Internet Engineering Task Force (IETF)
Request for Comments: 7361
Category: Standards Track
ISSN: 2070-1721

P. Dutta
F. Balus
Alcatel-Lucent
O. Stokes
Extreme Networks
G. Calvignac
Orange
D. Fedyk
Hewlett-Packard

September 2014

LDP Extensions for Optimized MAC Address Withdrawal in a Hierarchical Virtual Private LAN Service (H-VPLS)

Abstract

RFC 4762 describes a mechanism to remove or unlearn Media Access Control (MAC) addresses that have been dynamically learned in a Virtual Private LAN Service (VPLS) instance for faster convergence on topology changes. The procedure also removes MAC addresses in the VPLS that do not require relearning due to such topology changes. This document defines an enhancement to the MAC address withdraw procedure with an empty MAC list (RFC 4762); this enhancement enables a Provider Edge (PE) device to remove only the MAC addresses that need to be relearned. Additional extensions to RFC 4762 MAC withdraw procedures are specified to provide an optimized MAC flushing for the Provider Backbone Bridging (PBB) VPLS specified in RFC 7041.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7361.

Table of Contents

1. Introduction

A method of Virtual Private LAN Service (VPLS), also known as Transparent LAN Services (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 Media Access Control (MAC) addresses at the Provider Edge (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]. Throughout this document, this configuration is referred to as "regular" H-VPLS.

[RFC7041] describes how Provider Backbone Bridging (PBB) can be integrated with VPLS to allow for useful PBB capabilities while continuing to avoid the use of the Multiple Spanning Tree Protocol (MSTP) in the backbone. The combined solution, referred to as "PBB-VPLS", results in better scalability in terms of number of service instances, PWs, and C-MAC (Customer MAC) addresses that need to be handled in the VPLS PEs, depending on the location of the I-component in the PBB-VPLS topology.

A MAC address withdrawal mechanism for VPLS is described in [RFC4762] to remove or unlearn MAC addresses for faster convergence on topology changes in resilient H-VPLS topologies. Note that the H-VPLS topology discussed in [RFC4762] describes the two-tier hierarchy in VPLS as the basic building block of H-VPLS, but it is possible to have a multi-tier hierarchy in an H-VPLS.

Figure 1 is reproduced from [RFC4762] and illustrates dual-homing in H-VPLS.

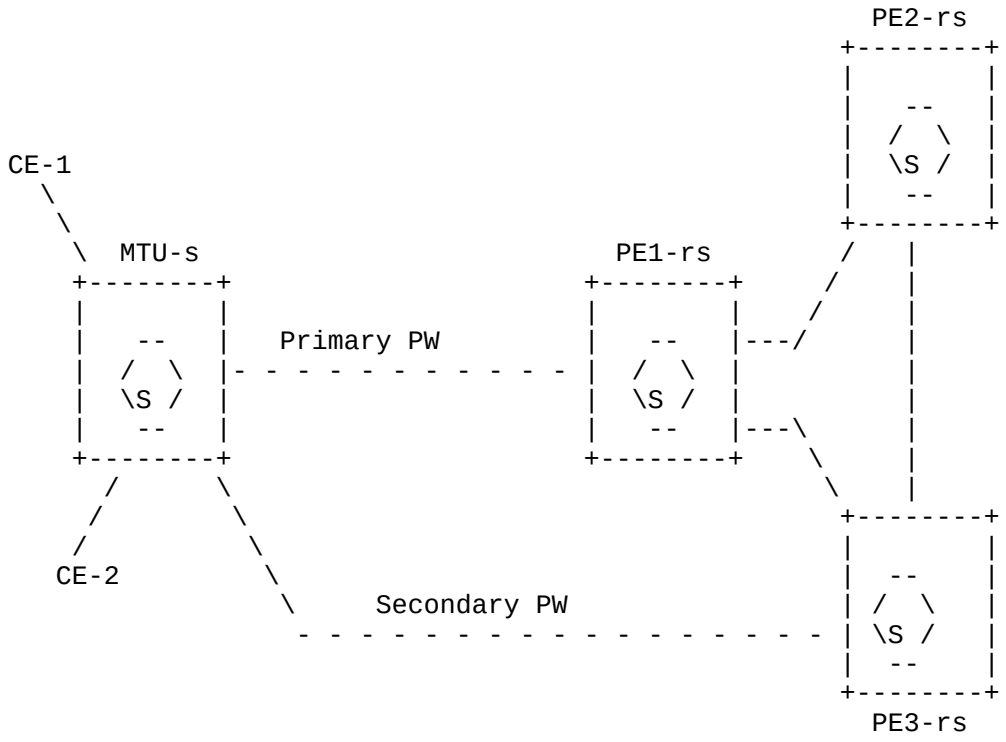


Figure 1: An Example of a Dual-Homed MTU-s

An example usage of the MAC flushing mechanism is the dual-homed H-VPLS where an edge device called the Multi-Tenant Unit switch (MTU-s) [RFC4762] is connected to two PE devices via a primary spoke PW and backup spoke PW, respectively. Such redundancy is designed to protect against the failure of the primary spoke PW or primary PE device. There could be multiple methods of dual-homing in H-VPLS that are not described in [RFC4762]. For example, note the following statement from Section 10.2.1 of [RFC4762].

> How a spoke is designated primary or secondary is outside the scope of this document. For example, a spanning tree instance running between only the MTU-s and the two PE-rs nodes is one possible method. Another method could be configuration.

This document intends to clarify several H-VPLS dual-homing models that are deployed in practice and various use cases of LDP-based MAC flushing in these models.

2. Terminology

This document uses the terminology defined in [RFC7041], [RFC5036], [RFC4447], and [RFC4762].

Throughout this document, "Virtual Private LAN Service" (VPLS) refers to the emulated bridged LAN service offered to a customer. "H-VPLS" refers to the hierarchical connectivity or layout of the MTU-s and the Provider Edge routing- and switching-capable (PE-rs) devices offering the VPLS [RFC4762].

The terms "spoke node" and "MTU-s" in H-VPLS are used interchangeably.

