MP tree, and then intersect again at a another node further downstream the tree. This may occur due to discrepancies in the routing algorithms used by different nodes, errors in path calculation or manual configuration, or network topology changes during the establishment of the P2MP LSP. Such re-merges are inefficient due to the unnecessary duplication of data and also consume additional network resources. Consequently one of the requirements for signaling P2MP LSPs is to choose a P2MP path that is re-merge free. In some deployments, it may also be required to signal P2MP-TE LSPs that are both re-merge and crossover free [RFC4875].

For the purposes of this document, a domain is considered to be any collection of network elements within a common sphere of address management or path computational responsibility. Examples of such domains include Interior Gateway Protocol (IGP) areas and Autonomous Systems (ASes). A border node is a node between different routing domains.

The re-merge free requirement becomes more acute to address when P2MP LSP spans multiple domains. This is because in an inter-domain environment, the ingress node may not have topological visibility into other domains to be able to compute and signal a re-merge free P2MP LSP. In that case, the border node for a new domain will be given loose next hops for one or more destinations in a P2MP LSP. A border node computes paths in its domain by individually expanding the loose next hops for the destinations when signaled to it. A border node can ensure that it computes the re-merge free paths while performing loose hop ERO expansions by individually grafting destinations. Note that the computed P2MP tree by a border node in this case may not be optimal. When processing a Path message, the border node may not have knowledge of all the destinations of the P2MP LSP; for example, in the case when not all S2L sub-LSPs pass through this border node. In that case, existing protocol mechanisms

do not provide sufficient information for it to be able to expand the loose hop(s) such that the overall P2MP LSP tree is guaranteed to be re-merge free.

[RFC4875] specifies two approaches to handle re-merge conditions. The first method is based on control plane handling the re-merge. In this case the node detecting the re-merge condition, i.e. the re-merge node initiates the removal of the re-merge sub-LSP(s) by sending a PathErr message(s) towards the ingress node. However, this can lead to a deadlock in setting up the P2MP LSP in certain cases; for example, when the first S2L setup causes the re-merge with all subsequent S2Ls in the tree. The second method is based on the data plane handling the re-merge condition. In this case, the re-merge node allows the re-merge condition to persist, but data from all but one incoming interface is dropped at the re-merge node. This ensures that duplicate data is not sent on any outgoing interface. However, network resources (such as bandwidth capacity) are wasted as long as re-merge condition persists which is inefficient.

[RFC4736] defines procedures and signaling extensions for reoptimizing an inter-domain P2P LSP. Specifically, an ingress node sends a "path re-evaluation request" to a border node by setting a flag (0x20) in SESSION\_ATTRIBUTES object in a Path message. A border node sends a PathErr code 25 (notify error defined in [RFC3209]) with sub-code 6 to indicate "preferable path exists" to the ingress node. The ingress node upon receiving this PathErr may initiate reoptimization of the LSP. [RFC4736] however does not define a procedure to reoptimize the entire P2MP LSP as a whole tree.

As per [RFC4875] Section 14, for a P2MP LSP, an ingress node may reoptimize the entire P2MP LSP by resignaling all destinations (Section 14.1, "Make-before-Break") or may reoptimize individual the destinations (Section 14.2 "Sub-Group-Based Re-Optimization"). Generally speaking make-before-break is considered available for "whole" P2MP LSP reoptimization, but it can also be used for reoptimizing physical routes for specific sub-LSP(s). The Sub-Group-Based reoptimization is not always applicable because it can lead to data duplication inside the backbone.

### 1.1. Summary of Solutions

This document defines RSVP-TE signaling procedures for P2MP LSP to handle the re-merge condition when using either the control plane or data plane approach. The procedures are applicable to both MPLS TE and GMPLS networks.

The control plane solution for the re-merge problem makes use of the crankback signaling mechanism of the RSVP protocol. [RFC5151]

describes such mechanisms for applying crankback to inter-domain P2P LSPs, but does not cover P2MP LSPs. Also, crankback mechanisms for P2MP LSPs are not addressed by [RFC4875]. This document describes how crankback signaling extensions for MPLS and GMPLS RSVP-TE defined in [RFC4920] can be used for setting up P2MP TE LSPs to resolve re-merges.

The data plane solution for the re-merge problem described in [RFC4875] is extended by using a new flag in the LSP\_ATTRIBUTES TLV (in a Path message) and a new flag in RRO Attributes Sub-object (in a Resv message) in RSVP. The LSP\_ATTRIBUTES TLV (in a Path message) and RRO Attributes Sub-object (in a Resv message) have been defined in [RFC5420]. This document describes how these new flags can be used to handle P2MP re-merge conditions efficiently.

For P2MP LSP, a border node may have loosely routed entire or part of the P2MP LSP by expanding EROs in Path messages of the destinations. Border node does not know with the signaling procedure defined in [RFC4736] if an ingress node is requesting a reoptimization for an individual destination(s) or reoptimization of the entire P2MP tree. Signaling extension and procedure are defined in this document to handle reoptimization of an individual destination(s) and the reoptimization of the entire P2MP tree. Basically, a new query message is defined in LSP\_ATTRIBUTES TLV to request for a "P2MP-TE Tree Re-evaluation" and a new sub-code is defined for PathErr message to indicate "Preferable P2MP-TE Tree Exists".

## 1.2. Path Computation Techniques

This document focuses on the case where the ingress node does not have full visibility of the topology of all domains and is therefore not able to compute the complete P2MP tree. Rather, it includes loose hops to traverse the domains for which it does not have full visibility and ingress border nodes(s) of each transit domain is responsible for expanding those loose hops.

The solution presented in this document do not guarantee optimization of the overall P2MP tree across all domains. Path Computation Element (PCE) can be used, instead, to address global optimization of the overall P2MP tree.

## 1.3. Use cases

Service providers having a network with multiple routing domains are interested to use the network for P2MP-TE LSPs. This allows the service providers to use the network to carry multicast and broadcast traffic (such as video). Service providers can deploy the VPLS and MVPN services in the network using inter-domain P2MP TE LSPs. The use

case is for P2MP TE LSPs across multiple routing domains that belong to a single administrative area. Use case for the Multiple administrative domains (e.g. autonomous systems) is outside the scope of this document.

## 2. 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 [RFC2119].

## 3. Control Plane Solution For Re-merge Handling

It is RECOMMENDED that boundary re-routing is requested for P2MP LSPs traversing multiple domains. This is because border nodes that are expanding loose hops are typically best placed to correct any re-merge errors that occur within their domain, not the ingress node.

### 3.1. Single Border Node For All S2Ls

It is RECOMMENDED that the ingress node of a P2MP LSP selects the same ingress border node in the loose hop ERO for all sibling S2L sub-LSPs that transit through a given domain. The reason is that it will increase the possibility of re-merge downstream if two or more border nodes have roles simultaneously to expand loose EROs. An ingress border node that performs the loose ERO expansion for individual sub-LSP(s) has the necessary state information for the destinations transiting through its domain to ensure computed P2MP tree is re-merge free.

### 3.2. Crankback and PathErr Signaling Procedure

As mentioned in [RFC4875], in order to avoid duplicate traffic, the re-merge node MAY initiate the removal of the re-merge S2L sub-LSPs by sending a PathErr message to the ingress node of the S2L sub-LSP.

Crankback procedures for rerouting around failures for P2P RSVP-TE LSPs are defined in [RFC4920]. These techniques can also be applied to P2MP LSPs to handle re-merge conditions, as described in this section.

If an ingress border node on the path of the P2MP LSP is unable to find a route that can supply the required resources or that is re-merge free, it MUST generate a PathErr message for the subset of the S2L sub-LSPs which it is not able to route. For this purpose the ingress border node SHOULD try to find a minimum subset of S2L sub-LSPs for which the PathErr needs to be generated towards the ingress

node. These are the S2L sub-LSPs on an incoming interface that has less number of S2L sub-LSPs compared to the second incoming interface that is causing the re-merge condition.

The RSVP-TE Notify messages do not include S2L\_SUB\_LSP objects and cannot be used to send errors for a subset of the S2L sub-LSPs in a Path message. For that reason, the error generating node SHOULD use a PathErr message rather than a Notify message to communicate the error. In the case of a re-merge error, the node SHOULD use the error code "Routing Problem" and the error value "ERO resulted in re-merge" as specified in [RFC4875].

A border node receiving a PathErr message for a set of S2L sub-LSPs MAY hold the message and attempt to signal an alternate path that can avoid re-merge through its domain for those S2L sub-LSPs that pass through it. However, in the case of a re-merge error for which some of the re-merging S2L sub-LSPs do not pass through the border node, it SHOULD propagate the PathErr upstream towards the ingress node. If the subsequent attempt by the border node is successful, the border node discards the held PathErr and follows the crankback roles of [RFC4920] and [RFC5151]. If repeated subsequent attempts by the border node are unsuccessful, the border node MUST send the held PathErr upstream towards the ingress node.

If the ingress node receives a PathErr message with error code "Routing Problem" and error value "ERO resulted in re-merge", then it SHOULD attempt to signal an alternate path through a different domain or through a different border node for the affected S2L sub-LSPs. The ingress node MAY use the error node information from the PathErr for this purpose.

However, it may be that the ingress node or an ingress border node does not have sufficient topology information to compute an Explicit Route that is guaranteed to avoid the re-merge link or node. In this case, Route Exclusions [RFC4874] may be particularly helpful. To achieve this, [RFC4874] allows the re-merge information to be presented as route exclusions to force avoidance of the re-merge link or node.

As discussed in [RFC4920] section 3.3, border node MAY keep the history of PathErrs. In case of P2MP LSPs, ingress node and border nodes may keep re-merge PathErrs in history table until S2L sub-LSPs have been successfully established or until local timer expires.

#### 4. Data Plane Solution For Re-merge Handling

As mentioned in [RFC4875], a node may accept the re-merging S2Ls but only send the data from one of these interfaces to its outgoing interfaces. That is, the node MUST drop data from all but one incoming interface causing the re-merge. This ensures that duplicate data is not sent on any outgoing interface. Note that data plane may be either programmed to drop the incoming traffic for the S2L sub-LSP or not programmed at all.

It is desirable to avoid the persistent re-merge condition associated with data plane based solution in the network in order to optimize bandwidth resources in the network.

The following sections define the RSVP-TE signaling extensions for "P2MP- TE Re-merge Recording Request" and "P2MP-TE Re-merge Present" messages.

##### 4.1. P2MP-TE Re-merge Recording Request Flag

In order to indicate to traversed nodes that P2MP-TE re-merge recording is desired, a new flag in the Attribute Flags TLV of the LSP\_ATTRIBUTES object defined in [RFC5420] is defined as follows:

Bit Number (to be assigned by IANA): P2MP-TE Re-merge Recording Request flag

The "P2MP-TE Re-merge Recording Request" flag is meaningful in a Path message and is inserted by the ingress node or a border node in the LSP\_ATTRIBUTES object.

If the "P2MP-TE Re-merge Recording Request" Flag is set, it implies that the "P2MP-TE Re-merge Present" flag defined in the next section MUST be used to indicate to the ingress and ingress border nodes of the transit domains that a re-merge condition is present for this S2L sub-LSP but accepted, and that incoming traffic is being dropped for this S2L sub-LSP.

The rules of the processing of the Attribute Flags TLV of the LSP\_ATTRIBUTES object follow [RFC5420].

##### 4.2. P2MP-TE Re-merge Present Flag

The "P2MP-TE Re-merge Present" Flag is the counter part of the "P2MP-TE Re-merge Recording Request" flag defined above. Specifically, RSVP signaling extension is defined to indicate to the



upstream node of the re-merge condition and that incoming traffic is being dropped for the given S2L.

When a node accepts a re-merge condition by dropping traffic from an incoming interface for an S2L due to the re-merge condition, and if it understands the "P2MP-TE Re-merge Recording Request" in the Attribute Flags TLV of the LSP\_ATTRIBUTES object of the Path message, the node MUST set the newly defined "P2MP-TE Re-merge Present" flag in the RRO Attributes sub-object defined in [RFC5420] in RRO.

The following new flag for RRO Attributes Sub-object is defined as follows:

Bit Number (same as bit number assigned for "P2MP-TE Re-merge Recording Request" flag): P2MP-TE Re-merge Present flag

The "P2MP-TE Re-merge Present" flag indicates that the S2L is causing a re-merge. The re-merge has been accepted but the incoming traffic on this S2L is dropped by the reporting node.

The rules of the processing of the RRO Attribute Sub-object in the Resv message follow [RFC5420].

#### 4.3. Signaling Procedure

When a node that does not support data plane based re-merge handling receives an S2L sub-LSP Path message with LSP Attributes sub-object that has "P2MP-TE Re-merge Recording Request" Flag set, and if the S2L is causing a re-merge condition, the node MUST reject the S2L sub-LSP Path message and send the PathErr with the error code "Routing Problem" and the error value "ERO resulted in re-merge" as specified in [RFC4875]. If a node is capable of data plane based re-merge handling but operator may have disabled it via a configuration, the the node MUST also reject the re-merge and send this PathErr.

When a Path message is received at a transit node for an S2L sub-LSP and "P2MP-TE Re-merge Recording Request" Flag is set in the LSP Attributes sub-object, the node MAY decide to accept the re-merge S2L sub-LSP based on the local policy and node capability. In this case, before the Resv message is sent to the upstream node for this S2L sub-LSP, the node MUST add the RRO Attributes sub-object in the Resv RRO if not already present and set the "P2MP-TE Re-merge Present" Flag if traffic from the incoming interface of this S2L sub-LSP will be dropped. This same incoming interface can still be used for a different S2L sub-LSP in the P2MP LSP to forward traffic and "P2MP-TE Re-merge Present" flag will not be set for that S2L sub-LSP. Note

that rules for adding or modifying the other RRO sub-objects do not change due to this flag.

When a transit node receives a Resv message for an S2L that is causing a re-merge condition, the node MUST set the "P2MP-TE Re-merge Present" flag in the RRO Attributes sub-object in the Resv message if it decides to drop the incoming traffic of this S2L. The "P2MP-TE Re-merge Present" flag in RRO Attribute sub-object is not set for the S2L(s) whose incoming interface is selected to receive and forward the traffic.

An ingress node MAY immediately start sending traffic on all S2Ls in up state even when re-merge conditions are present on some S2Ls of the P2MP LSP.

The proposed signaling extensions allow an ingress node and an ingress border node to have a complete view of the re-merge conditions on the entire S2L path and on all S2Ls of the P2MP tree. The ingress or ingress border node in this case can take appropriate actions to resolve the re-merge conditions and optimize network bandwidth resources usage. This can be achieved by computing and selecting alternate path(s) for the S2L(s) bypassing the re-merge node(s).

The proposed signaling extensions are equally applicable to single domain scenarios.

A node where re-merge is present, may decide to select a different incoming interface to forward traffic from in the future. In that case, a Resv change message with updated "P2MP-TE Re-merge Present" flag in the RRO is sent upstream for all effected S2Ls. For the new set of S2L sub-LSPs whose traffic from the incoming interface is dropped, "P2MP-TE Re-merge Present" flag will be set.

A border node due to local policy MAY remove the record route object from the Resv message of the S2L sub-LSP and propagate Resv message towards the ingress node. When such a policy is provisioned, the border node may attempt to correct the re-merge condition in its domain. If the border node is not able to resolve the re-merge condition, the border node SHOULD send the PathErr with the error code "Routing Problem" and the error value "ERO resulted in re-merge" as specified in [RFC4875].

## 5. Intra-domain P2MP-TE LSP Re-merge Handling

Re-merges between S2Ls in a single domain can occur due to provisioning errors or path computation errors in the environment

where IGP-TE or PCE is used. Re-merges can also happen in the environment where static routing or static path selection policy is applied at ingress (e.g., CSPF calculation is disabled due to some operational reasons), regardless of using loose or static hops. In either case, procedures described in this document are equally applicable to the intra-domain (i.e. single domain) P2MP-TE LSPs.

## 6. Reoptimization Handling

### 6.1. P2MP-TE Tree Re-evaluation Request Flag

In order to query border nodes to check if a preferable P2MP tree exists, a new flag is defined in Attributes Flags TLV of the LSP\_ATTRIBUTES object [RFC5420] as follows:

Bit Number (to be assigned by IANA): P2MP-TE Tree Re-evaluation Request flag

The "P2MP-TE Tree Re-evaluation Request" flag is meaningful in a Path message of an S2L sub-LSP and is inserted by the ingress node.

### 6.2. Preferable P2MP-TE Tree Exists Flag

In order to indicate to an ingress node that a preferable P2MP-TE tree is available, following new sub-code for PathErr code 25 (notify error) is defined:

Sub-code (to be assigned by IANA): Preferable P2MP-TE Tree Exists flag

When a preferable P2MP-TE tree is found, the border node MUST send "Preferable P2MP-TE Tree Exists" to the ingress node in order to reoptimize the entire P2MP LSP.

### 6.3. Signaling Procedure

Using signaling procedure defined in [RFC4736], an ingress node MUST initiate "path re-evaluation request" query to reoptimize a destination in a P2MP LSP. Note that this message MUST be used to reoptimize a single or a sub-set of the destinations in a P2MP LSP. Ingress node MUST send this query in a Path message for each destination it is reoptimizing.

When a Path message for a destination in a P2MP LSP with "path re-evaluation request" flag [RFC4736] is received at the border node,

it MUST re-compute the loose-hop ERO to see if a preferable path exists for that destination. A border node MUST send PathErr code 25 (notify error defined in [RFC3209]) with "preferable path exists" sub-code to indicate that a preferable path exists for the requested destination AND border node is capable of per destination reoptimization. A border node MUST terminate the path query. Alternatively, a border node not capable of per destination reoptimization MAY respond with "Preferable P2MP-TE Tree Exists" PathErr by checking for a preferable P2MP tree instead of a preferable single destination.

It is often desired to reoptimize the entire P2MP LSP. In order to query border nodes to check if a preferable P2MP tree exists, an ingress node MUST send a Path message with "P2MP-TE Tree Re-evaluation Request" defined in this document. An ingress node MAY send this message for all destinations in a P2MP LSP or a sub-set of the destinations.

A border node receiving the "P2MP-TE Tree Re-evaluation Request" MUST check for a preferable P2MP LSP for the destinations it is loosely routing by loose-hop ERO expansions. The border node if a preferable P2MP-TE tree is found, MUST reply with "Preferable P2MP-TE Tree Exists" sub-code defined in this document with PathErr 25 (notify error defined in [RFC3209]) and terminate the path query.

Note that a border node MAY send "Preferable P2MP-TE Tree Exists" with PathErr code 25 to indicate the ingress node in order to reoptimize the entire P2MP LSP message unsolicited or in a response to "path re-evaluation query" for a destination or in a response to "P2MP-TE Tree Re-evaluation Request" message.

If an ingress node initiated a "path re-evaluation request" query for a single destination for per S2L sub-LSP reoptimization and receives "Preferable P2MP-TE Tree Exists" PathErr, the ingress node MAY cancel the per S2L reoptimization and initiate P2MP-TE tree reoptimization. This may happen in case when a border node is not capable of per destination reoptimization.

Note that even if per destination reoptimization, not whole P2MP LSP Tree reoptimization, is sufficient, ingress node often needs to re-signal whole P2MP LSP tree to complete route optimization for that destination. In this case, make-before-break reoptimization scheme is used (see [RFC4875] Section 14.1), and all S2L sub-LSPs are re-signaled with a different LSP-ID. That is, the procedure of signaling a re-optimization by an ingress node is separate from the matter if PathErr reply was "Preferable Path Exists" or "Preferable P2MP-TE Tree Exists".

## 7. Compatibility

The LSP\_ATTRIBUTES TLV and RRO Attributes sub-object have been defined [RFC5420] with class numbers in the form 1lbbbbbb, which ensures compatibility with non-supporting nodes. Per [RFC2205], nodes not supporting this extension will ignore the TLV, sub-object and the new flags defined in this document but forward it, unexamined and unmodified, in all messages resulting from this message.

## 8. Security Considerations

This document does not introduce any additional security issues above those identified in [RFC3209], [RFC4875], [RFC5151], [RFC4920] and [RFC5920].

## 9. IANA Considerations

The following new flag is defined for the Attributes Flags TLV in the LSP\_ATTRIBUTES object [RFC5420]. The numeric values are to be assigned by IANA.

- o P2MP-TE Re-merge Recording Request Flag:

- Bit Number: To be assigned by IANA.
- Attribute flag carried in Path message: Yes
- Attribute flag carried in Resv message: No

The following new flag is defined for the RRO Attributes sub-object in the RECORD\_ROUTE object [RFC5420]. The numeric values are to be assigned by IANA.

- o P2MP-TE Re-merge Present Flag:

- Bit Number: To be assigned by IANA.
- Attribute flag carried in Path message: No
- Attribute flag carried in RRO Attributes sub-object in RRO of the Resv message: Yes

The following new flag is defined for the Attributes Flags TLV in the LSP\_ATTRIBUTES object [RFC5420]. The numeric values are to be assigned by IANA.

- o P2MP-TE Tree Re-evaluation Request Flag:
  - Bit Number: To be assigned by IANA.
  - Attribute flag carried in Path message: Yes
  - Attribute flag carried in Resv message: No

As defined in [RFC3209], the Error Code 25 in the ERROR\_SPEC object corresponds to a Notify Error PathErr. This document adds a new sub-code as follows for this PathErr:

- o Preferable P2MP-TE Tree Exists sub-code:
  - Sub-code for Notify PathErr code 25. To be assigned by IANA.

## 10. Acknowledgments

The authors would like to thank N. Neate for his contributions on the draft.

## 11. References

### 11.1. Normative References

- [RFC4875] Aggarwal, R., Papadimitriou, D., and S. Yasukawa, "Extensions to Resource Reservation Protocol Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.
- [RFC5151] Farrel, A., Ayyangar, A., and JP. Vasseur, "Inter-Domain MPLS and GMPLS Traffic Engineering -- Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 5151, February 2008.
- [RFC4920] Farrel, A., Satyanarayana, A., Iwata, A., Fujita, N., and G. Ash, "Crankback Signaling Extensions for MPLS and GMPLS RSVP-TE", RFC 4920, July 2007.
- [RFC5920] L. Fang, Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC4736] Vasseur, JP., Ikejiri, Y. and Zhang, R, "Reoptimization of Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Loosely Routed Label Switched Path (LSP)", RFC 4736, November 2006.
- [RFC5420] Farrel, A., Papadimitriou, D., Vasseur, JP., and A. Ayyangar, "Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)", RFC 5420, February 2009.

### 11.2. Informative References

- [RFC4726] Farrel, A., Vasseur, J., and A. Ayyangar, "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, November 2006.

Author's Addresses

Zafar Ali  
Cisco Systems

Email: zali@cisco.com

Rakesh Gandhi  
Cisco Systems

Email: rgandhi@cisco.com

Tarek Saad  
Cisco Systems

Email: tsaad@cisco.com

Robert H. Venator  
Defense Information Systems Agency

Email: robert.h.venator.civ@mail.mil

Yuji Kamite  
NTT Communications Corporation

Email: y.kamite@ntt.com



MPLS Working Group  
Internet Draft  
Intended status: Standard Track

I. Busi (Ed)  
Alcatel-Lucent  
H. van Helvoort (Ed)  
J. He (Ed)  
Huawei

Expires: July 2012

January 11, 2012

MPLS-TP OAM based on Y.1731  
draft-bhh-mpls-tp-oam-y1731-08.txt

## Abstract

This document describes methods to leverage Y.1731 [2] Protocol Data Units (PDU) and procedures (state machines) to provide a set of Operation, Administration, and Maintenance (OAM) mechanisms that meets the MPLS Transport Profile (MPLS-TP) OAM requirements as defined in [8].

In particular, this document describes the MPLS-TP technology specific encapsulation mechanisms to carry these OAM PDUs within MPLS-TP packets to provide MPLS-TP OAM capabilities in MPLS-TP networks.

## Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress".

