L2VPN Working Group

Pranjal Kumar Dutta
Marc Lasserre
Alcatel-Lucent

Internet Draft
Intended status: Standard
Expires: September 2009

Olen Stokes
Extreme Networks

March 8, 2009

LDP Extensions for Optimized MAC Address Withdrawal in H-VPLS
draft-pdutta-l2vpn-vpls-ldp-mac-opt-04.txt

Status of this Memo

Abstract

[RFC4762] describes a mechanism to remove or unlearn MAC addresses that have been dynamically learned in a VPLS Instance for faster convergence on topology change. The procedure also removes the MAC addresses in the VPLS that does not require relearning due to such topology change. This document defines an extension to MAC Address Withdrawal procedure with empty MAC List [RFC4762], which enables a Provider Edge(PE) device to remove only the MAC addresses that needs

to be relearned.

Conventions used in this document

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

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

Table of Contents

Terminology

This document uses the terminology defined in [RFC5036], [RFC4447] and [RFC4762]. Throughout this document VPLS means the emulated bridged LAN service offered to a customer. H-VPLS means the hierarchical connectivity or layout of MTU-s and PE devices offering the VPLS[RFC4762]. The terms spoke node and MTU-s in H-VPLS are used interchangeably.

1. Introduction

A method of Virtual Private LAN Service (VPLS), also known as Transparent LAN Service (TLS) is described in [RFC4762]. A VPLS is created using a collection of one or more point-to-point pseudowires (PWs) [RFC4664] configured in a flat, full-mesh topology. The mesh topology provides a LAN segment or broadcast domain that is fully capable of learning and forwarding on Ethernet MAC addresses at the PE devices.

This VPLS full mesh core configuration can be augmented with additional non-meshed spoke nodes to provide a Hierarchical VPLS (H-VPLS) service[RFC4762].

MAC Address Withdrawal mechanism is described in [RFC4762] to remove or unlearn MAC addresses for faster convergence on topology change in dual-homed H-VPLS[RFC4762]. In dual-homed H-VPLS, an edge device termed as MTU-s is connected to two PE devices via primary spoke PW and backup spoke PW respectively. Such redundancy is designed to protect against the failure of primary spoke PW or primary PE device. When MTU-s switches over to backup PW, it is required to flush the MAC addresses learned in the VPLS in upstream PE devices participating in full mesh, to avoid black holing of frames to those addresses. Note that forced switchover to backup PW can be also performed at MTU-s administratively due to maintenance activities on the primary spoke PW.

When backup PW is made active by MTU-s, it triggers LDP Address Withdraw Message with list of MAC addresses to be flushed. The message is forwarded over the LDP session(s) associated with the newly activated PW. In order to minimize the impact on LDP convergence time when MAC List TLV contains a large number of MAC addresses, it may be preferable to send a LDP Address Withdraw Message with an empty MAC List. Throughout this document the term MAC Flush Message is used to specify LDP Address Withdraw Message with empty MAC List described in [RFC4762] unless specified otherwise.

As per the processing rules in [RFC4762] a PE device on receiving a MAC flush message removes all MAC addresses associated with the specified VPLS instance (as indicated in the FEC TLV) except the MAC addresses learned over the newly activated PW. The PE device further triggers MAC flush message to each remote PE device

connected to it in the VPLS full mesh.

This method of MAC flushing is modeled after Topology Change Notification (TCN) in Rapid Spanning Tree Protocol (RSTP)[802.1w]. When a bridge switches from failed link to the backup link, the bridge sends out a TCN message over the newly activated link. The upstream bridge upon receiving this message flushes its entire MAC addresses except the ones received over this link and sends the TCN message out of its other ports in that spanning tree instance. The message is further relayed along the spanning tree by the other bridges.

When a PE device in the full-mesh of H-VPLS receives a MAC flush message it also flushes the MAC addresses which are not affected due to topology change, thus leading to unnecessary flooding and relearning. This document describes the problem and a solution to optimize the MAC flush procedure in [RFC4762] so that flush only the set of MAC addresses that require relearning when topology changes in H-VPLS.

[L2VPN-SIG] describes about inter-AS VPLS deployments models where Multi-Segment PWs (MS-PW) may be used. The solution proposed in this document is generic and is applicable to MS-PWs in H-VPLS.

