Multi-Service Interworking - IP Over MPLS

MFA Forum 16.0.0

MFA Forum Technical Committee

February 2007

Note: The user's attention is called to the possibility that implementation of the MFA Forum Technical Specification contained herein may require the use of inventions covered by patent rights held by third parties. By publication of this MFA Forum Technical Specification the MFA Forum makes no representation that the implementation of the specification will not infringe on any third party rights. The MFA Forum take no position with respect to any claim that has been or may be asserted by any third party, the validity of any patent rights related to any such claims, or the extent to which a license to use any such rights may not be available.

Editor:

Himanshu Shah

Ciena Corp.

For more information contact:

The MFA Forum
Suite 117
48377 Fremont Blvd.
Fremont, CA 94538 USA

Phone: +1 (510) 492-4056
FAX: +1 (510) 492-4001
E-Mail: info@mfaforum.org
WWW: http://www.mfaforum.org

Full Notice

Revision History

Version	Change	Date
MFA.16.0.0	Initial Revision	February 24, 2007

Table of Contents

1. Introduction

1.1 PURPOSE

This specification describes Multi-service Interworking for IP version 4 over an MPLS core network. The term "Multi-service Interworking" means that IP packets are transported across the MPLS core in a pseudo wire (PW) [RFC 3985], and the two attachment circuits (ACs) associated with the PW may employ different Layer 2 technologies (e.g. ATM on one AC and Ethernet on other).

The native service provided to end users is the IP service. This is a Layer 2-based IP service that is distinguished from a Layer 3 IP service by the fact that a PE forwards a CE's IP traffic based upon Layer 2 information rather than Layer 3 information. In particular, the PE device does not perform longest prefix match lookup of the destination IP address for frame forwarding, nor does it participate in routing protocols with the CE, for discovering network topology. In the context of Virtual Private Networks, this Layer 2 forwarding characteristic makes this a Layer 2 VPN service as opposed to an IP-VPN (L3VPN) service [RFC 4364].

The use of Layer 2 interworking for IP traffic enables carriers and service providers to introduce Ethernet and PPP as attachments while preserving existing ATM and Frame Relay infrastructure.

1.2 SCOPE

This specification:

- Describes point-to-point IP connectivity (Virtual Private Wire Service (VPWS[L2VPN-FRM])) for two Attachment Circuits using any combination of the following technologies: Ethernet, Frame Relay, ATM[1], or PPP
- Describes Layer 2 VPN service for IP with minimal or no changes to CE configurations
- Does not require a PE to participate in the CEs' routing protocols
- Supports the address resolution function for all supported Attachment Circuit types (Ethernet, ATM, Frame Relay, and PPP)
- Does not support IPv6 in this revision, since that would require a solution for neighbor discovery resolution
- Does not support native IS-IS; it only supports native IP version 4 encapsulations

1.3 OVERVIEW

IP Multi-service Interworking over MPLS is a Virtual Private Wire Service (VPWS). It provides point-to-point IP connectivity for two CE devices across an MPLS network. What distinguishes this from a homogeneous Layer 2 VPN is that it allows the type of Attachment Circuits used to connect each site to be different from each other, and it does so in a way that requires minimal configuration changes to the connecting CEs in some topologies and no changes in other topologies.

[1] This specification requires that an ATM AC use AAL5.

This specification describes how a PE device discovers a CE device that is attached by one of a variety of types of Attachment Circuit, and how a PE forwards a customer's IP traffic based upon the learned Layer 2-specific information. The solution requires the PE to participate in the different address resolution procedures used by CE devices on different types of Attachment Circuit, and to exchange the information with other PEs to enable a local PE to act as a proxy for a remote CE.

Section 1 briefly describes the scope, purpose and overview of the IP multi-service interworking solution. Section 2 defines terminology and provides references used in the document. Section 3 describes the reference model for IP multi-service interworking. Section 4 describes the two-sided model for IP multi-service interworking. Section 5 provides an overview of the principles of the interworking function. Section 6 discusses various encapsulations that are used by the interworking function. Section 7 describes the management of IP pseudo wires, and section 8 describes the configuration elements that are necessary to provision the IP multi-service interworking function.

Annex A discusses the single-sided model for IP multi-service interworking.

Annex B details the control plane elements of IP multi-service interworking.

2. Definitions, Terminology, and References

2.1 DEFINITIONS

Must, Shall or Mandatory — the item is an absolute requirement of this specification
Should — the item is desirable.
May or Optional — the item is not compulsory, and may be followed or ignored according to the needs of the implementer.

2.2 ACRONYMS

Acronym	Definition
AC	Attachment Circuit
AAL5	ATM Adaptation Layer Type 5
ARP	Address Resolution Protocol
ATM	Asynchronous Transfer Mode
BDR	Backup Designated Router
BECN	Backward Explicit Congestion Notification
BGP	Border Gateway Protocol
CE	Customer Edge
CPCS	Common Part Convergence Sublayer
CRC	Cyclic Redundancy Check
DA	Destination Address
DE	Discard Eligible
DLCI	Data Link Connection Identifier

Acronym	Definition
DMAC	Destination Media Access Control
DR	Designated Router
DSAP	Destination Service Access Point
FEC	Forwarding Equivalence Class
FECN	Forward Explicit Congestion Notification
FWD	Forwarder
HDLC	High-Level Data Link Connection
IANA	Internet Assigned Numbers Authority
IETF	Internet Engineering Task Force
IGP	Internet Gateway Protocol
IP	Internet Protocol
IPCP	IP Control Protocol
IWF	Inter Working Function
LDP	Label Distribution Protocol
LLC	Logical Link Control
L2VPN	Layer 2 VPN
L3VPN	Layer 3 VPN
LMI	Link Management Interface
LSP	Label Switched Path
MAC	Media Access Control
MPLS	Multi Protocol Label Switching
NCP	Network Control Protocol
NLPID	Network Layer Protocol Identifier
NSP	Native Service Processor
OSPF	Open Shortest Path First
OUI	Organizational Unique Identifier
PDU	Protocol Data Unit
PE	Provider Edge
PID	Protocol Identifier
PPP	Point to Point Protocol
PW	Pseudo Wire
PWE3	Pseudo Wire Emulation Edge to Edge
PVC	Permanent Virtual Circuit
RIP	Routing Information Protocol
SA	Source Address

