

Network Working Group  
Internet Draft  
Expiration Date: December 2003

Luca Martini  
Nasser El-Aawar  
Level 3 Communications, LLC.

Toby Smith  
Laurel Networks, Inc.  
Giles Heron  
PacketExchange Ltd.

Eric C. Rosen  
Cisco Systems, Inc.

June 2003

## Pseudowire Setup and Maintenance using LDP

[draft-ietf-pwe3-control-protocol-03.txt](#)

### Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#).

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 <http://www.ietf.org/shadow.html>.

### Abstract

Layer 2 services (such as Frame Relay, ATM, ethernet) can be "emulated" over an IP and/or MPLS backbone by encapsulating the layer 2 PDUs and then transmitting them over "pseudowires". It is also possible to use pseudowires to provide SONET circuit emulation over an IP and/or MPLS network. This document specifies a protocol for establishing and maintaining the pseudowires, using extensions to LDP. Procedures for encapsulating layer 2 PDUs are specified in a set of companion documents.

Martini, et al.

[Page 1]

## Table of Contents

<a href="#">1</a>	Specification of Requirements .....	<a href="#">3</a>
<a href="#">2</a>	Introduction .....	<a href="#">3</a>
<a href="#">3</a>	The Pseudowire Label .....	<a href="#">5</a>
<a href="#">4</a>	Details Specific to Particular Emulated Services .....	<a href="#">6</a>
<a href="#">4.1</a>	Frame Relay .....	<a href="#">6</a>
<a href="#">4.2</a>	ATM .....	<a href="#">6</a>
<a href="#">4.2.1</a>	ATM AAL5 SDU VCC Transport .....	<a href="#">6</a>
<a href="#">4.2.2</a>	ATM Transparent Cell Transport .....	<a href="#">7</a>
<a href="#">4.2.3</a>	ATM n-to-one VCC and VPC Cell Transport .....	<a href="#">7</a>
<a href="#">4.2.4</a>	OAM Cell Support .....	<a href="#">7</a>
<a href="#">4.2.5</a>	ILMI Support .....	<a href="#">8</a>
<a href="#">4.2.6</a>	ATM AAL5 PDU VCC Transport .....	<a href="#">9</a>
<a href="#">4.2.7</a>	ATM one-to-one VCC and VPC Cell Transport .....	<a href="#">9</a>
<a href="#">4.3</a>	Ethernet VLAN .....	<a href="#">9</a>
<a href="#">4.4</a>	Ethernet .....	<a href="#">9</a>
<a href="#">4.5</a>	HDLC and PPP .....	<a href="#">10</a>
<a href="#">4.6</a>	IP Layer2 Transport .....	<a href="#">10</a>
<a href="#">5</a>	LDP .....	<a href="#">10</a>
<a href="#">5.1</a>	The PwId FEC Element .....	<a href="#">11</a>
<a href="#">5.2</a>	The Generalized ID FEC Element .....	<a href="#">12</a>
<a href="#">5.2.1</a>	Attachment Identifiers .....	<a href="#">13</a>
<a href="#">5.2.2</a>	Encoding the Generalized ID FEC Element .....	<a href="#">14</a>
<a href="#">5.2.3</a>	Procedures .....	<a href="#">15</a>
<a href="#">5.3</a>	Signaling of Pseudo Wire Status .....	<a href="#">16</a>
<a href="#">5.3.1</a>	Use of Label Mappings. ....	<a href="#">16</a>
<a href="#">5.3.2</a>	Signaling PW status. ....	<a href="#">16</a>
<a href="#">5.4</a>	Interface Parameters Field .....	<a href="#">17</a>
<a href="#">5.4.1</a>	PW types for which the control word is REQUIRED .....	<a href="#">19</a>
<a href="#">5.4.2</a>	PW types for which the control word is NOT mandatory ...	<a href="#">20</a>
<a href="#">5.4.3</a>	Status codes .....	<a href="#">21</a>
<a href="#">5.5</a>	LDP label Withdrawal procedures .....	<a href="#">21</a>
<a href="#">5.6</a>	Sequencing Considerations .....	<a href="#">22</a>
<a href="#">5.6.1</a>	Label Mapping Advertisements .....	<a href="#">22</a>
<a href="#">5.6.2</a>	Label Mapping Release .....	<a href="#">23</a>
<a href="#">6</a>	Security Considerations .....	<a href="#">23</a>
<a href="#">7</a>	References .....	<a href="#">23</a>
<a href="#">8</a>	Author Information .....	<a href="#">24</a>
<a href="#">9</a>	Additional Contributing Authors .....	<a href="#">25</a>
<a href="#">10</a>	<a href="#">Appendix A</a> - C-bit Handling Procedures Diagram .....	<a href="#">28</a>

## 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 [\[7\]](#), [\[10\]](#), and [\[12\]](#) it is explained how to encapsulate a layer 2 Protocol Data Unit (PDU) for transmission over an IP and/or MPLS network. Those specifications that a "pseudowire header", consisting of a demultiplexor field, will be prepended to the encapsulated PDU. The pseudowire demultiplexor field is put on before transmitting a packet on a pseudowire. When the packet arrives at the remote endpoint of the pseudowire, the demultiplexor is what enables the receiver to identify the particular pseudowire on which the packet has arrived. To actually transmit the packet from one pseudowire endpoint to another, the packet may need to travel through a "PSN tunnel"; this will require an additional header to be prepended to the packet.

An accompanying document [\[8\]](#) also describes a method for transporting time division multiplexed (TDM) digital signals (TDM circuit emulation) over a packet-oriented MPLS network. The transmission system for circuit-oriented TDM signals is the Synchronous Optical Network (SONET)[\[5\]](#)/Synchronous Digital Hierarchy (SDH) [\[6\]](#). To support TDM traffic, which includes voice, data, and private leased line service, the pseudowires must emulate the circuit characteristics of SONET/SDH payloads. The TDM signals and payloads are encapsulated for transmission over pseudowires. To this encapsulation is prepended a pseudowire demultiplexor and a PSN tunnel header.

