Himanshu Shah
Ciena Networks

K.Arvind
Enterasys Networks

PPVPN Working Group
Internet Draft
draft-ietf-l2vpn-ipls-00.txt

Eric Rosen
Francois Le Faucheur
Cisco Systems

November 2003
Expires: May 2004

Giles Heron
PacketExchange,Ltd

Vasile Radoaca
Nortel Networks

IP-Only LAN Service (IPLS)

Status of this Memo

Abstract

A Virtual Private LAN Service (VPLS) [VPLS] is used to interconnect systems across a wide-area or metropolitan-area network, making it

appear to those systems as if they are interconnected on a private LAN. The systems which are interconnected in this way may themselves be LAN switches. If, however, the interconnected systems are NOT LAN switches, but rather are IP hosts or IP routers, certain simplifications are possible. We call this simplified type of virtual private LAN service an ?Ip-only LAN Service? (IPLS). In

IPLS, as in VPLS, LAN interfaces are run in promiscuous mode, and frames are forwarded based on their MAC Destination Addresses. However, the maintenance of the MAC forwarding tables is done via signaling, rather than via the ?MAC Address Learning? procedures of IEEE 802.1D. Further, Address Resolution Protocol (ARP) messages are proxied, rather than being carried transparently. This draft specifies the protocols and procedures for support of the IPLS service.

1.0 Conventions

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

RELATED DOCUMENTS
draft-ietf-ppvpn-l2-framework-03.txt
draft-ietf-ppvpn-vpls-ldp-01.txt
draft-ietf-pwe3-control-protocol-04.txt

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

Belongs in PPVPN

WHY IS IT TARGETED AT THIS WG

This document describes a mechanism to assist in Provider-Provisioned Layer 2 VPNs.

JUSTIFICATION

This document provides a detailed description for IP-only LAN Service (IPLS), which is discussed in the L2 PPVPN Framework [PPVPN-FWK]. The VPLS [VPLS] services of L2VPN require PE devices to function as MAC learning bridges. IPLS is a solution for a specific topology where MAC learning capabilities are not required for VPLS services, because user data traffic is restricted to IP, and the CE devices are not LAN switches.

Table of Contents

2.0 Overview

As emphasized in [VPLS], Ethernet has become popular as an access technology in Metropolitan and Wide Area Networks. [VPLS] describes how geographically dispersed customer LANs can be interconnected over a service provider?s network using Layer 2 VPNs. The VPLS service is provided by Provider Edge (PE) devices, and it is provided to Customer Edge (CE) devices. The VPLS architecture provides such services by incorporating bridging functions such as MAC address learning in the PE devices.

There are Provider Edge platforms, both existing and forthcoming, which have been designed primarily to be IP routers, rather than to be LAN switches. It can be fairly straightforward to add a MAC address lookup capability to these platforms, and to run their LAN interfaces in promiscuous mode, so that they can forward frames based on the MAC Destination Address of the frame. It is less straightforward to add the IEEE 802.1D MAC Address learning capability to these platforms. However, as discussed in [L2VPN-FWK], in scenarios where the CE devices are NOT LAN switches, but rather are IP hosts or IP routers, it is possible to provide the virtual private LAN service without requiring IEEE 802.1D MAC address learning/aging on the PE. Due to these restrictions, such a service is referred to as an "IP-Only LAN Service", or IPLS. Requirements for such an IPLS service are presented in [L2VPN-REQTS]. The purpose of this draft is to specify a solution optimized for this IPLS service.

Consequently, IPLS allows a service provider to provide a VPLS-like service by using PE routers that are not designed to perform general LAN bridging functions. However one must be willing to accept the restriction that the Virtual LAN service be used for IP traffic only, and not used to interconnect CE devices that are themselves LAN switches. This seems like an acceptable restriction in many environments, given that IP is the predominant type of traffic in today's networks.

In IPLS, a PE device implements multi-point LAN connectivity for IP traffic using the following key functions:

1. Discovery: Each Provider Edge (PE) device discovers IP/MAC address associations for the locally attached Customer Edge (CE) devices, for each IPLS instance configured on the PE device.

2. Pseudowire (PW) for Unicast Traffic: For each locally attached CE device in a given IPLS instance, a PE device sets up a pseudo-wire (VC-LSP) to each of the other PEs that supports the same IPLS instance.