Acronym	Definition
SAP	Service Access Point
SAR	Segmentation and Reassembly
SDU	Service Data Unit
SMAC	Source Media Access Control
SNAP	Service Network Access Point
SSAP	Source Service Access Point
SVC	Switched Virtual Circuit
TLV	Type Length Value
UI	User Information
VC	Virtual Circuit
VCC	Virtual Channel Connection
VLAN	Virtual Local Area Network
VPI	Virtual Path Identifier
VCI	Virtual Channel Identifier
VPN	Virtual Private Network
VPWS	Virtual Private Wire Service

2.3 REFERENCES

2.3.1 Normative References

The following is a list of standards on which this implementation agreement is based:

RFC 1112	IETF – Host Extensions for IP Multicasting, August 1989
RFC 2427	IETF – Multiprotocol interconnect over Frame Relay, September 1998
RFC 2684	IETF – Multiprotocol Encapsulation over ATM Adaptation Layer 5, September 1999
RFC 3985	IETF – Pseudo wire Emulation Edge-to-Edge (PWE3) Architecture, March 2005
RFC 4446	IETF - IANA allocations for PWE3, April 2006
RFC 4447	IETF – Pseudo wire setup and maintenance using LDP, April 2006
OAM-IW	MFA Forum 13.0.0 – Fault Management for Multi-service Interworking over MPLS Version 1.0, June 2006

2.3.2 Informative References

Following is a list of standards this implementation agreement refers to for information purposes.

RFC 1332 IETF – PPP Internet Control Protocol, May 1992.

RFC 1700	IETF - Assigned Numbers, October,
RFC 2225	IETF – Classical IP and ARP over ATM, April 1998.
RFC 3036	IETF - LDP Specification, January 2001
RFC 4364	1994IETF - BGP/MPLS IP Virtual Private Networks (VPNs), February 2006
IEEE-802.1Q	IEEE Standards for Local and metropolitan area networks – Virtual Bridged Local Area Networks, May 2003
MFA-FR-ATM	MFA Forum, MPLS2006.117.01 - Multi-Service Interworking - Frame Relay and ATM Service Interworking over MPLS, April 2006
RFC 4417	RFC 4717 IETF - Encapsulation Methods for Transport of Asynchronous Transfer Mode (ATM) over MPLS Networks, December 2006
L2VPN-FRM	IETF - Framework for Layer 2 VPN, RFC 4664, September 2006.

3. Interworking Reference Model

Figure 1 shows an interworking reference model, where CE, PE, Pseudo Wire (PW), NSP (Native Service Processor), PW Processor (including payload encapsulation, LDP signaling, etc. for PW processing), and Attachment Circuit (AC) are defined by the IETF in [RFC 3985]. Specifically, the AC is a physical or virtual circuit connection between a CE and a PE. As shown in Figure 1, Native Service 1 and Native Service 2, over Attachment Circuit 1 and Attachment Circuit 2, respectively, are interconnected to define an end-to-end service between the two CEs, with one or two Interworking Functions (IWF) residing in the PE(s).

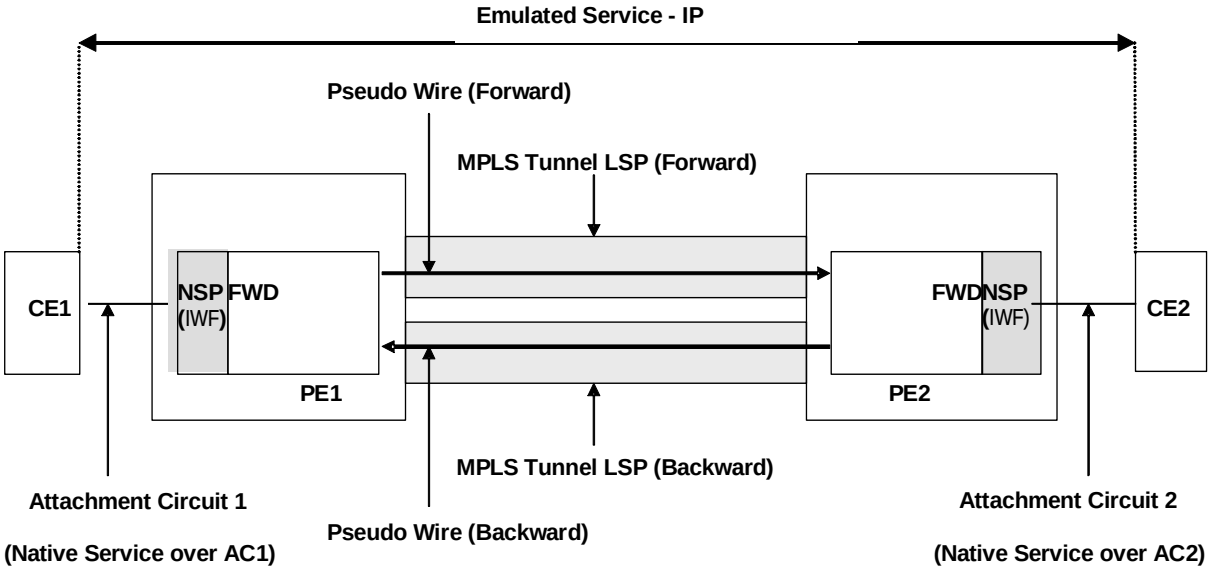


Figure 1 Interworking Reference Architecture

3.1 ATTACHMENT CIRCUITS

A CE device attaches to a PE device via a circuit known as an "Attachment Circuit" (AC). An Attachment Circuit may be a Frame Relay VC, an ATM VCC, an Ethernet port, a VLAN, an HDLC/PPP connection on a physical interface, etc. The CE device is an IP router or an IP host. The CE device may directly attach to the PE or may reside behind a Layer 2 switched network. However, the following restrictions are imposed:

- For a given Attachment Circuit there is only one CE IP device that subscribes to the Layer 2 IP service. It is further noted that when an AC is Frame Relay or ATM, the AC must be of type DLCI or VPI/VCI, respectively.
- For a given IP host address, the CE device does not connect to more than one PE through a Layer 2 switched network; for example, an Ethernet switch connected to a CE on one side cannot connect to two PEs on the other side via a single Attachment Circuit (e.g. Ethernet + VLAN).