In this document, we specify the use of the MPLS Label Distribution Protocol, LDP [\[RFC3036\]](#), as a protocol a protocol for setting up and maintaining the pseudowires. In particular, we define new TLVs for LDP, which enable LDP to identify pseudowires and to signal attributes of pseudowires. We specify how a pseudowire endpoint uses these TLVs in LDP to bind a demultiplexor field value to a pseudowire, and how it informs the remote endpoint of the binding. We also specify procedures for reporting pseudowire status changes, passing additional information about the pseudowire as needed, and for releasing the bindings.

In the protocol specified herein, the pseudowire demultiplexor field is an MPLS label. Thus the packets which are transmitted from one end of the pseudowire to the other are MPLS packets. Unless the pseudowire endpoints are immediately adjacent, these MPLS packets

must be transmitted through a PSN tunnel. Any sort of PSN tunnel can be used, as long as it is possible to transmit MPLS packets through it. The PSN tunnel can itself be an LSP, but it could equally well be an IP tunnel, a GRE tunnel, an IPsec tunnel, or any other sort of tunnel which can carry MPLS packets. Procedures for setting up and maintaining the PSN tunnels are outside the scope of this document.

This document deals only with the setup and maintenance of point-to-point pseudowires. Neither point-to-multipoint nor multipoint-to-point pseudowires are discussed.

QoS related issues are not discussed in this document.

The following two figures describe the reference models which are derived from [13] to support the Ethernet PW emulated services.

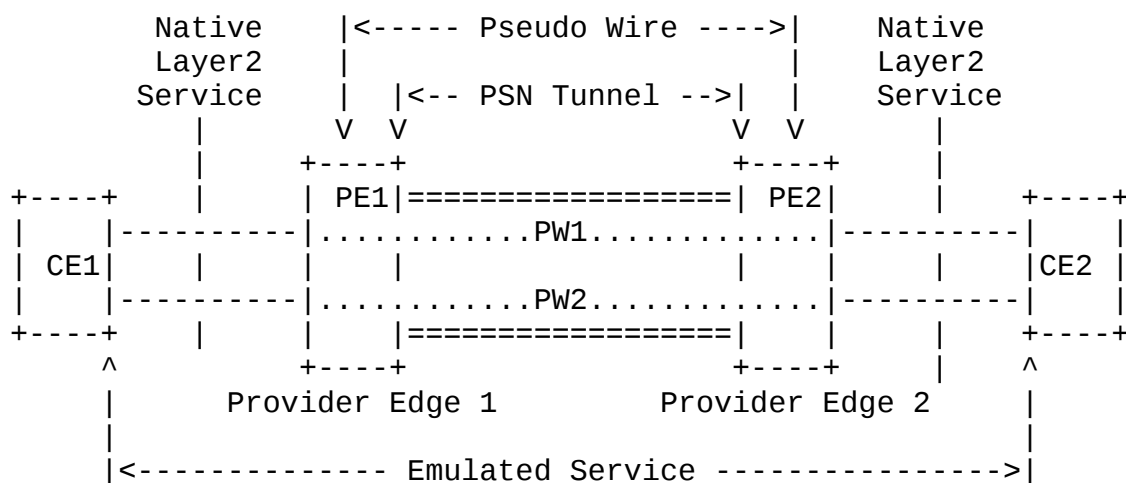


Figure 1: PWE3 Reference Model

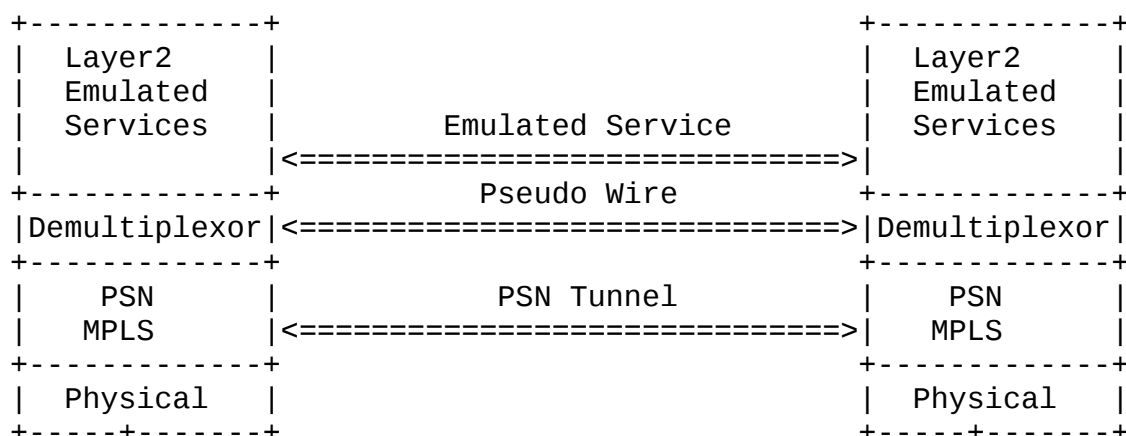


Figure 2: PWE3 Protocol Stack Reference Model

For the purpose of 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 of PE2.

### 3. The Pseudowire Label

Suppose it is desired to transport layer 2 PDUs from ingress LSR PE1 to egress LSR PE2, across an intervening PSN. We assume that there is a PSN tunnel from PE1 to PE2. That is, we assume that PE1 can cause a packet to be delivered to PE2 by encapsulating the packet in a "PSN tunnel header" and sending the result to one of its adjacencies. If the PSN tunnel is an MPLS Label Switched Path (LSP), then putting on a PSN tunnel encapsulation is a matter of pushing on an additional MPLS label; if the PSN tunnel is, e.g., a GRE tunnel, then putting on the tunnel encapsulation requires prepending an IP header and a GRE header.

We presuppose that an arbitrary number of pseudowires can be carried through a single PSN tunnel. Thus it is never necessary to maintain state in the network core for individual pseudowires. We do not presuppose that the PSN tunnels are point-to-point; although the pseudowires are point-to-point, the PSN tunnels may be multipoint-to-point. We do not presuppose that PE2 will even be able to determine the PSN tunnel through which a received packet was transmitted. (E.g., if the PSN tunnel is an LSP, and penultimate hop popping is used, when the packet arrives at PE2 it will contain no information identifying the tunnel.)

