A Framework for E-Tree Service over MPLS Network
draft-key-l2vpn-etree-frwk-00.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the
provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering
Task Force (IETF), its areas, and its working groups. Note that
other groups may also distribute working documents as Internet-
Drafts.

Internet-Drafts are draft documents valid for a maximum of six months
and may be updated, replaced, or obsoleted by other documents at any
time. It is inappropriate to use Internet-Drafts as reference
material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at

This Internet-Draft will expire on May 9, 2010.
Abstract

This document proposes a solution framework for supporting Metro Ethernet Forum (MEF) Ethernet Tree (E-Tree) services over a Multiprotocol Label Switching (MPLS) network. The objective is to provide a simple and effective approach to emulate E-Tree services in addition to Ethernet LAN (E-LAN) services on an existing MPLS network.

Conventions used in this document

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

1. Introduction.................................................... 4
  1.1. Objective and Scope........................................... 4
  1.2. Traditional Ethernet Network................................. 4
  1.3. MEF Multipoint Ethernet Services............................. 4
    1.3.1. Similarity between E-LAN and E-Tree.................... 5
    1.3.2. Difference between E-LAN and E-Tree.................... 5
  1.4. IETF Multipoint L2VPN Services.............................. 6
    1.4.1. Virtual Private LAN Service (VPLS)..................... 6
    1.4.2. Virtual Private Multicast Service (VPMS)................. 6
  1.5. Terminology................................................ 7

2. Reference Model................................................ 8

3. Use Cases...................................................... 10

4. Challenges.................................................... 12
  4.1. Generic E-Tree Service Definition.......................... 12
    4.1.1. Leaf-to-Leaf Communication Restriction................. 12
  4.2. Use Case Desirable Requirements............................ 12
    4.2.1. Point-to-Multipoint Bandwidth Optimisation............. 12
    4.2.2. Multicast Optimisation................................ 13
    4.2.3. MAC-based Forwarding Unnecessary....................... 13
    4.2.4. MAC-based Forwarding Security Concern.................. 13

5. MAC-based Forwarding E-Tree.................................... 15
  5.1. MAC-based Forwarding Any-to-Any Ethernet VPN............... 15
  5.2. Leaf-to-Leaf Communication Restriction.................... 15
    5.2.1. Per-payload Signaling on PW - From Leaf or Root........ 15
    5.2.2. Extension to VPLS..................................... 16
  5.3. Optional Enhancement - Point-to-Multipoint PW................ 17
  5.4. Optional Enhancement - Multicast in VPLS.................... 17

6. Non-MAC-based Forwarding E-Tree................................ 18
  6.1. Single Root, Broadcast Only - VPMS........................ 18
  6.2. Multiple Roots, Broadcast and Unicast...................... 18

7. Security Consideration........................................ 19

8. IANA Considerations........................................... 19

9. Acknowledgements.............................................. 19

10. References................................................ 19
    10.1. Normative References.................................... 19
    10.2. Informative References.................................. 20

Authors’ Addresses................................................ 21

Intellectual Property and Copyright Statements.................. 21
1. Introduction

1.1. Objective and Scope

This document proposes a solution framework for supporting Metro Ethernet Forum (MEF) Ethernet Tree (E-Tree) services over a MPLS network. The objective is to provide a simple and effective approach to emulate E-Tree services in addition to Ethernet LAN (E-LAN) services on an existing MPLS network.

This solution framework makes use of existing IETF specified mechanisms unless there are technical reasons why the existing mechanisms are insufficient or unnecessary.

This document does not intend to provide a full specification of the solution, but rather to identify the functional components of the overall solution, and for each component, whether it is MANDATORY or OPTIONAL, whether existing mechanism is sufficient, or whether relevant mechanism is already under development.

In this document, "current standard" refers to [RFC4385], [RFC4447], [RFC4448], [RFC4761] and [RFC4762].

1.2. Traditional Ethernet Network

In this document, traditional Ethernet network refers to the Ethernet bridge/switch network, not the Ethernet repeater/hub network.

Data frame is Ethernet frame.

Data forwarding is MAC-based forwarding.

It is important to note that in traditional Ethernet network unicast unknown, multicast and broadcast frames are forwarded in exactly the same way to every port except the ingress port.

An Ethernet host receiving a frame checks the destination address in the frame to decide whether it is the intended destination.