3.2 NATIVE SERVICES OVER ATTACHMENT CIRCUITS

A Native Service is defined as the service provided by the PE to the CE over the Attachment Circuit, to which a customer subscribes.

In this specification we consider the Native Service as IP over a data link transport. Nested data link transport, such as IP over Ethernet over ATM/FR, is outside the scope of this specification; for example, if the CE1-PE1 Attachment Circuit is Frame Relay, then from CE1's perspective IP over Frame Relay transport is the service provided to CE1. However, if the PE2-CE2 link is Ethernet, IP over Ethernet transport is the service provided to CE2. With Multi-service Interworking enabled, CE1 and CE2 receive end-to-end native IP service, transparent to the disparity between the data link layers of the Attachment Circuits.

The NSP module resides in the PE. It is responsible for processing the data received over the AC from the CE before presentation to the PW for transmission across the MPLS network. It is also responsible for processing the data received from a PW before it is sent out over the AC. This specification describes the following interworking functions that the NSP module is responsible for when performing multi-service IP interworking over MPLS. Note these functions are not required for every scenario:

- Data link encapsulation and decapsulation processing for the supported AC types and for the associated PW – e.g., identification and mapping of each service instance for each AC to and from the associated PW
- Packet discard of all non-IP packets, except for ARP and Inverse ARP frames, which are submitted to the IWF control plane for learning the data link address-to-IP address association
- Traffic Management Interworking - Mapping of traffic and service parameters (such as Drop Precedence or Congestion Indication, if any) between the supported AC and the corresponding PW
- PVC Management Interworking - Mapping PVC management indications between the ATM/FR/Ethernet AC and the corresponding PW

The IWFs in this interworking reference model (see Figure 1) are at times referred to as "IP IWFs" to distinguish them, where needed, from other types of interworking functions. In this specification, if the term "IWF" is not otherwise qualified, it is assumed to mean "IP IWF". The IP IWF is contrasted with other IWFs by way of performing IP-specific interworking, as

described in this specification; for example, a FRF8.2 IWF is different from an IP IWF, and both can co-exist.

The forwarder module in Figure 1 is responsible for forwarding frames between an AC and the PW corresponding to a given service instance. There is one such forwarder module per service instance. There is static mapping between a given AC and its associated PW such that the AC implies which PW is to be used for forwarding the frames to the remote PE (and vice versa - the PW implies which AC is to be used).

It should be noted that the frame adaptation functionality (formatting the frames based upon the AC type) and filtering and/or blocking of non-IP packets (which includes ARP frames, Inverse ARP frames, control frames, etc.) is done by the NSP module (more specifically by the IWF within this module) and not by the forwarder module. It should also be noted that the representation of the PE in Figure 1 is logical, and its modules can be implemented by different physical entities.

4. INTERWORKING MODELS

There are two models for IP Interworking discussed in this specification. In the first model the data link format utilized by each Attachment Circuit is removed at the respective PE. The PEs communicate over an IP PW. When a link layer frame arrives at one PE, the IP PDU is extracted, then encapsulated in an IP pseudo wire, and transported to the remote PE. From there the PDU is decapsulated from the IP pseudo wire, encapsulated in the appropriate data link format, and sent on the remote Attachment Circuit. This model is referred to as the two-sided model, and is described in section 4.1. Conformance of an implementation to this model is mandatory.

The second model does not use an IP pseudo wire. Instead, the PW is of the same type as one of the ACs (call it the remote AC). When a frame arrives at the remote PE from the remote AC, the PE adapts it to the PW format without decapsulating the embedded IP PDU. The IP Interworking Function resides only on the local PE, which is responsible for decapsulating the data link header of the PW payload and encapsulating the data link header of the local AC type. This model is referred to as the single-sided model, and is described in Annex A. Conformance of an implementation to this model is optional.

4.1 TWO-SIDED MODEL

In the two-sided model an IP IWF resides on each of the PEs. In this model, the Layer 2 Attachment Circuit control and management planes are terminated by the NSP at each PE. When a frame is received over the Attachment Circuit, the IP IWF processes it as follows.

- A Layer 2 control frame, such as an LMI frame over Frame Relay or a PAUSE frame over Ethernet, is submitted to the control plane for local processing.
- An address resolution protocol frame, such as ARP or inverse ARP, is submitted to the control plane for address learning purposes, and is then discarded.
- A frame that is neither an L2 control frame nor an address resolution protocol frame is examined to see if its payload is an IPv4 packet; if the payload is not an IPv4 packet then the frame is discarded.

IP packets are forwarded over an IP PW. The encapsulation format and the signaling of the IP PW are defined in [RFC 4447]. The IP PW type, 0x000B, is defined in [RFC 4446].

A frame received from the IP PW is submitted to the IP IWF. The IP IWF extracts the IP packet and adds a Layer 2 encapsulation using the Layer 2 addressing information collected by the control plane.

When a frame is received over the Attachment Circuit, the forwarding decision is based upon the Attachment Circuit information; for instance, if the ingress circuit is an Ethernet Attachment Circuit, the Ethernet port or Ethernet port plus VLAN ID is used as the key to the forwarding decision. When the ingress Attachment Circuit is Frame Relay or ATM, the DLCI or the ATM VPI/VCI information, respectively, is used as the key. When the ingress Attachment Circuit is PPP, the interface index is used as the key.

When a frame is received over the PW, the PW label is used to identify the egress AC and the associated data link header information.

