PALS Working Group
Internet Draft
Intended Status: Proposed Standard
Expires: April 26, 2015

Patrice Brissette
Kamran Raza
Sami Boutros
Cisco Systems, Inc.

Nick Del Regno
Matthew Turlington
Verizon

June 29, 2015

Handling Incoming Label Request for PW FEC Types
draft-brissette-pals-pw-fec-label-request-01

Abstract

This document clarifies the behavior of an LSR PE upon receiving an LDP Label Request message for Pseudowire (PW) FEC types. Furthermore, this document specifies the procedures to be followed by the LSR PE in order to answer such requests for a given PW FEC type.

Convention

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

Table of Contents

1 Introduction

Label Distribution Protocol (LDP) base specification [RFC5036] defines different LDP message types and their procedures for advertising label bindings. These procedures are generic and inherited by any FEC type that is advertised using these message types. For a given FEC type, any difference in behavior, compared to what is already specified in RFC 5036, needs to be spelled out clearly in the corresponding specification in which the FEC type is being introduced or extended.

[RFC4447] specifies mechanisms to setup pseudowires (PWs) using LDP. [RFC4447] does not specify any behavior change with regards to label binding distribution for PW FEC types in response to a corresponding Label Request message from a peer LSR PE. This implies that [RFC4447] inherits the base procedures defined in [RFC5036] for Label Request and associated response for a PW FEC type. The lack of specification in the area of Label Request in [RFC4447] has led to some interoperability issues between vendors due to different interpretation. For example, there are some implementations which do not honor and do not respond to an incoming Label Request for a PW FEC type, resulting in functionality impact. Some of these problems are very critical for the deployment of PW technologies. The following is a non-exhaustive list of some of the problems and potential breakages that may result due to the lack of support of incoming Label Request for a PW FEC:

- An LSR PE can not restart forwarding of packet with sequence number 1 as specified in section 4.1 of [RFC4385] with regards to Control Word Sequencing.

- An LSR PE may not be able to perform a PW consistency check as defined in section 4.1 of [RFC6667], resulting in LSR PEs becoming out-of-sync.

- Some implementations of LSR PE do not checkpoint PW label bindings learnt from peer(s) in their persistent memory and hence are not able to recover any peer state after their own restarts or switchovers. Such implementations typically require re-learning of peer's label bindings after their own failure and rely on Label Request mechanisms.

- The combination of Downstream Unsolicited mode and Conservative Label retention (used due to memory limitations) can lead to a situation where an LSR PE releases the label learnt from a peer for a PW that it may need later. Label Request is used to solve this issue. For example, consider an LSR PE operating in Label Conservative mode receiving a label binding for a non-locally configured/known PW. This LSR PE ignores such a label binding and later tries to re-learn it via Label Request procedure once PW is locally configured. The authors will like to remind the readers about the following fact: [RFC4447] does not mandate to use Label Liberal mode. Therefore it is possible that some implementation used Label Conservative mode.

This document clarifies the use of Label Request message and its procedures for PW FEC types and re-enforces the acceptable behavior to be implemented by an LSR PE.

2. Requirements

This document recommends the following action to be implemented by an LSR PE that supports a PW FEC Type (P2P or P2MP type):

- An LSR PE MUST respond to an incoming Label Request message for a PW FEC by sending its local binding for the PW via a Label Mapping message. If no such binding is available, the LSR PE SHOULD respond by sending a new status code "No PW" in a Notification message.

- An LSR PE MUST respond to an incoming Label Request message for a Wildcard FEC [RFC5036] by sending its local bindings for all its PWs via Label Mapping messages. This is in addition to label bindings corresponding to any other LDP FEC types configured and available at the LSR.

- An LSR PE MUST respond to an incoming Label Request message for a Typed Wildcard PW FEC [RFC6667] by sending its local bindings for all its PWs for the given FEC type via Label Mapping messages. For a given PW FEC type, this advertisement is to be scoped either for a specific PW type or for all PW types according to the received PW Typed Wildcard FEC.

3. Procedures

This document re-enforces the Label Request generic procedures, as defined by RFC 5036, for PW FEC types, and hence strongly recommends that an LSR PE receiving the PW Label Request message should respond either by sending its label binding in Label Mapping message(s) or with a Notification message indicating why it cannot satisfy the request.

An LSR PE should respond to a Label Request when corresponding PW FEC is resolved locally. The following sub sections define the meaning of a "resolution" for a given PW FEC type.

3.1 PWid FEC (FEC128)

A PWid FEC is resolved when a local label binding has been allocated after local configuration application.

[RFC6073] does not preclude setting up MS-PWs using FEC-128, therefore this procedure is also applicable to PEs acting as S-PEs.

3.2 Generalized PWid FEC (FEC129):

A Generalized PWid FEC is resolved at an ST-PE when SAII is locally configured, TAII is learnt statically or dynamically via discovery mechanisms, and a local label binding has been allocated.

This FEC is resolved at an TT-PE when SAII is locally configured, TAII is learnt statically or dynamically via discovery mechanisms, remote label binding is received, and a local label binding has been allocated.

Whereas, this FEC is resolved at an S-PE when remote label binding is received for PW segment, TAII is learnt statically or dynamically via discovery mechanisms, and a local label binding has been allocated.

3.3 Common to PWid and Generalized PWid FEC

A FEC is resolved at an S-PE when remote label binding is received for PW segment.

In the case of Generalized PWid FEC, TAII is learnt statically or dynamically via discovery mechanisms, and a local label binding has been allocated. Whereas PWid FEC is resolved when a local binding has been allocated.

3.4 P2MP PW Upstream FEC (FEC130):

Editor Note: Deferred for further study.

3.5 P2MP PW Downstream FEC (FEC132):

Editor Note: Deferred for further study.

3.5 PW Typed Wildcard FEC

The rules defined for individual PW FEC types apply equally when they are used under a PW Typed Wildcard FEC [RFC6667].

4 Acknowledgements

The authors would like to thank for Alexander Vainshtein its reviews and comments of this document.

5 Security Considerations

This document does not introduce any additional security constraints.

6 IANA Considerations

This document requires the assignment of a new LDP Status Code to be used in a Notification message to notify a peer LSR if lookup fails at receiving LSR for a PW FEC received in a Label Request message.

The value requested from the IANA managed LDP registry "LDP Status Code Name Space" is:

Range/Value	E	Description
0x00000032	0	No PW

7 References

7.1 Normative References

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

[RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed., "LDP Specification", RFC 5036, October 2007.

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

[RFC6667] Raza, K., Boutros, S., and Pignataro, C., "LDP Typed Wildcard FEC for PWid and Generalized PWid FEC", RFC 6667, July 2012.

7.2 Informative References

Authors’ Addresses

Patrice Brissette
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, ON K2K-3E8, Canada.
EMail: pbrisset@cisco.com

Kamran Raza
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, ON K2K-3E8, Canada.
EMail: skraza@cisco.com

Sami Boutros
Cisco Systems, Inc.
3750 Cisco Way,
San Jose, CA 95134, USA.
E-mail: sboutros@cisco.com

Nick Del Regno

Verizon
400 International Pkwy
Richardson, TX 75081, USA.
E-mail: nick.delregno@verizon.com

Matthew Turlington
Verizon
400 International Pkwy
Richardson, TX 75081, USA.
E-mail: matt.turlington@verizon.com

Networking Working Group   V. Govindan
Internet-Draft   C. Pignataro
Intended status: Standards Track   Cisco Systems
Expires: September 06, 2015   March 05, 2015

Advertising S-BFD Discriminators in L2TPv3
draft-gp-l2tpext-sbfd-discriminator-00.txt

Abstract

This document defines a new AVP for advertising one or more S-BFD Discriminators using the L2TPv3 Control Protocol AVP.

Requirements Language

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

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

[I-D.ietf-bfd-seamless-base] defines a simplified mechanism to use Bidirectional Forwarding Detection (BFD)[RFC5880]. This mechanism depends on network nodes knowing the BFD discriminators which each node in the network has reserved for this purpose. Use of the Layer2 Tunneling protocol Version 3 (L2TPv3) [RFC3931] is one possible means of advertising these discriminators. S-BFD requires the usage of unique discriminators within an administrative domain.

This document specifies the encoding to be used when S-BFD discriminators are advertised using L2TPv3.

2. S-BFD Target Discriminator ID AVP

This AVP is exchanged during session negotiation (ICRQ, ICRP, OCRQ, OCRP).

2.1. Encoding Format

The S-BFD Target Discriminator ID AVP, Attribute Type TBD, is an identifier used to advertise the S-BFD target discriminators supported by an LCCE for S-BFD Reflector operation. This AVP indicates that the advertiser implements a S-BFD reflector supporting the specified target discriminators and is ready for S-BFD Reflector operation. The receiving LCCE MAY use this AVP if it wants to monitor connectivity to the advertising LCCE using S-BFD or BFD.

The Attribute Value field for this AVP has the following format:

S-BFD Discriminator Advertisement (ICRQ, ICRP, ICCN, OCRQ, OCRP, OCCN):

```
                                        No. of octets
              +-----------------------------+
              | Discriminator Value(s)      |     4/Discriminator
              :                             :
              +-----------------------------+
```

An LCCE MAY include the S-BFD Discriminator Advertisement AVP in a L2TP Control Protocol message (ICRQ, ICRP, OCRQ, OCRP) [RFC3931]. Multiple S-BFD Discriminators AVPs MAY be advertised by a LCCE. If the other LCCE does not wish to monitor connectivity using S-BFD, it MAY safely discard this AVP without affecting the rest of session negotiation. While current use-cases [I-D.ietf-bfd-seamless-use-case] of S-BFD require advertisement of only one discriminator, the AVP encoding allows specification an arbitrary number of discrminators for extensibility. When multiple S-BFD discriminators are advertised, the mechanism to choose a subset of specific discriminator(s) is out of scope for this document.

The M bit of the L2TP Control Protocol Message (ICRQ, ICRP, OCRQ, OCRP) [RFC3931] MUST NOT be set inside the S-BFD Target Discriminator ID AVP advertisement.

3. IANA Considerations

This number space is managed by IANA as per [RFC3438].

A summary of the new AVPs requested for Attribute Type 0 follows:

Control Message Attribute Value Pairs

Attribute Type	Description
TBD	S-BFD Discriminators

4. Security Considerations

Security concerns for L2TP are addressed in [RFC3931]. Introduction of the S-BFD Discriminator Advertisement AVP introduces no new security risks for L2TP.

Advertisement of the S-BFD discriminators does make it possible for attackers to initiate S-BFD sessions using the advertised information. The vulnerabilities this poses and how to mitigate them are discussed in the Security Considerations section of [I-D.ietf-bfd-seamless-base].

5. Acknowledgements

Authors would like to thank Nobo Akiya, Stewart Bryant and Pawel Sowinski for providing core inputs for the document and for performing thorough reviews and providing number of comments. Authors would like to thank Nagendra Kumar for his reviews.

6. Contributing Authors

Mallik Mudigonda
Cisco Systems
Email: mmudigon@cisco.com

7. References

7.1. Normative References

[I-D.ietf-bfd-seamless-base]
Akiya, N., Pignataro, C., Ward, D., Bhatia, M., and J. Networks, "Seamless Bidirectional Forwarding Detection (S-BFD)", draft-ietf-bfd-seamless-base-04 (work in progress), January 2015.

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

[RFC3438] Townsley, W., "Layer Two Tunneling Protocol (L2TP) Internet Assigned Numbers Authority (IANA) Considerations Update", BCP 68, RFC 3438, December 2002.

[RFC3931] Lau, J., Townsley, M., and I. Goyret, "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, March 2005.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.

7.2. Informative References

[I-D.ietf-bfd-seamless-use-case] Aldrin, S., Bhatia, M., Mirsky, G., Kumar, N., and S. Matsushima, "Seamless Bidirectional Forwarding Detection (BFD) Use Case", draft-ietf-bfd-seamless-use-case-01 (work in progress), December 2014.

Authors’ Addresses

Vengada Prasad Govindan
Cisco Systems

Email: venggovi@cisco.com

Carlos Pignataro
Cisco Systems

Email: cpignata@cisco.com

Internet Engineering Task Force
Internet-Draft
Updates: 5885 (if approved)
Intended status: Standards Track
Expires: January 1, 2016

V. Govindan
C. Pignataro
Cisco Systems
June 30, 2015

Seamless BFD for VCCV
draft-gp-pals-seamless-vccv-01

Abstract

This document extends the procedures and Connectivity Verification (CV) types already defined for Bidirectional Forwarding Detection (BFD) for Virtual Circuit Connectivity Verification (VCCV) to define Seamless BFD (S-BFD) for VCCV. This document will be extended in future to include definition of procedures for S-BFD over Tunnels. This document extends the CV values defined in RFC5885.

Status of This Memo

Copyright Notice

Table of Contents

1. Background

BFD for VCCV [RFC5885] defines the CV types for BFD using VCCV, protocol operation and the required packet encapsulation formats. This document extends those procedures, CV type values to enable S-BFD [I-D.ietf-bfd-seamless-base] operation for VCCV.

The new S-BFD CV Types are PW demultiplexer-agnostic, and hence applicable for both MPLS and Layer Two Tunneling Protocol version 3 (L2TPv3) pseudowire demultiplexers. This document concerns itself with the S-BFD VCCV operation over single-segment pseudowires (SS-PWs). The scope of this document is as follows:

This specification describes procedures only for S-BFD asynchronous mode.

S-BFD Echo mode is outside the scope of this specification.

S-BFD operation for fault detection and status signaling is outside the scope of this specification.

1.1. Requirements Language

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

2. S-BFD Connectivity Verification

S-BFD protocol provides continuity check services by monitoring the S-BFD control packets sent and received over the VCCV channel of the PW. The term <Connectivity Verification> is used throughout this document to be consistent with [RFC5885].

This section defines the CV types to be used for S-BFD. It also defines the procedures for S-BFD discriminator advertisement for the SBD reflector and the procedure for S-BFD Initiator operation.

Two CV Types are defined for S-BFD. Table 1 summarizes the S-BFD CV Types, grouping them by encapsulation (i.e., with versus without IP/UDP headers) for fault detection only. S-BFD for fault detection and status signaling is outside the scope of this specification.

	Fault Detection Only	Fault Detection and Status Signaling
S-BFD, IP/UDP Encapsulation (with IP/UDP Headers)	TBD1 (Note1)	N/A
S-BFD, PW-ACH Encapsulation when using MPLS PW or L2SS Encapsulation when using L2TP PW (without IP/UDP Headers)	TBD2 (Note2)	N/A

Table 1: Bitmask Values for BFD CV Types

Two new bits are requested from IANA to indicate S-BFD operation.

2.1. Co-existence of S-BFD and BFD capabilites

Since the CV types for S-BFD and BFD are unique, BFD and S-BFD capabilities can be advertised concurrently.

2.2. S-BFD CV Operation

2.2.1. S-BFD Initiator Operation

The S-BFD Initiator SHOULD bootstrap S-BFD sessions after it learns the discriminator of the remote target identifier through one or more of the following methods:

1. Advertisements of S-BFD discriminators made through AVP/ TLVs defined in L2TP/ LDP.

2. Provisioning of S-BFD discriminators.

3. Probing remote S-BFD discriminators through S-BFD Alert discriminators [I-D.akiya-bfd-seamless-alert-discrim]

S-BFD Initiator operation MUST be according to the specifications in Section 7.2 of [I-D.ietf-bfd-seamless-base].

2.2.2. S-BFD Reflector Operation

When as pseudowire signalling protocol such as LDP or L2TPv3 is in use the S-BFD Reflector advertises its target discriminators using that signalling protocol. When static PWs are in use the target discriminator of S-BFD needs to be provisioned on the S-BFD Initiator nodes.

All point to point pseudowires are bidirectional, the S-BFD Reflector therefore reflects the S-BFD packet back to the Initiator using the VCCV channel of the reverse direction of the PW on which it was received.

It is observed that the reflector has enough information to reflect the S-BFD Async packet received by it back to the S-BFD initiator using the fields of the L2TPv3 headers.

S-BFD Reflector operation for BFD protocol fields MUST be according to the specifications in Section TBD of [I-D.ietf-bfd-seamless-base].

2.2.2.1. S-BFD Reflector Demultiplexing

TBD

2.2.2.2. S-BFD Reflector transmission of control packets

The procedures of S-BFD Reflector described in [I-D.ietf-bfd-seamless-base] apply for S-BFD using VCCV.

2.2.2.3. S-BFD Reflector advertisement of target discriminators using LDP

TBD.

2.2.2.4. S-BFD Reflector advertisement of target discriminators using L2TP

The S-BFD Reflector MUST use the AVP [I-D.gp-l2tpext-sbfd-discriminator] defined for advertising its target discriminators using L2TP.

2.2.2.5. Provisioning of S-BFD Reflector target discriminators

S-BFD target discriminators MAY be provisioned when static PWs are used.

2.2.2.6. Probing of S-BFD Reflector target discriminators using alert discriminators

S-BFD alert discriminators MAY be used to probe S-BFD target discriminators. If a node implements S-BFD reflector, it SHOULD respond to Alert discriminator requests received from potential S-BFD Initiators.

2.3. S-BFD Encapsulation

Unless specified differently below, the encapsulation of S-BFD packets is the identical the method specified in Sec.3.2 [RFC5885] and in [RFC5880] for the encapsulation of BFD packets.

- o IP/UDP BFD Encapsulation (BFD with IP/UDP Headers)

  * The destination UDP port for the IP encapsulated S-BFD packet MUST be 7784 [I-D.ietf-bfd-seamless-base].

  * The encapsulation of the S-BFD header fields MUST be according to Sec.7.2.2 of [I-D.ietf-bfd-seamless-base].

- o PW-ACH/ L2SS BFD Encapsulation (BFD without IP/UDP Headers)

  * The encapsulation of S-BFD packets using this format MUST be according to Sec.3.2 of [RFC5885] with the exception of the PW-ACH/ L2SS type.

  * When VCCV carries PW-ACH/ L2SS-encapsulated S-BFD (i.e., "raw" S-BFD), the PW-ACH (pseudowire CW's) or L2SS' Channel Type MUST be set to TBD2 to indicate "S-BFD Control, PW-ACH/ L2SS-encapsulated" (i.e., S-BFD without IP/UDP headers; see Section 5.3). This is to allow the identification of the encased S-BFD payload when demultiplexing the VCCV control channel.

2.4. S-BFD CV Types

3. Capability Selection

When multiple S-BFD CV Types are advertised, and after applying the rules in [RFC5885], the set that both ends of the pseudowire have in common is determined. If the two ends have more than one S-BFD CV Type in common, the following list of S-BFD CV Types is considered in the order of the lowest list number CV Type to the highest list number CV Type, and the CV Type with the lowest list number is used:

1. TBD1 - S-BFD IP/UDP-encapsulated, for PW Fault Detection only.

2. TBD2 - S-BFD PW-ACH/ L2SS-encapsulated (without IP/UDP headers), for PW Fault Detection only.

The order of capability selection between S-BFD and BFD is defined as follows:

Advertised capabilities of PE1/ PE2	BFD Only	SBFD Only	Both S-BFD and BFD
BFD Only	BFD	None (Note1)	BFD Only
S-BFD Only	None (Note1)	S-BFD	S-BFD only
Both S-BFD and BFD	BFD only	S-BFD only	Both SBFD and BFD

Table 2: Capability Selection Matrix for BFD and S-BFD

Note1: Can we mandate failing the bringup of the PW in case of a capability mismatch?

4. Security Considerations

Security measures described in [RFC5885] and [I-D.ietf-bfd-seamless-base] are to be followed.

5. IANA Considerations

5.1. MPLS CV Types for the VCCV Interface Parameters Sub-TLV

The VCCV Interface Parameters Sub-TLV codepoint is defined in [RFC4446], and the VCCV CV Types registry is defined in [RFC5085].

This section lists the new BFD CV Types.

IANA has augmented the "VCCV Connectivity Verification (CV) Types" registry in the Pseudowire Name Spaces reachable from [IANA]. These are bitfield values. CV Type values TBD are specified in Section 2 of this document.

MPLS Connectivity Verification (CV) Types:

```
Bit (Value)  Description                        Reference
===========  ===========                        ==============
TBD1(0xY)    S-BFD IP/UDP-encapsulated,         this document
             for PW Fault Detection only
TBD2(0xZ)    S-BFD PW-ACH/L2SS-encapsulated,    this document
             for PW Fault Detection only
```

5.2. L2TPv3 CV Types for the VCCV Capability AVP

This section lists the new requests for S-BFD CV Types to be added to the existing "VCCV Capability AVP" registry in the L2TP name spaces. The Layer Two Tunneling Protocol "L2TP" Name Spaces are reachable from [IANA]. IANA is requested to assign the following L2TPv3 Connectivity Verification (CV) Types in the VCCV Capability AVP Values registry.

```
VCCV Capability AVP (Attribute Type 96) Values
----------------------------------------------

L2TPv3 Connectivity Verification (CV) Types:

Bit (Value)  Description                   Reference
===========  ===========                   ==============
TBD1(0xY)    S-BFD IP/UDP-encapsulated,    this document
             for PW Fault Detection only
TBD2(0xZ)    S-BFD L2SS-encapsulated,    this document
             for PW Fault Detection only
```

5.3. PW Associated Channel Type

As per the IANA considerations in [RFC5586], IANA is requested to allocate the following Channel Types in the "MPLS Generalized Associated Channel (G-ACh) Types" registry:

IANA has reserved a new Pseudowire Associated Channel Type value as follows:

Registry:

```
                                                TLV
 Value   Description                            Follows  Reference
 ------  -------------------------------------  -------  ---------------
 TBD2    S-BFD Control, PW-ACH/L2SS             No       [This document]
         encapsulation
         (without IP/UDP Headers)
```

6. Acknowledgements

Authors would like to thank Nobo Akiya, Stewart Bryant, Pawel Sowinski and Greg Mirsky for providing the core inputs of this document and for performing thorough reviews and providing number of comments. Authors would also like to thank Yuanlong for comments received.

7. Contributing Authors

Mallik Mudigonda
Cisco Systems
Email: mmudigon@cisco.com

8. References

8.1. Normative References

[I-D.akiya-bfd-seamless-alert-discrim]
Akiya, N., Pignataro, C., and D. Ward, "Seamless Bidirectional Forwarding Detection (S-BFD) Alert Discriminator", draft-akiya-bfd-seamless-alert-discrim-03 (work in progress), October 2014.

[I-D.gp-l2tpext-sbfd-discriminator]
Govindan, V. and C. Pignataro, "Advertising S-BFD Discriminators in L2TPv3", draft-gp-l2tpext-sbfd-discriminator-00 (work in progress), March 2015.

[I-D.ietf-bfd-seamless-base]
Akiya, N., Pignataro, C., Ward, D., Bhatia, M., and J. Networks, "Seamless Bidirectional Forwarding Detection (S-BFD)", draft-ietf-bfd-seamless-base-05 (work in progress), June 2015.