When PE2 receives a packet over a pseudowire, it must be able to determine that the packet was in fact received over a pseudowire, and it must be able to associate that packet with a particular pseudowire. PE2 is able to do this by examining the MPLS label which serves as the pseudowire demultiplexor field shown in Figure 2. Call this label the "PW label".

So when PE1 sends a layer 2 PDU to PE2, it first pushes a PW label on its label stack, thereby creating an MPLS packet. It then (if PE1 is not adjacent to PE2) encapsulates that MPLS packet in a PSN tunnel header. (If the PSN tunnel is an LSP, this is just a matter of pushing on a second label.) The PW label is not visible again until the MPLS packet reaches PE2. PE2's disposition of the packet is based on the PW label.

Note that the PW label must always be at the bottom of the packet's label stack and labels MUST be allocated from the per-platform label space.

This document specifies a protocol for assigning and distributing the PW label. This protocol is LDP, extended as specified in the remainder of this document. An LDP session must be set up between the pseudowire endpoints. LDP MUST be used in its "downstream unsolicited" mode. LDP's "liberal label retention" mode SHOULD be used.

In addition to the protocol specified herein, static assignment of PW labels MAY be used, and implementations of this protocol SHOULD provide support for static assignment.

This document specifies all the procedures necessary to set up and maintain the pseudowires needed to support "unswitched" point-to-point services, where each endpoint of the pseudowire is provisioned with the identify of the other endpoint. There are also protocol mechanisms specified herein which can be used to support switched services, and which can be used to support other provisioning models. However, the use of the protocol mechanisms to support those other models and services is not described in this document.

## [4. Details Specific to Particular Emulated Services](#)

### [4.1. Frame Relay](#)

When emulating a frame relay service, the Frame Relay PDUs are encapsulated according to the procedures defined in [\[7\]](#). The PE MUST provide Frame Relay PVC status signaling to the Frame Relay network. If the PE detects a service affecting condition for a particular DLCI, as defined in [\[2\]](#) Q.933 Annex A.5 sited in IA FRF1.1, PE MUST communicate to the remote PE the status of the PW corresponds to the frame relay DLCI. The Egress PE SHOULD generate the corresponding errors and alarms as defined in [\[2\]](#) on the egress Frame relay PVC.

### [4.2. ATM](#)

#### [4.2.1. ATM AAL5 SDU VCC Transport](#)

ATM AAL5 CPCS-SDUs are encapsulated according to [\[10\]](#) ATM AAL5 CPCS-SDU mode. This mode allows the transport of ATM AAL5 CPCS-SDUs traveling on a particular ATM PVC across the network to another ATM PVC.

#### [4.2.2. ATM Transparent Cell Transport](#)

This mode is similar to the Ethernet port mode. Every cell that is received at the ingress ATM port on the ingress PE, PE1, is encapsulated according to [\[10\]](#), ATM cell mode n-to-one, and sent across the PW to the egress PE, PE2. This mode allows an ATM port to be connected to only one other ATM port. [\[10\]](#) ATM cell n-to-one mode allows for concatenation ( grouping ) of multiple cells into a single MPLS frame. Concatenation of ATM cells is OPTIONAL for transmission at the ingress PE, PE1. If the Egress PE PE2 supports cell concatenation the ingress PE, PE1, should only concatenate cells up to the "Maximum Number of concatenated ATM cells" parameter received as part of the FEC element.

#### [4.2.3. ATM n-to-one VCC and VPC Cell Transport](#)

This mode is similar to the ATM AAL5 VCC transport except that cells are transported. Every cell that is received on a pre-defined ATM PVC, or ATM PVP, at the ingress ATM port on the ingress PE, PE1, is encapsulated according to [\[10\]](#), ATM n-to-one cell mode, and sent across the LSP to the egress PE PE2. Grouping of ATM cells is OPTIONAL for transmission at the ingress PE, PE1. If the Egress PE PE2 supports cell concatenation the ingress PE, PE1, MUST only concatenate cells up to the "Maximum Number of concatenated ATM cells in a frame" parameter received as part of the FEC element.

#### [4.2.4. OAM Cell Support](#)

OAM cells MAY be transported on the VC LSP. When the PE is operating in AAL5 CPCS-SDU transport mode if it does not support transport of ATM cells, the PE MUST discard incoming MPLS frames on an ATM PW LSP that contain a PW label with the T bit set [\[10\]](#). When operating in AAL5 SDU transport mode an PE that supports transport of OAM cells using the T bit defined in [\[10\]](#), or an PE operating in any of the cell transport modes MUST follow the procedures outlined in [\[9\]](#) [section 8](#) for mode 0 only, in addition to the applicable procedures specified in [\[6\]](#).

##### [4.2.4.1. SDU/PDU OAM Cell Emulation Mode](#)

A PE operating in ATM SDU, or PDU transport mode, that does not support transport of OAM cells across an LSP MAY provide OAM support on ATM PVCs using the following procedures:

- Loopback cells response

If an F5 end-to-end OAM cell is received from a ATM VC, by either PE that is transporting this ATM VC, with a loopback indication value of 1, and the PE has a label mapping for the ATM VC, then the PE MUST decrement the loopback indication value and loop back the cell on the ATM VC. Otherwise the loopback cell MUST be discarded by the PE.

- AIS Alarm.

If an ingress PE, PE1, receives an AIS F4/F5 OAM cell, it MUST notify the remote PE of the failure. The remote PE, PE2, MUST in turn send F5 OAM AIS cells on the respective PVCs. Note that if the PE supports forwarding of OAM cells, then the received OAM AIS alarm cells MUST be forwarded along the PW as well.

- Interface failure.

If the PE detects a physical interface failure, or the interface is administratively disabled, the PE MUST notify the remote PE for all VCs associated with the failure.

- PSN/PW failure detection.

If the PE detects a failure in the PW, by receiving a label withdraw for a specific PW ID, or the targeted LDP session fails, or a PW status TLV notification is received, then a proper AIS F5 OAM cell MUST be generated for all the affected atm PVCs. The AIS OAM alarm will be generated on the ATM output port of the PE that detected the failure.

