Internet Draft
Document: draft-ietf-pwe3-framework-00.txt

Prayson Pate, Editor
Overture Networks, Inc.
XiPeng Xiao
Photuris Inc.
Craig White
Level 3 Communications, LLC.
Tricci So
Caspian Networks
Kireeti Kompella
Juniper Networks, Inc.
Andrew G. Malis
Vivace Networks
Thomas K. Johnson
Litchfield Communications
Thomas D. Nadeau
Stewart Bryant
Cisco Systems

Framework for Pseudo Wire Emulation Edge-to-Edge (PWE3)

draft-ietf-pwe3-framework-00.txt

Status of this Memo

Abstract

This document describes a framework for Pseudo Wire Emulation Edge-to-Edge (PWE3). It discusses the emulation of circuits (such as T1, E1, T3, E3 and SONET/SDH) and services (such as ATM and Frame Relay) over packet switched networks (PSNs) using IP or MPLS. It presents an architectural framework for pseudo wires (PWs), defines terminology, specifies the various protocol elements and their functions, overviews some of the services that will be supported and discusses how PWs fit into the broader context of protocols.

Copyright Notice

Table of Contents

1. Introduction

This document describes a framework for Pseudo Wire Emulation Edge-to-Edge (PWE3). It discusses the emulation of circuits (such as T1, E1, T3, E3 and SONET/SDH) and services (such as ATM and Frame Relay) over packet switched networks (PSNs) using IP or MPLS. It presents an architectural framework for pseudo wires (PWs), defines terminology, specifies the various protocol elements and their functions, overviews the services supported and discusses how PWs fit into the broader context of protocols. See [XIAO] for the requirements for PWs.

1.1. What Are Pseudo Wires?

1.1.1. Definition

PWE3 is a mechanism that emulates the essential attributes of a service (such as a T1 leased line or Frame Relay) over a PSN. The required functions of PWs include encapsulating service-specific bit-streams or PDUs arriving at an ingress port, and carrying them across a path or tunnel, managing their timing and order, and any other operations required to emulate the behavior and characteristics of the service as faithfully as possible.

From the customer perspective, the PW is perceived as an unshared link or circuit of the chosen service. However, there may be deficiencies that impede some applications from being carried on a PW. These limitations should be fully described in the appropriate service-specific Encapsulation, Emulation and Maintenance Documents (EEMDs) and Applicability Statements (ASes).

1.1.2. Functions

PWs provide the following functions in order to emulate the behavior

and characteristics of the desired service.

- Encapsulation of service-specific PDUs or circuit data arriving at an ingress port (logical or physical).

- Carrying the encapsulated data across a tunnel.

- Managing the signaling, timing, order or other aspects of the service at the boundaries of the PW.

- Service-specific status and alarm management.

EEMDs and/or ASes for each service will describe any shortfalls of the emulation's faithfulness.

1.2. Goals of This Document

- Description of the motivation for creating PWs, and some background on how they may be deployed.

- Description of an architecture and terminology for PWs.

- Description of the statistics and other network management information needed for tunnel operation and management.

- Whenever possible, relevant requirements from existing IETF documents and other sources will be incorporated by reference.

1.3. Non-Goals

The following are non-goals for this document:

- The detailed specification of the bits and bytes of the encapsulations of the various services. This description is contained in an EEMD and/or ES.

- The detailed definition of the protocols involved in PW setup and maintenance.

The following are outside the scope of PWE3:

- Discussion of the protection of the encapsulated content of the PW.

- Any multicast service not native to the emulated medium. Thus, Ethernet transmission to a "multicast" IEEE-48 address is in scope, while multicast services like MARS that are implemented on top of the medium are out of scope.

- Methods to signal or control the underlying PSN.

1.4. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALLNOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119. Below are the definitions for the terms used throughout the document.

Packet Switched Network
: A Packet Switched Network (PSN) is a network using IP or MPLS as the unit of switching.

Pseudo Wire Emulation Edge to Edge
: Pseudo Wire Emulation Edge to Edge (PWE3) is a mechanism that emulates the essential attributes of a service (such as a T1 leased line or Frame Relay) over a PSN.

Customer Edge
: A Customer Edge (CE) is a device where one end of an emulated service originates and terminates. The CE is not aware that it is using an emulated service rather than a "real" service.

Provider Edge
: A Provider Edge (PE) is a device that provides PWE3 to a CE.

Pseudo Wire
: A Pseudo Wire (PW) is a connection between two PEs carried over a PSN. The PE provides the adaptation between the CE and the PW.

PW End Service
: A Pseudo Wire End Service (PWES) is the interface between a PE and a CE. This can be a physical interface like a T1 or Ethernet, or a virtual interface like a VC or VLAN.

Pseudo Wire PDU
: A Pseudo Wire PDU is a PDU sent on the PW that

contains all of the data and control information necessary to provide the desired service.

PSN Tunnel — A PSN Tunnel is a tunnel inside which multiple PWs can be nested so that they are transparent to core network devices.