2. Problem Description

Figure.1 describes a dual-homed H-VPLS scenerio for a VPLS instance where the problem with the existing MAC flush method in [RFC4762] is explained.

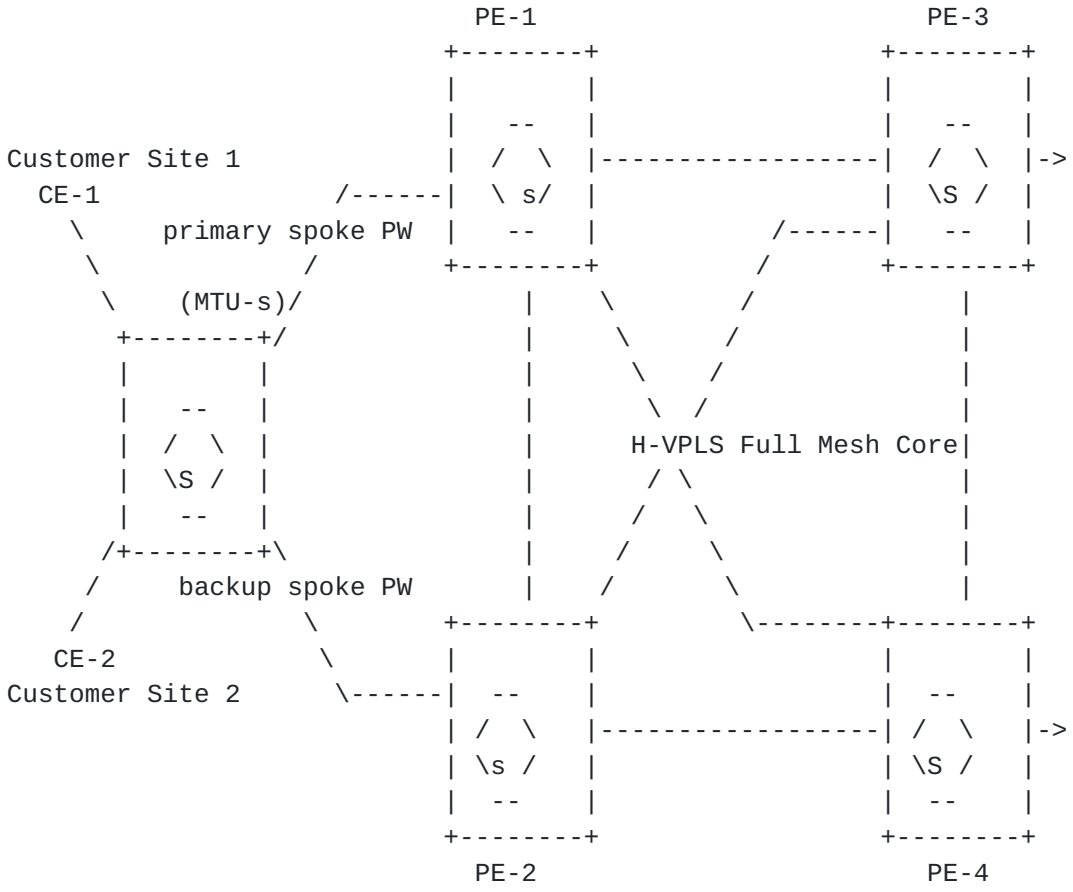

Figure 1: Dual homed MTU-s in two tier hierarchy H-VPLS

In Figure.1, the MTU-s is dual-homed to PE-1 and PE-2. Only the primary spoke PW is active at MTU-s, thus PE-1 acting as the primary device to reach the full mesh in the VPLS instance. The MAC addresses of nodes located at access sites (behind CE1 and CE2) are learned at PE-1 over the primary spoke PW. PE-2, PE-3 and PE-4 learn those MAC addresses on their respective mesh PWs terminating to PE-1.

When MTU-s switches to backup spoke PW and activates it, PE-2

becomes the primary device to reach the full mesh core. Traffic entering the H-VPLS from CE-1 and CE-2 is diverted by MTU-s to the backup spoke PW. For faster convergence MTU-s may desire to unlearn or remove the MAC addresses that have been learned in the upstream VPLS full-mesh through PE-1. MTU-s may send MAC flush message to PE-2 once the backup PW has been made active. As per the processing rules defined in [RFC4762], PE-2 flushes the entire MAC addresses learned in the VPLS from the PWs terminating at PE-1, PE-3 and PE-4 except the ones learned over the newly activated spoke PW.