"Spoke PW" refers to the Pseudowire (PW) that provides connectivity between MTU-s and PE-rs nodes.

"Mesh PW" refers to the PW that provides connectivity between PE-rs nodes in a VPLS full-mesh core.

"MAC flush message" refers to a Label Distribution Protocol (LDP) address withdraw message without a MAC List TLV.

A MAC flush message "in the context of a PW" refers to the message that has been received over the LDP session that is used to set up the PW used to provide connectivity in VPLS. The MAC flush message carries the context of the PW in terms of the Forwarding Equivalence Class (FEC) TLV associated with the PW [RFC4762] [RFC4447].

In general, "MAC flushing" refers to the method of initiating and processing MAC flush messages across a VPLS instance.

2.1. 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 [RFC2119].

3. Overview

When the MTU-s switches over to the backup PW, the requirement is to flush the MAC addresses learned in the corresponding Virtual Switch Instance (VSI) in peer PE devices participating in the full mesh, to avoid the black-holing of frames to those addresses. This is accomplished by sending an LDP address withdraw message -- a new message defined in this document -- from the PE that is no longer

connected to the MTU-s with the primary PW. The new message contains a list of MAC addresses to be removed and is sent to all other PEs over the corresponding LDP sessions.

In order to minimize the impact on LDP convergence time and scalability when a MAC List TLV contains a large number of MAC addresses, many implementations use an LDP address withdraw message with an empty MAC list. When a PE-rs switch in the full mesh of H-VPLS receives this message, it also flushes MAC addresses that are not affected due to the topology change, thus leading to unnecessary flooding and relearning. Throughout this document, the term "MAC flush message" is used to specify an LDP address withdraw message with an empty MAC list as described in [RFC4762]. The solutions described in this document are applicable only to LDP address withdraw messages with empty MAC lists.

In a VPLS topology, the core PWs remain active and learning happens on the PE-rs nodes. However, when the VPLS topology changes, the PE-rs must relearn using MAC address withdrawal or flushing. As per the MAC address withdrawal 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 PW associated with this signaling session over which the message was received. Throughout this document, we use the terminology "positive" MAC flushing or "flush-all-but-mine" for this type of MAC flush message and its actions.

This document introduces an optimized "negative" MAC flush message, described in Section 3.2, that can be configured to improve the response to topology changes in a number of Ethernet topologies where the Service Level Agreement (SLA) is dependent on minimal disruption and fast restoration of affected traffic. This new message is used in the case of Provider Backbone Bridging (PBB) topologies to restrict the flushing to a set of service instances (I-SIDs). It is also important to note that the MAC flush message described in [RFC4762], which is called "a positive MAC flush message" in this document, MUST always be handled for Backbone MACs (B-MACs) in cases where the core nodes change or fail. In dual-homed or multi-homed edge topologies, the procedures in this document augment [RFC4762] messages and provide less disruption for those cases.

3.1. MAC Flushing on Activation of Backup Spoke PW

This section describes scenarios where MAC flush withdrawal is initiated on activation of a backup PW in H-VPLS.

3.1.1. MAC Flushing Initiated by PE-rs

[RFC4762] specifies that on failure of the primary PW it is PE3-rs (Figure 1) that initiates MAC flushing towards the core. However, note that PE3-rs can initiate MAC flushing only when PE3-rs is dual-homing "aware" -- that is, there is some redundancy management protocol running between the MTU-s and its host PE-rs devices. The scope of this document is applicable to several dual-homing or multi-homing protocols. This document illustrates that multi-homing can be improved with negative MAC flushing. One example is BGP-based multi-homing in LDP-based VPLS, which uses the procedures defined in [VPLS-MH]. In this method of dual-homing, PE3-rs would neither forward any traffic to the MTU-s nor receive any traffic from the MTU-s while PE1-rs is acting as a primary (or designated forwarder).

3.1.2. MAC Flushing Initiated by MTU-s

When dual-homing is achieved by manual configuration in the MTU-s, the hosting PE-rs devices are dual-homing "agnostic", and PE3-rs cannot initiate MAC flush messages. PE3-rs can send or receive traffic over the backup PW, since the dual-homing control is with the MTU-s only. When the backup PW is made active by the MTU-s, the MTU-s triggers a MAC flush message. The message is sent over the LDP session associated with the newly activated PW. On receiving the MAC flush message from the MTU-s, PE3-rs (the PE-rs device with a now-active PW) would flush all the MAC addresses it has learned, except the ones learned over the newly activated spoke PW. PE3-rs further initiates a MAC flush message to all other PE devices in the core. Note that a forced switchover to the backup PW can also be invoked by the MTU-s due to maintenance or administrative activities on the former primary spoke PW.

The method of MAC flushing initiated by the MTU-s is modeled after Topology Change Notification (TCN) in the Rapid Spanning Tree Protocol (RSTP) [IEEE.802.1Q-2011]. When a bridge switches from a failed link to the backup link, the bridge sends out a TCN message over the newly activated link. Upon receiving this message, the upstream bridge flushes its entire list of MAC addresses, except the ones received over this link. The upstream bridge then 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.

The MAC flushing information is propagated in the control plane. The control-plane message propagation is associated with the data path and hence follows propagation rules similar to those used for forwarding in the LDP data plane. For example, PE-rs nodes follow the data-plane "split-horizon" forwarding rules in H-VPLS (refer to Section 4.4 of [RFC4762]). Therefore, a MAC flush message is propagated in the context of mesh PW(s) when it is received in the context of a spoke PW. When a PE-rs node receives a MAC flush message in the context of a mesh PW, then it is not propagated to other mesh PWs.

