Network Working Group   Luca Martini
Internet Draft   Nasser El-Aawar
Expiration Date: January 2003   Level 3 Communications, LLC.

Giles Heron   Steve Vogelsang
PacketExchange Ltd.   Laurel Networks, Inc.

Chris Liljenstolpe   Vasile Radoaca
Cable & Wireless   Nortel Networks

Daniel Tappan   Kireeti Kompella
Eric C. Rosen   Juniper Networks
Cisco Systems, Inc.

Andrew G. Malis   Tricci So
Vinai Sirkay   Chris Flores
Vivace Networks, Inc.   Consultant

XiPeng Xiao   David Zelig
Redback Networks   Corrigent Systems

Raj Sharma   Loa Andersson
Luminous Networks, Inc.   Utfors

Nick Tingle
Sunil Khandekar
TiMetra Networks
July 2002

Encapsulation Methods for Transport of Ethernet Frames Over IP and MPLS Networks

draft-martini-ethernet-encap-mpls-01.txt

Status of this Memo

Abstract

An Ethernet PW allows Ethernet/802.3 Protocol Data Units (PDUs) to be carried over Packet Switched Networks (PSNs) using IP, L2TP or MPLS transport. This enables Service Providers to leverage their existing PSN to offer Ethernet services.

This document describes methods for encapsulating Ethernet/802.3 PDUs for transport over an MPLS or IP network.

Table of Contents

1. Specification of Requirements

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

2. Introduction

In an MPLS or IP network, it is possible to use control protocols such as those specified in [MARTINI-TRANS] to set up "emulated vir¡ tual circuits" that carry the the Protocol Data Units of layer 2 pro¡ tocols across the network. A number of these emulated virtual cir¡ cuits may be carried in a single tunnel. This requires of course that the layer 2 PDUs be encapsulated. We can distinguish three lay¡ ers of this encapsulation:

- the "tunnel header", which contains the information needed to transport the PDU across the IP or MPLS network; this is header belongs to the tunneling protocol, e.g., MPLS, GRE, L2TP.

- the "demultiplexer field", which is used to distinguish individ¡ ual emulated virtual circuits within a single tunnel; this field must be understood by the tunneling protocol as well; it may be, e.g., an MPLS label or a GRE key field.

- the "emulated VC encapsulation", which contains the information about the enclosed layer 2 PDU which is necessary in order to properly emulate the corresponding layer 2 protocol.

This document specifies the emulated Virtual Circuit (VC) encapsula¡ tion for the ethernet protocols. Although different layer 2 protocols require different information to be carried in this encapsulation, an attempt has been made to make the encapsulation as common as possible for all layer 2 protocols. Other layer 2 protocols are described in separate documents. [MARTINI-ATM] [MARTINI-FRAME] [MARTINI-PPP]

This document also specifies the way in which the demultiplexer field is added to the emulated VC encapsulation when an MPLS label is used as the demultiplexer field.

The scope of this document also includes:

- Pseudo-wire (PW) requirements for emulating Ethernet trunking and switching behavior.

- PE-bound and CE-bound packet processing of Ethernet PDUs

- QoS and security considerations

- Inter-domain transport considerations for Ethernet PE

The following two figures describe the reference models which are derived from [PWE3-FRAME] [PWE3-REQ] to support the Ethernet PW emu¡ lated services.

```
         Native    |<----- Pseudo Wire ---->|  Native
        Ethernet   |                        |  Ethernet
           or      |  |<-- PSN Tunnel -->|  |    or
         VLAN      V  V                  V  V   VLAN
        Service  +----+                  +----+ Service
+----+    |      | PE1|==================| PE2|     |    +----+
|    |----------|............PW1.............|----------|    |
| CE1|    |      |    |                  |    |     |    |CE2 |
|    |----------|............PW2.............|----------|    |
+----+    |      |    |==================|    |     |    +----+
     ^           +----+                  +----+     |    ^
     |      Provider Edge 1         Provider Edge 2      |
     |                                                   |
     |<-------------- Emulated Service ---------------->|
```