In the H-VPLS core, PE devices are connected in full mesh unlike the spanning tree connectivity in bridges. So the MAC addresses that require flushing and relearning at PE-2 are only the MAC addresses those have been learned on the PW connected to PE-1.

PE-2 further relays MAC flush messages to all other PE devices in the full mesh. Same processing rule applies at all those PE devices. For example, at PE-3 the entire MAC addresses learned from the PWs connected to PE-1 and PE-4 are flushed and relearned subsequently. As the number of PE devices in the full-mesh increases, the number of unaffected MAC addresses flushed in a VPLS instance also increases, thus leading to unnecessary flooding and relearning. With large number of VPLS instances provisioned in the H-VPLS network topology the amount of unnecessary flooding and relearning increases.

3. Optimized MAC Flush Mechanism

The basic principle of the optimized MAC flush mechanism is explained with reference to Figure 1. On switching over to the backup spoke PW when MTU-s triggers MAC flush message to PE-2, it also communicates the unique PW endpoint identifier (PE-ID) in PE-1, the formerly active PE device. In VPLS a PW terminates on a Virtual Switching Instance (VSI) in a PE device. The PE-ID is relayed in all the subsequent MAC flush messages triggered by PE-2 to its peer PE devices in the full mesh. Each PE device that receives the message identifies the VPLS (From FEC TLV) and its respective PW that terminates in PE-1 (from PE-ID). Thus the PE device flushes only the MAC addresses learned from that PW connected to PE-1.

4. PE-ID TLV

This document defines a PW Endpoint Identifier (PE-ID) TLV for LDP [RFC5036]. The PE-ID TLV carries the unique identifier of a generic PW endpoint.

The encoding of PE-ID TLV follows standard LDP TLV encoding in [RFC5036]. A PE-ID TLV contains a list of one or more PE-ID Elements. Its encoding is:

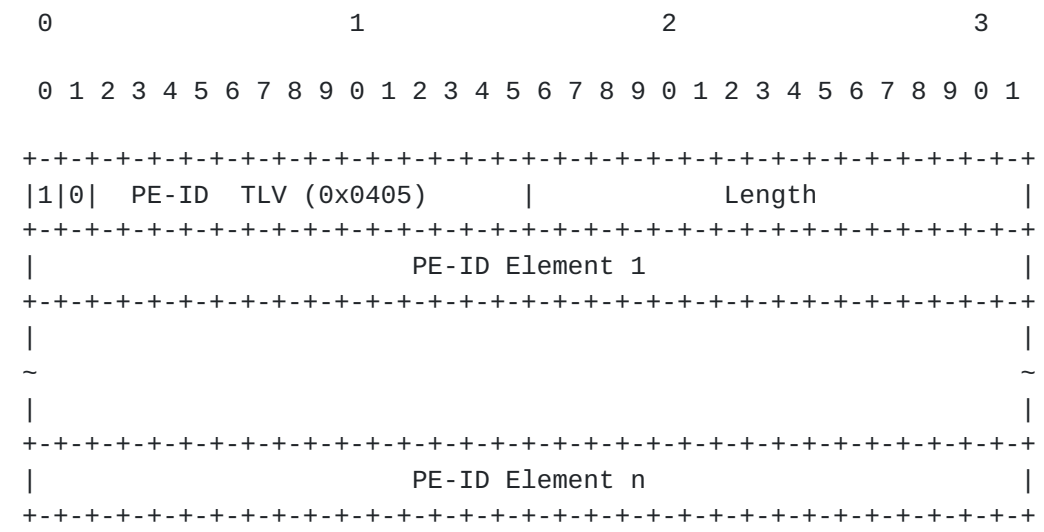

Figure 2. PE-ID TLV

U (Unknown) bit of thus LDP TLV MUST be set to 1. If the PE-ID TLV is not understood then it is ignored the receiving device.

F (Forward) MUST be set to 0. Since the LDP mechanism used here is targeted, the TLV is not forwarded if it is not understood by the receiving device.

The Type field MUST be set to 0x405 (subject to IANA approval). This identifies the TLV type as PE-ID TLV.

Length field specifies the total length in octets of the Value in PE-ID TLV.

