Provider Provisioned VPN Working Group
Internet Draft
Expiration Date: January 2002

Vach Kompella
Sunil Khandekar
Nick Tingle
TiMetra Networks

Giles Heron
PacketExchange

Juha Heinanen
Song Networks

Tom S. C. Soon
SBC Communications

Rick Wilder
Masergy

Luca Martini
Level3 Communications

Virtual Private Switched Network Services over an MPLS Network

draft-vkompella-ppvpn-vpsn-mpls-00.txt

Status of this Memo

Abstract

This document describes how to establish a Virtual Private Switched Network Service for Ethernet over an MPLS network.

Table of Contents

Status of this Memo....1
Abstract....1
Table of Contents....2
1. Placement of this Memo in Sub-IP Area....2
2. Introduction....3
3. Virtual Private Switched Network....3
4. Tunneled VPSN....3
4.1. MAC Address Learning....4
4.1.1. MAC Address Learning with Ethernet Encapsulation....5
4.1.2. MAC Address Learning with Ethernet VLAN Encapsulation....5
5. MAC Address Management....5
5.1. Aging MAC Addresses....5
5.2. MAC Address Signaling....6
5.2.1. MAC TLV....6
5.2.2. Address Message Containing MAC TLV....8
5.2.3. Address Withdraw Message Containing MAC TLV....8
5.2.4. LDP Session Failure or Termination....9
5.3. Discussion on MAC TLVs....9
5.3.1. Physical Network Reorganization....9
5.3.2. Spanning Tree Action....9
6. Signaling a VPSN....10
6.1. Adjacency and Session Management....10
7. Security Considerations....11
8. Intellectual Property Disclaimer....11
9. References....11
10. Authors' Addresses....11

1. Placement of this Memo in Sub-IP Area

RELATED DOCUMENTS

draft-heron-ppvpn-vpsn-reqmts-00.txt

WHERE DOES THIS FIT IN THE PICTURE OF THE SUB-IP WORK

This fits in the PPVPN box.

WHY IS IT TARGETTED AT THIS WG

This work fits in the PPVPN working group charter. It describes a service that uses an emulation of a Layer 2 medium to create a provider provisioned virtual private network [1], specifically, a Transparent LAN service.

JUSTIFICATION

We believe the WG should consider this draft because it specifies

signaling for a class of layer 2 VPN that has up to now not been sufficiently addressed in this WG.

2. Introduction

This document describes how to signal a Virtual Private Switched Network (VPSN), which, in the context of Ethernet, is very similar to a Transparent LAN Service (TLS). The VPSN operates over a packet switched network that has tunneling capabilities. The requirements for a VPSN are given in [2].

A VPSN provides the ability to mimic the behavior of an Ethernet LAN. Several proposals exist to create a tunneled Ethernet service [3][4], but they address point-to-point tunneling of Ethernet frames. Simply using point-to-point tunneling of Ethernet frames, it is possible to create a multi-point service, but it is extremely inefficient. By adding the ability to learn MAC addresses and associate them with tunnel endpoints, it is possible to create a more efficient VPSN.