Figure 1: PWE3 Ethernet/VLAN Interface Reference Configuration

```
 +-------------+                                +-------------+
 |  Emulated   |                                |  Emulated   |
 |  Ethernet   |                                |  Ethernet   |
 | (including  |         Emulated Service       | (including  |
 |  VLAN)      |<==============================>|  VLAN)      |
 |  Services   |                                |  Services   |
 +-------------+           Pseudo Wire          +-------------+
 |Demultiplexer|<==============================>|Demultiplexor|
 +-------------+                                +-------------+
 |    PSN      |            PSN Tunnel          |    PSN      |
 | MPLS or IP  |<==============================>| MPLS or IP  |
 +-------------+                                +-------------+
 |  Physical   |                                |  Physical   |
 +-----+-------+                                +-----+-------+
       |                                              |
       |                MPLS or IP Network            |
       |             _____    ___       _____         |
       |           _/    \___/   \    _/    \__       |
       |          /               \__/         \_     |
       |         /                               \    |
       +========/                                 |===+
                \                                 /
                 \                               /
                  \   ___      ____    __      _/
                   \_/   \____/   \___/  \____/
```

Figure 2: Ethernet PWE3 Protocol Stack Reference Model

For the purpose of this document R1 will be defined as the ingress router, and R2 as the egress router. A layer 2 PDU will be received at R1, encapsulated at R1, transported, decapsulated at R2, and transmitted out of R2.

3. Requirements for Ethernet Pseudo-Wire Emulation

An Ethernet PW emulates a single Ethernet link between exactly two endpoints. The following reference model describes the termination point of each end of the PW within the PE:

```
           +----------------------------------+
           |                PE                |
   +---+   +-+  +-----+  +------+  +------+  +-+
   |   |   |P|  |     |  |PW ter|  | PSN  |  |P|
   |   |<==|h|<=| NSP |<=|minati|<=|Tunnel|<=|h|<== From PSN
   |   |   |y|  |     |  |on    |  |      |  |y|
   | C |   +-+  +-----+  +------+  +------+  +-+
   | E |   |                                  |
   |   |   +-+  +-----+  +------+  +------+  +-+
   |   |   |P|  |     |  |PW ter|  | PSN  |  |P|
   |   |==>|h|=>| NSP |=>|minati|=>|Tunnel|=>|h|==> To PSN
   |   |   |y|  |     |  |on    |  |      |  |y|
   +---+   +-+  +-----+  +------+  +------+  +-+
           |                                  |
           +----------------------------------+
                       ^        ^
                       |        |
                       A        B
```

Figure 3: PW reference diagram

The PW terminates at a logical port within the PE, defined at point A in the above diagram. This port provides an Ethernet MAC service that will deliver each Ethernet packet that is received at point A, unaltered, to the point A in the corresponding PE at the other end of the PW.

The "NSP" function includes packet processing needed to translate the Ethernet packets that arrive at the CE-PE interface to/from the Ethernet packets that are applied to the PW termination point. Such functions may include stripping, overwriting or adding VLAN tags, physical port multi¡ plexing and demultiplexing, PW-PW bridging, L2 encapsulation, shaping, policing, etc.

The points to the left of A, including the physical layer between the CE and PE, and any adaptation (NSP) functions between it and the PW termi¡ nations, are outside of the scope of PWE3 and are not defined here.

"PW Termination", between A and B, represents the operations for setting up and maintaining the PW, and for encapsulating and decapsulating the Ethernet packets according to the PSN type in use. This document defines these operations, and the services offered and required at points A and B.

"PSN Tunnel" denotes the PSN tunneling technology that is being used: MPLS or GRE/IP.

A pseudo wire can be one of the two types: raw or tagged. This is a property of the emulated Ethernet link and indicates whether the pseudo

wire MUST contain an 802.1Q VLAN tag (i.e. tagged mode) or MAY contain a tag (i.e. raw mode).

3.1. Packet Processing

3.1.1. Encapsulation

The entire Ethernet frame without any preamble or FCS is transported as a single packet. A VC label is prepended to this and the packet is forwarded through a PSN tunnel (either MPLS or GRE/IP).

3.1.2. MTU Management

Ingress and egress PWESs MUST agree on their maximum MTU size to be transported over the PSN.

