Next-Generation Packet-Based Transport Networks (PTN)

By Reza Vaez-Ghaemi, Ph.D.

Introduction

Competition for market share among telecom, cable/multimedia service (MSO), and mobile operators further fuels the need for network investments. High-speed Internet services, IP video, and smartphone applications significantly increase the average revenue per user. For these high bandwidth services to become profitable, the cost per bit must drop drastically, which packet and optical technologies make possible to a degree. This white paper focuses on packet-based transport technologies.

Packet-based technologies have been deployed in Enterprise environments for many years, because Ethernet enables a cost-effective networking of computing equipment and peripherals in local area networks (LANs). However, Ethernet lacks key attributes necessary for deployment in wide area networks (WANs), which are characterized by large numbers of subscribers and services. It is mandatory to clearly differentiate and process various services. Some services such as voice and video are more sensitive to delay and/or delay variation than others. Fast restoration of services is another key attribute of a Carrier-grade technology. Also, WANs are large and may consist of multiple operators' networks; thus it is very critical to be able to manage faults and monitor performance for these networks.

This white paper focuses on next-generation packet-based transport networks (PTN) with a focus on emerging Multiple Protocol Labeling Switching (MPLS) technologies.

IP/MPLS

MPLS was originally developed by the Internet Engineering Task Force (IETF) to deliver a cost-efficient way of routing traffic in high-performance core networks. However, MPLS has since found strong application in service providers' core networks and as a platform for data services, such as Layer 3 or Layer 2 virtual private networks (VPN).

MPLS is essentially a labeling system designed to accommodate multiple protocols. Label Switched Paths (LSPs) are used to define the paths of packets in the network so that a connection-oriented mode is introduced into a connectionless network. The use of MPLS labels enables routers to avoid the processing overhead of deeply inspecting each data packet and performing resource-intensive route lookup operations based upon IP addresses.

An LSP is a sequence of MPLS nodes (as shown in Figure 1) that connects adjacent routers within the MPLS network. The nodes known as Label Switched Routers (LSR) switch traffic based on the MPLS label carried between the Layer 2 and Layer 3 headers.
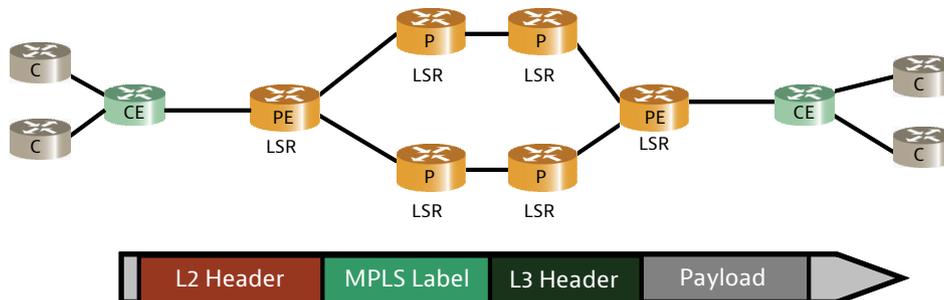

Figure 1. MPLS Networks

Layer 3 VPNs

Layer 3 VPNs, or RFC 2547bis VPNs, have been the most widely deployed MPLS technology. They use virtual routing instances (VRI) to create a separate routing table for each subscriber and use Border Gateway Protocol (BGP) to establish peering relations and signal the VPN-associated labels with each of the corresponding provider edge (PE) routers. This function results in a highly scalable network, because core (P) routers do not require any information about the VPNs.

Layer 3 VPNs are useful when subscribers want Layer 3 connectivity and would prefer to outsource their routing operations to an external provider to ensure use of a variety of Layer 2 interfaces on either side of a VPN. For example, Site A can use an Ethernet interface while Site B uses an asynchronous transfer mode (ATM)-interface; however, Sites A and B are part of a single VPN. It is relatively simple to implement multiple topologies with router filtering, including hub and spoke or full mesh as described here:

- Hub and spoke—The central site is configured to "learn" all the routes from the remote sites, while the remote sites are restricted to "learn" routes only from the central site.

