Network Working Group Luca Martini
Internet Draft Eric C. Rosen
Expiration Date: November 2006 Cisco Systems, Inc.

Giles Heron
Andrew G. Malis
Tellabs

May 2006

# Encapsulation Methods for Transport of PPP/HDLC Over MPLS Networks

draft-ietf-pwe3-hdlc-ppp-encap-mpls-09.txt

## Status of this Memo



## Abstract

A Pseudowire (PW) can be used to carry Point to Point Protocol (PPP), or High-Level Data Link Control (HDLC) Protocol Data Units over an Multi Protocol Label Switching (MPLS) network without terminating the PPP/HDLC protocol. This enables service providers to offer "emulated" HDLC, or PPP link services over existing MPLS networks. This document specifies the encapsulation of PPP/HDLC Packet Data Units (PDUs)

within a pseudo wire.

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

A PPP/HDLC Pseudowire (PW) allows PPP/HDLC Protocol Data Units (PDUs) to be carried over an MPLS network. In addressing the issues associated with carrying a PPP/HDLC PDU over an MPLS network, this document assumes that a Pseudowire (PW) has been set up by some means outside the scope of this document. This may be via manual configuration, or using the signaling protocol such as that defined in [CONTROL].

The following figure describes the reference models which are derived from [ARCH] to support the HDLC/PPP PW emulated services. The reader is also asummmed to be familiar with the content of the [ARCH] Document.

```
      |<-------------- Emulated Service ---------------->|
      |                                                  |
      |          |<------- Pseudo Wire ------>|          |
      |          |                            |          |
      |          |    |<-- PSN Tunnel -->|    |          |
      |          V    V                  V    V          |
      V    AC    +----+                  +----+    AC    V
+-----+    |     | PE1|==================| PE2|     |    +-----+
|     |----------|............PW1.............|----------|     |
| CE1 |    |     |    |                  |    |     |    | CE2 |
|     |----------|............PW2.............|----------|     |
+-----+  ^ |     |    |==================|    |     | ^  +-----+
      ^  |       +----+                  +----+     | |  ^
      |  |   Provider Edge 1         Provider Edge 2  |  |
      |  |                                            |  |
Customer |                                            | Customer
Edge 1   |                                            | Edge 2
         |                                            |
         |                                            |
   native HDLC/PPP service                   native HDLC/PPP service
```

Figure 1: PWE3 HDLC/PPP Interface Reference Configuration

This document specifies the emulated PW encapsulation for PPP, and HDLC, however quality of service related issues are not discussed in this document. For the purpose of the discussion in this document PE1 will be defined as the ingress router, and PE2 as the egress router. A layer 2 PDU will be received at PE1, encapsulated at PE1, transported, decapsulated at PE2, and transmitted out on the attachment circuit of PE2.