3.2. MAC Flushing on Failure

MAC flushing on failure, or "negative" MAC flushing, is introduced in this document. Negative MAC flushing is an improvement on the current practice of sending a MAC flush message with an empty MAC list as described in Section 3.1.1. We use the term "negative" MAC flushing or "flush-all-from-me" for this kind of flushing action as opposed to the "positive" MAC flush action in [RFC4762]. In negative MAC flushing, the MAC flushing is initiated by PE1-rs (Figure 1) on detection of failure of the primary spoke PW. The MAC flush message is sent to all participating PE-rs devices in the VPLS full mesh. PE1-rs should initiate MAC flushing only if PE1-rs is dual-homing aware. (If PE1-rs is dual-homing agnostic, the policy is to not initiate MAC flushing on failure, since that could cause unnecessary flushing in the case of a single-homed MTU-s.) The specific dual-homing protocols for this scenario are outside the scope of this document, but the operator can choose to use the optimized MAC flushing described in this document or the [RFC4762] procedures.

The procedure for negative MAC flushing is beneficial and results in less disruption than the [RFC4762] procedures, including when the MTU-s is dual-homed with a variety of Ethernet technologies, not just LDP. The negative MAC flush message is a more targeted MAC flush, and the other PE-rs nodes will flush only the specified MACs. This targeted MAC flush cannot be achieved with the MAC address withdraw message defined in [RFC4762]. Negative MAC flushing typically results in a smaller set of MACs to be flushed and results in less disruption for these topologies.

Note that in the case of negative MAC flushing the list SHOULD be only the MACs for the affected MTU-s. If the list is empty, then the negative MAC flush procedures will result in flushing and relearning all attached MTU-s devices for the originating PE-rs.

3.3. MAC Flushing in PBB-VPLS

[RFC7041] describes how PBB can be integrated with VPLS to allow for useful PBB capabilities while continuing to avoid the use of MSTP in the backbone. The combined solution, referred to as "PBB-VPLS", results in better scalability in terms of the number of service instances, PWs, and C-MACs that need to be handled in the VPLS PE-rs devices. This document describes extensions to LDP MAC flushing procedures described in [RFC4762] that are required to build desirable capabilities for the PBB-VPLS solution.

The solution proposed in this document is generic and is applicable when Multi-Segment Pseudowires (MS-PWs) [RFC6073] are used in interconnecting PE devices in H-VPLS. There could be other H-VPLS models not defined in this document where the solution may be applicable.

4. Problem Description

This section describes the problems in detail with respect to various MAC flushing actions described in Section 3.

4.1. MAC Flushing Optimization in VPLS Resiliency

This section describes the optimizations required in MAC flushing procedures when H-VPLS resiliency is provided by primary and backup spoke PWs.

4.1.1. MAC Flushing Optimization for Regular H-VPLS

Figure 2 shows a dual-homed H-VPLS scenario for a VPLS instance, where the problem with the existing MAC flushing method is as explained in Section 3.

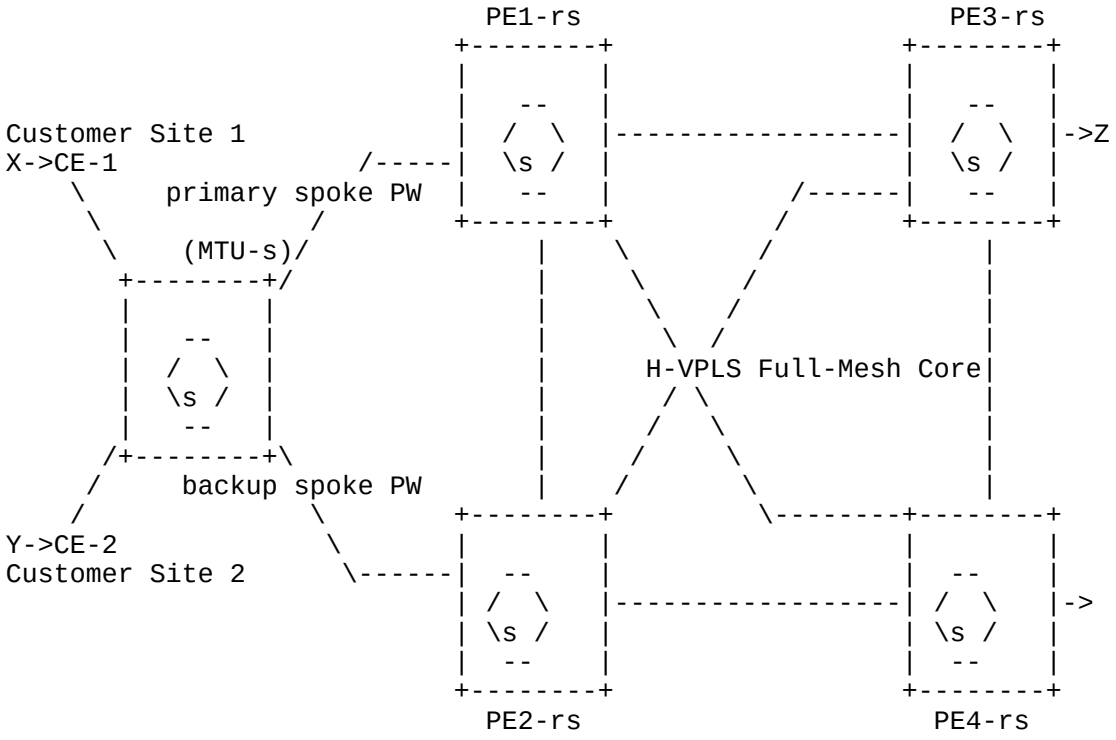


Figure 2: Dual-Homed MTU-s in Two-Tier Hierarchy H-VPLS

In Figure 2, the MTU-s is dual-homed to PE1-rs and PE2-rs. Only the primary spoke PW is active at the MTU-s; thus, PE1-rs is acting as the active device (designated forwarder) to reach the full mesh in the VPLS instance. The MAC addresses of nodes located at access sites (behind CE-1 and CE-2) are learned at PE1-rs over the primary spoke PW. Let’s say X represents a set of such MAC addresses located behind CE-1. MAC Z represents one of a possible set of other destination MACs. As packets flow from X to other MACs in the VPLS network, PE2-rs, PE3-rs, and PE4-rs learn about X on their respective mesh PWs terminating at PE1-rs. When the MTU-s switches to the backup spoke PW and activates it, PE2-rs becomes the active device (designated forwarder) to reach the full-mesh core for the MTU-s. Traffic entering the H-VPLS from CE-1 and CE-2 is diverted by the MTU-s to the spoke PW to PE2-rs. Traffic destined from PE2-rs,

PE3-rs, and PE4-rs to X will be black-holed until the MAC address aging timer expires (the default is 5 minutes) or a packet flows from X to other addresses through PE2-rs.