PW Domain — A PW Domain (PWD) is a collection of instances of PWs that are within the scope of a single homogeneous administrative domain (e.g. PW over MPLS network or PW over IP network etc.).

Path-oriented PW — A Path-oriented PW is a PW for which the network devices of the underlying PSN must maintain path state information.

Non-path-oriented PW — A Non-path-oriented PW is a PW for which the network devices of the underlying PSN need not maintain path state information.

Interworking — Interworking is used to express interactions between networks, between end systems, or between parts thereof, with the aim of providing a functional entity capable of supporting an end-to-end communication. The interactions required to provide a functional entity rely on functions and on the means to select these functions.

Interworking Function — An Interworking Function (IWF) is a functional entity that facilitates interworking between two dissimilar networks (e.g., ATM & MPLS, ATM & L2TP, etc.). A PE performs the IWF function.

Service Interworking — In Service Interworking, the IWF (Interworking Function) between two dissimilar protocols (e.g., ATM & MPLS, Frame Relay & ATM, ATM & IP, ATM & L2TP, etc.) terminates the protocol used in one network and translates (i.e. maps) its Protocol Control Information (PCI) to the PCI of the protocol used in other network for User, Control and Management Plane functions to the extent possible. In general, since not all functions may be supported in one or other

of the networks, the translation of PCI may be partial or non-existent. However, this should not result in any loss of user data since the payload is not affected by PCI conversion at the service interworking IWF.

Network Interworking
: In Network Interworking, the PCI (Protocol Control Information) of the protocol and the payload information used in two similar networks are transferred transparently by an IWF of the PE across the PSN. Typically the IWF of the PE encapsulates the information which is transmitted by means of an adaptation function and transfers it transparently to the other network.

Applicability Statement
: Each PW service will have an Applicability Statement (AS) that describes the applicability of PWs for that service.

Encapsulation, Emulation and Maintenance Documents
: Each PW service will have an Encapsulation, Emulation and Maintenance Document (EEMDs) that described the particulars of PWs for that service, as well as the degree of faithfulness to that service.

Inbound
: The traffic direction where information from a CE is adapted to a PW, and PW-PDUs are sent into the PSN.

Outbound
: The traffic direction where PW-PDUs are received on a PW from the PSN, re-converted back in the emulated service, and sent out to a CE.

CE Signaling
: CE (end-to-end) Signaling refers to messages sent and received by the CEs. It may be desirable or even necessary for the PE to participate in or monitor this signaling in order to effectively emulate the service.

PE/PW Maintenance
: PE/PW Maintenance is used by the PEs to set up,

maintain and tear down the PW. It may be coupled with CE signaling in order to effectively manage the PW.

PSN Tunnel Signaling
PSN Tunnel Signaling is used to set up, maintain and remove the underlying PSN tunnel. An example would be LDP in MPLS for maintaining LSPs.

2. Background and Motivation

Why is anyone interested in PWs? This section gives some background on where networks are today and why they are changing. It also talks about the motivation to provide converged networks while continuing to support existing services. Finally it discusses how PWs can be a solution for this dilemma.

2.1. Current Network Architecture

2.1.1. Multiple Networks

For any given service provider delivering multiple services, the current "network" usually consists of parallel or "overlay" networks. Each of these networks implements a specific service, such as voice, Frame Relay, Internet access, etc. This is quite expensive, both in terms of capital expense as well as in operational costs. Furthermore, the presence of multiple networks complicates planning. Service providers wind up asking themselves these questions:

- Which of my networks do I build out?

- How many fibers do I need for each network?

- How do I efficiently manage multiple networks?

A converged network helps service providers answer these questions in a consistent and economical fashion.

2.1.2. Convergence Today

There are some examples of convergence in today's network:

- Frame Relay is frequently carried over ATM networks using [FRF.5] interworking.

- T1, E1 and T3 circuits are sometimes carried over ATM networks using [ATMCES].

- Voice is carried over ATM (using AAL2), Frame Relay (using FRF.11 VoFR), IP (using VoIP) and MPLS (using VoMPLS) networks.

Deployment of these examples range from limited (ATM CES) to fairly common (FRF.5 interworking) to rapidly growing (VoIP).

2.1.3. The Emerging Converged Network

Many service providers are finding that the new IP-based and MPLS-based switching systems are much less costly to acquire, deploy and maintain than the systems that they replace. The new systems take advantage of advances in technology in these ways:

- The newer systems leverage mass production of ASICs and optical interfaces to reduce capital expense.

- The bulk of the traffic in the network today originates from packet sources. Packet switches can economically switch and deliver this traffic natively.

- Variable-length switches have lower system costs than ATM due to simpler switching mechanisms as well as elimination of segmentation and reassembly (SAR) at the edges of the network.

- Deployment of services is simpler due to the connectionless nature of IP services or the rapid provisioning of MPLS applications.