Figure 2 below illustrates the two-sided model for IP interworking over MPLS. It shows that a PE-to-CE Attachment Circuit can be any one of Ethernet, ATM, Frame Relay or PPP at either end. The figure also shows an IP IWF residing on both PEs, and an IP PW extending between the PEs.

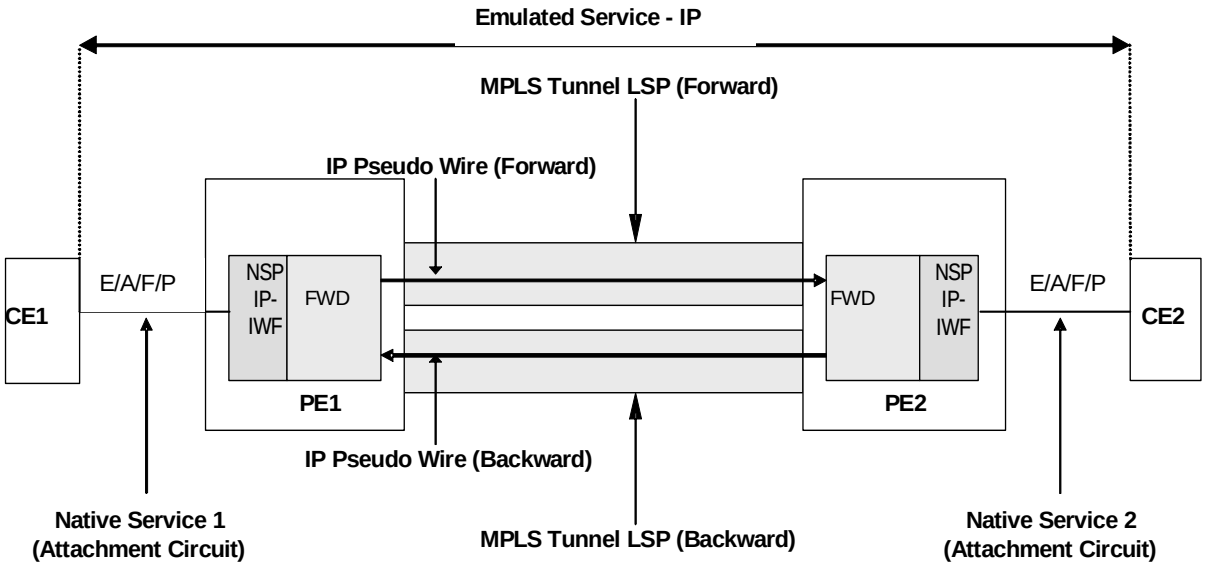


Figure 2: Two-sided IP Multi-Service Interworking

While the two-sided model is sufficiently general that it could be supported by several PW types, the PW type specified here for the two-sided model is the IP PW type, as illustrated in Figure 2. Compared to the other PW types, the IP PW type best represents a canonical PW that can be used across PEs, in that any PE can support IP services without any knowledge of the AC types or Layer 2-specific PW types that might be supported at other PEs. Additionally, the use of the IP PW type results in efficient use of network resources in that only the IP packets, without additional Layer 2 encapsulations, need to be transported over the pseudo wire.

4.2 CE CONFIGURATION CHANGES

When a CE is connected to a PE via an Ethernet AC, configuration changes are required if the OSPF routing protocol is enabled on that Ethernet AC. The issue is that IP multi-service interworking supports only point-to-point connectivity. However, OSPF normally executes special procedures (such as DR, BDR election) when it is enabled on a broadcast subnet like Ethernet. In order to disable this procedures, OSPF allows explicit configuration to treat the underlying network as point-to-point. This specification requires that when OSPF is running over an Ethernet AC, the CE must be configured to treat the underlying network as point-to-point.

5. IWF for IP

The IP interworking function in the PE operates between an AC of one type and a PW of a disparate type. In the two-sided model we have an IP PW, and thus we have IP interworking functions in both PEs. The main elements of this function are,

- Data link header manipulation
- Data link address resolution

The data link header manipulation involves removing the data link encapsulation of an IP packet at the ingress and adding the appropriate data link encapsulation at the egress. The data link address resolution procedure involves:

- Learning the *data link address-to-IP address association* of the attached CE
- Exchanging this information with the remote PE
- Using a proxy function to inform the local CE of the remote CE's *data link address-to-IP address association*

The data link address-to-IP address resolution protocol is different for each Attachment Circuit type; for instance, Ethernet uses ARP, Frame Relay and ATM use Inverse ARP, and PPP uses IPCP [RFC 1332]. Thus, the Attachment Circuit type dictates what address resolution protocol the PE will participate in to learn the address associations.

The details of data link address resolution functions are defined in Annex B.

6. Encapsulation Formats

This section describes the frame encapsulations used for IP multi-service interworking. The following AC types are supported:

- Ethernet
- Frame Relay
- ATM
- PPP

6.1 IP MULTI-SERVICE INTERWORKING FOR ETHERNET AC

The Ethernet AC is identified by the Ethernet port, or the combination of the Ethernet port and the VLAN ID. An IP packet identified by Ethertype 0x0800 in the MAC header is processed as follows:

- The Ethernet MAC header is removed
- The CRC field is removed
- PW label and MPLS tunnel labels are pushed onto the IP PDU.

Other packets, except ARP and link control frames, are dropped. ARP and link control frames are submitted to the control plane.

In some Ethernet networks it is possible that IP packets are encapsulated in a SNAP/SAP format. In this encapsulation the Type/Length field following the DMAC, SMAC contains the length of the PDU. The 802.2 header that follows the MAC header contains 0xAA as DSAP and SSAP, and the UI field is set to 0x03. The five-byte SNAP header follows the 802.2 header; it is set to 0x00-00-00 (OUI field) and 0x0800 (Ethertype for IP). The PE identifies these as IP frames, and processes them as follows:

- The Ethernet MAC header is removed
- The 802.2 header is removed
- The SNAP header is removed
- The CRC field is removed
- PW label and MPLS tunnel labels are pushed onto the IP PDU.