- Full mesh topologies would result in all the sites having the ability to "learn" or import routes from every other site.

Service providers are currently providing up to several hundred VPNs each of which can contain as many as 1000 sites. A wide variety of routing protocols are available to deploy on the subscriber access link, such as a (customer edge) CE-to-PE link. Most VPNs have been deployed with static routes followed by BGP routing. Layer 3 VPNs offer advanced capabilities providing for hierarchical VPNs, allowing service providers to provide connectivity across multiple administrative networks. Layer 3 VPNs may also be referred to as BGP VPN, although BGP can also be used in conjunction with a Layer 2 VPN as described in the next section.

Layer 2 VPNs

Layer 2 VPNs, also known as Martini VPNs, refer to the ability of service provider customers to provide Layer 2 circuits over an MPLS-enabled IP backbone. It is important to understand the three major components of Layer 2 VPNs:

1. Layer 2 Transport over MPLS—Layer 2 circuit carried transparently over a MPLS-enabled IP backbone (also known as AToM, or any transport over MPLS)

2. Virtual Private Wire Service (VPWS)

3. Virtual Private LAN Service (VPLS)

The predominant Layer 2 circuits include Ethernet, ATM, frame relay, Point-to-Point Protocol (PPP), and high-level data link control (HDLC). AToM and Layer 3 VPNs are based on the same concepts, but AToM uses a directed Label Distribution Protocol (LDP) session to distribute the virtual circuit (VC) labels (analog to BGP VPN label). Consequently, core routers are not required to have the knowledge of the circuit on a per-subscriber basis, resulting in a very scalable architecture. Prior to the availability of AToM, service providers had to build different

networks to provide Layer 2 connectivity, such as an ATM and a frame relay network, resulting in increased operational and capital expenses. Layer 2 VPNs on MPLS now enable them to combine these different networks, creating significant savings in terms of operational and capital expenses.

VPWS (Pseudowire)

Pseudowire (PW) emulates the operations of a Layer 2 point-to-point service over a Packet-Switched Network (PSN). The emulated service may be ATM, frame relay, synchronous optical network/synchronous digital hierarchy (SONET/SDH), or Ethernet. The underlying PSN can be Ethernet, MPLS, Internet Protocol (IPv4 or IPv6), or Layer 2 Tunneling Protocol Version 3 (L2TPv3).

VPWS, or PW, specifications include:

- PWE3 (IETF RFC 3985): Pseudowire Emulation Edge-to-Edge (PWE3) describes the emulation of services such as frame relay, ATM, Ethernet, Time Division Multiplexing (TDM), and SONET/SDH over PSNs using IP or MPLS. It presents the architectural framework for PWs, defines terminology, and specifies the various protocol elements and their functions. PWE3 is a mechanism that emulates the essential attributes of a telecommunications service (such as a T1 leased line or frame relay) over a PSN.

- SATOP (IETF RFC 4553): The Structure Agnostic TDM over Packet (SATOP) document describes the method for encapsulating TDM bit-streams (T1, E1, T3, and E3) as PWs over PSN. It addresses only structure-agnostic transport, that is the protocol completely disregards any structure that may possibly be imposed on these signals, in particular the structure imposed by standard TDM framing (G.704). This emulation is referred to as emulation of unstructured TDM circuits in RFC 4197 and suits applications where the PEs routers have no need to interpret TDM data or to participate in the TDM signaling.

- CESoPSN (IETF RFC 5086): The Circuit Emulation Service over Packet-Switched Networks (CESoPSN) document describes a method for encapsulating structured (NxDS0) TDM signals as PWs over PSNs. In this regard, it complements similar work for structure-agnostic emulation of TDM bit-streams [RFC 4553]. Emulation of NxDS0 circuits saves PSN bandwidth and supports DS0-level grooming and distributed cross-connect applications. It also enhances resilience of CE devices to the effects of packet loss in the PSN.