2.1.4. Transition to a Packet-Optimized Converged Network

The greatest assets for many service providers are the physical communications links that they own. The time and costs associated with acquiring the necessary rights of way, getting the required governmental approvals, and physically installing the cabling over a variety of terrains and obstacles represents a significant asset that is difficult to replace. Their greatest on-going costs are the operational expenses associated with maintaining and operating their networks. In order to maximize the return on their assets and minimize their operating costs, service providers often look to consolidate the delivery of multiple service types onto a single networking technology.

The first generation converged network was based on TDM (time-division multiplexing) technology. Voice, video, and data traffic have been carried successfully across TDM/DACS-based networks for decades. TDM technology has some significant drawbacks as a converged networking technology. Operational costs for TDM networks remain relatively high because the provisioning of end-to-end TDM

circuits is typically a tedious and labor-intensive task. In addition, TDM switching does not make the best use of the communications links. This is because fixed assignment of timeslots does not allow for the statistical multiplexing of bursty data traffic (i.e. temporarily unused bandwidth on one timeslot cannot be dynamically re-allocated to another service).

The second generation of converged network was based on ATM technology. Today many service providers convert voice, video, and data traffic into fixed-length cells for carriage across ATM-based networks. ATM improves upon TDM technology by providing the ability to statistically multiplex different types of traffic onto communications links. In addition, ATM SPVC technology is often used to automatically provision end-to-end services, providing an additional advantage over traditional TDM networks. However, ATM has several significant drawbacks. One of the most frequently cited problems with ATM is the so-called cell-tax, which refers to the 5 bytes out of 53 used as an ATM cell header. Another significant problem with ATM is the AAL5 SAR, which becomes extremely difficult to implement above 1 Gbps. There are also issues with the long-term scalability of ATM, especially as a switching layer beneath IP.

As packet traffic takes up a larger and larger portion of the available network bandwidth, it becomes increasingly useful to optimize public networks for the Internet Protocol. However, many service providers are confronting several obstacles in engineering packet-optimized networks. Although Internet traffic is the fastest growing traffic segment, it does not generate the highest revenue per bit. For example, Frame Relay traffic currently generates a higher revenue per bit than do native IP services. Private line TDM services still generate even more revenue per bit than does Frame Relay. In addition, there is a tremendous amount of legacy equipment deployed within public networks that does not communicate using the Internet Protocol. Service providers continue to utilize non-IP equipment to deploy a variety of services, and see a need to interconnect this legacy equipment over their IP-optimized core networks.

2.2. PWE3 as a Path to Convergence

How do service providers realize the capital and operational benefits of a new packet-based infrastructure, while leveraging the existing base of SONET (Synchronous Optical Network) gear, and while also

protecting the large revenue stream associated with this equipment? How do they move from mature Frame Relay or ATM networks, while still being able to provide these lucrative services?

One possibility is the emulation of circuits or services via PWs. Circuit emulation over ATM and interworking of Frame Relay and ATM have already been standardized. Emulation allows existing circuits and/or services to be carried across the new infrastructure, and thus enables the interworking of disparate networks. [ATMCES] provides some insight into the requirements for such a service:

> There is a user demand for carrying certain types of constant bit rate (CBR) or "circuit" traffic over Asynchronous Transfer Mode (ATM) networks. As ATM is essentially a packet- rather than circuit-oriented transmission technology, it must emulate circuit

> characteristics in order to provide good support for CBR traffic.
>
> A critical attribute of a Circuit Emulation Service (CES) is that the performance realized over ATM should be comparable to that experienced with the current PDH/SDH technology.

Implemented correctly, PWE3 can provide a means for supporting today's services over a new network.

2.3. Suitable Applications for PWE3

What makes an application suitable (or not) for PWE3 emulation? When considering PWs as a means of providing an application, the following questions must be considered:

- Is the application sufficiently deployed to warrant emulation?

- Is there interest on the part of service providers in providing an emulation for the given application?

- Is there interest on the part of equipment manufacturers in providing products for the emulation of a given application?

- Are the complexities and limitations of providing an emulation worth the savings in capital and operational expenses?

If the answer to all four questions is "yes", then the application is likely to be a good candidate for PWE3. Otherwise, there may not be sufficient overlap between the customers, service providers, equipment manufacturers and technology to warrant providing such an emulation.

2.4. Summary

To maximize the return on their assets and minimize their operational costs, many service providers are looking to consolidate the delivery of multiple service offerings and traffic types onto a single IP-optimized network.

In order to create this next-generation converged network, standard methods must be developed to emulate existing telecommunications formats such as Ethernet, Frame Relay, ATM, and TDM over IP-optimized core networks. This document describes a framework accomplishing this goal.