1.3. MEF Multipoint Ethernet Services

MEF defines two multipoint Ethernet Service types:
- E-LAN (Ethernet LAN), multipoint-to-multipoint service
- E-Tree (Ethernet Tree), rooted-multipoint service

According to MEF’s technical specification, a generic E-LAN/E-Tree service is always bidirectional in the sense that ingress frames can originate at any endpoint in the service. However, some application scenarios of E-Tree may have unidirectional traffic only. Section 3 will discuss about different use cases.
For full specification, please refer to MEF’s "Ethernet Services Definitions - Phase 2" [MEF6.1] and "Ethernet Services Attributes - Phase 2" [MEF10.1].

1.3.1. Similarity between E-LAN and E-Tree

Data frame MUST be Ethernet frame.

Data forwarding can be MAC-based forwarding or something else, to be specified by service provider in the particular service definition.

Extract from [MEF6.1]:

<table>
<thead>
<tr>
<th>EVC Service Attribute</th>
<th>E-LAN/E-Tree Service Type Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicast Service Frame</td>
<td>Deliver Unconditionally or Deliver Conditionally.</td>
</tr>
<tr>
<td>Delivery</td>
<td>If Delivered Conditionally, MUST specify the delivery criteria.</td>
</tr>
<tr>
<td>Multicast Service Frame</td>
<td>Deliver Unconditionally or Deliver Conditionally.</td>
</tr>
<tr>
<td>Delivery</td>
<td>If Delivered Conditionally, MUST specify the delivery criteria.</td>
</tr>
<tr>
<td>Broadcast Service Frame</td>
<td>Deliver Unconditionally or Deliver Conditionally.</td>
</tr>
<tr>
<td>Delivery</td>
<td>If Delivered Conditionally, MUST specify the delivery criteria.</td>
</tr>
</tbody>
</table>

It is important to note that it is not a must for a MEF multipoint Ethernet service (E-LAN or E-Tree) to use MAC-based forwarding. This document presents a solution framework for MAC-based forwarding E-Tree in section 5, and also discusses non-MAC-based forwarding E-Tree in section 6.

1.3.2. Difference between E-LAN and E-Tree

Within the context of a multipoint Ethernet service, each endpoint is designated as either a Root or a Leaf. A Root can communicate with all other endpoints in the same multipoint Ethernet service, however a Leaf can only communicate with Roots but not Leafs.

The only difference between E-LAN and E-Tree is:
- E-LAN has Root endpoints only, which implies there is no communication restriction between endpoints
- E-Tree has both Root and Leaf endpoints, which implies there is a need to enforce communication restriction between Leaf endpoints
Extract from [MEF10.1]:
The UNI Type MUST have the value either "Root" or "Leaf." If the type of EVC is Point-to-Point or Multipoint-to-Multipoint, then the UNI Type MUST equal "Root."

Extract from [MEF10.1]:
An ingress Service Frame mapped to the EVC at a Leaf UNI MUST NOT result in an egress Service Frame at another Leaf UNI but MAY result in an egress Service Frame at some or all of the Root UNIs.

1.4. IETF Multipoint L2VPN Services

1.4.1. Virtual Private LAN Service (VPLS)

VPLS is a L2VPN service that provides multipoint-to-multipoint connectivity for Ethernet across an IP or MPLS-enabled IP Packet Switched Network. VPLS emulates the Ethernet VLAN functionality of traditional Ethernet network.

VPLS is a current IETF standard, please refer to [RFC4761] [RFC4762].

Data frame is Ethernet frame.

Data forwarding is MAC-based forwarding.

VPLS can be used to emulate E-LAN service over MPLS network provided that the E-LAN service uses MAC-based forwarding as service frame delivery attribute. Considerable number of service providers have adopted this approach to provide E-LAN services to customers.

1.4.2. Virtual Private Multicast Service (VPMS)

VPMS is a L2VPN service that provides point-to-multipoint connectivity across a variety of link layers, including Frame Relay, ATM, Ethernet, PPP, etc., across an IP or MPLS-enabled IP Packet Switched Network.

In the Ethernet use case, VPMS provides single coverage of receiver membership, i.e. there is no distinct differentiation for multiple multicast groups. Destination address in Ethernet frame is not used in data forwarding.

VPMS MUST support unidirectional point-to-multipoint traffic from a sender to multiple receivers and MAY support reverse traffic in a point-to-point manner.

VPMS is currently under development. Please refer to [Draft VPMS Frmwk].
1.5. Terminology

**E-Tree**

An Ethernet VPN in which each Root AC can communicate with every other AC, whereas Leaf ACs can only communicate with Root ACs. Each AC on an E-Tree construct is designated as either a Root AC or a Leaf AC. There can be multiple Root ACs and Leaf ACs per E-Tree construct.