The list of current Internet-Drafts can be accessed at  
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at  
<http://www.ietf.org/shadow.html>

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

## Table of Contents

1. Introduction.....	4
1.1. Contributing Authors.....	5
2. Conventions used in this document.....	5
2.1. Terminology.....	6
3. Encapsulation of OAM PDU in MPLS-TP.....	6
4. MPLS-TP OAM Packet Formats.....	7
4.1. Continuity Check Message (CCM).....	8
4.1.1. MEG ID Formats.....	9
4.2. OAM Loopback (LBM/LBR).....	9
4.2.1. Format of MEP and MIP ID TLVs.....	12
4.3. Alarm Indication Signal (AIS).....	16
4.4. Lock Reporting (LCK).....	16
4.5. Test (TST).....	17
4.6. Loss Measurement (LMM/LMR).....	17
4.7. One-way delay measurement (1DM).....	17
4.8. Two-way delay Measurement Message/Reply (DM).....	17
4.9. Client Signal Fail (CSF).....	18
5. MPLS-TP OAM Procedures.....	18
5.1. Continuity Check Message (MT-CCM) procedures.....	18
5.2. OAM Loopback (MT-LBM/LBR) procedures.....	20
5.3. Alarm Indication Signal (MT-AIS) procedures.....	21
5.4. Lock Reporting (LCK).....	22
5.5. Test (TST).....	23
5.6. Loss Measurement (LMM/LMR).....	23
5.7. One-way delay measurement (1DM).....	23
5.8. Two-way delay Measurement Message/Reply (DM).....	23
5.9. Client Signal Fail (CSF).....	23
6. Security Considerations.....	23
7. IANA Considerations.....	23
8. Acknowledgments.....	23
9. References.....	25
9.1. Normative References.....	25
9.2. Informative References.....	25

## 1. Introduction

This document describes the method for leveraging Y.1731 [2] Protocol Data Units (PDUs) and procedures to provide a set of Operation, Administration, and Maintenance (OAM) mechanisms that meet the MPLS Transport Profile (MPLS-TP) OAM requirements as defined in [8].

This version of the draft does not introduce any technical change to the -06 version of this draft.

ITU-T Recommendation Y.1731 [2] specifies:

- o OAM PDUs and procedures that meet the transport networks requirements for OAM
- o Encapsulation mechanisms to carry these OAM PDUs within Ethernet frames to provide Ethernet OAM capabilities in Ethernet networks

Although Y.1731 is focused on Ethernet OAM, the definition of OAM PDUs and procedures are technology independent and can also be used in other packet technologies (e.g., MPLS-TP) provided that the technology specific encapsulation is defined.

The OAM toolset defined in Y.1731 [2] serves as a benchmark for a high performance, comprehensive suite of packet transport OAM capabilities. It can be provided by lightweight protocol design and supports operational simplicity by providing commonality with the established operation models utilized in other transport network technologies (e.g., SDH/SONET and OTN).

This document describes mechanisms for MPLS-TP OAM that reuse the same OAM PDUs and procedures defined in Y.1731 [2], together with the necessary MPLS-TP technology specific encapsulation mechanisms.

The advantages offered by this toolset are summarized below:

- o Simplify the operations for the network operators and service providers that have to test and maintain a single general OAM protocol set when operating LSP, PW and VPLS networks.
- o Accelerate the market adoption of MPLS-TP since Y.1731 is already mature, supported, and deployed.
- o Reduce the complexity and increase the reuse of code for implementation in packet transport devices that may support both

Ethernet and MPLS-TP capabilities, e.g. VPLS and H-VPLS applications.

It is worth noting that multi-vendor interoperable implementations of the OAM mechanisms described in this document already exist to meet the essential OAM requirements for MPLS-TP deployments in PTN applications as described in [9].

Ethernet OAM is also defined by IEEE 802.1ag [14]. IEEE 802.1ag and ITU-T Y.1731 have been developed in cooperation by IEEE and ITU. They support a common subset of OAM functions. ITU-T Y.1731 further extends this common subset with additional OAM mechanisms that are important for the transport network (e.g. AIS, DM, LM).

This document does not deprecate existing MPLS and PW OAM mechanisms nor preclude definition of other MPLS-TP OAM tools.

The mechanisms described in this document, when used to provide MPLS-TP PW OAM functions, are open to support the OAM message mapping procedures defined in [10]. In order to support those procedures, the PEs MUST map the states of the procedures defined in Y.1731 to the PW defect states defined in [10].

The mapping procedures are outside the scope of this document.

In the rest of this document the term "OAM PDU" is used to indicate an OAM PDU whose format and associated procedures are defined in Y.1731 [2] and that this document proposes to be used to provide MPLS-TP OAM functions.

### 1.1. Contributing Authors

Italo Busi, Huub van Helvoort, Jia He, Christian Addeo, Alessandro D'Alessandro, Simon Delord, John Hoffmans, Ruiquan Jing, Kam Lam, Wang Lei, Han Li, Vishwas Manral, Masahiko Mizutani, Manuel Paul, Josef Roese, Vincenzo Sestito, Yuji Tochio, Munefumi Tsurusawa, Maarten Visser, Rolf Winter

### 2. 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 [1].

## 2.1. Terminology

ACH	Associated Channel Header
G-ACh	Generic Associated Channel
GAL	G-ACh Label
ME	Maintenance Entity
MEL	MEG Level
MEG	Maintenance Entity Group
MEP	Maintenance End Point
MIP	Maintenance Intermediate Point
PTN	Packet Transport Network
TLV	Type Length Value

## 3. Encapsulation of OAM PDU in MPLS-TP

Although Y.1731 is focused on Ethernet OAM, the definition of OAM PDUs and procedures are technology independent.

When used to provide Ethernet OAM capabilities, these PDUs are encapsulated into an Ethernet frame where an Ethernet header is prepended to the OAM PDUs.

The MAC DA is used to identify the MEPs and MIPs where the OAM PDU needs to be processed. The EtherType is used to distinguish OAM frames from user data frames.

Within MPLS-TP OAM Framework [6], OAM packets are distinguished from user data packets using the GAL and ACH [5] construct and they are addressed to MEPs or MIPs using existing MPLS forwarding mechanisms (i.e. label stacking and TTL expiration). It is therefore possible to reuse the OAM PDUs defined in [2] within MPLS-TP and encapsulate them within ACH.

A single ACH Channel Type (0xFFFF) is required to identify the presence of Y.1731 OAM PDU. Within the OAM PDU, the OpCode field, defined in [2], allows identifying the specific OAM PDU.

OAM PDUs are encapsulated using the ACH, according to [5], as described in Figure 1 below.

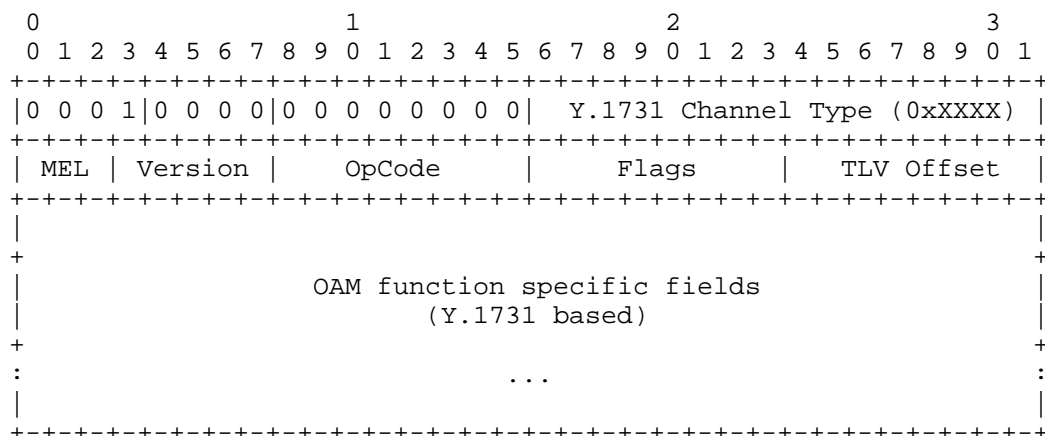


Figure 1 G-ACh Packet carrying a Y.1731 PDU

Moreover, MPLS-TP relies upon a different mechanism for supporting tandem connection monitoring (i.e. label stacking) than the fixed MEL (Maintenance Entity Group Level) field used in Ethernet.

Therefore in MPLS-TP the MEL field is allowed not to be used for supporting tandem connection monitoring.

When OAM PDUs are used in MPLS-TP, the MEL field MUST be set on transmission and checked at reception for compliancy with Y.1731 [2].

The MEL value to set and check MUST be configurable. The DEFAULT value MUST be "111". With co-routed bidirectional transport paths, the configured MEL MUST be the same in both directions.

The OpCode field identifies the type of the OAM PDU.

The setting of the Version, Flags and TLV Offset is OpCode specific and described in Y.1731 [2].

#### 4. MPLS-TP OAM Packet Formats

This section describes the OAM functions that can be supported reusing the OAM PDUs and procedures defined in Y.1731 [2] to meet MPLS-TP OAM Requirements, as defined in [8].

This document is proposing not to use the Y.1731 MCC OAM PDU in MPLS-TP. The solution proposed in [7], where MCC PDU is directly encapsulated within an ACH with a PID, SHOULD be used instead.

The LTM/LTR OAM PDUs, as currently defined Y.1731 [2], are tracing the path for a specific MAC address: this tool is therefore addressing a different requirement than the "Route Tracing" functional requirement described in section 2.2.4 of RFC 5860 [8]. Their purpose is to test the MAC Address Forwarding tables. Due to the fact that MPLS-TP forwarding is not based on the MAC Address Forwarding tables, these tools are not applicable to MPLS-TP as currently defined.

Procedures for supporting the route tracing MPLS-TP OAM functional requirement (section 2.2.4 of RFC 5860 [8]) are outside the scope of this document.

#### 4.1. Continuity Check Message (CCM)

The CCM PDU is defined in Y.1731 [2]. When encapsulated within MPLS-TP as described in section 3, it can be used to support the following MPLS-TP OAM functional requirements:

- o Pro-active continuity check (section 2.2.2 of RFC 5860 [8]);
- o Pro-active connectivity verification (section 2.2.3 of RFC 5860 [8]);
- o Pro-active remote defect indication (section 2.2.9 of RFC 5860 [8]);
- o Pro-active packet loss measurement (section 2.2.11 of RFC 5860 [8]).

Procedures for transmitting and receiving CCM PDUs are defined in Y.1731 [2] and described in section 5.1.

It is worth noting that the use of CCM does not require any additional status information other than the configuration parameters and defect states.

The transmission period of the CCM MUST always be the configured period and MUST not change unless the operator reconfigures it. This is a fundamental requirement to allow deterministic and predictable



protocol behavior: in transport networks the operator configures and fully controls the repetition rate of pro-active CC-V.

In order to perform pro-active Connectivity Verification, the CCM packet contains a globally unique identifier of the source MEP, as described in [6].

The source MEP for LSPs, PWs and Sections is identified by combining a globally unique MEG ID (see section 4.1.1) with a MEP ID that is unique within the scope of the Maintenance Entity Group.

#### 4.1.1. MEG ID Formats

The generic format for MEG ID is defined in Figure A-1 of Y.1731 [2]. Different formats of MEG ID are allowed: the MEG ID format type is identified by the MEG ID Format field.

The format of the ICC-based MEG ID is defined in Annex A of Y.1731 [2]. This format is applicable to MPLS-TP Sections, LSPs and PWs.

MPLS-TP supports also IP-based format for MEG ID. These formats are still under definition in [12] and therefore outside the scope of this document.

#### 4.2. OAM Loopback (LBM/LBR)

The LBM/LBR PDUs, defined in Y.1731 [2]. When encapsulated within MPLS-TP, as described in section 3, they can be used to support the following MPLS-TP OAM functional requirements:

- o On-demand bidirectional connectivity verification (section 2.2.3 of RFC 5860 [8]);
- o Bidirectional in-service or out-of-service diagnostic test (section 2.2.5 of RFC 5860 [8]).

Procedures for transmitting and receiving LBM/LBR PDUs are defined in Y.1731 [2] and described in section 5.2.

It is worth noticing that these OAM PDUs cover different functions than those defined in [11].

When the LBM/LBR is used for out-of-service diagnostic test, it is REQUIRED that the transport path is locked on both MEPs before the diagnostic test is performed. In transport networks, the transport

path is locked on both sides by network management operations. However, single-ended procedures as defined in [11] MAY be used.

In order to allow proper identification of the target MEP/MIP the LBM is addressed to, the LBM PDU MUST include the Target MEP/MIP ID TLV: this TLV MUST be present in an LBM PDU and MUST be located at the top of the TLVs (i.e., it MUST start at the offset indicated by the TLV Offset field).

A LBM packet with the Target MIP/MEP ID equal to the ID of receiving MIP or MEP is considered to be a valid LBM packet. Every field in the LBM packet is copied to the LBR packet, only the OpCode field is changed from LBM to LBR.

To allow proper identification of the actual MEP/MIP that has replied to an LBM PDU, the LBR PDU MUST include the Replying MEP/MIP ID TLV: this TLV MUST be present in an LBR PDU and it MUST be located at the top of the TLVs (i.e., it MUST start at the offset indicated by the TLV Offset field).

In order to simplify hardware based implementations, these TLVs have been defined to have a fixed position (as indicated by the TLV Offset field) and a fixed length (see clause 4.2.1).

It is worth noting that the MEP/MIP identifiers used in the Target MEP/MIP ID and in the Replying MEP/MIP ID TLVs SHOULD be unique within the scope of the MEG. When LBM/LBR OAM is used for connectivity verification purposes, there are some misconnectivity cases that could not be easily located by simply relying upon these TLVs. In order to locate these misconnectivity configurations, the LBM PDU SHOULD carry a Requesting MEP ID TLV that provides a globally unique identification of the MEP that has originated the LBM PDU. When the Requesting MEP ID TLV is present in the LBM PDU, the replying MIP/MEP MUST check that the received requesting MEP identifier matches with the expected requesting MEP identifier before replying. In this case, the LBR PDU MUST carry the Requesting MEP ID TLV confirming to the MEP the LBR PDU is sent to that the Requesting MEP ID TLV in the LBM PDU has been checked before replying.

When LBM/LBR OAM is used for bidirectional diagnostic tests, the Requesting MEP ID TLVs MUST NOT be included.

The format of the LBM and LBR PDUs are shown in Figure 2 and in Figure 3.

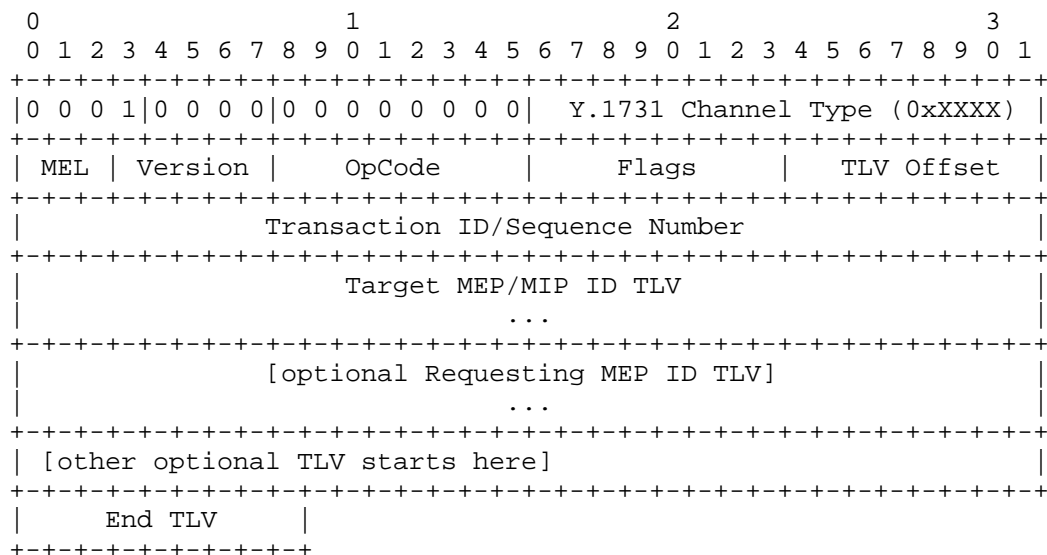


Figure 2 LBM Packet Format

The OpCode MUST be set to 0x03 (LBM). The TLV Offset MUST be set to 0x04. The formats of the Target MEP/MIP ID TLV and of the Requesting MEP ID TLV are defined in 4.2.1.

The Target MEP/MIP ID MUST be always present as the first TLV within the LBM PDU. When present, the Requesting MEP ID TLV MUST immediately follow the Target MEP/MIP ID TLV.

When the LBM packet is sent to a target MIP, the source MEP MUST know the hop count to the target MIP and set the TTL field accordingly, as described in [6].

This solution allows supporting per-node and per-interface MIP implementations as described in section 3.4 of [6]:

- o In the case of a per-node MIP implementation, the LBM packet is processed in the per-node MIP if the Target MEP/MIP ID matches the per-node MIP identifier; otherwise, the LBM packet is dropped;

- o In the case of a per-interface MIP implementation, the LBM packet is processed in the ingress MIP if the Target MEP/MIP ID matches the ingress MIP identifier; otherwise, the LBM packet is forwarded to the egress port(s) together (i.e., fate sharing) with the user data packets. The LBM packet is processed in the egress MIP if the Target MEP/MIP ID matches the egress MIP identifier; otherwise, the LBM packet is dropped.

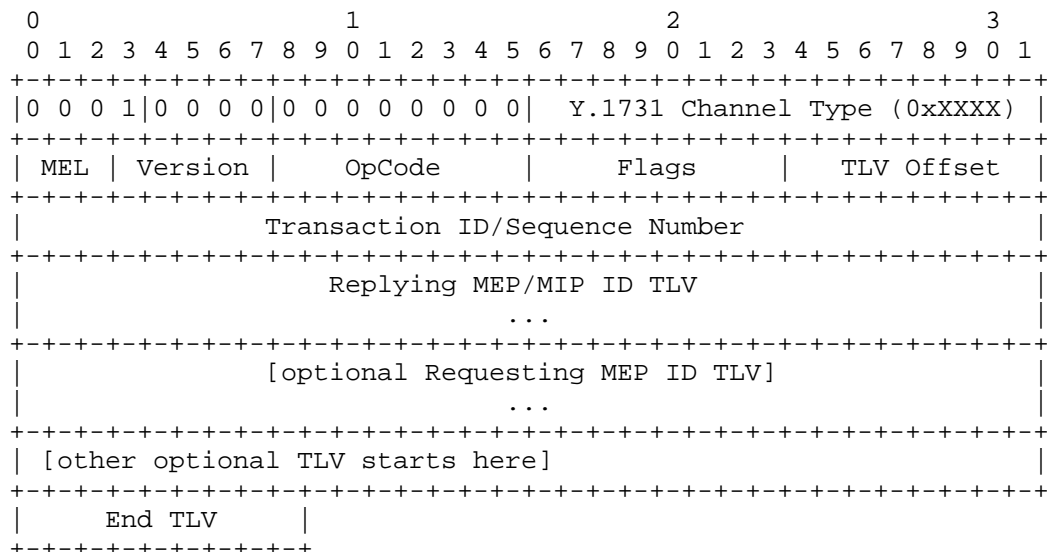


Figure 3 LBR Packet Format

The Replying MEP/MIP ID TLV MUST be present as the first TLV within the LBR PDU. When present, the Requesting MEP ID TLV MUST follow the Replying MEP/MIP ID TLV within the LBR PDU.

#### 4.2.1. Format of MEP and MIP ID TLVs

The format of the Target and Replying MIP/MEP ID TLVs are shown in Figure 4 and Figure 5.

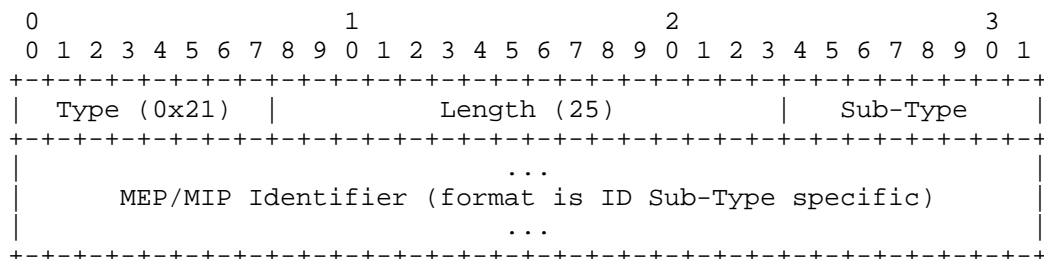


Figure 4 Target MEP/MIP ID TLV format

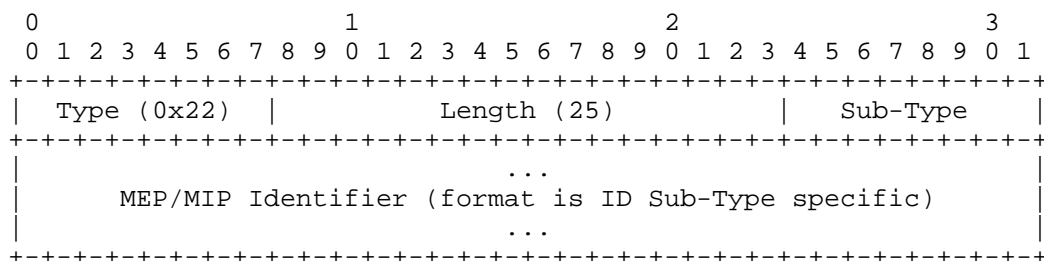


Figure 5 Replying MEP/MIP ID TLV format

Different formats of MEP/MIP identifiers MAY be used: the format type is described by the MEP/MIP ID Sub-Type field.

The "Discovery ingress/node MEP/MIP" and the "Discovery egress MEP/MIP" identifiers MAY only be used within the LBM PDU (and MUST NOT appear in an LBR PDU) for discovering the identifiers of the MEPs or of the MIPs located at a given TTL distance from the MEP originating the LBM PDU.

The format of the Target MEP/MIP ID TLV carrying a "Discovery ingress/node MEP/MIP" is shown in Figure 6.

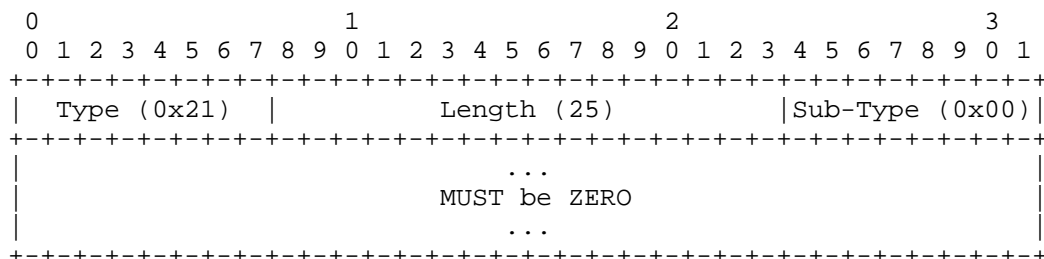


Figure 6 Target MEP/MIP ID TLV format (discovery ingress/node MEP/MIP)

The format of the Target MEP/MIP ID TLV carrying a "Discovery egress MEP/MIP" is shown in Figure 7.

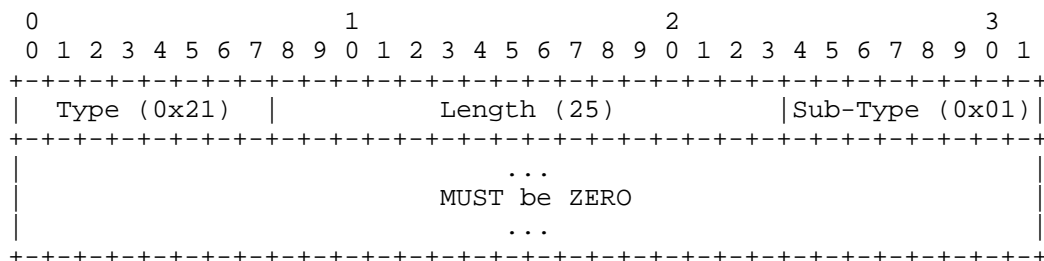


Figure 7 Target MEP/MIP ID TLV format (discovery egress MEP/MIP)

The format of the Target or Replying MEP/MIP ID TLV carrying an "ICC-based MEP ID" is shown in Figure 8.

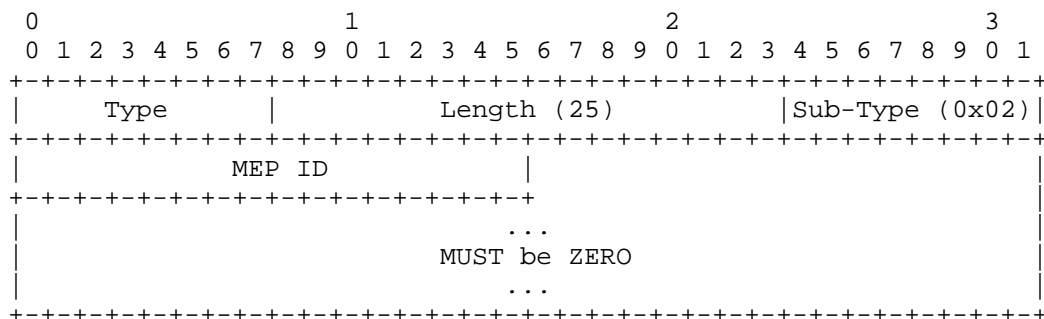


Figure 8 Target or Replying MEP/MIP ID TLV format (ICC-based MEP ID)

The MEP ID is a 16-bit integer value identifying the transmitting MEP within the MEG.