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          C
```

Figure 2: PW reference diagram

The PW terminates at a logical port within the PE, defined at point B in the above diagram. This port provides an HDLC Native Service Processing function that will deliver each PPP/HDLC packet that is received at point A, unaltered, to the point A in the corresponding PE at the other end of the PW.

The Native Service Processing (NSP) function includes packet processing that is required for the PPP/HDLC packets that are forwarded to the PW termination point. Such functions may include bit stuffing, PW-PW bridging, L2 encapsulation, shaping, policing, etc. These functions are specific to the native packet technology , and may not be required for the PW emulation service.

The points to the left of B, including the physical layer between the CE and PE, and any adaptation (NSP) functions between it and the PW terminations, 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 PPP/HDLC packets as necessary to transmit them across the MPLS network.

## 3. Applicability Statement

PPP/HDLC transport over PW service is not intended to perfectly emulate the traditional PPP or HDLC service, but it can be used for some applications that require PPP or HDLC transport service.

The applicability statements in [FRAME] also apply to the Frame Relay port mode PW described in this document.

The following are notable differences between traditional PPP/HDLC service, and the protocol described in this document:

- Packet ordering can be preserved using the OPTIONAL sequence field in the control word, however implementations are not required to support this feature.

- The Quality of Service model for traditional PPP/HDLC links can be emulated, however this is outside the scope of this document.

- A Frame Relay Port mode PW, or HDLC PW, does not process any packet relay status messages or alarms as described in [Q922] [Q933]

- The HDLC Flags are processed locally in the PE connected to the attachment circuit.

The HDLC mode is suitable for port to port transport of Frame Relay UNI or NNI traffic. Since all packets are passed in a largely transparent manner over the HDLC PW, any protocol which has HDLC-like framing may utilize the HDLC PW mode, including PPP, Frame-Relay, X.25, etc. Exceptions include cases where direct access to the HDLC interface is required, or modes which operate on the flags, Frame Check Sequence (FCS) , or bit/byte unstuffing that is performed before sending the HDLC PDU over the PW. An example of this is PPP Asynchronous-Control-Character-Map (ACCM) negotiation.

For PPP since media-specific framing is not carried the following options will not operate correctly if the PPP peers attempt to negotiate them:

- Frame Check Sequence (FCS) Alternatives
- Address-and-Control-Field-Compression (ACFC)
- Asynchronous-Control-Character-Map (ACCM)

Note also that PW LSP Interface MTU negotiation as specified in [CONTROL] is not affected by PPP MRU advertisement. Thus if a PPP peer sends a PDU with a length in excess of that negotiated for the PW tunnel that PDU will be discarded by the ingress router.

## 4. General encapsulation method

This section describes the general encapsulation format for PPP and HDLC packets over MPLS pseudo wires.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               PSN Transport Header (As Required)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Pseudo Wire Header                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Control Word                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     PPP/HDLC Service Payload                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: General format for PPP/HDLC encapsulation over PSNs

The PSN Transport Header depends on the particular tunneling technology in use. This header is used to transport the encapsulated PPP/HDLC information through the packet switched core.

The Pseudo Wire Header identifies a particular PPP/HDLC service on a tunnel. In case of MPLS the Pseudo Wire Header is the MPLS label at the bottom of the MPLS label stack.

The Control Word is inserted before the PPP/HDLC service payload. It may contain a length and sequence number.

### 4.1. The Control Word

There are four requirements that may need to be satisfied when transporting layer 2 protocols over an MPLS PSN:

- -i. Sequentiality may need to be preserved.
- -ii. Small packets may need to be padded in order to be transmitted on a medium where the minimum transport unit is larger than the actual packet size.
- -iii. Control bits carried in the header of the layer 2 packet may need to be transported.
- -iv. Creating an in-band associated channel for operation and maintenance communications.

The Control Word defined in this section is based on the Generic PW MPLS Control Word as defined in [CW]. It provides the ability to sequence individual packets on the PW, avoidance of equal-cost

multiple-path load-balancing (ECMP) [RFC2992], and enables OAM mechanisms including [VCCV].