3. Architecture of Pseudo Wires

3.1. Network Reference Model

Figure 1 below shows the network reference model for PWs. As shown, the PW provides an emulated service between the customer edges (CEs).

```
                       |<------- Pseudo Wire ------>|
                       |    |<-- PSN Tunnel -->|    |
                PW     V    V                  V    V     PW
           End Service +----+                  +----+ End Service
      +-----+    |     | PE1|==================| PE2|     |    +-----+
      |     |----------|............PW1.............|----------|     |
      | CE1 |    |     |    |                  |    |     |    | CE2 |
      |     |----------|............PW2.............|----------|     |
      +-----+    |     |    |==================|    |     |    +-----+
    Customer^          +----+                  +----+     |    ^Customer
     Edge 1 |      Provider Edge 1         Provider Edge 2     | Edge 1
            |<-------------- Emulated Service ---------------->|
```

Figure 1: PWE3 Network Reference Model

Any bits or packets presented at the PW End Service (PWES) are encapsulated in a PW-PDU and carried across the underlying network. The PEs perform the encapsulation, decapsulation, order management, timing and any other functions required by the service. In some cases the PWES can be treated as a virtual interfaces into a further processing (like switching or bridging) of the original service before the physical connection to the CE. Examples include Ethernet bridging, SONET cross-connect, translation of locally-significant identifiers such as VCI/VPI, etc. to other service type, etc. The underlying PSN is not involved in any of these service-specific operations.

3.2. Maintenance Reference Model

Figure 2 below shows the maintenance reference model for PWs.

```
         |<------- CE (end-to-end) Signaling ------>|
         |     |<---- PW/PE Maintenance ----->|     |
         |     |     |<-- PSN Tunnel -->|     |     |
         |     |     |    Signaling     |     |     |
         |     V     V  (out of scope)  V     V     |
         v     +-----+                  +-----+     v
   +-----+     | PE1 |==================| PE2 |     +-----+
   |     |-----|.............PW1..............|-----|     |
   | CE1 |     |     |                  |     |     | CE2 |
   |     |-----|.............PW2..............|-----|     |
   +-----+     |     |==================|     |     +-----+
               +-----+                  +-----+
   Customer   Provider                 Provider   Customer
    Edge 1     Edge 1                   Edge 2     Edge 2
```

Figure 2: PWE3 Maintenance Reference Model

- The CE (end-to-end) signaling is between the CEs. This signaling includes Frame Relay PVC status signaling, ATM SVC signaling, etc.

- The PW/PE Maintenance is used between the PEs to set up, maintain and tear down PWs, including any required coordination of parameters between the two ends.

- The PSN Tunnel signaling controls the underlying PSN. An example would be LDP in MPLS for maintaining LSPs. This type of signaling is not within the scope of PWE3.

3.3. Maintenance Architecture

<Editor's Note: Luca Martini has asked that a section be added describing the architecture of the maintenance protocol(s). This section will contain that architecture. Some possible topics are listed below.>

3.3.1. PW Setup

How are PWs set up?

3.3.2. PW Integrity

How do we ensure the sanity or connectivity of the PW over the PSN?

3.3.3. Service Maintenance

How does CE maintenance behavior e.g. FR LMI, ATM AIS/RDI, etc. affect the PW?

3.4. Protocol Stack Reference Model

Figure 3 below shows the protocol stack reference model for PWs.

```
+-----------------+                               +-----------------+
|Emulated Service|       Emulated Service        |Emulated Service|
|(TDM, ATM, etc).|<============================>|(TDM, ATM, etc.)|
+-----------------+           Pseudo Wire         +-----------------+
|  Encapsulation |<============================>|  Encapsulation |
+-----------------+           PSN Tunnel          +-----------------+
|    IP/MPLS     |<============================>|    IP/MPLS     |
+-----------------+                               +-----------------+
|    Physical    |                               |    Physical    |
+--------+--------+        ______________         +---------+-------+
         |                /              \                  |
         +===============/       PSN      \===============+
                         \                /
                          \______________/
```

Figure 3: PWE3 Protocol Stack Reference Model

The PW provides the CE with what appears to be a connection to its peer at the far end. Bits or PDUs from the CE are passed through an encapsulation layer.

3.5. Logical Protocol Layering Model

3.5.1. Protocol Layers

The logical protocol-layering model needed to support a PW is shown in Figure 4 below.

```
+---------------------------+
|          Payload          |
+---------------------------+
|    Encapsulation Layer    |
+---------------------------+
|     Multiplexing Layer    |
+---------------------------+
|        PSN Header         |
+---------------------------+
|       MAC/Datalink        |
+---------------------------+
|         Physical          |
+---------------------------+
```

Figure 4: Logical Protocol Layering Model

The logical protocol-layering model shown in Figure 4 above minimizes the differences between the PWs operating over different PSN types.

The payload is transported over the Encapsulation Layer that carries any information that is not available in the payload itself and which is needed by the PW outbound interface to send the reconstructed service to the CE. If needed, this layer also provides support for real-time processing, sequencing and indication of length.

The Multiplexing Layer provides the ability to deliver multiple PWs over a single PSN tunnel.

The PSN header, MAC/datalink and physical layer definitions are outside the scope of this framework.

3.5.2. Instantiation of the Protocol Layers

The instantiation of the logical protocol-layering model of Figure 4 is shown in Figure 5 below.