The format of the Target or Replying MEP/MIP ID TLV carrying an "ICC-based MIP ID" is shown in Figure 9.

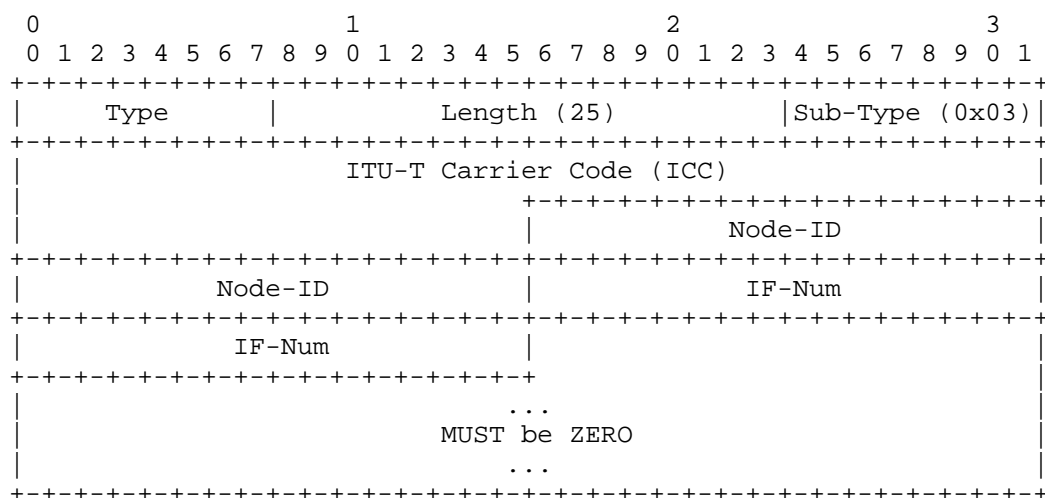


Figure 9 Target or Replying MEP/MIP ID TLV format (ICC-based MIP ID)

The ITU-T Carrier Code (ICC) is a code assigned to a network operator/service provider and maintained by the ITU-T Telecommunication Standardization Bureau (TSB) as per [13].

The Node-ID is a numeric identifier of the node where the MIP is located. Its assignment is a matter for the organization to which the ICC has been assigned, provided that uniqueness within that organization is guaranteed.

The IF-Num is a numeric identifier of the Access Point (AP) toward the server layer trail, which can be either an MPLS-TP or a non MPLS-TP server layer, where a per-interface MIP is located. Its assignment is a matter for the node the MIP is located, provided that uniqueness within that node is guaranteed. Note that the value 0 for IF-Num is reserved to identify per-node MIPs.

MPLS-TP supports also IP-based format for MIP and MEP identifiers. These formats are still under definition in [12] and therefore outside the scope of this document.

The format of the Requesting MEP ID TLVs is shown in Figure 10.

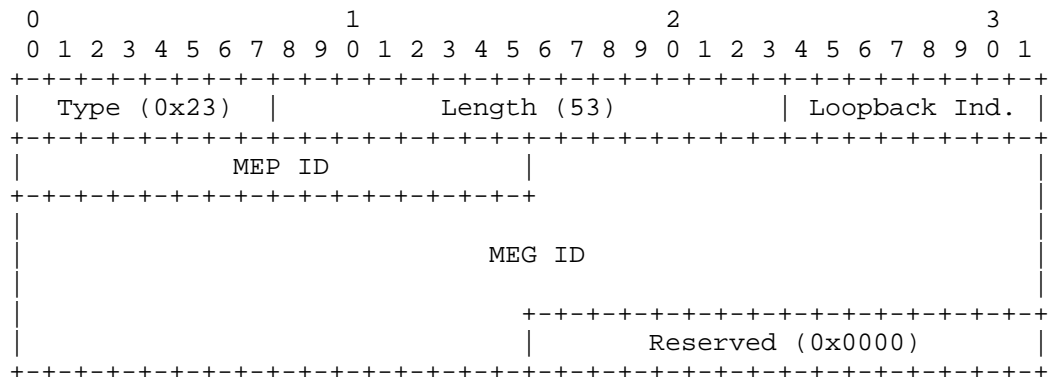


Figure 10 Requesting MEP ID TLV format

The MEP ID and MEG ID carry the globally unique MEP ID as defined in section 4.1.1.

The Reserved bits MUST be set to all-ZEROes in transmission and ignored in reception.

The Loopback Indication MUST be set to 0x0000 when this TLV is inserted in an LBM PDU and SHOULD be set to 0x0001 in the LBR PDU. This is used to indicate that the value of this TLV has been checked by the node that generated the LBR PDU.

#### 4.3. Alarm Indication Signal (AIS)

The AIS PDU is defined in Y.1731 [2]. When encapsulated within MPLS-TP, as described in section 3, it can be used to support the alarm reporting MPLS-TP OAM functional requirement (section 2.2.8 of RFC 5860 [8]).

Procedures for transmitting and receiving AIS PDUs are defined in Y.1731 [2] and described in section 5.3.

#### 4.4. Lock Reporting (LCK)

The LCK PDU is defined in Y.1731 [2]. When encapsulated within MPLS-TP, as described in section 3, it can be used to support the lock reporting MPLS-TP OAM functional requirement (section 2.2.7 of RFC 5860 [8]).



Procedures for transmitting and receiving LCK PDUs are defined in Y.1731 [2] and described in section 5.4.

#### 4.5. Test (TST)

The TST PDU is defined in Y.1731 [2]. When encapsulated within MPLS-TP, as described in section 3, it can be used to support the uni-directional in-service or out-of-service diagnostic tests MPLS-TP OAM functional requirement (section 2.2.8 of RFC 5860 [8]).

Procedures for transmitting and receiving TST PDUs are defined in Y.1731 [2] and described in section 5.5.

#### 4.6. Loss Measurement (LMM/LMR)

The LMM/LMR PDUs are defined in Y.1731 [2]. When encapsulated within MPLS-TP, as described in section 3, they can be used to support on-demand packet loss measurement MPLS-TP OAM functional requirement (section 2.2.11 of RFC 5860 [8]).

Procedures for transmitting and receiving LMM/LMR PDUs are defined in Y.1731 [2] and described in section 5.6.

#### 4.7. One-way delay measurement (1DM)

The 1DM PDU is defined in Y.1731 [2]. When encapsulated within MPLS-TP, as described in section 3, it can be used to support the on-demand one-way packet delay measurement MPLS-TP OAM functional requirement (section 2.2.12 of RFC 5860 [8]).

It can also be used to support proactive one-way delay measurement MPLS-TP OAM functional requirement (section 2.2.12 of RFC 5860 [8]).

Procedures for transmitting and receiving 1DM PDUs are defined in Y.1731 [2] and described in section 5.7.

#### 4.8. Two-way delay Measurement Message/Reply (DM)

The DMM/DMR PDUs are defined in Y.1731 [2]. When encapsulated within MPLS-TP, as described in section 3, they can be used to support on-demand two-ways packet delay measurement MPLS-TP OAM functional requirement (section 2.2.12 of RFC 5860 [8]).

They can also be used to support proactive two-ways packet delay measurement MPLS-TP OAM functional requirement (section 2.2.12 of RFC 5860 [8]).

Procedures for transmitting and receiving DMM/DMR PDUs are defined in Y.1731 [2] and described in section 5.8.

#### 4.9. Client Signal Fail (CSF)

The CSF PDU is defined in Y.1731 Amendment 1 [3]. When encapsulated within MPLS-TP, as described in section 3, it can be used to support the client failure indication MPLS-TP OAM functional requirement (section 2.2.10 of RFC 5860 [8]).

Procedures for transmitting and receiving CSF PDUs are defined in Y.1731 Amendment 1 [3] and described in section 5.9.

### 5. MPLS-TP OAM Procedures

The high level procedures for processing Y.1731 OAM PDUs are described in [2] and [3]. The technology independent procedures are also applicable to MPLS-TP OAM.

More detailed and formal procedures for processing Y.1731 OAM PDUs are defined in G.8021 [4]. Although the description in [4] is Ethernet-specific, the technology independent procedures are also applicable to MPLS-TP OAM.

This section describes the MPLS-TP OAM procedures based on the technology independent ones defined in [2], [3] and [4].

#### 5.1. Continuity Check Message (MT-CCM) procedures

The MT-CCM PDU format is defined in section 4.1.

When CCM generation is enabled, the MEP MUST generate CCM OAM packets with the periodicity and the PHB configured by the operator:

- o MEL field MUST be set to the configured value (see section 3);
- o Version field MUST be set to 0 (see section 3);
- o OpCode field MUST be set to 0x01 (see section 4.1);

- o RDI flag MUST be set, if the MEP asserts signal file. Otherwise, it MUST be cleared;
- o Reserved flags MUST be set to 0 (see section 4.1);
- o Period field MUST be set according to the configured periodicity (see Table 9-3 of [2]);
- o TLV Offset field MUST be set to 70 (see section 4.1);
- o Sequence Number MUST be set to 0 (see section 4.1);
- o MEP ID and MEG ID fields MUST carry the configured values;
- o The TxFCf field MUST carry the current value of the counter for in-profile data packets transmitted towards the peer MEP, when pro-active loss measurement is enabled. Otherwise it MUST be set to 0.
- o The RxFCb field MUST carry the current value of the counter for in-profile data packets received from the peer MEP, if pro-active loss measurement is enabled. Otherwise it MUST be set to 0.
- o The TxFCb field MUST carry the value of TxFCf of the last received CCM PDU from the peer MEP, if pro active loss measurement is enabled. Otherwise it MUST be set to 0.
- o Reserved field MUST be set to 0 (see section 4.1);
- o End TLV MUST be inserted after the Reserved field (see section 4.1).

The transmission period of the CCM is always the configured period and does not change unless the operator reconfigures it.

When a MEP receives a CCM OAM packet, it checks the various fields (see Figure 8-19 of [4]). The following defects are detected as described in clause 6.1 of [4]: dLOC, dUNL, dMMG, dUNM, dUNP, dUNPr and dRDI.

If the Version, MEL, MEG and MEP fields are valid and pro-active loss measurement is enabled, the values of the packet counters are processed as described in clause 8.1.7.4 of [4].

## 5.2. OAM Loopback (MT-LBM/LBR) procedures

The MT-LBM/LBR PDU formats are defined in section 4.2.

When an out-of-service OAM loopback function is performed, client data traffic is disrupted in the diagnosed ME. The MEP configured for the out-of-service test MUST transmit MT-LCK packets in the immediate client (sub-)layer, as described in section 5.4.

When an in-service OAM loopback function is performed, client data traffic is not disrupted and the packets with MT-LBM/LBR information are transmitted in such a manner that a limited part of the service bandwidth is utilized. The periodicity for packets with MT-LBM/LBR information is pre-determined.

When on-demand OAM loopback is enabled at a MEP, the (requesting) MEP MUST generate and send to one of the MIPs or the peer MEP MT-LBM OAM packets with the periodicity and the PHB configured by the operator:

- o MEL field MUST be set to the configured value (see section 3);
- o Version field MUST be set to 0 (see section 3);
- o OpCode field MUST be set to 0x03 (see section 4.2);
- o Flags field MUST be set to all-ZEROes (see section 4.2);
- o TLV Offset field MUST be set to 4 (see section 4.2);
- o Transaction field is a 4-octet field that contains the transaction ID/sequence number for the loop-back measurement;
- o Target MEP/MIP-ID and Originator MEP-ID fields are set to carry the configured values;
- o Optional TLV field whose length and contents are configurable at the requesting MEP. The contents can be a test pattern and an optional checksum. Examples of test patterns include pseudo-random bit sequence (PRBS) ( $2^{31}-1$ ) as specified in sub-clause 5.8/O.150, all '0' pattern, etc. For bidirectional diagnostic test application, configuration is required for a test signal generator and a test signal detector associated with the MEP;
- o End TLV field is set to all-ZEROes (see section 4.2).

Whenever a valid MT-LBM packet is received by a (receiving) MIP or a (receiving) MEP, an MT-LBR packet is generated and transmitted by the receiving MIP/MEP to the requesting MEP:

- o MEL field MUST be copied from the received MT-LBM PDU;
- o Version field MUST be copied from the received MT-LBM PDU;
- o OpCode field MUST be set to 2 (see section 4.2);
- o Flags field MUST be copied from the received MT-LBM PDU;
- o TLV Offset field MUST be copied from the received MT-LBM PDU;
- o Transaction field MUST be copied from the received MT-LBM PDU;
- o The Target MEP/MIP-ID and Originator MEP-ID fields are set to the value which is copied from the last received MT-LBM PDU;
- o The Optional TLV field MUST be copied from the received MT-LBM PDU;
- o End TLV field MUST be inserted after the last TLV field and it MUST be copied from the last received MT-LBM PDU.

### 5.3. Alarm Indication Signal (MT-AIS) procedures

The MT-AIS PDU format is described in section 4.3.

When the server layer trail termination sink asserts signal fail, it notifies the server/MT\_A\_Sk function that raises the aAIS consequent action. The aAIS is cleared when the server layer trail termination clears the signal fail condition and notifies the server/MT\_A\_Sk.

When the aAIS consequent action is raised, the server/MT\_A\_Sk MUST continuously generate MPLS-TP OAM packets carrying the AIS PDU until the aAIS consequent action is cleared:

- o MEL field MUST be set to the configured value (see section 3):
- o Version field MUST be set to 0 (see section 3):
- o OpCode MUST be set to 0x21 (see section 4.3):
- o Reserved flags MUST be set to 0 (see section 4.3):

- o Period field MUST be set according to the configure periodicity (see Table 9-4 of [2]);
- o TLV Offset MUST be set to 0 (see section 4.3):
- o End TLV MUST be inserted after the TLV Offset field (see section 4.3).

The DEFAULT periodicity for MT-AIS is once per second.

The generated AIS packets MUST be inserted in the incoming stream, i.e., the output stream contains the incoming packets and the generated AIS packets.

When a MEP receives an AIS packet with the correct MEL value, it MUST detect the dAIS defect as described in clause 6.1 of [4].

#### 5.4. Lock Reporting (LCK)

The MT-LCK PDU format is described in section 4.4.

When the access to the server layer trail is administratively locked by the operator, the server/MT\_A\_So and server/MT\_A\_Sk functions raise the aLCK consequent action. The aLCK is cleared when the access to the server layer trail is administratively unlocked.

When the aLCK consequent action is raised, the server/MT\_A\_So and server/MT\_A\_Sk MUST continuously generate, on both directions, MPLS-TP OAM packets carrying the LCK PDU until the aLCK consequent action is cleared:

- o MEL field MUST be set to the configured value (see section 3):
- o Version field MUST be set to 0 (see section 3):
- o OpCode MUST be set to 0x23 (see section 4.4):
- o Reserved flags MUST be set to 0 (see section 4.4):
- o Period field MUST be set according to the configure periodicity (see Table 9-4 of [2]);
- o TLV Offset MUST be set to 0 (see section 4.4):

- o End TLV MUST be inserted after the TLV Offset field (see section 4.4).

The DEFAULT periodicity for MT-LCK is once per second.

When a MEP receives an LCK packet with the correct MEL value, it detects the dLCK defect as described in clause 6.1 of [4].

#### 5.5. Test (TST)

#### 5.6. Loss Measurement (LMM/LMR)

#### 5.7. One-way delay measurement (1DM)

#### 5.8. Two-way delay Measurement Message/Reply (DM)

#### 5.9. Client Signal Fail (CSF)

### 6. Security Considerations

Spurious OAM messages, such as those defined in this document, potentially could form a vector for a denial of service attack. However, since these messages are carried in a control channel, one would have to gain access to a node providing the service in order to launch such an attack. Since transport networks are usually operated as a walled garden, such threats are less likely.

### 7. IANA Considerations

IANA is requested to allocate a Channel Type value 0xXXXX to identify an associated channel carrying all the OAM PDUs that are defined in section 4

[Editor's note - The value 0x8902 has been proposed to keep the channel type identical to the EtherType value used in Ethernet OAM]

### 8. Acknowledgments

The authors gratefully acknowledge the contributions of Malcolm Betts, Zhenlong Cui, Feng Huang, Kam Lam, Jian Yang, Haiyan Zhang for the definition of extensions to LBM/LBR required for supporting on-demand connectivity verification OAM functions.

The authors would like to thank all the members of the CCSA for their comments and support.

The authors would also like to thank Brian Branscomb, Feng Huang, Kam Lam, Fang Li, Akira Sakurai and Yaakov Stein for their comments and enhancements to the text.

This document was prepared using 2-Word-v2.0.template.dot.



## 9. References

### 9.1. Normative References

- [1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [2] ITU-T Recommendation Y.1731 (02/08), "OAM functions and mechanisms for Ethernet based networks", February 2008
- [3] ITU-T Recommendation Y.1731 Amendment 1 (07/10), "OAM functions and mechanisms for Ethernet based networks", July 2010
- [4] ITU-T Recommendation G.8021 (12/07), "Characteristics of Ethernet transport network equipment functional blocks", December 2007
- [5] Vigoureux, M., Bocci, M., Swallow, G., Ward, D., Aggarwal, R., "MPLS Generic Associated Channel", RFC 5586, June 2009
- [6] Busi, I., Allan, D., " Operations, Administration and Maintenance Framework for MPLS-based Transport Networks", draft-ietf-mpls-tp-oam-framework-11 (work in progress), February 2011
- [7] Beller, D., Farrel, A., "An Inband Data Communication Network For the MPLS Transport Profile", RFC 5718, January 2010

### 9.2. Informative References

- [8] Vigoureux, M., Betts, M., Ward, D., "Requirements for OAM in MPLS Transport Networks", RFC 5860, May 2010
- [9] Li, F., Li, H., D'Alessandro, A., Jing, R., Wang, G., "Operator Considerations on MPLS-TP OAM Mechanisms", draft-fang-mpls-tp-oam-considerations-02 (work in progress), July 2011
- [10] Nadeau, T., et al., "Pseudo Wire (PW) OAM Message Mapping", draft-ietf-pwe3-oam-msg-map-16 (work in progress), April 2011
- [11] Boutros, S., et al., "Operating MPLS Transport Profile LSP in Loopback Mode", draft-ietf-mpls-tp-li-lb-02 (work in progress), June 2011

- [12] Swallow, G., Bocci, M., " MPLS-TP Identifiers", draft-ietf-mpls-tp-identifiers-02 (work in progress), July 2010
- [13] ITU-T Recommendation M.1400 (07/06), " Designations for interconnections among operators' networks", July 2006
- [14] IEEE Standard 802.1ag-2007, "IEEE Standard for Local and Metropolitan Area Networks: Connectivity Fault Management", September 2007

#### Author's Addresses

Italo Busi (Editor)  
Alcatel-Lucent

Email: Italo.Busi@alcatel-lucent.com

Huub van Helvoort (Editor)  
Huawei Technologies

Email: hhelvoort@huawei.com

Jia He (Editor)  
Huawei Technologies

Email: hejia@huawei.com

#### Contributing Authors' Addresses

Christian Addeo  
Alcatel-Lucent

Email: Christian.Addeo@alcatel-lucent.com

Alessandro D'Alessandro  
Telecom Italia

Email: alessandro.dalessandro@telecomitalia.it

Simon Delord  
Telstra

Email: [simon.a.delord@team.telstra.com](mailto:simon.a.delord@team.telstra.com)

John Hoffmans  
KPN

Email: [john.hoffmans@kpn.com](mailto:john.hoffmans@kpn.com)

Ruiquan Jing  
China Telecom

Email: [jingrq@ctbri.com.cn](mailto:jingrq@ctbri.com.cn)

Hing-Kam (Kam) Lam  
Alcatel-Lucent

Email: [Kam.Lam@alcatel-lucent.com](mailto:Kam.Lam@alcatel-lucent.com)

Wang Lei  
China Mobile Communications Corporation

Email: [wangleiyj@chinamobile.com](mailto:wangleiyj@chinamobile.com)

Han Li  
China Mobile Communications Corporation

Email: [lihan@chinamobile.com](mailto:lihan@chinamobile.com)

Vishwas Manral  
IPInfusion Inc

Email: [vishwas@ipinfusion.com](mailto:vishwas@ipinfusion.com)

Masahiko Mizutani  
Hitachi, Ltd.

Email: [masahiko.mizutani.ew@hitachi.com](mailto:masahiko.mizutani.ew@hitachi.com)

Manuel Paul  
Deutsche Telekom

Email: [Manuel.Paul@telekom.de](mailto:Manuel.Paul@telekom.de)

Josef Roese  
Deutsche Telekom

Email: [Josef.Roese@t-systems.com](mailto:Josef.Roese@t-systems.com)

Vincenzo Sestito  
Alcatel-Lucent

Email: [vincenzo.sestito@alcatel-lucent.com](mailto:vincenzo.sestito@alcatel-lucent.com)

Yuji Tochio  
Fujitsu

Email: [tochio@jp.fujitsu.com](mailto:tochio@jp.fujitsu.com)

Munefumi Tsurusawa  
KDDI R&D Labs

Email: [tsuru@kddilabs.jp](mailto:tsuru@kddilabs.jp)

Maarten Visser  
Huawei Technologies

Email: [maarten.vissers@huawei.com](mailto:maarten.vissers@huawei.com)

Internet-Draft

MPLS-TP OAM based on Y.1731

January 2012

Rolf Winter  
NEC

Email: Rolf.Winter@nw.neclab.eu



Network Working Group  
Internet Draft  
Intended status: Experimental  
Expires: October 26, 2011

A. Zamfir  
Z. Ali  
Cisco Systems  
D. Papadimitriou  
Alcatel-Lucent  
April 27, 2011

## Component Link Recording and Resource Control for TE Links

draft-ietf-mppls-explicit-resource-control-bundle-10.txt

### Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on October 26, 2011.

### Abstract

Record Route is a useful administrative tool that has been used extensively by the service providers. However, when TE links are bundled, identification of label resource in Record Route object (RRO) is not sufficient to determine the component link within a TE link that is being used by a given LSP. In other words, when link bundling is used, resource recording requires mechanisms to specify the component link identifier, along with the TE link identifier and Label. As it is not possible to record component link in the RRO, this document defines the extensions to RSVP-TE [RFC3209] and [RFC3473] to specify component link identifiers for resource recording purposes.

A.Zamfir et al. Expires October 26, 2011 [Page 1]  
Component Link Recording & Resource Control for TE Links April 27, 2011

This document also defines the Explicit Route object (ERO) counterpart of the RRO extension. The ERO extensions are needed to perform explicit label/resource control over bundled TE link. Hence, this document defines the extensions to RSVP-TE [RFC3209] and [RFC3473] to specify component link identifiers for explicit

resource control and recording over TE link bundles.

## Table of Contents

1. Introduction.....	3
Conventions used in this document.....	4
2. Terminology.....	4
3. LSP Resource Recording.....	4
3.1. Component Interface Identifier RRO Subobject.....	5
3.2. Processing of Component Interface ID RRO Subobject.....	6
4. Signaling Component Interface Identifier in ERO.....	7
4.1. Component Interface ID ERO Subobject.....	7
4.2. Processing of Component Interface ID ERO Subobject.....	8
5. Backward Compatibility.....	10
6. Security Considerations.....	10
7. IANA Considerations.....	10
8. References.....	11
8.1. Normative References.....	11
9. Acknowledgments.....	12
10. Copyright.....	13

## 1. Introduction

In GMPLS networks [RFC3945] where unbundled (being either Packet-Switching Capable, Layer2-Switching Capable, Time Division Multiplexing or Lambda-Switching Capable) Traffic Engineering (TE) Links are used, one of the types of resources that an LSP originator could control and record are the component links used by non-neighboring nodes on the LSP path. The resource control and recording is done by the use of the EXPLICIT\_ROUTE object (ERO) and RECORD\_ROUTE object (RRO), respectively.

Link Bundling, introduced in [RFC4201], is used to improve routing scalability by reducing the amount of TE related information that need to be flooded and handled by IGP in a TE network. This is accomplished by aggregating and abstracting the TE Link components.



In some cases the component link selection/recording within a TE link is left as a local decision (ERO and RRO contains only TE links). However there are cases when it is desirable for a non-local (e.g., LSP head-end) node to make this selection. The use of such information has found since so far three main applications (while not excluding others unknown at the time of writing): diagnostic, association of component specific attributes for which the bundled information is too coarse (e.g., Shared Risk Link Groups) and thus blocking SRLG-disjoint LSP establishment, allocation of labels at network edges, and notification in case of failures. The latter is useful when a single TE link interconnects two parts of the network. In case one of its components fails notifying a complete TE link failure leaves the network disconnected. In either case, it is required to know which component link within a bundled TE link has been used for a given LSP. For these cases, the TE Link and the Label currently specified in the ERO/RRO are not enough and the component link needs to be specified along with the label. In the case of bi-directional Label Switched Paths (LSP) both upstream and downstream information may be specified. Therefore, explicit resource control and recording over a bundled TE link also requires ability to specify a component link within the TE link.