   For instance, if PEx and PEy both support IPLS I, and PEy is locally attached to CEw and CEz, PEy will initiate the setup of two Pseudowires between itself and PEx. One of these will be used to carry unicast traffic from any of PEx?s CE devices to CEw. The other will be used to carry unicast traffic from any of PEx?s CE devices to CEz.

Note that these Pseudowires carry traffic only in one direction. Further, while the Pseudowire implicitly identifies the destination CE of the traffic, it does not identify the source CE; packets from many CEs may be freely intermixed on a given Pseudowire.

3. Pseudowires for Multicast Traffic: In addition, every PE supporting a given IPLS instance will set up a special ?multicast Pseudowire? to every other PE in that IPLS instance. If, in the above example, one of PEx?s CE devices sends a multicast packet, PEx would forward the multicast packet to PEy on the special multicast Pseudowire. PEy would then send a copy of that packet to CEw and a copy to CEz.

   Thus when a PE sends a multicast packet across the network, it sends one copy to each remote PE (supporting the given IPLS instance). If a particular remote PE has more than one CE device in that IPLS instance, the remote PE must replicate the packet and send one copy to each of its local CEs.

   As with the Pseudowires that are used for unicast traffic, packets travel in only one direction on these Pseudowires, and packets from different sources may be freely intermixed.

4. Signaling: The necessary Pseudowires can be set up and maintained using the LDP-based signaling procedures described in [PWE3-CONTROL] and/or [ROSEN-SIG]. Use of other signaling procedures is for further study.

   A PE may assign the same label to each of the unicast Pseudowires that lead to a given CE device, in effect creating a multipoint-to-point Pseudowire.

   Similarly, a PE may assign the same label to each of the multicast Pseudowires for a given IPLS instances, in effect creating a multipoint-to-point Pseudowire.

   When setting up a Pseudowire to be used for unicast traffic, the PE must also signal the IP address and the MAC address of the corresponding CE device.

5. Proxy ARP: Distribution of IP/MAC address associations to remote PE devices via PW signaling enables each PE device to function as a proxy ARP server for CE devices attached to other PE devices. This makes it possible for any CE device to ARP for

the MAC addresses of remote CE devices.

6. Forwarding: A PE device programs its Forwarding Information Base using the CE MAC addresses and VC labels signaled through the PW signaling. Unicast IP traffic from the local CEs is then switched to the proper VC-LSP based on the destination MAC address. Multicast IP traffic from the local CEs is replicated by the local PE over all the Attachment Circuits (except the one it came in) and all the multicast VC-LSPs for that IPLS instance. Remote PEs that receive the multicast packets over the multicast VC-LSPs then replicates onto all its Attachment Circuits for that IPLS instance.

Both VPLS [VPLS] and IPLS require the ingress PE to forward a frame based on its destination MAC address. However, two key differences between VPLS and IPLS can already be noted from the above description:

- In VPLS, MAC entries are placed in the FIB of the ingress PE as a result of IEEE 802.1D MAC address learning (which occurs in the data plane) while in IPLS MAC entries are placed in the FIB as a result of Pseudowire signaling operations (control plane).
- In VPLS, the egress PE looks up a frame?s MAC destination address to determine the customer-facing interface out which the frame must be sent; in IPLS, the choice of interface is based entirely on the VC-label.

The following sections describe the details of the IPLS scheme.

2.1 Terminology

IPLS	Ip-only LAN service (a type of Virtual Private LAN Service that is restricted to IP traffic only).
IPLS Network	A collection of PE nodes supporting the IPLS service and the associated mechanisms described in this document, including the Extended LDP based PW signaling between them.
IPLS Service	A single service instance of IPLS emulating a LAN segment for IP data traffic.
MPt-Pt PW	Multipoint-to-Point Pseudowire. A Pseudowire

that carries traffic from remote PE devices to a PE device that signals the Pseudowire. The signaling PE device advertises the same VC-label to all remote PE devices that participate in the IPLS service instance. In IPLS, for a given IPLS instance, a MPt-Pt PW used for IP unicast traffic is established by a PE for each CE device locally attached to that PE. It is a unidirectional tree whose leaves consist of the remote PE peers (which connect at least one Attachment Circuit associated with the same IPLS instance) and whose root is the signaling PE. Traffic flows from the leaves towards the root.