**Root AC**

An ingress frame at a Root AC can be delivered to one or more of any of the other ACs in the E-Tree. Please note that this AC is bidirectional.

**Leaf AC**

Ingress frame at a Leaf AC can only be delivered to one or more Root ACs in the E-Tree. Ingress frame at a Leaf AC MUST NOT be delivered to any Leaf ACs in the E-Tree. Please note that this AC is bidirectional.
2. Reference Model

Figure 1 below describes a generic reference model where PE1, PE2 and PE3 need to establish an E-Tree construct between different Ethernet endpoints. Each PE has 2 Root ACs and 2 Leaf ACs connected to a VSI. These VSIs are then linked together via Ethernet PWs.

In most use cases, an E-Tree construct has only a few Root ACs but many Leaf ACs. There may be only Root ACs or only Leaf ACs on a PE.

Figure 1: E-Tree Reference Model
With an E-Tree construct:

- A Root AC can receive from and transmit to any other ACs.
- A Leaf AC can receive from and transmit to any Root ACs.
- A Leaf AC cannot receive from and transmit to any other Leaf ACs.

This applies to all traffic, including Unicast Known, Unicast Unknown, Broadcast and Multicast.

When an Ethernet Frame is received on PE1 via AC1, the frame can be transmitted to any other local ACs on PE1 and via Ethernet PWs to any remote ACs on PE2 and PE3.

However when an Ethernet frame is received on PE1 via AC3, the frame can be transmitted to any other local Root ACs on PE1 and via Ethernet PWs to any remote Root ACs on PE2 and PE3, but the frame cannot be transmitted to any local Leaf ACs on PE1 nor any remote Leaf ACs on PE2 and PE3.
3. Use Cases

Table 1 below presents some major use cases.

| Use Case                  | Root              | Leaf            |
|---------------------------+-------------------+-----------------|
| 1 | Broadcast Video         | Video Source     | Subscriber      |
|   | (unidirectional only)   |                  |                 |
| 2 | Broadcast/Multicast Video plus Control Channel | Video Source | Subscriber |
| 3 | Internet Access         | BNG Router       | Subscriber      |
| 4 | IEEE 1588 PTPv2 Clock Synchronisation | PTP Server | PTP Client |
| 5 | Mobile Backhaul         | RAN NC           | RAN BS          |
| 6 | Hub & Spoke VPN         | Hub Site         | Spoke Site      |
| 7 | Wholesale Access        | Customer’s Interconnect | Customer’s Subscriber |
| 8 | Device Management       | Management System | Managed Device  |

Table 1: E-Tree Use cases

Common to all use cases, direct Leaf-to-Leaf communication is not required. For Mobile backhaul, this may not be valid for LTE X2 interfaces in the future.

If direct Leaf-to-Leaf communication is not allowed due to security concern, then E-Tree should be used to prohibit communication between Leaf endpoints, otherwise E-LAN is also a feasible option.

Also common to the use cases mentioned above, there may be single or multiple Root endpoints in one E-Tree service. The need for multiple Root endpoints is usually driven by redundancy requirement. Whether a particular E-Tree service needs to support single or multiple Root endpoints depends on the target application.
A generic E-Tree service supports the following traffic flows:
- Unicast bidirectional Root to/from Root
- Unicast bidirectional Root to/from Leaf
- Broadcast/Multicast unidirectional Root to all Roots and Leafs
- Broadcast/Multicast unidirectional Leaf to all Roots

A particular E-Tree service may need to support all the above or only a subset depending on the target application.

Among the use cases mentioned above, broadcast video draws most attention. Actually, broadcast video is a representing example for content delivery in general, such as news feed, financial data feed, etc.
4. Challenges

4.1. Generic E-Tree Service Definition

This section highlights why the current standard VPLS is insufficient for emulating E-Tree service over MPLS network.

4.1.1. Leaf-to-Leaf Communication Restriction

Current standard VPLS treats all ACs equal (i.e. not classified into Root or Leaf) and provides any-to-any connectivity among all ACs. The current standard VPLS does not include any mechanism of communication restriction between specific ACs, therefore is insufficient for emulating generic E-Tree service over MPLS network.

In order to fulfil the generic E-Tree service definition, extensions to the current VPLS standard and related PWE3 standard are required. Such extensions should have minimal impact on the emulated E-LAN services already in operation.