Another important assumption of this document is that the identifier space used for component link identification are unique for a given node (following [RFC4201]). The reason stems as follows: most experimental developments started with TE links composed by a single component link and then only bundling was added by grouping them. Component links were thus identified such that they could mimic the behavior of TE link processing. This also justifies the experimental status of this document.

This document defines extensions to and describes the use of RSVP-TE

A.Zamfir et al. Expires October 26, 2011 [Page 3]  
Component Link Recording & Resource Control for TE Links April 27, 2011

[RFC3209], [RFC3471], [RFC3473] to specify the component link identifier for resource recording and explicit resource control over TE link bundles. Specifically, in this document, component interface identifier RRO and ERO subobjects are defined to complement their Label RRO and ERO counterparts. Furthermore, procedures for processing component interface identifier RRO and ERO subobjects and how they can co-exist with the Label RRO and ERO subobjects are specified.

#### Conventions used in this document

In examples, "C:" and "S:" indicate lines sent by the client and server respectively.

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].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC 2119 significance.

In this document, the characters ">>" preceding an indented line(s)

indicates a compliance requirement statement using the key words listed above. This convention aids reviewers in quickly identifying or finding the explicit compliance requirements of this RFC.

## 2. Terminology

o) TE Link: Unless specified otherwise, it refers to a bundled Traffic Engineering link as defined in [RFC4201]. Furthermore, the terms TE Link and bundled TE Link are used interchangeably in this document.

o) Component (interface) link: refers (locally) to a link that is part of a bundled TE link as described in RFC4201.

o) Component Interface Identifier: Refers to an ID used to uniquely identify a Component Interface. on a bundled link a combination of <component link identifier, label> is sufficient to unambiguously identify the appropriate resources used by an LSP. The IDs used for component link identification are unique for a given node [RFC4201].

## 3. LSP Resource Recording

LSP Resource Recording refers to the ability to record the resources used by an LSP.

A.Zamfir et al. Expires October 26, 2011 [Page 4]  
Component Link Recording & Resource Control for TE Links April 27, 2011

The procedure for unbundled numbered TE links is described in [RFC3209] and for unbundled unnumbered TE links in [RFC3477]. For the purpose of recording LSP resources used over bundled TE Links, the Component Interface Identifier RRO sub-object is introduced.

### 3.1. Component Interface Identifier RRO subobject

A new subobject of the Record Route object (RRO) is used to record component interface identifier of a (bundled) TE Link. This subobject 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
+-----+-----+-----+-----+-----+-----+-----+-----+
|L|      Type      |      Length      |U| Reserved (must be zero) |
+-----+-----+-----+-----+-----+-----+-----+-----+
|
//   IPv4, IPv6 or unnumbered Component Interface Identifier   //
|
+-----+-----+-----+-----+-----+-----+-----+-----+

```

L: 1 bit

This bit must be set to 0.

Type

Type 10 (TBD): Component Interface identifier IPv4  
Type 11 (TBD): Component Interface identifier IPv6  
Type 12 (TBD): Component Interface identifier Unnumbered

Length

#### Component Link Record. & Resource Control for TE Link Bundles

The Length contains the total length of the subobject in bytes, including the Type and Length fields. The Length is 8 bytes for the Component Interface identifier IPv4 and Component Interface identifier Unnumbered types. For Component Interface identifier IPv6 type of sub-object, the length field is 20 bytes.

U: 1 bit

This bit indicates the direction of the component interface. It is set to 0 for the downstream interface. It is set to 1 for the upstream interface and is only used for bi-directional LSPs.

A.Zamfir et al. Expires October 26, 2011 [Page 5]  
Component Link Recording & Resource Control for TE Links April 27, 2011

### 3.2. Processing of Component Interface identifier RRO Subobject

If a node desires component link recording, the "Component Link Recording desired" flag (value TBD) should be set in the LSP\_ATTRIBUTES object, object that is defined in [RFC5420].

Setting of "Component Link Recording desired" flag is independent of the Label Recording flag in SESSION\_ATTRIBUTE object as specified in [RFC3209]. Nevertheless, the following combinations are valid:

- 1) If both Label and Component Link flags are clear, then neither Labels nor Component Links are recorded.
- 2) If Label Recording flag is set and Component Link flag is clear, then only Label Recording is performed as defined in [RFC3209].
- 3) If Label Recording flag is clear and Component Link flag is set, then Component Link Recording is performed as defined in this document.
- 4) If both Label Recording and Component Link flags are set, then Label Recording is performed as defined in [RFC3209] and also Component Link recording is performed as defined in this document.

In most cases, a node initiates recording for a given LSP by adding the RRO to the Path message. If the node desires Component Link recording and if the outgoing TE link is bundled, then the initial RRO contains the Component Link identifier (numbered or unnumbered) as selected by the sender. As well, the Component Link Recording desired flag is set in the LSP\_ATTRIBUTE object. If the node also desires label recording, it sets the Label\_Recording flag in the SESSION\_ATTRIBUTE object.

When a Path message with the "Component Link Recording desired" flag set is received by an intermediate node, if a new Path message is to be sent for a downstream bundled TE link, the node adds a new Component Link subobject to the RECORD\_ROUTE object (RRO) and

appends the resulting RRO to the Path message before transmission. Note also that, unlike Labels, Component Link identifiers are always known on receipt of the Path message.

When the destination node of an RSVP session receives a Path message with an RRO and the "Component Link Recording desired" flag set, this indicates that the sender node needs TE route as well as component link recording. The destination node initiates the RRO

process by adding an RRO to Resv messages. The processing mirrors that of the Path messages. The Component Interface Record subobject is pushed onto the RECORD\_ROUTE object (RRO) prior to pushing on the node's IP address. A node MUST NOT push on a Component Interface Record subobject without also pushing on the IP address or unnumbered Interface Id subobject that identifies the TE Link.

When component interfaces are recorded for unidirectional LSPs, the downstream interface is the one identified by the Component Interface subobject. For bi-directional LSPs, component interface RRO subobjects for both downstream and upstream interfaces MUST be included.

#### 4. Signaling Component Interface Identifier in ERO

##### 4.1. Component Interface Identifier ERO subobject

A new OPTIoNAL subobject of the EXPLICIT\_ROUTE object (ERO) is used to specify component interface identifier of a bundled TE Link. This Component Interface Identifier subobject 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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|L|   Type   |   Length   |U|   Reserved (MUST be zero)   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                                     |
|//  IPv4, IPv6 or unnumbered Component Interface Identifier  //
|                                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

L: 1 bit

This bit must be set to 0.

Type

Type 10 (TBD): Component Interface identifier IPv4  
Type 11 (TBD): Component Interface identifier IPv6  
Type 12 (TBD): Component Interface identifier Unnumbered

Length

The Length contains the total length of the subobject in bytes, including the Type and Length fields. The Length is 8 bytes for the Component Interface identifier types: IPv4

and Component Interface identifier Unnumbered. For Component Interface identifier IPv6 type of sub-object, the length field

A.Zamfir et al. Expires October 26, 2011 [Page 7]  
Component Link Recording & Resource Control for TE Links April 27, 2011

is 20 bytes.

U: 1 bit

This bit indicates the direction of the component interface. It is 0 for the downstream interface. It is set to 1 for the upstream interface and is only used for bi-directional LSPs.

#### 4.2. Processing of Component Interface Identifier ERO Subobject

The Component Interface Identifier ERO subobject follows a subobject containing the IP address, or the link identifier [RFC3477], associated with the TE link on which it is to be used. It is used to identify the component of a bundled TE Link.

The following SHOULD result in "Bad EXPLICIT\_ROUTE object" error being sent upstream by a node processing an ERO that contains the Component Interface ID sub-object:

- o) The first component interface identifier subobject is not preceded by a sub-object containing an IPv4 or IPv6 address, or an interface identifier [RFC3477], associated with a TE link.
- o) The Component Interface Identifier ERO subobject follows a subobject that has the L-bit set.
- o) on unidirectional LSP setup, there is a Component Interface Identifier ERO subobject with the U-bit set.
- o) Two Component Interface Identifier ERO subobjects with the same U-bit values exist.

If a node implements the component interface identifier subobject, it MUST check if it represents a component interface in the bundled TE Link specified in the preceding subobject that contains the IPv4/IPv6 address or interface identifier of the TE Link. If the content of the component interface identifier subobject does not match a component interface in the TE link, a "Bad EXPLICIT\_ROUTE object" error SHOULD be reported as "Routing Problem" (error code 24).

If U-bit of the subobject being examined is cleared (0) and the upstream interface specified in this subobject is acceptable, then the value of the upstream component interface is translated locally in the TLV of the IF\_ID RSVP\_HOP object [RFC3471]. The local decision normally used to select the upstream component link is bypassed except for local translation into the outgoing interface identifier from the received incoming remote interface identifier.

A.Zamfir et al. Expires October 26, 2011 [Page 8]  
Component Link Recording & Resource Control for TE Links April 27, 2011

If this interface is not acceptable, a "Bad EXPLICIT\_ROUTE object"

error SHOULD be reported as "Routing Problem" (error code 24).

If the U-bit of the subobject being examined is set (1), then the value represents the component interface to be used for upstream traffic associated with the bidirectional LSP. Again, if this interface is not acceptable or if the request is not one for a bidirectional LSP, then a "Bad EXPLICIT\_ROUTE object" error SHOULD be reported as "Routing Problem" (error code 24). otherwise, the component interface IP address/ identifier is copied into a TLV sub-object as part of the IF\_ID RSVP\_HOP object. The local decision normally used to select the upstream component link is bypassed except for local translation into the outgoing interface identifier from the received incoming remote interface identifier.

The IF\_ID RSVP\_HOP object constructed as above MUST be included in the corresponding outgoing Path message.

Note that, associated with a TE Link sub-object in the ERO, either the (remote) upstream component interface or the (remote) downstream component interface or both may be specified. As specified in [RFC4201] there is no relationship between the TE Link type (numbered or unnumbered) and the Link type of any one of its components.

The Component Interface Identifier ERO subobject is optional. Similarly, presence of the Label ERO sub-objects is not mandatory [RFC3471], [RFC3473]. Furthermore, component interface identifier ERO subobject and Label ERO subobject may be included in the ERO independently of each other. one of the following alternatives applies:

- o) When both sub-objects are absent, a node may select any appropriate component link within the TE link and any label on the selected component link.
- o) When the Label subobject is only present for a bundled link, then the selection of the component link within the bundle is a local decision and the node may select any appropriate component link, which can assume the label specified in the Label ERO.
- o) When only the component interface identifier ERO subobject is present, a node MUST select the component interface specified in the ERO and may select any appropriate label value at the specified component link.
- o) When both component interface identifier ERO subobject and Label ERO subobject are present, the node MUST select the locally

corresponding component link and the specified label value on that component link. When present, both subobjects may appear in any relative order to each other but they MUST appear after the TE Link subobject that they refer to.

After processing, the component interface identifier subobjects are removed from the ERO.

Inferred from above, the interface subobject should never be the

first subobject in a newly received message. If the component interface subobject is the first subobject in a received ERO, then it SHOULD be treated as a "Bad strict node" error.

Note: Information to construct the Component Interface ERO subobject MAY come from the same mean used to populate the label ERO subobject. Procedures by which an LSR at the head-end of an LSP obtains the information needed to construct the Component Interface subobject are outside the scope of this document.

## 5. Backward Compatibility

The extensions specified in this document do not affect the processing of the RRO, ERO at nodes that do not support them. A node that does not support the Component Interface RRO subobject but that does support Label subobject SHOULD only insert the Label subobject in the RRO as per [RFC3471] and [RFC3473].

A node that receives an ERO that contains a Component Link ID subobject SHOULD send "Bad EXPLICIT\_ROUTE object" if it does not implement this subobject.

Per [RFC3209], Section 4.4.5, a non-compliant node that receives an RRO that contains Component Interface Identifier sub-objects should ignore and pass them on. This limits the full applicability of if nodes traversed by the LSP are compliant with the proposed extensions.

## 6. Security Considerations

An implementation of the extensions described in this document does exposes the component interface identifiers to other nodes in the network. If this is considered confidential information the mechanisms described in [RFC5920] should be considered.

## 7. IANA Considerations

This document introduces the following RSVP protocol elements:

A.Zamfir et al. Expires October 26, 2011 [Page 10]  
Component Link Recording & Resource Control for TE Links April 27, 2011

o) Component Interface Identifier RRO subobject of the RECORD\_ROUTE object (RRO):

- . IANA registry: RSVP PARAMETERS
- . Registry Name: Class Names, Class Numbers, and Class Types
- . Reference: [RFC3936]
- . Following subobjects have been added to the existing entry for:

### 21 RECORD\_ROUTE

- . Type 10 (TBD): Component Interface identifier IPv4
- . Type 11 (TBD): Component Interface identifier IPv6
- . Type 12 (TBD): Component Interface identifier Unnumbered

o) Component Interface Identifier subobject of the EXPLICIT\_ROUTE object (ERO).

- . IANA registry: RSVP PARAMETERS

- . Registry Name: Class Names, Class Numbers, and Class Types
- . Reference: [RFC3936]
- . Following subobjects have been added to the existing entry for:

20 EXPLICIT\_ROUTE

- . Type 10 (TBD): Component Interface identifier IPv4
- . Type 11 (TBD): Component Interface identifier IPv6
- . Type 12 (TBD): Component Interface identifier Unnumbered

o) A new "Component Link Recording desired" flag (value TBD) of the LSP\_ATTRIBUTES object [RFC5420]:

- . Bit Flag: 0x80
- . Name: Local Component Link Recording desired

## 8. References

### 8.1. Normative References

- [RFC2205] R.Braden, et al., "Resource ReSerVation Protocol (RSVP) Version 1, Functional Specification", RFC 2205, September 1997.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

A.Zamfir et al. Expires October 26, 2011 [Page 11]  
Component Link Recording & Resource Control for TE Links April 27, 2011

- [RFC2234] Crocker, D. and Overell, P. (Editors), "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, Internet Mail Consortium and Demon Internet Ltd., November 1997.
- [RFC3209] D.Awduche, et al., "Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3471] L.Berger, et al., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] L.Berger, et al., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3477] K.Kompella, et al., "Signaling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, January 2003.
- [RFC3945] E.Mannie, et al., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [RFC4201] K.Kompella, et al., "Link Bundling in MPLS Traffic Engineering", RFC 4201, January 2003.



[RFC5420] A.Farrel, et al., "Encoding of Attributes for Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) Establishment Using Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 5420.

[RFC5920] L.Fang, "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

## 9. Acknowledgments

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).

A.Zamfir et al. Expires October 26, 2011 [Page 12]  
Component Link Recording & Resource Control for TE Links April 27, 2011

## 10. Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Authors Addresses

Anca Zamfir  
Cisco Systems Inc.  
Email: [ancaz@cisco.com](mailto:ancaz@cisco.com)

Zafar Ali  
Cisco systems Inc.  
Email: [zali@cisco.com](mailto:zali@cisco.com)

Dimitri Papadimitriou  
Alcatel-Lucent Bell  
Email: [dimitri.papadimitriou@alcatel-lucent.com](mailto:dimitri.papadimitriou@alcatel-lucent.com)



Network Working Group  
Internet Draft  
Category: Standards Track  
Expiration Date: August 2011

R. Aggarwal  
Juniper Networks

J. L. Le Roux  
France Telecom

February 02, 2011

## MPLS Upstream Label Assignment for LDP

draft-ietf-mpls-ldp-upstream-10.txt

### Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at  
<http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at  
<http://www.ietf.org/shadow.html>.

### Copyright and License Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as

described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

#### Abstract

This document describes procedures for distributing upstream-assigned labels for Label Distribution Protocol (LDP). It also describes how these procedures can be used for avoiding branch Label Switching Router (LSR) traffic replication on a LAN for LDP point-to-multipoint (P2MP) Label Switched Paths (LSPs).

## Table of Contents

1	Specification of requirements .....	3
2	Introduction .....	3
3	LDP Upstream Label Assignment Capability .....	4
4	Distributing Upstream-Assigned Labels in LDP .....	5
4.1	Procedures .....	5
5	LDP Tunnel Identifier Exchange .....	6
6	LDP Point-to-Multipoint LSPs on a LAN .....	10
7	IANA Considerations .....	12
7.1	LDP TLVs .....	12
7.2	Interface Type Identifiers .....	12
8	Security Considerations .....	12
9	Acknowledgements .....	13
10	References .....	13
10.1	Normative References .....	13
10.2	Informative References .....	13
11	Author's Address .....	14

## 1. Specification of requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

## 2. Introduction

This document describes procedures for distributing upstream-assigned labels [RFC5331] for Label Distribution Protocol (LDP) [RFC5036]. These procedures follow the architecture for MPLS Upstream Label Assignment described in [RFC5331].

This document describes extensions to LDP that a Label Switching Router (LSR) can use to advertise to its neighboring LSRs whether the LSR supports upstream label assignment.

This document also describes extensions to LDP to distribute

upstream-assigned labels.

The usage of MPLS upstream label assignment using LDP for avoiding branch LSR traffic replication on a LAN for LDP point-to-multipoint (P2MP) Label Switched Paths (LSPs) [MLDP] is also described.

### 3. LDP Upstream Label Assignment Capability

According to [RFC5331], upstream-assigned label bindings MUST NOT be used unless it is known that a downstream LSR supports them. This implies that there MUST be a mechanism to enable an LSR to advertise to its LDP neighbor LSR(s) its support of upstream-assigned labels.

A new Capability Parameter, the LDP Upstream Label Assignment Capability, is introduced to allow an LDP peer to exchange with its peers, its support of upstream label assignment. This parameter follows the format and procedures for exchanging Capability Parameters defined in [RFC5561].

Following is the format of the LDP Upstream Label Assignment Capability Parameter:

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|1|0| Upstream Lbl Ass Cap(IANA)|           Length (= 1)           |
+-----+-----+-----+-----+-----+-----+-----+-----+
|1| Reserved           |
+-----+-----+-----+-----+

```

If an LSR includes the Upstream Label Assignment Capability in LDP Initialization Messages it implies that the LSR is capable of both distributing upstream-assigned label bindings and receiving upstream-assigned label bindings. The reserved bits MUST be set to zero on transmission and ignored on receipt. The Upstream Label Assignment Capability Parameter MUST be carried only in LDP initialization messages and MUST be ignored if received in LDP Capability messages.

#### 4. Distributing Upstream-Assigned Labels in LDP

An optional LDP TLV, Upstream-Assigned Label Request TLV, is introduced. To request an upstream-assigned label an LDP peer MUST include this TLV in a Label Request 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| Upstream Ass Lbl Req (TBD)|          Length          |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Reserved                 |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