#### 4.2.5. ILMI Support

An MPLS edge PE MAY provide an ATM ILMI to the ATM edge switch. If an ingress PE receives an ILMI message indicating that the ATM edge switch has deleted a VC, or if the physical interface goes down, it MUST send a PW status notification message for all PWs associated with the failure. When a PW label mapping is withdrawn, or PW status notification message is received the egress PE SHOULD notify its client of this failure by deleting the VC using ILMI.



#### [4.2.6.](#) ATM AAL5 PDU VCC Transport

ATM AAL5 CPCS-PDUs are encapsulated according to [\[10\]](#) ATM AAL5 CPCS-PDU mode. This mode allows the transport of ATM AAL5 CPCS-PDUs traveling on a particular ATM PVC across the network to another ATM PVC. This mode supports fragmentation of the ATM AAL5 CPCS-PDU in order to maintain the position of the OAM cells with respect to the user cells. Fragmentation may also be performed to maintain the size of the packet carrying the AAL5 PDU within the MTU of the link.

#### [4.2.7.](#) ATM one-to-one VCC and VPC Cell Transport

This mode is similar to the ATM AAL5 n-to-one cell transport except an encapsulation method that maps one ATM VCC or one ATM VPC to one Pseudo-Wire is used. Every cell that is received on a pre-defined ATM PVC, or ATM PVP, at the ingress ATM port on the ingress PE, PE1, is encapsulated according to [\[10\]](#), ATM one-to-one cell mode, and sent across the LSP to the egress PE PE2. Grouping of ATM cells is OPTIONAL for transmission at the ingress PE, PE1. If the Egress PE PE2 supports cell concatenation the ingress PE, PE1, MUST only concatenate cells up to the "Maximum Number of concatenated ATM cells in a frame" parameter received as part of the FEC element.

#### [4.3.](#) Ethernet VLAN

The Ethernet frame will be encapsulated according to the procedures in [\[12\]](#) tagged mode. It should be noted that if the VLAN identifier is modified by the egress PE, according to the procedures outlined above, the Ethernet spanning tree protocol might fail to work properly. If the PE detects a failure on the Ethernet physical port, or the port is administratively disabled, it MUST send PW status notification message for all PWs associated with the port. This mode uses service-delimiting tags to map input ethernet frames to respective PWs.

#### [4.4.](#) Ethernet

The Ethernet frame will be encapsulated according to the procedures in [\[12\]](#) "ethernet raw mode". If the PE detects a failure on the Ethernet input port, or the port is administratively disabled, the PE MUST send a corresponding PW status notification message.

#### [4.5.](#) HDLC and PPP

HDLC and PPP frames are encapsulated according to the procedures in [\[11\]](#). If the MPLS edge PE detects that the physical link has failed, or the port is administratively disabled, it MUST send a PW status notification message that corresponds to the HDLC or PPP PW.

#### [4.6.](#) IP Layer2 Transport

This mode switches IP packets into a Pseudo-Wire. the encapsulation used is according to [\[3\]](#). IP interworking is implementation specific, part of the NSP function [\[13\]](#), and is outside the scope of this document.

### [5.](#) LDP

The PW label bindings are distributed using the LDP downstream unsolicited mode described in [\[1\]](#). The PEs will establish an LDP session using the Extended Discovery mechanism described in [\[1, section 2.4.2 and 2.5\]](#).

An LDP Label Mapping message contains a FEC TLV, a Label TLV, and zero or more optional parameter TLVs.

The FEC TLV is used to indicate the meaning of the label. In the current context, the FEC TLV would be used to identify the particular pseudowire that a particular label is bound to. In this specification, we define two new FEC TLVs to be used for identifying pseudowires. When setting up a particular pseudowire, only one of these FEC TLVs is used. The one to be used will depend on the particular service being emulated and on the particular provisioning model being supported.

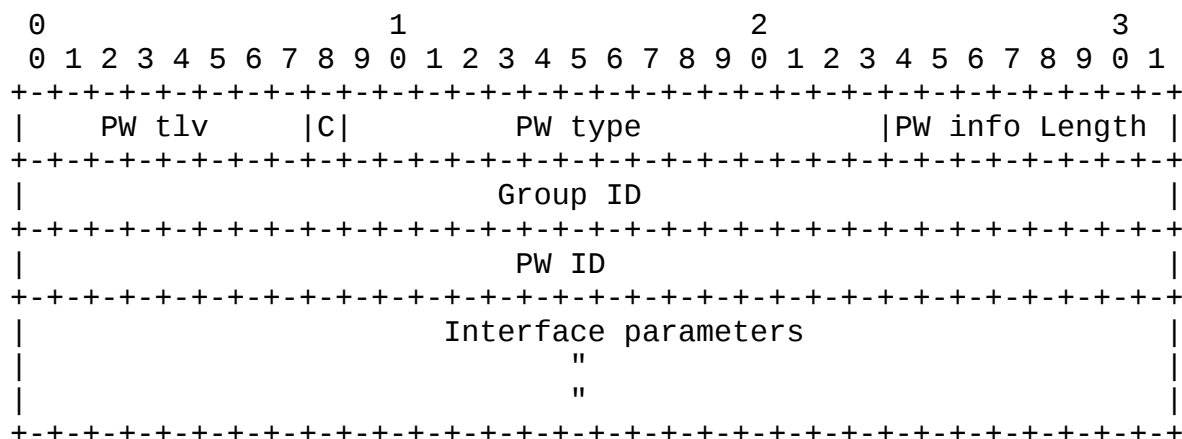
LDP allows each FEC TLV to consist of a set of FEC elements. For setting up and maintaining pseudowires, however, each FEC TLV MUST contain exactly one FEC element.

LDP has several kinds of label TLVs. For setting up and maintaining pseudowires, the Generic Label TLV MUST be used.

### 5.1. The Pwid FEC Element

The Pwid FEC element may be used whenever both pseudowire endpoints have been provisioned with the same 32-bit identifier for the pseudowire.