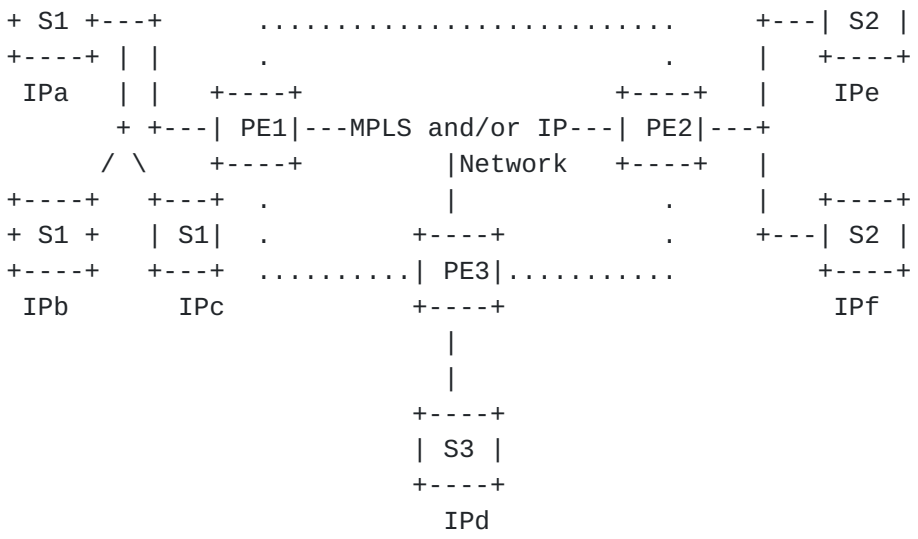
Multicast PW — Multicast Pseudowire. A special kind of MPt-Pt PW that carries only IP multicast/broadcast traffic. In the IPLS architecture, for each IPLS instance supported by a PE, that PE device establishes exactly one Multicast PW.

CE — Customer Edge device. In this document, a CE is any IP node (host or router) connected to the IPLS LAN service.

Replication Tree — The collection of all Multicast Pseudowires and attachment circuits that are members of an IPLS service instance on a given PE. When the PE on an attachment circuit receives a multicast/broadcast packet, the PE device sends a copy of the packet to every Multicast Pseudowire and attachment circuit of the replication tree, excluding the attachment circuit on which the packet was received.

3.0 Topology

The Customer Edge (CE) devices are IP nodes (hosts or routers) that are connected to PE devices either directly, or via an Ethernet network. We assume that the PE/CE connection may be regarded by the PE as an ?interface? to which one or more CEs are attached. This interface may be the physical LAN interface or a VLAN. The Provider Edge (PE) routers are MPLS Label Edge Routers (LERs) that serve as Pseudowire endpoints.



In the above diagram, an IPLS instance is shown with three sites: site S1, site S2 and site S3. In site S3, the CE device is directly connected to its PE. In the other two sites, there are multiple CEs connected to a single PE. More precisely, the CEs at these sites are on an Ethernet (switched at site 1 and shared at site 2) network (or VLAN), and the PE is attached to that same Ethernet network or VLAN). We impose the following restriction: if one or more CEs attach to a PE by virtue of being on a common LAN or VLAN, there MUST NOT be more than one PE on that LAN or VLAN.

PE1, PE2 and PE3 are shown as connected via an MPLS network; however, other tunneling technologies, such as GRE, L2TP, etc., could also be used to carry the Pseudowires.

An IPLS instance is a single broadcast domain, such that each IP end station (e.g., IPa) appears to be co-located with other IP end stations (e.g., IPb though IPf) on the same subnet. The IPLS service is transparent to the CE devices and requires no changes to them.

4.0 Configuration

Each PE router is configured with one or more IPLS service instances, and each IPLS service instance is associated with a unique VPN-Id. For a given IPLS service instance, a set of Attachment Circuits is identified. Each Attachment Circuit can be associated with only one IPLS instance. An Attachment Circuit, in this document, is either a customer-facing Ethernet port, or a particular VLAN (identified by an IEEE 802.1Q VLAN ID) on a customer-facing Ethernet port.