An optional LDP TLV, Upstream-Assigned Label TLV is introduced to signal an upstream-assigned label. Upstream-Assigned Label TLVs are carried by the messages used to advertise, release and withdraw upstream assigned label mappings.

```

      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| Upstream Ass Label (TBD) |          Length          |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Reserved                 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Label                    |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

The Label field is a 20-bit label value as specified in [RFC3032] represented as a 20-bit number in a 4 octet field as specified in section 3.4.2.1 of RFC5036 [RFC5036].

##### 4.1. Procedures

Procedures for Label Mapping, Label Request, Label Abort, Label Withdraw and Label Release follow [RFC5036] other than the modifications pointed out in this section.

A LDP LSR MUST NOT distribute the Upstream Assigned Label TLV to a neighboring LSR if the neighboring LSR had not previously advertised the Upstream Label Assignment Capability in its LDP Initialization messages. A LDP LSR MUST NOT send the Upstream Assigned Label Request TLV to a neighboring LSR if the neighboring LSR had not previously advertised the Upstream Label Assignment Capability in its LDP Initialization messages.

As described in [RFC5331] the distribution of upstream-assigned labels is similar to either ordered LSP control or independent LSP control of the downstream assigned labels.

When the label distributed in a Label Mapping message is an upstream-assigned label, the Upstream Assigned Label TLV MUST be included in the Label Mapping message. When an LSR receives a Label Mapping message with an Upstream Assigned Label TLV and it does not recognize the TLV, it MUST generate a Notification message with a status code of "Unknown TLV" [RFC5036]. If it does recognize the TLV but is unable to process the upstream label, it MUST generate a Notification message with a status code of "No Label Resources". If the Label Mapping message was generated in response to a Label Request message, the Label Request message MUST contain an Upstream Assigned Label Request TLV. A LSR that generates an upstream assigned label request to a neighbor LSR, for a given FEC, MUST NOT send a downstream label mapping to the neighbor LSR for that FEC unless it withdraws the upstream-assigned label binding. Similarly if an LSR generates a downstream assigned label request to a neighbor LSR, for a given FEC, it MUST NOT send an upstream label mapping to that LSR for that FEC, unless it aborts the downstream assigned label request.

The Upstream Assigned Label TLV may be optionally included in Label Withdraw and Label Release messages that withdraw/release a particular upstream assigned label binding.

## 5. LDP Tunnel Identifier Exchange

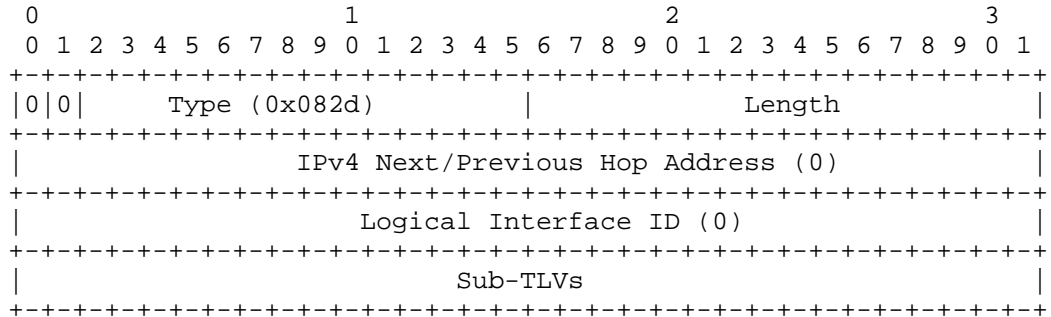
As described in [RFC5331] an upstream LSR Ru MAY transmit an MPLS packet, the top label of which (L) is upstream-assigned, to a downstream LSR Rd, by encapsulating it in an IP or MPLS tunnel. In this case the fact that L is upstream-assigned is determined by Rd by the tunnel on which the packet is received. There must be a mechanism for Ru to inform Rd that a particular tunnel from Ru to Rd will be used by Ru for transmitting MPLS packets with upstream-assigned MPLS labels.

When LDP is used for upstream label assignment, the Interface ID TLV [RFC3472] is used for signaling the Tunnel Identifier. If Ru uses an IP or MPLS tunnel to transmit MPLS packets with upstream assigned labels to Rd, Ru MUST include the Interface ID TLV in the Label Mapping messages along with the Upstream Assigned Label TLV. The IPv4/v6 Next/Previous Hop Address and the Logical Interface ID fields in the Interface ID TLV SHOULD be set to 0 by the sender and ignored by the receiver. The Length field indicates the total length of the TLV, i.e., 4 + the length of the value field in octets. A value field whose length is not a multiple of four MUST be zero-padded so

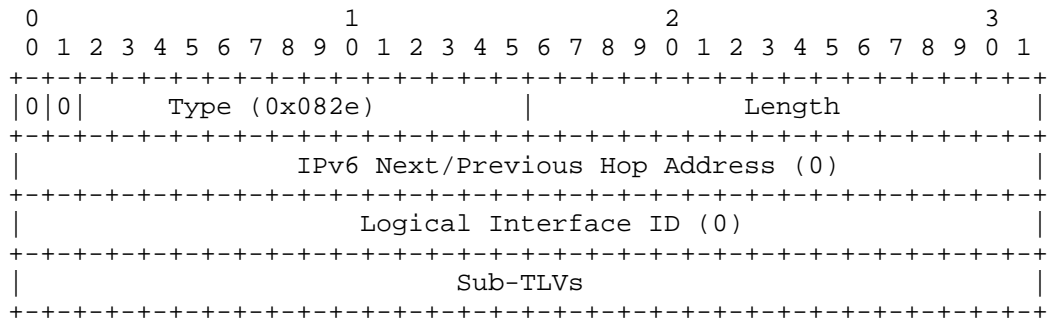


that the TLV is four- octet aligned.

Hence the IPv4 Interface ID TLV has the following format:



The IPv6 Interface ID TLV has the following format:

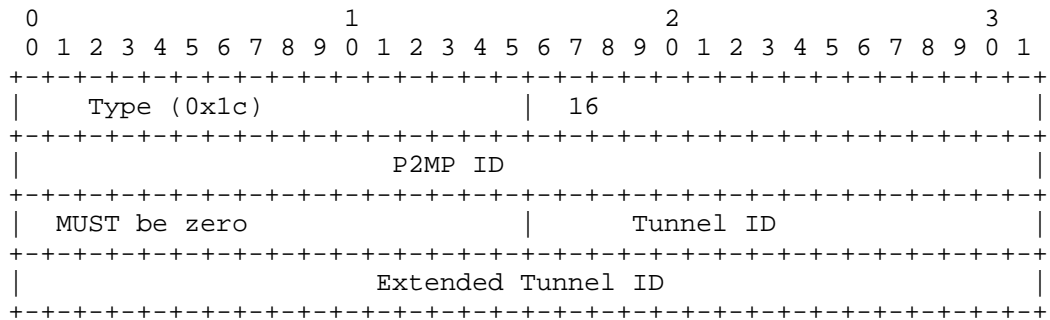


As shown in the above figures the Interface ID TLV carries sub-TLVs. Four new Interface ID sub-TLVs are introduced to support RSVP-TE P2MP LSPs, LDP P2MP LSPs, IP Multicast Tunnels and context labels. The sub-TLV value in the sub-TLV acts as the tunnel identifier.

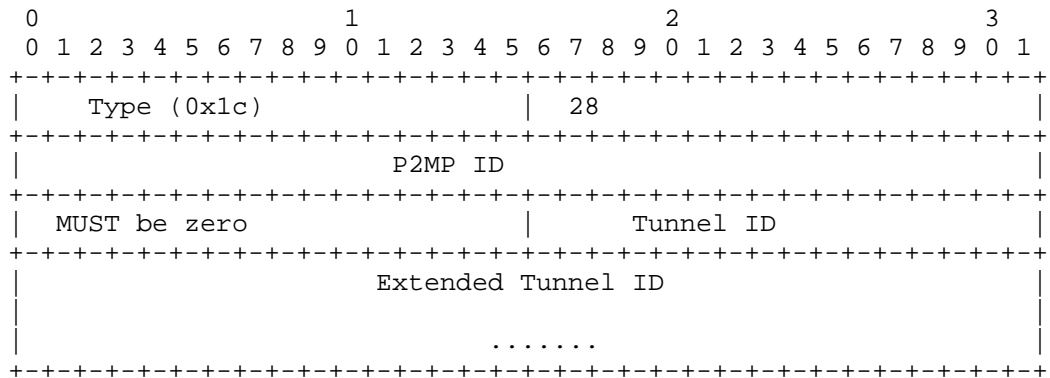
Following are the sub-TLVs that are introduced:

1. RSVP-TE P2MP LSP TLV. Type = 28 (To be assigned by IANA). Value of the TLV is the RSVP-TE P2MP LSP SESSION Object [RFC4875].

Below is the RSVP-TE P2MP LSP TLV format when carried in the IPv4 Interface ID TLV:

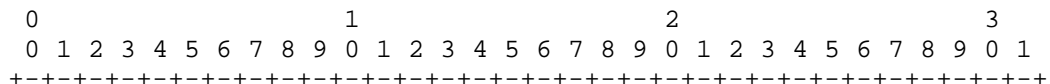


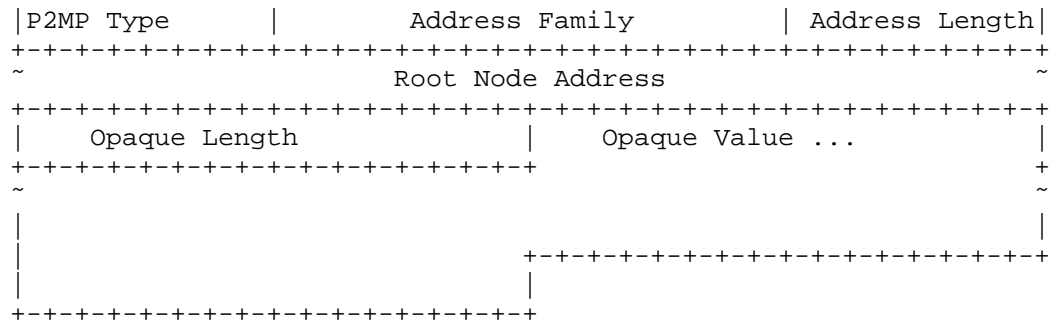
Below is the RSVP-TE P2MP LSP TLV format when carried in the IPv6 Interface ID TLV:



This TLV identifies the RSVP-TE P2MP LSP. It allows Ru to tunnel an "inner" LDP P2MP LSP, the label for which is upstream assigned, over an "outer" RSVP-TE P2MP LSP that has leaves <Rd1...Rdn>. The RSVP-TE P2MP LSP IF\_ID TLV allows Ru to signal to <Rd1...Rdn> the binding of the inner LDP P2MP LSP to the outer RSVP-TE P2MP LSP. The control plane signaling between Ru and <Rd1...Rdn> for the inner P2MP LSP uses targeted LDP signaling messages

2. LDP P2MP LSP TLV. Type = 29 (To be assigned by IANA). Value of the TLV is the LDP P2MP FEC as defined in [MLDP] and has to be set as per the procedures in [MLDP]. Here is the format of the LDP P2MP FEC as defined in [MLDP]:





The Address Family MUST be set to IPv4, the Address Length MUST be set to 4 and the Root Node Address MUST be set to an IPv4 address when the LDP P2MP LSP TLV is carried in the IPv4 Interface ID TLV. The Address Family MUST be set to IPv6, the Address Length MUST be set to 16 and the Root Node Address MUST be set to an IPv6 address when the LDP P2MP LSP TLV is carried in the IPv6 Interface ID TLV.

The TLV value identifies the LDP P2MP LSP. It allows Ru to tunnel an "inner" LDP P2MP LSP, the label for which is upstream assigned, over an "outer" LDP P2MP LSP that has leaves <Rd1...Rdn>. The LDP P2MP LSP IF\_ID TLV allows Ru to signal to <Rd1...Rdn> the binding of the inner LDP P2MP LSP to the outer LDP- P2MP LSP. The control plane signaling between Ru and <Rd1...Rdn> for the inner P2MP LSP uses targeted LDP signaling messages

3. IP Multicast Tunnel TLV. Type = 30 (To be assigned by IANA) In this case the TLV value is a <Source Address, Multicast Group Address> tuple. Source Address is the IP address of the root of the tunnel i.e. Ru, and Multicast Group Address is the Multicast Group Address used by the tunnel. The addresses MUST be IPv4 addresses when the IP Multicast Tunnel TLV is included in the IPv4 Interface ID TLV. The addresses MUST be IPv6 addresses when the IP Multicast Tunnel TLV is included in the IPv6 Interface ID TLV.

4. MPLS Context Label TLV. Type = 31 (To be assigned by IANA). In this case the TLV value is a <Source Address, MPLS Context Label> tuple. The Source Address belongs to Ru and the MPLS Context Label is an upstream assigned label, assigned by Ru. The Source Address MUST be set to an IPv4 address when the MPLS Context Label TLV is carried in the IPv4 Interface ID TLV. The Source Address MUST be set to an IPv6 address when the MPLS Context Label TLV is carried in the IPv6 Interface ID TLV. This allows Ru to tunnel an "inner" LDP P2MP LSP, the label of which is upstream assigned, over an "outer" one-hop MPLS LSP, where the outer one-hop LSP has the following property:

- + The label pushed by Ru for the outer MPLS LSP is an upstream assigned context label, assigned by Ru. When <Rd1...Rdn> perform an MPLS label lookup on this label a combination of this label and the incoming interface MUST be sufficient for <Rd1...Rdn> to uniquely determine Ru's context specific label space to lookup the next label on the stack in. <Rd1...Rdn> MUST receive the data sent by Ru with the context specific label assigned by Ru being the top label on the label stack.

Currently the usage of the context label TLV is limited only to LDP P2MP LSPs on a LAN as specified in the next section. The context label TLV MUST NOT be used for any other purposes.

Note that when the outer P2MP LSP is signaled with RSVP-TE or MLDP the above procedures assume that Ru has a priori knowledge of all the <Rd1, ... Rdn>. In the scenario where the outer P2MP LSP is signaled using RSVP-TE, Ru can obtain this information from RSVP-TE. However, in the scenario where the outer P2MP LSP is signaled using MLDP, MLDP does not provide this information to Ru. In this scenario the procedures by which Ru could acquire this information are outside the scope of this document.

## 6. LDP Point-to-Multipoint LSPs on a LAN

This section describes one application of upstream label assignment using LDP. Further applications are to be described in separate documents.

[MLDP] describes how to setup P2MP LSPs using LDP. On a LAN the solution relies on "ingress replication". A LSR on a LAN, that is a branch LSR for a P2MP LSP, (say Ru) sends a separate copy of a packet that it receives on the P2MP LSP to each of the downstream LSRs on the LAN (say <Rd1...Rdn> that are adjacent to it in the P2MP LSP.

It is desirable for Ru to send a single copy of the packet for the LDP P2MP LSP on the LAN, when there are multiple downstream routers on the LAN that are adjacent to Ru in that LDP P2MP LSP. This requires that each of <Rd1...Rdn> must be able to associate the label L, used by Ru to transmit packets for the P2MP LSP on the LAN, with that P2MP LSP. It is possible to achieve this using LDP upstream-assigned labels with the following procedures.

Consider an LSR Rd that receives the LDP P2MP FEC [MLDP] from its downstream LDP peer. Further the upstream interface to reach LSR Ru which is the next-hop to the P2MP LSP root address, Pr, in the LDP P2MP FEC, is a LAN interface, Li. Further Rd and Ru support upstream-assigned labels. In this case Rd instead of sending a Label Mapping

message as described in [MLDP] sends a Label Request message to Ru. This Label Request message MUST contain an Upstream Assigned Label Request TLV.

On receiving this message, Ru sends back a Label Mapping message to Rd with an upstream-assigned label. This message also contains an Interface ID TLV with a MPLS Context Label sub-TLV, as described in the previous section, with the value of the MPLS label set to a value assigned by Ru on interface Li as specified in [RFC5331]. Processing of the Label Request and Label Mapping messages for LDP upstream-assigned labels is as described in section 4.1. If Ru receives a Label Request for an upstream assigned label for the same P2MP FEC from multiple downstream LSRs on the LAN, <Rd1...Rdn>, it MUST send the same upstream-assigned label to each of <Rd1...Rdn>.

Ru transmits the MPLS packet using the procedures defined in [RFC5331] and [RFC5332]. The MPLS packet transmitted by Ru contains as the top label the context label assigned by Ru on the LAN interface, Li. The bottom label is the upstream label assigned by Ru to the LDP P2MP LSP. The top label is looked up in the context of the LAN interface, Li, [RFC5331] by a downstream LSR on the LAN. This lookup enables the downstream LSR to determine the context specific label space to lookup the inner label in.

Note that <Rd1...Rdn> may have more than one equal cost next-hop on the LAN to reach Pr. It MAY be desirable for all of them to send the label request to the same upstream LSR and they MAY select one upstream LSR using the following procedure:

1. The candidate upstream LSRs are numbered from lower to higher IP address
2. The following hash is performed:  $H = (\text{Sum Opaque value}) \bmod N$ , where N is the number of candidate upstream LSRs. Opaque value is defined in [MLDP] and comprises the P2MP LSP identifier.
3. The selected upstream LSR U is the LSR that has the number H.

This allows for load balancing of a set of LSPs among a set of candidate upstream LSRs, while ensuring that on a LAN interface a single upstream LSR is selected. It is also to be noted that the procedures in this section can still be used by Rd and Ru if other LSRs on the LAN do not support upstream label assignment. Ingress replication and downstream label assignment will continue to be used for LSRs that do not support upstream label assignment.

## 7. IANA Considerations

### 7.1. LDP TLVs

IANA maintains a registry of LDP TLVs at the registry "Label Distribution Protocol" in the sub-registry called "TLV Type Name Space".

This document defines a new LDP Upstream Label Assignment Capability TLV (Section 3). IANA is requested to assign the value 0x0507 to this TLV.

This document defines a new LDP Upstream-Assigned Label TLV (Section 4). IANA is requested to assign the type value of 0x204 to this TLV.

This document defines a new LDP Upstream-Assigned Label Request TLV (Section 4). IANA is requested to assign the type value of 0x205 to this TLV.

### 7.2. Interface Type Identifiers

[RFC3472] defines the LDP Interface ID IPv4 and IPv6 TLV. These top-level TLVs can carry sub-TLVs dependent on the interface type. These sub-TLVs are assigned "Interface ID Types". IANA maintains a registry of Interface ID Types for use in GMPLS in the registry "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Parameters" and sub-registry "Interface\_ID Types". IANA is requested to make corresponding allocations from this registry as follows:

- RSVP-TE P2MP LSP TLV (requested value 28)
- LDP P2MP LSP TLV (requested value 29)
- IP Multicast Tunnel TLV (requested value 30)
- MPLS Context Label TLV (requested value 31)

## 8. Security Considerations

The security considerations discussed in RFC 5036, RFC 5331 and RFC 5332 apply to this document.

More detailed discussion of security issues that are relevant in the context of MPLS and GMPLS, including security threats, related defensive techniques, and the mechanisms for detection and reporting, are discussed in "Security Framework for MPLS and GMPLS Networks

[MPLS-SEC].

## 9. Acknowledgements

Thanks to Yakov Rekhter for his contribution. Thanks to Ina Minei and Thomas Morin for their comments. The hashing algorithm used on LAN interfaces is taken from [MLDP]. Thanks to Loa Andersson, Adrian Farrel and Eric Rosen for their comments and review.

## 10. References

### 10.1. Normative References

[RFC5331] R. Aggarwal, Y. Rekhter, E. Rosen, "MPLS Upstream Label Assignment and Context Specific Label Space", RFC5331

[RFC5332] T. Eckert, E. Rosen, R. Aggarwal, Y. Rekhter, RFC5332

[RFC2119] "Key words for use in RFCs to Indicate Requirement Levels.", Bradner, March 1997

[RFC5036] L. Andersson, et. al., "LDP Specification", RFC5036.

[RFC4875] R. Aggarwal, D. Papadimitriou, S. Yasukawa [Editors], "Extensions to RSVP-TE for Point to Multipoint TE LSPs", RFC 4875

[MLDP] I. Minei et. al, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", draft-ietf-mpls-ldp-p2mp-08.txt

### 10.2. Informative References

[RFC5561] B. Thomas, K. Raza, S. Aggarwal, R. Aggarwal, JL. Le Roux, "LDP Capabilities", RFC5561

[MPLS-SEC] L. fang, ed, "Security Framework for MPLS and GMPLS Networks", draft-ietf-mpls-mpls-and-gmpls-security-framework-07.txt

[RFC3032] E. Rosen et. al, "MPLS Label Stack Encoding", RFC 3032

[RFC3472] Ashwood-Smith, P. and L. Berger, Editors, " Generalized Multi-Protocol Label Switching (GMPLS) Signaling - Constraint-based Routed Label Distribution Protocol (CR-LDP) Extensions", RFC 3472, January 2003.

## 11. Author's Address

Rahul Aggarwal  
Juniper Networks  
1194 North Mathilda Ave.  
Sunnyvale, CA 94089  
Phone: +1-408-936-2720  
Email: rahul@juniper.net

Jean-Louis Le Roux  
France Telecom  
2, avenue Pierre-Marzin  
22307 Lannion Cedex  
France  
E-mail: jeanlouis.leroux@orange-ftgroup.com



MPLS Working Group

Internet Draft

Intended status: Standard Track

Expires: February 16, 2012

Z. Ali

G. Swallow

Cisco Systems, Inc.

R. Aggarwal

Juniper Networks

August 17, 2011

Non Penultimate Hop Popping Behavior and out-of-band mapping for  
RSVP-TE Label Switched Paths  
draft-ietf-mpls-rsvp-te-no-php-oob-mapping-09.txt

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on February 16, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Expires February 2012

[Page 1]

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

## Abstract

There are many deployment scenarios which require Egress Label Switching Router (LSR) to receive binding of the Resource ReserVation Protocol Traffic Engineered (RSVP-TE) Label Switched Path (LSP) to an application, and payload identification, using some "out-of-band" (OOB) mechanism. This document defines protocol mechanisms to address this requirement. The procedures described in this document are equally applicable for point-to-point (P2P) and point-to-multipoint (P2MP) LSPs.

## 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].

## Table of Contents

Copyright Notice .....	1
1. Introduction .....	3
2. RSVP-TE signaling extensions .....	4
2.1. Signaling non-PHP behavior .....	4
2.2. Signaling OOB Mapping Indication .....	5
2.3. Relationship between OOB and non-PHP flags .....	7
2.4. Egress Procedure for label binding .....	7
3. Security Considerations .....	8
4. IANA Considerations .....	8
4.1. Attribute Flags for LSP_ATTRIBUTES object .....	8
4.2. New RSVP error sub-code .....	9

5. Acknowledgments .....	9
6. References .....	9
6.1. Normative References .....	9
6.2. Informative References .....	10

## 1. Introduction

When Resource ReserVation Protocol Traffic Engineered (RSVP-TE) is used for applications like Multicast Virtual Private Network (MVPN) [MVPN] and Virtual Private LAN Service (VPLS) [RFC4761], an Egress Label Switching Router (LSR) receives the binding of the RSVP-TE Label Switched Path (LSP) to an application, and payload identification, using an "out-of-band" (OOB) mechanism (e.g., using Border Gateway Protocol (BGP)). In such cases, the Egress LSR cannot make correct forwarding decision until such OOB mapping information is received. Furthermore, in order to apply the binding information, the Egress LSR needs to identify the incoming LSP on which traffic is coming. Therefore, non Penultimate Hop Popping (non-PHP) behavior is required to apply OOB mapping. Non-PHP behavior requires the egress LSRs to assign a non-NULL label for the LSP being signaled.