- TDM over MPLS (ITU Y.1413): International Telecommunications Union (ITU) Recommendation Y.1413 focuses on the required functions for network interworking between TDM and MPLS; specifically, the user plane interworking mechanisms and procedures for transport. In particular it specifies a list of requirements, interworking scenarios, and interworking encapsulation formats and semantics for TDM-MPLS network interfaces. Given that TDM connections are inherently point to point, this interworking defines a single connection between two interworking functions (IWFs). This Recommendation only addresses TDM rates up to and including T3 and E3.

Note that an LSP may carry traffic corresponding to several customers between the same peering routers. Also, multiple LSPs may exist on the same physical wire thus requiring a means of multiplexing and de-multiplexing traffic.

The PW encapsulation technique uses two labels to enable multiplexing/demultiplexing traffic—the outer label is used for LSP tunnel identification and the second (inner) label is used for customer identification, which technically is called the VC. These two labels are inserted as part of the MPLS header after the Ethernet header in the outgoing frame1. For the payload, the entire incoming Ethernet frame from the customer is encapsulated within the payload of the outgoing MPLS frame, thereby increasing the resulting size of the frame to be transported in the MPLS network. The adjacent routers along an LSP must agree on the tunnel label used; which is accomplished as part of MPLS signaling, for example, either by LDP or by Resource Reservation Protocol with Traffic Engineering (RSVP-TE), if traffic engineering is desired.

VPLS

Switched Ethernet network architectures have proven to be successful in delivering high-performance, low-cost Layer 2 VPN multipoint services. However, as the size of these switched Ethernet networks continue to grow, the limitations on the scalability of this architecture have become increasingly apparent, including:

• limited VLAN address space per switched Ethernet domain

• scalability of spanning tree protocols (IEEE 802.1d) for network redundancy and traffic engineering

• Ethernet MAC (media access control) address learning rate, which is important to minimize broadcast traffic resulting from unknown MAC addresses.

To address the limitations of both MPLS L3 VPNs and Ethernet switching, innovations in network technology for delivering multipoint connectivity services have led to the development of a new technology known as VPLS, which is an MPLS application used to provide multipoint-to-multipoint Layer 2 VPN services. VPLS allows geographically dispersed sites to share an Ethernet broadcast domain by connecting each site to an MPLS-based network. In contrast to Layer 2 MPLS VPNs, which allow only point-to-point Layer 2 MPLS tunnels, VPLS allows a full mesh of sites or any-to-any (multipoint) connectivity that is achieved by setting up point-to-point PW between a node and every other node in the VPLS instance.

Because the two customers are distinct, each customer belongs to a unique VPLS instance as Figure 2 shows. Each border node in the VPLS network is called a PE router, because that router denotes the demarcation point of a service provider's network. The switch/router on the customer premises communicating with the PE router is designated the CE device. The connection between the PE and CE routers is also referred to as the attachment circuit (AC). The VPLS network consists of a full mesh of PWs established between the PE routers in the VPLS network. The LSPs that form a PW are unidirectional in nature. Therefore, a pair of LSPs is created, one in each direction, between a pair of PE routers.
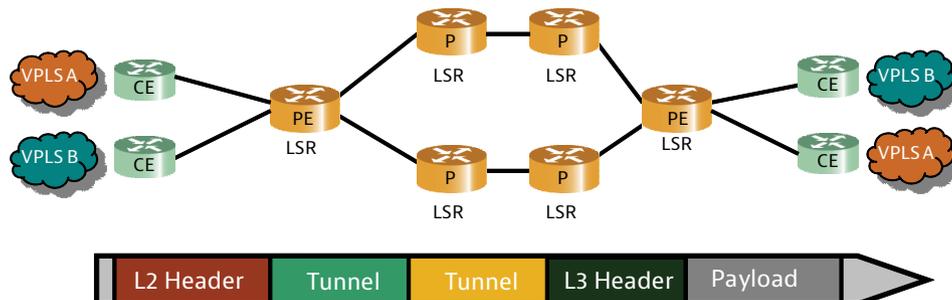