Where possible, the components shown below use existing IETF standards and common work in progress. Otherwise, the goal was to call for the design of components that have the wider application within the IETF.

```
       +===========================================+ - - - - - - - - - -
       |      Payload (circuit/cell/packet)        | Payload
       +===========================================+ - - - - - - - - - -
       |   Encapsulation-specific information      |
       |           Sequencing (optional)           | Encapsulation
       |       Length (payload/PSN-specific)       | Layer
       |  Timing Information (payload-specific)    |
       +============+==============================+ - - - - - - - - - -
       |  L2TPv3    |        MPLS shim             | Multiplexing Layer
       +------------+-------+----------------------+ - - - - - - - - - -
       |    IPv4/v6         |        MPLS          | PSN Header
       +====================+======================+ - - - - - - - - - -
       |               MAC/Data-link               |
       +===========================================+
       |                 Physical                  |
       +===========================================+
```

Figure 5: Instantiated Protocol Layering

3.5.2.1. Payload

The payload is generically classified into three types: circuit, cell and packet. Within these generic types there will be specific service types. For example, the generic packet type includes Frame Relay and Ethernet. The design of the encapsulation layer, and the choice between transporting the payload in a native or intermediate format, will be defined in the service-specific EEMD and/or AS.

3.5.2.1.1. Circuit Payload

A circuit payload is a sampled set of bits relayed across the PW. This service will need sequencing and real-time support.

3.5.2.1.2. Cell Payload

A cell payload is one, or more, ATM cells relayed across the PW. This service will need sequence number support, and may also need real-time support. The payload may consist of a single cell or a cluster of cells. The cells to be incorporated in the payload may be selected by filtering on VCI/VPI. In the case of a trunked interface the payload may be considered complete on the expiry of a timer or when a fixed number of cells have been received or both.

3.5.2.1.3. Packet Payload

A packet payload would normally be relayed across the PW as a single unit. A packet payload may need sequencing or sequencing and real-time support. A packet payload may additionally require

fragmentation

3.5.2.2. Sequencing

Support for sequencing depends on the payload type, and may be omitted if not needed. For example, Frame Relay always requires the preservation of packet order. However, where the Frame Relay service is only being used to carry IP, it may be desirable to relax that constraint in return for reduced per-packet processing cost.

[RTP] or encapsulation-specific sequence numbers may be used for sequencing.

3.5.2.3. Timing

A suitable real-time protocol is RTP [RTP]. This is an extensible protocol designed to be extended to carry new payload types over a transport layer running over an IP network. It includes a control protocol for managing the timing service. RTP is not currently defined for operation over L2TP. A short document describing the correct method of multiplexing the RTP control and data over LT2Pv3 is required.

3.5.2.4. Multiplexing Layer

Suitable protocols for use as a multiplexing layer are L2TPv3 [LAU] and MPLS [MPLS].

- The updated definition of L2TP in [LAU] is specified to operate directly over a PSN, and in the limiting case the L2TP data encapsulation reduces to a four-octet opaque data field. L2TPv3 is specified for operation in both manually configured and negotiated mode. The associated control protocol is extensible and may be used to signal PW-specific configuration or run-time information.

- MPLS may also be used in the multiplexing layer. In this case, an MPLS tag is used as a "shim" to identify the particular PW of interest.

3.5.2.5. PSN Layer

The three PSN types within the scope of the IETF are IPv4, IPv6 and

MPLS. IPv4 and IPv6 both provide the necessary switching, length and fragmentation services needed to support all IETF specified transport protocols. L2TPv3 is specified to run directly over IPv4 and IPv6. When the PSN is IPv4 or IPv6 no PSN convergence layer is needed.

MPLS provides switching service, but no length or fragmentation service. When MPLS is used as the PSN, the encapsulation must provide length and fragmentation services, if needed.

3.6. Architecture Assumptions

1) The current design is focused on a point-to-point and same-to-same service interface at both end of the PW. Only network interworking will be performed at the edge or the PW. Support for service interworking is for out of scope.

2) The initial design of PWE3 is focused on a single homogeneous administrative PWD (e.g. PW over MPLS or PW over IP etc. ONLY). Interworking between different PW types and the support of inter-domain PWs are for further study.

3) The design of PW will not perfectly emulate the characteristics of the native service. It will be dependent on both the emulated service, as well as on the network implementation. An EEMD and AS shall be created for each service to describe the degree of faithfulness of a PW to the native service.

4) Only the permanent emulated circuit type (e.g. PVC/PVP) is considered initially. The switched emulated circuit type (e.g. SVC/SVP) will be for further study.

5) The creation and placement of the PSN tunnel to support the PW is not within the scope.

6) The current PW encapsulation approach considerations are focused on IPv4, IPv6 and MPLS. Other encapsulation approaches are for further study.

7) Current PW service applications are focused on Ethernet (i.e. Ethernet II (DIX), 802.3 "raw", Ethernet 802.2, Ethernet SNAP, 802.3ac VLAN, 802.1Q), Frame Relay, ATM and TDM (e.g. DS1, DS3,

E1, SONET/SDH etc.).