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

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, February 2006.

[RFC4446] Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)", BCP 116, RFC 4446, April 2006.

[RFC5085] Nadeau, T. and C. Pignataro, "Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires", RFC 5085, December 2007.

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

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.

[RFC5885] Nadeau, T. and C. Pignataro, "Bidirectional Forwarding Detection (BFD) for the Pseudowire Virtual Circuit Connectivity Verification (VCCV)", RFC 5885, June 2010.

8.2. Informative References

[IANA] Internet Assigned Numbers Authority, "Protocol Registries", <http://www.iana.org>.

Authors' Addresses

Vengada Prasad Govindan
Cisco Systems

Email: venggovi@cisco.com

Carlos Pignataro
Cisco Systems

Email: cpignata@cisco.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 21, 2016

F. Zhang, Ed.
Huawei
B. Wu, Ed.
ZTE Corporation
E. Bellagamba, Ed.
Ericsson
M. Chen, Ed.
Huawei
July 20, 2015

LDP Extensions for Proactive Operations, Administration and Maintenance Configuration of Dynamic MPLS Transport Profile PseudoWire

draft-ietf-pals-mpls-tp-oam-config-03

Abstract

This document defines extensions to LDP to configure proactive OAM functions for both SS-PW and MS-PW when the PW control plane is used.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

There are two documents that define MultiProtocol Label Switching (MPLS) Pseudowire (PW). [RFC3985] defines Single Segment PW (SS-PW) and [RFC5659] defines Multi-Segment PW (MS-PW). The two documents explain how to provide emulated services over an MPLS Packet Switched Network (PSN). The MPLS Transport Profile (MPLS-TP) is described in [RFC5291], which describes a profile of MPLS that introduces the operational models that were typically used in transport networks, while providing additional Operations, Administration and Maintenance (OAM), survivability and other maintenance functions that were not previously supported by IP/MPLS network. The MPLS-TP requirements are defined in [RFC5860].

The MPLS-TP OAM mechanisms are described in [RFC6371], which can be categorized into proactive and on-demand OAM. Proactive OAM refers to OAM operations that are either configured to be carried out periodically and continuously or preconfigured to act on certain events (e.g., alarm signals). In contrast, on-demand OAM is initiated manually and for a limited amount of time, usually for operations such as diagnostics to investigate into a defect condition.

When a control plane is not used the OAM functions are typically configured from the Network Management System (NMS). When a control plane is used, requirement 51 in [RFC5654] requires that it MUST be able to support configuration of the OAM functions. The control plane is also required to be able to configure, maintain and modify, as well as activation/deactivation of maintenance points.

For MPLS-TP OAM configuration, two companion documents exists. [RFC7260] and [RFC7487] define extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) to support the establishment and configuration of OAM entities along with Label Switched Path (LSP) signaling. [I-D.ietf-mpls-lsp-ping-mpls-tp-oam-conf] defines extensions to LSP Ping [RFC4379] to support the configuration of proactive MPLS-TP OAM functions.

This document defines extensions to the Label Distribution Protocol (LDP) to configure proactive MPLS-TP PW OAM functions for both Point to Point SS-PW and MS-PW when the PW control plane is used. The extensions defined in this document are aligned with those companion documents. Extensions to Point to Multi-Point (P2MP) PW are for future study and 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].

2.1. Acronyms

AIS: Alarm indication signal

BFD: Bidirectional Forwarding Detection

CC: Continuity Check

CV: Connectivity Verification

DM: Delay Measurement

FEC: Forwarding Equivalence Class

FMS: Fault Management Signal

G-ACh: Generic Associated Channel

LDI: Link Down Indication

LDP: Label Distribution Protocol

LM: Loss Measurement

LSP: Label Switched Path

MEP: Maintenance Entity Group End Point

MIP: Maintenance Entity Group Intermediate Point

MPLS-TP: MPLS Transport Profile

MS-PW: Multi-Segment PseudoWire

NMS: Network Management System

OAM: Operations, Administration and Maintenance

P2MP: Point to Multi-Point

PE: Provider Edge

PHB: Per-Hop Behavior

PM: Performance Monitoring

PSN: Packet Switched Network

PW: Pseudowire

S-PE: Switching Provider Edge

SS-PW: Single-Segment Pseudo Wire

T-PE: Terminating Provider Edge

TLV: Type Length Value

3. PW OAM Configuration

This document defines two new TLVs, the PW OAM Administration TLV and the PW OAM Functions TLV.

The PW OAM Administrations TLV is used to setup/remove MIP and MEP functions and to control whether OAM alarm function should be suppressed or not.

The PW OAM Functions TLV is used to configure, enable and disable OAM functions that include Continuity Check (CC), Connectivity Verification (CV), Packet Loss/Delay/Throught and PM and Fault Management Signal(FMS). More details about the new TLVs can be found in Section 4.

3.1. OAM Configuration for SS-PW

3.1.1. Establishment of OAM Entities and Functions

OAM entities and functions can be setup, configured and activated either when the PW first is signalled or on an already established PW. This section describes how the OAM entities and functions are setup and configured with the signalling of a PW.

For the case where OAM entities and functions are setup and configured after establishment of a PW, the procedures are identical to the "adjustment of OAM parameters" procedures, more detail can be found in Section 3.1.2.

Given that a SS-PW needs to be setup between PE1 and PE2 (see Figure 1) . OAM functions MUST be setup and enabled in the appropriate order so that spurious alarms can be avoided.

```
            +-------+               +-------+
            |       |               |       |
            |       |---------------|       |
            |       |               |       |
            +-------+               +-------+
               PE1                     PE2
```

Figure 1 SS-PW OAM Configuration Scheme

Figure 1: SS-PW OAM Configuration Scheme

Fist, the ingress PE (e.g., PE1) must setup the OAM sink function and prepare to receive OAM messages. Until the PW is fully established, any OAM alarm SHOULD be suppressed.

To achieve this, a Label Mapping message with the "OAM Alarms Enabled" flag cleared is sent. In the message, the "OAM MEP Entities Desired" flag is set. Since there is no MIPs for a SS-PW, the "OAM MIP Entities desired" flag MUST be cleared. In addition, to configure and enable particular OAM functions, the PW OAM Functions TLV and relevant sub-TLVs MUST be included.

When the Label Mapping message is received by PE2, PE2 needs to check whether it supports the requested OAM configuration. If it does not support the requested OAM configuration, a Label Release message MUST be returned to PE1, with a Status Code set to "PW OAM Parameters Rejected". The PW signalling is complete and the PW will not be established. If the requested OAM parameters and configuration are supported, PE2 will establish and configure the requested OAM entities.

If PE2 fails to establish and configure the OAM entities, a Label Release message will be returned to PE1, with a Status Code set to "PW MEP Configuration Failed". The PW signalling is complete and the PW will not be established.

If the OAM entities are setup and configured successfully, the OAM sink and source functions is setup and the OAM sink function will be configured to receive OAM messages.

Since the OAM alarm is disabled, no alarms will be generated. The OAM source function can start to send OAM messages. PE2 will then reply a Label Mapping message to PE1, the PW OAM Administration TLV and PW OAM Configuration TLV from the received Label Mapping message MUST be copied and carried in the Label Mapping message.

When PE1 receives this Label Mapping message, PE1 completes any pending OAM configuration and enables the OAM source function to send OAM messages.

For PE1, the OAM entities and functions are now setup and configured, and OAM messages may be exchanged. The OAM alarms can be safely enabled. The initiator PE (PE1) will send another Label Mapping message with "OAM Alarms Enabled" flag set to PE2, this will allow PE2 to enable the OAM alarm function.

When the Label Mapping message is received by PE2, the OAM alarm will be enabled. PE2 then sends a Notification message with the Status Code set to "PW OAM Alarms enabled" to PE1.

When the Notification message is received by PE1, PE1 enables the OAM alarm function. At this point, data-plane OAM is fully functional.

3.1.2. Adjustment of OAM Parameters

Existing OAM parameters may be changed during the life time of a PW. To achieve this, PE1 sends a Label Mapping message with the updated OAM parameters to PE2.

To avoid spurious alarms, it is important that OAM sink and source functions on both PEs are updated in a synchronized way. First, the alarms of the OAM sink function should be suppressed. After that, new OAM parameters can be adjusted. Subsequently, the parameters of the OAM source function can be updated. Finally, the alarms can be enabled again.

Consequently, the ingress PE needs to keep its OAM sink and source functions running without any changes until the OAM parameters are updated. However, in order to suppress spurious alarms, it also need to disable the alarm functions before the Label Mapping message, with the "OAM Alarms Enabled" flag cleared and the updated PW OAM Function TLV, is sent. The OAM alarm function needs to be disabled until the corresponding response message is received.

On receipt of the Label Mapping message, PE2 needs to check whether the updated parameters can be supported. If they can be supported, PE2 needs first disable the OAM alarms firstly and then update the OAM parameters. When the update is done, a Notification message needs to be sent to PE1, with the Status Code set to "PW MEP Configuration Succeed", to acknowledge the update. If PE2 can not support the update, a Notification message with Status Code set to "PW OAM Parameters Rejected" will be sent to PE1.

When PE1 receives the Notification message with the Status Code set to "PW MEP Configuration Succeed", PE1 will update using the new OAM parameters. After the OAM parameters are updated and the OAM is running with the new parameter settings, OAM alarms are still disabled. A subsequent Label Mapping message with "OAM Alarms Enabled" flag set will be sent to re-enable OAM alarms. If the Status Code of the received Notification message is "PW OAM Parameters Rejected", it will not update the OAM parameters. The OAM alarms are just enabled again and the OAM will keep running with the old parameters. PE1 can also re-try changing the OAM parameters using a different set of parameters.

When PE2 received the Label Mapping message with "OAM Alarms Enabled" flag set, it will enable the OAM alarms and reply a Notification message with Status Code set to "PW OAM Alarms Enabled". When received the Notification message, PE1 will enable the OAM alarms again.

3.1.3. Deleting OAM Entities

In some cases it may be useful to remove all OAM entities and functions from a PW without actually tearing down the connection. The deleting procedures are defined as below:

First, the ingress PE (e.g., PE1) disables its own the OAM alarms and then sends a Label Mapping message to the remote PE (e.g., PE2) with the "OAM Alarms Enabled" flag set but with all other OAM configurations unchanged.

When received the Label Mapping message, PE2 disables the OAM alarm and then send a Notification message with Status Code set to "PW OAM Alarms Disabled" to PE1.

When received the confirmation from PE2, it’s safe to delete the OAM entities. PE1 will delete the OAM entities related to the PW and send another Label Mapping message with the "OAM MEP Entities desired" flag cleared to PE2.

When PE2 received the Label Mapping message, it will delete all OAM entities related to the PW and then reply a Notification message with the Status Code set to "PW MEP Entities Disabled" to PE1.

3.2. OAM Configuration for MS-PW

3.2.1. Establishment of OAM Entities and Functions

Given that a MS-PW needs to be setup between T-PE1 and T-PE2, across S-PE1 and S-PE2 (see Figure 2) . OAM functions MUST be setup and enabled in the appropriate order so that spurious alarms can be avoided.

```
+-------+        +-------+        +-------+        +-------+
|       |        |       |        |       |        |       |
|      A|--------|B     C|--------|D     E|--------|F      |
|       |        |       |        |       |        |       |
+-------+        +-------+        +-------+        +-------+
  T-PE1            S-PE1            S-PE2            T-PE2
```

Figure 2 MS-PW Scenario

Figure 2: MS-PW OAM Configuration Scheme

Fist, T-PE1 must setup the OAM sink function and prepare to receive OAM messages. Until the PW is fully established, any OAM alarm SHOULD be suppressed.

To achieve this, a Label Mapping message with the "OAM Alarms Enabled" flag cleared is sent. If the S-PEs are expected to setup and configure the MIP entities, the "OAM MIP Entities desired" flag MUST be set. In the message, the "OAM MEP Entities Desired" flag is set. In addition, to configure and enable particular OAM functions, the PW OAM Functions TLV and relevant sub-TLVs MUST be included.

On receipt of the Label Mapping message, S-PE(e.g., S-PE1) needs to check whether it supports MIP function. If S-PE1 does not support MIP function, a Notification message will be sent to T-PE1, with the Status Code set to "PW MIP Configuration Failed". If S-PE1 supports MIP function, it will establish and configure the MIP entities according to the "OAM MIP Entities desired" flag in the PW OAM Administration TLV. No matter whether S-PE1 supports MIP function, it will relay the Label Mapping message downstream to the next S-PE. All the subsequent S-PEs along the PW will perform the same operations as S-PE1 does until the Label Mapping message reaches the remote T-PE (T-PE2).

When the Label Mapping message is received by the remote T-PE (T-PE2), T-PE2 needs to check whether it support the requested OAM configuration. If it does not support the requested OAM configuration, a Label Release message MUST be returned to its upstream PE, with a Status Code set to "PW MEP Configuration Failed". The signalling is complete and the PW will not be established. If the requested OAM parameters and configuration are supported, T-PE2 will establish and configure requested OAM entities.

If T-PE2 fails to establish and configure the OAM entities, a Label Release message MUST be replied to its upstream PE, with a Status Code set to "PW MEP Configuration Failed". The signalling is complete and the PW will not be established.

If the OAM entities established and configured successfully, the OAM sink and source functions are setup and the OAM sink function will be configured to receive OAM messages. Since the OAM alarm is disabled, no alarms will be generated. The OAM source function can start to send OAM messages. T-PE2 will then reply a Label Mapping message, the PW OAM Administration TLV and PW OAM Function TLV from the received Label Mapping message MUST be copied and carried in the returned Label Mapping message.

S-PEs will relay the Label Mapping message upstream until it reaches T-PE1. When the Label Mapping message is received by T-PE1, T-PE1 will complete any pending OAM configuration and enables the OAM source function to send OAM messages.

For T-PE1, the OAM entities and functions are now setup and configured, and OAM messages may be exchanged. The OAM alarms can be safely enabled. T-PE1 will send another Label Mapping message with "OAM Alarms Enabled" flag set to enable the OAM alarm function.

When the Label Mapping message is received by S-PEs, S-PEs will enable the OAM alarm and relay the Label Mapping message downstream until it reaches T-PE2.

When the Label Mapping message is received by T-PE2, the OAM alarm will be enabled. T-PE2 then sends a Notification message to T-PE1, with the Status Code set to "PW OAM Alarms Enabled". Once the Notification message is received by T-PE1, T-PE1 enables the OAM alarm function. At this point, data-plane OAM is fully functional.

3.2.2. Adjustment of OAM Parameters

Existing OAM parameters may be changed during the life time of a PW. To achieve this, the T-PE1 needs to send a Label Mapping message with the updated OAM parameters to adjust and update OAM parameters.

To avoid spurious alarms, it is important that OAM sink and source functions on both sides are updated in a synchronized way. Fist, the alarms of the OAM sink function should be suppressed. After that, new OAM parameters can be adjusted. Subsequently, the parameters of the OAM source function can be updated. Finally, the alarms can be enabled again.

Consequently, T-PE1 needs to keep its OAM sink and source functions running without any changes until the OAM parameters are updated. However, in order to suppress spurious alarms, it also need to disable the alarm functions before the Label Mapping message, with the "OAM Alarms Enabled" flag cleared and the updated PW OAM Function TLV, is sent. The OAM alarm function needs to be disabled until the corresponding response message is received.

When the Label Mapping message is received by S-PEs, each S-PE just disables the OAM alarm and relay the Label Mapping message downstream until the Label Mapping message reaches the remote T-PE (T-PE2).

On receipt of the Label Mapping message, T-PE2 needs to check whether the updated parameters can be supported. If they can be supported, T-PE2 needs first disable the OAM alarms and then update the OAM parameters. When the update is done, a Notification message needs to be sent to T-PE1, with the Status Code set to "PW MEP Configuration Succeed", to acknowledge the update. If T-PE2 can not support the update, a Notification message with Status Code set to "PW OAM Parameters Rejected" will be sent T-PE1.

When T-PE1 receives the Notification message with the Status Code set to "PW MEP Configuration Succeed", T-PE1 will update using the new OAM parameters. After the OAM parameters are updated and the OAM is running with the new parameter settings, OAM alarms are still disabled. A subsequent Label Mapping message with "OAM Alarms Enabled" flag set will be sent to re-enable OAM alarms. If the Status Code of the received Notification message is "PW OAM Parameters Rejected", it will not update the OAM parameters. The OAM alarms are just enabled again and the OAM will keep running with the old parameters. T-PE1 can also re-try changing the OAM parameters using a different set of parameters.

When S-PEs receives the Label Mapping message, they will enable the OAM alarms and relay the Label Mapping message downstream.

When T-PE2 receives the Label Mapping message with the "OAM Alarms Enabled" flag set, it will enable the OAM alarms and reply a Notification message with Status Code set to "PW OAM Alarms Enabled". When received the Notification message, T-PE1 will enable the OAM alarms again.

3.2.3. Deleting OAM Entities

In some cases it may be useful to remove all OAM entities and functions from a PW without actually tearing down the connection. The deleting procedures are defined as below:

First, T-PE1disables its own the OAM alarms and then sends a Label Mapping message to the remote PE (e.g., T-PE2) with the "OAM Alarms Enabled" flag cleared but with all other OAM configurations unchanged.

When received the Label Mapping message, S-PEs will disable the OAM alarm and relay the Mapping message downstream until the Label Mapping message reaches the remote T-PE (T-PE2).

When received the Label Mapping message, T-PE2 will disable the OAM alarm and then reply a Notification message with Status Code set to "PW OAM Alarms Disabled" to T-PE1.

When received the confirmation from T-PE2, it’s safe to delete the OAM entities. T-PE1 will delete the all OAM entities associated with the PW and send another Label Mapping message with both the "OAM MEP Entities desired" and "OAM MIP Entities desired" flags cleared to the remote T-PE.

When received the Label Mapping message, S-PE (e.g., S-PE1) will delete all the OAM entities associated with the PW and relay the Label Mapping message downstream. Subsequent S-PEs will do the same operations until the Label Mapping message reaches the remote T-PE.

When T-PE2 receives the Label Mapping message, it will delete all OAM entities associated with the PW and then reply a Notification message with the Status Code set to "PW MEP Entities Disabled" to T-PE1.

4. LDP Extensions

4.1. PW OAM Administration TLV

The PW OAM Administration TLV is used to configure and enable the MEP, MIP and Alarm functions. It can be sent with the Label Mapping message. The format of the TLV is as follows:

```
  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|           Type            |          Length(=4)           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                OAM Administration Flags                       |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

PW OAM Administration TLV

The PW OAM Administration TLV type is TBD1.

The Length field is 2 octets in length. It defines the length in octets of OAM Administration Flags filed, it's value is 4.

The OAM Administration Flags is a bitmap with the length of 4 octets.

This document defines the following flags:

OAM Administration Flags bit#	Description
0	OAM MIP Entities Desired
1	OAM MEP Entities Desired
2	OAM Alarms Enabled
3-31	Reserved

The "OAM MIP Entities Desired" flag is used to direct the S-PE(s) along a PW to establish (when set) or delete (when cleared) the OAM MIP entities. This flag only applies to MS-PW scenario. For SS-PW case, this flag MUST be cleared when sent, and SHOULD be ignored when received.

The "OAM MEP Entities Desired" flag is used to request the remote T-PE to establish (when set) or delete (when cleared) the OAM entities.

The "OAM Alarms Enabled" flag is used to request the received PEs to enable (when set) or disable (when cleared) OAM alarms function.

Reserved (3-31 bits): MUST be set to zero on transmission and SHOULD be ignored on receipt.

4.2. PW OAM Functions TLV

The PW OAM Functions TLV is defined to configure and enable specific OAM functions, it is carried in Label Mapping message when used. The format of the TLV is as follows:

```
 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|           Type            |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      OAM Function Flags                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                           sub-TLVs                            ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

PW OAM Functions TLV

The PW OAM Functions TLV contains a number of flags indicating which OAM functions should be activated and OAM function specific sub-TLVs with configuration parameters for particular functions.

The PW OAM Functions TLV type is TBD2.

The Length field is 2 octets in length. It defines the length in octets of OAM Function Flags and sub-TLVs fields.

The OAM Function Flags is a bitmap with the length of 4 octets.

This document defines the following flags:

OAM Function Flags bit#	Description
0	Continuity Check (CC)
1	Connectivity Verification (CV)
2	Fault Management Signals (FMS)
3	Performance Monitoring (PM) Loss
4	Performance Monitoring (PM) Delay
5	Performance Monitoring (PM) Throughput
6-31	Reserved

The sub-TLVs corresponding to the different OAM function flags are as follows.

- o BFD Configuration sub-TLV MUST be included if the CC and/or the CV OAM Function flag is set. Furthermore, if the CV flag is set, the CC flag MUST be set as well.

- o Performance Monitoring sub-TLV MUST be included if the PM Loss/ Delay OAM Function flag is set.

- o Fault Management Signal (FMS) sub-TLV MAY be included if the FMS OAM Function flag is set. If the Fault Management Signal sub-TLV is not included, the default configuration values are used.

4.2.1. BFD Configuration sub-TLV

The BFD Configuration Sub-TLV (depicted below) is defined for BFD-OAM-specific configuration parameters. The BFD Configuration Sub-TLV is carried as a sub-TLV of the PW OAM Functions TLV.

This sub-TLV accommodates generic BFD OAM information and carries sub- TLVs.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      BFD Conf. Type (1)       |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Vers.|N|S|I|G|U|B|       Reserved (set to all 0s)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                           sub-TLVs                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

BFD Configuration sub-TLV

Type: indicates a new type, the BFD Configuration Sub-TLV (suggested value 1).

Length: Indicates the length of the Value field in octets.

Version: Identifies the BFD protocol version. If the egress LSR does not support the version, a Notification message MUST be generated, with the Status Code set to "OAM Problem/ Unsupported BFD Version".

BFD Negotiation (N): If set, timer negotiation/re-negotiation via BFD Control messages is enabled. When cleared, it is disabled.

Symmetric Session (S): If set, the BFD session MUST use symmetric timing values.

Integrity (I): If set, BFD Authentication MUST be enabled. If the BFD Configuration Sub-TLV does not include a BFD Authentication Sub-TLV, the authentication MUST use Keyed SHA1 with an empty pre-shared key (all 0s). If the egress LSR does not support BFD Authentication, an Notification message MUST be generated, with Status Code set to "OAM Problem/BFD Authentication unsupported".

Encapsulation Capability (G): If set, it shows the capability of encapsulating BFD messages into The G-Ach channel. If both the G bit and U bit are set, configuration gives precedence to the G bit. If the egress LSR does not support any of the ingress LSR Encapsulation Capabilities, an Notification message MUST be generated, with the Status Code set to "OAM Problem/Unsupported BFD Encapsulation format".