For example, if a packet flows from MAC Z to MAC X after the backup spoke PW is active, packets from MAC Z travel from PE3-rs to PE1-rs and are dropped. However, if a packet with MAC X as source and MAC Z as destination arrives at PE2-rs, PE2-rs will now learn that MAC X is on the backup spoke PW and will forward to MAC Z. At this point, traffic from PE3-rs to MAC X will go to PE2-rs, since PE3-rs has also learned about MAC X. Therefore, a mechanism is required to make this learning more timely in cases where traffic is not bidirectional.

To avoid traffic black-holing, the MAC addresses that have been learned in the upstream VPLS full mesh through PE1-rs must be relearned or removed from the MAC Forwarding Information Bases (FIBs) in the VSIs at PE2-rs, PE3-rs, and PE4-rs. If PE1-rs and PE2-rs are dual-homing agnostic, then on activation of the standby PW from the MTU-s, a MAC flush message will be sent by the MTU-s to PE2-rs that will flush all the MAC addresses learned in the VPLS instance at PE2-rs from all other PWs except the PW connected to the MTU-s.

PE2-rs further relays the MAC flush messages to all other PE-rs devices in the full mesh. The same processing rule applies for all those PE-rs devices: all the MAC addresses are flushed except the ones learned on the PW connected to PE2-rs. For example, at PE3-rs all of the MAC addresses learned from the PWs connected to PE1-rs and PE4-rs are flushed and relearned subsequently. Before the relearning happens, flooding of unknown destination MAC addresses takes place throughout the network. As the number of PE-rs 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 a large number of VPLS instances provisioned in the H-VPLS network topology, the amount of unnecessary flooding and relearning increases. An optimization, described below, is required that will flush only the MAC addresses learned from the respective PWs between PE1-rs and other PE devices in the full mesh, to minimize the relearning and flooding in the network. In the example above, only the MAC addresses in sets X and Y (shown in Figure 2) need to be flushed across the core.

The same case is applicable when PE1-rs and PE2-rs are dual-homing aware and participate in a designated forwarder election. When PE2-rs becomes the active device for the MTU-s, then PE2-rs MAY initiate MAC flushing towards the core. The receiving action of the MAC flush message in other PE-rs devices is the same as in MAC flushing initiated by the MTU-s. This is the behavior specified in [RFC4762].

4.1.2. MAC Flushing Optimization for Native Ethernet Access

The analysis in Section 4.1.1 applies also to the native Ethernet access into a VPLS. In such a scenario, one active endpoint and one or more standby endpoints terminate into two or more VPLS or H-VPLS PE-rs devices. Examples of this dual-homed access are ITU-T [ITU.G8032] access rings or any proprietary multi-chassis Link Aggregation Group (LAG) emulations. Upon failure of the active native Ethernet endpoint on PE1-rs, an optimized MAC flush message is required to be initiated by PE1-rs to ensure that on PE2-rs, PE3-rs, and PE4-rs only the MAC addresses learned from the respective PWs connected to PE1-rs are being flushed.

4.2. Black-Holing Issue in PBB-VPLS

In a PBB-VPLS deployment, a B-component VPLS (B-VPLS) may be used as infrastructure to support one or more I-component instances. The B-VPLS control plane (LDP Signaling) and learning of Backbone MACs (B-MACs) replace the I-component control plane and learning of Customer MACs (C-MACs) throughout the MPLS core. This raises an additional challenge related to black-hole avoidance in the I-component domain as described in this section. Figure 3 describes the case of a Customer Edge (CE) device (node A) dual-homed to two I-component instances located on two PBB-VPLS PEs (PE1-rs and PE2-rs).

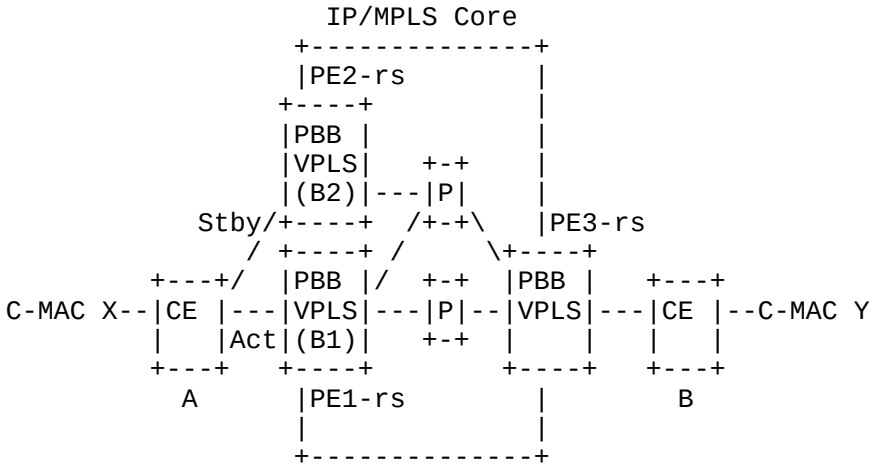


Figure 3: PBB Black-Holing Issue - CE Dual-Homing Use Case

The link between PE1-rs and CE-A is active (marked with A), while the link between CE-A and PE2-rs is in standby/blocked status. In the network diagram, C-MAC X is one of the MAC addresses located behind

CE-A in the customer domain, C-MAC Y is behind CE-B, and the B-VPLS instances on PE1-rs are associated with B-MAC B1 and PE2-rs with B-MAC B2.