8) Within the single administrative PWD, the design of the PW assumes the inheritance of the security mechanism that has been applied to the emulated services. The PW also inherits any security features from the PSN e.g. IPsec for an IP PSN. No PW-specific security mechanism will be specified.

4. Design Considerations

4.1. PW-PDU Validation

It is a common practice to use a checksum, CRC or FCS to assure end-to-end integrity of frames. The PW service-specific mechanisms must define whether the packet's checksum shall be preserved across the PWD or be removed at the ingress PE and then be re-calculated at the egress PE. The former approach saves work, while the later saves bandwidth.

For protocols like ATM and Frame Relay, the checksum is only applicable to a single link. This is because the circuit identifiers (e.g. Frame Relay DLCI or ATM VPI/VCI) have only local significance and are changed on each hop or span. If the circuit identifier (and thus checksum) is going to change as a part of the PW emulation, it would be more efficient to strip and re-calculate the checksum.

Other PDU headers (e.g. UDP in IP) do not change during transit. It would make sense to preserve these types of checksums.

The EEMD for each protocol must describe the validation scheme to be used.

4.2. PW-PDU Sequencing

One major consideration of PW design is how to ensure in-sequence delivery of packets, if needed. The design of the PW for each protocol must consider the support of the PSN for in-order delivery as well as the requirements of the particular application. for example, IP is connectionless and does not guarantee in-order delivery. When using IP, a PW sequence number may be needed for some applications (such as TDM).

4.3. Session Multiplexing

One way to facilitate scaling is to increase the number of PWs per underlying tunnel. There are two ways to achieve this:

- For a service like Relay or ATM, all of the VCs on a given port could be lumped together. VCs would not be distinguishable within the PWD.

- Service SDUs could be distinguished within a PW-PDU by port, channel or VC identifiers. This approach would allow for switching or grooming in the PWD.

4.4. Security

Each EEMD must specify a means to protect the control of the PWE and the PE/PW signaling. The security-related protection of the encapsulated content of the PW is outside of scope.

4.5. Encapsulation Control

4.5.1. Scalability

Different service types may be required between CEs. Support of multiple services implies that a range of PWD label spaces may be needed. If the PWD spans a PSN supporting traffic engineering, then the ability to supporting label stacking would be desirable.

4.5.2. Service Integration

It may be desirable to design a PW to transport a variety of services which have different transport characteristics. To achieve this integration it may be useful to allow the service requirements to be mapped to the tunneling label in such a way that the PWD can apply the appropriate service and transport management to the PW.

4.6. Statistics

The PE can tabulate statistics that help monitor the state of the network, and to help with measurement of SLAs. Typical counters include:

- Counts of PW-PDUs sent and received, with and without errors.

- Counts of PW-PDUs lost (TDM only).

- Counts of service PDUs sent and received, with and without errors (non-TDM).

- Service-specific interface counts.

These counters would be contained in a MIB, they should not replicate existing MIB counters.

4.7. Traceroute

Tracing functionality is desirable for emulated circuits and services, because it allows verification and remediation of the operation and configuration of the forwarding plane. [BONICA] describes the requirements for a generic route tracing application. Applicability of these requirements to PWE3 is an interesting problem, as many of the emulated services have no equivalent function. In general, there is not a way to trace the forwarding plane of an TDM or Frame Relay PVC. ATM does provide an option in the loopback OAM cell to return each intermediate hop (see [I.610]).

There needs to be a mechanism through which upper layers can ask emulated services to reveal their internal forwarding details. A common mechanism for all emulated services over a particular PSN may be possible. For example, if MPLS is the PSN, the path for a VC LSP could be revealed via the signaling from the underlying TE tunnel LSP, or perhaps via the proposed MPLS OAM. However, when we are trying to trace the entire emulated service, starting from the CE (e.g. an ATM VCC), then a uniform approach probably will not work and different approaches would be required for different emulated services.

4.8. Congestion Considerations

The PSN carrying the PW may be subject to congestion. The congestion characteristics will vary with the PSN type, the network architecture and configuration, and the loading of the PSN.

Each PW EEMD and/or AS will have to specify whether it needs an appropriate mechanism for operating in the presence of this

congestion, including methods of mapping between its native congestion reporting and avoidance mechanisms, and those provided by the PW.

4.9. PW SNMP MIB Architecture

This section describes the general architecture for SNMP MIBs used to manage PW services and the underlying PSN. The intent here is to provides a clear picture of how all of the pertinent MIBs fit together to form a cohesive management framework for deploying PWE3 services.