[CW] states, "If a PW is sensitive to packet mis-ordering and is being carried over an MPLS PSN that uses the contents of the MPLS payload to select the ECMP path, it MUST employ a mechanism which prevents packet mis-ordering." This is necessary due to the fact that ECMP implementations may examine the first nibble after the MPLS label stack to determine whether the labeled packet is IP or not. Thus, if the PPP protocol number of an PPP packet carried over the PW without a control word present begins with 0x4 or 0x6, it could be mistaken for an IPv4 or IPv6 packet. This could, depending on the configuration and topology of the MPLS network, lead to a situation where all packets for a given PW do not follow the same path. This may increase out-of-order packets on a given PW, or cause OAM packets to follow a different path than actual traffic.

The features that the control word provides may not be needed for a given PPP/HDLC PW. For example, ECMP may not be present or active on a given MPLS network, strict packet sequencing may not be required, etc. If this is the case, the control word provides little value and is therefore optional. Early PPP/HDLC PW implementations have been deployed that do not include a control word or the ability to process one if present. To aid in backwards compatibility, future implementations MUST be able to send and receive packets without the control word present.

In all cases the egress PE MUST be aware of whether the ingress PE will send a control word over a specific PW. This may be achieved by configuration of the PEs, or by signaling, as defined in [CONTROL].

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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 0 0 0|0 0 0 0|Res|   Length  |     Sequence Number           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: MPLS PWE3 Control Word

In the above diagram the first 4 bits are set to 0 in indicate a CW [CW].

The next 4 bits provide space for carrying protocol specific flags. These are not used for HDLC/PPP and they MUST be set to 0 when transmitting, and MUST be ignored upon receipt.

The next 2 bits are reserved for future use, and MUST be set to 0 on transmission and MUST be ignored on reception.

The next 6 bits provide a length field, which is used as follows: If the packet's length (defined as the length of the layer 2 payload plus the length of the control word) is less than 64 bytes, the length field MUST be set to the packet's length. Otherwise the length field MUST be set to zero. The value of the length field, if not zero, is used to remove any padding that may have been added by the MPLS network. If the control word is used, and padding was added to the packet while transiting the MPLS network, then when the packet reaches the egress PE the padding MUST be removed before forwarding the packet.

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.[CW]

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

The procedures described in section 4 of [CW] MUST be followed to process the sequence number field.

## 4.2. MTU Requirements

The network MUST be configured with an MTU that is sufficient to transport the largest encapsulation packets. When MPLS is used as the tunneling protocol, for example, this is likely to be 12 or more bytes greater than the largest packet size. The methodology described in [FRAG] MAY be used to fragment encapsulated packets that exceed the PSN MTU. However if [FRAG] is not used then if the ingress router determines that an encapsulated layer 2 PDU exceeds the MTU of the PSN tunnel through which it must be sent, the PDU MUST be dropped.

If a packet is received on the attachment circuit that exceeds the interface MTU subTLV value [CONTROL], it MUST be dropped. It is also recommended that PPP devices MUST NOT negotiate PPP MRUs larger than that of the AC MTU.

## 5. Protocol-Specific Details

### 5.1. HDLC

HDLC mode provides port to port transport of HDLC encapsulated traffic. The HDLC PDU is transported in its entirety, including the HDLC address, control and protocol fields, but excluding HDLC flags and the FCS. Bit/Byte stuffing is undone. The control word is OPTIONAL. If the control word is used then the flag bits in the control word are not used, and MUST be set to 0 when transmitting, and MUST be ignored upon receipt.

When the PE detects a status change in the attachment circuit status, such as an attachment circuit physical link failure, or the AC is administratively disabled, the PE MUST send the appropriate PW status notification message that corresponds to the HDLC AC status. In a similar manner, the local PW status MUST also be reflected in a respective PW status notification message as described in [CONTROL].

The PW of type 0x0006 "HDLC" will be used to transport HDLC packets. The IANA allocation registry of "Pseudowire Type" is defined in the IANA allocation document for PWs [BCP116] along with initial allocated values.

### 5.2. Frame Relay Port Mode

Figure 5 illustrates the concept of frame relay port mode or many-to-one mapping which is an OPTIONAL capability.

Figure 5a shows two frame relay devices physically connected with a frame relay UNI or NNI. Between their two ports P1 and P2, n frame relay VCs are configured.

Figure 5b shows the replacement of the physical frame relay interface with a pair of PEs and a PW between them. The interface between a FR device and a PE is either a FR UNI or NNI. The set of n FR VCs between the two FR ports P1 and P2 which are controlled by the same signaling channel using DLCI=0, are mapped into one PW. The standard frame relay Link Management Interface (LMI) procedures happen directly between the CEs. Hence with port mode we have many-to-one mapping between FR VCs and a PW.