Encapsulation Capability (U): If set, it shows the capability of encapsulating BFD messages into UDP packets. If both the G bit and U bit are set, configuration gives precedence to the G bit. If the egress LSR does not support any of the ingress LSR Encapsulation Capabilities, a Notification message MUST be generated, with the Status Code set to "OAM Problem/Unsupported BFD Encapsulation Format".

Bidirectional (B): If set, it configures BFD in the Bidirectional mode. If it is not set, it configures BFD in unidirectional mode. In the second case, the source node does not expect any Discriminator values back from the destination node.

Reserved: Reserved for future specifications; set to 0 on transmission and ignored when received.

The BFD Configuration Sub-TLV MUST include the following sub-TLVs in the Label Mapping message:

o BFD Identifiers Sub-TLV; and

o Negotiation Timer Parameters Sub-TLV if the N flag is cleared.

The BFD Configuration Sub-TLV MUST include the following sub-TLVs in the "response" Label Mapping message:

o BFD Identifiers Sub-TLV; and

o Negotiation Timer Parameters Sub-TLV if:

  * the N and S flags are cleared; or if

  * the N flag is cleared and the S flag is set and the Negotiation Timer Parameters Sub-TLV received by the egress contains unsupported values. In this case, an updated Negotiation Timer Parameters Sub-TLV containing values supported by the egress LSR MUST be returned to the ingress.

4.2.1.1. Local Discriminator sub-TLV

The Local Discriminator sub-TLV is carried as a sub-TLV of the BFD Configuration sub-TLV and is depicted below.

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  Lcl. Discr. Type (1) (IANA)  |         Length (4)            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                       Local Discriminator                     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Local Discriminator sub-TLV

Type: indicates a new type, the Local Discriminator sub-TLV (suggested value 1).

Length: indicates the length of the Value field in octets (4).

Local Discriminator: A unique, nonzero discriminator value generated by the transmitting system and referring to itself, used to demultiplex multiple BFD sessions between the same pair of systems.

4.2.1.2. Negotiation Timer Parameters sub-TLV

The Negotiation Timer Parameters sub-TLV is carried as a sub-TLV of the BFD Configuration sub-TLV and is depicted below.

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  Timer Neg.  Type (2) (IANA)  |          Length (16)          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |         Acceptable Min. Asynchronous TX interval              |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |         Acceptable Min. Asynchronous RX interval              |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |               Required Echo TX Interval                       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Negotiation Timer Parameters sub-TLV

Type: indicates a new type, the Negotiation Timer Parameters sub-TLV (suggested value 2).

Length: indicates the length of the Value field in octets (12).

Acceptable Min. Asynchronous TX interval: in case of S (symmetric) flag set in the BFD Configuration sub-TLV, defined in Section 4.2.1, it expresses the desired time interval (in microseconds) at which the ingress PE intends to both transmit and receive BFD periodic control packets. If the receiving PE cannot support such value, it SHOULD reply with an interval greater than the one proposed.

In case of S (symmetric) flag cleared in the BFD Configuration sub-TLV, this field expresses the desired time interval (in microseconds) at which T-PE intends to transmit BFD periodic control packets in its transmitting direction.

Acceptable Min. Asynchronous RX interval: in case of S (symmetric) flag set in the BFD Configuration sub-TLV, this field MUST be equal to Acceptable Min. Asynchronous TX interval and has no additional meaning respect to the one described for "Acceptable Min. Asynchronous TX interval".

In case of S (symmetric) flag cleared in the BFD Configuration sub-TLV, it expresses the minimum time interval (in microseconds) at which T-PEs can receive BFD periodic control packets. In case this value is greater than the value of Acceptable Min. Asynchronous TX interval received from the other edge LSR, such T-PE MUST adopt the interval expressed in this Acceptable Min. Asynchronous RX interval.

Required Echo TX Interval: the minimum interval (in microseconds) between received BFD Echo packets that this system is capable of supporting, less any jitter applied by the sender as described in [RFC5880] sect. 6.8.9. This value is also an indication for the receiving system of the minimum interval between transmitted BFD Echo packets. If this value is zero, the transmitting system does not support the receipt of BFD Echo packets. If the receiving system cannot support this value, a Notification message MUST be generated, with the Status Code set to "Unsupported BFD TX Echo rate interval". By default the value is set to 0.

4.2.1.3. BFD Authentication sub-TLV

The BFD Authentication sub-TLV is carried as a sub-TLV of the BFD Configuration sub-TLV and is depicted below.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    BFD Auth. Type (3) (IANA)  |          Length = 8           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Auth Type   |  Auth Key ID  |         Reserved (0s)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

BFD Authentication sub-TLV

Type: indicates a new type, the BFD Authentication sub-TLV (suggested value 3).

Length: indicates the length of the Value field in octets (4).

Auth Type: indicates which type of authentication to use. The same values as are defined in section 4.1 of [RFC5880] are used. If the egress PE does not support this type, an Notification message MUST be generated, with the Status Code set to "OAM Problem/Unsupported BFD Authentication Type".

Auth Key ID: indicates which authentication key or password (depending on Auth Type) should be used. How the key exchange is performed is out of scope of this document. If the egress PE does not support this Auth Key ID, a Notification message MUST be generated, with the Status Code set to "OAM Problem/Mismatch of BFD Authentication Key ID".

Reserved: Reserved for future specification and set to 0 on transmission and ignored when received.

4.2.1.4. Traffic Class Sub-TLV

The Traffic Class sub-TLV is carried as a sub-TLV of the BFD Configuration sub-TLV or Fault Management Signal sub-TLV (Section 4.2.3) and is depicted as below.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Traffic Class sub-Type (104)  |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  TC |                Reserved (set to all 0s)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type, the Traffic Class sub-TLV (suggested value 4).

Length: Indicates the length of the Value field in octets (4).

Traffic Class (TC): Identifies the TC [RFC5462] for periodic continuity monitoring messages or packets with fault management information. If the Traffic Class sub-TLV is present, then the value of the TC field MUST be used as the value of the TC field of an MPLS label stack entry. If the Traffic Class sub-TLV is absent from BFD Configuration sub-TLV or Fault Management Signal sub-TLV, then selection of the TC value is a local decision.

4.2.2. Performance Monitoring sub-TLV

If the PW OAM Functions TLV has either the L (Loss), D (Delay) or T (Throughput) flag set, the Performance Measurement sub-TLV MUST be present. Failure to include the correct sub-TLVs MUST result in a Notification message (with the Status Code set to "OAM Problem/ Configuration Error") being generated. The Performance Measurement sub-TLV provides the configuration information mentioned in Section 7 of [RFC6374]. It includes support for the configuration of quality thresholds and, as described in [RFC6374], "the crossing of which will trigger warnings or alarms, and result reporting and exception notification will be integrated into the system-wide network management and reporting framework." In case the values need to be different than the default ones the Performance Measurement sub-TLV MAY include the following sub-TLVs:

o "MPLS-PW PM Loss sub-TLV" if the L flag is set in the "PW OAM Functions TLV";

o "MPLS-PW PM Delay sub-TLV" if the D flag is set in the "PW OAM Functions TLV ".

The "Performance Monitoring sub-TLV" depicted below is carried as a sub-TLV of the "PW OAM Functions TLV"

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Perf Monitoring Type (IANA)|         Length                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|D|L|J|Y|K|C|           Reserved (set to all 0s)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                           sub-TLVs                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Performance Monitoring sub-TLV

Sub-type: indicates a new sub-type, the Performance Management sub-TLV (suggested value 2).

Length: indicates the length of the Value field in octets (4).
Configuration Flags, for the specific function description please refer to [RFC6374]:

o D: Delay inferred/direct (0=INFERRED, 1=DIRECT). If the egress LSR does not support specified mode, a Notification message MUST be generated, with the Status Code set to "OAM Problem/Unsupported Delay Mode".

- L: Loss inferred/direct (0=INFERRED, 1=DIRECT). If the egress LSR does not support specified mode, a Notification message MUST be generated, with the Status Code set to "OAM Problem/Unsupported Loss Mode".

- J: Delay variation/jitter (1=ACTIVE, 0=NOT ACTIVE). If the egress LSR does not support Delay variation measurements and the J flag is set, a Notification message MUST be generated, with the Status Code set to "OAM Problem/Delay variation unsupported".

- Y: Dyadic (1=ACTIVE, 0=NOT ACTIVE). If the egress LSR does not support Dyadic mode and the Y flag is set, a Notification message MUST be generated, with the Status Code set to "OAM Problem/Dyadic mode unsupported".

- K: Loopback (1=ACTIVE, 0=NOT ACTIVE). If the egress LSR does not support Loopback mode and the K flag is set, a Notification message MUST be generated, with the Status Code set to "OAM Problem/ Loopback mode unsupported".

- C: Combined (1=ACTIVE, 0=NOT ACTIVE). If the egress LSR does not support Combined mode and the C flag is set, a Notification message MUST be generated, with the Status Code set to "OAM Problem/ Combined mode unsupported".

Reserved: Reserved for future specification and set to 0 on transmission and ignored when received.

4.2.2.1. PM Loss Sub-TLV

The PM Loss sub-TLV depicted below is carried as a sub-TLV of the Performance Monitoring sub-TLV.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  PM Loss Type (1) (IANA)      |         Length                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| OTF |T|B|                    RESERVED                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Measurement Interval                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Test Interval                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Loss Threshold                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

PM Loss sub-TLV

Type: indicates a new type, the PM Loss sub-TLV (suggested value 1).

Length: indicates the length of the parameters in octets.

OTF: Origin Timestamp Format of the Origin Timestamp field described in [RFC6374]. By default it is set to IEEE 1588 version 1. If the egress PE cannot support this value, a Notification message MUST be generated, with the Status Code set to "OAM Problem/Unsupported Timestamp Format".

Configuration Flags, please refer to [RFC6374] for further details:

- o T: Traffic-class-specific measurement indicator. Set to 1 when the measurement operation is scoped to packets of a particular traffic class (DSCP value), and 0 otherwise. When set to 1, the DS field of the message indicates the measured traffic class. By default it is set to 1.

- o B: Octet (byte) count. When set to 1, indicates that the Counter 1-4 fields represent octet counts. When set to 0, indicates that the Counter 1-4 fields represent packet counts. By default it is set to 0.

Measurement Interval: the time interval (in microseconds) at which LM query messages MUST be sent on both directions. If the T-PE receiving the Mapping message can not support such value, it can reply back with a higher interval. By default it is set to (100) as per [RFC6375]..

Test Interval: test messages interval as described in [RFC6374]. By default it is set to (10) as per [RFC6375].

Loss Threshold: the threshold value of measured lost packets per measurement over which action(s) SHOULD be triggered.

4.2.2.2. PM Delay Sub-TLV

The PM Delay sub-TLV depicted below is carried as a sub-TLV of the PW OAM Functions TLV.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  PM Delay Type (2) (IANA)     |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| OTF |T|B|                    RESERVED                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Measurement Interval                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Test Interval                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Delay Threshold                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

PM Delay sub-TLV

Type: indicates a new type, the PM Delay sub-TLV (suggested value 2).

Length: indicates the length of the parameters in octets.

OTF: Origin Timestamp Format of the Origin Timestamp field described in [RFC6374]. By default it is set to IEEE 1588 version 1. If the egress LSR cannot support this value, a Notification message MUST be generated, with the Status Code set to "OAM Problem/Unsupported Timestamp Format".

Configuration Flags, please refer to [RFC6374] for further details:

- o T: Traffic-class-specific measurement indicator. Set to 1 when the measurement operation is scoped to packets of a particular traffic class (DSCP value), and 0 otherwise. When set to 1, the DS field of the message indicates the measured traffic class. By default it is set to 1.

- o B: Octet (byte) count. When set to 1, indicates that the Counter 1-4 fields represent octet counts. When set to 0, indicates that the Counter 1-4 fields represent packet counts. By default it is set to 0.

Measurement Interval: the time interval (in microseconds) at which LM query messages MUST be sent on both directions. If the T-PE receiving the Mapping message can not support such value, it can reply back with a higher interval. By default it is set to (1000) as per [RFC6375].

Test Interval: test messages interval as described in [RFC6374]. By default it is set to (10) as per [RFC6375].

Delay Threshold: the threshold value of measured two-way delay (in milliseconds) over which action(s) SHOULD be triggered.

4.2.3. Fault Management Signal Sub-TLV

The Fault Management Signal sub-TLV depicted below is carried as a sub-TLV of the PW OAM Functions TLV. When both working and protection paths are configured, both PWs SHOULD be configured with identical settings of the E flag, T flag, and the refresh timer.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       FMS sub-type (300)      |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|E|S|T|           Reserved            |     Refresh Timer       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                           sub-TLVs                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Fault Management Signal sub-TLV

Type: indicates a new type, the Fault Management Signal sub-TLV (suggested value 4).

Length: indicates the length of the parameters in octets (8).

FMS Signal Flags are used to enable the FMS signals at end point MEPs and the Server MEPs of the links over which the PW is forwarded. In this document only the S flag pertains to Server MEPs.

The following flags are defined:

- o E: Enable Alarm Indication Signal (AIS) and Lock Report (LKR) signaling as described in [RFC6427]. Default value is 1 (enabled). If the egress MEP does not support FMS signal generation, a Notification message MUST be generated, with the

Status Code set to "OAM Problem/Fault management signaling unsupported".

- o S: Indicate to a server MEP that it should transmit AIS and LKR signals on the client PW. Default value is 0 (disabled). If a Server MEP which is capable of generating FMS messages is for some reason unable to do so for the PW being signaled, a Notification message MUST be generated, with the Status Code set to "OAM Problem/ Unable to create fault management association".

- o T: Set timer value, enabled the configuration of a specific timer value. Default value is 0 (disabled).

- o Remaining bits: Reserved for future specification and set to 0.

Refresh Timer: indicates the refresh timer of fault indication messages, in seconds. The value MUST be between 1 to 20 seconds as specified for the Refresh Timer field in [RFC6427]. If the egress PE receiving the Label Mapping message cannot support the value it SHOULD reply with a higher timer value.

Fault Management Signal sub-TLV MAY include Traffic Class sub-TLV (Section 4.2.1.4). If TC sub-TLV is present, the value of the TC field MUST be used as the value of the TC field of a PW label stack entry for FMS messages. If the TC sub-TLV is absent, then selection of the TC value is local decision.

5. IANA Considerations

5.1. TLVs

IANA is requested to assign two new TLV types from the registry "TLV Type Name Space" in the "Label Distribution Protocol (LDP) Parameters" registry.

Value	TLV	References
TBD1	PW OAM Administration TLV	this document
TBD2	PW OAM Functions TLV	this document

The sub-TLV space and assignments for the PW OAM Functions TLV will be the same as that for the MPLS OAM Functions TLV. Sub-types for the MPLS OAM Functions TLV and the PW OAM Functions TLV MUST be kept the same. Any new sub-type added to the MPLS OAM Functions TLV MUST apply to the PW OAM Functions TLV as well.

5.1.1. PW OAM Configuration Sub-TLV

IANA is requested to create a registry of "Pseudowire OAM Configuration Sub-TLV types". These are 16 bit values. Sub-TLV types 1 through 8 are specified in this document. Sub-TLV types 0 and 65535 are reserved. Sub-TLV 9 through 65534 are to be assigned by IANA, using the "Expert Review" policy defined in [RFC5226].

Value	Sub-TLV	References
1	BFD Configuration sub-TLV	this document
2	Performance Monitoring sub-TLV	this document
3	Fault Management Signal sub-TLV	this document

5.1.1.1. BFD Configuration sub-TLVs

IANA is requested to create a registry of "Pseudowire OAM BFD Configuration Sub-TLV types". These are 16 bit values. Sub-TLV types 1 through 3 are specified in this document. Sub-TLV types 0 and 65535 are reserved. Sub-TLV 4 through 65534 are to be assigned by IANA, using the "Expert Review" policy defined in [RFC5226].

Value	Sub-TLV	References
1	Local Discriminator sub-TLV	this document
2	Negotiation Timer Parameters sub-TLV	this document
3	BFD Authentication sub-TLV	this document
4	Traffic Class Sub-TLV	this document

5.1.1.2. Performance Monitoring sub-TLVs

IANA is requested to create a registry of "Pseudowire OAM Performance Monitoring Sub-TLV types". These are 16 bit values. Sub-TLV types 1 through 2 are specified in this document. Sub-TLV types 0 and 65535 are reserved. Sub-TLV 3 through 65534 are to be assigned by IANA, using the "Expert Review" policy defined in [RFC5226].

Value	Sub-TLV	References
1	PM Loss TLV	this document
2	PM Delay TLV	this document

5.2. OAM Configuration Status Code

IANA is requested to assign the following LDP status codes from the registry "STATUS CODE NAME SPACE" in the "Label Distribution Protocol (LDP) Parameters" registry.

Range/Value	E	Description	Reference
TBD3	0	"PW OAM Alarms Enabled"	This document
TBD4	0	"PW OAM Alarms Disabled"	This document
TBD5	0	"PW MEP Configuration Failed"	This document
TBD6	0	"PW MEP Configuration Succeed"	This document
TBD7	0	"PW MEP Entities Disabled"	This document
TBD8	0	"PW MIP Configuration Failed"	This document
TBD9	0	"PW OAM Parameters Rejected"	This document
TBD10	0	"OAM Problem/Unsupported BFD Version"	This document
TBD11	0	"OAM Problem/Unsupported BFD Encapsulation format"	This document
TBD12	0	"OAM Problem/Unsupported BFD Authentication Type"	This document
TBD13	0	"OAM Problem/Mismatch of BFD Authentication Key ID	This document
TBD14	0	"OAM Problem/Unsupported Timestamp Format"	This document
TBD15	0	"OAM Problem/Unsupported Delay Mode"	This document
TBD16	0	"OAM Problem/Unsupported Loss Mode"	This document
TBD17	0	"OAM Problem/Delay variation unsupported"	This document
TBD18	0	"OAM Problem/Dyadic mode unsupported"	This document
TBD19	0	"OAM Problem/Loopback mode unsupported"	This document
TBD20	0	"OAM Problem/Combined mode unsupported"	This document
TBD21	0	"OAM Problem/Fault management signaling unsupported"	This document
TBD22	0	"OAM Problem/Unable to create fault management association"	This document

6. Security Considerations

Security considerations relating to LDP are described in section 5 of [RFC5036] and section 11 of [RFC5561]. Security considerations relating to use of LDP in setting up PWs is described in section 8 of [RFC4447].

This document defines new TLV/sub-TLV types, and OAM configuration procedures intended for use with MPLS-TP, which do not raise any additional security issues.

7. Acknowledgement

The authors would like to thank Andrew Malis, Greg Mirsky, Luca Martini, Matthew Bocci, Thomas Nadeau for their valuable comments and discussions, especially would like to thank Eric Gray for his review of this document.

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.

[RFC4447] Martini, L., Ed., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, DOI 10.17487/RFC4447, April 2006, <http://www.rfc-editor.org/info/rfc4447>.

[RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed., "LDP Specification", RFC 5036, DOI 10.17487/RFC5036, October 2007, <http://www.rfc-editor.org/info/rfc5036>.

[RFC5561] Thomas, B., Raza, K., Aggarwal, S., Aggarwal, R., and JL. Le Roux, "LDP Capabilities", RFC 5561, DOI 10.17487/RFC5561, July 2009, <http://www.rfc-editor.org/info/rfc5561>.

8.2. Informative References

[I-D.ietf-mpls-lsp-ping-mpls-tp-oam-conf] Bellagamba, E., Mirsky, G., Andersson, L., Skoldstrom, P., Ward, D., and J. Drake, "Configuration of Proactive Operations, Administration, and Maintenance (OAM) Functions for MPLS-based Transport Networks using LSP Ping", draft-ietf-mpls-lsp-ping-mpls-tp-oam-conf-09 (work in progress), January 2015.

[RFC3985] Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, DOI 10.17487/RFC3985, March 2005, <http://www.rfc-editor.org/info/rfc3985>.

[RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, DOI 10.17487/RFC4379, February 2006, <http://www.rfc-editor.org/info/rfc4379>.

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

[RFC5291] Chen, E. and Y. Rekhter, "Outbound Route Filtering Capability for BGP-4", RFC 5291, DOI 10.17487/RFC5291, August 2008, <http://www.rfc-editor.org/info/rfc5291>.

[RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", RFC 5462, DOI 10.17487/RFC5462, February 2009, <http://www.rfc-editor.org/info/rfc5462>.

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

[RFC5659] Bocci, M. and S. Bryant, "An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge", RFC 5659, DOI 10.17487/RFC5659, October 2009, <http://www.rfc-editor.org/info/rfc5659>.

[RFC5860] Vigoureux, M., Ed., Ward, D., Ed., and M. Betts, Ed., "Requirements for Operations, Administration, and Maintenance (OAM) in MPLS Transport Networks", RFC 5860, DOI 10.17487/RFC5860, May 2010, <http://www.rfc-editor.org/info/rfc5860>.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <http://www.rfc-editor.org/info/rfc5880>.

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

[RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, DOI 10.17487/RFC6374, September 2011, <http://www.rfc-editor.org/info/rfc6374>.

[RFC6375] Frost, D., Ed. and S. Bryant, Ed., "A Packet Loss and Delay Measurement Profile for MPLS-Based Transport Networks", RFC 6375, DOI 10.17487/RFC6375, September 2011, <http://www.rfc-editor.org/info/rfc6375>.

[RFC6427] Swallow, G., Ed., Fulignoli, A., Ed., Vigoureux, M., Ed., Boutros, S., and D. Ward, "MPLS Fault Management Operations, Administration, and Maintenance (OAM)", RFC 6427, DOI 10.17487/RFC6427, November 2011, <http://www.rfc-editor.org/info/rfc6427>.

[RFC7260] Takacs, A., Fedyk, D., and J. He, "GMPLS RSVP-TE Extensions for Operations, Administration, and Maintenance (OAM) Configuration", RFC 7260, DOI 10.17487/RFC7260, June 2014, <http://www.rfc-editor.org/info/rfc7260>.

[RFC7487] Bellagamba, E., Takacs, A., Mirsky, G., Andersson, L., Skoldstrom, P., and D. Ward, "Configuration of Proactive Operations, Administration, and Maintenance (OAM) Functions for MPLS-Based Transport Networks Using RSVP-TE", RFC 7487, DOI 10.17487/RFC7487, March 2015, <http://www.rfc-editor.org/info/rfc7487>.