Figure 2. VPLS Networks

Hierarchical VPLS

The previously mentioned VPLS architecture involves setting up a full mesh between all participating PE routers. Scalability suggests limiting the number of nodes added to this full mesh. An alternative architecture involves a hub-and-spoke where service provides place "smart" multi-tenant unit (MTU) switches at the edge of the network that are connected to the PE router via a spoke connection as shown in Figure 3. The PE thus acts as a hub device aggregating traffic from multiple MTU devices.
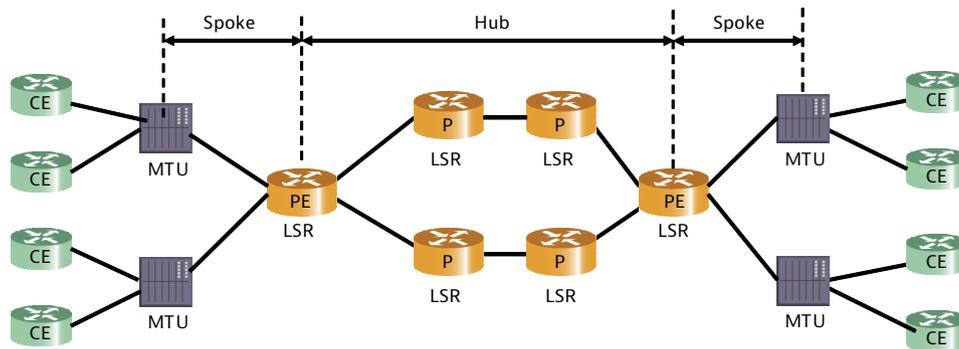

Figure 3. H-VPLS Architecture

In terms of its capabilities, the MTU is a switch that can perform functions such as Q-in-Q (802.1q tunnel tags). Therefore, traffic from each customer can be assigned a unique provider VLAN tag and sent over a Q-in-Q tunnel to the PE device. The PE router can then map the provider VLAN tag (outer VLAN tag) to a distinct VPLS instance. Alternatively, a Virtual Leased Line (VLL) can also be deployed between the MTU and the PE router. The hub-and-spoke scheme has the advantage of separating manageability of the core network from that of the network edge. In addition, the MTU can also be used to switch local traffic without sending traffic to the PE.

MPLS-TP

IP/MPLS technology is mature enough for various application scenarios, because it provides powerful traffic engineering capabilities. However, its equipment and operational costs have prevented service providers from wide-scale deployment of it in metro and aggregation networks. Transport maintenance engineers and technicians are familiar with SONET/SDH operations and maintenance procedures and would require additional training learn the IP/MPLS planning and administration. Additionally, IP/MPLS lacks key SONET/SDH-like operations, administration, and maintenance functions for performance monitoring and fault management.

MPLS-TP is a transport profile of MPLS whose definition has been driven by the IETF. It is designed for use as a network layer technology in transport networks. Its design will be a continuation of the work started by the transport network experts of the ITU-T, specifically SG15. It is a connection-oriented packet-switched application that will offer a dedicated MPLS implementation by removing features that are irrelevant to connection-oriented applications and adding mechanisms that provide support for critical transport functionality.

MPLS-TP simplifies the application scenarios of MPLS with decreased equipment, operation, and maintenance costs. The data plane is separated from the control plane, which leads to higher network stability, reliability, and flexibility. With a strong OAM and protection switching function, the MPLS-TP-based PTN could achieve the same reliability and resilience level as SDH/next-generation SDH (NG-SDH).