There are other applications that require non-PHP behavior. When RSVP-TE point-to-multipoint (P2MP) LSPs are used to carry IP multicast traffic non-PHP behavior enables a leaf LSR to identify the P2MP TE LSP, on which traffic is received. Hence the egress LSR can determine whether traffic is received on the expected P2MP LSP and discard traffic that is not received on the expected P2MP LSP. Non-PHP behavior is also required to determine the context of upstream assigned labels when the context is a MPLS LSP. Non-PHP behavior may also be required for MPLS-TP LSPs [RFC5921].

This document defines two new flags in the Attributes Flags TLV of the LSP\_ATTRIBUTES object defined in [RFC5420]: one flag for communication of non-PHP behavior, and one flag to indicate that the binding of the LSP to an application and payload identifier (payload-Id) needs to be learned via an out-of-band mapping mechanism. As there is one-to-one correspondence between bits in the Attribute Flags TLV and the RRO Attributes subobject, corresponding flags to be carried in RRO Attributes subobject are also defined.

The procedures described in this document are equally applicable for P2P and P2MP LSPs. Specification of the OOB communication mechanism(s) is beyond the scope of this document.

## 2. RSVP-TE signaling extensions

This section describes the signaling extensions required to address the above-mentioned requirements.

### 2.1. Signaling non-PHP behavior

In order to request non-PHP behavior for an RSVP-TE LSP, this document defines a new flag in the Attributes Flags TLV of the LSP\_ATTRIBUTES object defined in [RFC5420]:

Bit Number (to be assigned by IANA): non-PHP behavior requested flag.

In order to indicate to the Ingress LSR that the Egress LSR recognizes the "non-PHP behavior requested flag", the following new bit is defined in the Flags field of the Record Route object (RRO) Attributes subobject:

Bit Number (same as bit number assigned for non-PHP behavior requested flag): Non-PHP behavior acknowledgement flag.

An Ingress LSR sets the "non-PHP behavior requested flag" to signal the egress LSRs SHOULD assign non-NULL label for the LSP being signaled. This flag MUST NOT be modified by any other LSRs in the network. LSRs other than the Egress LSRs SHOULD ignore this flag.

If an egress LSR receiving the Path message, supports the LSP\_ATTRIBUTES object and the Attributes Flags TLV, and also recognizes the "non-PHP behavior requested flag", it MUST allocate a non-NULL local label. The egress LSR MUST also set the "Non-PHP behavior acknowledgement flag" in the Flags field of the RRO Attribute subobject.

If the egress LSR

- supports the LSP\_ATTRIBUTES object but does not recognize the Attributes Flags TLV; or
- supports the LSP\_ATTRIBUTES object and recognize the Attributes Flags TLV, but does not recognize the "non-PHP behavior requested flag";

then it silently ignores this request according to the processing rules of [RFC5420].

An ingress LSR requesting non-PHP behavior SHOULD examine "Non-PHP behavior acknowledgement flag" in the Flags field of the RRO Attribute subobject and MAY send a Path Tear to the Egress which has not set the "Non-PHP behavior acknowledgement flag". An ingress LSR requesting non-PHP behavior MAY also examine the label value corresponding to the Egress LSR(s) in the RRO, and MAY send a Path Tear to the Egress which assigns a Null label value.

When signaling a P2MP LSP, a source node may wish to solicit individual response to the "non-PHP behavior requested flag" from the leaf nodes. Given the constraints on how the LSP\_ATTRIBUTES may be carried in Path and Resv Messages according to RFC5420, in this situation the source node MUST use a separate Path message for each leaf in networks where [ATTRIBUTE-BNF] is not supported. In networks with [ATTRIBUTE-BNF] deployed either separate Path message for each leaf or multiple leafs per Path message MAY be used by the source node.

## 2.2. Signaling OOB Mapping Indication

This document defines a single flag to indicate that the normal binding mechanism of an RSVP session is overridden. The actual out-of-band mappings are beyond the scope of this document. The flag is carried in the Attributes Flags TLV of the LSP\_ATTRIBUTES object defined in [RFC5420] and is defined as follows:

Bit Number (to be assigned by IANA): OOB mapping indication flag.

In order to indicate to the Ingress LSR that the Egress LSR recognizes the "OOB mapping indication flag", the following new bit is defined in the Flags field of the Record Route object (RRO) Attributes subobject:

Bit Number (same as bit number assigned for OOB mapping indication flag): OOB mapping acknowledgement flag.

An Ingress LSR sets the OOB mapping indication flag to signal the Egress LSR that binding of RSVP-TE LSP to an application and payload identification is being signaled out-of-band. This flag MUST NOT be modified by any other LSRs in the network. LSRs other than the Egress LSRs SHOULD ignore this flag.

When an Egress LSR which supports the "OOB mapping indication flag", receives a Path message with that flag set, the Egress LSR MUST set the "OOB mapping acknowledgement flag" in the Flags field of the RRO Attribute subobject. The rest of the RSVP signaling proceeds as normal. However, the LSR MUST have received the OOB mapping before accepting traffic on the LSP. This implies that the Egress LSR MUST NOT setup forwarding state for the LSP before it receives the OOB mapping.

Note that the payload information SHOULD be supplied by the OOB mapping. If the egress LSR receives the payload information from OOB mapping then the LSR MUST ignore L3PID in the Label Request Object [RFC3209].

If the egress LSR

- supports the LSP\_ATTRIBUTES object but does not recognize the Attributes Flags TLV; or
- supports the LSP\_ATTRIBUTES object and recognizes the Attributes Flags TLV, but does not recognize the "OOB mapping indication flag";

then it silently ignores this request according to the processing rules of [RFC5420].

An ingress LSR requesting OOB mapping SHOULD examine "OOB mapping acknowledgement flag" in the Flags field of the RRO Attribute subobject and MAY send a Path Tear to the Egress which has not set the "OOB mapping acknowledgement flag".

When signaling a P2MP LSP, a source node may wish to solicit individual response to the "OOB mapping indication flag" from the leaf nodes. Given the constraints on how the LSP\_ATTRIBUTES may be carried in Path and Resv Messages according to RFC5420, in this situation the source node MUST use a separate Path message for each leaf in networks where [ATTRIBUTE-BNF] is not supported. In

networks with [ATTRIBUTE-BNF] deployed either separate Path message for each leaf or multiple leafs per Path message MAY be used by the source node.

In deploying applications where Egress LSR receives the binding of the RSVP-TE LSP to an application, and payload identification, using OOB mechanism, it is important to recognize that the OOB mapping is sent asynchronously with respect to the signaling of RSVP-TE LSP. Egress LSR only installs forwarding state for the LSP after it receives the OOB mapping. In deploying applications using OOB mechanism, an Ingress LSR may need to know when the Egress is properly setup for forwarding (i.e., has received the OOB mapping). How the Ingress LSR determines that the LSR is properly setup for forwarding at the Egress LSR is beyond the scope of this document. Nonetheless, if the OOB mapping is not received by the Egress LSR within a reasonable time, the procedure defined in section 2.4 to tear down the LSP is followed.

### 2.3. Relationship between OOB and non-PHP flags

"Non-PHP behavior desired" and "OOB mapping indication" flags can appear and be processed independently of each other. However, as mentioned earlier, in the context of the applications discussed in this document, OOB mapping requires non-PHP behavior. An Ingress LSR requesting the OOB mapping MAY also set the "non-PHP behavior requested flag" in the LSP\_ATTRIBUTES object in the Path message.

### 2.4. Egress Procedure for label binding

RSVP-TE signaling completion and the OOB mapping information reception happen asynchronously at the Egress. As mentioned in Section 2.2, Egress waits for the OOB mapping before accepting traffic on the LSP. Nonetheless, MPLS OAM mechanisms, e.g., LSP Ping and Trace route as defined in [RFC4379], [P2MP-OAM], are expected to work independent of OOB mapping learning process.

In order to avoid unnecessary use of the resources and possible black-holing of traffic, an Egress LSR MAY send a Path Error message if the OOB mapping information is not received within a reasonable time. This Path Error message SHOULD include the error code/sub-code "Notify Error/ no OOB mapping received" for all affected LSPs. If notify request was included when the LSP was initially setup, Notify message (as defined in [RFC3473]) MAY also be used for delivery of this information to the Ingress LSR. An Egress LSR MAY implement a cleanup timer for this purpose. The

time-out value is a local decision at the Egress, with a RECOMMENDED default value of 60 seconds.

### 3. Security Considerations

Addition of "non-PHP behavior" adds a variable of attacks on the label assigned by the Egress node. As change in the value of the egress label reported in the RRO can cause the LSP to be torn down, additional security considerations for protecting label assigned by the Egress node are required. Security mechanisms as identified in [RFC5920], [RFC2205], [RFC3209], [RFC3473], [RFC5420] and [RFC4875] can be used for this purpose. This document does not introduce any additional security issues above those identified in [RFC5920], [RFC2205], [RFC3209], [RFC3473], [RFC5420] and [RFC4875].

### 4. IANA Considerations

The following changes to the Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Parameters registry are required.

#### 4.1. Attribute Flags for LSP\_ATTRIBUTES object

The following new flags are defined for the Attributes Flags TLV in the LSP\_ATTRIBUTES object. The numeric values are to be assigned by IANA.

o Non-PHP behavior flag:

This flag is used in the Attributes Flags TLV in a Path message. The flag has corresponding new flag to be used in the RRO Attributes subobject. As per [RFC5420], the bit numbering in the Attribute Flags TLV and the RRO Attributes subobject is identical. That is, the same attribute is indicated by the same bit in both places. This flag is not allowed in the Attributes Flags TLV in a Resv message. Specifically, Attributes of this flag are as follows:

- Bit Number: To be assigned by IANA.
- Attribute flag carried in Path message: Yes
- Attribute flag carried in Resv message: No
- Attribute flag carried in RRO message: Yes



o OOB mapping flag:

This flag is used in the Attributes Flags TLV in a Path message. The flag has corresponding new flag to be used in the RRO Attributes subobject. As per [RFC5420], the bit numbering in the Attribute Flags TLV and the RRO Attributes subobject is identical. That is, the same attribute is indicated by the same bit in both places. This flag is not allowed in the Attributes Flags TLV in a Resv message. Specifically, Attributes of this flag are as follows:

- Bit Number: To be assigned by IANA.
- Attribute flag carried in Path message: Yes
- Attribute flag carried in Resv message: No
- Attribute flag carried in RRO message: Yes

#### 4.2. New RSVP error sub-code

For Error Code = 25 "Notify Error" (see [RFC3209]) the following sub-code is defined.

Sub-code	Value
-----	-----
No OOB mapping received	to be assigned by IANA.

#### 5. Acknowledgments

The authors would like to thank Yakov Rekhter for his suggestions on the draft.

#### 6. References

##### 6.1. Normative References

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

- [RFC5420] A. Farrel, D. Papadimitriou, J. P. Vasseur and A. Ayyangar, "Encoding of Attributes for Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) Establishment Using RSVP-TE", RFC 5420, February 2006.
- [RFC3209] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC4875] R. Aggarwal, D. Papadimitriou, S. Yasukawa, et al, "Extensions to RSVP-TE for Point-to-Multipoint TE LSPs", RFC 4875.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003..
- [RFC2205] R. Braden, Ed., "Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification", RFC 2205, September 1997.
- [ATTRIBUTE-BNF] Berger, L. and Swallow, G., "LSP Attributes Related Routing Backus-Naur Form", draft-ietf-ccamp-attribute-bnf, work in progress.

## 6.2. Informative References

- [MVPN] E. Rosen, R. Aggarwal et al, "Multicast in MPLS/BGP IP VPNs", draft-ietf-l3vpn-2547bis-mcast-10.txt, work in progress.
- [RFC4761] Kompella, K., Ed., and Y. Rekhter, Ed., "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.
- [RFC5921] M. Bocci, S. Bryant, et al, "A Framework for MPLS in Transport Networks", RFC 5921, January 2007.
- [RFC5920] L. Fang, Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

- [RFC4379] K. Kompella, and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.
- [P2MP-OAM] S. Saxena, Ed., G. Swallow, Z. Ali, A. Farrel, S. Yasukawa, T. Nadeau, "Detecting Data Plane Failures in Point-to-Multipoint Multiprotocol Label Switching (MPLS) - Extensions to LSP Ping", draft-ietf-mpls-p2mp-lsp-ping-17.txt, work in progress.

#### Author's Addresses

Zafar Ali  
Cisco Systems, Inc.  
Email: zali@cisco.com

George Swallow  
Cisco Systems, Inc.  
Email: swallow@cisco.com

Rahul Aggarwal  
Juniper Networks  
rahul@juniper.net

Expires February 2012

[Page 11]

Internet Engineering Task Force  
Internet-Draft  
Intended status: Informational  
Expires: December 9, 2011

R. Martinotti  
D. Caviglia  
Ericsson  
N. Sprecher  
Nokia Siemens Networks  
A. D'Alessandro  
A. Capello  
Telecom Italia  
Y. Suemura  
NEC Corporation of America  
June 7, 2011

Interworking between MPLS-TP and IP/MPLS  
draft-martinotti-mpls-tp-interworking-02

Abstract

Purpose of this ID is to illustrate interworking scenarios between network(s) supporting MPLS-TP and network(s) supporting IP/MPLS. Main interworking aspects, issues and open points are highlighted.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 9, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
1.1. Scope of this document . . . . .	3
2. Conventions used in this document . . . . .	3
3. Acronyms . . . . .	3
4. Problem Statement . . . . .	4
5. Terminology . . . . .	4
6. Elements used in the figures . . . . .	5
7. Interconnectivity Options . . . . .	6
7.1. Network Layering model . . . . .	8
7.1.1. OAM Implication of the Layering model . . . . .	8
7.1.2. Layering model control plane consideration . . . . .	8
7.2. Network Layering scenarios . . . . .	8
7.2.1. Port based transparent transport of IP/MPLS . . . . .	8
7.2.2. VLAN based transparent transport of IP/MPLS . . . . .	12
7.2.3. Port based transport of IP/MPLS with Link Layer removal . . . . .	12
7.2.4. IP/MPLS / MPLS-TP hybrid edge node . . . . .	15
7.2.5. MPLS-TP carried over IP/MPLS . . . . .	18
7.3. Network Partitioning Model . . . . .	18
7.3.1. Connectivity constraints of the partitioning model . . . . .	18
7.3.2. OAM Implications of the partitioning model . . . . .	19
7.4. Network Partitioning scenarios . . . . .	19
7.4.1. Border Node - Multisegment Pseudowire . . . . .	20
7.4.2. Border Node - LSP stitching . . . . .	22
7.4.3. Border Link - Multisegment Pseudowire . . . . .	25
7.4.4. Border Link - LSP stitching . . . . .	27
8. Acknowledgements . . . . .	30
9. IANA Considerations . . . . .	30
10. Contributing Authors . . . . .	30
11. Security Considerations . . . . .	32
12. References . . . . .	32
12.1. Normative References . . . . .	32
12.2. Informative References . . . . .	32
Appendix A. Additional Stuff . . . . .	32
Authors' Addresses . . . . .	33

## 1. Introduction

### 1.1. Scope of this document

This document illustrates the most likely interworking scenarios between MPLS-TP and IP/MPLS. For each of the examined scenarios interworking aspects, limitations, issues and open points, with particular focus on OAM capabilities, are provided.

The main architectural construct considered in this document foresees PWE3 Protocol Stack Reference Model and MPLS Protocol Stack Reference Model. See [RFC 5921] for details.

## 2. 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 [RFC2119].

## 3. Acronyms

AC Attachment circuit  
CE Customer Edge  
CLI Client  
CP Control Plane  
DP Data Plane  
ETH Ethernet MAC Layer  
ETY Ethernet Physical Layer  
IWF Interworking Function  
LER Label Edge Router  
LSP Label Switched Path  
LSR Label Switch Router

MAC Media Access Control  
MEP Maintenance Association End Point  
MIP Maintenance Association Intermediate Point  
MP Management Plane  
MS-PW Multi Segment PW  
NE Network Element  
OAM Operations, Administration and Maintenance  
PE Provider Edge  
PHY Physical Layer  
PSN Packet Switched Network  
PW Pseudowire  
SRV Server  
SS-PW Single Segment PW  
S-PE Switching Provider Edge  
T-PE Terminating Provider Edge

#### 4. Problem Statement

This document addresses interworking issues between MPLS-TP network and IP/MPLS network. The network decomposition can envisage network layering and/or network partitioning.

The presented scenarios are not intended to be comprehensive, for instance more complex scenarios can be created composing those described in this document.

#### 5. Terminology

As far as this document is concerned, the following terminology is used:

- o IP/MPLS NE: a NE that supports IP/MPLS functions
- o IP/MPLS Network: a network in which IP/MPLS NEs are deployed
- o MPLS-TP NE: a NE that supports MPLS-TP functions
- o MPLS-TP Network: a network in which MPLS-TP NEs are deployed
- o Node: either MPLS-TP NE, IP/MPLS NE or CE
- o Ingress direction: from client to network
- o Egress direction: from network to client

For each of the scenarios described in this document, two paragraphs may appear, one related to possible issues already envisaged by the authors (Open Issues), the other related to aspects still left for further study and/or definition (Open Points).

This Section provides some terminology about network layering and partitioning. Primarily source of those definitions is [ITU-T G.805]. Readers already familiar with these concepts can skip this Section.

## 6. Elements used in the figures

A legenda of the symbols, which are most used in the following Sections, is provided, in order to facilitate comprehension of the scenarios.