PE-ID Element 1 to PE-ID Element n:
    There are several types of PE-ID Elements. The PE-ID Element

Encoding depends on the type of the PE-ID Element. A PE-ID Element uniquely identifies a PW Endpoint.

A PE-ID Element value is encoded as 1 octet field that specifies the element type, 1 octet field that identifies the length in octets of the element value, and a variable length field that is type dependent element value.

The PE-ID Element value encoding is:

PE-ID Element Type name	Type	Length	Value
FEC-128 specific	0x01	2 octets	See below.
FEC-129 specific	0x02	Variable	See below.

The type of PE-ID Element depends on the type of FEC Element used to provision the respective PW.

[RFC4447] defines two types of FEC elements that may be used for provisioning PWs - Pwid FEC (type 128) and the Generalized ID (GID) FEC (type 129).

The Pwid FEC element includes a fixed-length 32 bit value called the PWid. The same PWid value must be configured on the local and remote PE prior to PW setup.

The GID FEC element includes TLV fields for attachment individual identifiers (AII) that, in conjunction with an attachment group identifier (AGI), serve as PW endpoint identifiers. The endpoint identifier on the local PE (denoted as <AGI, source AII or SAII>) is called the source attachment identifier (SAI) and the endpoint identifier on the remote PE (denoted as <AGI, target AII or TAII>) is called the target attachment identifier (TAI). The SAI and TAI can be distinct values. This is useful for provisioning models where the local PE (with a particular SAI) does not know and must somehow learn (e.g. via MP-BGP auto-discovery) of remote TAI values prior to launching PW setup messages towards the remote PE.

FEC-128 specific PE-ID Element

This sub-type is to be used to identify a PW endpoint only if Pwid FEC Element is used for signaling the PW. The encoding of this PE-ID element is as follows:

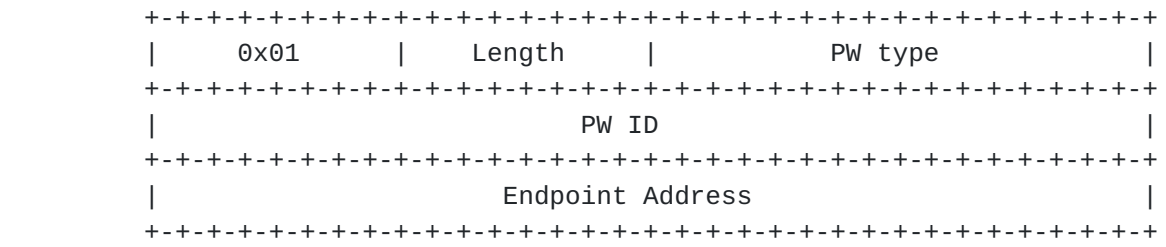

Figure 3. FEC-128 specific PE-ID Element

PW type
The PW Type value from PWid FEC element.

PW ID
The PW ID value from the Pwid FEC element.

Endpoint Address
32-bit LSR-ID from the LDP-ID used in LDP signaling session by a PW endpoint.

FEC-129 specific PE-ID element
This sub-type is to be used to indentify a PW endpoint only if GID FEC Element is used for signaling the PW. The encoding of this PE-ID element is as follows:

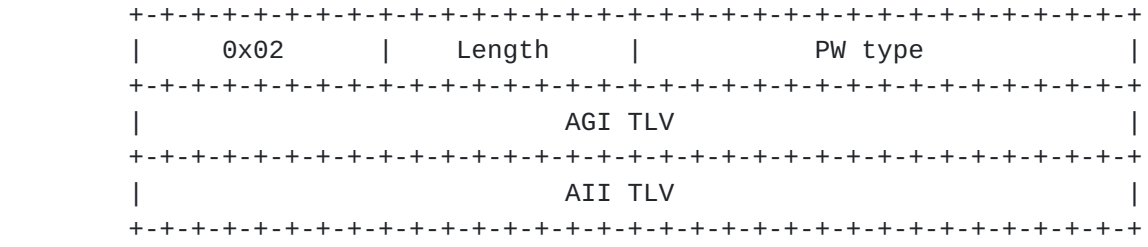

Figure 4. FEC-129 specific PE-ID Element

PW type
The PW Type value from GID FEC element.