In addition, the Ethernet PDU may also carry a VLAN tag [IEEE 802.1Q]. When present, the Ethertype value of 0x0800 is preceded by the combination of a Tag Protocol Identifier (0x8100) and a 2 byte Tag Control Information field (which includes the 12-bit VLAN ID). The processing of a VLAN-tagged frame at the ingress PE is as follows:

- The Ethernet MAC DA and SA are removed
- The VLAN tag is removed
- The Ethertype for IP (0x0800) is removed
- The CRC field is removed
- PW label and MPLS tunnel labels are pushed onto the IP PDU.

Only IP packets are carried over the IP PW. When these packets arrive at a PE with an egress Ethernet AC they are processed as follows:

- The MPLS tunnel label, if present, is removed
- The PW label is removed
- Ethertype 0x0800 is inserted
- If the Attachment Circuit is represented by a VLAN identifier, the VLAN tag is inserted.
- The remainder of the MAC header is constructed as follows.
    - The source MAC address of the PE is added
    - If the destination IP address is unicast, the destination MAC address field is set to the local CE's MAC address
    - If the destination IP address is multicast, the destination MAC address is constructed based upon [RFC 1112].

6.2 IP MULTI-SERVICE INTERWORKING FOR FRAME RELAY AC

The Frame Relay Attachment Circuit is identified by the Frame Relay port and the DLCI value. An IP packet, as identified by NLPID value 0xCC, is processed at the ingress PE as follows:

- Frame Relay Q.922 and [RFC 2427] headers are removed
- The Frame Check Sequence is removed
- PW label and MPLS tunnel labels are pushed onto the IP PDU.

All other frames, except inverse ARP/ARP and link control frames, are dropped. ARP, inverse ARP, and link control frames are submitted to the control plane.

Only IP packets are carried over the IP PW. When these packets arrive at a PE with an egress Frame Relay AC they are processed as follows.

- The MPLS tunnel label, if present, is removed
- The PW label is removed
- Q.922 header and [RFC 2427] headers are inserted, with NLPID set to 0xCC. The DE and C/R bits in the Q.922 header are set to zero. The FECN and BECN bits in the Q.922 header may be set based upon local conditions. The DLCI field in the Q.922 header is set to the DLCI value associated with the Attachment Circuit
- A Frame Check Sequence is added.

6.3 IP MULTI-SERVICE INTERWORKING FOR ATM AC

The ATM Attachment Circuit is identified by the ATM port and VPI/VCI value. At the ingress PE, ATM cells received on the attachment circuit are reassembled to construct an ATM AAL5 PDU. The AAL5 PDU is then confirmed as containing an IP packet based upon the Attachment Circuit (i.e. VC-based multiplexing) or the LLC+OUI+PID values (where LLC=0xAA-AA-03, OUI=0x00-00-00, PID=0x0800 as defined in [RFC 2684]). The ATM AAL5 PDU is processed at the ingress PE as follows:

- The ATM AAL5 CPCS-PDU trailer is removed
- The RFC 2684 header, if present, is removed
- PW label and MPLS tunnel labels are pushed onto the IP packet.

All other frames, except inverse ARP, ARP and link control frames, are dropped. ARP, inverse ARP, and link control frames are submitted to the control plane.

Only IP packets are carried over the IP PW. When these packets arrive at an egress PE with an ATM AC they are processed as follows:

- The MPLS tunnel label, if present, is removed
- The PW label is removed
- The [RFC 2684] header, with LLC, OUI and PID values as defined above, is inserted if necessary.
- An ATM AAL5 CPCS-PDU trailer is appended
- The ATM AAL5 PDU is segmented into ATM cells.

6.4 IP MULTI-SERVICE INTERWORKING FOR PPP AC

The PPP Attachment Circuit is identified by the PPP port value. At the ingress PE an IP packet received from the PPP AC is identified by a PPP header value of 0x0021. There are no ARP packets. All packets other than IP and NCP packets are dropped. NCP packets are submitted to the control plane. An IP packet is processed at the ingress PE as follows:

- The PPP header (2 bytes) is removed
- The PPP FCS is removed
- PW label and MPLS tunnel labels are pushed onto the IP PDU.

Only IP packets are carried over the IP PW. When these packets arrive at an egress PE with a PPP AC they are processed as follows:

- The MPLS tunnel label, if present, is removed
- The PW label is removed
- A PPP header with value 0x0021 is inserted
- The PPP FCS is appended.

7. Management Plane Interworking

Management plane interworking for defect states and notifications is specified in [OAM-IW].

8. Service Provisioning

The key service provisioning elements in this specification are:

- Configuration of the Attachment Circuit
- Configuration of the pseudo wire
- Configuration related to IP interworking.

The configuration related to the Attachment Circuit and the pseudo wire are common for any Layer 2 interworking service, and are outside the scope of this specification. The configuration elements for IP interworking are identified as follows:

- IP address of the CE – used for the mediation and proxy functions
- Data link address of the CE – used for mediation and data link header manipulation
- Data link address of the PE – used for the proxy function.

The IP interworking service can be provisioned statically by explicitly configuring the above information. The procedures associated with the static configuration are described in Annex B.2.

9. Security

Authentication and data integrity in the exchange of signaling and control messages are issues that are of the utmost concern to many network operators. While this specification does not address these issues in detail, some of the issues surrounding control plane and data plane security for LDP-based pseudo wire setup are addressed in [RFC 4447]. Others may be addressed in a future release of this document or in another normative document issued by the MFA Forum or other Standards Development Organization.

9.1 CONTROL PLANE SECURITY

The integrity and authentication of the signaling plane is described in security section 8.2 of [RFC 4447].

9.2 DATA PLANE SECURITY

Data integrity is a client-to-client connection issue. Section 8.1 of [RFC 4447] recommends various data plane security procedures to be implemented by the network elements.