3.1.3. Frame Ordering

In general, applications running over Ethernet do not require strict frame ordering. However the IEEE definition of 802.3 [802.3] requires that frames from the same conversation are delivered in sequence. Moreover, the PSN cannot (in the general case) be assumed to provide or to guarantee frame ordering. Therefore if strict frame ordering is required, the control word defined below MUST be utilized and its sequence number processing enabled.

3.1.4. Frame Error Processing

An encapsulated Ethernet frame traversing a psuedo-wire may be dropped, corrupted or delivered out-of-order. Per [PWE3-REQ], packet-loss, corruption, and out-of-order delivery is considered to be a "generalized bit error" of the psuedo-wire. Therefore, the native Ethernet frame error processing mechanisms MUST be extended to the corresponding psuedo-wire service. Therefore, if a PE device receives an Ethernet frame containing hardware level CRC errors, framing errors, or a runt condition, the frame MUST be discarded on input. Note that this processing is part of the NSP function and is outside the scope of this draft.

3.1.5. IEEE 802.3x Flow Control Interworking

In a standard Ethernet network, the flow control mechanism is optional and typically configured between the two nodes on a point-to-point link (e.g. between the CE and the PE). IEEE 802.3x PAUSE frames MUST NOT be carried across the PW. See Appendix A for notes on CE-PE flow control.

3.2. Maintenance

It is desirable to have a signaling mechanism for establishing Ethernet PWs and for detecting failure of an Ethernet PW. It is recommended that the procedures defined in [MARTINI-TRANS] be used for this purpose.

3.3. Management

The PW management model of Ethernet PW follows the general management guidelines for PW management as appear in [PW-MIB] and defined in [PWE3-REQ], [PWE3-FRAME]. It is composed of 3 components. [PW-MIB] defines the parameters common to all types of PW and PSNs, for example common counters, error handling, some maintenance protocol parameters etc. For each type of PSN there is a separate module that defines the association of the PW to the PSN tunnel, see example in [PW-MPLS-MIB] for the MPLS PSN. For Ethernet PW, an additional MIB module [PW-ENET-MIB] defines the Ethernet specific parameters required to be configured or monitored.

The above modules enable both manual configuration and the use of maintenance procedures to set up the Ethernet PW and monitor PW state where applicable.

As specified in [PWE3-REQ] and [PWE3-FRAME], an implementation SHOULD support the relevant PW MIB modules for PW set-up and monitoring. Other mechanisms for PW set up (command line interface for example) MAY be supported.

3.4. QoS Considerations

The ingress PE MAY consider the user priority (PRI) field [802.1Q] of the VLAN tag header when determining the value to be placed in the Quality of Service field of the encapsulating protocol (e.g., the EXP fields of the MPLS label stack). In a similar way, the egress PE MAY consider the Quality of Service field of the encapsulating protocol when queuing the packet for CE-bound.

A PE MUST support the ability to carry the Ethernet PW as a best effort service over the PSN. Transparency of PRI bits (if sent from CE to PE) between CE devices, regardless of the COS support of the PSN. Where the 802.1Q VLAN field is added at the PE, a default PRI setting of zero MUST be supported, a configured default value is recommended.

A PE may support additional QOS support by means of one or more of the following methods:

-i. One COS per PW End Service (PWES), mapped to a single COS PW at the PSN.
-ii. Multiple COS per PWES mapped to a single PW with multiple COS at the PSN.
-iii. Multiple COS per PWES mapped to multiple PWs at the PSN.

Examples of the cases above and details of the service mapping considerations are described in Appendix B.

The PW guaranteed rate at the PSN level is PW provider policy based on agreement with the customer, and may be different from the Ethernet physical port rate. Consideration of Ethernet flow control was discussed above.

3.5. Security Considerations

This document specifies the security consideration regarding the encapsulation for the PW. In terms of encapsulation, security of the encapsulated packets depends on the nature of the protocol that is carried by these packets, while the encapsulation itself shall not affect the related security issues.

Nevertheless, the security limitations of the PE and/or the PW MUST not restrict the security implementation choices of the user of the PWE3 (i.e. users should be able to implement IPSEC or any other appropriate security mechanism in addition to the security inherent in the PW)".