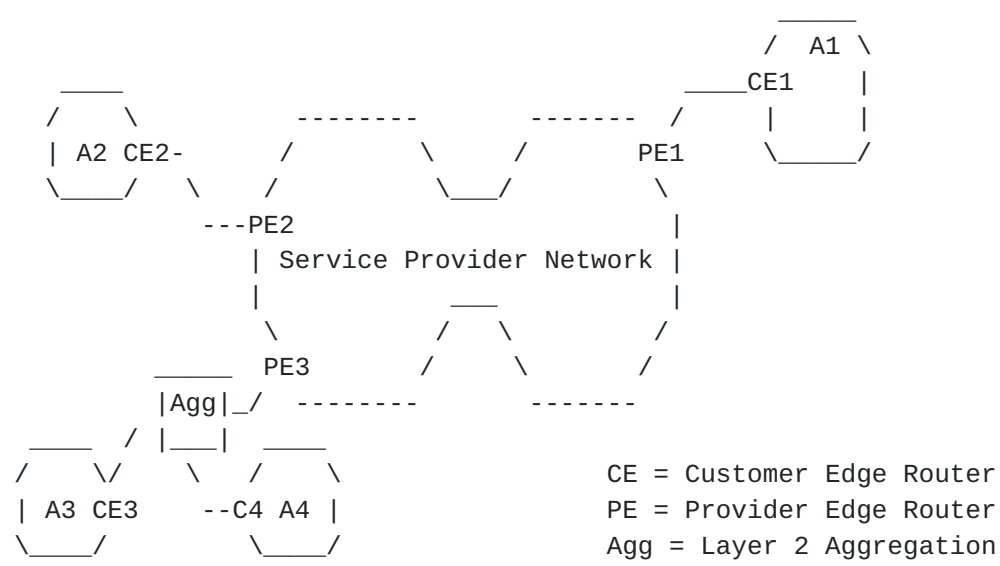
3. Virtual Private Switched Network

In a VPSN, packets are carried across a transit network so that the source and destination networks operate as if inter-connected by a LAN. In the past, ATM has been used to provide TLS. Rather than overlay a TLS-like service over ATM, this document describes how a VPSN can be created over a tunneled network, where the tunneling technology is not hardware-specific. This document describes extensions to [3] to provide VPSN service. The encapsulation of the packets will follow the Ethernet or the Ethernet VLAN specifications given in [5]. All nodes of a VPSN MUST use the same encapsulation.

Throughout this document, we will use LDP as the signaling protocol for the VPSN. Targeted LDP between the PEs is all that is assumed. A future document addresses signaling using Multi-protocol extensions to BGP. We make no assumption about the nature of the transport tunnels that actually carry the traffic between PEs. For example, they may be traffic engineered tunnels set up with RSVP-TE, LDP, GRE or MPLS/IP tunnels.

4. Tunneled VPSN

Consider the following provider network, over which a VPSN service is to be provisioned. Customer A has three sites, with three local area networks, A1, A2, and A3, that need to be connected. CEs represent customer edge routers and PEs represent provider edge routers.



The general problem is MAC address discovery. If it is not known where the destination MAC addresses for a certain location are, then the only way for a packet sourced, say at A1, to reach one of the nodes in A2 or A3, is to broadcast (or multicast) the packet. Broadcasts and multicasts are hard to do over a tunneled network, where the efficiencies of sending single packets are lost. However, this is not unlike the problem of interconnecting multiple LANs at a single location. The problem is solved using learning bridges which learn MAC addresses and are able to squelch the transmission of packets to a MAC address on a particular subnet because they know that MAC address is not there.

Using the network below as an example setup, we describe two different methods for making VPSN services possible. Then we describe the signaling required to make it all work.

4.1. MAC Address Learning

Initially, the VPSN is set up so that PE1, PE2 and PE3 have a full-mesh of tunnels between them for carrying tunneled traffic. The VPSN service is assigned a VCID (a 32-bit quantity that is unique across the provider network across all VPSNs). The VC Type for VPSNs, the Ethernet VLAN VC type, is associated with the VCID. Unlike [3], a VCID is unique within a provider domain. Allocation of domain-wide unique VCIDs is outside the scope of this draft.

For the above example, say PE1 signals VC Label 102 to PE2 and 103 to PE3, and PE2 signals VC Label 201 to PE1 and 203 to PE3.

Assume a packet from A1 is bound for A2. When it leaves CE1, say it

has a source MAC address of M1 and a destination MAC of M2. If PE1 does not know where M2 is, it will multicast the packet to PE2 and

PE3. When PE2 receives the packet, it will have an inner label of 201. PE2 can conclude that the source MAC address M1 is behind PE1, since it distributed the label 201 to PE1. It can therefore associate MAC address M1 with VC Label 102.

Note that the two different encapsulations, Ethernet and Ethernet VLAN, lead to two slightly different learning algorithms. We describe them below.