```
Node:
----- Direct connection
- - - Virtual connection
..... one or more direct connections

Layers:
|      Termination
+      Connection
<->   Stitching

OAM:
> or < MEP
O      MIP
```

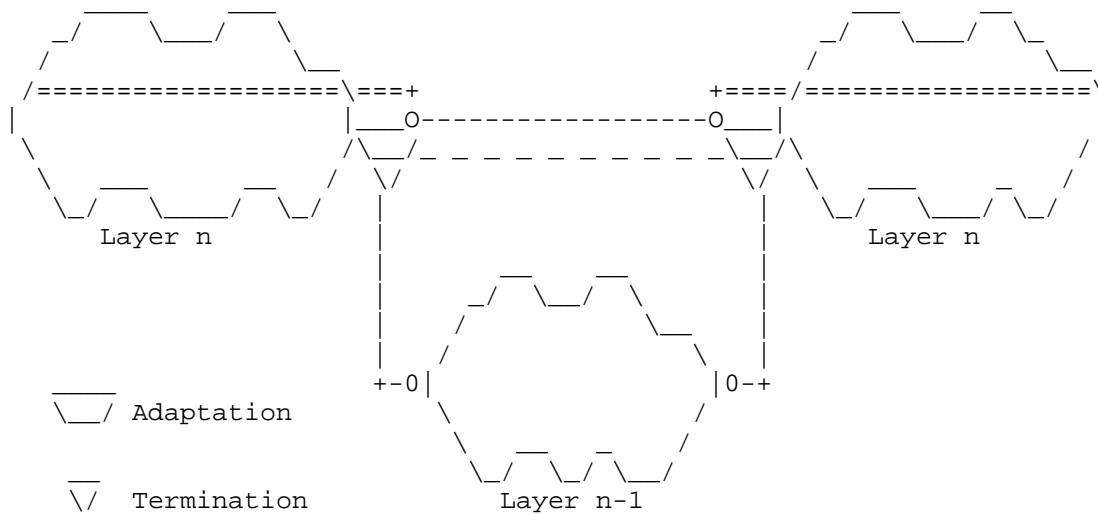
Figure 1

## 7. Interconnectivity Options

The MPLS-TP project adds dataplane OAM functionality to the MPLS tool set that permits executive action to be delegated to the dataplane. This provides the option of running MPLS without a control plane while still providing carrier grade resiliency options for connection oriented operation. Connection oriented operation alone does not offer the scalability to offer contemporary multipoint service solutions, but the combination of MPLS-TP connection oriented backhaul and IP/MPLS service capabilities permits the deployment of networks that scale significantly beyond the boundaries of current control plane scaling.

This section describes the methods in which IP/MPLS and MPLS-TP domains can interconnect. The network decomposition can envisage network layering and/or network partitioning. The presented scenarios are not intended to be comprehensive, for instance more complex scenarios can be created composing those described in this document. The various elements introduced in this section will be referred to in later sections.

The following figure illustrates the Network Layering concept, as it is described in Section 7.1:

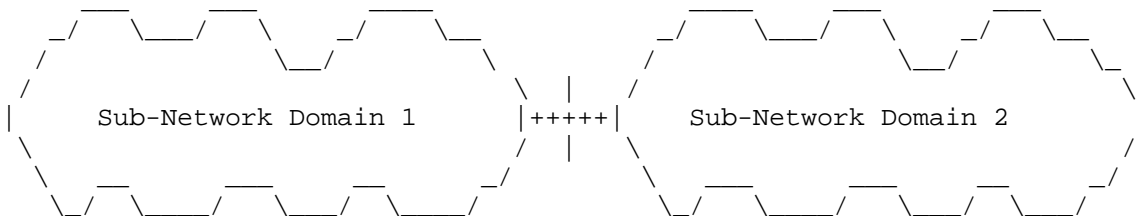


Network Layering

Figure 2

Layer n is carried over Layer n-1, via adaptation and termination functions. Some readers will also call this concept "Overlay model".

The following figure illustrates the Network Partitioning concept, as it is described in Section 7.3:



Network Partitioning

Figure 3

The boundary between the two subnetworks can be a link (as defined by [ITU-T G.805]), but also a Node, which in this case SHALL be able to

handle the technologies of both subnetworks.

The two subnetworks are at the same level. Some readers will also call this concept "Peer model".

#### 7.1. Network Layering model

Two relationship are considered: the IP/MPLS network is carried over the MPLS-TP one, the MPLS-TP network is carried over the IP/MPLS one. This version of the draft focuses on the former relationship. In the MPLS-TP architecture, the pseudo wire is the primary unit of carriage of non-MPLS-TP payloads. This provides a clean demarcation between MPLS-TP operations and transported payloads.

##### 7.1.1. OAM Implication of the Layering model

The overlay model has the virtue of uniform deployment of OAM capabilities and encapsulations at all MIPs and MEPs at a given layer in the label stack. The IP/MPLS architecture does include OAM transactions originated by MIPs so the layer interworking function for MPLS-TP servers is simplified.

##### 7.1.2. Layering model control plane consideration

The interworking between an IP/MPLS domain and an MPLS-TP domain highly depends on the implemented model (i.e. layering or partitioning) and different scenarios can be implemented depending on a number of different aspects.

In the case of layering model, the first aspect consists on the provisioning of the LSP at the N-1 layer (MPLS-TP layer). Two possible scenarios are foreseen: pre-configuration of the MPLS-TP LSP or induced provisioning. The pre-configuration of the MPLS-TP LSP can be performed either manually via NMS or via the MPLS-TP control plane signaling and the MPLS-TP LSP can be exported to the IP/MPLS domain as a forwarding adjacency. On the other side the signaling messages at the IP/MPLS layer, upon reaching the border of the MPLS-TP domain, can induce the signaling of the MPLS-TP LSP via RSVP-TE. Other use cases depend on how the IP/MPLS is carried over the MPLS-TP domain and are analyzed scenario by scenario in the following sections.

#### 7.2. Network Layering scenarios

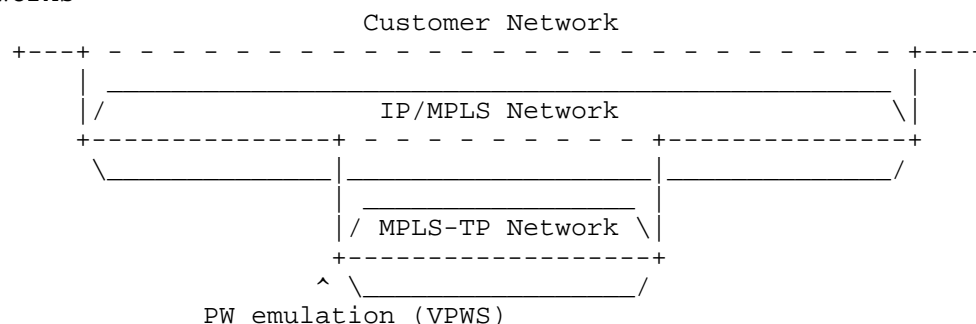
##### 7.2.1. Port based transparent transport of IP/MPLS

This scenario foresees an IP/MPLS network carried over an MPLS-TP network. The selection of the route over the MPLS-TP network is done

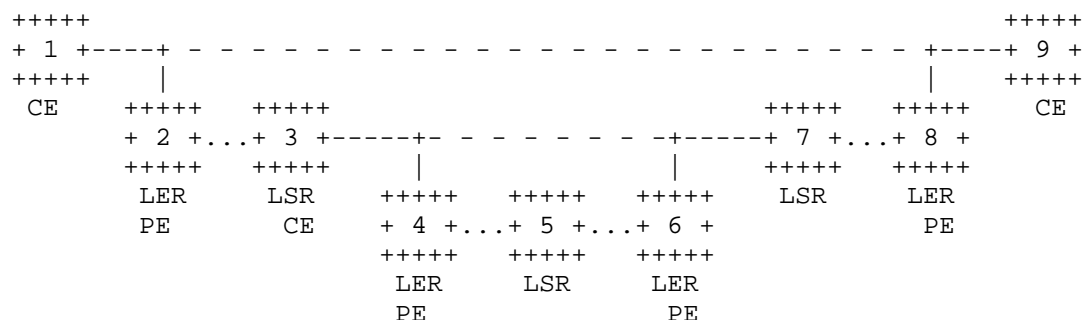
on a per port basis. The interworking is done via Link Layer (e.g. Ethernet) encapsulation in PW over MPLS-TP (as per PWE3 Protocol Stack Reference Model). MPLS-TP LSPs are pre-configured with respect to IP/MPLS LSPs and IP/MPLS LSRs may be seen one hop away.

The following figure illustrates the functional interworking among the networks:

Networks:



Nodes:



Port based transparent transport - Networks view

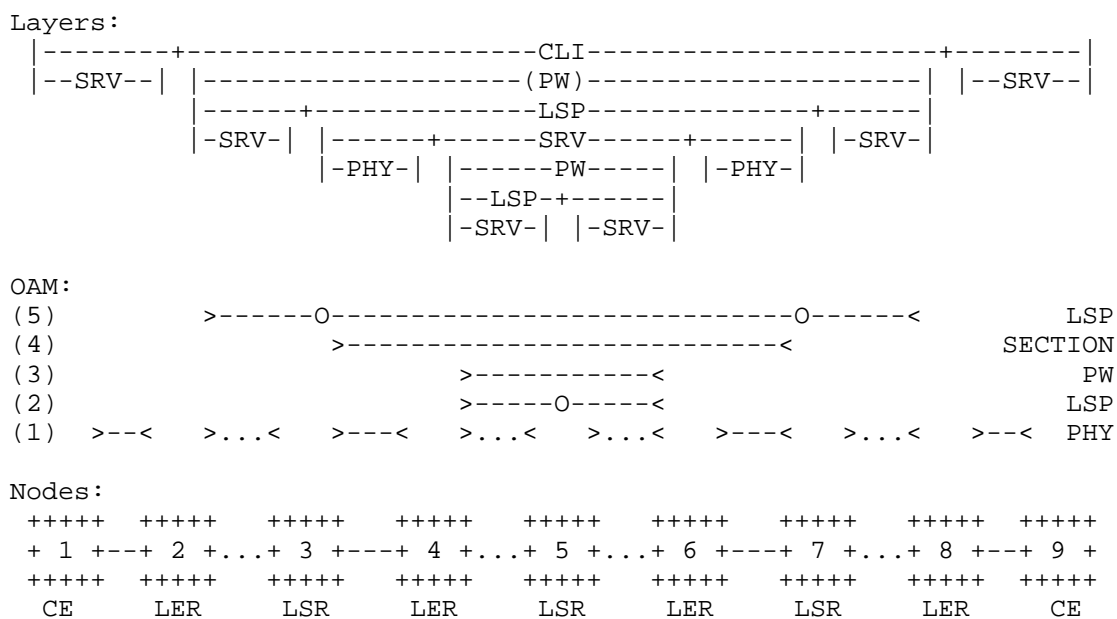
Figure 4

The LSR 3 and 7 are one hop away from the IP/MPLS layer point of view, CP/MP of IP/MPLS is transparently transported by MPLS-TP network.

In case the Link Layer is Ethernet, the service provided by the MPLS-TP network could be an E-Line service realized via VPWS. The LER4 and 6 do not need to know that above the Ethernet layer there is an MPLS LSP.

## 7.2.1.1. OAM Considerations

The following figure illustrates the stacking relationship among the technology layers and OAM relationship among the networks:



Port based transparent transport - Layers and OAM view

Figure 5

Several levels of OAM are shown in the previous figure, these are not comprehensive and any subset of them MAY be configured in a network. A brief description of the different levels is provided:

- (5) Edge-to-Edge MPLS OAM on IP/MPLS network (at LSP level)
- (4) Section OAM on IP/MPLS network
- (3) Edge-to-Edge MPLS-TP OAM on MPLS-TP network (at PW level)
- (2) Edge-to-Edge MPLS-TP OAM on MPLS-TP network (at LSP level)

(1) Physical level OAM (MAY be of several kinds)

In case of fault detected at the MPLS-TP LSP (2) level, the corresponding server MEP asserts a signal fail condition and notifies that to the co-located MPLS-TP client/server adaptation function which then generates OAM packets with AIS information in the downstream direction to allow the suppression of secondary alarms at the MPLS-TP MEP in the client (sub-) layer, which in this example correspond to PW layer (3).

Note that the OAM layers not directly related to MPLS-TP network have been reported just for completeness of the scenario; however their behavior and interworking are out of scope of this document. For MPLS-TP Alarm reporting detailed description, please refer to [draft-ietf-mpls-tp-oam-framework].

#### 7.2.1.2. Control Plane considerations

In this case the interconnection between the IP/MPLS domain and the MPLS-TP domain consists of a link. This does not allow a transparent transport of the IP control messages (e.g. LDP) over the MPLS-TP LSPs due to the fact that the egress node of the MPLS-TP domain is not able to route IP packets on its interfaces. The IP control messages need to be carried over an Ethernet frame over a PWE3 before being injected into the MPLS-TP LSP. In other words they are forwarded with two labels, the PWE3 one (S=0) and the LSP one (S=1). The IP control message, upon reaching the egress LER of the MPLS-TP domain, can be correctly forwarded to the ingress node of the IP/MPLS domain.

#### 7.2.1.3. Services view

There are two service models supported by the overlay model when combined with Ethernet PWs. The first is simple p2p encapsulation and transport of all traffic presented to the MPLS-TP on a given interface. This is of limited utility due to the number of ports required to achieve the desired level of network interconnect across the MPLS-TP core.

The second is that the MPLS-TP LER maps VLANs to distinct PWs such that multiple IP/MPLS adjacencies can be supported over each interface between the IP/MPLS LSR and the MPLS-TP LER. This potentially can require a large number of IP/MPLS adjacencies overlaying the core.

In both cases the service can be unprotected or protected.

#### 7.2.1.4. Resiliency considerations

In the scenario where the service is unprotected, resiliency is fully delegated to the IP/MPLS network, which will depend on a combination of routing convergence and/or FRR to maintain service. This will be at the expense of routing stability.

A protected service can offer significant improvements in routing stability with the exception that the link between the IP/MPLS LSRs and the MPLS-TP LERs and the MPLS-TP LERs themselves are single points of failure. There is an advantage in that the single points of failures are adjacent to the MPLS LSRs such that there is a high probability of such failures manifesting themselves immediately in the form of a physical layer loss-of-signal failure and thus accelerating recovery. Multiple failure scenarios may also result in the IP/MPLS overlay having to take action to recover connectivity but this would be gated by whatever OAM detection mechanisms were employed by the IP/MPLS layer as there is no equivalent of MPLS-TP LDI across the interconnect interface.

#### 7.2.2. VLAN based transparent transport of IP/MPLS

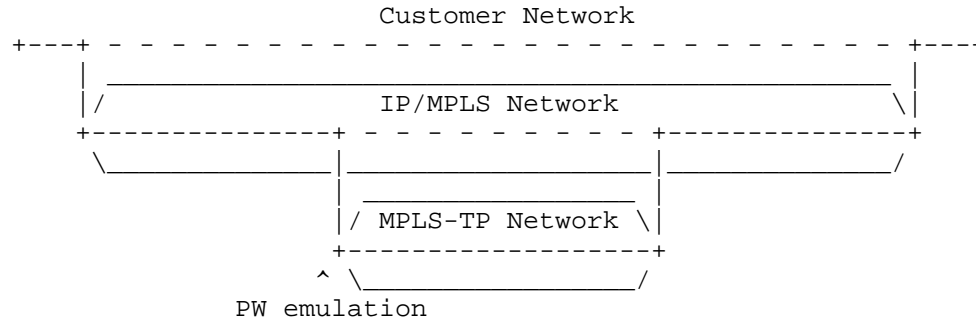
This scenario is analogous to the previous one. The interconnection between the IP/MPLS LSRs and the MPLS-TP PEs is done via .1Q Tagged Ethernet, and VLANs are used to select the routes over the p2p Ethernet connectivity services over MPLS-TP (VPWS). The interworking is done via Ethernet encapsulation in PW over MPLS-TP (as per PWE3 Protocol Stack Reference Model). This VLAN based interconnection may be used in order to reduce the number of physical interfaces between the two networks. The same considerations of previous scenarios apply.

#### 7.2.3. Port based transport of IP/MPLS with Link Layer removal

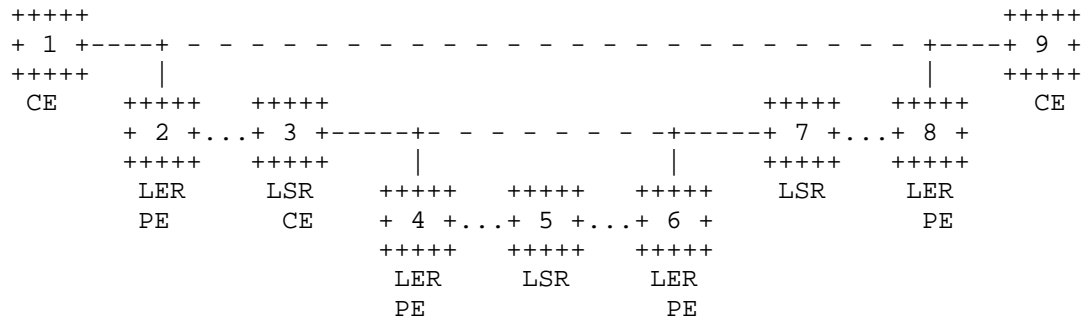
This scenario foresees an IP/MPLS network carried over an MPLS-TP network. The selection of the route over the MPLS-TP network is done on a per port basis. The physical interface between the IP/MPLS and the MPLS-TP network may be of different kind (e.g. Ethernet, POS); the interworking is done via Link Layer removal and client packet (MPLS and IP) encapsulation in PW over MPLS-TP (as per PWE3 Protocol Stack Reference Model). MPLS-TP LSPs are pre-configured with respect to IP/MPLS LSPs and are seen as routing adjacencies by the IP/MPLS network.

The following figure illustrates the functional interworking among the networks:

Networks:



Nodes:



Port based transport with Link Layer removal - Networks view

Figure 6

The LSR 3 and 7 are one hop away from the IP/MPLS layer point of view.

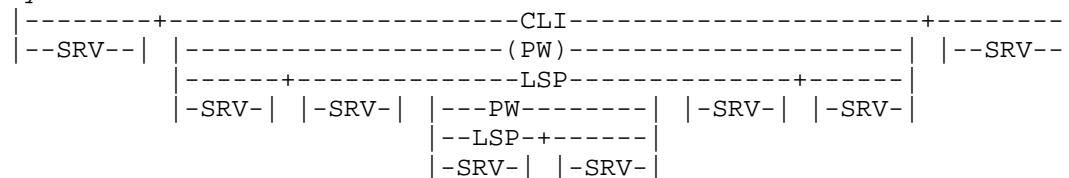
The service provided by the MPLS-TP network is p2p; client traffic is separated on a per port basis, so that (for example) all traffic coming from LSR 3 on the interface to LER 4 is transparently transported via LER 6 to LSR 7 and viceversa. The client traffic to be encapsulated is both MPLS packets (DP) and IP packets (DP, CP and MP). The encapsulation may be performed via PWs, that is, one PW is needed for MPLS and one for IP between any given port pair or directly using the LSP label stacking. The encapsulation via PW is required such that the IP/MPLS section preserves PHY like properties and to operationally isolate TP and IP/MPLS operation (e.g. reserved label handling link GAL and Router Alert).



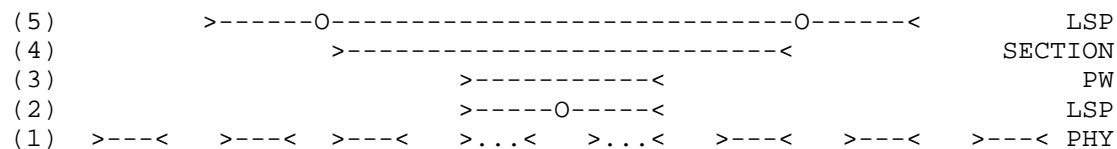
#### 7.2.3.1. OAM considerations

The following figure illustrates the stacking relationship among the technology layers and OAM relationship among the networks:

Layers:



OAM:



Nodes :



## Port based transport with Link Layer removal - Layers and OAM view

Figure 7

Several levels of OAM are possible, a subset of them is shown in the previous figure, however these are not comprehensive, any subset of them MAY be configured in a network. A brief description of the levels is provided:

- (5) Edge-to-Edge MPLS OAM on IP/MPLS network (at LSP level)
- (4) Section MPLS OAM
- (3) Edge-to-Edge MPLS-TP OAM on MPLS-TP network (at PW level)
- (2) Edge-to-Edge MPLS-TP OAM on MPLS-TP network (at LSP level)

(1) Physical level OAM (MAY be of several kinds)

#### 7.2.3.2. Control Plane considerations

In the case of transparent transport of the IP/MPLS over the MPLS-TP domain there are no differences, from a control plane point of view, with respect to the case of Ethernet encapsulation over MPLS-TP. Same considerations carried out in section 5.1.3.1.1.2 apply to this section.

#### 7.2.3.3. Services view

The service model for the transparent transport mode is simple p2p encapsulation and transport of all traffic presented to the MPLS-TP on a given interface. This is of limited utility due to the number of ports required to achieve the desired level of network interconnect across the MPLS-TP core. It would potentially also require a correspondingly high number of IP/MPLS adjacencies to overlay the core.

The service can be unprotected or protected.

#### 7.2.3.4. Resiliency considerations

The resiliency considerations are the same as for the overlay model.

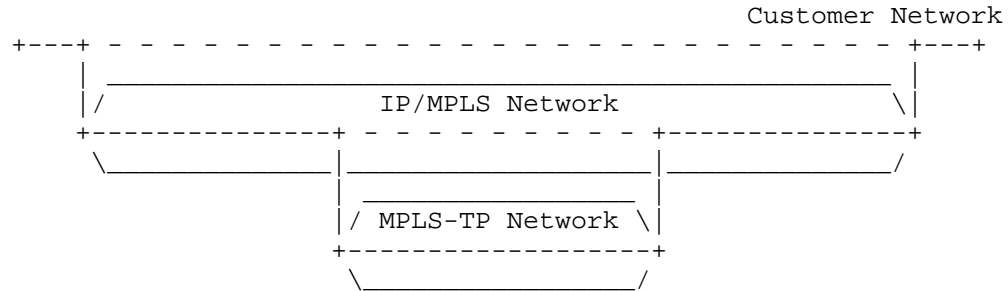
#### 7.2.4. IP/MPLS / MPLS-TP hybrid edge node

In this scenario the physical interface between the IP/MPLS and the MPLS-TP network is generic and may be other than Ethernet (e.g. POS); the interworking is done via client LSP packet encapsulation as per MPLS labeled or IP traffic over MPLS-TP as per RFC 5921. MPLS-TP LSPs are pre-configured with respect to IP/MPLS LSPs and are seen as routing adjacencies between the hybrid edge nodes by the IP/MPLS network.

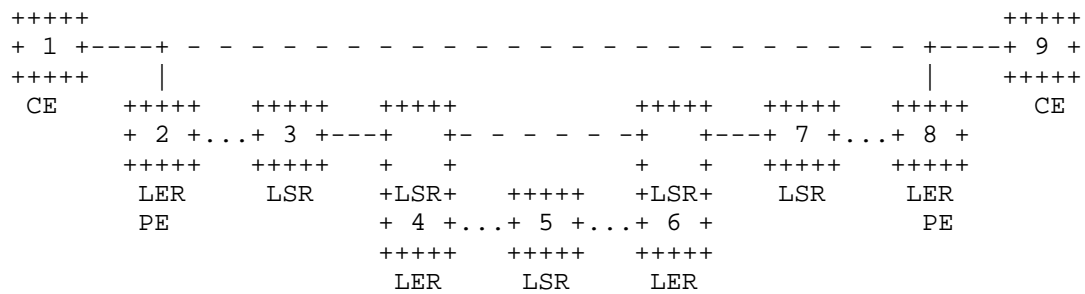
The service that is offered to the IP/MPLS network is that of a multi-point MPLS VPN.

The following figure illustrates the functional interworking among the networks.

Networks:



Nodes:



IP/MPLS encapsulation over MPLS-TP - Networks view

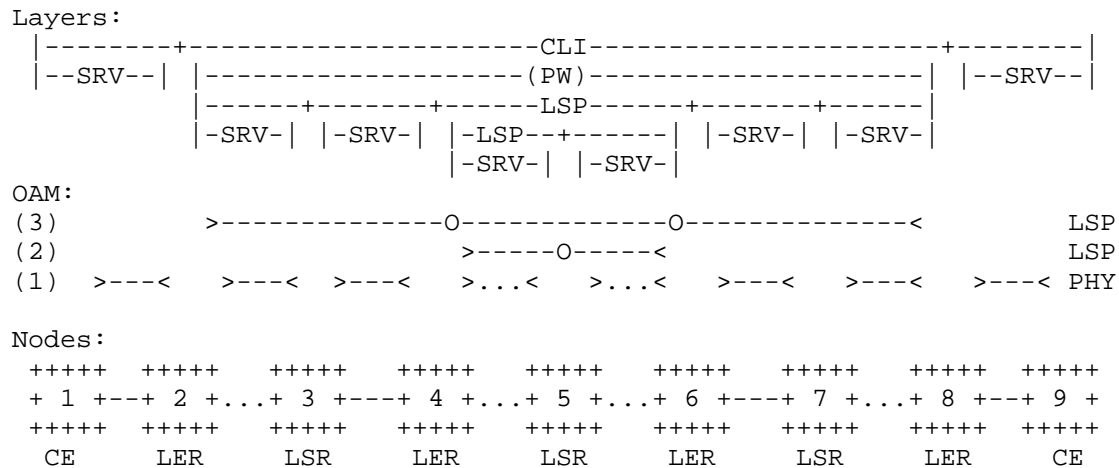
Figure 8

The Node 4 and 6 in the above figure act as dual function:

- o LSR of client IP/MPLS network
- o LER of server MPLS-TP subnetwork

#### 7.2.4.1. OAM considerations

The following figure illustrates the stacking relationship among the technology layers and OAM relationship among the networks:



IP/MPLS encapsulation over MPLS-TP - Layers and OAM view

Figure 9

Several levels of OAM are possible, a subset of them is shown in the previous figure, however these are not comprehensive, any subset of them MAY be configured in a network. A brief description of the levels is provided:

- (3) Edge-to-Edge MPLS OAM on IP/MPLS network (at LSP level)
- (2) Edge-to-Edge MPLS-TP OAM on MPLS-TP network (at LSP level)
- (1) Physical level OAM (MAY be of several kinds)

#### 7.2.4.2. Control Plane considerations

This case is different from the previous two because the interconnection between the IP/MPLS domain and the MPLS-TP domain consists of a node. This lead to the fact that IP control messages do not need to be carried over a PWE3 along the MPLS-TP domain but can be directly carried over an LSP. In other words they are forwarded with a single LSP label (S=1) and , upon reaching the hybrid node between the MPLS-TP domain and the next IP/MPLS domain , the signaling can be carried on.

#### 7.2.4.3. Services view

The service model for the hybrid edge node model is that the MPLS-TP network appears to the IP/MPLS network as a complete IP/MPLS

subnetwork. This has the virtue of collapsing the number of IP/MPLS adjacencies required to overlay the core.

The service can be unprotected or protected. And the protection can be a combination of MPLS-TP resiliency and IP/MPLS recovery actions.

#### 7.2.4.4. Resiliency considerations

The resiliency considerations are similar to that of the overlay model. However the extension of the control plane to the hybrid node means the lack of a dataplane LDI equivalent is mitigated, the IP/MPLS domain having been extended to reach the MPLS-TP OAM domain such that LDI indications from core failures can interwork directly the the control plane and accelerate recovery actions.

#### 7.2.5. MPLS-TP carried over IP/MPLS

TODO

### 7.3. Network Partitioning Model

In the rest of this Section the following assumptions apply:

- o Customer network is carried partly over IP/MPLS subnetwork (e.g. via PW encapsulation) and partly over MPLS-TP subnetwork.
- o An example of server layer of MPLS is Ethernet.

For the purposes of this Section, MPLS-TP subnetwork is deployed between a CE and an IP/MPLS subnetwork. Other kinds of deployment are possible (not shown in this document), for instance:

- o More than two subnetworks are deployed between the CEs
- o MPLS-TP can be deployed between two subnetworks

#### 7.3.1. Connectivity constraints of the partitioning model

The partitioning model is constrained to interconnecting LSPs or PWs with common behavioral characteristics. As MPLS-TP is constrained to connection oriented behavior the portion of the LSP that transits an IP/MPLS subnetwork will need to be effectively constrained to the same profile, that is connection oriented, and no PHP or merging. No ECMP or transit of LAG cannot be guaranteed which means OAM fate sharing may not exist in IP/MPLS subnetworks and the end-to-end OAM may only serve to coordinate dataplane resiliency actions between MEPs with respect to faults in the MPLS-TP subnetworks.

### 7.3.2. OAM Implications of the partitioning model

The partitioning model requires the concatenations of path segments that do not necessarily have common OAM components and have a number of possible implementations. At the simplest level configuration of common OAM capabilities and encapsulation between the MEPs in the MEG is required. The set that is common to the MEPs in the MEG may not necessarily be supported by the MIPs, and knowledge of MIP capability will not figure into MEP negotiation, so the MEPs may select a common mode that is not common with that supported by the MIPs.

The primary consequence being that MPLS-TP MIP originated transactions, or messages targeted to MIPs using MPLS-TP encapsulations will not be guaranteed to provide a uniform quality of information as not all MIPs will support MPLS-TP OAM extensions, and as noted will not participate in MEP-MEP configuration or negotiation.

This means that GAL encapsulated OAM may only serve to coordinate dataplane resiliency actions between MEPs with respect to faults in the MPLS-TP subnetworks and faults in the IP/MPLS subnetwork are recovered by IP/MPLS mechanisms (e.g. FRR). Edge to edge monitoring of MPLS/MPLS-TP networks may be implemented using an edge to edge LSP OAM/PW OAM, in order not to need a gateway/translation function on the border node between the two domains.

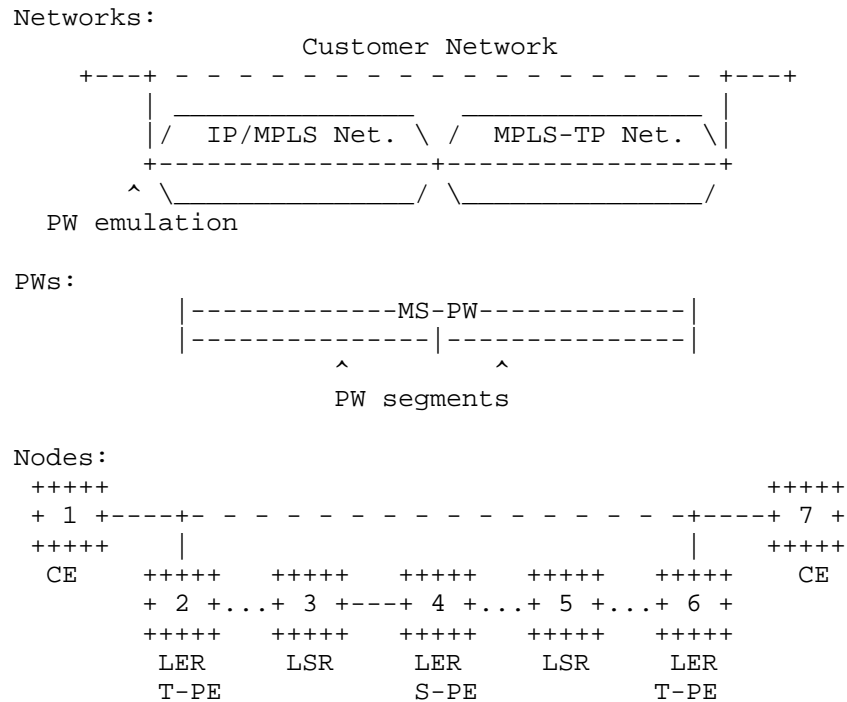
### 7.4. Network Partitioning scenarios

The main features to be taken into account in deploying a partitioned network are the following:

- o Border Node or Border Link
- o MultiSegment Pseudowire or LSP Stitching
- o Network Interworking
- o End-to-End OAM support
- o Interaction between DP of IP/MPLS and DP of MPLS-TP
- o Interaction between CP of IP/MPLS and MP of MPLS-TP
- o Interaction between CP of IP/MPLS and CP of MPLS-TP
- o Interaction between MP of IP/MPLS and MP of MPLS-TP

## 7.4.1. Border Node - Multisegment Pseudowire

The following figure illustrates the functional interworking among the networks:

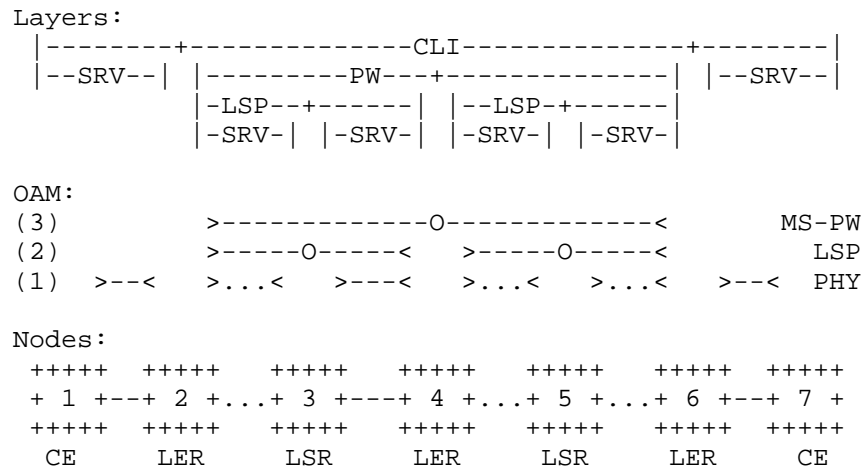


Border Node - Multisegment Pseudowire - Networks and PWs view

Figure 10

## 7.4.1.1. OAM considerations

The following figure illustrates the stacking relationship among the technology layers and OAM relationship among the networks:



Border Node - Multisegment Pseudowire - Layers and OAM view

Figure 11

Several levels of OAM are possible, a subset of them is shown in the previous figure, however these are not comprehensive, any subset of them MAY be configured in a network. A brief description of the different levels is provided:

- (3) Edge-to-Edge MPLS/MPLS-TP OAM on whole network (at PW level)
- (2) Edge-to-Edge MPLS OAM and Edge-to-Edge MPLS-TP OAM on each network partition respectively (at LSP level)
- (1) Physical level OAM (MAY be of several kind)

#### Open Points:

- o Interworking between LSP OAM (2) and MS-PW OAM (3) is still to be cleared/defined
- o Edge-to-Edge MS-PW OAM (3) must be configured on different subnetworks

#### 7.4.1.2. Control Plane considerations

TODO



#### 7.4.1.3. Services view

The generalized service model for all partitioning models is a p2p connection for the PW client.

#### 7.4.1.4. Resiliency considerations

The PW can be configured to be protected or unprotected at the PW layer. If it is unprotected it is dependent on the underlying domains (MPLS-TP or IP/MPLS) resiliency mechanisms to offer subnetwork protection, but the S-PE is a single point of failure. A protected PW can be set up such that the working and protection PWs traverse physically diverse S-PEs.

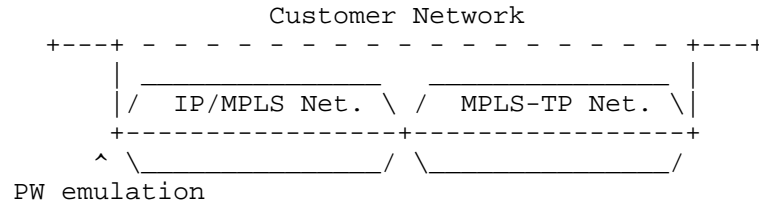
Implementing E2E protection at the PW layer requires CC flows on the PW which for large numbers of PWs may have scaling implications.

When the PW is protected, the border node as an MS-PW stitching point permits the interworking of MPLS-TP fault indications with the PW signaling in the IP/MPLS domain such that fast E2E protection switching can be coordinated without requiring fast CC/CV OAM flows in the PW layer.