It is required that PEs will have user separation between different PW and different virtual ports that the PWs are connected to. For example: if two PWs are connected to the same physical port and associated to different virtual ports (i.e. VLANs), it is required that packets from one VC will not be forwarded to the VLAN that is associated to the second VCs.

A received packet is associated with a PW by means of the VC label. However this mechanism provides no guarantee that the packet was sent

by the peer PE. Further checks may be useful to protect against mis-configuration and connection hijacking.

The PE must be able to be protected from malformed, or maliciously altered, customer traffic. This includes, but is not limited to, illegal VLAN use, short packets, long packets, etc.

Security achieved by access control of MAC addresses is out of scope of this document.

Additional security requirements related to the use of PW in a switching (virtual bridging) environment are not discussed here as they are not within the scope of this draft.

In the case of a PW crossing from one autonomous system to another, through a private interconnection, security considerations are much the same as in the intra-domain case. However in some cases the PW may travel through a third-party autonomous system, or across a pub¡ lic interconnection point. In these cases there may be a requirement to encrypt the user data using a method appropriate to the PSN tun¡ neling mechanism.

4. General encapsulation method

4.1. The Control Word

When carrying Ethernet over an IP or MPLS backbone sequentiality may need to be preserved. The OPTIONAL control word defined here addresses this requirement. Implementations MUST support sending no control word, and MAY support sending a control word.

In all cases the egress router must be aware of whether the ingress router will send a control word over a specific virtual circuit. This may be achieved by configuration of the routers, or by signal¡ ing, for example as defined in [MARTINI-TRANS].

The control word is defined as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Reserved            |       Sequence Number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

In the above diagram the first 16 bits are reserved for future use. They MUST be set to 0 when transmitting, and MUST be ignored upon receipt.

The next 16 bits provide a sequence number that can be used to guarantee ordered packet delivery. The processing of the sequence number field is OPTIONAL.

The sequence number space is a 16 bit, unsigned circular space. The sequence number value 0 is used to indicate an unsequenced packet.

4.1.1. Setting the sequence number

For a given emulated VC, and a pair of routers R1 and R2, if R1 sup¡ ports packet sequencing then the following procedures should be used:

- the initial packet transmitted on the emulated VC MUST use sequence number 1
- subsequent packets MUST increment the sequence number by one for each packet
- when the transmit sequence number reaches the maximum 16 bit value (65535) the sequence number MUST wrap to 1

If the transmitting router R1 does not support sequence number pro¡ cessing, then the sequence number field in the control word MUST be set to 0.

4.1.2. Processing the sequence number

If a router R2 supports receive sequence number processing, then the following procedures should be used:

When an emulated VC is initially set up, the "expected sequence num¡ ber" associated with it MUST be initialized to 1.

When a packet is received on that emulated VC, the sequence number should be processed as follows:

- if the sequence number on the packet is 0, then the packet passes the sequence number check

- otherwise if the packet sequence number >= the expected sequence number and the packet sequence number - the expected sequence number < 32768, then the packet is in order.

- otherwise if the packet sequence number < the expected sequence number and the expected sequence number - the packet sequence number >= 32768, then the packet is in order.

- otherwise the packet is out of order.

If a packet passes the sequence number check, or is in order then, it can be delivered immediately. If the packet is in order, then the expected sequence number should be set using the algorithm:

```
expected_sequence_number := packet_sequence_number + 1 mod 2**16
if (expected_sequence_number = 0) then expected_sequence_number := 1;
```

Packets which are received out of order MAY be dropped or reordered at the discretion of the receiver.

If a router R2 does not support receive sequence number processing, then the sequence number field MAY be ignored.

4.2. MTU Requirements

The network MUST be configured with an MTU that is sufficient to transport the largest encapsulation frames. If MPLS is used as the tunneling protocol, for example, this is likely to be 8 or more bytes greater than the largest frame size. Other tunneling protocols may have longer headers and require larger MTUs. If the ingress router determines that an encapsulated layer 2 PDU exceeds the MTU of the tunnel through which it must be sent, the PDU MUST be dropped. If an egress router receives an encapsulated layer 2 PDU whose payload length (i.e., the length of the PDU itself without any of the encap¡ sulation headers), exceeds the MTU of the destination layer 2 inter¡ face, the PDU MUST be dropped.