```
              +------+                          +--------+
              | FR   |                          |   FR  |
              |device|         FR UNI/NNI       | device|
              |    [P1]------------------------[P2]     |
              |      |      carrying n FR VCs   |       |
              +------+                          +--------+

                 [Pn]: A port
```

Figure 5a: FR interface between two FR devices

```
                   |<---------------------------->|
                    |                              |
                     +----+                  +----+
   +------+          |    |     One PW       |    |         +------+
   |      |          |    |==================|    |         |      |
   |  FR  |    FR    | PE1| carrying n FR VCs| PE2|    FR   |  FR  |
   |device|----------|    |                  |    |---------|device|
   | CE1  | UNI/NNI  |    |                  |    | UNI/NNI | CE2  |
   +------+          +----+                  +----+         +------+
          |                                                 |
          |<----------------------------------------------->|
                                n FR VCs
```

Figure 5b: Pseudo-wires replacing the FR interface

FR VCs are not visible individually to a PE; there is no configuration of individual FR VC in a PE. A PE processes the set of FR VCs assigned to a port as an aggregate.

FR port mode provides transport between two PEs of a complete FR frame using the same encapsulation as described above for HDLC mode.

Although frame relay port mode shares the same encapsulation as HDLC mode, a different PW type is allocated in [BCP116]: 0x000F  Frame-Relay Port mode.

All other aspects of this PW type are identical to the HDLC PW encapsulation described above.

### 5.3. PPP

PPP mode provides point to point transport of PPP encapsulated traffic, as specified in [PPP]. The PPP PDU is transported in its entirety, including the protocol field (whether compressed using Protocol Field Compression or not), but excluding any media-specific framing information, such as HDLC address and control fields or FCS.

If the OPTIONAL control word is used then the flag bits in the control word are not used, and MUST be set to 0 when transmitting, and MUST be ignored upon receipt.

When the PE detects a status change in the attachment circuit (AC) status, such as an attachment circuit physical link failure, or the AC is administratively disabled, the PE MUST send the appropriate PW status notification message that corresponds to the PPP AC status. It should be noted that PPP negotiation status is transparent to the PW, and MUST NOT be communicated to the remote MPLS PE. In a similar manner, the local PW status MUST also be reflected in a respective PW status notification message as described in [CONTROL].

A PW of type 0x0007 "PPP" will be used to transport PPP packets.

The IANA allocation registry of "Pseudowire Type" is defined in the IANA allocation document for PWs [BCP116] along with initial allocated values.

## 6. Using an MPLS Label as the Demultiplexer Field

To use an MPLS label as the demultiplexer field, a 32-bit label stack entry [MPLSENCAP] is simply prepended to the emulated PW encapsulation, and hence will appear as the bottom label of an MPLS label stack. This label may be called the "PW label". The particular emulated PW identified by a particular label value must be agreed by the ingress and egress LSRs, either by signaling (e.g, via the methods of [CONTROL]) or by configuration. Other fields of the label stack entry are set as described below.

### 6.1. 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 PW label.  If more than one MPLS label is imposed by the ingress LSR, the EXP field of any labels higher in the stack MUST also carry the same value.

### 6.2. MPLS Shim S Bit Value

The ingress LSR, PE1, MUST set the S bit of the PW label to a value of 1 to denote that the PW label is at the bottom of the stack.