4.2. Use Case Desirable Requirements

There are quite a variety of use cases for E-Tree. For some use cases, the generic MEF E-Tree service definition is good enough. For some other use cases, there are desirable requirements beyond that.

The challenges discussed in this section are not related to the generic E-Tree service definition but the desirable requirements of specific use cases.

4.2.1. Point-to-Multipoint Bandwidth Optimisation

The current standard VPLS uses point-to-point PW between PEs. For unicast unknown/broadcast/multicast frame, the ingress PE replicates the frame on every PW towards remote PE belonging to the same VPLS instance. Depending on the mapping between the logical topology of the E-Tree service and the physical topology of the network, multiple PWs may transverse same physical link, result in multiple copies of the same payload frame on the physical link. Such approach is inefficient in terms of bandwidth usage.

For some use cases, for example broadcast/multicast video, due to nature of the application, there is significant volume of point-to-multipoint traffic. Bandwidth optimisation for such traffic within the network becomes a concern from the service provider perspective.
4.2.2. Multicast Optimisation

The current standard VPLS does not maintain information about multicast group membership and treats multicast frame in exactly the same way as broadcast frame. The ingress PE replicates a multicast frame on every PW towards remote PE belonging to the same VPLS instance. The remote PE then forward the frame to every local AC of the same VPLS instance.

A multicast frame will be forwarded to all ACs, including those not a member of the specific multicast group. Unnecessary traffic consumes bandwidth on access link and may become a concern from the customer perspective. In some cases, it may also be a security concern as the multicast frame may be forwarded to an endpoint other than the intended destinations.

A multicast frame will be forwarded to a remote PE with no member of the specific multicast group. Unnecessary traffic consumes bandwidth in the network and may become a concern from the service provider perspective.

For some use cases, for example multicast video, due to nature of the application, there is significant volume of multicast traffic. The above becomes a real concern from both the customer and service provider perspectives.

4.2.3. MAC-based Forwarding Unnecessary

For some use cases, for example broadcast video, due to nature of the application, there is only broadcast unidirectional traffic from Root to all other endpoints. It is unnecessary to use destination address for data forwarding. Deliver unconditionally for ingress frame at Root endpoint may be a simpler approach than MAC-based forwarding.

4.2.4. MAC-based Forwarding Security Concern

MAC-based forwarding will make an unicast frame from a Root destined for a specific Leaf being forwarded to other endpoints in addition to the intended destination when the frame is classified as unicast unknown, may be due to MAC address aged out or MAC address table overflow.

MAC address spoofing may cause an unicast frame from a Root destined for a specific Leaf being forwarded to an endpoint different from the intended destination.

If such unicast frame carries sensitive information strictly for the intended destination only, then the MAC-based forwarding may cause a security concern from the customer perspective.
For some use cases, for example Internet access and wholesale access, this is a valid concern.

There are some possible mitigations:
- For every Leaf endpoint of the particular E-Tree service, deploy a service provider controlled router between the Leaf endpoint and the customer network
- Customer to deploy security mechanism above Ethernet Layer, for example IPsec, SSL, SSH

Whether the MAC-based forwarding really becomes a security concern depends on the particular application and the deployment scenario. This is unlikely to be a critical concern in most cases.
5. MAC-based Forwarding E-Tree

As mentioned in section 1.3.1. E-Tree can use MAC-based forwarding or something else for data forwarding. This section presents a solution framework for MAC-based forwarding E-Tree. Section 6 will discuss other variants.

This is a VPLS-based solution. Functional components of the solution are identified and discussed in the subsections.

5.1. MAC-based Forwarding Any-to-Any Ethernet VPN

This is a MANDATORY component.

This component is the current standard VPLS and PWE3 as specified in [RFC4385] [RFC4447] [RFC4448] [RFC4761] [RFC4762], which provides any-to-any connectivity among all ACs in one VPLS instance.

This is the base component. All other MANDATORY/OPTIONAL components are to be added on top of this component.

5.2. Leaf-to-Leaf Communication Restriction

This is a MANDATORY component.

This component is a minimal extension to the current VPLS and PWE3 standards, with the objective to provide a simple and effective way to support E-Tree services in addition to E-LAN services using VPLS on a MPLS network.