The PE router can optionally be configured with a local MAC address to be used as source MAC address when packets are forwarded from a Pseudowire to an Attachment Circuit. By default, a PE uses the MAC address of the customer-facing Ethernet interface for this purpose.

5.0 Discovery

The discovery process includes:

- Remote PE discovery
- VPN (i.e., IPLS) membership discovery
- IP CE end station discovery

This draft does not discuss the remote PE discovery or VPN membership discovery. This information can either be user configured or can be obtained using auto-discovery techniques described in [BGP-Discovery] or other methods. However, the discovery of the CE is an important operational step in the IPLS model and is described below.

5.1 CE discovery

Each PE actively detects the presence of local CEs by snooping IP and ARP frames received over the Attachment Circuits. During the discovery phase, the PE examines each broadcast/multicast Ethernet frame. For IP frames (for example IGP discovery/multicast/broadcast packets typically 224.x.x.x addresses), the CE?s (source) MAC address is extracted from the Ethernet header and the (source) IP address is obtained from the IP header. For ARP frames, the source MAC and IP address are determined from the ARP PDU.

For each CE, the PE maintains a <Attachment Circuit identification info, VPN-Id, IP address, MAC address> tuple.

Once discovered, the presence/active-status of a CE is monitored continuously by examining the received ARP frames and by periodically generating ARP requests. The absence of an ARP response from a CE after a configurable number of such ARP requests, is interpreted as a loss of connectivity with the CE.

6.0 Pseudowire Creation

6.1 Receive Unicast Multipoint-to-point Pseudowire

As the PE discovers each locally attached CE, a unicast Multipoint-

to-point Pseudowire (MPt-Pt PW) associated exclusively with that CE is created by distributing the CE?s IP address and MAC address along with a VC-Label to all the remote PE peers that participate in the same IPLS instance. Note that the same value of a VC-label should be distributed to all the remote PE peers for a given CE. The MP-Pt PW thus created is used by remote PEs to send unicast IP traffic to a specific CE.

(The same functionality can be provided by a set of point-to-point PWs, so the PE is not required to send the same VC-label to all the other PEs. For convenience however, we will speak in the following only of multipoint-to-point PWs, without pointing out each time that a set of point-to-point PWs could be used instead.)

The PE forwards a frame received over this MPt-Pt PW to the associated Attachment Circuit.

6.2 Receive Replication Multipoint-to-point Pseudowire

When a PE is configured to participate in an IPLS instance, it advertises a "multicast" VC-label to every other PE that is a member of the same IPLS. The advertised VC-label value is the same for each PE, which creates a multipoint-to-point Pseudowire for IP multicast traffic. There is only one multicast MPt-Pt PW per PE for each IPLS instance and this Pseudowire is used exclusively to carry multicast/broadcast IP traffic from the remote PEs to this PE for this IPLS instance.

Note that no special functionality is expected from this Pseudowire. We sometimes call it a ?multicast Pseudowire? because we use it only to carry multicast traffic. The Pseudowire itself need not provide any different service than any of the unicast Pseudowires.

In particular, the Receive multicast MPt-PT PW does not perform any replication of frames itself. Rather, it is there to signify to the PE that the PE needs to replicate a copy of a frame received over this MPt-Pt PW onto all the attachment circuits that are associated with the IPLS instance of the MPt-Pt PW.

The use of Pseudowires, which are specially optimized for multicast, is for further study.

6.3 Send Multicast Replication tree

The PE creates a send replication tree for each IPLS instance, which consists of the collection of all attachment circuits and all the ?multicast? Pseudowires of this IPLS instance.

Any broadcast/multicast frame received over an attachment circuit is replicated to all the other attachment circuits and all Pseudowires of the send replication tree of the IPLS instance of the incoming Attachment Circuit.

7.0 Proxy ARP

As part of the signaling of the unicast multipoint-to-point pseudo-wire (See Section 8), each PE distributes to its remote PE peers the CE IP address/MAC address associations that it has discovered. The remote PE peers then build and maintain a database of these associations.