Authors’ Addresses

Fei Zhang (editor)
Huawei

Email: zhangfei7@huawei.com

Bo Wu (editor)
ZTE Corporation

Email: wu.bo@zte.com.cn

Elisa Bellagamba (editor)
Ericsson
Farogatan 6
Kista, 164 40
Sweden

Phone: +46 761440785
Email: elisa.bellagamba@ericsson.com

Mach(Guoyi) Chen (editor)
Huawei

Email: mach.chen@huawei.com

Network Working Group M. Chen
Internet-Draft W. Cao
Intended status: Standards Track Huawei
Expires: January 22, 2017 A. Takacs
Ericsson
P. Pan
July 21, 2016

LDP Extensions for Pseudowire Binding to LSP Tunnels
draft-ietf-pals-mpls-tp-pw-over-bidir-lsp-09.txt

Abstract

Many transport services require that user traffic, in the form of Pseudowires (PW), be delivered via either a single co-routed bidirectional tunnel or two unidirectional tunnels that share the same routes. This document defines an optional extension to Label Distribution Protocol (LDP) that enables the binding between PWs and the underlying Traffic Engineering (TE) tunnels. The extension applies to both single-segment and multi-segment PWs.

Requirements Language

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

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Pseudo Wire Emulation Edge-to-Edge (PWE3) [RFC3985] is a mechanism to emulate layer 2 services, such as Ethernet Point-to-Point circuits. Such services are emulated between two Attachment Circuits, and the Pseudowire (PW)-encapsulated layer 2 service payload is transported via Packet Switching Network (PSN) tunnels between Provider Edges (PEs). PWE3 typically uses Label Distribution Protocol (LDP) [RFC5036] or Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) [RFC3209] LSPs as PSN tunnels. The PEs select and bind the Pseudowires to PSN tunnels independently. Today, there is no standardized protocol-based provisioning mechanism to associate PWs to PSN tunnels, such associations must be managed via provisioning or other private methods.

PW-to-PSN Tunnel binding has become increasingly common and important in many deployment scenarios, , as it allows service providers to provide service level agreements to their customers for such traffic attributes as bandwidth, latency, and availability.

The requirements for explicit control of PW-to-LSP mapping has been described in Section 5.3.2 of [RFC6373]. Figure 1 illustrates how PWs can be bound to particular LSPs.

```
          +------+                  +------+
---AC1 ---|..............PWs...............|---AC1---
---...----| PE1  |=======LSPs=======| PE2  |---...---
---ACn ---|      |-------Links------|      |---ACn---
          +------+                  +------+
```

Figure 1: Explicit PW-to-LSP binding scenario

There are two PEs (PE1 and PE2) connected through multiple parallel links that may be on different physical fibers. Each link is managed and controlled as a bi-directional LSP. At each PE, there are a large number of bi-directional user flows from multiple Ethernet interfaces (access circuits in the figure). Each user flow uses a pair of uni-directional PWs to carry bi-directional traffic. The operators need to make sure that the user flows (that is, the PW-pairs) are carried on the same fiber or bidirectional LSP.

There are a number of reasons behind this requirement. First, due to delay and latency constraints, traffic going over different fibers may require a large amount of expensive buffer memory to compensate for the differential delay at the headend nodes. Further, the operators may apply different protection mechanisms on different parts of the network (e.g., to deloy 1:1 protection in one part and 1+1 protection in other parts). As such, operators may prefer to have a user’s traffic traverse the same fiber. That implies that both forwarding and reserve direction PWs that belong to the same user flow need to be mapped to the same co-routed bi-directional LSP or two LSPs with the same route.

Figure 2 illustrates a scenario where PW-LSP binding is not applied.

```
                +----+   +--+ LSP1 +--+   +----+
     +-----+    | PE1|===|P1|======|P2|===| PE2|    +-----+
     |     |----|    |   +--+      +--+   |    |----|     |
     | CE1 |    |............PW...............|    | CE2 |
     |     |----|    |      +--+          |    |----|     |
     +-----+    |    |======|P3|==========|    |    +-----+
                +----+      +--+ LSP2     +----+
```

Figure 2: Inconsistent SS-PW to LSP binding scenario

LSP1 and LSP2 are two bidirectional connections on diverse paths. The operator needs to deliver a bi-directional flow between PE1 and PE2. Using existing mechanisms, it's possible that PE1 may select LSP1 (PE1-P1-P2-PE2) as the PSN tunnel for traffic from PE1 to PE2, while selecting LSP2 (PE2-P3-PE1) as the PSN tunnel for traffic from PE2 to PE1.

Consequently, the user traffic is delivered over two disjoint LSPs that may have very different service attributes in terms of latency and protection. This may not be acceptable as a reliable and effective transport service to the customer.

A similar problem may also exist in multi-segment PWs (MS-PWs), where user traffic on a particular PW may hop over different networks on forward and reverse directions.

One way to solve this problem is by introducing manual provisioning at each PE to bind the PWs to the underlying PSN tunnels. However, this is prone to configuration errors and does not scale.

This document introduces an automatic solution by extending Forwarding Equivalence Class (FEC) 128/129 PW based on [RFC4447].

2. LDP Extensions

This document defines a new optional TLV, PSN Tunnel Binding TLV, to communicate tunnel/LSPs selection and binding requests between PEs. The TLV carries a PW's binding profile and provides explicit or implicit information for the underlying PSN tunnel binding operation.

The binding operation applies in both single-segment (SS) and multi-segment (MS) scenarios.

The extension supports two types of binding requests:

1. Strict binding: the requesting PE will choose and explicitly indicate the LSP information in the requests; the receiving PE

MUST obey the requests, otherwise, the PW will not be established.

2. Co-routed binding: the requesting PE will suggest an underlying LSP to a remote PE. On receive, the remote PE has the option to use the suggested LSP, or reply the information for an alternative.

In this document, the terminology of "tunnel" is identical to the "TE Tunnel" defined in Section 2.1 of [RFC3209], which is uniquely identified by a SESSION object that includes Tunnel end point address, Tunnel ID and Extended Tunnel ID. The terminology "LSP" is identical to the "LSP tunnel" defined in Section 2.1 of [RFC3209], which is uniquely identified by the SESSION object together with SENDER_TEMPLATE (or FILTER_SPEC) object that consists of LSP ID and Tunnel endpoint address.

2.1. PSN Tunnel Binding TLV

PSN Tunnel Binding TLV is an optional TLV and MUST be carried in the LDP Label Mapping message [RFC5036] if PW to LSP binding is required. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|U|F| PSN Tunnel Binding (TBD1) |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|C|S|T|    Unallocated flags    |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                        PSN Tunnel Sub-TLV                     ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: PSN Tunnel Binding TLV

The U-bit and F-bit are defined in Section 3.3 [RFC5036]. Since the PSN Tunnel Binding TLV is an optional TLV, the U-bit MUST be set to 1 so that a receiver MUST silently ignore this TLV if unknown to it, and continue processing the rest of the message.

A receiver of this TLV is not allowed to forward the TLV further when it does not know the TLV. So, the F-bit MUST be set to 0.

The PSN Tunnel Binding TLV type is TBD1.

The Length field is 2 octets in length. It defines the length in octets of the value field (including Flags, Reserved, sub-TLV fields).

The Flag field is 2 octets in length, three flags are defined in this document. The rest unallocated flags MUST be set to zero when sending, and MUST be ignored when received.

C (Co-routed path) bit: This informs the remote T-PE/S-PEs about the properties of the underlying LSPs. When set, the remote T-PE/S-PEs SHOULD select co-routed LSP (as the forwarding tunnel) as the reverse PSN tunnel. If there is no such tunnel available, it may trigger the remote T-PE/S-PEs to establish a new LSP.

S (Strict) bit: This instructs the PEs with respect to the handling of the underlying LSPs. When set, the remote PE MUST use the tunnel/LSP specified in the PSN Tunnel Sub-TLV as the PSN tunnel on the reverse direction of the PW, or the PW will fail to be established.

Either the C-bit or the S-bit MUST be set. The C-bit and S-bit are mutually exclusive from each other, and cannot be set in the same message. If "both C-bit and S-bit are set" or "both C-bit and S-bit are clear" are received, a Label Release message with status code set to "The C-bit or S-bit unknown" (TBD5) MUST be replied, and the PW will not be established.

T (Tunnel Representation) bit: This indicates the format of the LSP tunnels. When the bit is set, the tunnel uses the tunnel information to identify itself, and the LSP Number fields in the PSN Tunnel sub-TLV (Section 2.1.1) MUST be set to zero. Otherwise, both tunnel and LSP information of the PSN tunnel are required. The default is set. The motivation for the T-bit is to support the MPLS protection operation where the LSP Number fields may be ignored.

The Reserved field is 2 octets in length and is left for future use.

2.1.1. PSN Tunnel Sub-TLV

PSN Tunnel Sub-TLVs are designed for inclusion in the PSN Tunnel Binding TLV to specify the tunnel/LSPs to which a PW is required to bind.

Two sub-TLVs are defined: the IPv4 and IPv6 Tunnel sub-TLVs.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type(TBD2)  |    Length     |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Source Global ID                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Source Node ID                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Source Tunnel Number     |     Source LSP Number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Destination Global ID                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Destination Node ID                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Destination Tunnel Number   |    Destination LSP Number     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 0                   1                   2                   3
```

Figure 4: IPv4 PSN Tunnel sub-TLV 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type(TBD3)  |    Length     |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Source Global ID                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                        Source Node ID                         ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Source Tunnel Number     |       Source LSP Number       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Destination Global ID                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                      Destination Node ID                      ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Destination Tunnel Number   |    Destination LSP Number     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: IPv6 PSN Tunnel sub-TLV format

The definition of Source and Destination Global/Node IDs and Tunnel/ LSP Numbers are derived from [RFC6370]. This is to describe the underlying LSPs. Note that the LSPs in this notation are globally unique. The ITU-T style identifiers [RFC6923] are not used in this document.

As defined in Section 4.6.1.2 and Section 4.6.2.2 of [RFC3209], the "Tunnel endpoint address" is mapped to Destination Node ID, and "Extended Tunnel ID" is mapped to Source Node ID. Both IDs can be IPv4 or IPv6 addresses. The Node IDs are routable addresses of the ingress LSR and egress LSR of the Tunnel/LSP.

A PSN Tunnel sub-TLV could be used to either identify a tunnel or a specific LSP. The T-bit in the Flag field defines the distinction as such that, when the T-bit is set, the Source/Destination LSP Number fields MUST be zero and ignored during processing. Otherwise, both Source/Destination LSP Number fields MUST have the actual LSP IDs of specific LSPs.

Each PSN Tunnel Binding TLV MUST only have one such sub-TLV. When sending, only one sub-TLV MUST be carried. When received, if there are more than one sub-TLVs carried, only the first sub-TLV MUST be used, the rest sub-TLVs MUST be ignored.

3. Theory of Operation

During PW setup, the PEs may choose to select desired forwarding tunnels/LSPs, and inform the remote T-PE/S-PEs about the desired reverse tunnels/LSPs.

Specifically, to set up a PW (or PW Segment), a PE may select a candidate tunnel/LSP to act as the PSN tunnel. If none is available or satisfies the constraints, the PE will trigger and establish a new tunnel/LSP. The selected tunnel/LSP information is carried in the PSN Tunnel Binding TLV and sent with the Label Mapping message to the target PE.

Upon the reception of the Label Mapping message, the receiving PE will process the PSN Tunnel Binding TLV, determine whether it can accept the suggested tunnel/LSP or to find the reverse tunnel/LSP that meets the request, and respond with a Label Mapping message, which contains the corresponding PSN Tunnel Binding TLV.

It is possible that two PEs may request PSN binding to the same PW or PW segment over different tunnels/LSPs at the same time. There may cause collisions of tunnel/LSPs selection as both PEs assume the active role.

As defined in (Section 7.2.1, [RFC6073]), each PE may be categorized into active and passive roles:

1. Active PE: the PE which initiates the selection of the tunnel/ LSPs and informs the remote PE;

2. Passive PE: the PE which obeys the active PE’s suggestion.

In the remaining of this document, we will elaborate the operation for SS-PW and MS-PW:

1. SS-PW: In this scenario, both PEs for a particular PW may assume the active roles.

2. MS-PW: One PE is active, while the other is passive. The PWs are setup using FEC 129.

4. PSN Binding Operation for SS-PW

As illustrated in Figure-5, both PEs (say, PE1 and PE2) of a PW may independently initiate the setup. To perform PSN binding, the Label Mapping messages MUST carry a PSN Tunnel Binding TLV, and the PSN Tunnel sub-TLV MUST contains the desired tunnel/LSPs of the sender.

```
                  +----+        LSP1        +----+
         +-----+  | PE1|====================| PE2|    +-----+
         |     |----|    |                    |    |----|     |
         | CE1 |    |............PW...............|    | CE2 |
         |     |----|    |                    |    |----|     |
         +-----+    |    |====================|    |    +-----+
                  +----+       LSP2         +----+
```

Figure 6: PSN binding operation in SS-PW environment

As outlined previously, there are two types of binding request: co-routed and strict.

In strict binding, a PE (e.g., PE1) will mandate the other PE (e.g., PE2) to use a specified tunnel/LSP (e.g. LSP1) as the PSN tunnel on the reverse direction. In the PSN Tunnel Binding TLV, the S-bit MUST be set, the C-bit MUST be cleared, and the Source and Destination IDs/Numbers MUST be filled.

On receive, if the S-bit is set, as well as following the processing procedure defined in Section 5.3.3 of [RFC4447], the receiving PE (i.e. PE2) needs to determine whether to accept the indicated tunnel/LSP in PSN Tunnel Sub-TLV.

If the receiving PE (PE2) is also an active PE, and may have initiated the PSN binding requests to the other PE (PE1), if the received PSN tunnel/LSP is the same as it has been sent in the Label Mapping message by PE2, then the signaling has converged on a mutually agreed Tunnel/LSP. The binding operation is completed.

Otherwise, the receiving PE (PE2) MUST compare its own Node ID against the received Source Node ID as unsigned integers. If the received Source Node ID is larger, the PE (PE2) will reply with a Label Mapping message to complete the PW setup and confirm the binding request. The PSN Tunnel Binding TLV in the message MUST contain the same Source and Destination IDs/Numbers as in the received binding request, in the appropriate order (where the source is PE2 and PE1 becomes the destination). On the other hand, if the receiving PE (PE2) has a Node ID that is larger than the Source Node ID carried in the PSN Tunnel Binding TLV, it MUST reply with a Label Release message with status code set to "Reject - unable to use the suggested tunnel/LSPs" and the received PSN Tunnel Binding TLV, and the PW will not be established.

To support co-routed binding, the receiving PE can select the appropriated PSN tunnel/LSP for the reverse direction of the PW, so long as the forwarding and reverse PSNs share the same route (links and nodes).

Initially, a PE (PE1) sends a Label Mapping message to the remote PE (PE2) with the PSN Tunnel Binding TLV, with C-bit set, S-bit cleared, and the appropriate Source and Destination IDs/Numbers. In case of unidirectional LSPs, the PSN Tunnel Binding TLV may only contain the Source IDs/Numbers, the Destination IDs/Numbers are set to zero and left for PE2 to complete when responding the Label Mapping message.

On receive, since PE2 is also an active PE, and may have initiated the PSN binding requests to the other PE (PE1), if the received PSN tunnel/LSP has the same route as the one that has been sent in the Label Mapping message to PE1, then the signaling has converged. The binding operation is completed.

Otherwise, PE2 needs to compare its own Node ID against the received Source Node ID as unsigned integers. If the received Source Node ID is larger, PE2 needs to find/establish a tunnel/LSP that meets the co-routed constraint, and reply with a Label Mapping message with a PSN Binding TLV that contains the Source and Destination IDs/Numbers of the tunnel/LSP. On the other hand, if the receiving PE (PE2) has a Node ID that is larger than the Source Node ID carried in the PSN Tunnel Binding TLV, it MUST reply with a Label Release message with status code set to "Reject - unable to use the suggested tunnel/LSPs" (TBD4) and the received PSN Tunnel Binding TLV.

In addition, if the received PSN tunnel/LSP end points do not match the PW end points, PE2 MUST reply with a Label Release message with status code set to "Reject - unable to use the suggested tunnel/LSPs" (TBD4) and the received PSN Tunnel Binding TLV MUST also be carried.

In both strict and co-routed bindings, if T-bit is set, the LSP Number field MUST be set to zero. Otherwise, the field MUST contain the actual LSP number for the related PSN LSP.

After a PW is established, the operators may choose to move the PWs from the current tunnel/LSPs to other tunnel/LSPs. Also the underlying PSN tunnel may break due to a network failure. When either of these scenarios occur, a new Label Mapping message MUST be sent to notify the remote PE of the changes. Note that when the T-bit is set, the working LSP broken will not provide this update if there are protection LSPs in place.

The message may carry a new PSN Tunnel Binding TLV, which contains the new Source and Destination Numbers/IDs. The handling of the new message should be identical to what has been described in this section.

However, if the new Label Mapping message does not contain the PSN Tunnel Binding TLV, it declares the removal of any co-routed/strict constraints. The current independent PW to PSN binding will be used.

Further, as an implementation option, the PEs may choose not to remove the traffic from an operational PW, until the completion of the underlying PSN tunnel/LSP changes.

5. PSN Binding Operation for MS-PW

MS-PW uses FEC 129 for PW setup. We refer the operation to Figure-6.