4.3. Tagged Mode

In this mode each frame MUST include an 802.1Q field. All frames in a PW MUST have the same 802.1Q tag value. Note that the tag may be overwritten by the NSP function at ingress or at egress.

Note that when using the signaling procedures defined in [MARTINI-TRANS], such a PW should be signaled as being of type "Ethernet VLAN".

4.4. Raw Mode

In this mode each frame MAY include an 802.1Q field. Multiple 802.1Q tag values MAY be transported over the same PW.

Note that when using the signaling procedures defined in [MARTINI-TRANS], such a PW should be signaled as being of type "Ethernet".

5. Using an MPLS Label as the Demultiplexer Field

To use an MPLS label as the demultiplexer field, a 32-bit label stack entry [MPLS-LABEL] is simply prepended to the emulated VC encapsula¡ tion, and hence will appear as the bottom label of an MPLS label stack. This label may be called the "VC label". The particular emu¡ lated VC identified by a particular label value must be agreed by the ingress and egress LSRs, either by signaling (e.g, via the methods of [MARTINI-TRANS]) or by configuration. Other fields of the label stack entry are set as follows.

5.1. MPLS Shim EXP Bit Values

If it is desired to carry Quality of Service information, the Quality of Service information SHOULD be represented in the EXP field of the VC label. If more than one MPLS label is imposed by the ingress LSR, the EXP field of any labels higher in the stack SHOULD also carry the same value.

5.2. MPLS Shim S Bit Value

The ingress LSR, R1, MUST set the S bit of the VC label to a value of 1 to denote that the VC label is at the bottom of the stack.

5.3. MPLS Shim TTL Values

The ingress LSR, R1, SHOULD set the TTL field of the VC label to a value of 255.

6. Security Considerations

This document specifies only encapsulations, and not the protocols used to carry the encapsulated packets across the network. Each such protocol may have its own set of security issues, but those issues are not affected by the encapsulations specified herein.

Specific security issues related to encapsulation are addressed in the requirements section above.

7. Intellectual Property Disclaimer

This document is being submitted for use in IETF standards discus¡ sions.

8. References

[MARTINI-TRANS] "Transport of Layer 2 Frames Over MPLS", Martini, L., et al., draft-martini-l2circuit-trans-mpls-09.txt, (work in progress), June 2002.

[MARTINI-ATM] "Encapsulation Methods for Transport of ATM Cells/Frame Over IP and MPLS Networks", Martini L., et al., draft-martini-atm-encap-mpls-00.txt, (work in progress), June 2002.

[MARTINI-FRAME] "Encapsulation Methods for Transport of Frame-Relay Over IP and MPLS Networks", Martini, L., et al., draft-martini-frame-encap-mpls-00.txt, (work in progress), June 2002.

[MARTINI-PPP] "Encapsulation Methods for Transport of PPP/HDLC Frames Over IP and MPLS Networks", Martini L., et al., draft-martini-ppp-hdlc-encap-mpls-00.txt, (work in progress), April 2002.

[PWE3-REQ] "Requirements for Pseudo Wire Emulation Edge-to-Edge (PWE3)", Xiao, X., McPherson, D., Pate, P., White, C., Kompella, K., Gill, V., Nadeau, T., draft-pwe3-requirements-03.txt, (work in progress), June 2002.

[PWE3-FRAME] "Framework for Pseudo Wire Emulation Edge-to-Edge (PWE3)", Pate, P., Xiao, X., So, T., Malis, A., Nadeau, T., White, C., Kompella, K., Johnson, T., Bryant, S., draft-pate-pwe3-framework-03.txt, (work in progress), June 2002.

[PW-MIB] "Pseudo Wire (PW) Management Information Base using SMIv2", Zelig, D., Mantin, S., Nadeau, T., Danenberg, D., draft-zelig-pw-mib-02.txt, (work in progress), February 2002.

[PW-MPLS-MIB] "Pseudo Wire (PW) over MPLS PSN Management Information Base", Zelig D., Mantin, S., Nadeau, T., Danenberg, D., Malis, A., draft-zelig-pw-mpls-mib-01.txt, (work in progress), February 2002.