Annex A

SINGLE-SIDED MODEL

(Normative)

In the single-sided model, only one PE includes an IP IWF. Note that if the heterogeneous Attachment Circuits used are FR and ATM, then an alternative method for providing the IP service is to use [MFA-FR-ATM].

The PE without the IP IWF (e.g. PE1 in Figure 3) treats the PW as a native service and forwards a frame from its local Attachment Circuit to the PW according to the encapsulation rules associated with that PW type. This PE treats the remote PE as if it were delivering a homogeneous Layer 2 VPN. However, there may be instances where OAM behavior is not consistent with a homogeneous Layer 2 VPN. The other PE, which includes the IP IWF (e.g. PE2 in Figure 3) treats the PW as the same data link type as the remote Attachment Circuit, and hence is responsible for replacing the data link encapsulation of the frame received over the PW with the data link encapsulation of its local Attachment Circuit (and vice versa). In essence, the NSP function on PE1 provides the processing required to extend the native service between the AC and the homogeneous PW, while PE2 provides NSP functions for two circuits: the remote Attachment Circuit as represented by the PW, and the local Attachment Circuit.

When a frame is received over the Attachment Circuit or PW, the IP IWF (at PE2) processes it as follows:

- A Layer 2 control frame, such as an LMI frame over Frame Relay, a GMRP or STP frame over Ethernet, etc., is submitted to the control plane for local processing. These are not forwarded on the PW or AC.
- An address resolution (such as ARP or inverse ARP) protocol frame is submitted to the control plane for address learning purposes. These are not forwarded on the PW or AC.
- All other frames are checked to confirm if they contain IP packets or not. IP packets are processed and forwarded with the proper encapsulation on the PW or AC. Non-IP packets are discarded.

Figure 3 shows that an Ethernet PW is extended between PE1 and PE2, and that the IP IWF resides only at PE2. CE2 attaches to PE2 via an ATM, Frame Relay or PPP Attachment Circuit. The PE2 IP IWF treats the Ethernet PW as a logical Ethernet port, or Ethernet port plus VLAN ID, and performs interworking functions such as,

- Learns the remote CE's IP address-to-MAC address association
- Acts as a proxy to the local CE on behalf of the remote CE
- Acts as a proxy to the remote CE on behalf of the local CE.

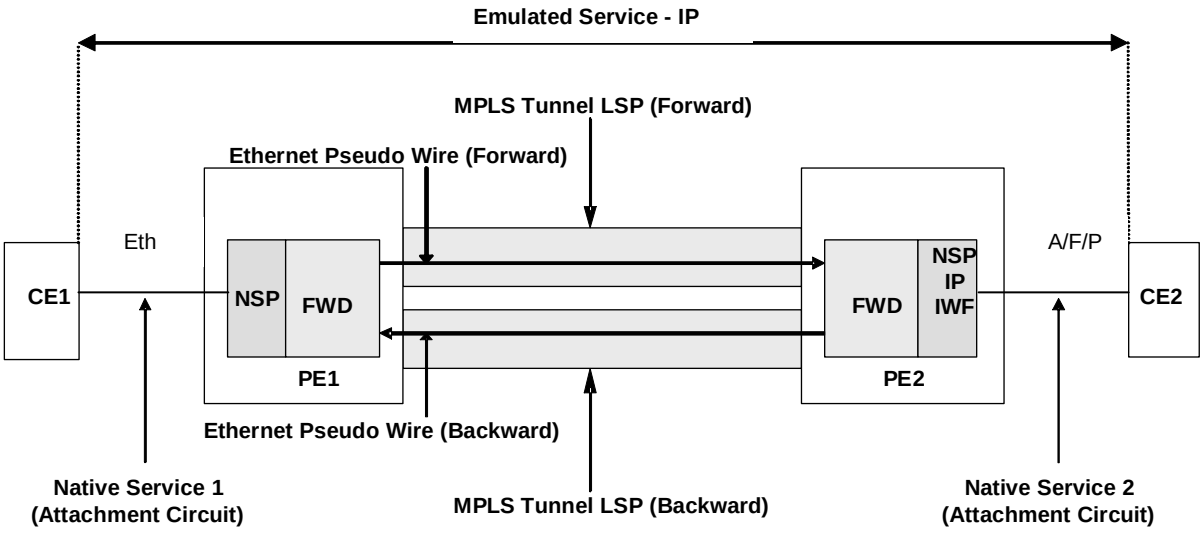


Figure 3: Single-Sided IP Multi-Service Interworking with Ethernet PW

Figure 4 shows a case where one AC is of type ATM, an ATM PW is extended between PE1 and PE2, and the IP IWF resides only at PE2. CE2 attaches to PE2 via an Ethernet, Frame Relay, or PPP link. The PE2 IP IWF treats the ATM PW as a logical ATM VCC and performs interworking functions as described in section 5.