When a PE receives an ARP request from a local CE for a remote CE, it searches for the destination IP address in the database associated with the CE?s IPLS instance. If a match is found, the PE sends an ARP response with the MAC address of the remote CE. This enables the local CE to send unicast IP frames addressed directly to the MAC address of the remote CE.

8.0 Signaling

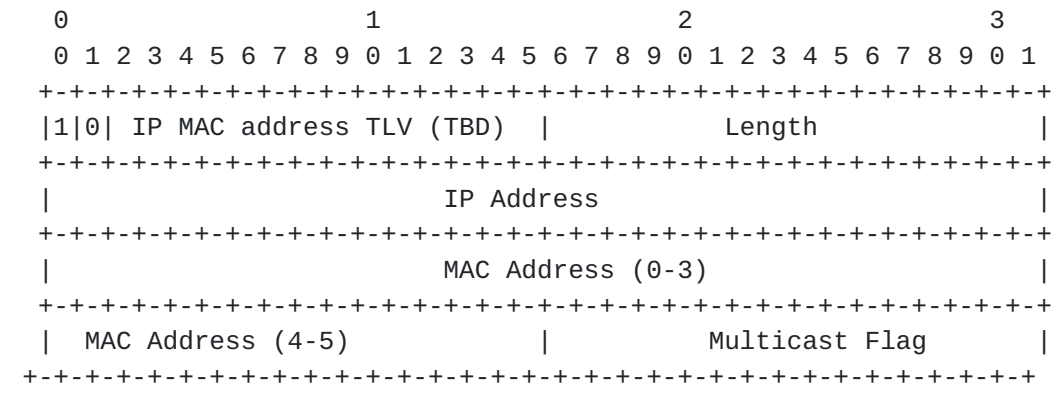
The [PWE3-CONTROL] uses Label Distribution Protocol (LDP) transport to exchange VC-FEC in the label mapping message in a downstream unsolicited mode. The VC-FEC comes in two flavors; Pwid and Generalized ID FEC elements. These FEC elements define some fields that are common between them. The discussions below refer to these common fields for IPLS related extensions.

8.1 IPLS PW Signaling

The IPLS uses user IP data as payload over the Pseduowire. The use of such encapsulation is identified by VC type field of the VC-FEC as the value 0x000B [PWE3-IANA].

In addition, this document defines an IP MAC address TLV that must be included in the optional TLV field of the label mapping message when advertising VC-FEC for the IPLS. Such use of optional TLV in the label mapping message to extend the attributes of the VC-FEC has also been specified in the [PWE3-Control].

When processing a received VC-FEC, the PE matches the VC-Id and VC-type with the locally configured VC-id to determine if the VC-FEC is of type IPLS. If matched, it further checks the presence of IP address TLV. If an IP MAC address TLV is absent, a label release message is issued to reject the PW establishment.



The Length field is defined as the sum of length of the IP address (4) and length of MAC address (6) and multicast flag (2) and is set to value 12.

The non-zero unicast value of the IP address field denotes IP address of advertising PE?s attached CE device.

The non-zero unicast value of the MAC address field denotes MAC address of the advertising PE?s attached CE device.

The Multicast Flag value of 1 indicates that the advertised VC-Label represents a ?multicast? PW. The Multicast Flag value of 0 indicates that advertised VC-label represents a ?unicast? PW. As explained earlier, the term ?multicast PW? only means that the PW carries IP broadcast/multicast traffic and does not refer to a multicast LSP in the traditional sense.

The Multicast Flag must be zero, if present, when the IP and MAC address parameters are present (and their value is non-zero). When Multicast Flag is set to 1, the values in IP and MAC Address fields are set to null and are ignored.

8.2 Signaling Advertisement Processing

A PE should process a received [PWE3-CONTROL] advertisement with VC-type of IPLS as follows,

- Verify the IPLS VPN membership by matching the VPN-Id signaled in the AGI field or the PW-ID field with all the VPN-Ids configured on the PE. Discard and release the VC label if VPN-Id is not found.
- Distribute the received IP address-to-MAC address binding by sending a gratuitous ARP response on all the attachment circuits associated with the VPN-Id.
- Program the Forwarding Information Base (FIB) such that when a packet is received from an attachment circuit with its destination MAC address matching the advertised MAC address,

the packet is forwarded out over the tunnel to the

advertising PE with the advertised VC-label as the inner label.