For this purpose a new type of FEC element is defined. The FEC element type is 128 [[note1](#)], and is defined as follows:



- PW type

A 15 bit quantity containing a value which represents the type of PW. Assigned Values are specified in "IANA Allocations for pseudo Wire Edge to Edge Emulation (PWE3)" [[14](#)].

- Control word bit (C)

The highest order bit (C) of the PW type is used to flag the presence of a control word (defined in [[7](#)]) as follows:

bit 15 = 1 control word present on this VC.  
 bit 15 = 0 no control word present on this VC.

Please see the section "C-Bit Handling Procedures" for further explanation.

- PW information length

Length of the PW ID field and the interface parameters field in octets. If this value is 0, then it references all PWs using the specified group ID and there is no PW ID present, nor any interface parameters.

- Group ID

An arbitrary 32 bit value which represents a group of PWs that is used to create groups in the VC space. The group ID is intended to be used as a port index, or a virtual tunnel index. To simplify configuration a particular PW ID at ingress could be part of the virtual tunnel for transport to the egress router. The Group ID is very useful to send wild card label withdrawals, or PW wild card status notification messages to remote PEs upon physical port failure.

- PW ID

A non-zero 32-bit connection ID that together with the PW type, identifies a particular PW. Note that the PW ID and the PW type must be the same at both endpoints.

- Interface parameters

This variable length field is used to provide interface specific parameters, such as CE-facing interface MTU.

Note that as the "interface parameters" are part of the FEC, the rules of LDP make it impossible to change the interface parameters once the pseudowire has been set up. Thus the interface parameters field must not be used to pass information, such as status information, which may change during the life of the pseudowire. Optional parameter TLVs should be used for that purpose.

Using the PWid FEC, each of the two pseudowire endpoints independently initiates the set up of a unidirectional LSP. An outgoing LSP and an incoming LSP are bound together into a single pseudowire if they have the same PW ID and PW type.

## 5.2. The Generalized ID FEC Element

There are cases where the PWid FEC element cannot be used, because both endpoints have not been provisioned with a common 32-bit PWid. In such cases, the "Generalized ID FEC Element" is used instead. This is FEC type 129 (provisionally, subject to assignment by IANA). It differs from the PWid FEC element in that the PWid and the group id are eliminated, and their place is taken by a generalized identifier field as described below. The Generalized ID FEC element includes a PW type field, a C bit, and an interface parameters field; these three fields are identical to those in the PWid FEC, and are used as discussed in the previous section.

### 5.2.1. Attachment Identifiers

As discussed in [13], a pseudowire can be thought of as connecting two "forwarders". The protocol used to setup a pseudowire must allow the forwarder at one end of a pseudowire to identify the forwarder at the other end. We use the term "attachment identifier", or "AI", to refer to the field which the protocol uses to identify the forwarders. In the PWid FEC, the PWid field serves as the AI. In this section we specify a more general form of AI which is structured and of variable length.

Every Forwarder in a PE must be associated with an Attachment Identifier (AI), either through configuration or through some algorithm. The Attachment Identifier must be unique in the context of the PE router in which the Forwarder resides. The combination <PE router, AI> must be globally unique.

It is frequently convenient to a set of Forwarders as being members of a particular "group", where PWs may only be set up among members of a group. In such cases, it is convenient to identify the Forwarders relative to the group, so that an Attachment Identifier would consist of an Attachment Group Identifier (AGI) plus an Attachment Individual Identifier (AII).

An Attachment Group Identifier may be thought of as a VPN-id, or a VLAN identifier, some attribute which is shared by all the Attachment VCs (or pools thereof) which are allowed to be connected.

The details of how to construct the AGI and AII fields identifying the pseudowire endpoints are outside the scope of this specification. Different pseudowire application, and different provisioning models, will require different sorts of AGI and AII fields. The specification of each such application and/or model must include the rules for constructing the AGI and AII fields.

As previously discussed, a (bidirectional) pseudowire consists of a pair of unidirectional LSPs, one in each direction. If a particular pseudowire connects PE1 with PE2, the LSP in the PE1-->PE2 direction can be identified as:

<PE1, <AGI, AII1>, PE2, <AGI, AII2>>,

and the LSP in the PE2-->PE1 direction can be identified by:

<PE2, <AGI, AII2>, PE1, <AGI, AII1>>.

Note that the AGI must be the same at both endpoints, but the AII will in general be different at each endpoint. Thus from the

perspective of a particular PE, each pseudowire has a local or "Source AII", and a remote or "Target AII". The pseudowire setup protocol can carry all three of these quantities:

- Attachment Group Identifier (AGI).
- Source Attachment Individual Identifier (SAII)
- Target Attachment Individual Identifier (TAII)

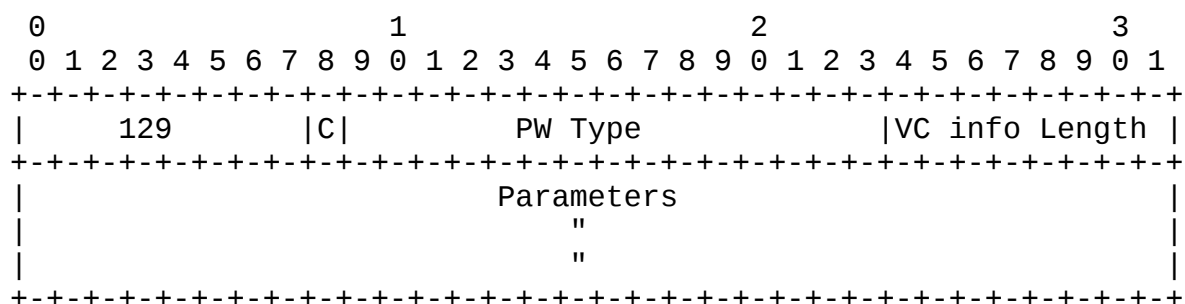
If the AGI is non-null, then the Source AI (SAI) consists of the AGI together with the SAII, and the Target AI (TAI) consists of the TAI together with the AGI. If the AGI is null, then the SAII and TAI are the SAI and TAI respectively.

The interpretation of the SAI and TAI is a local matter at the respective endpoint.