Figure 4 shows the essential feature enhancements of MPLS-TP which are:

• MPLS forwarding plane restrictions and enhancement

• Control Plane: static or dynamic

• Enhanced OAM functionality

MPLS (pre-MPLS-TP)   MPLS-TP enhancement

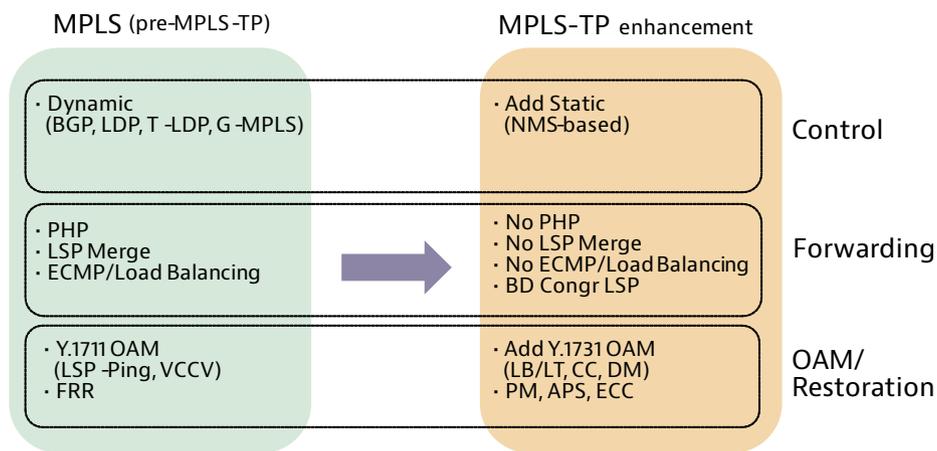| MPLS (pre-MPLS-TP)	MPLS-TP enhancement	
· Dynamic (BGP, LDP, T-LDP, G-MPLS)	· Add Static (NMS-based)	Control
· PHP · LSP Merge · ECMP/Load Balancing	· No PHP · No LSP Merge · No ECMP/Load Balancing · BD Congr LSP	Forwarding
· Y.1711 OAM (LSP-Ping, VCCV) · FRR	· Add Y.1731 OAM (LB/LT, CC, DM) · PM, APS, ECC	OAM/ Restoration

Figure 4. MPLS-TP enhancements

Forwarding Plane

To improve the network transport and management capabilities, several elements of the MPLS are adjusted in MPLS-TP.

• No LSP Merge: IP/MPLS-based architectures allow LSP merge for packets traveling the same path, which helps improve the efficiency in the transport of the traffic but results in loss of LSP headend information. This headend information is critical for providing enhanced OAM capabilities that service providers desire. MPLS-TP does not allow LSP merges leading to enhanced end-to-end OAM capability.

• No Penultimate Hop Popping (PHP): Popping of the outer MPLS label, allowed in IP/MPLS-based networks, causes loss of context at the adjacent P-router. The outer MPLS label is used as the OAM identifier (packet context) for the enhanced OAM capabilities offered by MPLS-TP, but is not allowed in MPLS-TP networks. Also, PHP does not offer any apparent benefits for Layer 2 VPNs.

• No Load Balancing/Equal Cost Multiple Path (ECMP): ECMP is a forwarding mechanism for routing packets along multiple paths of equal cost to achieve nearly equal distribution of link load sharing. However, this mechanism leads to troubleshooting problems, because the customer's actual path varies between the packets.

• Bidirectional Congruent LSP: This element allows MPLS-TP-based networks to emulate classical transport networks—transmit and receive follow the same path through the network. It simplifies the operations for bidirectional connectivity, which also improves jitter performance because of reduced packet delay variance. It also simplifies troubleshooting and improves using the OAM enhancements provided by MPLS-TP.