[PW-ENET-MIB] "Ethernet Pseudo Wire (PW) Management Information Base", Zelig, D., Nadeau, T., draft-zelig-pw-enet-mib-00.txt, (work in progress) February 2002.

[802.3] IEEE, ISO/IEC 8802-3: 2000 (E), "IEEE Standard for Information technology -- Telecommunications and information exchange between systems -- Local and metropolitan area networks -- Specific requirements -- Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications", 2000.

[802.1Q] ANSI/IEEE Standard 802.1Q, "IEEE Standards for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks", 1998.

[MPLS-LABEL] "MPLS Label Stack Encoding", Rosen, E., Rekhter, E., Tappan, D., Fedorkow, G., Farinacci, D., Li, T., Conta, A., RFC 3032.

9. Author Information

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

Nasser El-Aawar
Level 3 Communications, LLC.
1025 Eldorado Blvd.
Broomfield, CO, 80021
e-mail: nna@level3.net

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

Dan Tappan
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA, 01824
e-mail: tappan@cisco.com

Eric Rosen
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA, 01824
e-mail: erosen@cisco.com

Steve Vogelsang
Laurel Networks, Inc.
Omega Corporate Center
1300 Omega Drive
Pittsburgh, PA 15205
e-mail: sjv@laurelnetworks.com

Andrew G. Malis
Vivace Networks, Inc.
2730 Orchard Parkway
San Jose, CA 95134
e-mail: Andy.Malis@vivacenetworks.com

Vinai Sirkay
Vivace Networks, Inc.
2730 Orchard Parkway
San Jose, CA 95134
e-mail: sirkay@technologist.com

Vasile Radoaca
Nortel Networks
600 **Technology Park**
Billerica MA 01821
e-mail: vasile@nortelnetworks.com

Chris Liljenstolpe
Cable & Wireless
11700 **Plaza America Drive**
Reston, VA 20190
e-mail: chris@cw.net

Kireeti Kompella
Juniper Networks
1194 **N. Mathilda Ave**
Sunnyvale, CA 94089
e-mail: kireeti@juniper.net

Tricci So
e-mail: tricciso@yahoo.ca

XiPeng Xiao
Redback Networks
300 **Holger Way,**
San Jose, CA 95134
e-mail: xipeng@redback.com

Chris Flores
Austin, Texas
e-mail: chris_flores@hotmail.com

David Zelig
Corrigent Systems
126, Yigal Alon St.
Tel Aviv, ISRAEL
e-mail: davidz@corrigent.com

Raj Sharma
Luminous Netwokrs, Inc.
10460 **Bubb Road**
Cupertino, CA 95014
e-mail: raj@luminous.com

Nick Tingle
TiMetra Networks
274 **Ferguson Drive**
Mountain View, CA 94043
e-mail: nick@timetra.com

Sunil Khandekar
TiMetra Networks
274 **Ferguson Drive**
Mountain View, CA 94043
email: sunil@timetra.com

Loa Andersson
Utfors
P.O. Box 525,
SE-169 29 Solna, Sweden
e-mail: loa.andersson@utfors.se

Appendix A - Interoperability Guidelines

Configuration Options

The following is a list of the configuration options for a point-to-point Ethernet PW based on the reference points of Figure 3:

Service and Encap on A	Encap on C	Operation at B ingress/egress	Remarks
1) Raw	Raw - Same as A		
2) Tag1	Tag2	Optional change of VLAN value	VLAN can be 0-4095 Change allowed in both directions
3) No Tag	Tag	Add/remove Tag field	Tag can be 0-4095 (note i)
4) Tag	No Tag	Remove/add Tag field	(note ii)

Figure 4: Configuration Options

Allowed combinations:

Raw and other services are not allowed on the same physical port (A). All other combinations are allowed, except that conflicting VLANs on (A) are not allowed.

Notes:

-i. Mode #3 MAY be limited to adding VLAN NULL only, since change of VLAN or association to specific VLAN can be done at the PW CE-bound side.