The association of two unidirectional LSPs into a single bidirectional pseudowire depends on the SAI and the TAI. Each application and/or provisioning model which uses the Generalized ID FEC element must specify the rules for performing this association.

#### 5.2.2. Encoding the Generalized ID FEC Element

FEC element type 129 is used. The FEC element is encoded as follows:



additional Parameters are:

- SAII, encoded as a one byte length field followed by the SAII.

- TAI, encoded as a one byte length field followed by the TAI.
- AGI, encoded as a one byte length field followed by the AGI.

The SAI, TAI, and AGI are simply carried as octet strings. The length byte specifies the size of the field, excluding the length byte itself. The null string can be sent by setting the length byte to 0.

### 5.2.3. Procedures

In order for PE1 to begin signaling PE2, PE1 must know the address of the remote PE2, and a TAI. This information may have been configured at PE1, or it may have been learned dynamically via some autodiscovery procedure.

To begin the signaling procedure, a PE (PE1) that has knowledge of the other endpoint (PE2) initiates the setup of the LSP in the incoming (PE2-->PE1) direction by sending a Label Mapping message containing the FEC type 129. The FEC element includes the SAI, AGI, and TAI.

What happens when PE2 receives such a Label Mapping message?

PE2 interprets the message as a request to set up a PW whose endpoint (at PE2) is the Forwarder identified by the TAI. From the perspective of the signaling protocol, exactly how PE2 maps AIs to Forwarders is a local matter. In some VPWS provisioning models, the TAI might, e.g., be a string which identifies a particular Attachment Circuit, such as "ATM3VPI4VCI5", or it might, e.g., be a string such as "Fred" which is associated by configuration with a particular Attachment Circuit. In VPLS, the TAI would be a VPN-id, identifying a particular VPLS instance.

If PE2 cannot map the TAI to one of its Forwarders, then PE2 sends a Label Release message to PE1, with a Status Code meaning "invalid TAI", and the processing of the Mapping message is complete.

If the Label Mapping Message has a valid TAI, PE2 must decide whether to accept it or not. The procedures for so deciding will depend on the particular type of Forwarder identified by the TAI. Of course, the Label Mapping message may be rejected due to standard LDP error conditions as detailed in [LDP].

If PE2 decides to accept the Label Mapping message, then it has to make sure that an LSP is set up in the opposite (PE1-->PE2) direction. If it has already signaled for the corresponding LSP in

that direction, nothing more need be done. Otherwise, it must initiate such signaling by sending a Label Mapping message to PE1. This is very similar to the Label Mapping message PE2 received, but with the SAI and TAI reversed.

### [5.3. Signaling of Pseudo Wire Status](#)

#### [5.3.1. Use of Label Mappings.](#)

The PEs MUST send PW label mapping messages to their peers as soon as the PW is configured and administratively enabled, regardless of the CE-facing interface state. The PW label should not be withdrawn unless the user administratively configures the CE-facing interface down (or the PW configuration is deleted entirely). A simple label withdraw method MAY also be supported as an alternative. In any case if the Label mapping is not available the PW MUST be considered in the down state.

#### [5.3.2. Signaling PW status.](#)

The PE devices use an LDP TLV to indicate status to their remote peers. This PW Status TLV contains more information than the alternative simple Label Withdraw message.

The format of the PW Status TLV is:

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								
1 0										PW Status (0x0???)										Length																			
										Status Code																													

Where status is a 4 octet bit field is specified in the PW IANA Allocations document [\[14\]](#)

Each bit in the status code field can be set individually to indicate more than a single failure at once. Each fault can be cleared by sending an appropriate status message with the respective bit cleared. The presence of the lowest bit (PW Not Forwarding) acts only as a generic failure indication when there is a link-down event for which none of the other bits apply.