- When the advertised VC-label is ?multicast?, add the VC-label to the send multicast replication tree for the VPN-Id. This enables sending a copy of a multicast/broadcast IP frame from the attachment circuit to this Pseudowire.

8.3 Requesting for IP to MAC binding

It is possible that in some cases, some CEs may remain undetected in the absence of any multicast/broadcast IP or ARP packet generation. If a local CE needs to converse with a remote CE in this undetected set, it will proceed to generate ARP requests. The Proxy ARP scheme described so far will be unable to resolve the ARP request, since the address to be resolved would not have been discovered (signaled) yet.

In order to address such situations, an optional Address Resolution Request TLV can be included in the LDP?s Notification Message. This TLV contains an IP address parameter that represents the destination IP address that needs to be resolved. The PE may use some intelligent mechanisms (e.g., the number of ARP requests received for unknown IP destination within a certain interval exceeds a threshold) to detect the need for such advertisement. When the need is detected, the PE generates Notification Messages to all remote PEs in the IPLS, with the IP address parameter in the Address Resolution Request TLV set to the destination IP address to be resolved.

A PE that supports the Address Resolution Request TLV must, on receiving a notification message with this TLV, generate an ARP request message using the received IP address as the destination, and some already known IP and MAC address as the source (in the ARP PDU) on all Attachment Circuits associated with the IPLS instance.

In essence, this is a request to remote PEs to generate an ARP request on their Attachment Circuits to locate a specific CE and advertise a Label Mapping message back to the requesting PE. This can be seen as reverting to the usual full broadcasting of ARP messages throughout the Emulated LAN in rare cases when Proxy ARP fails.

The definition of Address Resolution Request TLV for the Notification message is the subject of future study.

8.4 CE MAC Address

Throughout this document we have referenced remote CE?s MAC address to be the 48-bit physical MAC address. The MAC address is learned and signaled by the remote PE while local PE uses the signaled MAC address to proxy ARP request for remote CE?s IP address, program the address in the FIB and use it as a key to forward packet from the Attachment Circuit to the Pseudowire.

Alternatively, it is also desirable to allow local PE to generate a unique 48-bit MAC address for the remote CE instead of using the signaled MAC address by the remote PE. The local PE would then use the generated MAC address for ARP proxy, programming the FIB and as a key to forward packets from the Attachment Circuit to the Pseudowire. By permitting address generation to represent each remote CE, local PE can use a key lookup algorithm that is most suitable for its architecture. For example, PE could use only 32-bits of the 48-bit MAC DA as the key for fast table lookup.

Which mechanism local PE uses to represent remote CE (i.e. using signaled MAC address or locally generated MAC address), is of local matter to the PE and has no bearing on the IPLS functionality.

9.0 Forwarding

9.1 Non-IP traffic

In an IPLS VPN, only IP traffic is forwarded by a PE. ARP frames are directed to the control plane in the PE and the rest of the frames are dropped silently. If the CEs must pass non-IP traffic to each other, they must do so through IP tunnels that terminate at the CEs themselves.

9.2 Unicast IP Traffic

In IPLS, IP traffic is forwarded from the Attachment Circuit to the PW based on the destination MAC address of the layer 2 frame (and not based on the IP Header).

To do so, the PE associates a Forwarding Information Base (FIB) with each IPLS instance and programs the FIB in the following manner:

- The PE programs its FIB when a CE (and its MAC address) is discovered on one of its Attachment Circuit such that a frame received on any other Attachment Circuit with destination to this CE, the frame is forwarded to the corresponding Attachment Circuit.
- The PE programs its FIB with the PW label such that a frame

received over that unicast PW is forwarded to the corresponding Attachment Circuit prepended with CE?s MAC address as the destination.

- The PE programs its FIB when processing a received PW signal such that a frame received from any of its Attachment Circuits associated with the same IPLS instance, is forwarded to the PW if the destination MAC address matches the one advertised in the PW signal.

The PE identifies the FIB associated with an IPLS instance based on the Attachment Circuit or the PW label. When a frame is received from an Attachment Circuit, PE uses destination MAC address as the

lookup key. When a frame is received from PW, PE uses VC-Label as the lookup key. The frame is dropped if the lookup fails.