```
             +-----+ LSP1 +-----+ LSP2 +-----+ LSP3 +-----+
     +---+   |T-PE1|======|S-PE1|======|S-PE2|======|T-PE2|   +---+
     |   |---|     |      |     |      |     |      |     |---|   |
     |CE1|   |......................PW....................|   |CE2|
     |   |---|     |      |     |      |     |      |     |---|   |
     +---+   |     |======|     |======|     |======|     |   +---+
             +-----+ LSP4 +-----+ LSP5 +-----+ LSP6 +-----+
```

Figure 7: PSN binding operation in MS-PW environment

When an active PE (that is, T-PE1) starts to signal a MS-PW, a PSN Tunnel Binding TLV MUST be carried in the Label Mapping message and sent to the adjacent S-PE (that is, S-PE1). The PSN Tunnel Binding TLV includes the PSN Tunnel sub-TLV that carries the desired tunnel/LSP of T-PE1's.

For strict binding, the initiating PE MUST set the S-bit, clear the C-bit and indicate the binding tunnel/LSP to the next-hop S-PE.

When S-PE1 receives the Label Mapping message, S-PE1 needs to determine if the signaling is for forward or reverse direction, as defined in Section 6.2.3 of [RFC7267].

If the Label Mapping message is for forward direction, and S-PE1 accepts the requested tunnel/LSPs from T-PE1, S-PE1 MUST save the tunnel/LSP information for reverse-direction processing later on. If the PSN binding request is not acceptable, S-PE1 MUST reply with a Label Release Message to the upstream PE (T-PE1) with Status Code set to "Reject - unable to use the suggested tunnel/LSPs" (TBD4).

Otherwise, S-PE1 relays the Label Mapping message to the next S-PE (that is, S-PE2), with the PSN Tunnel sub-TLV carrying the information of the new PSN tunnel/LSPs selected by S-PE1. S-PE2 and subsequent S-PEs will repeat the same operation until the Label Mapping message reaches to the remote T-PE (that is, T-PE2).

If T-PE2 agrees with the requested tunnel/LSPs, it will reply with a Label Mapping message to initiate to the binding process on the reverse direction. The Label Mapping message contains the received PSN Tunnel Binding TLV for confirmation purposes.

When its upstream S-PE (S-PE2) receives the Label Mapping message, the S-PE relays the Label Mapping message to its upstream adjacent S-PE (S-PE1), with the previously saved PSN tunnel/LSP information in the PSN Tunnel sub-TLV. The same procedure will be applied on subsequent S-PEs, until the message reaches to T-PE1 to complete the PSN binding setup.

During the binding process, if any PE does not agree to the requested tunnel/LSPs, it can send a Label Release Message to its upstream adjacent PE with Status Code set to "Reject - unable to use the suggested tunnel/LSPs" (TBD4). In addition, if the received PSN tunnel/LSP end points do not match the PW Segment end points, the receiving PE MUST reply with a Label Release message with status code set to "Reject - unable to use the suggested tunnel/LSPs" (TBD4) and the received PSN Tunnel Binding TLV MUST also be carried.

For co-routed binding, the initiating PE (T-PE1) MUST set the C-bit, reset the S-bit and indicates the suggested tunnel/LSP in PSN Tunnel sub-TLV to the next-hop S-PE (S-PE1).

During the MS-PW setup, the PEs have the option of ignoring the suggested tunnel/LSP, and to select another tunnel/LSP for the segment PW between itself and its upstream PE in reverse direction only if the tunnel/LSP is co-routed with the forward one. Otherwise, the procedure is the same as the strict binding.

The tunnel/LSPs may change after a MS-PW being established. When a tunnel/LSP has changed, the PE that detects the change SHOULD select an alternative tunnel/LSP for temporary use while negotiating with other PEs following the procedure described in this section.

6. PSN Tunnel Select Considerations

As stated in Section 1 of this document, the PSN tunnel that is used for binding can be either a co-routed bi-directional LSP or two LSPs with the same route. The co-routed bi-directional LSP has the characteristics that both directions not only cross the same nodes and links but have the same life span. But for the two LSPs case, even if they have the same route at the beginning, it cannot be guaranteed that they will always have the same route all the time. For example, when Fast ReRoute (FRR) [RFC4090] is deployed for the LSPs, link or node failure may make the two LSPs use different routes. So, if the network supports co-routed bi-directional LSPs, it is RECOMMENDED that a co-routed bi-directional LSP should be used; otherwise, two LSPs with same route may be used.

7. Security Considerations

The ability to control which LSP is used to carry traffic from a PW can be a potential security risk both for denial of service and traffic interception. It is RECOMMENDED that PEs do not accept the use of LSPs identified in the PSN Tunnel Binding TLV unless the LSP end points match the PW or PW segment end points. Furthermore, it is RECOMMENDED that PEs implement the LDP security mechanisms described in [RFC5036] and [RFC5920].

8. IANA Considerations

8.1. LDP TLV Types

This document defines a new TLV [Section 2.1 of this document] for inclusion in LDP Label Mapping message. IANA is requested to assign TLV type value (TBD1) to the new defined TLVs from LDP "TLV Type Name Space" registry.

8.1.1. PSN Tunnel Sub-TLVs

This document defines two sub-TLVs [Section 2.1.1 of this document] for PSN Tunnel Binding TLV. IANA is required to create a new PWE3 registry ("PSN Tunnel Sub-TLV Name Space") for PSN Tunnel sub-TLVs and to assign Sub-TLV type values to the following sub-TLVs:

IPv4 PSN Tunnel sub-TLV - TBD2

IPv6 PSN Tunnel sub-TLV - TBD3

In addition, the values 0 and 255 in this new registry should be reserved, and values 1-254 will be allocated by IETF Review.

8.2. LDP Status Codes

This document defines two new LDP status codes, IANA is requested to assigned status codes to these new defined codes from the LDP "STATUS CODE NAME SPACE" registry.

"Reject - unable to use the suggested tunnel/LSPs" - TBD4

"The C-bit or S-bit unknown" -TBD5

The E bit is set to one for both new codes.

9. Acknowledgements

The authors would like to thank Adrian Farrel, Kamran Raza, Xinchun Guo, Mingming Zhu and Li Xue for their comments and help in preparing this document. Also this draft benefits from the discussions with Nabil Bitar, Paul Doolan, Frederic Journay, Andy Malis, Curtis Villamizar, Luca Martini, Alexander Vainshtein, Huub van Helvoort, Daniele Ceccarelli and Stewart Byant.

We would especially like to acknowledge Ping Pan, a co-author on the early versions of this document. It was a privilege to have known him.

The coauthors of this document, the working group chairs, the responsible AD, and the PALS Working Group wish to dedicate this RFC to the memory of our friend and colleague Ping Pan, in recognition for his devotion and hard work at the IETF.

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4447] Martini, L., Ed., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, DOI 10.17487/RFC4447, April 2006, <http://www.rfc-editor.org/info/rfc4447>.

[RFC6370] Bocci, M., Swallow, G., and E. Gray, "MPLS Transport Profile (MPLS-TP) Identifiers", RFC 6370, DOI 10.17487/RFC6370, September 2011, <http://www.rfc-editor.org/info/rfc6370>.

10.2. Informative References

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

[RFC3985] Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, DOI 10.17487/RFC3985, March 2005, <http://www.rfc-editor.org/info/rfc3985>.

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

[RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed., "LDP Specification", RFC 5036, DOI 10.17487/RFC5036, October 2007, <http://www.rfc-editor.org/info/rfc5036>.

[RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, DOI 10.17487/RFC5920, July 2010, <http://www.rfc-editor.org/info/rfc5920>.

[RFC6073] Martini, L., Metz, C., Nadeau, T., Bocci, M., and M. Aissaoui, "Segmented Pseudowire", RFC 6073, DOI 10.17487/RFC6073, January 2011, <http://www.rfc-editor.org/info/rfc6073>.

[RFC6373] Andersson, L., Ed., Berger, L., Ed., Fang, L., Ed., Bitar, N., Ed., and E. Gray, Ed., "MPLS Transport Profile (MPLS-TP) Control Plane Framework", RFC 6373, DOI 10.17487/RFC6373, September 2011, <http://www.rfc-editor.org/info/rfc6373>.

[RFC6923] Winter, R., Gray, E., van Helvoort, H., and M. Betts, "MPLS Transport Profile (MPLS-TP) Identifiers Following ITU-T Conventions", RFC 6923, DOI 10.17487/RFC6923, May 2013, <http://www.rfc-editor.org/info/rfc6923>.

[RFC7267] Martini, L., Ed., Bocci, M., Ed., and F. Balus, Ed., "Dynamic Placement of Multi-Segment Pseudowires", RFC 7267, DOI 10.17487/RFC7267, June 2014, <http://www.rfc-editor.org/info/rfc7267>.

Authors' Addresses

Mach(Guoyi) Chen
Huawei

Email: mach.chen@huawei.com

Wei Cao
Huawei

Email: wayne.caowei@huawei.com

Attila Takacs
Ericsson
Laborc u. 1.
Budapest 1037
Hungary

Email: attila.takacs@ericsson.com

Ping Pan

Internet Engineering Task Force
Internet Draft
Intended status: Internet Standard
Expires: January 5, 2017
Obsoletes: 6723, 4447

Luca Martini Ed.
Giles Heron Ed.

Cisco

July 5, 2016

Pseudowire Setup and Maintenance using the Label Distribution Protocol

draft-ietf-pals-rfc4447bis-05.txt

Status of this Memo

Abstract

Layer 2 services (such as Frame Relay, Asynchronous Transfer Mode, and Ethernet) can be "emulated" over an MPLS backbone by encapsulating the Layer 2 Protocol Data Units (PDU) and then transmitting them over "pseudowires". It is also possible to use pseudowires to provide low-rate Time Division Multiplexed and

Synchronous Optical NETworking circuit emulation over an MPLS-enabled network. This document specifies a protocol for establishing and maintaining the pseudowires, using extensions to the Label Distribution Protocol (LDP). Procedures for encapsulating Layer 2 PDUs are specified in a set of companion documents.

This document has been written to address errata in a previous version of this standard.

Table of Contents

1. Introduction

[RFC4619], [RFC4717], [RFC4618], and [RFC4448] explain how to encapsulate a Layer 2 Protocol Data Unit (PDU) for transmission over an MPLS-enabled network. Those documents specify that a "pseudowire header", consisting of a demultiplexor field, will be prepended to the encapsulated PDU. The pseudowire demultiplexor field is prepended before transmitting a packet on a pseudowire. When the packet arrives at the remote endpoint of the pseudowire, the demultiplexor is what enables the receiver to identify the particular pseudowire on which the packet has arrived. To transmit the packet from one pseudowire endpoint to another, the packet may need to travel through a "Packet Switched Network (PSN) tunnel"; this will require that an additional header be prepended to the packet.

Accompanying documents [RFC4842], [RFC4553] specify methods for transporting time-division multiplexing (TDM) digital signals (TDM circuit emulation) over a packet-oriented MPLS-enabled network. The transmission system for circuit-oriented TDM signals is the Synchronous Optical Network [ANSI] (SONET)/Synchronous Digital Hierarchy (SDH) [ITUG]. To support TDM traffic, which includes voice, data, and private leased-line service, the pseudowires must emulate the circuit characteristics of SONET/SDH payloads. The TDM signals and payloads are encapsulated for transmission over pseudowires. A pseudowire demultiplexor and a PSN tunnel header is prepended to this encapsulation.

[RFC4553] describes methods for transporting low-rate time-division multiplexing (TDM) digital signals (TDM circuit emulation) over PSNs, while [RFC4842] similarly describes transport of high-rate TDM (SONET/SDH). To support TDM traffic, the pseudowires must emulate the circuit characteristics of the original T1, E1, T3, E3, SONET, or SDH signals. [RFC4553] does this by encapsulating an arbitrary but constant amount of the TDM data in each packet, and the other methods encapsulate TDM structures.

In this document, we specify the use of the MPLS Label Distribution Protocol, LDP [RFC5036], as a protocol for setting up and maintaining the pseudowires. In particular, we define new TLVs, FEC elements, parameters, and codes for LDP, which enable LDP to identify pseudowires and to signal attributes of pseudowires. We specify how a pseudowire endpoint uses these TLVs in LDP to bind a demultiplexor field value to a pseudowire, and how it informs the remote endpoint of the binding. We also specify procedures for reporting pseudowire status changes, for passing additional information about the pseudowire as needed, and for releasing the bindings. These procedures are intended to be independent of the underlying version of IP used for LDP signaling.

In the protocol specified herein, the pseudowire demultiplexor field is an MPLS label. Thus, the packets that are transmitted from one end of the pseudowire to the other are MPLS packets, which must be transmitted through an MPLS tunnel. However, if the pseudowire endpoints are immediately adjacent and penultimate hop popping behavior is in use, the MPLS tunnel may not be necessary. Any sort of PSN tunnel can be used, as long as it is possible to transmit MPLS packets through it. The PSN tunnel can itself be an MPLS LSP, or any other sort of tunnel that can carry MPLS packets. Procedures for setting up and maintaining the MPLS tunnels are outside the scope of this document.

This document deals only with the setup and maintenance of point-to-point pseudowires. Neither point-to-multipoint nor multipoint-to-point pseudowires are discussed.

QoS-related issues are not discussed in this document.

The following two figures describe the reference models that are derived from [RFC3985] to support the PW emulated services.

```
         |<-------------- Emulated Service ---------------->|
         |                                                  |
         |          |<------- Pseudowire ------->|          |
         |          |                            |          |
         |Attachment|    |<-- PSN Tunnel -->|    |Attachment|
         |  Circuit V    V                  V    V  Circuit |
         V   (AC)   +----+                  +----+   (AC)   V
   +-----+    |     | PE1|==================| PE2|     |    +-----+
   |     |----------|............PW1.............|----------|     |
   | CE1 |    |     |    |                  |    |     |    | CE2 |
   |     |----------|............PW2.............|----------|     |
   +-----+  ^ |     |    |==================|    |     | ^  +-----+
         ^  |       +----+                  +----+     | |  ^
         |  |   Provider Edge 1         Provider Edge 2  |  |
         |  |                                            |  |
   Customer |                                            | Customer
   Edge 1   |                                            | Edge 2
            |                                            |
      native service                               native service
```

Figure 1: PWE3 Reference Model

```
   +-----------------+                            +-----------------+
   |Emulated Service |                            |Emulated Service |
   |(e.g., TDM, ATM) |<==== Emulated Service ===>|(e.g., TDM, ATM) |
   +-----------------+                            +-----------------+
   |    Payload      |                            |    Payload      |
   |  Encapsulation  |<====== Pseudowire =======>|  Encapsulation  |
   +-----------------+                            +-----------------+
   |PW Demultiplexer |                            |PW Demultiplexer |
   |   PSN Tunnel,   |<======= PSN Tunnel ======>|  PSN Tunnel,    |
   | PSN & Physical  |                            | PSN & Physical  |
   |     Layers      |                            |    Layers       |
   +-------+---------+        ___________         +---------+-------+
           |                /                               |
           +===============/     PSN       ===============+
                                         /
                            _____________/
```

Figure 2: PWE3 Protocol Stack Reference Model

For the purpose of this document, PE1 will be defined as the ingress router, and PE2 as the egress router. A layer 2 PDU will be received at PE1, encapsulated at PE1, transported and decapsulated at PE2, and transmitted out of PE2.

2. Changes from RFC4447

The changes in this document are mostly minor fixes to spelling and grammar, or clarifications to the text, which were either noted as errata to [RFC4447] or found by the editors.

Additionally a new section (7.3) on control-word renegotiation by label request message has been added, obsoleting [RFC6723]. The diagram of C-bit handling procedures has also been removed. A note has been added in section 6.3.2 to clarify that the C-bit is part of the FEC.

A reference has also been added to [RFC7358] indicating the use of downstream unsolicited mode to distribute PW FEC label bindings, independent of the negotiated label advertisement mode of the LDP session.

3. 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].

4. The Pseudowire Label

Suppose that it is desired to transport Layer 2 PDUs from ingress LSR PE1 to egress LSR PE2, across an intervening MPLS-enabled network. We assume that there is an MPLS tunnel from PE1 to PE2. That is, we assume that PE1 can cause a packet to be delivered to PE2 by encapsulating the packet in an "MPLS tunnel header" and sending the result to one of its adjacencies. The MPLS tunnel is an MPLS Label Switched Path (LSP); thus, putting on an MPLS tunnel encapsulation is a matter of pushing on an MPLS label.

We presuppose that a large number of pseudowires can be carried through a single MPLS tunnel. Thus it is never necessary to maintain state in the network core for individual pseudowires. We do not presuppose that the MPLS tunnels are point to point; although the pseudowires are point to point, the MPLS tunnels may be multipoint to point. We do not presuppose that PE2 will even be able to determine the MPLS tunnel through which a received packet was transmitted. (For example, if the MPLS tunnel is an LSP and penultimate hop popping is used, when the packet arrives at PE2, it will contain no information identifying the tunnel.)

When PE2 receives a packet over a pseudowire, it must be able to determine that the packet was in fact received over a pseudowire, and it must be able to associate that packet with a particular pseudowire. PE2 is able to do this by examining the MPLS label that serves as the pseudowire demultiplexor field shown in Figure 2. Call this label the "PW label".

When PE1 sends a Layer 2 PDU to PE2, it creates an MPLS packet by adding the PW label to the packet, thus creating the first entry of the label stack. If the PSN tunnel is an MPLS LSP, the PE1 pushes another label (the tunnel label) onto the packet as the second entry of the label stack. The PW label is not visible again until the MPLS packet reaches PE2. PE2’s disposition of the packet is based on the PW label.

If the payload of the MPLS packet is, for example, an ATM AAL5 PDU, the PW label will generally correspond to a particular ATM VC at PE2. That is, PE2 needs to be able to infer from the PW label the outgoing interface and the VPI/VCI value for the AAL5 PDU. If the payload is a Frame Relay PDU, then PE2 needs to be able to infer from the PW label the outgoing interface and the DLCI value. If the payload is an Ethernet frame, then PE2 needs to be able to infer from the PW label the outgoing interface, and perhaps the VLAN identifier. This process is uni-directional and will be repeated independently for bi-directional operation. When using the PWid FEC Element, it is REQUIRED that the same PW ID and PW type be assigned for a given circuit in both directions. The group ID (see below) MUST NOT be required to match in both directions. The transported frame MAY be modified when it reaches the egress router. If the header of the transported Layer 2 frame is modified, this MUST be done at the egress LSR only. Note that the PW label must always be at the bottom of the packet's label stack, and labels MUST be allocated from the per-platform label space.

This document does not specify a method for distributing the MPLS tunnel label or any other labels that may appear above the PW label on the stack. Any acceptable method of MPLS label distribution will do. This document specifies a protocol for assigning and distributing the PW label. This protocol is LDP, extended as specified in the remainder of this document. An LDP session must be set up between the pseudowire endpoints. LDP MUST exchange PW FEC label bindings in downstream unsolicited mode, independent of the negotiated label advertisement mode of the LDP session according to the specifications in specified in [RFC7358]. LDP's "liberal label retention" mode SHOULD be used. However all the LDP procedures that are specified in [RFC5036], and that are also applicable to this protocol specification MUST be implemented.

This document requires that a receiving LSR MUST respond to a Label Request message with either a Label Mapping for the requested label or with a Notification message that indicates why it cannot satisfy the request. These procedures are specified in [RFC5036] section 3.5.7 "Label Mapping Message", and 3.5.8 "Label Request Message". Note that sending these responses is a stricter requirement than is specified in [RFC5036] but these response messages are REQUIRED to ensure correct operation of this protocol.

In addition to the protocol specified herein, static assignment of PW labels may be used, and implementations of this protocol SHOULD provide support for static assignment. PW encapsulation is always symmetrical in both directions of traffic along a specific PW, whether the PW uses an LDP control plane or not.

This document specifies all the procedures necessary to set up and maintain the pseudowires needed to support "unswitched" point to point services, where each endpoint of the pseudowire is provisioned with the identity of the other endpoint. There are also protocol mechanisms specified herein that can be used to support switched services and other provisioning models. However, the use of the protocol mechanisms to support those other models and services is not described in this document.

5. Details Specific to Particular Emulated Services

5.1. IP Layer 2 Transport

This mode carries IP packets over a pseudowire. The encapsulation used is according to [RFC3032]. The PW control word MAY be inserted between the MPLS label stack and the IP payload. The encapsulation of the IP packets for forwarding on the attachment circuit is implementation specific, is part of the native service processing (NSP) function [RFC3985], and is outside the scope of this document.

6. LDP

The PW label bindings are distributed using the LDP downstream unsolicited mode described in [RFC5036]. The PEs will establish an LDP session using the Extended Discovery mechanism described in [LDP, sectionn 2.4.2 and 2.5].

An LDP Label Mapping message contains an FEC TLV, a Label TLV, and zero or more optional parameter TLVs.

The FEC TLV is used to indicate the meaning of the label. In the current context, the FEC TLV would be used to identify the particular pseudowire that a particular label is bound to. In this specification, we define two new FEC TLVs to be used for identifying pseudowires. When setting up a particular pseudowire, only one of these FEC TLVs is used. The one to be used will depend on the particular service being emulated and on the particular provisioning model being supported.

LDP allows each FEC TLV to consist of a set of FEC elements. For setting up and maintaining pseudowires, however, each FEC TLV MUST contain exactly one FEC element.

The LDP base specification has several kinds of label TLVs, including the Generic Label TLV, as specified in [RFC5036], section 3.4.2.1. For setting up and maintaining pseudowires, the Generic Label TLV MUST be used.

6.1. The PWid FEC Element

The PWid FEC element may be used whenever both pseudowire endpoints have been provisioned with the same 32-bit identifier for the pseudowire.

For this purpose, a new type of FEC element is defined. The FEC element type is 0x80 and is defined as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  PWid (0x80)  |C|         PW type             |PW info Length |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Group ID                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             PW ID                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Interface Parameter  Sub-TLV                   |
|                              "                                |
|                              "                                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- PW type

  A 15 bit quantity containing a value that represents the type of PW. Assigned Values are specified in "IANA Allocations for pseudo Wire Edge to Edge Emulation (PWE3)" [RFC4446].

- Control word bit (C)

  The bit (C) is used to flag the presence of a control word as follows:

      C = 1 control word present on this PW.
      C = 0 no control word present on this PW.

  Please see the section "Control Word" for further explanation.

- PW information length

  Length of the PW ID field and the interface parameters sub-TLV in octets. If this value is 0, then it references all PWs using the specified group ID, and there is no PW ID present, nor are there any interface parameter sub-TLVs.

- Group ID

  An arbitrary 32 bit value which represents a group of PWs that is used to create groups in the PW space. The group ID is intended to be used as a port index, or a virtual tunnel index. To simplify configuration a particular PW ID at ingress could be part of a Group ID assigned to the virtual tunnel for transport to the egress router. The Group ID is very useful for sending wild card label withdrawals, or PW wild card status notification messages to remote PEs upon physical port failure.

- PW ID

  A non-zero 32-bit connection ID that together with the PW type identifies a particular PW. Note that the PW ID and the PW type MUST be the same at both endpoints.

- Interface Parameter Sub-TLV

  This variable length TLV is used to provide interface specific parameters, such as attachment circuit MTU.

  Note that as the "interface parameter sub-TLV" is part of the FEC, the rules of LDP make it impossible to change the interface parameters once the pseudowire has been set up. Thus the interface parameters field must not be used to pass information, such as status information, that may change during the life of the pseudowire. Optional parameter TLVs should be used for that purpose.

Using the PWid FEC, each of the two pseudowire endpoints independently initiates the setup of a unidirectional LSP.   An outgoing LSP and an incoming LSP are bound together into a single pseudowire if they have the same PW ID and PW type.

6.2. The Generalized PWid FEC Element

The PWid FEC element can be used if a unique 32-bit value has been assigned to the PW, and if each endpoint has been provisioned with that value. The Generalized PWid FEC element requires that the PW endpoints be uniquely identified; the PW itself is identified as a pair of endpoints. In addition, the endpoint identifiers are structured to support applications where the identity of the remote endpoints needs to be auto-discovered rather than statically configured.

The "Generalized PWid FEC Element" is FEC type 0x81.

The Generalized PWid FEC Element does not contain anything corresponding to the "Group ID" of the PWid FEC element. The functionality of the "Group ID" is provided by a separate optional LDP TLV, the "PW Group ID TLV", described below. The Interface Parameters field of the PWid FEC element is also absent; its functionality is replaced by the optional Interface Parameters TLV, described below.

6.2.1. Attachment Identifiers

As discussed in [RFC3985], a pseudowire can be thought of as connecting two "forwarders". The protocol used to set up a pseudowire must allow the forwarder at one end of a pseudowire to identify the forwarder at the other end. We use the term "attachment identifier", or "AI", to refer to the field that the protocol uses to identify the forwarders. In the PWid FEC, the PWid field serves as the AI. In this section, we specify a more general form of AI that is structured and of variable length.

Every Forwarder in a PE must be associated with an Attachment Identifier (AI), either through configuration or through some algorithm. The Attachment Identifier must be unique in the context of the PE router in which the Forwarder resides. The combination <PE router IP address, AI> must be globally unique.

It is frequently convenient to regard a set of Forwarders as being members of a particular "group", where PWs may only be set up among members of a group. In such cases, it is convenient to identify the Forwarders relative to the group, so that an Attachment Identifier would consist of an Attachment Group Identifier (AGI) plus an Attachment Individual Identifier (AII).

An Attachment Group Identifier may be thought of as a VPN-id, or a VLAN identifier, some attribute that is shared by all the Attachment PWs (or pools thereof) that are allowed to be connected.

The details of how to construct the AGI and AII fields identifying the pseudowire endpoints are outside the scope of this specification. Different pseudowire applications, and different provisioning models, will require different sorts of AGI and AII fields. The specification of each such application and/or model must include the rules for constructing the AGI and AII fields.

As previously discussed, a (bidirectional) pseudowire consists of a pair of unidirectional LSPs, one in each direction. If a particular pseudowire connects PE1 with PE2, the PW direction from PE1 to PE2 can be identified as:

<PE1, <AGI, AII1>, PE2, <AGI, AII2>>,

and the PW direction from PE2 to PE1 can be identified by:

<PE2, <AGI, AII2>, PE1, <AGI, AII1>>.

Note that the AGI must be the same at both endpoints, but the AII will in general be different at each endpoint. Thus, from the perspective of a particular PE, each pseudowire has a local or "Source AII", and a remote or "Target AII". The pseudowire setup protocol can carry all three of these quantities:

- Attachment Group Identifier (AGI).

- Source Attachment Individual Identifier (SAII)

- Target Attachment Individual Identifier (TAII)

If the AGI is non-null, then the Source AI (SAI) consists of the AGI together with the SAII, and the Target AI (TAI) consists of the TAII together with the AGI. If the AGI is null, then the SAII and TAII are the SAI and TAI, respectively.

The interpretation of the SAI and TAI is a local matter at the respective endpoint.

The association of two unidirectional LSPs into a single bidirectional pseudowire depends on the SAI and the TAI. Each application and/or provisioning model that uses the Generalized PWid FEC element must specify the rules for performing this association.

6.2.2. Encoding the Generalized PWid FEC Element

FEC element type 0x81 is used. The FEC element is encoded as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Gen PWid (0x81)|C|         PW Type             |PW info Length |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   AGI Type    |    Length     |      Value                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                    AGI  Value (contd.)                        ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   AII Type    |    Length     |      Value                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                   SAII  Value (contd.)                        ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   AII Type    |    Length     |      Value                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                   TAII Value (contd.)                         ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

This document does not specify the AII and AGI type field values; specification of the type field values to be used for a particular application is part of the specification of that application. IANA has assigned these values using the method defined in the [RFC4446] document.

The SAII, TAII, and AGI are simply carried as octet strings. The length byte specifies the size of the Value field. The null string can be sent by setting the length byte to 0. If a particular application does not need all three of these sub-elements, it MUST send all the sub-elements but set the length to 0 for the unused sub-elements.

The PW information length field contains the length of the SAII, TAII, and AGI, combined in octets. If this value is 0, then it references all PWs using the specific grouping ID (specified in the PW Group ID TLV). In this case, there are no other FEC element fields (AGI, SAII, etc.) present, nor any interface parameters TLVs.

Note that the interpretation of a particular field as AGI, SAII, or TAII depends on the order of its occurrence. The type field identifies the type of the AGI, SAII, or TAII. When comparing two occurrences of an AGI (or SAII or TAII), the two occurrences are considered identical if the type, length, and value fields of one are identical, respectively, to those of the other.

6.2.2.1. Interface Parameters TLV

This TLV MUST only be used when sending the Generalized PW FEC. It specifies interface-specific parameters. Specific parameters, when applicable, MUST be used to validate that the PEs and the ingress and egress ports at the edges of the circuit have the necessary capabilities to interoperate with each other.