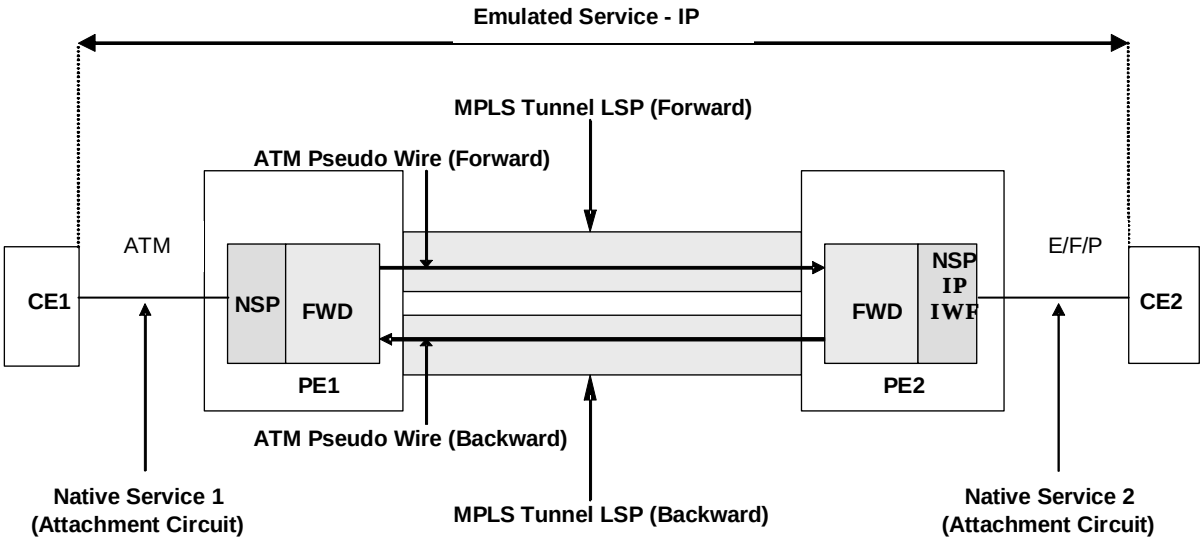


Figure 4: Single-Sided Multi-Service Interworking with ATM PW

The example in Figure 3 would be suitable for customer networks that predominantly support Ethernet access along with a small number of sites with WAN (ATM, Frame Relay, PPP) access, while the example in Figure 4 would be suitable for customer networks that predominantly support FR or ATM access along with a small number of sites with Ethernet access (Figure 4 could be extended to show use of a FR PW when the ACs are of type FR and Ethernet). For these types of network, use of the single-sided model minimizes the number of upgrades required when deploying Multi-service Interworking for IP over MPLS.

Figure 5 illustrates an example of the single-sided model, where IP and non-IP IWFs (e.g. FRF8.2 IWF at PE1 and IP IWF at PE2) coexist in certain configurations. In this case PE2 continues to provide IP interworking for the PW while PE1 remains responsible for L2 interworking between the PW and its local AC; for example, PE1 may execute a FRF8.2 IWF between a FR AC and an ATM PW, while PE2 executes an IP IWF between the ATM PW and an ETH AC. Here PE2 makes no distinction between the cases of (1) PE1 with an ATM AC and ATM PW, and (2) PE1 with a FR AC and ATM PW. In both cases the PW is ATM. If the ATM PW utilizes N:1 Cell Mode encapsulation, there is a SAR function in PE2. If the ATM PW utilizes SDU Mode encapsulation, there is no SAR function in PE2 [RFC 4417].

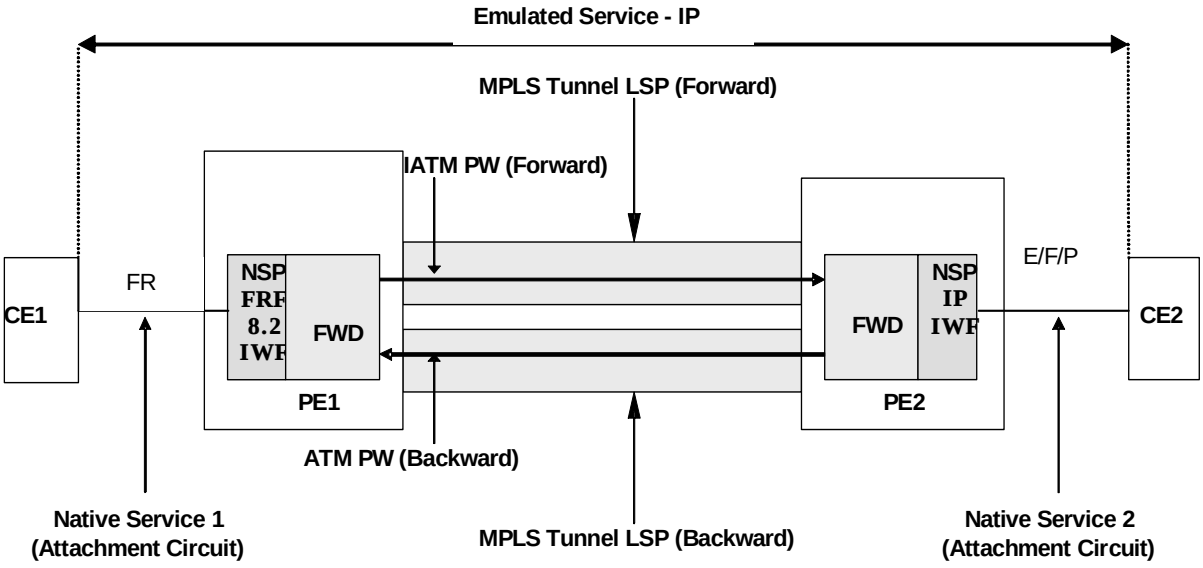


Figure 5: Single-Sided Model with FRF8.2 IWF at one end and IP IWF at the other

Annex B

CONTROL PLANE DESCRIPTION

(Normative)

The following sections describe:

- How a PE device learns the *IP address-to-data link address* association of CE devices on different link types.
- How the CE IP address information is exchanged between PEs using PW signaling through LDP
- How a local PE uses the signaled information to proxy for the remote CE

B.1 LAYER 2 ADDRESS LEARNING

As discussed earlier, an important step in providing IP Multi-service Interworking is for a PE to learn the binding between the IP and Layer 2 addresses of the attached CE device. This specification mandates that each implementation support the administrative configuration of such a binding in the PE. The configuration of customer-specific information in a service provider device (such as a PE) may be unwieldy, but is necessary for interoperability.

B.2 CE IP ADDRESS SIGNALING BETWEEN PES

It is a requirement of this specification that a PE support the manual configuration of both the local CE's IP address and the remote CE's IP address. It is also a requirement of this specification that a PE support manual configuration of the IP PW labels (known as static PW). However, when LDP is used for dynamic setup of the IP PW, this document requires that a PE must support the manual configuration of the IP address of the local CE and communicate this address to the remote PE via signaling as described below.

B.2.1 When to Signal an IP address of a CE

A PE device advertises the IP address of the attached CE only when the encapsulation type of the pseudo wire is IP Layer 2 Transport (the value 0x0000B, as defined in [RFC 4446]) and the IP address of the attached CE has been configured. The LDP Label Mapping message is sent only after the configuration of the IP address of the attached CE.