The Status TLV is transported to the remote PW peer via the LDP notification message. The format of the Notification Message is:



```

      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|   Notification   (0x0001)   |   Message Length   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     Message ID         |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           PW FEC TLV   or Generalized ID FEC Element   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     PW Status TLV       |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

The PW FEC TLV SHOULD not include the interface parameters as they are ignored in the context of this message. When a PE's CE-facing interface encounters an error, use of the PW status message allows the PE to send a single status message, using a PW FEC TLV with only the group ID set, to denote this change in status for all affected PW connections.

As mentioned above the Group ID field can be used to send a status notification for all PWs associated with a particular group ID. This procedure is OPTIONAL, and if it is implemented the LDP Notification message should be as follows: the PW information length field is set to 0, the PW ID field is not present, and the interface parameters field is not present. For the purpose of this document this is called the "wild card PW status notification procedure", and all PEs implementing this design are REQUIRED to accept such a notification message, but are not required to send it.

#### 5.4. Interface Parameters Field

This field specifies interface specific parameters. When applicable, it MUST be used to validate that the PEs, and the ingress and egress ports at the edges of the circuit, have the necessary capabilities to interoperate with each other. The field structure 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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Parameter ID |   Length   |   Variable Length Value   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     Variable Length Value   |
|                                     "                         |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

The parameter ID Values are specified in "IANA Allocations for pseudo Wire Edge to Edge Emulation (PWE3)" [[14](#)].

The Length field is defined as the length of the interface parameter including the parameter id and length field itself. Processing of the interface parameters should continue when encountering unknown interface parameters and they MUST be silently ignored.

- Interface MTU

A 2 octet value indicating the MTU in octets. This is the Maximum Transmission Unit, excluding encapsulation overhead, of the egress packet interface that will be transmitting the decapsulated PDU that is received from the MPLS network. This parameter is applicable only to PW types 1, 2, 4, 5, 6, 7, 14, and 15 and is REQUIRED for these PW types. If this parameter does not match in both directions of a specific PW, that PW MUST NOT be enabled.

- Maximum Number of concatenated ATM cells

A 2 octet value specifying the maximum number of concatenated ATM cells that can be processed as a single PDU by the egress PE. An ingress PE transmitting concatenated cells on this PW can concatenate a number of cells up to the value of this parameter, but MUST NOT exceed it. This parameter is applicable only to PW types 3, 9, and 0x0a, and is REQUIRED for these PWC types. This parameter does not need to match in both directions of a specific PW.

- Optional Interface Description string

This arbitrary, OPTIONAL, interface description string is used to send a human-readable administrative string describing the interface to the remote. This parameter is OPTIONAL, and is applicable to all PW types. The interface description parameter string length is variable, and can be from 0 to 80 octets. Human-readable text MUST be provided in the UTF-8 charset using the Default Language [[RFC2277](#)].

- Payload Bytes

A 2 octet value indicating the number of TDM payload octets contained in all packets on the CEM stream, from 48 to 1,023 octets. All of the packets in a given CEM stream have the same number of payload bytes. Note that there is a possibility that the packet size may exceed the SPE size in the case of an STS-1 SPE, which could cause two pointers to be needed in the CEM

header, since the payload may contain two J1 bytes for consecutive SPEs. For this reason, the number of payload bytes must be less than or equal to 783 for STS-1 SPEs.

- CEP Options.

An optional 16 Bit value of CEM Flags. See [8] for the definition of the bit values.

- Requested VLAN ID.

An Optional 16 bit value indicating the requested VLAN ID. This parameter MAY be used by an PE that is incapable of rewriting the 802.1Q ethernet VLAN tag on output. If the ingress PE receives this request it MAY rewrite the VLAN ID tag in input to match the requested VLAN ID. If this is not possible, and the VLAN ID does not already match configured ingress VLAN ID the PW should not be enabled. This parameter is applicable only to PW type 4.

- CEP/TDM bit rate.

This 32-bit integer is mandatory for CEP. For other PWs carrying TDM traffic it is mandatory if the bit-rate cannot be directly inferred from the service type. If present, it expresses the bit rate of the attachment circuit as known to the advertizing PE in "units" of 64 kbit/s. I.e., the value 26 must be used for CEP carrying VT1.5 SPE, 35 - for CEP carrying a VT2 SPE, 99 - for VT6 SPE, 783 - for STS-1 SPE and  $n \times 783$  - for STS-nc,  $n = 3, 12, 48, 192$ . Attempts to establish a PWC between a pair of TDM ports with different bit-rates MUST be rejected with the appropriate status code (see section "Status codes" below).

- Frame-Relay DLCI length.

An optional 16 bit value indicating the length of the frame-relay DLCI field. This OPTIONAL interface parameter can have value of 2, or 4, with the default being equal to 2. If this interface parameter is not present the default value of 2 is assumed.

#### 5.4.1. PW types for which the control word is REQUIRED

The Label Mapping messages which are sent in order to set up these PWs MUST have  $c=1$ . When a Label Mapping message for a PW of one of these types is received, and  $c=0$ , a Label Release MUST be sent, with an "Illegal C-bit" status code. In this case, the PW will not be enabled.

#### 5.4.2. PW types for which the control word is NOT mandatory

If a system is capable of sending and receiving the control word on PW types for which the control word is not mandatory, then each such PW endpoint MUST be configurable with a parameter that specifies whether the use of the control word is PREFERRED or NOT PREFERRED. For each PW, there MUST be a default value of this parameter. This specification does NOT state what the default value should be.

If a system is NOT capable of sending and receiving the control word on PWC types for which the control word is not mandatory, then it behaves as exactly as if it were configured for the use of the control word to be NOT PREFERRED.

If a Label Mapping message for the PW has already been received, but no Label Mapping message for the PW has yet been sent, then the procedure is the following:

- i. If the received Label Mapping message has c=0, send a Label Mapping message with c=0, and the control word is not used.
- ii. If the received Label Mapping message has c=1, and the PW is locally configured such that the use of the control word is preferred, then send a Label Mapping message with c=1, and the control word is used.
- iii. If the received Label Mapping message has c=1, and the PW is locally configured such that the use of the control word is not preferred or the control word is not supported, then act as if no Label Mapping message for the PW had been received (i.e., proceed to the next paragraph).

If a Label Mapping message for the PW has not already been received (or if the received Label Mapping message had c=1 and either local configuration says that the use of the control word is not preferred or the control word is not supported), then send a Label Mapping message in which the c bit is set to correspond to the locally configured preference for use of the control word. (I.e., set c=1 if locally configured to prefer the control word, set c=0 if locally configured to prefer not to use the control word or if the control word is not supported).

The next action depends on what control message is next received for that PW. The possibilities are:

- i. A Label Mapping message with the same c bit value as specified in the Label Mapping message that was sent. PW setup is now complete, and the control word is used if c=1 but not used if c=0.

- ii. A Label Mapping message with c=1, but the Label Mapping message that was sent has c=0. In this case, ignore the received Label Mapping message, and continue to wait for the next control message for the PW.
- iii. A Label Mapping message with c=0, but the Label Mapping message that was sent has c=1. In this case, send a Label Withdraw message with a "Wrong c-bit" status code, followed by a Label Mapping message that has c=0. PW setup is now complete, and the control word is not used.
- iv. A Label Withdraw message with the "Wrong c-bit" status code. Treat as a normal Label Withdraw, but do not respond. Continue to wait for the next control message for the PW.

If at any time after a Label Mapping message has been received, a corresponding Label Withdraw or Release is received, the action taken is the same as for any Label Withdraw or Release that might be received at any time. Note that receiving a Label Withdraw should not cause a corresponding Label Release to be sent.

If both endpoints prefer the use of the control word, this procedure will cause it to be used. If either endpoint prefers not to use the control word, or does not support the control word, this procedure will cause it not to be used. If one endpoint prefers to use the control word but the other does not, the one that prefers not to use it has no extra protocol to execute, it just waits for a Label Mapping message that has c=0.

The diagram in [Appendix A](#) illustrates the above procedure.

#### [5.4.3](#). Status codes

[RFC 3036](#) has a range of Status Code values which are assigned by IANA on a First Come, First Served basis. These additional status codes, and assigned Values are specified in "IANA Allocations for pseudo Wire Edge to Edge Emulation (PWE3)" [[14](#)].

#### [5.5](#). LDP label Withdrawal procedures

As mentioned above the Group ID field can be used to withdraw all PW labels associated with a particular group ID. This procedure is OPTIONAL, and if it is implemented the LDP label withdraw message should be as follows: the PW information length field is set to 0, the PW ID field is not present, and the interface parameters field is not present. For the purpose of this document this is called the "wild card withdraw procedure", and all PEs implementing this design are REQUIRED to accept such a withdraw message, but are not required

to send it.

The interface parameters field MUST NOT be present in any LDP PW label withdrawal message or release message. A wildcard release message MUST include only the group ID. A Label Release message initiated from the imposition router must always include the PW ID.

## [5.6.](#) Sequencing Considerations

In the case where the router considers the sequence number field in the control word, it is important to note the following when advertising labels

### [5.6.1.](#) Label Mapping Advertisements

After a label has been withdrawn by the disposition router and/or released by the imposition router, care must be taken to not re-advertise (re-use) the released label until the disposition router can be reasonably certain that old packets containing the released label no longer persist in the MPLS network.

This precaution is required to prevent the imposition router from restarting packet forwarding with sequence number of 1 when it receives the same label mapping if there are still older packets persisting in the network with sequence number between 1 and 32768. For example, if there is a packet with sequence number= $n$  where  $n$  is in the interval $[1,32768]$  traveling through the network, it would be possible for the disposition router to receive that packet after it re-advertises the label. Since the label has been released by the imposition router, the disposition router SHOULD be expecting the next packet to arrive with sequence number to be 1. Receipt of a packet with sequence number equal to  $n$  will result in  $n$  packets potentially being rejected by the disposition router until the imposition router imposes a sequence number of  $n+1$  into a packet. Possible methods to avoid this is for the disposition router to always advertise a different PW label, or for the disposition router to wait for a sufficient time before attempting to re-advertise a recently released label. This is only an issue when sequence number processing at the disposition router is enabled.

### 5.6.2. Label Mapping Release

In situations where the imposition router wants to restart forwarding of packets with sequence number 1, the router shall 1) Send to disposition router a label mapping release, and 2) Send to disposition router a label mapping request. When sequencing is supported, advertisement of a PW label in response to a label mapping request MUST also consider the issues discussed in the section on Label Mapping Advertisements.

## 6. Security Considerations

This document does not affect the underlying security issues of MPLS.

## 7. References

- [1] "LDP Specification." L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas. January 2001. [RFC3036](#)
- [2] ITU-T Recommendation Q.933, and Q.922 Specification for Frame Mode Basic call control, ITU Geneva 1995
- [3] "MPLS Label Stack Encoding", E. Rosen, Y. Rekhter, D. Tappan, G. Fedorkow, D. Farinacci, T. Li, A. Conta. [RFC3032](#)
- [4] "IEEE 802.3ac-1998" IEEE standard specification.
- [5] American National Standards Institute, "Synchronous Optical Network Formats," ANSI T1.105-1995.
- [6] ITU Recommendation G.707, "Network Node Interface For The Synchronous Digital Hierarchy", 1996.
- [7] "Frame Relay over Pseudo-Wires", [draft-ietf-pwe3-frame-relay-01.txt](#). ( work in progress )
- [8] "SONET/SDH Circuit Emulation Service Over Packet (CEP)", [draft-ietf-pwe3-sonet-01.txt](#) ( Work in progress )
- [9] ATM Forum Specification fb-fbatm-0151.000 (2000) ,Frame Based ATM over SONET/SDH Transport (FAST)
- [10] "Encapsulation Methods for Transport of ATM Cells/Frame Over IP and MPLS Networks", [draft-ietf-pwe3-atm-encap-02.txt](#) ( work in progress )

[11] "Encapsulation Methods for Transport of PPP/HDLC Frames Over IP and MPLS Networks", [draft-ietf-pwe3-hdlc-ppp-encap-00.txt](#). ( work in progress )

[12] "Encapsulation Methods for Transport of Ethernet Frames Over IP/MPLS Networks", [draft-ietf-pwe3-ethernet-encap-01.txt](#). ( work in progress )

[13] "PWE3 Architecture" Bryant, et al., [draft-ietf-pwe3-arch-04.txt](#) ( work in progress ), August 2003.

[14] "IANA Allocations for pseudo Wire Edge to Edge Emulation (PWE3)" Martini, Townsley, [draft-ietf-pwe3-iana-allocation-01.txt](#) ( work in progress ), February 2003

[RFC2434] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations section in RFCs", [BCP 26](#), [RFC 2434](#), October 1998.

[RFC2277] Alvestrand, H., "IETF Policy on Character Sets and Languages", [BCP 18](#), [RFC 2277](#), January 1998.

[note1] FEC element type 128 is pending IANA approval.

[note2] Status codes assignment is pending IANA approval.

## 8. Author Information

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

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



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

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

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

## 9. Additional Contributing Authors

Dimitri Stratton Vlachos  
Mazu Networks, Inc.  
125 Cambridgepark Drive  
Cambridge, MA 02140  
e-mail: d@mazunetworks.com

Jayakumar Jayakumar,  
Cisco Systems Inc.  
225, E.Tasman, MS-SJ3/3,  
San Jose, CA, 95134  
e-mail: jjayakum@cisco.com

Alex Hamilton,  
Cisco Systems Inc.  
285 W. Tasman, MS-SJCI/3/4,  
San Jose, CA, 95134  
e-mail: tahamilt@cisco.com

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

John Shirron  
Omega Corporate Center  
1300 Omega Drive  
Pittsburgh, PA 15205  
Laurel Networks, Inc.  
e-mail: jshirron@laurelnetworks.com

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

Andrew G. Malis  
Vivace Networks, Inc.  
2730 Orchard Parkway  
San Jose, CA 95134  
Phone: +1 408 383 7223  
Email: 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](mailto:chris@cw.net)

Dave Cooper  
Global Crossing  
960 Hamlin Court  
Sunnyvale, CA 94089  
e-mail: [dcooper@gblix.net](mailto:dcooper@gblix.net)

Kireeti Kompella  
Juniper Networks  
1194 N. Mathilda Ave  
Sunnyvale, CA 94089  
e-mail: [kireeti@juniper.net](mailto:kireeti@juniper.net)

10. Appendix A - C-bit Handling Procedures Diagram