PW ID
The PW ID value from the GID FEC element

AGI TLV
The AGI from the corresponding GID Element

AII TLV
The AII associated with the PW endpoint.

5. Application of PE-ID TLV in Optimized MAC Flush

For optimized MAC flush, the PE-ID TLV MAY be sent as an OPTIONAL parameter in existing LDP Address Withdraw Message with empty MAC List. The PE-ID TLV carries the unique PW endpoint identifier in a VPLS as described in section 4.

It is to note that for optimized MAC flush the PE-ID TLV carries sufficient information for identifying the VPLS instance and the unique VSI Identifier. For backward compatibility with MAC flush procedures in [RFC4762] both FEC TLV and PE-ID TLV should be sent in the MAC flush message. However the inclusion of the FEC-TLV should be based on what would be the desired effect should the PE-ID not be understood by the receiver. In cases where the desired action when the PE-ID is not understood would be to behave as described in [RFC4762], then the FEC TLV SHOULD be included. In cases where the desired action when the PE-ID is not understood is to take no action, then the FEC TLV SHOULD NOT be included. The PE-ID TLV SHOULD carry the unique VSI identifier in the VPLS instance (specified in the FEC TLV). The PE-ID TLV SHOULD be placed after the existing TLVs in MAC Flush message in [RFC4762].

5.1 PE-ID TLV Processing Rules

This section describes the processing rules of PE-ID TLV that SHOULD be followed in the context of MAC flush procedures in an H-VPLS.

When an MTU-s triggers MAC flush after activation of backup spoke PW, it MAY send the PE-ID TLV that identifies VSI in the formerly active PE device. There may be cases where a PE device in full mesh initiates MAC flush torwards the core when it detects a spoke PW failure. In such a case the PE-ID TLV in MAC flush message MAY identify its own VSI in the VPLS instance. Irrespective of whether it is the MTU-s or PE device that initiates the MAC flush, a PE device receiving the PE-ID TLV SHOULD follow the same processing rules as described in this section.

If MS-PW signaled with GID Element is used in VPLS then a MAC flush message is processed only at the T-PE nodes. In this section, a PE device signifies only T-PE in MS-PW case unless specified otherwise.

When a PE device receives a MAC flush with PE-ID TLV, it SHOULD flush all the MAC addresses learned in the VPLS from the PW that terminates in the remote VSI identified by the PE-ID element.

If a PE-ID element received in the MAC flush message identifies the local VSI, it SHOULD flush the MAC addresses learned from its local spoke PW(s) in the VPLS instance.

If a PE device receives a MAC flush with the PE-ID TLV option And a valid MAC address list, it SHOULD ignore the option and deal with MAC addresses explicitly as per [RFC4762].

If a PE device that doesn't support PE-ID TLV receives a MAC flush message with this option, it MUST ignore the option and follow the processing rules as per [RFC4762].

5.2 Optimized MAC Flush Procedures

This section explains the optimized MAC flush procedure in the scenario in Figure.1.

When the backup PW is activated by MTU-s, it may send MAC flush message to PE-2 with the optional PE-ID TLV. The PE-ID element carries the VSI identifier in PE-1 for the VPLS. Upon receipt of the MAC flush message, PE-2 identifies the VPLS instance that requires MAC flush from the FEC element in the FEC TLV. From the PE-ID TLV, PE-2 identifies the PW in the VPLS that terminates in PE-1. PE-2

removes all MAC addresses learned from that PW.

PE-2 relays MAC flush messages with the received PE-ID to all its peer PE devices. When the message is received at PE-3, it identifies the PW that terminates in the remote VSI in PE-1. PE-3 removes all MAC addresses learned on the PW that terminated in PE1.

There may be redundancy scenerios where a PE device in the full mesh may be required to initiate optimized MAC Address Withdrawal. Figure 5. shows a redundant H-VPLS topology to protect against failure of MTU-s device. Provider RSTP may be used as selection algorithm for active and backup PWs in order to maintain the connectivity between MTU devices and PE devices at the edge. It is assumed that PE devices can detect failure on PWs in either direction through OAM mechanisms such as VCCV procedures for instance.