4.9.1. MIB Layering

The SNMP MIBs created for PWE3 should fit the architecture shown in Figure 6.

```
               +-----------+ +-----------+     +-----------+
     Service   |    CEM    | | Ethernet  |     |    ATM    |
      Layer    |Service MIB| |Service MIB| ... |Service MIB|
               +-----------+ +-----------+     +-----------+
                       \           |             /
                         \         |           /
   - - - - - - - - - - - - \ - - - | - - - - / - - - - - - -
                             \     |       /
               +---------------------------------------------+
    Generic PW |             Generic PW MIBs                 |
      Layer    +---------------------------------------------+
                             /     |       \
   - - - - - - - - - - - - / - - - | - - - - \ - - - - - - -
                         /         |           \
                       /           |             \
               +-----------+ +-----------+     +-----------+
     PSN VC    |GRE VC MIB | |L2TP VC MIB|     |MPLS VC MIB|
      Layer    +-----------+ +-----------+     +-----------+
                     |             |                 |
   - - - - - - - - - | - - - - - - | - - - - - - - - | - - -
                     |             |                 |
               +-----------+ +-----------+     +-----------+
       PSN     |GRE MIB(s) | |L2TP MIB(s)|     |MPLS MIB(s)|
      Layer    +-----------+ +-----------+     +-----------+
```

Figure 6: Relationship of SNMP MIBs

Figure 7 shows an example for a TDM PW carried over MPLS.

```
                  +------------------+
                  |    SONET MIB     | RFC2558
                  +------------------+
                           |
                  +------------------+
       Service    |SONET Service MIB| pw-cem-mib
        Layer     +------------------+
    - - - - - - - - - - - | - - - - - - - - - - - - - - -
                  +------------------+
    Generic PW    | Generic PW MIBS | pw-tc-mib
        Layer     +------------------+ pw-mib
    - - - - - - - - - - - | - - - - - - - - - - - - - - -
                  +------------------+
        PSN VC    |   MPLS VC MIBS   | pw-mpls-mib
        Layer     +------------------+
    - - - - - - - - - - - | - - - - - - - - - - - - - - -
                  +------------------+
         PSN      |    MPLS MIBs     |   mpls-te-mib
        Layer     +------------------+   mpls-lsr-mib
```

Figure 7: Service-specific Example for MIBs

Note that there is a separate MIB for each emulated service as well as one for each underlying PSN. These MIBs may be used in various combinations as needed.

4.9.2. Service Layer MIBs

The first layer is referred to as the Service Layer. It contains MIBs for PWE3 services such as Ethernet, ATM, circuits and Frame Relay. This layer contains those corresponding MIBs used to mate or adapt those emulated services to the underlying services. This working group should not produce any MIBs for managing the general service; rather, it should produce just those MIBs that are used to interface or adapt the emulated service onto the PWE3 management framework. For example, the standard SONET MIB [SONETMIB] is designed and maintained by another working group. Also, the SONET MIB is designed to manage the native service without PW emulation. Since the PWE3 working group is chartered to produce the corresponding adaptation MIB, in this case, it would produce the PW-CEM-MIB [PWMPLSMIB] that would be used to adapt SONET services to the underlying PSN that carries the PWE3 service.

4.9.3. Generic PW MIBs

The second layer is referred to as the Generic PW Layer. This layer is composed of two MIBs: the PWE-TC-MIB [PWTCMIB] and the PWE-MIB [PWMIB]. These MIBs are responsible for providing general PWE3 counters and service models used for monitoring and configuration of

PWE3 services over any supported PSN service. That is, this MIB

provides a general model of PWE3 abstraction for management purposes. This MIB is used to interconnect the Service Layer MIBs to the PSN VC Layer MIBs. The latter will be described in the next section. This layer also provides the PW-TC-MIB [PWTCMIB]. This MIB contains common SMI textual conventions [SMIv2] that may be used by any PW MIB.

4.9.4. PSN VC Layer MIBs

The third layer in the PWE3 management architecture is referred to as the PSN VC layer. This layer is comprised of MIBs that are specifically designed to interface general PWE3 services (VCs) onto those underlying PSN services. In general this means that the MIB provides a means with which an operator can map the PW service onto the native PSN service. For example, in the case of MPLS, it is required that the general VC service be layered onto MPLS LSPs or Traffic Engineered (TE) Tunnels [MPLS]. In this case, the PW-MPLS-MIB [PWMPLSMIB] was created to adapt the general PWE3 circuit services onto MPLS. Like the Service Layer described above the PWE3 working group should produce these MIBs.

4.9.5. PSN Layer MIBs

The fourth and final layer in the PWE3 management architecture is referred to as the PSN layer. This layer is comprised of those MIBs that control the PSN service-specific services. For example, in the case of the MPLS [MPLS] PSN service, the MPLS-LSR-MIB [LSRMIB] and the MPLS-TE-MIB [TEMIB] are used to interface the general PWE3 VC services onto native MPLS LSPs and/or TE tunnels to carry the emulated services. In addition, the MPLS-LDP-MIB [LDPMIB] may be used to reveal the MPLS labels that are distributed over the MPLS PSN in order to maintain the PW service. The MIBs in this layer are produced by other working groups that design and specify the native PSN services. These MIBs should contain the appropriate mechanisms for monitoring and configuring the PSN service such that the emulated PWE3 service will function correctly.

5. Acknowledgments

This document was created by the PWE3 Framework design team. Valuable input was also provided by John Rutemiller, David Zelig, Durai Chinnaiah, Jayakumar Jayakumar, Ron Bonica, Ghassem Koleyni, and Eric Rosen.