As the packets flow from C-MAC X to C-MAC Y through PE1-rs with B-MAC B1, the remote PE-rs devices participating in the B-VPLS with the same I-SID (for example, PE3-rs) will learn the C-MAC X associated with B-MAC B1 on PE1-rs. Under a failure condition of the link between CE-A and PE1-rs and on activation of the link to PE2-rs, the remote PE-rs devices (for example, PE3-rs) will forward the traffic destined for C-MAC X to B-MAC B1, resulting in PE1-rs black-holing that traffic until the aging timer expires or a packet flows from X to Y through PE2-rs (B-MAC B2). This may take a long time (the default aging timer is 5 minutes) and may affect a large number of flows across multiple I-components.

A possible solution to this issue is to use the existing LDP MAC flushing method as specified in [RFC4762] to flush the B-MAC associated with the PE-rs in the B-VPLS domain where the failure occurred. This will automatically flush the C-MAC-to-B-MAC association in the remote PE-rs devices. This solution has the disadvantage of producing a lot of unnecessary MAC flushing in the B-VPLS domain as there was no failure or topology change affecting the Backbone domain.

A better solution -- one that would propagate the I-component events through the backbone infrastructure (B-VPLS) -- is required in order to flush only the C-MAC-to-B-MAC associations in the remote PBB-VPLS-capable PE-rs devices. Since there are no I-component control-plane exchanges across the PBB backbone, extensions to the B-VPLS control plane are required to propagate the I-component MAC flushing events across the B-VPLS.

5. Solution Description

This section describes the solution for the problem space described in Section 4.

5.1. MAC Flushing Optimization for VPLS Resiliency

The basic principle of the optimized MAC flush mechanism is explained with reference to Figure 2. The optimization is achieved by initiating MAC flushing on failure as described in Section 3.2.

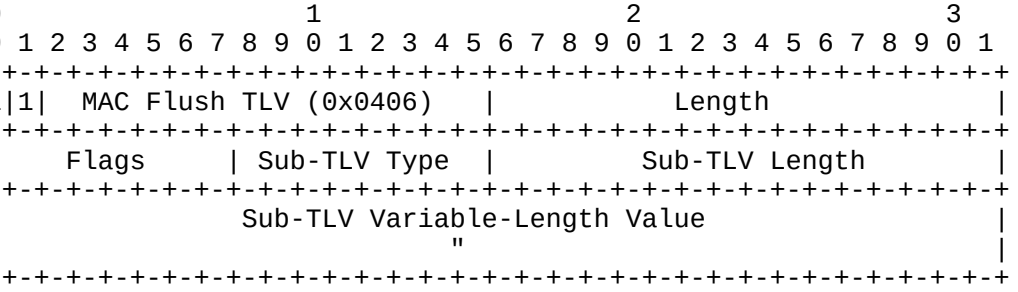
PE1-rs would initiate MAC flushing towards the core on detection of failure of the primary spoke PW between the MTU-s and PE1-rs (or status change from active to standby [RFC6718]). This method is referred to as "MAC flushing on failure" throughout this document.

The MAC flush message would indicate to receiving PE-rs devices to flush all MACs learned over the PW in the context of the VPLS for which the MAC flush message is received. Each PE-rs device in the full mesh that receives the message identifies the VPLS instance and its respective PW that terminates in PE1-rs from the FEC TLV received in the message and/or LDP session. Thus, the PE-rs device flushes only the MAC addresses learned from that PW connected to PE1-rs, minimizing the required relearning and the flooding throughout the VPLS domain.

This section defines a generic MAC Flush Parameters TLV for LDP [RFC5036]. Throughout this document, the MAC Flush Parameters TLV is also referred to as the "MAC Flush TLV". A MAC Flush TLV carries information on the desired action at the PE-rs device receiving the message and is used for optimized MAC flushing in VPLS. The MAC Flush TLV can also be used for the [RFC4762] style of MAC flushing as explained in Section 3.

5.1.1. MAC Flush Parameters TLV

The MAC Flush Parameters TLV is described below:



The U-bit and F-bit [RFC5036] are set to forward if unknown so that potential intermediate VPLS PE-rs devices unaware of the new TLV can just propagate it transparently. In the case of a B-VPLS network that has PBB-VPLS in the core with no I-components attached, this message can still be useful to edge B-VPLS devices that do have the I-components with the I-SIDs and understand the message. The MAC Flush Parameters TLV type is 0x0406, as assigned by IANA. The encoding of the TLV follows the standard LDP TLV encoding described in [RFC5036].

The TLV value field contains a 1-byte Flag field used as described below. Further, the TLV value MAY carry one or more sub-TLVs. Any sub-TLV definition for the above TLV MUST address the actions in combination with other existing sub-TLVs.

The detailed format for the Flags bit vector is described below:

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|C|N|    MBZ    | (MBZ = MUST Be Zero)
+-+-+-+-+-+-+-+-+
```

The 1-byte Flag field is mandatory. The following flags are defined:

C-flag: Used to indicate the context of the PBB-VPLS component in which MAC flushing is required. For PBB-VPLS, there are two contexts of MAC flushing -- the Backbone VPLS (B-component VPLS) and the Customer VPLS (I-component VPLS). The C-flag MUST be ZERO (C = 0) when a MAC flush action for the B-VPLS is required and MUST be set (C = 1) when the MAC flush action for the I-component is required. In the regular H-VPLS case, the C-flag MUST be ZERO (C = 0) to indicate that the flush applies to the current VPLS context.

N-flag: Used to indicate whether a positive (N = 0, flush-all-but-mine) or negative (N = 1, flush-all-from-me) MAC flush action is required. The source (mine/me) is defined as the PW associated with either the LDP session on which the LDP MAC withdraw was received or the B-MAC(s) listed in the B-MAC Sub-TLV. For the optimized MAC flush procedure described in this section, the flag MUST be set (N = 1).

Detailed usage in the context of PBB-VPLS is explained in Section 5.2.

MBZ flags: The rest of the flags SHOULD be set to zero on transmission and ignored on reception.

The MAC Flush TLV SHOULD be placed after the existing TLVs in the [RFC4762] MAC flush message.

5.1.2. Application of the MAC Flush TLV in Optimized MAC Flushing

When optimized MAC flushing is supported, the MAC Flush TLV MUST be sent in an existing LDP address withdraw message with an empty MAC list but from the core PE-rs on detection of failure of its local/primary spoke PW. The N-bit in the TLV MUST be set to 1 to indicate flush-all-from-me. If the optimized MAC flush procedure is used in a Backbone VPLS or regular VPLS/H-VPLS context, the C-bit MUST be ZERO (C = 0). If it is used in an I-component context, the C-bit MUST be set (C = 1). See Section 5.2 for details of its usage in the context of PBB-VPLS.

Note that the assumption is that the MAC Flush TLV is understood by all devices before it is turned on in any network. See Section 6 ("Operational Considerations").

When optimized MAC flushing is not supported, the MAC withdraw procedures defined in [RFC4762], where either the MTU-s or PE2-rs sends the MAC withdraw message, SHOULD be used. This includes the case where the network is being changed to support optimized MAC flushing but not all devices are capable of understanding optimized MAC flush messages.

In the case of B-VPLS devices, the optimized MAC flush message SHOULD be supported.

5.1.3. MAC Flush TLV Processing Rules for Regular VPLS

This section describes the processing rules of the MAC Flush TLV that MUST be followed in the context of optimized MAC flush procedures in VPLS.

When optimized MAC flushing is supported, a multi-homing PE-rs initiates a MAC flush message towards the other related VPLS PE-rs devices when it detects a transition (failure or a change to standby) in its active spoke PW. In such a case the MAC Flush TLV MUST be sent with N = 1. A PE-rs device receiving the MAC Flush TLV SHOULD follow the same processing rules as those described in this section.

Note that if a Multi-Segment Pseudowire (MS-PW) is used in VPLS, then a MAC flush message is processed only at the PW Terminating Provider Edge (T-PE) nodes, since PW Switching Provider Edge S-PE(s) traversed by the MS-PW propagates the MAC flush messages without any action. In this section, a PE-rs device signifies only a T-PE in the MS-PW case.

When a PE-rs device receives a MAC Flush TLV with N = 1, it SHOULD flush all the MAC addresses learned from the PW in the VPLS in the context on which the MAC flush message is received. It is assumed that when these procedures are used all nodes support the MAC flush message. See Section 6 ("Operational Considerations") for details.

When optimized MAC flushing is not supported, a MAC Flush TLV is received with N = 0 in the MAC flush message; in such a case, the receiving PE-rs SHOULD flush the MAC addresses learned from all PWs in the VPLS instance, except the ones learned over the PW on which the message is received.

Regardless of whether optimized MAC flushing is supported, if a PE-rs device receives a MAC flush message with a MAC Flush TLV option (N = 0 or N = 1) and a valid MAC address list, it SHOULD ignore the option and deal with MAC addresses explicitly as per [RFC4762].

5.1.4. Optimized MAC Flush Procedures

This section expands on the optimized MAC flush procedure in the scenario shown in Figure 2.

When optimized MAC flushing is being used, a PE-rs that is dual-homing aware SHOULD send MAC address messages with a MAC Flush TLV and N = 1, provided the other PEs understand the new messages. Upon receipt of the MAC flush message, PE2-rs identifies the VPLS instance that requires MAC flushing from the FEC element in the FEC TLV. On receiving N = 1, PE2-rs removes all MAC addresses learned from that PW over which the message is received. The same action is performed by PE3-rs and PE4-rs.

Figure 4 shows another redundant H-VPLS topology to protect against failure of the MTU-s device. In this case, since there is more than a single MTU-S, a protocol such as provider RSTP [IEEE.802.1Q-2011] may be used as the selection algorithm for active and backup PWs in order to maintain the connectivity between MTU-s devices and PE-rs devices at the edge. It is assumed that PE-rs devices can detect failure on PWs in either direction through OAM mechanisms (for instance, Virtual Circuit Connectivity Verification (VCCV) procedures).

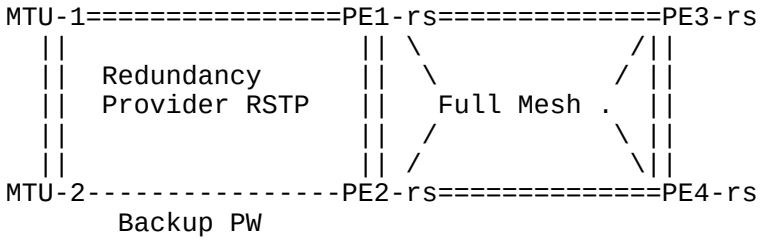


Figure 4: Redundancy with Provider RSTP

MTU-1, MTU-2, PE1-rs, and PE2-rs participate in provider RSTP. Configuration using RSTP ensures that the PW between MTU-1 and PE1-rs is active and the PW between MTU-2 and PE2-rs is blocked (made backup) at the MTU-2 end. When the active PW failure is detected by RSTP, it activates the PW between MTU-2 and PE2-rs. When PE1-rs detects the failing PW to MTU-1, it MAY trigger MAC flushing into the full mesh with a MAC Flush TLV that carries N = 1. Other PE-rs

devices in the full mesh that receive the MAC flush message identify their respective PWs terminating on PE1-rs and flush all the MAC addresses learned from it.

[RFC4762] describes a 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-rs devices. To provide redundancy against failure of the inter-domain spoke, full mesh of inter-domain spokes can be set up between border PE-rs devices, and provider RSTP may be used for selection of the active inter-domain spoke. In the case of inter-domain spoke PW failure, MAC withdrawal initiated by PE-rs MAY be used for optimized MAC flush procedures within individual domains.