9.3 Broadcasts and Multicast forwarding

When the destination MAC address is either a broadcast or multicast, a copy of the frame is sent to the control plane for CE discovery purposes (see section 5.1).

When a multicast/broadcast IP frame is received from an Attachment Circuit, a PE replicates it onto the Send Multicast Replication Tree (See section 6.3). When a multicast/broadcast IP frame is received from a Pseudowire, the PE forwards a copy of the frame to all attachment circuits associated with the IPLS VPN instance involved.

It is important to note that PEs participating in an IPLS VPN are responsible for translating a multicast IP address to a multicast Ethernet MAC address when forwarding frames from a ?multicast? Pseudowire to the Attachment Circuits. (The translation consists of recognizing the multicast IP address (224.x1.x2.x3) and appending the least significant three bytes of the IP address to 0x01-00-05 to construct the MAC address, e.g., 0x01-00-5E-x1-x2-x3 [RFC-1112]).

All other IP packets received over the ?multicast? MPt-Pt PW (such as directed broadcasts, subnet broadcasts, etc) are forwarded over Attachment Circuits using a broadcast MAC address.

9.4 Encapsulation

The Ethernet MAC header of a frame received from an Attachment Circuit is stripped before forwarding the frame to the appropriate Pseudowire. However, the MAC header is retained when a unicast or broadcast IP frame is directed to one or more Attachment Circuit(s). An IP frame received over a Pseudowire is prepended with a MAC

header before transmitting it on the appropriate Attachment Circuit(s). The fields in the MAC header are filled in as follows:

- The destination MAC address is the MAC address associated with the VC label in the FIB when the Pseudowire is unicast
- The destination MAC address is a multicast MAC address derived from the IP multicast address or the broadcast MAC address when the VC label is ?multicast?
- The source MAC address is the PE?s own local MAC address or a MAC address which has been specially configured on the PE for this use.
- The Ethernet Type field is 0x0800
- The frame may get IEEE802.1Q tagged based on the VLAN information associated with the Attachment Circuit.

An FCS field is appended to the frame.

10.0 Attaching to IPLS via ATM or FR

In addition to (i) an Ethernet port and a (ii) combination of Ethernet port and a VLAN ID, an Attachment Circuit to IPLS may also be (iii) an ATM or FR VC carrying encapsulated bridged Ethernet frames or (iv) the combination of an ATM or FR VC and a VLAN ID.

The ATM/FR VC is just used as a way to transport Ethernet frames between a customer site and the PE. The PE terminates the ATM/FR VC and operates on the encapsulated Ethernet frames exactly as if those were received on a local Ethernet interface. When a frame is propagated from Pseudowire to a ATM or FR VC, PE prepends the Ethernet frame with the appropriate bridged encapsulation header as define in [RFC 1487] and [RFC 1490] respectively. Operation of an IPLS over ATM/FR VC is exactly as described above, with the exception that the attachment circuit is then identified via the ATM VCI/VPI or Frame Relay DLCI (instead of via a local Ethernet port ID), or a combination of those with a VLAN ID.

11.0 VPLS vs IPLS

The VPLS approach proposed in [VPLS] provides VPN services for IP as well as other protocols. The IPLS approach described in this draft is similar to VPLS in many respects:

- It provides a Provider Provisioned Virtual LAN service with multipoint capability where a CE connected via a single attachment circuit can reach many remote CEs
- It appears as a broadcast domain and a single subnet
- forwarding is based on destination MAC addresses

However, unlike VPLS, IPLS is restricted to IP traffic only. By restricting the scope of the service to the predominant type of traffic in today's environment, IPLS eliminates the need for service provider edge routers to implement some bridging functions such as MAC address learning in the data path (by, instead, distributing MAC information in the control plane). Thus this solution offers a number of benefits:

- Facilitates Virtual LAN services in instances where PE devices cannot or cannot efficiently (or are specifically configured not to) perform MAC address learning.
- Does not require flooding of ARP frames normally. Also, unknown Unicast frames are never flooded as would be the case in VPLS.
- Encapsulation is more efficient (MAC header is stripped) while traversing the backbone network.
- PE devices are not burdened with the processing overhead associated with traditional bridging (e.g., STP processing, etc.). Note however that some of these overheads (e.g., STP processing) could optionally be turned-off with a VPLS solution in the case where it is known that only IP devices are interconnected.