4.1.1. MAC Address Learning with Ethernet Encapsulation

When the Ethernet encapsulation is used, the PE is service unaware, i.e., it does not distinguish between frames that have 802.1q tags and those that do not. The model is one that allows overlaying multiple VLANs over a single VPSN. In this model, a PE MUST learn based on both the 802.1q tag and the MAC address. This is to take care of unfortunate duplicate MAC addresses used in different customer VLANs.

4.1.2. MAC Address Learning with Ethernet VLAN Encapsulation

When the Ethernet VLAN encapsulation is used, the PE is service aware, i.e., it associates that particular VLAN with the VPSN. The PE can safely learn based on the MAC address alone.

5. MAC Address Management

From the above description, it is clear that MAC addresses are being learned at multiple locations. For example, the CEs may learn MAC addresses through ARP (for IPv4 traffic), since the VPSN service behaves like a LAN. CEs can be out of sync with PEs that also have to learn MAC addresses and associate them with VC Labels. (This is one reason why a CE may know the MAC address of another CE router, but the PE routers may need to relearn them).

5.1. Aging MAC Addresses

PEs that learn remote MAC addresses need to have an aging mechanism to remove unused entries associated with a VC Label. This is important both for conservation of memory as well as for administrative purposes. For example, if customer site A1 has another CE connected to PE1, and CE1 is shut down, eventually, the other PEs should unlearn CE1's MAC address.

As with existing LAN bridges, two aging timers SHOULD be implemented on a PE. First, a local aging of MAC addresses learned from the

customer-facing network SHOULD be implemented with a shorter value of the timer. Second, a remote aging of MAC addresses learned

during the operation of the VPSN SHOULD be implemented with a considerably longer timer value. The remote aging timer keeps entries around longer, since the loss of an entry entails a broadcast across the VPSN to discover the MAC address location.

As packets arrive from the customer-facing network, local MAC addresses SHOULD be remembered, along with aging. The aging timer for MAC address M SHOULD be reset when a packet is received from the customer-facing with source MAC address M.

As packets arrive from the remote PEs, remote MAC addresses SHOULD be learned. The aging timer for a remote MAC address M SHOULD be reset when a packet arrives from a remote PE with source MAC address M.

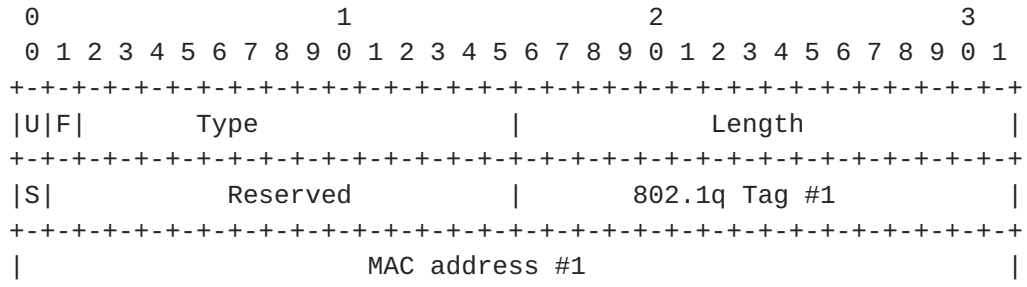
5.2. MAC Address Signaling

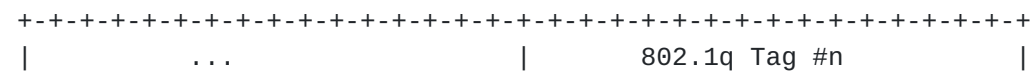
There should be a more proactive manner of installing MAC address associations and removing them for faster convergence. We introduce a MAC TLV that is used to specify a list of MAC addresses that can be added or removed using the Address Message and the Address Withdraw Message, respectively.