Further, the procedures are applicable to any native Ethernet access topologies multi-homed to two or more VPLS PE-rs devices. The text in this section applies for the native Ethernet case where active/standby PWs are replaced with the active/standby Ethernet endpoints. An optimized MAC flush message can be generated by the VPLS PE-rs that detects the failure in the primary Ethernet access.

5.2. LDP MAC Flush Extensions for PBB-VPLS

The use of an address withdraw message with a MAC List TLV is proposed in [RFC4762] as a way to expedite removal of MAC addresses as the result of a topology change (e.g., failure of a primary link of a VPLS PE-rs device and, implicitly, the activation of an alternate link in a dual-homing use case). These existing procedures apply individually to B-VPLS and I-component domains.

When it comes to reflecting topology changes in access networks connected to I-components across the B-VPLS domain, certain additions should be considered, as described below.

MAC switching in PBB is based on the mapping of Customer MACs (C-MACs) to one or more Backbone MACs (B-MACs). A topology change in the access (I-component domain) should just invoke the flushing of C-MAC entries in the PBB PEs' FIB(s) associated with the I-component(s) impacted by the failure. There is a need to indicate the PBB PE (B-MAC source) that originated the MAC flush message to selectively flush only the MACs that are affected.

These goals can be achieved by including the MAC Flush Parameters TLV in the LDP address withdraw message to indicate the particular domain(s) requiring MAC flushing. On the other end, the receiving PEs SHOULD use the information from the new TLV to flush only the related FIB entry/entries in the I-component instance(s).

At least one of the following sub-TLVs MUST be included in the MAC Flush Parameters TLV if the C-flag is set to 1:

- PBB B-MAC List Sub-TLV:

  Type: 0x0407

  Length: Value length in octets. At least one B-MAC address MUST be present in the list.

  Value: One or a list of 48-bit B-MAC addresses. These are the source B-MAC addresses associated with the B-VPLS instance that originated the MAC withdraw message. It will be used to identify the C-MAC(s) mapped to the B-MAC(s) listed in the sub-TLV.

- PBB I-SID List Sub-TLV:

  Type: 0x0408

  Length: Value length in octets. Zero indicates an empty I-SID list. An empty I-SID list means that the flushing applies to all the I-SIDs mapped to the B-VPLS indicated by the FEC TLV.

  Value: One or a list of 24-bit I-SIDs that represent the I-component FIB(s) where the MAC flushing needs to take place.

5.2.1. MAC Flush TLV Processing Rules for PBB-VPLS

The following steps describe the details of the processing rules for the MAC Flush TLV in the context of PBB-VPLS. In general, these procedures are similar to the VPLS case but are tailored to PBB, which may have a large number of MAC addresses. In PBB, there are two sets of MAC addresses: Backbone (outer) MACs (B-MACs) and Customer (inner) MACs (C-MACs). C-MACs are associated to remote B-MACs by learning. There are also I-SIDs in PBB; I-SIDs are similar to VLANs for the purposes of the discussion in this section. In order to achieve behavior that is similar to the Regular VPLS case, there are some differences in the interpretation of the optimized MAC flush message.

1. Positive flush of C-MACs. This is equivalent to [RFC4762] MAC flushing in a PBB context. In this case, the N-bit is set to 0; the C-bit is set to 1, and C-MACs are to be flushed. However, since C-MACs are related to B-MACs in an I-SID context, further refinement of the flushing scope is possible.

- If an I-SID needs to be flushed (all C-MACs within that I-SID), then I-SIDs are listed in the appropriate TLV. If all I-SIDs are to have the C-MACs flushed, then the I-SID TLV can be empty. It is typical to flush a single I-SID only, since each I-SID is associated with one or more interfaces (typically one, in the case of dual-homing). In the PBB case, flushing the I-SID is equivalent to the empty MAC list discussed in [RFC4762].

- If only a set of B-MAC-to-C-MAC associations needs to be flushed, then a B-MAC list can be included to further refine the list. This can be the case if an I-SID component has more than one interface and a B-MAC is used to refine the granularity. Since this is a positive MAC flush message, the intended behavior is to flush all C-MACs except those that are associated with B-MACs in the list.

  Positive flush of B-MACs is also useful for propagating flush from other protocols such as RSTP.

2. Negative flush of C-MACs. This is equivalent to optimized MAC flushing. In this case, the N-bit is set to 1; the C-bit is set to 1, and a list of B-MACs is provided so that the respective C-MACs can be flushed.

   - The I-SID list SHOULD be specified. If it is absent, then all I-SIDs require that the C-MACs be flushed.

   - A set of B-MACs SHOULD be listed, since B-MAC-to-C-MAC associations need to be flushed and listing B-MACs scopes the flush to just those B-MACs. Again, this is typical usage, because a PBB VPLS I-component interface will have one associated I-SID and typically one (but possibly more than one) B-MAC, each with multiple remotely learned C-MACs. The B-MAC list is included to further refine the list for the remote receiver. Since this is a negative MAC flush message, the intended behavior is to flush all remote C-MACs that are associated with any B-MACs in the list (in other words, from the affected interface).

The processing rules on reception of the MAC flush message are:

- On Backbone Core Bridges (BCBs), if the C-bit is set to 1, then the PBB-VPLS SHOULD NOT flush their B-MAC FIBs. The B-VPLS control plane SHOULD propagate the MAC flush message following the data-plane split-horizon rules to the established B-VPLS topology.

- On Backbone Edge Bridges (BEBs), the following actions apply:

  - The PBB I-SID list is used to determine the particular I-SID FIBs (I-component) that need to be considered for flushing action. If the PBB I-SID List Sub-TLV is not included in a received message, then all the I-SID FIBs associated with the receiving B-VPLS SHOULD be considered for flushing action.

  - The PBB B-MAC list is used to identify from the I-SID FIBs in the previous step to selectively flush B-MAC-to-C-MAC associations, depending on the N-flag specified below. If the PBB B-MAC List Sub-TLV is not included in a received message, then all B-MAC-to-C-MAC associations in all I-SID FIBs (I-component) as specified by the I-SID List are considered for required flushing action, again depending on the N-flag specified below.

  - Next, depending on the N-flag value, the following actions apply:

    - N = 0: all the C-MACs in the selected I-SID FIBs SHOULD be flushed, with the exception of the resultant C-MAC list from the B-MAC list mentioned in the message ("flush all but the C-MACs associated with the B-MAC(s) in the B-MAC List Sub-TLV from the FIBs associated with the I-SID list").

    - N = 1: all the resultant C-MACs SHOULD be flushed ("flush all the C-MACs associated with the B-MAC(s) in the B-MAC List Sub-TLV from the FIBs associated with the I-SID list").

5.2.2. Applicability of the MAC Flush Parameters TLV

If the MAC Flush Parameters TLV is received by a Backbone Edge Bridge (BEB) in a PBB-VPLS that does not understand the TLV, then an undesirable MAC flushing action may result. It is RECOMMENDED that all PE-rs devices participating in PBB-VPLS support the MAC Flush Parameters TLV. If this is not possible, the MAC Flush Parameters TLV SHOULD be disabled, as mentioned in Section 6 ("Operational Considerations").

"Mac Flush TLV" and its formal name -- "MAC Flush Parameters TLV" -- are synonymous. The MAC Flush TLV is applicable to the regular VPLS context as well, as explained in Section 3.1.1. To achieve negative MAC flushing (flush-all-from-me) in a regular VPLS context, the MAC Flush Parameters TLV SHOULD be encoded with C = 0 and N = 1 without

inclusion of any Sub-TLVs. A negative MAC flush message is highly desirable in scenarios where VPLS access redundancy is provided by Ethernet ring protection as specified in the ITU-T G.8032 [ITU.G8032] specification.

6. Operational Considerations

As mentioned earlier, if the MAC Flush Parameters TLV is not understood by a receiver, then an undesirable MAC flushing action would result. To avoid this, one possible solution is to develop an LDP-based capability negotiation mechanism to negotiate support of various MAC flushing capabilities between PE-rs devices in a VPLS instance. A negotiation mechanism was discussed previously and was considered outside the scope of this document. Negotiation is not required to deploy this optimized MAC flushing as described in this document.

VPLS may be used with or without the optimization. If an operator wants the optimization for VPLS, it is the operator’s responsibility to make sure that the VPLS devices that are capable of supporting the optimization are properly configured. From an operational standpoint, it is RECOMMENDED that implementations of the solution provide administrative control to select the desired MAC flushing action towards a PE-rs device in the VPLS. Thus, in the topology described in Figure 2, an implementation could support PE1-rs, sending optimized MAC flush messages towards the PE-rs devices that support the solution and the PE2-rs device initiating the [RFC4762] style of MAC flush messages towards the PE-rs devices that do not support the optimized solution during upgrades. The PE-rs that supports the MAC Flush Parameters TLV MUST support the RFC 4762 MAC flushing procedures, since this document only augments them.

In the case of PBB-VPLS, this operation is the only method supported for specifying I-SIDs, and the optimization is assumed to be supported or should be turned off, reverting to flushing using [RFC4762] at the Backbone MAC level.

7. IANA Considerations

7.1. New LDP TLV

IANA maintains a registry called "Label Distribution Protocol (LDP) Parameters" with a sub-registry called "TLV Type Name Space".

IANA has allocated three new code points as follows:

Value	Description	Reference	Notes
0x0406	MAC Flush Parameters TLV	[RFC7361]	
0x0407	PBB B-MAC List Sub-TLV	[RFC7361]	
0x0408	PBB I-SID List Sub-TLV	[RFC7361]	

7.2. New Registry for MAC Flush Flags

IANA has created a new sub-registry under "Label Distribution Protocol (LDP) Parameters" called "MAC Flush Flags".

IANA has populated the registry as follows:

Bit Number	Hex	Abbreviation	Description	Reference
0	0x80	C	Context	[RFC7361]
1	0x40	N	Negative MAC flushing	[RFC7361]
2-7			Unassigned	

Other new bits are to be assigned by Standards Action [RFC5226].

8. Security Considerations

Control-plane aspects:

LDP security (authentication) methods as described in [RFC5036] are applicable here. Further, this document implements security considerations as discussed in [RFC4447] and [RFC4762]. The extensions defined here optimize the MAC flushing action, and so the risk of security attacks is reduced. However, in the event that the configuration of support for the new TLV can be spoofed, sub-optimal behavior will be seen.

Data-plane aspects:

This specification does not have any impact on the VPLS forwarding plane but can improve MAC flushing behavior.

9. Contributing Author

The authors would like to thank Marc Lasserre, who made a major contribution to the development of this document.

Marc Lasserre
Alcatel-Lucent
EMail: marc.lasserre@alcatel-lucent.com

10. Acknowledgements

The authors would like to thank the following people who have provided valuable comments, feedback, and review on the topics discussed in this document: Dimitri Papadimitriou, Jorge Rabadan, Prashanth Ishwar, Vipin Jain, John Rigby, Ali Sajassi, Wim Henderickx, Paul Kwok, Maarten Vissers, Daniel Cohn, Nabil Bitar, Giles Heron, Adrian Farrel, Ben Niven-Jenkins, Robert Sparks, Susan Hares, and Stephen Farrell.

11. References

11.1. 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, April 2006.

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

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

11.2. Informative References

[IEEE.802.1Q-2011] IEEE, "IEEE Standard for Local and metropolitan area networks -- Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks", IEEE Std 802.1Q, 2011.

[ITU.G8032] International Telecommunication Union, "Ethernet ring protection switching", ITU-T Recommendation G.8032, February 2012.

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

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

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

[RFC6718] Muley, P., Aissaoui, M., and M. Bocci, "Pseudowire Redundancy", RFC 6718, August 2012.

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

[VPLS-MH] Kothari, B., Kompella, K., Henderickx, W., Balus, F., Uttaro, J., Palislamovic, S., and W. Lin, "BGP based Multi-homing in Virtual Private LAN Service", Work in Progress, July 2014.

Authors' Addresses

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

EMail: pranjal.dutta@alcatel-lucent.com

Florin Balus
Alcatel-Lucent
701 E Middlefield Road
Mountain View, CA 94043
USA

EMail: florin.balus@alcatel-lucent.com

Olen Stokes
Extreme Networks
2121 RDU Center Drive
Suite 300
Morrisville, NC 27650
USA

EMail: ostokes@extremenetworks.com

Geraldine Calvignac
Orange
2, avenue Pierre-Marzin
Lannion Cedex, 22307
France

EMail: geraldine.calvignac@orange.com

Don Fedyk
Hewlett-Packard Company
USA

EMail: don.fedyk@hp.com