- Loops (perhaps through backdoor links) are minimized since a PE could easily reject (via label release) a duplicate IP to MAC address advertisement.

12.0 IP Protocols

The solution described in this document offers IPLS service for IPv4 traffic only. For this reason, the MAC Header is not carried over the Pseudowire. It is reconstructed by the PE when receiving a packet from a Pseudowire and the Ethertype 0x0800 is used in the MAC Header since IPv4 is assumed.

However, this solution may be extended to carry other types of important traffic such as ISIS and IPv6 which are not encapsulated in Etherent with the use of Ethertype 0x0800. In order to permit the propagation of such packets correctly, one may create a separate set of Pseudowires, or pass protocol information in the "control word" of a "multiprotocol" Pseudowire, or encapsulate the Ethernet MAC Header in the Pseudowire. The selection of appropriate multiplexing/demultiplexing scheme is the subject of future study. The current document focuses on IPLS service for IPv4 traffic.

13.0 Dual Homing with IPLS

As stated in previous sections, IPLS prohibits connection of a common LAN or VLAN to more than one PE. Alternatively, CE device by itself can connect to more than one instance of IPLS through two separate LAN or VLAN connections to separate PEs. To the CE IP device, these separate connections appear as a connection to two IP subnets. The failure of reachability through one subnet is then resolved via other subnet by the IP protocols.

14.0 Acknowledgements

Authors would like to thank Nigel Burmeister and others at Tenor Networks for their valuable comments.

15.0 Security Considerations

The security aspects of this solution will be discussed at a later time.

16.0 References

[L2VPN-REQ] Augustyn, W. et.al "Service Requirements for Layer 2 Provider Provisioned Virtual Private Networks", draft-ietf-l2vpn-requirements-00.txt, Work in Progress, Internet Draft, May 2003.

[L2VPN-FMWK] Andersson, et.al, draft-ietf-l2vpn-l2-framework-03.txt, L2VPN framework, October 2003, (work in progress).

[PWE3-CONTROL] Martini et. Al., ?Pseudowire Setup and Maintenance using LDP?, draft-ietf-pwe3-control-protocol-04.txt, October 2003 (work in progress)

[PWE3-IANA] Martini et. Al., ?IANA Allocations for pseudo Wire Edge to Edge Emulation (PWE3)?, draft-ietf-pwe3-iana-allocation-02.txt, October 2003 (work in progress)

[VPLS] Lasserre et al, ?Virtual Private LN Service over MPLS?, draft-ietf-ppvpn-vpls-ldp-01.txt, November 2003 (work in progress).

[BGP-Discovery] ?Using BGP as an Auto-Discovery Mechanism for Provider Provisioned VPNs?, Ould-Brahim et al., draft-ietf-ppvpn-bgpvpn-auto-04.txt, May 2003, (work in progress).

[ARP] Plummer, D., "An Ethernet Address Resolution Protocol: Or Converting Network Protocol Addresses to 48.bit Ethernet

Addresses for Transmission on Ethernet Hardware", STD 37, RFC 826, November 1982.

[PROXY-ARP] Postel, J., "Multi-LAN Address Resolution", RFC 925, October 1984.

[RFC-1112] Deering, S., ?Host Extensions for IP Multicasting?, RFC 1112, August, 1989.

IPR Notice

Author's Address

Himanshu Shah

Ciena Networks
35 Nagog Park,
Acton, MA 01720
Email: hshah@ciena.com

K.Arvind
Enterasys Networks
50 Minuteman Rd, Suite 100
Andover, MA 01810
Email: karvind@enterasys.com

Eric Rosen
Cisco Systems
300 Apollo Drive,
Chelmsford, MA 01824
Email: erosen@cisco.com

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

Francois Le Faucheur
Cisco Systems, Inc.
Village d'Entreprise Green Side - Batiment T3
400, Avenue de Roumanille
06410 Biot-Sophia Antipolis
France
Email: flefauch@cisco.com

Vasile Radoaca
Nortel Networks
Email: vasile@nortelnetworks.com