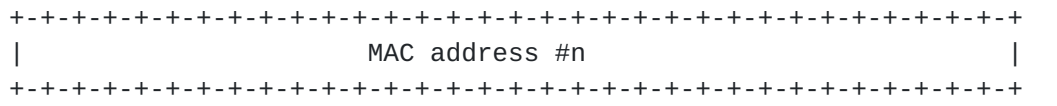
The Address Withdraw message with MAC TLVs SHOULD be supported in order to uninstall learned MAC addresses that have moved or gone away more quickly. It is not quite as essential that the Address message with MAC TLVs be supported. Once a MAC address is unlearned, re-learning occurs through flooding, so the Address message only prevents flooding. The Address message MAY be supported.

5.2.1. MAC TLV

MAC addresses can be signaled using an LDP Address Message. We define a new TLV, the MAC TLV. Its format is described below. The encoding of a MAC TLV address is a 2-byte 802.1q tag, followed by the 6-byte MAC address encoding specified by IEEE 802 documents [6]. The 802.1q tag and the MAC address MUST appear in pairs. If no tag is required, the value of the tag field MUST be zero.




...
802.1q Tag #n



U bit
Unknown bit. This bit MUST be set to 0. If the MAC address format is not understood, then the TLV is not understood, and MUST be ignored.

F bit
Forward bit. This bit MUST be set to 0. Since the LDP mechanism used here is Targeted, the TLV MUST NOT be forwarded.

Type
Type field. This field MUST be set to 0x0404 (subject to IANA approval). This identifies the TLV type as MAC TLV.

Length
Length field. This field specifies the total length of the TLV, including the Type and Length fields.

S bit
Static bit. In an Address Message, this bit indicates whether the MAC addresses specified in the TLV are static or dynamic. If the bit is set, the addresses are static, and MUST NOT be aged out. If it is clear, the addresses are dynamic, and SHOULD be aged. The S bit has no significance in an Address Withdraw Message, and MUST be zero.

Reserved
Reserved bits. They MUST NOT be interpreted at the receiver, and MUST be set to zero by the sender.

802.1q Tag
The 802.1q Tag. The value MUST be zero if the Ethernet VLAN encapsulation is used. If the Ethernet encapsulation is used, and the Ethernet address is associated with a VLAN, it MUST be set to the VLAN tag. If the Ethernet encapsulation is used, and the MAC address is not associated with a VLAN, it MUST be set to zero. Since an 802.1q tag is 12-bits, the high 4 bits of the field MUST be set to zero.

MAC Address
The MAC address being signaled.

The LDP Address Message contains a FEC TLV (to identify the VPSN in consideration), a MAC Address TLV and optional parameters. No optional parameters have been defined for MAC Address signaling.

The LDP Address Withdraw Message contains a FEC TLV (to identify the

VPSN in consideration), a MAC Address TLV and optional parameters.

No optional parameters have been defined for the MAC Address Withdraw signaling.

5.2.2. Address Message Containing MAC TLV

The processing for MAC TLVs received in an Address Message is:

For each (q-tag, MAC address) pair in the MAC TLV:
- If a mapping between the (q-tag, MAC address) pair and a VC label exists, then, remove the existing mapping, and replace it with the new association. If the S bit is on in the TLV, then each (q-tag, MAC address) is not aged. Otherwise, an aging timer with the remote aging timer value SHOULD be started.
- If a mapping does not exist, then install a new mapping between (q-tag, MAC Address) pair and VC label. If the S bit is on in the TLV, then each the mapping is not aged. Otherwise, an aging timer with the remote aging timer value SHOULD be started.

The scope of a MAC TLV is the VPSN specified in the FEC TLV in the Address Message.

Note that each MAC TLV contains a number of (q-tag, MAC address) pairs with the same property, i.e., either static or dynamic. A single Address Message MAY contain multiple MAC TLVs. The number of MAC addresses can be deduced from the length field in the TLV.

5.2.3. Address Withdraw Message Containing MAC TLV

When MAC addresses are being removed explicitly, e.g., an adjacent CE router has been disconnected, an Address Withdraw Message can be sent with the list of MAC addresses to be withdrawn.