If the two CE devices are attached to the same PE, for example where one CE is connected to an Ethernet port and the other to a Frame Relay port, the IP addresses are learned in the same manner described above. However, since the CE devices are local, the distribution of IP addresses for these CE devices is a local step.

B.2.2 LDP-Based Distribution

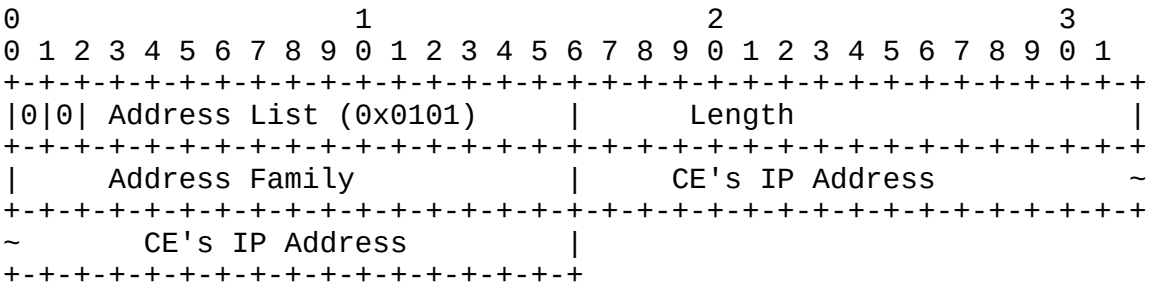
The PW control protocol [RFC 4447] uses Label Distribution Protocol (LDP) transport to exchange PW FEC information in the Label Mapping message in Downstream Unsolicited mode. The PW FEC comes in two flavors: PW ID and Generalized ID FEC elements; there are some common fields between them. The discussions below refer to these common fields for IP PW (note that RFC 4447 refers to IP PW as PW of type "IP Layer 2 Transport").

In addition to PW FEC, this specification defines an IP address TLV that must be included in the optional parameter field of the Label Mapping message when advertising the PW FEC for IP Layer 2 Transport (i.e. IP PW). The use of optional parameters in the Label Mapping message to extend the attributes of the PW FEC is specified in the [RFC 4447].

When processing a received PW FEC, the PE matches the PW ID and PW type with the locally-configured PW ID to determine if the PW FEC is of type IP Layer 2 Transport. If there is a match, it further checks for the presence of an IP address TLV in the optional parameter field. If absent, a Label Release message is issued.

We use the Address List TLV as defined in the LDP Specification [RFC 3036] to signal the IP address of the local CE. This IP address TLV must be included in the optional parameter field of the Label Mapping message.

Encoding of the IP Address TLV is:



Length
The Length is set to 6 bytes; 2 bytes for address family and 4 bytes of IP address.

Address Family
Two-octet quantity containing a value from the ADDRESS FAMILY NUMBERS section of Assigned Numbers [RFC1700], which encodes the address contained in the Address field.

CE's IP Address
IP address of the CE attached to the advertising PE. The encoding of the individual address depends upon the Address Family.

The following address encoding is valid for this specification:

Address Family	Address Encoding
IPv4 (0x0001)	4 octet full IPv4 address

The IP address field is set to the IP address of its local CE device.

B.3 LAYER 2 ADDRESS NOTIFICATION AND PROXY FUNCTION

Once a PE has learned the IP and Layer 2 address association of the remote CE, the PE communicates this information to the locally-attached CE. The PE either initiates an address resolution request to, or responds to an outstanding request from the local CE. The procedures used by the PE depend on the local AC type.

It is important to note that even when a PE has all the required information about the local CE and the remote CE, it still needs to perform proxy functions to communicate the information about the remote CE to the local CE. In some cases, the local CE may have been configured with the information of the remote CE. However, since the PE has no means to know the configuration status of the local CE, it must attempt to provide this information about the remote CE.

B.3.1 Ethernet Data Link

Once the PE learns the remote CE's IP address (as described above), the PE may choose to generate an unsolicited ARP message to notify the Ethernet CE about the binding of the remote CE's IP address with the PE's own MAC address.

Whenever the Ethernet CE generates an ARP request, the PE must proxy an ARP response using its own MAC address as the source hardware address and remote (i.e. Frame Relay/ATM/PPP) CE's IP address as the source protocol address. The PE must respond only to those ARP requests whose destination protocol address matches the remote CE's IP address.

B.3.2 Frame Relay Data Link

When a PE receives information about the remote CE from the remote PE, the proxy function takes the following actions based on the state of the corresponding AC.

- If the DLCI is inactive, the PE activates the DLCI via LMI and issues an inverse ARP request.
- If the DLCI is active and the PE has not received an inverse ARP request from the local CE, it issues an inverse ARP request
- If the DLCI is active and the PE has already received an inverse ARP request, it issues an inverse ARP response.

In the inverse ARP request or inverse ARP response, the PE includes the remote CE's IP address as the source protocol address.

B.3.3 ATM Data Link

When a PE receives information about the remote CE from the remote PE, the proxy function takes the following actions based on the state of the corresponding AC.

- If the ATM interface is active and the VCC is of type PVC, VCC is active. If VCC is of type SVC the VCC becomes active when the signaling completes. When VCC becomes active, PE issues an inverse ATM ARP request.
- If the ATM VCC is active and the PE has not received an inverse ATM ARP request from the local CE, it issues an inverse ARP request
- If the ATM VCC is active and the PE has already received an inverse ATM ARP request, it issues an inverse ATM ARP response.

In the inverse ARP request or inverse ARP response, the PE includes the remote CE's IP address as the source protocol address.

B.3.4 PPP Data Link

When a PE receives information about the remote CE from the remote PE, it notifies the local CE of the IP address of the remote CE by sending a Configure-Request and setting the IP-Address option to the IP address of the remote CE.

END OF DOCUMENT