6. References

[L2TP] W.M. Townsley, A. Valencia, A. Rubens, G. Singh Pall, G. Zorn, B. Palter, "Layer Two Tunneling Protocol (L2TP)", RFC 2661, August 1999.

[RTP] H. Schulzrinne et al, "RTP: A Transport Protocol for Real-Time Applications", RFC1889, January 1996.

[MPLS] E. Rosen, "Multiprotocol Label Switching Architecture", RFC3031, January 2001.

[IP] DARPA, "Internet Protocol", RFC791, September 1981.

[CONGEST] S. Floyd, "Congestion Control Principles", RFC2914, September 2000.

[SONETMIB] K. Tesink, "Definitions of Managed Objects for the SONET/SDH Interface Type", RFC2558, March 1999.

[SMIv2] Case et al, "Structure of Management Information for Version 2 of the Simple Network Management Protocol (SNMPv2)", RFC1902, January 1996.

[MALIS] Malis et al, "SONET/SDH Circuit Emulation over Packet (CEP)" (draft-malis-pwe3-sonet-01.txt), work in progress, November 2001.

[XIAO] Xiao et al, "Requirements for Pseudo Wire Emulation Edge-to-Edge (PWE3)" (draft-pwe3-requirements-02.txt), work in progress, November 2001.

[BONICA] Bonica et al, "Tracing Requirements for Generic Tunnels" (draft-bonica-tunneltrace-01.txt), work in progress, February 2001.

[LAU] J Lau et al, Layer Two Tunneling Protocol "L2TP", (draft-ietf-l2tpext-l2tp-base-01.txt) work in progress, October 2001.

[PWMIB] Zelig et al, "Pseudo Wire (PW) Management Information Base Using SMIv2", (draft-zelig-pw-mib-00.txt), work in progress, July 2001.

[PWTCMIB] Nadeau et al, "Definitions for Textual Conventions and OBJECT-IDENTITIES for Pseudo-Wires Management" (draft-nadeau-pw-tc-mib-00.txt), work in progress, July 2001.

[PWMPLSMIB] Danenberg et al, "SONET/SDH Circuit Emulation Service Over MPLS (CEM) Management Information Base Using SMIv2", (draft-danenberg-pw-cem-mib-00.txt), work in progress, July 2001.

[LSRMIB] Srinivasan et al, "MPLS Label Switch Router Management Information Base Using SMIv2", draft-ietf-mpls-lsr-mib-07.txt, work in progress, January 2001.

[TEMIB] Srinivasan et al, "Traffic Engineering Management Information Base Using SMIv2",

<draft-ietf-mpls-te-mib-05.txt>, work in progress, November 2000.

[LDP-MIB] Cucchiara, J., Sjostrand, H., and Luciani, J., "Definitions of Managed Objects for the Multiprotocol Label Switching, Label Distribution Protocol (LDP)", <draft-ietf-mpls-ldp-mib-08.txt>, work in progress, August 2001.

[ATMCES] ATM Forum, "Circuit Emulation Service Interoperability Specification Version 2.0" (af-vtoa-0078-000), January 1997.

[FRF.5] O'Leary et al, "Frame Relay/ATM PVC Network Interworking Implementation Agreement", Frame Relay Forum FRF.5, December 20, 1994. ITU Recommendation Q.933, Annex A, Geneva, 1995.

[I.363.1] ITU, "B-ISDN ATM Adaptation Layer specification: Type 1 AAL", Recommendation I.363.1, August, 1996.

[I.610] ITU, "B-ISDN Operation and Maintenance Principles and Functions", ITU Recommendation I.610, February, 1999.

[802.3] IEEE, "ISO/IEC 8802-3: 2000 (E), 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.

7. Security Considerations

It may be desirable to define methods for ensuring security during exchange of encapsulation control information at an administrative boundary of the PSN.

8. IANA Considerations

There are no IANA considerations for this document.

9. Authors' Addresses

Prayson Pate
Overture Networks, Inc.
P. O. Box 14864
RTP, NC, USA 27709
prayson.pate@overturenetworks.com

XiPeng Xiao
Photuris, Inc.
2025 Stierlin Court
Mountain View, CA 94043
xiaoxipe@cse.msu.edu

Tricci So
Caspian Networks
170 Baytech Dr.
San Jose, CA 95134
tso@caspiannetworks.com

Craig White
Level 3 Communications, LLC.
1025 Eldorado Blvd.
Broomfield, CO, 80021
Craig.White@Level3.com

Kireeti Kompella
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
kireeti@juniper.net

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

Thomas K. Johnson
Litchfield Communications
76 Westbury Park Rd.
Watertown, CT 06795
tom_johnson@litchfieldcomm.com

Thomas D. Nadeau
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA 01824
tnadeau@cisco.com

Stewart Bryant
Cisco Systems Ltd
4, The Square
Stockley Park
Uxbridge UB11 1BL UK
stbryant@cisco.com

10. Full Copyright Section