```
 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|  PW Intf P. TLV (0x096B)  |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Sub-TLV Type  |    Length     |    Variable Length Value      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Variable Length Value                 |
|                             "                                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

A more detailed description of this field can be found in the section "Interface Parameters Sub-TLV", below.

6.2.2.2. PW Group ID TLV

```
 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| PW Group ID TLV (0x096C)  |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             Value                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The PW Group ID is an arbitrary 32-bit value that represents an arbitrary group of PWs. It is used to create group PWs; for example, a PW Grouping ID can be used as a port index and assigned to all PWs that lead to that port. Use of the PW Group ID enables a PE to send "wild card" label withdrawals, or "wild card" status notification messages, to remote PEs upon physical port failure.

Note Well: The PW Group ID is different from and has no relation to, the Attachment Group Identifier.

The PW Group ID TLV is not part of the FEC and will not be advertised except in the PW FEC advertisement. The advertising PE MAY use the wild card withdraw semantics, but the remote PEs MUST implement support for wild card messages. This TLV MUST only be used when sending the Generalized PW ID FEC.

To issue a wild card command (status or withdraw):

- Set the PW Info Length to 0 in the Generalized PWid FEC Element.
- Send only the PW Group ID TLV with the FEC (no AGI/SAII/TAII is sent).

6.2.3. Signaling Procedures

In order for PE1 to begin signaling PE2, PE1 must know the address of the remote PE2, and a TAI. This information may have been configured at PE1, or it may have been learned dynamically via some autodiscovery procedure.

The egress PE (PE1), which has knowledge of the ingress PE, initiates the setup by sending a Label Mapping Message to the ingress PE (PE2). The Label Mapping message contains the FEC TLV, carrying the Generalized PWid FEC Element (type 0x81). The Generalized PWid FEC element contains the AGI, SAII, and TAII information.

Next, when PE2 receives such a Label Mapping message, PE2 interprets the message as a request to set up a PW whose endpoint (at PE2) is the Forwarder identified by the TAI. From the perspective of the signaling protocol, exactly how PE2 maps AIs to Forwarders is a local matter. In some Virtual Private Wire Services (VPWS) provisioning models, the TAI might, for example, be a string that identifies a particular Attachment Circuit, such as "ATM3VPI4VCI5", or it might, for example, be a string, such as "Fred", that is associated by configuration with a particular Attachment Circuit. In VPLS, the AGI could be a VPN-id, identifying a particular VPLS instance.

If PE2 cannot map the TAI to one of its Forwarders, then PE2 sends a Label Release message to PE1, with a Status Code of "Unassigned/Unrecognized TAI", and the processing of the Label Mapping message is complete.

The FEC TLV sent in a Label Release message is the same as the FEC TLV received in the Label Mapping being released (but without the interface parameter TLV). More generally, the FEC TLV is the same in all LDP messages relating to the same PW. In a Label Release this means that the SAII is the remote peer’s AII and the TAII is the sender’s local AII.

If the Label Mapping Message has a valid TAI, PE2 must decide whether to accept it. The procedures for so deciding will depend on the particular type of Forwarder identified by the TAI. Of course, the Label Mapping message may be rejected due to standard LDP error conditions as detailed in [RFC5036].

If PE2 decides to accept the Label Mapping message, then it has to make sure that a PW LSP is set up in the opposite (PE1-->PE2) direction. If it has already signaled for the corresponding PW LSP in that direction, nothing more needs to be done. Otherwise, it must initiate such signaling by sending a Label Mapping message to PE1. This is very similar to the Label Mapping message PE2 received, but the SAI and TAI are reversed.

Thus, a bidirectional PW consists of two LSPs, where the FEC of one has the SAII and TAII reversed with respect to the FEC of the other.

6.3. Signaling of Pseudowire Status

6.3.1. Use of Label Mapping Messages

The PEs MUST send Label Mapping Messages to their peers as soon as the PW is configured and administratively enabled, regardless of the attachment circuit state. The PW label should not be withdrawn unless the operator administratively configures the pseudowire down (or the PW configuration is deleted entirely). Using the procedures outlined in this section, a simple label withdraw method MAY also be supported as a legacy means of signaling PW status and AC status. In any case, if the label-to-PW binding is not available the PW MUST be considered in the down state.

Once the PW status negotiation procedures are completed and if they result in the use of the label withdraw method for PW status communication, and this method is not supported by one of the PEs, then that PE must send a Label Release Message to its peer with the following error:

"Label Withdraw PW Status Method Not Supported"

If the label withdraw method for PW status communication is selected for the PW, it will result in the Label Mapping Message being advertised only if the attachment circuit is active. The PW status signaling procedures described in this section MUST be fully implemented.

6.3.2. Signaling PW Status

The PE devices use an LDP TLV to indicate status to their remote peers. This PW Status TLV contains more information than the alternative simple Label Withdraw message.

The format of the PW Status TLV is:

```
 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|     PW Status (0x096A)    |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Status Code                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The status code is a 4 octet bit field as specified in the PW IANA Allocations document [RFC4446]. The length specifies the length of the Status Code field in octets (equal to 4).

Each bit in the status code field can be set individually to indicate more than a single failure at once. Each fault can be cleared by sending an appropriate Notification message in which the respective bit is cleared. The presence of the lowest bit (PW Not Forwarding) acts only as a generic failure indication when there is a link-down event for which none of the other bits apply.

The Status TLV is transported to the remote PW peer via the LDP Notification message as described in [RFC5036]. The format of the Notification Message for carrying the PW Status is as follows:

```
 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|   Notification (0x0001)     |      Message Length           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Message ID                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Status (TLV)                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       PW Status TLV                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           PWId FEC TLV or Generalized ID FEC TLV              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 PW Group ID TLV (Optional)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The Status TLV status code is set to 0x00000028, "PW status", to indicate that PW status follows. Since this notification does not refer to any particular message, the Message Id field is set to 0.

The PW FEC TLV SHOULD NOT include the interface parameter sub-TLVs, as they are ignored in the context of this message. However the PW

FEC TLV MUST include the C-bit, where aplicable, as it is part of the FEC. When a PE’s attachment circuit encounters an error, use of the PW Notification Message allows the PE to send a single "wild card" status message, using a PW FEC TLV with only the group ID set, to denote this change in status for all affected PW connections. This status message contains either the PW FEC TLV with only the group ID set, or else it contains the Generalized FEC TLV with only the PW Group ID TLV.

As mentioned above, the Group ID field of the PWid FEC element, or the PW Grouping ID TLV used with the Generalized PWid FEC element, can be used to send a status notification for all arbitrary sets of PWs. This procedure is OPTIONAL, and if it is implemented, the LDP Notification message should be as follows: If the PWid FEC element is used, the PW information length field is set to 0, the PW ID field is not present, and the interface parameter sub-TLVs are not present. If the Generalized FEC element is used, the AGI, SAII, and TAII are not present, the PW information length field is set to 0, the PW Group ID TLV is included, and the Interface Parameters TLV is omitted. For the purpose of this document, this is called the "wild card PW status notification procedure", and all PEs implementing this design are REQUIRED to accept such a notification message but are not required to send it.

6.3.3. Pseudowire Status Negotiation Procedures

When a PW is first set up, the PEs MUST attempt to negotiate the usage of the PW status TLV. This is accomplished as follows: A PE that supports the PW Status TLV MUST include it in the initial Label Mapping message following the PW FEC and the interface parameter sub-TLVs. The PW Status TLV will then be used for the lifetime of the pseudowire. This is shown in the following diagram:

```
  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
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +                 PWId FEC or Generalized ID FEC                +
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                       Interface Parameters                    |
 |                              "                                |
 |                              "                                |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |0|0| Generic Label (0x0200)    |      Length                   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |     Label                                                     |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |1|0|     PW Status (0x096A)    |            Length             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                          Status Code                          |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

If a PW Status TLV is included in the initial Label Mapping message for a PW, then if the Label Mapping message from the remote PE for that PW does not include a PW status TLV, or if the remote PE does not support the PW Status TLV, the PW will revert to the label withdraw method of signaling PW status. Note that if the PW Status TLV is not supported by the remote peer, the peer will automatically ignore it, since the I (ignore) bit is set in the TLV. The PW Status TLV, therefore, will not be present in the corresponding FEC advertisement from the remote LDP peer, which results in exactly the above behavior.

If the PW Status TLV is not present following the FEC TLV in the initial PW Label Mapping message received by a PE, then the PW Status TLV will not be used, and both PEs supporting the pseudowire will revert to the label withdraw procedure for signaling status changes.

If the negotiation process results in the usage of the PW status TLV, then the actual PW status is determined by the PW status TLV that was sent within the initial PW Label Mapping message. Subsequent updates of PW status are conveyed through the notification message.

6.4. Interface Parameters Sub-TLV

This field specifies interface-specific parameters. When applicable, it MUST be used to validate that the PEs and the ingress and egress ports at the edges of the circuit have the necessary capabilities to interoperate with each other. The field structure is defined as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Sub-TLV Type  |    Length     |    Variable Length Value      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Variable Length Value                 |
|                             "                                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The interface parameter sub-TLV type values are specified in "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)" [RFC4446].

The Length field is defined as the length of the interface parameter including the parameter id and length field itself. Processing of the interface parameters should continue when unknown interface parameters are encountered, and they MUST be silently ignored.

- Interface MTU sub-TLV type

  A 2 octet value indicating the MTU in octets. This is the Maximum Transmission Unit, excluding encapsulation overhead, of the egress packet interface that will be transmitting the decapsulated PDU that is received from the MPLS-enabled network. This parameter is applicable only to PWs transporting packets and is REQUIRED for these PW types. If this parameter does not match in both directions of a specific PW, that PW MUST NOT be enabled.

- Optional Interface Description string sub-TLV type

  This arbitrary, and OPTIONAL, interface description string is used to send a human-readable administrative string describing the interface to the remote. This parameter is OPTIONAL, and is applicable to all PW types. The interface description parameter string length is variable, and can be from 0 to 80 octets. Human-readable text MUST be provided in the UTF-8 charset using the Default Language [RFC2277].

6.5. LDP label Withdrawal procedures

As mentioned above, the Group ID field of the PWid FEC element, or the PW Grouping ID TLV used with the Generalized PWid FEC element, can be used to withdraw all PW labels associated with a particular PW group. This procedure is OPTIONAL, and if it is implemented, the LDP Label Withdraw message should be as follows: If the PWid FEC element is used, the PW information length field is set to 0, the PW ID field is not present, the interface parameter sub-TLVs are not present, and the Label TLV is not present. If the Generalized FEC element is used, the AGI, SAII, and TAII are not present, the PW information length field is set to 0, the PW Group ID TLV is included, the Interface Parameters TLV is not present, and the Label TLV is not present. For the purpose of this document, this is called the "wild card withdraw procedure", and all PEs implementing this design are REQUIRED to accept such withdrawn message but are not required to send it. Note that the PW Group ID TLV only applies to PWs using the Generalized ID FEC element, while the Group ID only applies to PWid FEC element.

The interface parameter sub-TLVs, or TLV, MUST NOT be present in any LDP PW Label Withdraw or Label Release message. A wild card Label Release message MUST include only the group ID, or Grouping ID TLV. A Label Release message initiated by a PE router must always include the PW ID.

7. Control Word

7.1. PW Types for which the Control Word is REQUIRED

The Label Mapping messages that are sent in order to set up these PWs MUST have C=1. When a Label Mapping message for a PW of one of these types is received and C=0, a Label Release message MUST be sent, with an "Illegal C-bit" status code. In this case, the PW will not be enabled

7.2. PW Types for which the Control Word is NOT mandatory

If a system is capable of sending and receiving the control word on PW types for which the control word is not mandatory, then each such PW endpoint MUST be configurable with a parameter that specifies whether the use of the control word is PREFERRED or NOT PREFERRED. For each PW, there MUST be a default value of this parameter. This specification does NOT state what the default value should be.

If a system is NOT capable of sending and receiving the control word on PW types for which the control word is not mandatory, then it behaves exactly as if it were configured for the use of the control word to be NOT PREFERRED.

If a Label Mapping message for the PW has already been received but no Label Mapping message for the PW has yet been sent, then the procedure is as follows:

- -i. If the received Label Mapping message has C=0, send a Label Mapping message with C=0; the control word is not used.
- -ii. If the received Label Mapping message has C=1, and the PW is locally configured such that the use of the control word is preferred, then send a Label Mapping message with C=1; the control word is used.
- -iii. If the received Label Mapping message has C=1, and the PW is locally configured such that the use of the control word is not preferred or the control word is not supported, then act as if no Label Mapping message for the PW had been received (That is: proceed to the next paragraph).

If a Label Mapping message for the PW has not already been received (or if the received Label Mapping message had C=1 and either local configuration says that the use of the control word is not preferred or the control word is not supported), then send a Label Mapping message in which the C-bit is set to correspond to the locally configured preference for use of the control word. (That is, set C=1 if locally configured to prefer the control word, and set C=0 if locally configured to prefer not to use the control word or if the control word is not supported).

The next action depends on what control message is next received for that PW. The possibilities are as follows:

- -i. A Label Mapping message with the same C-bit value as specified in the Label Mapping message that was sent. PW setup is now complete, and the control word is used if C=1 but is not used if C=0.

- -ii. A Label Mapping message with C=1, but the Label Mapping message that was sent has C=0. In this case, ignore the received Label Mapping message and continue to wait for the next control message for the PW.

- -iii. A Label Mapping message with C=0, but the Label Mapping message that was sent has C=1. In this case, send a Label Withdraw message with a "Wrong C-bit" status code, followed by a Label Mapping message that has C=0. PW setup is now complete, and the control word is not used.

-iv. A Label Withdraw message with the "Wrong C-bit" status code. Treat as a normal Label Withdraw, but do not respond. Continue to wait for the next control message for the PW.

If at any time after a Label Mapping message has been received a corresponding Label Withdraw or Release is received, the action taken is the same as for any Label Withdraw or Release that might be received at any time.

If both endpoints prefer the use of the control word, this procedure will cause it to be used. If either endpoint prefers not to use the control word or does not support the control word, this procedure will cause it not to be used. If one endpoint prefers to use the control word but the other does not, the one that prefers not to use it has no extra protocol to execute; it just waits for a Label Mapping message that has C=0.

7.3. Control-Word Renegotiation by Label Request Message

It is possible that after the PW C-bit negotation procedure described above is completed, the local PE is re-provisioned with a different control word preference. Therefore once the Control-Word negotation procedures are completed, the procedure can be restarted as follows:

-i. If local PE has previously sent a Label Mapping message, it MUST send a Label Withdraw message to remote PE and wait until it has received a Label Release message from the remote PE.

-ii. the local PE MUST send a label release message to the remote PE for the specific label associated with the FEC that was advertized for this specific PW. Note: the above-mentioned steps of the Label Release message and Label Withdraw message are not required to be excuted in any specific sequence.

-iii. The local PE MUST send a Label Request message to the peer PE, and then MUST wait until it receives a Label Mapping message containing the remote PE’s currently configured preference for use of the control word.

Once the remote PE has successfully processed the Label Withdraw message and Label Release messages, it will reset the C-bit negotation state machine and its use of the control word with the locally configured preference.

From this point on the local and remote PEs will follow the C-bit negotaiation procedures defined in the previous section.

The above C-bit renegotation process SHOULD NOT be interupted until it is completed, or unpredictable results might occur.

7.4. Sequencing Considerations

In the case where the router considers the sequence number field in the control word, it is important to note the following details when advertising labels.

7.4.1. Label Advertisements

After a label has been withdrawn by the output router and/or released by the input router, care must be taken not to advertise (re-use) the same released label until the output router can be reasonably certain that old packets containing the released label no longer persist in the MPLS-enabled network.

This precaution is required to prevent the imposition router from restarting packet forwarding with a sequence number of 1 when it receives a Label Mapping message that binds the same FEC to the same label if there are still older packets in the network with a sequence number between 1 and 32768. For example, if there is a packet with a sequence number=n, where n is in the interval [1,32768] traveling through the network, it would be possible for the disposition router to receive that packet after it re-advertises the label. Since the label has been released by the imposition router, the disposition router SHOULD be expecting the next packet to arrive with a sequence number of 1. Receipt of a packet with a sequence number equal to n will result in n packets potentially being rejected by the disposition router until the imposition router imposes a sequence number of n+1 into a packet. Possible methods to avoid this are for the disposition router always to advertise a different PW label, or for the disposition router to wait for a sufficient time before attempting to re-advertise a recently released label. This is only an issue when sequence number processing is enabled at the disposition router.

7.4.2. Label Release

In situations where the imposition router wants to restart forwarding of packets with sequence number 1, the router shall 1) send to the disposition router a Label Release Message, and 2) send to the disposition router a Label Request message. When sequencing is supported, advertisement of a PW label in response to a Label Request message MUST also consider the issues discussed in the section on Label Advertisements.

8. IANA Considerations

The authors request that IANA remove this section before publication and that IANA update any references to [RFC4447] in their registries to refer to this document.

9. Security Considerations

This document specifies the LDP extensions that are needed for setting up and maintaining pseudowires. The purpose of setting up pseudowires is to enable Layer 2 frames to be encapsulated in MPLS and transmitted from one end of a pseudowire to the other. Therefore we treat the security considerations for both the data plane and the control plane.

9.1. Data-Plane Security

With regard to the security of the data plane, the following areas must be considered:

- MPLS PDU inspection.
- MPLS PDU spoofing.
- MPLS PDU alteration.
- MPLS PSN protocol security.
- Access Circuit security.
- Denial of service prevention on the PE routers.

When an MPLS PSN is used to provide pseudowire service, there is a perception that security MUST be at least equal to the currently deployed Layer 2 native protocol networks that the MPLS/PW network combination is emulating. This means that the MPLS-enabled network SHOULD be isolated from outside packet insertion in such a way that it SHOULD NOT be possible to insert an MPLS packet into the network directly. To prevent unwanted packet insertion, it is also important to prevent unauthorized physical access to the PSN, as well as unauthorized administrative access to individual network elements.

As mentioned above, an MPLS-enabled network should not accept MPLS packets from its external interfaces (i.e., interfaces to CE devices or to other providers' networks) unless the top label of the packet was legitimately distributed to the system from which the packet is being received. If the packet's incoming interface leads to a different SP (rather than to a customer), an appropriate trust relationship must also be present, including the trust that the other SP also provides appropriate security measures.

The three main security problems faced when using an MPLS-enabled network to transport PWs are spoofing, alteration, and inspection. First, there is a possibility that the PE receiving PW PDUs will get a PDU that appears to be from the PE transmitting the PW into the PSN, but that was not actually transmitted by the PE originating the PW. (That is, the specified encapsulations do not by themselves enable the decapsulator to authenticate the encapsulator.) A second problem is the possibility that the PW PDU will be altered between the time it enters the PSN and the time it leaves the PSN (i.e., the specified encapsulations do not by themselves assure the decapsulator of the packet’s integrity.) A third problem is the possibility that the PDU’s contents will be seen while the PDU is in transit through the PSN (i.e., the specification encapsulations do not ensure privacy.) How significant these issues are in practice depends on the security requirements of the applications whose traffic is being sent through the tunnel, and how secure the PSN itself is.

9.2. Control-Plane Security

General security considerations with regard to the use of LDP are specified in section 5 of [RFC5036]. Those considerations also apply to the case where LDP is used to set up pseudowires.

A pseudowire connects two attachment circuits. It is important to make sure that LDP connections are not arbitrarily accepted from anywhere, or else a local attachment circuit might get connected to an arbitrary remote attachment circuit. Therefore, an incoming LDP session request MUST NOT be accepted unless its IP source address is known to be the source of an "eligible" LDP peer. The set of eligible peers could be pre-configured (either as a list of IP addresses, or as a list of address/mask combinations), or it could be discovered dynamically via an auto-discovery protocol that is itself trusted. (Obviously, if the auto-discovery protocol were not trusted, the set of "eligible peers" it produces could not be trusted.)

Even if an LDP connection request appears to come from an eligible peer, its source address may have been spoofed. Therefore, some means of preventing source address spoofing must be in place. For example, if all the eligible peers are in the same network, source address filtering at the border routers of that network could eliminate the possibility of source address spoofing.

The LDP MD5 authentication key option, as described in section 2.9 of [RFC5036], MUST be implemented, and for a greater degree of security, it must be used. This provides integrity and authentication for the LDP messages and eliminates the possibility of source address spoofing. Use of the MD5 option does not provide privacy, but privacy of the LDP control messages is not usually considered important. As the MD5 option relies on the configuration of pre-shared keys, it does not provide much protection against replay attacks. In addition, its reliance on pre-shared keys may make it very difficult to deploy when the set of eligible neighbors is determined by an auto-configuration protocol.

When the Generalized PWid FEC Element is used, it is possible that a particular LDP peer may be one of the eligible LDP peers but may not be the right one to connect to the particular attachment circuit identified by the particular instance of the Generalized PWid FEC element. However, given that the peer is known to be one of the eligible peers (as discussed above), this would be the result of a configuration error, rather than a security problem. Nevertheless, it may be advisable for a PE to associate each of its local attachment circuits with a set of eligible peers rather than have just a single set of eligible peers associated with the PE as a whole.

10. Interoperability and Deployment

Section 2.2. of [RFC6410] specifies four requirements that an Internet Standard must meet. This section documents how this document meets those requirements.

The pseudowire technology was first deployed in 2001 and has been widely deployed by many carriers. [RFC7079] documents the results of a survey of PW implementations, with specific emphasis on Control Word usage. [EANTC] documents a public multi-vendor interoperability test of MPLS and Carrier Ethernet equipment, which included testing of Ethernet, ATM and TDM pseudowires.

The errata against [RFC4447] are generally editorial in nature and have been addressed in this document.

All features in this specification have been implemented by multiple vendors.

No IPR disloures have been made to the IETF related to this document, to RFC4447 or RFC6723, or to the Internet-Drafts that resulted in RFC4447 and RFC6723.

11. Acknowledgments

The authors wish to acknowledge the contributions of Vach Kompella, Vanson Lim, Wei Luo, Himanshu Shah, and Nick Weeds. The authors wish to also acknowledge the contribution of the authors of rfc6723, Lizhong Jin, Raymond Key, Simon Delord, Tom Nadeau, Sami Boutros, whose work has been incorporated in this revised document.