5.2.1. Per-payload Signaling on PW - From Leaf or Root

Let’s look at the scenario illustrated in Figure 2 below. VPLS is used to emulate an E-Tree service over a MPLS network.
When PE2 receives a frame from PE1 via the PW,
- PE2 does not know which AC on PE1 is the ingress AC
- PE2 does not know whether the ingress AC is a Leaf AC or not
- PE2 does not have sufficient information to enforce the Leaf-to-Leaf communication restriction

A simple fix is to carry additional one bit of information (0 or 1) for each payload Ethernet frame on the Ethernet PW
- Indicate whether the frame is from a Leaf AC on ingress PE or not
- Based on this one bit of information, egress PE can decide whether the frame can be forwarded to a local Leaf AC or not

Extension to current PWE3 standard [RFC4448] is proposed. The work in progress [Draft CW L-bit] provides a precise specification on how to use a specific bit within the Control Word, refer as "CW L-bit" to indicate whether the payload Ethernet frame comes from a Root AC or a Leaf AC.

5.2.2. Extension to VPLS

Extension to current VPLS standard [RFC4761] [RFC4762] is proposed. The work in progress [Draft VPLS ETree] provides a precise specification on how to enforce the Leaf-to-Leaf communication restriction locally on a PE.

The [Draft VPLS ETree] introduces the following:
- AC Type, either Root or Leaf
- Use of Control Word to carry the "CW L-bit"
- Additional "Set CW L-bit" action and "Forward or Drop" decision in data forwarding
It is important to note that the "Set CW L-bit" action and "Forward or Drop" decision specified in [Draft VPLS ETree] are in addition to and performed after the following:
- MAC-based forwarding decision as per current standard
- Loop free VPLS "split horizon" rule (MUST NOT forward traffic from one PW to another PW in the same VPLS mesh) as per current standard

5.3. Optional Enhancement - Point-to-Multipoint PW

This is an OPTIONAL component, applicable only when there is significant volume of point-to-multipoint traffic.

Point-to-Multipoint pseudowire (P2MP PW) is a PW attached to a source used to distribute Layer 1 or Layer 2 format traffic to a set of receivers. P2MP PW is unidirectional but optionally bidirectional.

By using P2MP PW, the ingress PE is not responsible for replicating the payload frame on each P2P PW towards egress PE, instead the network elements along the physical path participate in replication. The replication is done by the underlying point-to-multipoint label switched path (P2MP LSP).

Extension to current VPLS standard will be required to specify how P2MP PW and P2P PW should be used and how MAC learning works on P2MP PW.

P2MP PW is currently under development. Please refer to [Draft P2MP PW Req] [Draft P2MP PW Sig].

5.4. Optional Enhancement - Multicast in VPLS

This is an OPTIONAL component, applicable only when there is significant volume of multicast traffic.

In the current standard VPLS, multicast frame is treated in exactly the same way as broadcast frame. Although this is the standard MAC-based forwarding of traditional Ethernet network, it is less than ideal as more IP multicast applications become available.

Multicast in VPLS is currently under development, with the objective to provide efficient ways to support IP multicast services over VPLS. Please refer to [RFC5501] [Draft Mcast VPLS].
6. Non MAC-based Forwarding E-Tree

This section presents some variants of E-Tree services which do not use MAC-based forwarding as the service frame delivery attribute.

6.1. Single Root, Broadcast Only - VPMS

This is in response to the challenge in section 4.2.3. MAC-based Forwarding Unnecessary.

VPMS provides single coverage of receiver membership. Destination address in Ethernet frame is not used in data forwarding.

For E-Tree service of single Root and only unidirectional broadcast traffic from the Root, for example certain broadcast video or similar content delivery applications, VPMS will be a much more simple and effective solution than VPLS.

VPMS is currently under development. Please refer to [Draft VPMS Frmwk].

6.2. Multiple Roots, Broadcast and Unicast

This is in response to the challenge in section 4.2.4. MAC-based Forwarding Security Concern.

This will be added in next version of this document.
7. Security Considerations

This will be added in next version of this document.

8. IANA Considerations

This will be added in next version of this document.

9. Acknowledgements

The authors would like to thank Lizhong Jin, Lucy Yong and Wim Henderickx for their valuable comments.

10. References

10.1. Normative References

[MEF6.1] Metro Ethernet Forum, Ethernet Services Definitions - Phase 2, April 2008
[MEF10.1] Metro Ethernet Forum, Ethernet Services Attributes - Phase 2, November 2006
[RFC2119] Bradner, S., Key words for use in RFCs to Indicate Requirement Levels, BCP 14, RFC 2119, March 1997
10.2. Informative References