-ii. Mode #4 exists in layer 2 switches, but is not recommended when operating with PW since it may not preserve the user's PRI bits. If there is a need to remove the VLAN tag (for TLS at the other end of the PW) it is recommended to use mode #2 with tag2=0 (NULL VLAN) on the PW and use mode #3 at the other end of the PW.

IEEE 802.3x Flow Control Considerations

If the receiving node becomes congested, it can send a special frame, called the PAUSE frame, to the source node at the opposite end of the connection. The implementation MUST provide a mechanism for terminat¡ ing PAUSE frames locally (i.e. at the local PE). It MUST operate as follows:

PAUSE frames received on a local Ethernet port SHOULD cause the PE device to buffer, or to discard, further Ethernet frames for that port until the PAUSE condition is cleared. Optionally the PE MAY simply discard PAUSE frames.

If the PE device wishes to pause data received on a local Ethernet port (perhaps because its own buffers are filling up or because it has received notification of congestion within the PSN) then it MAY issue a PAUSE frame on the local Ethernet port, but MUST clear this condition when willing to receive more data.

Appendix B - QoS Details

Section 3.7 describes various modes for supporting PW QOS over the PSN. Examples of the above for a point to point VLAN service are:

- The classification to the PW is based on VLAN field only, regard¡ less of the user PRI bits. The PW is assigned a specific COS (marking, scheduling, etc.) at the tunnel level.

- The classification to the PW is based on VLAN field, but the PRI bits of the user is mapped to different COS marking (and network behavior) at the PW level. Examples are DiffServ coding in case of IP PSN, and E-LSP in MPLS PSN.

- The classification to the PW is based on VLAN field and the PRI bits, and packets with different PRI bits are mapped to different PWs. An example is to map a PWES to different L-LSPs in MPLS PSN in order to support multiple COS service over an L-LSP capable network.

  The specific value to be assigned at the PSN for various COS is not specified and is application specific.

Adaptation of 802.1Q COS to PSN COS

It is not required that the PSN will have the same COS definition of COS as defined in [802.1Q], and the mapping of 802.1Q COS to PSN QOS is application specific and depends on the agreement between the cus¡ tomer and the PW provider. However, the following principles adopted from 802.1Q table 8-2 MUST be met when applying set of PSN COS based on user's PRI bits.

User Priority	#of available classes of service							
	1	2	3	4	5	6	7	8
0 Best Effort (Default)	**0**	**0**	**0**	**1**	**1**	**1**	**1**	**2**
1 Background	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
2 Spare	0	0	0	0	0	0	0	1
3 Excellent Effort	0	0	0	1	1	2	2	3
4 Controlled Load	**0**	**1**	**1**	**2**	**2**	**3**	**3**	**4**
5 Interactive Multimedia	**0**	**1**	**1**	**2**	**3**	**4**	**4**	**5**
6 Interactive Voice	**0**	**1**	**2**	**3**	**4**	**5**	**5**	**6**
7 Network Control	0	1	2	3	4	5	6	7

Figure 5: IEEE 802.1Q COS Service Mapping

Drop precedence

The 802.1P standard does not support drop precedence, therefore from the PW PE-bound point of view there is no mapping required. It is however possible to mark different drop precedence for different PW packets based on the operator policy and required network behavior. This functionality is not discussed further here.

PSN COS labels interaction with VC label COS marking

Marking of COS bits at the VC level is not required if the PSN tunnel is PE to PE based, since only the PSN COS marking is visible to the PSN network. In cases where the VC multiplexing field is carried without an external tunnel (for example directly connected PEs with PHP, or PEs connected using GRE/IP), the rules stated above for tun¡ nel COS marking apply also for the VC level.

In summary, the rules for COS marking shall be as follows:

- If there is only a VC label then, it shall contain the appropri¡ ate CoS value (e.g. MPLS between PEs which are directly adjacent to each other).

- If the VC label and PSN tunnel labels are both being used, then the CoS marking on the PSN header shall be marked with the cor¡ rect CoS value.

- If the PSN marking is stripped at a node before the PE, the PSN marking MUST be copied to the VC label. An example is MPLS PSN with the use of PHP.

PSN QOS support and signaling of QOS is out of scope of this doc¡ ument.