12. Normative References

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

[RFC5036] "LDP Specification." Andersson, L. Ed., Minei, I. Ed., Thomas, B. Ed. January 2001. RFC5036, October 2007

[RFC3032] "MPLS Label Stack Encoding", Rosen E., Rekhter Y., Tappan D., Fedorkow G., Farinacci D., Li T., Conta A.. RFC3032

[RFC4446] "IANA Allocations for pseudo Wire Edge to Edge Emulation (PWE3)" Martini L. RFC4446, April 2006

[RFC7358] "Label Advertisement Discipline for LDP Forwarding Equivalence Classes (FECs)", Raza K., Boutros S., Martini L., RFC7358, October 2014

13. Informative References

[RFC2277] Alvestrand, H., "IETF Policy on Character Sets and Languages", BCP 18, RFC 2277, January 1998.

[RFC3985] "PWE3 Architecture" Bryant, et al., RFC3985.

[RFC4842] "Synchronous Optical Network/Synchronous Digital Hierarchy (SONET/SDH) Circuit Emulation over Packet (CEP)", A. Malis, P. Pate, R. Cohen, Ed., D. Zelig, RFC4842, April 2007

[RFC4553] "Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)", Vainshtein A. Ed., Stein, YJ. Ed. RFC4553, June 2006

[RFC4619] "Encapsulation Methods for Transport of Frame Relay over Multiprotocol Label Switching (MPLS) Networks", Martini L. Ed. C. Kawa Ed., A. Malis, Ed. RFC4619, September 2006

[RFC4717] "Encapsulation Methods for Transport of Asynchronous Transfer Mode (ATM) over MPLS Networks", Martini L., Jayakumar J., Bocci M., El-Aawar N., Brayley J., Koleyni G. RFC4717, December 2006

[RFC4618] "Encapsulation Methods for Transport of PPP/High-Level Data Link Control (HDLC) Frames over MPLS Networks", Martini L. Rosen E., Heron G., Malis A. RFC4618, September 2006

[RFC4448] "Encapsulation Methods for Transport of Ethernet over MPLS Networks", Martini L. Ed., Rosen E., El-Aawar N., Heron G. RFC4448, April 2006.

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

[RFC6410] "Reducing the Standards Track to Two Maturity Levels", Housley R., Crocker D., Burger E. RFC6410, October 2011

[RFC6723] "Update of the Pseudowire Control-Word Negotiation Mechanism", Jin L. Ed., Key R. Ed., Delord S., Nadeau T., Boutros S. RFC5723, September 2012

[RFC6410] "Reducing the Standads Track to Two Maturity Levels", Housley R., Crocker D., Burger E. RFC6410, October 2011

[RFC7079] "The Pseudowire (PW) and Virtual Circuit Connectivity Verification (VCCV) Implementation Survey Results", Del Regno N., Malis A. RFC7079, November 2013

[ANSI] American National Standards Institute, "Synchronous Optical Network Formats", ANSI T1.105-1995.

[ITUG] ITU Recommendation G.707, "Network Node Interface For The Synchronous Digital Hierarchy", 1996.

[EANTC] The European Advanced Networking Test Center "MPLS and Carrier Ethernet: Service - Connect - Transport. Public Multi-Vendor Interoperability Test", February 2009.

14. Author Information

Luca Martini
Cisco Systems, Inc.
1899 Wynkoop Street
Suite 600
Denver, CO, 80202
e-mail: lmartini@cisco.com

Giles Heron
Cisco Systems
10 New Square
Bedfont Lakes
Feltham
Middlesex
TW14 8HA
UK
e-mail: giheron@cisco.com

15. Additional Historical Contributing Authors

This historical list is from the original RFC, and is not updated. It is intended for recognition of their work on RFC4447.

Nasser El-Aawar
Level 3 Communications, LLC.
1025 Eldorado Blvd.
Broomfield, CO, 80021
e-mail: nna@level3.net

Eric C. Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA 01719
e-mail: erosen@cisco.com

Dan Tappan
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA 01719
e-mail: tappan@cisco.com

Toby Smith
Google
6425 Penn Ave. #700
Pittsburgh, PA 15206
e-mail: tob@google.com

Dimitri Vlachos
Riverbed Technology
e-mail: dimitri@riverbed.com

Jayakumar Jayakumar,
Cisco Systems Inc.
3800 Zanker Road, MS-SJ02/2,
San Jose, CA, 95134
e-mail: jjayakum@cisco.com

Alex Hamilton,
Cisco Systems Inc.
485 East Tasman Drive, MS-SJC07/3,
San Jose, CA, 95134
e-mail: tahamilt@cisco.com

Steve Vogelsang
ECI Telecom
Omega Corporate Center
1300 Omega Drive
Pittsburgh, PA 15205
e-mail: stephen.vogelsang@ecitele.com

John Shirron
ECI Telecom
Omega Corporate Center
1300 Omega Drive
Pittsburgh, PA 15205
e-mail: john.shirron@ecitele.com

Andrew G. Malis
Verizon
60 Sylvan Rd.
Waltham, MA 02451
e-mail: andrew.g.malis@verizon.com

Vinai Sirkay
Reliance Infocomm
Dhirubai Ambani Knowledge City
Navi Mumbai 400 709
e-mail: vinai@sirkay.com

Vasile Radoaca
Nortel Networks
600 Technology Park
Billerica MA 01821
e-mail: vasile@nortelnetworks.com

Chris Liljenstolpe
149 Santa Monica Way
San Francisco, CA 94127
e-mail: ietf@cdl.asgaard.org

Dave Cooper
Global Crossing
960 Hamlin Court
Sunnyvale, CA 94089
e-mail: dcooper@gblx.net

Kireeti Kompella
Juniper Networks
1194 N. Mathilda Ave
Sunnyvale, CA 94089
e-mail: kireeti@juniper.net

Expiration Date: January 2017

PALS Workgroup
Internet-Draft
Intended status: Standards Track
Expires: May 25, 2017

P. Jain
Cisco Systems, Inc.
S. Boutros
VMWare, Inc.
S. Aldrin
Google Inc.
November 21, 2016

Definition of P2MP PW TLV for LSP-Ping Mechanisms
draft-jain-pals-p2mp-pw-lsp-ping-02

Abstract

LSP-Ping is a widely deployed Operation, Administration, and Maintenance (OAM) mechanism in MPLS networks. This document describes a mechanism to verify connectivity of Point-to-Multipoint (P2MP) Pseudowires (PW) using LSP Ping.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

A Point-to-Multipoint (P2MP) Pseudowire (PW) emulates the essential attributes of a unidirectional P2MP Telecommunications service such as P2MP ATM over PSN. Requirements for P2MP PW are described in [RFC7338]. P2MP PWs are carried over P2MP MPLS LSP. The Procedures for P2MP PW signaling using BGP are described in [RFC7117] and LDP for single segment P2MP PWs are described in [I-D.ietf-pwe3-p2mp-pw]. Many P2MP PWs can share the same P2MP MPLS LSP and this arrangement is called Aggregate P-tree. The aggregate P2MP trees require an upstream assigned label so that on the tail of the P2MP LSP, the traffic can be associated with a VPN or a VPLS instance. When a P2MP MPLS LSP carries only one VPN or VPLS service instance, the arrangement is called Inclusive P-Tree. For Inclusive P-Trees, P2MP MPLS LSP label itself can uniquely identify the VPN or VPLS service being carried over P2MP MPLS LSP. The P2MP MPLS LSP can also be used in Selective P-Tree arrangement for carrying multicast traffic. In a Selective P-Tree arrangement, traffic to each multicast group in a VPN or VPLS instance is carried by a separate unique P-tree. In Aggregate Selective P-tree arrangement, traffic to a set of multicast groups from different VPN or VPLS instances is carried over a same shared P-tree.

The P2MP MPLS LSP are setup either using P2MP RSVP-TE [RFC4875] or Multipoint LDP (mDLP) [RFC6388]. Mechanisms for fault detection and isolation for data plane failures for P2MP MPLS LSPs are specified in

[RFC6425]. This document describes a mechanism to detect data plane failures for P2MP PW carried over P2MP MPLS LSPs.

This document defines a new P2MP Pseudowire sub-TLV for Target FEC Stack for P2MP PW. The P2MP Pseudowire sub-TLV is added in Target FEC Stack TLV by the originator of the Echo Request to inform the receiver at P2MP MPLS LSP tail, of the P2MP PW being tested.

Multi-segment Pseudowires support is out of scope of this document at present and may be included in future.

2. 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].

3. Terminology

ATM: Asynchronous Transfer Mode

LSR: Label Switching Router

MPLS-OAM: MPLS Operations, Administration and Maintenance

P2MP-PW: Point-to-Multipoint PseudoWire

PW: PseudoWire

TLV: Type Length Value

4. Identifying a P2MP PW

This document introduces a new LSP Ping Target FEC Stack sub-TLV, P2MP Pseudowire sub-TLV, to identify the P2MP PW under test at the P2MP LSP Tail/Bud node.

4.1. P2MP Pseudowire Sub-TLV

The P2MP Pseudowire sub-TLV has the format shown in Figure 1. This TLV is included in the echo request sent over P2MP PW by the originator of request.

The Attachment Group Identifier (AGI) in P2MP Pseudowire Sub-TLV as described in Section 3.4.2 in [RFC4446], identifies the VPLS instance. The Originating Router’s IP address is the IPv4 or IPv6 address of the P2MP PW root.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| AGI Type     |   AGI Length  |                                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                                |
~                          AGI Value                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| IP Addr Len |                                                 |
+-+-+-+-+-+-+-+                                                 |
~               Originating Routers IP Addr                     ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: P2MP Pseudowire sub-TLV format

For Inclusive and Selective P2MP MPLS P-trees, the echo request is sent using the P2MP MPLS LSP label.

For Aggregate Inclusive and Aggregate Selective P-trees, the echo request is sent using a label stack of [P2MP MPLS P-tree label, upstream assigned P2MP PW label]. The P2MP MPLS P-tree label is the outer label and upstream assigned P2MP PW label is inner label.

5. Encapsulation of OAM Ping Packets

The LSP Ping Echo request IPv4/UDP packets will be encapsulated with the MPLS label stack as described in previous sections, followed by the GAL Label [RFC6426]. The GAL label will be followed by the ACH with the Pseudowire Associated Channel Type 16 bit value in the ACH set to IPv4 indicating that the carried packet is an IPv4 packet.

6. Operations

In this section, we explain the operation of the LSP Ping over P2MP PW. Figure 2 shows a P2MP PW PW1 setup from T-PE1 to remote PEs (T-PE2, T-PE3 and T-PE4). The transport LSP associated with the P2MP PW1 can be MLDP P2MP MPLS LSP or P2MP TE tunnel.

```
                 |<--------------P2MP PW---------------->|
          Native |                                       |  Native
         Service |     |<--PSN1->|      |<--PSN2->|      |  Service
          (AC)   V     V         V      V         V      V   (AC)
            |    +-----+         +------+         +------+    |
            |    |     |         |   P1 |=========|T-PE2 |AC3 |    +---+
            |    |     |         |   .......PW1.........>|-------->|CE3|
            |    |T-PE1|=========|   .  |=========|      |    |    +---+
            |    |  .......PW1........  |         +------+    |
            |    |  .  |=========|   .  |         +------+    |
            |    |  .  |         |   .  |=========|T-PE3 |AC4 |    +---+
    +---+   |AC1 |  .  |         |   .......PW1.........>|-------->|CE4|
    |CE1|------->|...  |         |      |=========|      |    |    +---+
    +---+   |    |  .  |         +------+         +------+    |
            |    |  .  |         +------+         +------+    |
            |    |  .  |=========|   P2 |=========|T-PE4 |AC5 |    +---+
            |    |  .......PW1..............PW1.........>|-------->|CE5|
            |    |     |=========|      |=========|      |    |    +---+
            |    +-----+         +------+         +------+    |
```

Figure 2: P2MP PW

When an operator wants to perform a connectivity check for the P2MP PW1, the operator initiate a LSP-Ping request with the Target FEC Stack TLV containing P2MP Pseudowire sub-TLV in the echo request packet. For an Inclusive P2MP P-tree arrangement, the echo request packet is sent over the P2MP MPLS LSP with {P2MP P-tree label, GAL} MPLS label stack and IP ACH Channel header. For an Aggregate Inclusive P-tree arrangement, the echo request packet is sent over the P2MP MPLS LSP with {P2MP P-tree label, P2MP PW upstream assigned label, GAL} MPLS label stack and IP ACH Channel header. The intermediate P router will do swap and replication based on the MPLS LSP label. Once the echo request packet reaches remote terminating PEs, T-PE1s will use the GAL label and the IP ACH Channel header to determine that the packet is IPv4 OAM Packet. The T-PEs will process the packet and perform checks for the P2MP Pseudowire sub-TLV present in the Target FEC Stack TLV as described in Section 4.4 in [RFC4379] and respond according to [RFC4379] processing rules.

7. Controlling Echo Responses

The procedures described in [RFC6425] for preventing congestion of Echo Responses (Echo Jitter TLV) and limiting the echo reply to a single egress node (Node Address P2MP Responder Identifier TLV) can be applied to P2MP PW LSP Ping.

8. Security Considerations

The proposal introduced in this document does not introduce any new security considerations beyond that already apply to [RFC6425].

9. IANA Considerations

This document defines a new sub-TLV type to be included in Target FEC Stack TLV (TLV Type 1) [RFC4379] in LSP Ping.

IANA is requested to assign a sub-TLV type value to the following sub-TLV from the "Multiprotocol Label Switching (MPLS) Label Switched Paths (LSPs) Parameters - TLVs" registry, "TLVs and sub- TLVs" sub-registry:

o P2MP Pseudowire sub-TLV

10. Acknowledgments

The authors would like to thank Shaleen Saxena, Michael Wildt, Tomofumi Hayashi, Danny Prairie for their valuable input and comments.

11. References

11.1. Normative References

[I-D.ietf-pwe3-p2mp-pw]
Sivabalan, S., Boutros, S., and L. Martini, "Signaling Root-Initiated Point-to-Multipoint Pseudowire using LDP", draft-ietf-pwe3-p2mp-pw-04 (work in progress), March 2012.

[RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, DOI 10.17487/RFC4379, February 2006, <http://www.rfc-editor.org/info/rfc4379>.

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

[RFC6425] Saxena, S., Ed., Swallow, G., Ali, Z., Farrel, A., Yasukawa, S., and T. Nadeau, "Detecting Data-Plane Failures in Point-to-Multipoint MPLS - Extensions to LSP Ping", RFC 6425, DOI 10.17487/RFC6425, November 2011, <http://www.rfc-editor.org/info/rfc6425>.

[RFC6426] Gray, E., Bahadur, N., Boutros, S., and R. Aggarwal, "MPLS On-Demand Connectivity Verification and Route Tracing", RFC 6426, DOI 10.17487/RFC6426, November 2011, <http://www.rfc-editor.org/info/rfc6426>.

[RFC7117] Aggarwal, R., Ed., Kamite, Y., Fang, L., Rekhter, Y., and C. Kodeboniya, "Multicast in Virtual Private LAN Service (VPLS)", RFC 7117, DOI 10.17487/RFC7117, February 2014, <http://www.rfc-editor.org/info/rfc7117>.

11.2. Informative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, DOI 10.17487/RFC4875, May 2007, <http://www.rfc-editor.org/info/rfc4875>.

[RFC5085] Nadeau, T., Ed. and C. Pignataro, Ed., "Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires", RFC 5085, DOI 10.17487/RFC5085, December 2007, <http://www.rfc-editor.org/info/rfc5085>.

[RFC6388] Wijnands, IJ., Ed., Minei, I., Ed., Kompella, K., and B. Thomas, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", RFC 6388, DOI 10.17487/RFC6388, November 2011, <http://www.rfc-editor.org/info/rfc6388>.

[RFC7338] Jounay, F., Ed., Kamite, Y., Ed., Heron, G., and M. Bocci, "Requirements and Framework for Point-to-Multipoint Pseudowires over MPLS Packet Switched Networks", RFC 7338, DOI 10.17487/RFC7338, September 2014, <http://www.rfc-editor.org/info/rfc7338>.

Authors' Addresses

Parag Jain
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, ON K2K-3E8
Canada

Email: paragj@cisco.com

Sami Boutros
VMWare, Inc.
USA

Email: sboutros@vmware.com

Sam Aldrin
Google Inc.
USA

Email: aldrin.ietf@gmail.com

PALS Working Group H. Shah
Internet-Draft Ciena Corporation
Intended status: Standards Track P. Brissette
Expires: January 7, 2016 R. Rahman
K. Raza
Cisco Systems, Inc.
Z. Li
Z. Shunwan
W. Haibo
Huawei Technologies
I. Chen
Ericsson
M. Bocci
Alcatel-Lucent
J. Hardwick
Metaswitch
S. Esale
Junipr Networks
K. Tiruveedhula
T. Singh
Juniper Networks
I. Hussain
Infinera Corporation
B. Wen
J. Walker
Comcast
N. Delregno
L. Jalil
M. Joecylyn
Verizon
July 06, 2015

YANG Data Model for MPLS-based L2VPN
draft-shah-pals-mpls-l2vpn-yang-00.txt

Abstract

This document describes a YANG data model for Layer 2 VPN services over MPLS networks. These services include Virtual Private Wire Service (VPWS), Virtual Private LAN service (VPLS) and Ethernet Virtual Private Service (EVPN) that uses LDP and BGP signaled Pseudowires. This document mainly focuses on L2VPN VPWS, other services are for future investigations.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

The Network Configuration Protocol (NETCONF) [RFC6241] is a network management protocol that defines mechanisms to manage network devices. YANG [RFC6020] is a modular language that represents data structures in an XML or JSON tree format, and is used as a data modeling language for the NETCONF.

This document introduces a YANG data model for MPLS based Layer 2 VPN services (L2VPN) [RFC4664] as well as switching between the local attachment circuits. The L2VPN services include point-to-point VPWS and Multipoint VPLS and EVPN services. These services are realized by signaling Pseudowires across MPLS networks using LDP [RFC4447][RFC4762] or BGP[RFC4761].

The Yang data model in this document defines Ethernet based Layer 2 services. Other Layer 2 services, such as ATM, Frame Relay, TDM, etc are included in the scope but will be covered as the future work items. The Ethernet based Layer 2 services will leverage the definitions used in other standards organizations such as IEEE 802.1 and Metro Ethernet Forum (MEF).

The goal is to propose a data object model consisting of building blocks that can be assembled in different order to realize different services. The definition work is undertaken initially by a smaller working group with members representing various vendors and service providers. The VPWS service definitions are covered first followed by VPLS services that build on the data blocks defined for VPWS.

The data model is defined for following constructs that are used for managing the services:

- o Configuration
- o Operational State
- o Executables (Actions)
- o Notifications

The document is organized to first define the data model for the configuration, operational state, actions and notifications of VPWS. The L2VPN data object model defined in this document uses the instance centric approach whereby VPWS service attributes are specified for a given VPWS instance.

2. 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].

3. L2VPN YANG Model

3.1. Overview

One single top level container, mpls-l2vpn, is defined as a parent for three different second level containers that are vpws-instances, vpls-instances, and common building blocks of AC-templates(Attachment Circuit templates) and pseudowire-templates. This document defines the vpws-instances and templates for AC and Pseudowires. The definition of vpls-instances and evpn-instances is left for future revisions.

The L2VPN services have been defined in the IETF L2VPN working group but leverages the pseudowire technologies that were defined in the PWE3 working group. A large number of RFCs from these working groups cover this subject matter. Hence, it is prudent that this document state the scope of the MPLS L2VPN object model definitions.

The following documents are within the scope. This is not an exhaustive list but a representation of documents that are covered for this work:

- Requirements for Pseudo-wire Emulation Edge-to-Edge (PWE3) [RFC3916]

- Pseudo-wire Emulation Edge-to-Edge (PWE3) Architecture [RFC3985]

- IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3) [RFC4446]

- Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP) [RFC4447]

- Encapsulation Methods for Transport of Ethernet over MPLS Networks [RFC4448]

- Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN [RFC4385]
- Requirements for Multi-Segment Pseudowire Emulation Edge-to-Edge (PWE3) [RFC5254]
- An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge [RFC5659]
- Segmented Pseudowire [RFC6073]
- Framework for Layer 2 Virtual Private Networks [RFC4664]
- Service Requirements for Layer 2 Provider-Provisioned Virtual Private Networks [RFC4665]
- Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling [RFC4761]
- Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling [RFC4762]
- Attachment Individual Identifier (AII) Types for Aggregation [RFC5003]
- Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs) [RFC6074]
- Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network [RFC6391]
- Layer 2 Virtual Private Networks Using BGP for Auto-Discovery and Signaling [RFC6624]
- Extensions to the Virtual Private LAN Service (VPLS) Provider Edge (PE) Model for Provider Backbone Bridging [RFC7041]
- LDP Extensions for Optimized MAC Address Withdrawal in a Hierarchical Virtual Private LAN Service (H-VPLS) [RFC7361]
- Using the generic associated channel label for Pseudowire in the MPLS Transport Profile [RFC6423]
- Pseudowire status for static pseudowire [RFC6478]

Note that while pseudowire over MPLS-TP related work is in scope, the initial effort will only address definitions of object model for VPWS services that are commonly deployed.

The ietf work in L2VPN and PWE3 working group relating to L2TP, OAM, multicast (e.g. p2mp, etree, etc) and access specific protocols such as G.8032, MSTP, etc is out-of-scope for this document.

The following is the high level view of the L2VPN data model.

```
template-ref AC // AC
                   template
                   attributes

template-ref PW // PW
                   template
                   attributes

vpws-instance name // container

        svc-type

        // list of AC and PW being used
        AC-1 // container
                template-ref AC
                attribute-override
        PW-2 // container
                template-ref PW
                attribute-override
        PW-3 // container
                template-ref PW
                attribute-override

        // ONLY 2 endpoints!!!
        endpoint-A // container
                        AC-1 // reference

        endpoint-Z // container
                redundancy-grp // container
                        PW-2 // reference
                        PW-3 // reference
```

Figure 1

3.2. L2VPN Common

3.2.1. ac-templates

The ac-templates container contains a list of ac-template. Each ac-template defines a list of AC attributes that are part of native services but associated and processed within the context of L2VPN. For instance, Ethernet VLAN tag imposition, disposition and translation or CVID-bundling would be part of this template.

3.2.2. pw-templates

The pw-templates container contains a list of pw-template. Each pw-template defines a list of common pseudowire attributes such as PW MTU, control word support etc.

3.3. VPWS

3.3.1. ac list

Each VPWS instance defines a list of AC which are cross-connected by the service. Each entry of the AC consists of one ac-template with predefined attributes and values, but also defines attributes that override the attributes defined in referenced ac-template.

3.3.2. pw list

Each VPWS instance defines a list of PW which are cross-connected by the service. Each entry of the PW consists of one pw-template with pre-defined attributes and values, but also defines attributes that override those defined in referenced pw-template. No restrictions are placed on type of signaling (i.e. LDP or BGP) used for a given PW. It is entirely possible to define two PWs, one signaled by LDP and other by BGP.