The processing for MAC TLVs received in an Address Withdraw Message is:

For each (q-tag, MAC address) pair in the TLV:
- Remove the association between the (q-tag, MAC address) pair and VC label. It does not matter whether the MAC address was installed as a static or dynamic address.

The scope of a MAC TLV is the VPSN specified in the FEC TLV in the Address Withdraw Message.

The number of MAC addresses can be deduced from the length field in the TLV. The address list MAY be empty. In this case, the S bit and the 15-bit Reserved field are not sent, i.e., the length field

would be set to 4. This tells the receiving LSR to delete any MAC

addresses learned from the sending LSR for the VPSN specified by the FEC TLV.

5.2.4. LDP Session Failure or Termination

When a targeted LDP session is torn down or terminated, all associated MAC address mappings MUST be removed.

5.3. Discussion on MAC TLVs

Several standard bridging issues can be handled with MAC Address registrations and withdrawals without having to resort to a spanning tree protocol. They arise from topology changes that move a MAC address from one PE to another either because of a physical reorganization, or because of a spanning tree action on the customer-facing network. These TLVs MAY be triggered by some conditions that can be vendor-specific, depending on the types of CE-PE interactions the vendor supports.

For example, suppose a CE router connects in to a VPSN at PE1. All the PEs participating in the VPSN learn of its MAC address and that it is served by PE1. If the CE router disconnects from PE1 and connects to PE2, all the PE routers need to re-learn its new position. If the other PEs wait until a packet from PE2 is sent carrying M as the source MAC, then only those PEs who receive such packets will be able to update their MAC mapping tables. Instead, a PE SHOULD send an LDP Address Withdraw message with a MAC TLV to flush out entries. There are two cases where it is advisable to send MAC TLVs.

5.3.1. Physical Network Reorganization

When a CE router, CE1, connects in to a different PE than it used to, the PEs participating in the VPSN should all discover this change as soon as possible. Suppose CE1 connects in to PE1 initially. All PEs in the VPSN learn of CE1's MAC address. At some point, CE1 is disconnected from PE1, and reconnected at PE2. PE2 locally learns that CE1 is behind it. Since it already has CE1's MAC address in its table, registered as being served by PE1, PE2 SHOULD send out an Address Withdraw message with CE1's MAC address.

5.3.2. Spanning Tree Action

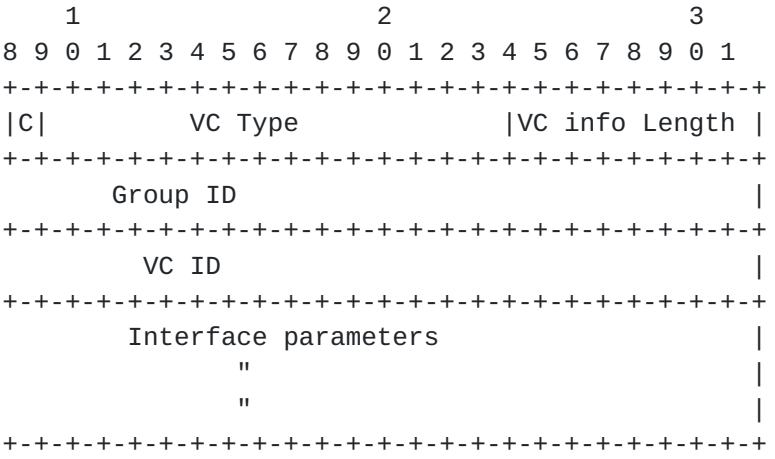
A CE router, CE1, may be multi-homed into the provider network at

PE1 and PE2. If spanning tree is run on the customer-facing network, then one of the links CE1-PE1 or CE1-PE2 will be blocked.

Suppose CE1-PE1 is active and CE1-PE2 is blocked. Now, if CE1-PE1 goes down, CE1-PE2 is unblocked, and CE1's MAC address appears to be served by PE2. PE2 SHOULD send out an Address Withdraw message with CE1's MAC address.