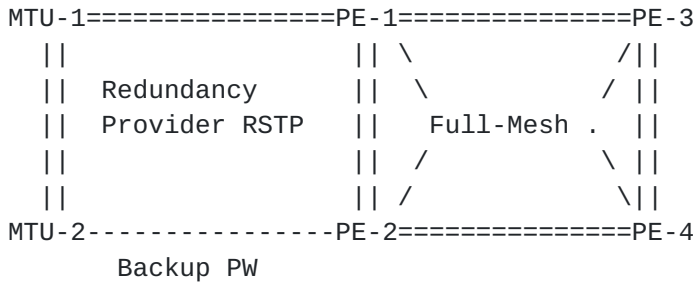

Figure 5. Redundancy with Provider RSTP

MTU-1, MTU-2, PE-1 and PE-2 participate in provider RSTP. By configuration in RSTP it is ensured that the PW between MTU-1 and PE-1 is active and the PW between MTU-2 and PE-2 is blocked (made backup) at MTU-2 end. When the active PW failure is detected by RSTP, it activates the PW between MTU-2 and PE-2. When PE-1 detects the failing PW to MTU-1, it may trigger MAC flush into the full mesh with PE-ID TLV that carries its own VSI identifier in the VPLS. Other PE devices in the full mesh that receive the MAC flush message identify their respective PWs terminating on PE-1 and flush all the MAC addresses learned from it.

By default, MTU-2 should still trigger MAC flush as currently

defined in [RFC4762] after the backup PW is made active by RSTP. Mechanisms to prevent two copies of MAC withdraws to be sent in such scenerios is out of scope of this document.

[RFC4762] describes multi-domain VPLS service where fully meshed VPLS networks (domains) are connected together by a single spoke PW per VPLS service between the VPLS "border" PE devices. To provide redundancy against failure of the inter-domain spoke, full mesh of inter-domain spokes can be setup between border PE devices and provider RSTP may be used for selection of the active inter-domain spoke. In case of inter-domain spoke PW failure, PE initiated MAC withdrawal may be used for optimized MAC flushing within individual domains.

6. General Applicability of PE-ID TLV

This document defines the PE-ID TLV and its application in optimized flushing of MAC addresses in H-VPLS on topology change. The use of PE-ID TLV in MAC flush mechanism is OPTIONAL.

Different H-VPLS redundancy scenarios are possible. The use of the PE-ID TLV applies to H-VPLS dual-homing scenarios described in this document. The use of the PE-ID TLV in different scenarios is out of scope. The protocols used in redundancy scenarios are outside the scope. The document doesn't specify the device that should trigger optimized MAC flush. The method to select the appropriate PE-ID in various redundancy scenarios is out of scope.

There may be other L2VPN applications where PE-ID TLV may be applicable and such applications are outside the scope of this document.

7. Security Considerations

- Control plane aspects
  - LDP security (authentication) methods as described in [RFC5036] is applicable here. Further this document implements security considerations as in [RFC4447] and [RFC4762].

- Data plane aspects

- This specification does not have any impact on the VPLS forwarding plane.

8. IANA Considerations

The Type field in PE-ID TLV is defined as 0x405 and is subject to IANA approval.

9. Acknowledgments

The authors would like to thank the following people who have provided valuable comments and feedback on the topics discussed in this document:

Florin Balus
Prashanth Ishwar
Vipin Jain
John Rigby
Ali Sajassi

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

10. References

10.1. Normative References

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

[RFC4762] Lasserre, M. and Kompella, V., "Virtual Private LAN Services using LDP", RFC 4762, January 2007.

[RFC5036] Andersson, L., et al. "LDP Specification",RFC5036, October 2007.

10.2. Informative References

[L2VPN-SIG] Rosen, E., et al. .Provisioning, Autodiscovery, and Signaling in L2VPNs., work in progress.

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

[RFC4447] Martini. and et al., "Pseudowire Setup and Maintenance Using Label Distribution Protocol (LDP)", RFC 4447, April 2006.

[802.1w] "IEEE Standard for Local and metropolitan area networks. Common specifications Part 3: Media Access Control (MAC) Bridges. Amendment 2: Rapid Reconfiguration", IEEE Std 802.1w-2001.

Author's Addresses

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

Email: pdutta@alcatel-lucent.com

Marc Lasserre
Alcatel-Lucent

Email: marc.lasserre@alcatel-lucent.com

Olen Stokes
Extreme Networks
PO Box 14129
RTP, NC 27709
USA

Email: ostokes@extremenetworks.com

Copyright Notice

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.