3.3.3. redundancy-grp choice

The redundancy-grp is a generic redundancy construct which can hold primary and backup members of AC and PWs. This flexibility permits combinations of -

- o primary and backup AC
- o primary and backup PW
- o primary AC and backup PW
- o primary PW and backup AC

3.3.4. endpoint container

The endpoint container holds AC, PW or redundancy-grp references. The core aspect of endpoint container is its flexible personality based on what user decides to include in it. It is future-proofed with possible extensions that can be included in the endpoint container such as Integrated Route Bridging (IRB), PW Headend, Virtual Switch Instance, etc.

3.3.5. vpws-instances container

The vpws-instances container contains a list of vpws-instance. Each entry of the vpws-instance represents a layer-2 cross-connection of two endpoints. This model defines three possible types of endpoints, ac, pw, and redundancy-grp, and allows a vpws-instance to cross-connect any one type of endpoint to all other types of endpoint.

The augmentation of ietf-mpls-l2vpn module is TBD. All IP addresses defined in this module are currently scoped under global VRF/table.

```
module: ietf-mpls-l2vpn
+--rw mpls-l2vpn
   +--rw common
   |  +--rw pw-templates
   |  |  +--rw pw-template* [name]
   |  |     +--rw name              string
   |  |     +--rw mtu?              uint32
   |  |     +--rw cw-negotiation?   cw-negotiation-type
   |  |     +--rw tunnel-policy?    string
   |  +--rw ac-templates
   |     +--rw ac-template* [name]
   |        +--rw name    string
   +--rw vpws-instances
   |  +--rw vpws-instance* [instance-name]
   |     +--rw instance-name     string
   |     +--rw description?      string
   |     +--rw service-type?     l2vpn-service-type
   |     +--rw discovery-type?   l2vpn-discovery-type
   |     +--rw signaling-type    l2vpn-signaling-type
   |     +--rw bgp-parameters
   |     |  +--rw common
   |     |  |  +--rw route-distinguisher?   string
   |     |  |  +--rw vpn-targets* [rt-value]
   |     |  |     +--rw rt-value    string
   |     |  |     +--rw rt-type     bgp-rt-type
   |     |  +--rw discovery
   |     |  |  +--rw vpn-id?   string
```

```
   |     |  +--rw signaling
   |     |     +--rw site-id?      uint16
   |     |     +--rw site-range?   uint16
   |     +--rw pw* [name]
   |     |  +--rw name                string
   |     |  +--rw cw-negotiation?   cw-negotiation-type
   |     |  +--rw template?          pw-template-ref
   |     |  +--rw vccv-ability?      boolean
   |     |  +--rw tunnel-policy?     string
   |     |  +--rw request-vlanid?   uint16
   |     |  +--rw vlan-tpid?         string
   |     |  +--rw ttl?               uint8
   |     |  +--rw (pw-type)?
   |     |     +--:(ldp-pw)
   |     |     |  +--rw peer-ip?           inet:ip-address
   |     |     |  +--rw pw-id?             uint32
   |     |     |  +--rw transmit-label?   uint32
   |     |     |  +--rw receive-label?    uint32
   |     |     |  +--rw icb?               boolean
   |     |     +--:(bgp-pw)
   |     |     |  +--rw remote-pe-id?     inet:ip-address
   |     |     +--:(bgp-ad-pw)
   |     |        +--rw remote-ve-id?     uint16
   |     +--rw ac* [name]
   |     |  +--rw name                         string
   |     |  +--rw template?                    ac-template-ref
   |     |  +--rw pipe-mode?                   enumeration
   |     |  +--rw link-discovery-protocol?  link-discovery-protocol-type
   |     +--rw endpoint-a
   |     |  +--rw (ac-or-pw-or-redundancy-grp)?
   |     |     +--:(ac)
   |     |     |  +--rw ac?                 -> ../../ac/name
   |     |     +--:(pw)
   |     |     |  +--rw pw?                 -> ../../pw/name
   |     |     +--:(redundancy-grp)
   |     |     |  +--rw (primary)
   |     |     |  |  +--:(primary-pw)
   |     |     |  |  |  +--rw primary-pw?        -> ../../pw/name
   |     |     |  |  +--:(primary-ac)
   |     |     |  |     +--rw primary-ac?        -> ../../ac/name
   |     |     |  +--rw (backup)
   |     |     |  |  +--:(backup-pw)
   |     |     |  |  |  +--rw backup-pw?         -> ../../pw/name
   |     |     |  |  +--:(backup-ac)
   |     |     |  |     +--rw backup-ac?         -> ../../ac/name
   |     |     |  +--rw protection-mode?   enumeration
   |     |     +--:(reroute-mode)
   |     |     |  +--rw reroute-mode?      enumeration
```

```
   |     |     +--:(reroute-delay)
   |     |     |  +--rw reroute-delay?     uint16
   |     |     +--:(dual-receive)
   |     |     |  +--rw dual-receive?      boolean
   |     |     +--:(revert)
   |     |     |  +--rw revert?            boolean
   |     |     +--:(revert-delay)
   |     |        +--rw revert-delay?      uint16
   |     +--rw endpoint-z
   |        +--rw (ac-or-pw-or-redundancy-grp)?
   |           +--:(ac)
   |           |  +--rw ac?                -> ../../ac/name
   |           +--:(pw)
   |           |  +--rw pw?                -> ../../pw/name
   |           +--:(redundancy-grp)
   |           |  +--rw (primary)
   |           |  |  +--:(primary-pw)
   |           |  |  |  +--rw primary-pw?        -> ../../pw/name
   |           |  |  +--:(primary-ac)
   |           |  |     +--rw primary-ac?        -> ../../ac/name
   |           |  +--rw (backup)
   |           |  |  +--:(backup-pw)
   |           |  |  |  +--rw backup-pw?         -> ../../pw/name
   |           |  |  +--:(backup-ac)
   |           |  |     +--rw backup-ac?         -> ../../ac/name
   |           |  +--rw protection-mode?   enumeration
   |           +--:(reroute-mode)
   |           |  +--rw reroute-mode?      enumeration
   |           +--:(reroute-delay)
   |           |  +--rw reroute-delay?     uint16
   |           +--:(dual-receive)
   |           |  +--rw dual-receive?      boolean
   |           +--:(revert)
   |           |  +--rw revert?            boolean
   |           +--:(revert-delay)
   |              +--rw revert-delay?      uint16
   +--rw vpls-instances
```

Figure 2

4. YANG Module

The L2VPN configuration container is logically divided into following high level config areas:

```
<CODE BEGINS> file "ietf-mpls-l2vpn@2015-06-30.yang"
module ietf-mpls-l2vpn {
  namespace "urn:ietf:params:xml:ns:yang:ietf-mpls-l2vpn";
  prefix "mpls-l2vpn";

  import ietf-inet-types {
    prefix "inet";
  }

  organization  "ietf";
  contact       "ietf";
  description   "mpls-l2vpn";
  revision "2015-06-30" {
    description "Initial revision";
    reference   "";
  }

  /* identities */

  identity link-discovery-protocol {
    description "Base identiy from which identities describing " +
                "link discovery protocols are derived.";
  }

  identity lacp {
    base "link-discovery-protocol";
    description "This identity represents LACP";
  }

  identity lldp {
    base "link-discovery-protocol";
    description "This identity represents LLDP";
  }

  identity bpdu {
    base "link-discovery-protocol";
    description "This identity represens BPDU";
  }

  identity cpd {
    base "link-discovery-protocol";
    description "This identity represents CPD";
  }

  identity udld {
    base "link-discovery-protocol";
    description "This identity represens UDLD";
  }
```

```
/* typedefs */

typedef l2vpn-service-type {
  type enumeration {
    enum ethernet {
      description "Ethernet service";
    }
    enum ATM {
      description "Asynchronous Transfer Mode";
    }
    enum FR {
      description "Frame-Relay";
    }
    enum TDM {
      description "Time Division Multiplexing";
    }
  }
  description "L2VPN service type";
}

typedef l2vpn-discovery-type {
  type enumeration {
    enum manual {
      description "Manual configuration";
    }
    enum bgp-ad {
      description "Border Gateway Protocol (BGP) auto-discovery";
    }
  }
  description "L2VPN discovery type";
}

typedef l2vpn-signaling-type {
  type enumeration {
    enum static {
      description "Static configuration of labels (no signaling)";
    }
    enum ldp {
      description "Label Distribution Protocol (LDP) signaling";
    }
    enum bgp {
      description "Border Gateway Protocol (BGP) signaling";
    }
  }
  description "L2VPN signaling type";
}

typedef bgp-rt-type {
```

```
    type enumeration {
      enum import {
        description "For import";
      }
      enum export {
        description "For export";
      }
      enum both {
        description "For both import and export";
      }
    }
    description "BGP route-target type. Import from BGP YANG";
  }

  typedef cw-negotiation-type {
    type enumeration {
      enum "non-preferred" {
        description "No preference for control-word";
      }
      enum "preferred" {
        description "Prefer to have control-word negotiation";
      }
    }
    description "control-word negotiation preference type";
  }

  typedef link-discovery-protocol-type {
    type identityref {
      base "link-discovery-protocol";
    }
    description "This type is used to identify " +
                "link discovery protocol";
  }

  typedef pw-template-ref {
    type leafref {
      path "/l2vpn/common/pw-templates/pw-template/name";
    }
    description "pw-template-ref";
  }

  typedef ac-template-ref {
    type leafref {
      path "/l2vpn/common/ac-templates/ac-template/name";
    }
    description "ac-tempalte-ref";
  }
```

```
     /* groupings */

     grouping vpws-endpoint {
       description
         "A vpws-endpoing could either be an ac or a pw";
       choice ac-or-pw-or-redundancy-grp {
         description "A choice ofattachment circuit or " +
                     "pseudowire or redundancy group";
         case ac {
           leaf ac {
             type leafref {
               path "../../ac/name";
             }
             description "reference to an attachment circuit";
           }
         }
         case pw {
           leaf pw {
             type leafref {
               path "../../pw/name";
             }
             description "reference to a pseudowire";
           }
         }
         case redundancy-grp {
           choice primary {
             mandatory true;
             description "primary options";
             case primary-pw  {
               leaf primary-pw {
                 type leafref {
                   path "../../pw/name";
                 }
                 description "primary pseudowire";
               }
             }
             case primary-ac {
               leaf primary-ac {
                 type leafref {
                   path "../../ac/name";
                 }
                 description "primary attachment circuit";
               }
             }
           }
           choice backup {
             mandatory true;
             description "backup options";
```

```
            case backup-pw {
              leaf backup-pw {
                type leafref {
                  path "../../pw/name";
                }
                description "backup pseudowire";
              }
            }
            case backup-ac {
              leaf backup-ac {
                type leafref {
                  path "../../ac/name";
                }
                description "backup attachment circuit";
              }
            }
          }
          leaf protection-mode {
            type enumeration {
              enum "frr" {
                value 0;
                description "fast reroute";
              }
              enum "master-slave" {
                value 1;
                description "master-slave";
              }
              enum "independent" {
                value 2;
                description "independent";
              }
            }
            description "protection-mode";
          }
        }
        leaf reroute-mode {
          type enumeration {
            enum "immediate" {
              value 0;
              description "immediate reroute";
            }
            enum "delayed" {
              value 1;
              description "delayed reroute";
            }
            enum "never" {
              value 2;
              description "never reroute";
```

```
            }
          }
          description "reroute-mode";
        }
        leaf reroute-delay {
          when "../reroute-mode = 'delayed'" {
            description
              "Specify amount of time to delay reroute " +
              "only when delayed route is configured";
          }
          type uint16;
          description
            "amount of time to delay reroute";
        }
        leaf dual-receive {
          type boolean;
          description
            "allow extra traffic to be carried by backup";
        }
        leaf revert {
          type boolean;
          description
            "allow forwarding to revert to primary " +
            "after restoring primary";
          /* This is called "revertive" during the discussion. */
        }
        leaf revert-delay {
          when "../revert = 'true'" {
            description
              "Specify the amount of time to wait to revert " +
              "to primary only if reversion is configured";
          }
          type uint16;
          description
            "amount ot time to wait to revert to primary";
          /* This is called "wtr" during discussion. */
        }
      }
    }

    /* We can define vpls-endpoing-grp that has the same structure as
     * vpws-endpoing-grp, but has more endpoint options.
     */

    /* L2VPN YANG Model */

    container l2vpn {
      description "l2vpn";
```

```
       container common {
         description "common l2pn attributes";
         container pw-templates {
           description "pw-templates";
           list pw-template {
             key "name";
             description "pw-template";
             leaf name {
               type string;
               description "name";
             }
             leaf mtu {
               type uint32;
               description "pseudowire mtu";
             }
             leaf cw-negotiation {
               type cw-negotiation-type;
               default "preferred";
               description
                 "control-word negotiation preference";
             }
             leaf tunnel-policy {
               type string;
               description "tunnel policy name";
             }
           }
         }
         container ac-templates {
           description "attachment circuit templates";
           /* To be fleshed out in future revisions */
           list ac-template {
             key "name";
             description "ac-template";
             leaf name {
               type string;
               description "name";
             }
           }
         }
       }
       container vpws-instances {
         description "vpws-instances";
         list vpws-instance {
           key "instance-name";
           description "A VPWS instance";
           leaf instance-name {
             type string;
             description "Name of VPWS instance";
```

```
          }
          leaf description {
            type string;
            description "Description of the VPWS instance";
          }
          leaf service-type {
            type l2vpn-service-type;
            default ethernet;
            description "VPWS service type";
          }
          leaf discovery-type {
            type l2vpn-discovery-type;
            default manual;
            description "VPWS discovery type";
          }
          leaf signaling-type {
            type l2vpn-signaling-type;
            mandatory true;
            description "VPWS signaling type";
          }
          container bgp-parameters {
            description "Parameters for BGP";
            container common {
              when "../../discovery-type = 'bgp-ad'" {
                description "Check discovery type: " +
                            "Can only configure BGP discovery if " +
                            "discovery type is BGP-AD";
              }
              description "Common BGP parameters";
              leaf route-distinguisher {
                type string;
                description "BGP RD";
              }
              list vpn-targets {
                key rt-value;
                description "Route Targets";
                leaf rt-value {
                  type string;
                  description "Route-Target value";
                }
                leaf rt-type {
                  type bgp-rt-type;
                  mandatory true;
                  description "Type of RT";
                }
              }
            }
            container discovery {
```

```
              when "../../discovery-type = 'bgp-ad'" {
                description "BGP parameters for discovery: " +
                            "Can only configure BGP discovery if " +
                            "discovery type is BGP-AD";
              }
              description "BGP parameters for discovery";
              leaf vpn-id {
                type string;
                description "VPN ID";
              }
            }
            container signaling {
              when "../../signaling-type = 'bgp'" {
                description "Check signaling type: " +
                            "Can only configure BGP signaling if " +
                            "signaling type is BGP";
              }
              description "BGP parameters for signaling";
              leaf site-id {
                type uint16;
                description "Site ID";
              }
              leaf site-range {
                type uint16;
                description "Site Range";
              }
            }
          }
          list pw {
            key "name";
            description "pseudowire";
            leaf name {
              type string;
              description "pseudowire name";
            }
            leaf cw-negotiation {
              type cw-negotiation-type;
              default "preferred";
              description "Override the control-word negotiation " +
                          "preference specified in the " +
                          "pseudowire template.";
            }
            leaf template {
              type pw-template-ref;
              description "pseudowire template";
            }
            leaf vccv-ability {
              type boolean;
```

```
              description "vccvability";
            }
            leaf tunnel-policy {
              type string;
              description "Used to override the tunnel policy name " +
                          "specified in the pseduowire template";
            }
            leaf request-vlanid {
              type uint16;
              description "request vlanid";
            }
            leaf vlan-tpid {
              type string;
              description "vlan tpid";
            }
            leaf ttl {
              type uint8;
              description "time-to-live";
            }
            choice pw-type {
              description "A choice of pseudowire type";
              case ldp-pw {
                leaf peer-ip {
                  type inet:ip-address;
                  description "peer IP address";
                }
                leaf pw-id {
                  type uint32;
                  description "pseudowire id";
                }
                leaf transmit-label {
                  type uint32;
                  description "transmit lable";
                }
                leaf receive-label {
                  type uint32;
                  description "receive label";
                }
                leaf icb {
                  type boolean;
                  description "inter-chassis backup";
                }
              }
              case bgp-pw {
                leaf remote-pe-id {
                  type inet:ip-address;
                  description "remote pe id";
                }
```

```
              }
              case bgp-ad-pw {
                leaf remote-ve-id {
                  type uint16;
                  description "remote ve id";
                }
              }
            }
          }
          list ac {
            key "name";
            description "attachment circuit";
            leaf name {
              type string;
              description "name";
            }
            leaf template {
              type ac-template-ref;
              description "attachment circuit template";
            }
            leaf pipe-mode {
              type enumeration {
                enum "pipe" {
                  value 0;
                  description "regular pipe mode";
                }
                enum "short-pipe" {
                  value 1;
                  description "short pipe mode";
                }
                enum "uniform" {
                  value 2;
                  description "uniform pipe mode";
                }
              }
              description "pipe mode";
            }
            leaf link-discovery-protocol {
              type link-discovery-protocol-type;
              description "link discovery protocol";
            }
          }
          container endpoint-a {
            description "endpoint-a";
            uses vpws-endpoint;
          }
          container endpoint-z {
            description "endpoint-z";
```

```
            uses vpws-endpoint;
          }
        }
      }
      container vpls-instances {
        /* To be fleshed out in future revisions */
        description "vpls-instances";
      }
    }
  }

  <CODE ENDS>
```

Figure 3

5. Security Considerations

The configuration, state, action and notification data defined in this document are designed to be accessed via the NETCONF protocol [RFC6241]. The lowest NETCONF layer is the secure transport layer and the mandatory-to-implement secure transport is SSH [RFC6242]. The NETCONF access control model [RFC6536] provides means to restrict access for particular NETCONF users to a pre-configured subset of all available NETCONF protocol operations and content.

The security concerns listed above are, however, no different than faced by other routing protocols. Hence, this draft does not change any underlying security issues inherent in [I-D.ietf-netmod-routing-cfg]

6. IANA Considerations

None.

7. Acknowledgments

The authors would like to acknowledge TBD for their useful comments.

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.

8.2. Informative References

[RFC3916] Xiao, X., McPherson, D., and P. Pate, "Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)", RFC 3916, September 2004.

[RFC3985] Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, March 2005.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, February 2006.

[RFC4446] Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)", BCP 116, RFC 4446, April 2006.

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

[RFC4448] Martini, L., Rosen, E., El-Aawar, N., and G. Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks", RFC 4448, April 2006.

[RFC4664] Andersson, L. and E. Rosen, "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, September 2006.

[RFC4665] Augustyn, W. and Y. Serbest, "Service Requirements for Layer 2 Provider-Provisioned Virtual Private Networks", RFC 4665, September 2006.

[RFC4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007.

[RFC5003] Metz, C., Martini, L., Balus, F., and J. Sugimoto, "Attachment Individual Identifier (AII) Types for Aggregation", RFC 5003, September 2007.

[RFC5254] Bitar, N., Bocci, M., and L. Martini, "Requirements for Multi-Segment Pseudowire Emulation Edge-to-Edge (PWE3)", RFC 5254, October 2008.

[RFC5659] Bocci, M. and S. Bryant, "An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge", RFC 5659, October 2009.

[RFC6020] Bjorklund, M., "YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)", RFC 6020, October 2010.

[RFC6073] Martini, L., Metz, C., Nadeau, T., Bocci, M., and M. Aissaoui, "Segmented Pseudowire", RFC 6073, January 2011.

[RFC6074] Rosen, E., Davie, B., Radoaca, V., and W. Luo, "Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs)", RFC 6074, January 2011.

[RFC6241] Enns, R., Bjorklund, M., Schoenwaelder, J., and A. Bierman, "Network Configuration Protocol (NETCONF)", RFC 6241, June 2011.

[RFC6242] Wasserman, M., "Using the NETCONF Protocol over Secure Shell (SSH)", RFC 6242, June 2011.

[RFC6391] Bryant, S., Filsfils, C., Drafz, U., Kompella, V., Regan, J., and S. Amante, "Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network", RFC 6391, November 2011.

[RFC6423] Li, H., Martini, L., He, J., and F. Huang, "Using the Generic Associated Channel Label for Pseudowire in the MPLS Transport Profile (MPLS-TP)", RFC 6423, November 2011.

[RFC6478] Martini, L., Swallow, G., Heron, G., and M. Bocci, "Pseudowire Status for Static Pseudowires", RFC 6478, May 2012.

[RFC6536] Bierman, A. and M. Bjorklund, "Network Configuration Protocol (NETCONF) Access Control Model", RFC 6536, March 2012.

[RFC6624] Kompella, K., Kothari, B., and R. Cherukuri, "Layer 2 Virtual Private Networks Using BGP for Auto-Discovery and Signaling", RFC 6624, May 2012.

[RFC7041] Balus, F., Sajassi, A., and N. Bitar, "Extensions to the Virtual Private LAN Service (VPLS) Provider Edge (PE) Model for Provider Backbone Bridging", RFC 7041, November 2013.

[RFC7361] Dutta, P., Balus, F., Stokes, O., Calvignac, G., and D. Fedyk, "LDP Extensions for Optimized MAC Address Withdrawal in a Hierarchical Virtual Private LAN Service (H-VPLS)", RFC 7361, September 2014.

Authors' Addresses

Himanshu Shah
Ciena Corporation

Email: hshah@ciena.com

Patrice Brissette
Cisco Systems, Inc.

Email: pbrisset@cisco.com

Reshad Rahman
Cisco Systems, Inc.

Email: rrahman@cisco.com

Kamran Raza
Cisco Systems, Inc.

Email: skraza@cisco.com

Zhenbin Li
Huawei Technologies

Email: lizhenbin@huawei.com

Zhuang Shunwan
Huawei Technologies

Email: Zhuangshunwan@huawei.com

Wang Haibo
Huawei Technologies

Email: rainsword.wang@huawei.com

Ing-When Chen
Ericsson

Email: ing-wher.chen@ericsson.com

Mathew Bocci
Alcatel-Lucent

Email: mathew.bocci@alcatel-lucent.com

Jonathan Hardwick
Metaswitch

Email: jonathan.hardwick@metaswitch.com

Santosh Esale
Junipr Networks

Email: sesale@juniper.net

Kishore Tiruveedhula
Juniper Networks

Email: kishoret@juniper.net

Tapraj Singh
Juniper Networks

Email: tsingh@juniper.net

Iftekar Hussain
Infinera Corporation

Email: ihussain@infinera.com

Bin Wen
Comcast

Email: Bin_Wen@cable.comcast.com

Jason Walker
Comcast

Email: jason_walker2@cable.comcast.com

Nick Delregno
Verizon

Email: nick.deregno@verizon.com

Luay Jalil
Verizon

Email: luay.jalil@verizon.com

Maria Joecylyn
Verizon

Email: joecylyn.malit@verizon.com