6. Signaling a VPSN

As described in [3], LDP will be used in Downstream Unsolicited mode to distribute VC labels. LDP will be used with targeted peers. The FEC TLV is defined as in [3].



VC, C, VC Type, VC Info Length, Group ID, Interface parameters
As defined in [3].

VCID
The VCID is defined much as in [3]. The only difference is that it is globally unique within a provider domain, independently of the VC Type.

6.1. Adjacency and Session Management

As in normal LDP Downstream Unsolicited operation, when a PE is configured as part of a VPSN, it issues a Targeted Hello to each other PE in the VPSN. The actual discovery of those other PEs that are part of the VPSN will be addressed in companion drafts, using BGP advertisements and LDP signaling, respectively.

When a Targeted Hello is received, if the receiving PE is not configured to be part of the VPSN, it MAY send back a Label Release. However, the Liberal Label Retention SHOULD be used [3], wherein a PE that does not participate in a VPSN may still retain a received VC label. The VC would be set up only when the PE is configured to be a member of the VPSN, and reciprocates with its own VC label mapping.

Note that the act of instantiating a VPSN on a PE triggers LDP session setup and VC label exchange, and since VC label exchange

occurs in both directions, Liberal Label Retention Mode is not necessary.

7. Security Considerations

No new security issues result from this draft. It is recommended in [2] that LDP security (authentication) methods [7] be applied. This would prevent unauthorized participation by a PE in a VPSN. Traffic separation for VPSNs is maintained using VC labels. However, for additional levels of security, the customer MAY deploy end-to-end security, which is out of the scope of this draft.

8. Intellectual Property Disclaimer

This document is being submitted for use in IETF standards discussions.

9. References

[1] "Service Requirements for Provider Provisioned Virtual Private Networks", M. Carugi, et al. June 2001. draft-ietf-ppvpn-requirements-01.txt. Work in progress.

[2] "Requirements for Virtual Private Switched Networks", G. Heron, et al. July 2001. draft-heron-ppvpn-vpsn-reqmts-00.txt. Work in progress.

[3] "Transport of Layer 2 Frames Over MPLS", L. Martini, et al. May 2001. draft-martini-l2circuit-trans-mpls-06.txt. Work in progress.

[4] "MPLS-Based Layer 2 VPNs", K. Kompella, et al. draft-kompella-ppvpn-l2vpn-00.txt. June 2001. Work in progress.

[5] "Encapsulation Methods for Transport of Layer 2 Frames Over MPLS", L. Martini, et al. draft-martini-l2circuit-encap-mpls-02.txt. February 2001. Work in progress.

[6] IEEE STD 802.3-2000. October 2000.

[7] "LDP Specification", L. Andersson, et al. RFC 3036. January 2001.

[8] IEEE STD 802.1Q-1998. December 1998.

10. Authors' Addresses

Vach Kompella
TiMetra Networks
274 Ferguson Dr.
Mountain View, CA 94043
Email: vkompella@timetra.com

Nick Tingle
TiMetra Networks
274 Ferguson Dr.
Mountain View, CA 94043
Email: ntingle@timetra.com

Sunil Khandekar
TiMetra Networks
274 Ferguson Dr.
Mountain View, CA 94043
Email: sunil@timetra.com

Giles Heron
PacketExchange Ltd.
The Truman Brewery
91 Brick Lane
LONDON E1 6QL
United Kingdom
Email: giles@packetexchange.net

Juha Heinanen
Song Networks, Inc.

Tom S. C. Soon
SBC Technology Resources Inc.
4698 Willow Road
Pleasanton, CA 94588
sxsoon@tri.sbc.com

Rick Wilder
Masergy Inc.
2901 Telestar Ct.
Falls Church, VA 22042

Luca Martini
Level 3 Communications, LLC.
1025 Eldorado Blvd.
Broomfield, CO, 80021
Email: luca@level3.net