## 7. Congestion Control

As explained in [ARCH], the PSN carrying the PW may be subject to congestion, with congestion characteristics depending on PSN type, network architecture, configuration, and loading. During congestion the PSN may exhibit packet loss that will impact the service carried by the PPP/HLDC PW. In addition, since PPP/HDLC PWs carry an unspecified type of services across the PSN, they cannot behave in a TCP-friendly manner prescribed by [RFC2914]. In the presence of services that reduce transmission rate, PPP/HDLC PWs will thus consume more than their fair share and SHOULD be halted.

Whenever possible, PPP/HDLC PWs should be run over traffic-engineered PSNs providing bandwidth allocation and admission control mechanisms. IntServ-enabled domains providing the Guaranteed Service (GS) or DiffServ-enabled domains using EF (expedited forwarding) are examples of traffic-engineered PSNs. Such PSNs will minimize loss and delay while providing some degree of isolation of the PPP/HDLC PW's effects from neighboring streams.

The PEs SHOULD monitor for congestion (by using explicit congestion notification, [VCCV], or by measuring packet loss) in order to ensure that the service using the PPP/HDLC PW may be maintained. When significant congestion is detected the PPP/HDLC PW SHOULD be administratively disabled. If the PW has been set up using the protocol defined in [CONTROL], then procedures specified in [CONTROL] for status notification can be used to disable packet transmission on the ingress PE from the egress PE. The PW may be restarted by manual intervention, or by automatic means after an appropriate waiting time.

## 8. IANA Considerations

This document has no new IANA Actions. All necessary IANA actions have already been included in [BCP116].

## 9. Security Considerations

The PPP and HDLC pseudowire type is subject to all of the general security considerations discussed in [ARCH][CONTROL]. This document specifies only encapsulations, and not the protocols that may be used to carry the encapsulated packets across the MPLS network. Each such protocol may have its own set of security issues, but those issues are not affected by the encapsulations specified herein.

## 10. Intellectual Property Statement



## 11. Full Copyright Statement

## 12. Normative References

[CONTROL] "Pseudowire Setup and Maintenance using LDP", Martini, L., et al., draft-ietf-pwe3-control-protocol-16.txt, ( work in progress ), April 2005

[MPLSENCAP] "MPLS Label Stack Encoding", E. Rosen, Y. Rekhter, D. Tappan, G. Fedorkow, D. Farinacci, T. Li, A. Conta. RFC3032

[BCP116] Martini L., "IANA Allocations for pseudo Wire Edge to Edge Emulation (PWE3)", RFC4446, BCP116, April 2006

[CW] "PWE3 Control Word for use over an MPLS PSN", S. Bryant, G. Swallow, D. McPherson, draft-ietf-pwe3-cw-06.txt, ( work in progress ), October 2005.

[PPP] "The Point-to-Point Protocol (PPP)", RFC 1661.

[FRAME] "Frame Relay over Pseudo-Wires", draft-ietf-pwe3-frame-relay-06.txt. December 2005, (work in progress )

## 13. Informative References

[ARCH] "PWE3 Architecture" Bryant, et al.,RFC3985

[FRAG] "PWE3 Fragmentation and Reassembly", A. Malis,W. M. Townsley, draft-ietf-pwe3-fragmentation-08.txt ( work in progress ) February 2005

[VCCV] Nadeau, T., et al."Pseudo Wire Virtual Circuit Connection Verification (VCCV)", Internet Draft draft-ietf-pwe3-vccv-08.txt, October 2005. (work in progress)

[RFC2992] RFC-2992:  Analysis of an Equal-Cost Multi-Path Algorithm, C. Hopps, November 2000

[RFC2914]  S. Floyd, "Congestion Control Principles" RFC 2914

[Q922] ITU-T Recommendation Q.922 Specification for Frame Mode Basic call control, ITU Geneva 1995

[Q933] ITU-T Recommendation Q.933 Specification for Frame Mode
Basic call control, ITU Geneva 2003

## 14. Author Information

Luca Martini
Cisco Systems, Inc.
9155 East Nichols Avenue, Suite 400
Englewood, CO, 80112
e-mail: lmartini@cisco.com

Giles Heron
Tellabs
Abbey Place
24-28 Easton Street
High Wycombe
Bucks
HP11 1NT
UK
e-mail: giles.heron@tellabs.com

Eric C. Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA 01719
E-mail: erosen@cisco.com

Andrew G. Malis
Tellabs
90 Rio Robles Dr.
San Jose, CA 95134
e-mail: Andy.Malis@tellabs.com

## 15. Contributing Author Information

Yeongil Seo
463-1 KT Technology Lab
Jeonmin-dong Yusung-gu
Daegeon, Korea
email: syi1@kt.co.kr

Toby Smith
Laurel Networks, Inc.
Omega Corporate Center
1300 Omega Drive
Pittsburgh, PA 15205
e-mail: tob@laurelnetworks.com