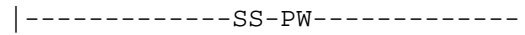
#### 7.4.2. Border Node - LSP stitching

The following figure illustrates the functional interworking among the networks:

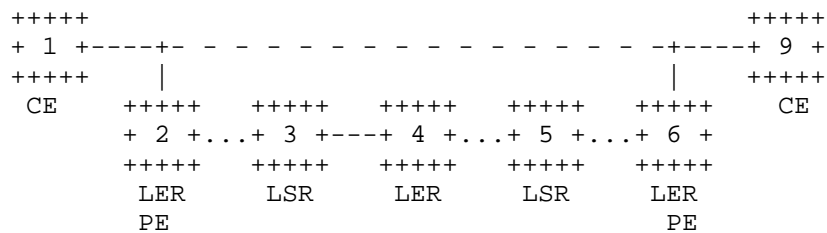
Networks:



PWs:



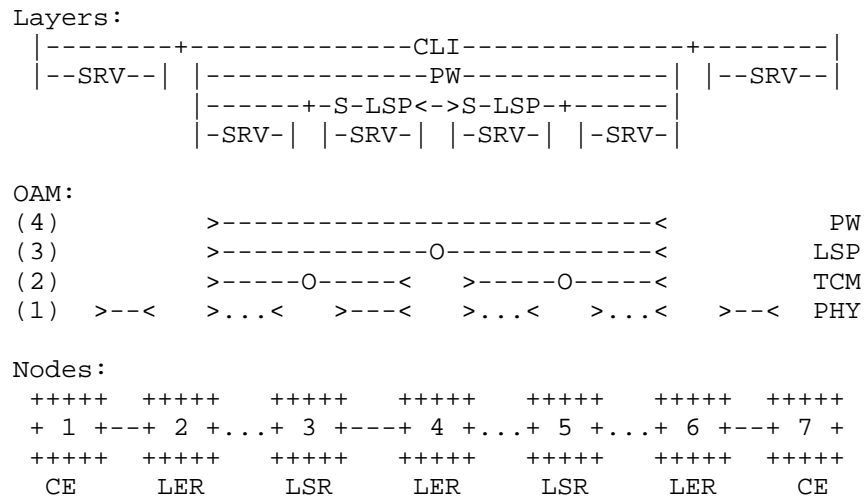
Nodes:



Border Node - LSP stitching - Networks and PWs view

Figure 12

The following figure illustrates the stacking relationship among the technology layers and OAM relationship among the networks:



Border Node - LSP stitching - Layers and OAM view

Figure 13

Note: in this case a SS-PW extends over the subnetworks as the stitched LSP does. TCM can be used to monitor the LSP segments.

Several levels of OAM are possible, a subset of them is shown in the previous figure, however these are not comprehensive, any subset of them MAY be configured in a network. A brief description of the different levels is provided:

- (4) Edge-to-Edge MPLS/MPLS-TP OAM on whole network (at PW level)
- (3) Edge-to-Edge MPLS/MPLS-TP OAM on whole network (at LSP level)
- (2) Edge-to-Edge MPLS OAM and Edge-toEdge MPLS-TP OAM on each network partition respectively (at TCM level)
- (1) Physical level OAM (MAY be of several kind)

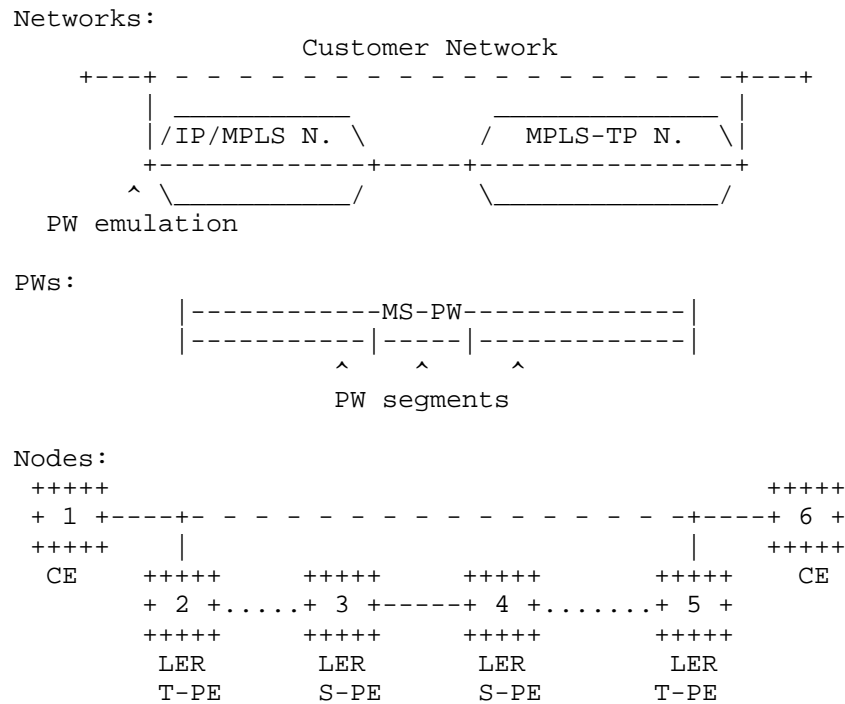
#### Open Points:

- o Edge-to-Edge LSP OAM (3) must be configured on different subnetworks
- o Edge-to-Edge PW OAM (4) must be configured on different subnetworks

- o Interworking between TCM OAM (2) and LSP OAM (3) is still to be cleared/defined

#### 7.4.3. Border Link - Multisegment Pseudowire

The following figure illustrates the functional interworking among the networks:

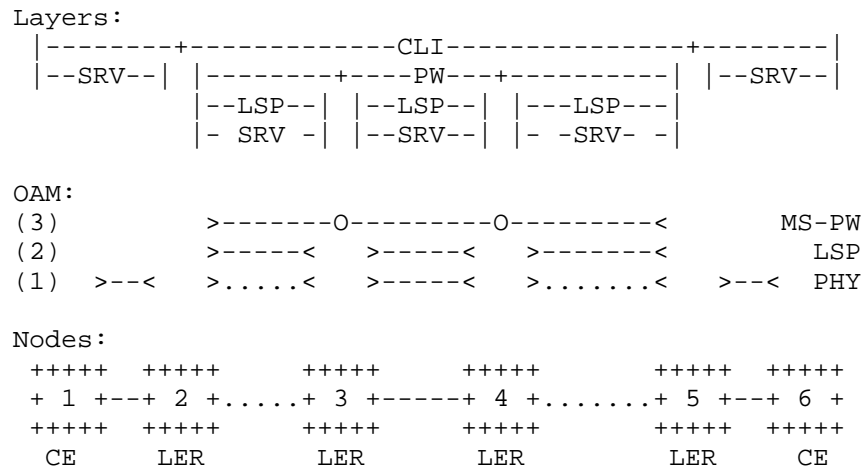


Border Link - Multisegment Pseudowire - Networks view

Figure 14

##### 7.4.3.1. OAM considerations

The following figure illustrates the stacking relationship among the technology layers and OAM relationship among the networks:



Border Link - Multisegment Pseudowire - Layers and OAM view

Figure 15

Several levels of OAM are possible, a subset of them is shown in the previous figure, however these are not comprehensive, any subset of them MAY be configured in a network. A brief description of the different levels is provided:

- (3) Edge-to-Edge MPLS/MPLS-TP OAM on whole network (at PW level)
- (2) Edge-to-Edge MPLS OAM, Border MPLS OAM and Edge-toEdge MPLS-TP OAM on each network partition respectively (at LSP level)
- (1) Physical level OAM (MAY be of several kinds)

#### Open Points:

- o Interworking between LSP OAM (2) and MS-PW OAM (3) is still to be cleared/defined
- o LSP between Node 3 and 4 could be avoided, however in this case PW over Ethernet should be specified.
- o Edge-to-Edge MS-PW OAM (3) must be configured on different subnetworks

#### 7.4.3.2. Control Plane considerations

TODO

#### 7.4.3.3. Services view

The generalized service model for all partitioning models is a p2p connection for the PW client.

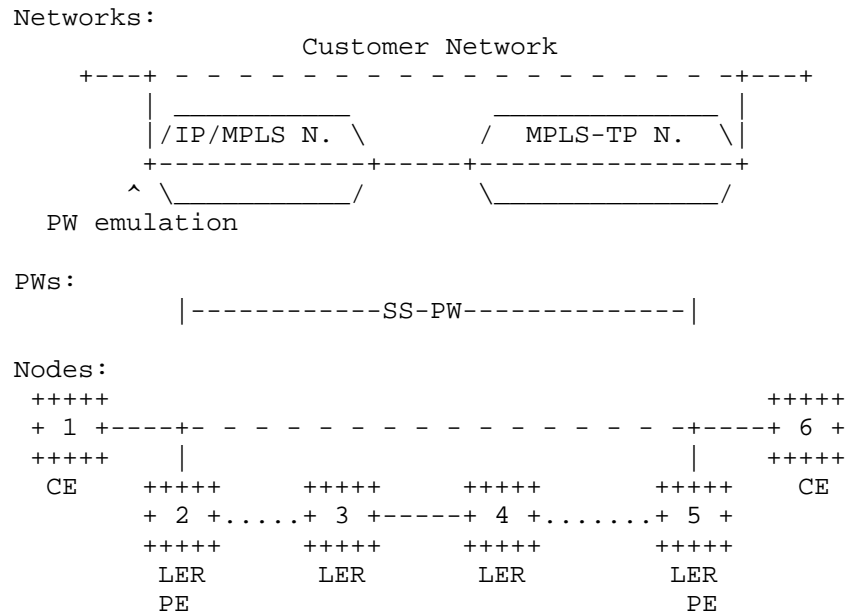
#### 7.4.3.4. Resiliency considerations

The PW can be configured to be protected or unprotected at the PW layer. If it is unprotected it is dependent on the underlying domains (MPLS-TP or IP/MPLS) resiliency mechanisms to offer subnetwork protection, but the border S-PEs and border link are all single points of failure. A protected PW can be set up such that the working and protection PWs traverse physically diverse border links.

Implementing E2E protection at the PW layer requires CC flows on the PW which for largenumbers of PWs may have scaling implications.

#### 7.4.4. Border Link - LSP stitching

The following figure illustrates the functional interworking among the networks:

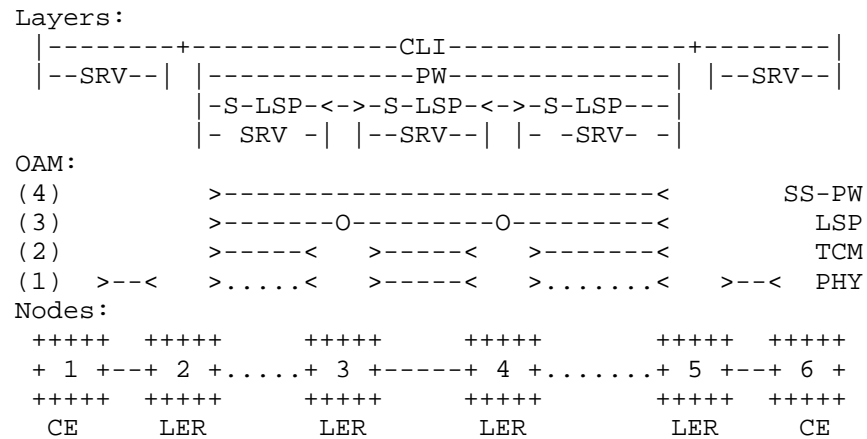


Border Link - LSP stitching - Networks view

Figure 16

#### 7.4.4.1. OAM considerations

The following figure illustrates the stacking relationship among the technology layers and OAM relationship among the networks:



Border Link - LSP stitching - Layers and OAM view

Figure 17

Note: in this case a SS-PW extends over the subnetworks as the stitched LSP does. TCM can be used to monitor the LSP segments.

Several levels of OAM are possible, a subset of them is shown in the previous figure, however these are not comprehensive, any subset of them MAY be configured in a network. A brief description of the different levels is provided:

- (4) Edge-to-Edge MPLS/MPLS-TP OAM on whole network (at PW level)
- (3) Edge-to-Edge MPLS/MPLS-TP OAM on whole network (at LSP level)
- (2) Edge-to-Edge MPLS OAM, Border MPLS OAM and Edge-to-Edge MPLS-TP OAM on each network partition respectively (at TCM level)
- (1) Physical level OAM (MAY be of several kinds)

#### 7.4.4.2. Control Plane considerations

TODO

#### 7.4.4.3. Services view

The generalized service model for all partitioning models is a p2p connection for the PW client.



#### 7.4.4.4. Resiliency considerations

The LSP can be configured to be protected end to end, have subnetwork protection or be unprotected at the LSP layer. In the subnetwork protection scenario the border S-PEs and the borderlink are all single points of failure.

When GAL/GACH encapsulated OAM is deployed at (a minimum) of the LSP MEPs, it is possible to envision interworking of the MPLS-TP LSP and LSPs in the IP/MPLS domain set up with RSVP-TE and/or with LDP. In the latter case the MPLS-TP LSP maps to a FEC rather than a specific LSP but the MPLS\_TP LSP would need to appear as a FEC in LDP with associated scaling impacts.

Open Points:

- o Edge-to-Edge LSP OAM (3) must be configured on different subnetworks
- o Edge-to-Edge PW OAM (4) must be configured on different subnetworks
- o Interworking between TCM OAM (2) and LSP OAM (3) is still to be cleared/defined
- o Interaction between IP/MPLS and MPLS-TP CPs is still to be cleared/defined

#### 8. Acknowledgements

The authors gratefully acknowledge the input of Attila Takacs.

#### 9. IANA Considerations

This memo includes no request to IANA.

#### 10. Contributing Authors

David Allan  
Ericsson  
Holger Way  
San Jose  
U.S.

Email: david.i.allan@ericsson.com

Elisa Bellagamba  
Ericsson  
Torshamnsgatan 48  
Stockholm 164 80  
Sweden

Email: elisa.bellagamba@ericsson.com

Daniele Ceccarelli  
Ericsson  
Via A. Negrone 1/A  
Genova - Sestri Ponente 16153  
Italy

Email: daniele.ceccarelli@ericsson.com

David Saccon  
Ericsson  
Holger Way  
San Jose  
U.S.

Email: david.sacson@ericsson.com

John Volkering  
Ericsson

Email: john.volkering@ericsson.com

## 11. Security Considerations

This document does not introduce any additional security aspects beyond those applicable to PWE3 and MPLS.

## 12. References

### 12.1. Normative References

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

### 12.2. Informative References

- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
- [RFC5654] Niven-Jenkins, B., Brungard, D., Betts, M., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, September 2009.
- [RFC5921] Bocci, M., Bryant, S., Frost, D., Levrau, L., and L. Berger, "A Framework for MPLS in Transport Networks", RFC 5921, July 2010.
- [RFC5860] Vigoureux, M., Ward, D., and M. Betts, "Requirements for Operations, Administration, and Maintenance (OAM) in MPLS Transport Networks", RFC 5860, May 2010.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [ITU-T G.805] "Generic functional architecture of transport networks", ID ITU-T G.805, March 2000.

## Appendix A. Additional Stuff

This becomes an Appendix.

Authors' Addresses

Riccardo Martinotti  
Ericsson  
Via A. Negrone 1/A  
Genova - Sestri Ponente 16153  
Italy  
  
Email: riccardo.martinotti@ericsson.com

Diego Caviglia  
Ericsson  
Via A. Negrone 1/A  
Genova - Sestri Ponente 16153  
Italy  
  
Email: diego.caviglia@ericsson.com

Nurit Sprecher  
Nokia Siemens Networks  
3 Hanagar St. Neve Ne'eman B  
Hod Hasharon 45241  
Israel  
  
Email: nurit.sprecher@nsn.com

Alessandro D'Alessandro  
Telecom Italia  
Via Reiss Romoli, 274  
Torino 10148  
Italy  
  
Email: alessandro.dalessandro@telecomitalia.it

Alessandro Capello  
Telecom Italia  
Via Reiss Romoli, 274  
Torino 10148  
Italy  
  
Email: alessandro.capello@telecomitalia.it

Yoshihiko Suemura  
NEC Corporation of America  
14040 Park Center Road  
Herndon, VA 20171  
USA

Email: Yoshihiko.Suemura@necam.com



Network Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: March 5, 2013

P. Dutta  
Alcatel-Lucent  
G. Heron  
Cisco Systems  
T. Nadeau  
Juniper Networks  
September 01, 2012

Targeted LDP Hello Reduction  
draft-pdutta-mppls-tldp-hello-reduce-04

Abstract

Targeted LDP (t-LDP) Hellos are used for establishing adjacencies with non-directly connected peers. After an LDP session is established to a Targeted Peer, there are deployment scenerios where it is not necessary to send Targeted LDP Hellos at the configured intervals. This document proposes a mechanism to turn off or reduce the rate of exchange of Targeted LDP Hellos after LDP session is established to a peer.

Requirements Language

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

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on March 5, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the

document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
2. Terminology . . . . .	4
3. Targeted LDP Hello Reduction Procedure . . . . .	4
4. IANA Considerations . . . . .	6
5. Security Considerations . . . . .	6
6. Operational Considerations . . . . .	6
7. Acknowledgements . . . . .	7
8. References . . . . .	7
8.1. Normative References . . . . .	7
8.2. Informative References . . . . .	7
Authors' Addresses . . . . .	8



## 1. Introduction

LDP Hello messages are exchanged as part of the LDP discovery mechanism [RFC5036]. There are two types of LDP discovery mechanism described in [RFC5036]- Basic Discovery and Extended Discovery.

To engage in LDP Basic Discovery on an interface, an LSR periodically sends LDP Link Hellos out the interface to the well-known LDP discovery port for the "all routers on this subnet" group multicast address. Receipt of an LDP Link Hello on an interface, identifies a hello adjacency with a potential LDP peer reachable at the link level on the interface. Thus an LSR may establish hello adjacency with multiple peers discovered over a single interface and must continue to transmit hellos at regular intervals even after hello adjacency is established to a peer.

Extended discovery is used to support LDP sessions between non-directly connected LSRs. An LDP Targeted Hello is sent to a specific address rather than to the "all routers" group multicast address for the ongoing interface. Receipt of a LDP Targeted Hello identifies a hello adjacency with a potential LDP peer at network level.

In Extended discovery there can be only one Targeted Hello Adjacency between two peers. Note that throughout this document "peer" means the LDP LSR designated by a unique LDP Identifier. Once the LDP session is operational between two targeted LDP peers, periodic session Keepalives are used to maintain the LDP session. There are certain deployment scenarios where after the session is operational the periodic Targeted Hellos between the LSRs become redundant, as session Keepalives in turn serves the intent of each LSR to maintain its adjacency to its peer. Moreover additional mechanisms such as centralized BFD [RFC5880] may be used to track liveness of ldp sessions.

When an LSR maintains a large number of LDP sessions (thousands) to Targeted peers, it is an additional burden to send and receive Targeted Hellos for all peers at periodic intervals. In MPLS deployments at access or mobility backhaul or in Seamless MPLS [I-D.ietf-mpls-seamless-mpls] , there can be very large volume of LDP sessions (e.g 10,000) with targeted LDP adjacencies to each base station (or last mile in a MPLS domain).

Another problem with targeted hello adjacency arises is Denial Of Service (DoS) attacks. It is possible that existing hello adjacencies can get lost due to DoS attack on LDP Hello receiver by spurious hello packets. Unlike TCP sessions it is not always possible to provide per peer protection for UDP based hellos. Implementations can use methods to protect existing adjacencies while

throttling spurious adjacencies but such methods may not be available in low cost MPLS devices deployed in access. So it is important to avoid dependency on Targeted LDP hellos on t-LDP adjacency maintenance as far as possible. Reduction of Hellos provide probabilistically better resilience on maintenance of hello adjacencies during sporadic hello attacks.

This document proposes an OPTIONAL mechanism to turn off Targeted LDP Hellos after a LDP session is established to a targeted peer, without changes in the procedures defined in [RFC5036]. The solutions described in this document may not be applicable in scenerios where Session Keepalives or BFD may not act as substitute for Targeted LDP Hellos. Refer to section 6 for operational considerations while deploying the solution described in this document.

## 2. Terminology

This document uses the terminology defined in [RFC3031] and [RFC5036].

## 3. Targeted LDP Hello Reduction Procedure

The Targeted LDP Hello Reduction procedure uses the existing Common Hello Parameters TLV defined in [RFC5036]. Figure 1. shows the encoding of the TLV from [RFC5036] for reference.

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  Common Hello Parm(0x0400)										Length																													
										Hold Time										T R  Reserved																			

Figure 1. Common Hello Parameters TLV.

By definition in [RFC5036], a value of 0 means use the default, which is 45 seconds for Targeted Hellos. A value of 0xffff means infinite.

The procedure to be followed for Targeted LDP Hello Reduction between a pair of LSRs is as follows:

1. An LSR starts transmitting periodic targeted hellos to its peer in order to establish the targeted hello adjacency. Each LSR proposes its configured hello hold time in the Common Hello Parameters TLV in its hello message to the peer. The hold time used

between a pair of LSRs is the minimum of the hold times proposed in their Hellos.

2. If the Hello is acceptable by receiving LSR, it establishes targeted hello adjacency with the source LSR. Establishment of Hello adjacency establishes the LDP session between peering LSRs.

3. After the LDP session is ESTABLISHED [RFC5036], each LSR MAY start proposing "relaxed" hold time (higher than configured) in Common Hello Parameters TLV in the subsequent Hello Messages.

Each LSR increases the advertised hold time by some factor after sending a set of Hellos (let's say 5) advertising consistent hold time. As the process of relaxing the advertised hold time continues, after a certain period of time an LSR reaches the maximum holdtime value of 0xffff. Thus after the session is ESTABLISHED, the hello hold time between the LSRs gets negotiated to infinite. Note that the Targeted Hello Adjacency continue to exist and only the adjacency hold times are now infinite.

4. If there are any changes in any parameters associated with a t-LDP Hello adjacency (e.g Configuration Sequence Number etc) then an updated Hello MUST be sent immediately without any changes to the "current" hold time (e.g infinite) that was advertised in the last Hello Message. Since hellos are not reliable, after any parameter change an implementation may send a set of hellos (let's say 5) at configured intervals (or faster) to reflect the change. But those hellos would continue to advertise infinite hold time and would fall back to reduced transmission rate after those 5 packets are sent.

5. If the LDP session between two LSRs fails leading to tearing down of adjacency, then each LSR reverts to advertising their configured hello hold time and repeats procedure 1 to 3. This also applies when LDP session restarts gracefully [RFC3478] when peering LSRs are graceful restart capable. Thus the reduction procedures allow an operator to configure very aggressive Targeted LDP Hello Holdtime to expedite bringing up a large number of LDP sessions in the event of failure but reduces the overhead of hello adjacency maintenance by manifold when sessions are ESTABLISHED. It is desirable to configure aggressive hold times in order to tear down spurious hello adjacencies sooner.

6. When a t-LDP adjacency with a remote LSRs has negotiated to infinite hold time and then remote LSR decides to tear down the adjacency without impacting the established LDP Session then local LSR would not be able to detect that remote node is no longer accepting hellos. It is RECOMMENDED that when a LSR that implements the Hello Reduction procedures send one or a set of contiguous hellos

(let's say 3) advertising hold time of 1 second while bringing down t-LDP Hello adjacency. This graceful closure procedure would cause the hello holdtimes at receiving LSR to be renegotiated to 1 second, which would eventually lead to tear down of the adjacency (due to timeout) by receiving LSR.

It is RECOMMENDED that each peering LSR implements the Targeted LDP Hello Reduction procedure; otherwise negotiated hello hold time between the LSRs does not fall back to the infinite hold time in step 3.

Note that it is not mandatory to advertise infinite hold time after session is established but can be any value that is significantly larger than configured hello hold time. However, it is RECOMMENDED to reach Infinite holdtime after session setup to derive maximum advantage from the procedure described above.

The Hello Reduction procedures does not apply to Basic Discovery (Link LDP Hellos) as Link LDP Hellos need to be sent over an interval continually in order to discover and set up sessions with new peers, especially over a multi-access interface.

#### 4. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

#### 5. Security Considerations

- Control plane aspects
- LDP security (authentication) methods as described in [RFC5036] is applicable here.
- Data plane aspects
- This specification does not have any impact on the MPLS forwarding plane setup by LDP.

#### 6. Operational Considerations

The method proposed in the document reduces significant burden on an LDP LSR that maintains Targeted LDP sessions to a large number (in

thousands) of peers. Further, if BFD [RFC5880] [RFC5883] is used for tracking connectivity to peers it is desirable to turn off Targeted LDP hellos after the LDP session is setup. However there are scenerios where tunring off Targeted LDP Hellos may not be desirable. Such scenerios are as follows:

1. When transport address of the LDP session is different from the IP addresses used to exchange t-LDP Hellos then Session Keepalives are not substitute for reachability or liveliness of adjacency. It is possible to use BFD to track the reachability of IP addresses used for t-LDP Hellos in which case t-LDP Hellos may be redundant. However if an implementation/deployment uses t-LDP hellos for purposes other than liveliness tracking then it is not recommended to turn on t-LDP hello reduction procedures.

2. While t-LDP Hello Reduction Procedures are deployed, it may be possible that t-LDP Hellos are disabled at remote LSR without bringing down the LDP session. If the remote LSR does not implement the procedure for graceful teardown of hello adjacency as described in step 6 in section 3 then it is possible that local LSR may not be able detect that remote LSR is no longer accepting Hellos and thus Hello adjacency would continue to exist in local LSR. It is also possible that the hello(s) sent during graceful cloure of adjacency may get lost (since LDP Hellos are not reliable) and thus local LSR may not detect the loss of adjacency with remote LSR.

## 7. Acknowledgements

The authors would like acknowledge the detailed review and the comments, suggestions from Markus Jork, Thomas Beckhaus, Lizhong Zin and Eric Rosen.

## 8. References

### 8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC5036] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, October 2007.

### 8.2. Informative References

- [I-D.ietf-mpls-seamless-mpls]  
Leymann, N., Decraene, B., Filsfils, C., Konstantynowicz,

M., and D. Steinberg, "Seamless MPLS Architecture",  
draft-ietf-mpls-seamless-mpls-01 (work in progress),  
March 2012.

- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
- [RFC3478] Leelanivas, M., Rekhter, Y., and R. Aggarwal, "Graceful Restart Mechanism for Label Distribution Protocol", RFC 3478, February 2003.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.
- [RFC5883] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for Multihop Paths", RFC 5883, June 2010.

#### Authors' Addresses

Pranjal Kumar Dutta  
Alcatel-Lucent  
701 E Middlefield Road  
Mountain View, CA 94043  
USA

Email: pranjal.dutta@alcatel-lucent.com

Giles Heron  
Cisco Systems  
9-11 New Square  
Bedfont Lakes, Feltham, Middlesex TW14 8HA  
United Kingdom

Email: giheron@cisco.com

Thomas Nadeau  
Juniper Networks

Email: tnadeau@juniper.net