Control Plane

Within the context of MPLS-TP, the control plane is the mechanism used to set up an LSP automatically across a packet-switched network domain. The use of a control plane protocol is optional in MPLS-TP. Some operators may prefer to configure the LSPs and PWs using a Network Management System in the same way that it would be used to provision SONET. In this case, no IP or routing protocol is used. Conversely, it is possible to use a dynamic control plane with MPLS-TP so that LSPs and PWs are set up by the network using Generalized (G)-MPLS and Targeted Label Distribution Protocol (T-LDP), respectively. G-MPLS is based on the traffic engineering (TE) extensions to MPLS (MPLS-TE). It may also be used to set up the OAM function and define recovery mechanisms. T-LDP is part of the PW architecture and is widely used today to signal PWs and their status.

OAM and Survivability

The functions of OAM and survivability for MPLS-TP networks are intended to reduce network operational complexity associated with network performance monitoring and management, fault management, and protection switching. These functions are required in order to operate without any IP layer functions. One of the goals of MPLS-TP OAM is to provide the tools needed to monitor and manage the network with the same attributes offered by legacy transport technologies. For example, the OAM is designed to travel on the exact same path as the data. In other words, MPLS-TP OAM monitors PWs or LSPs. It provides an enhancement to IP/MPLS OAM capabilities provided in ITU Y.1711 that deliver basic fault monitoring and troubleshooting features at the LSP level, such as virtual circuit connectivity verification (VCCV) and LSP Ping.

MPLS-TP OAM operates in the context of Maintenance Entities (MEs) that are a relationship between two points of a point-to-point transport path or a root and a leaf of a point-to-multipoint transport path to which maintenance and monitoring operations apply. These two points are called Maintenance Entity Group (MEG) End Points (MEPs). In between these two points, zero or more intermediate points may exist, called Maintenance Entity Group Intermediate Points (MIPs). Another OAM functional component is referred to as MEG, which is a collection of one or more MEs that belong to the same transport path and that are maintained and monitored as a group.

MEPs activate and control all of the OAM functionality for the MEG. A MEP is capable of originating and terminating OAM messages for fault management and performance monitoring. These OAM messages are encapsulated into an OAM packet using the Generic Associated Channel (G-Ach) as defined in RFC 5586; in this case, the G-Ach message is an OAM message and the channel type indicates an OAM message. A MEP terminates all the OAM packets it receives from the MEG to which it belongs. The MEG to which the OAM packet belongs is inferred from the MPLS or PW label or, in case of MPLS-TP section, the MPLS-TP port the OAM packet has been received with the Generic Alert Label (GAL) at the top of the label stack.

Two important components of the OAM mechanisms (shown in Figure 5) are the G-Ach and the GAL. As their names indicate, they allow operators to send any type of control traffic into a PW or an LSP. Assigned the label value 14, the GAL is used in MPLS to indicate an OAM message, such as LSP Ping. MPLS-TP enhances this concept with a new label value of 13 which is used to flag a G-Ach.
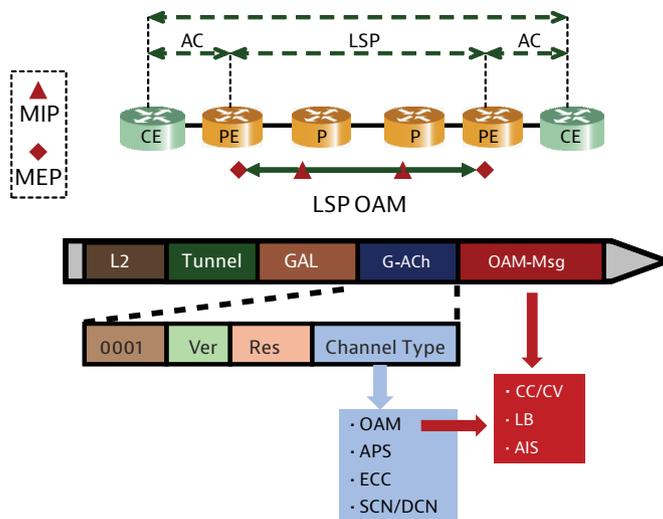

Figure 5. MPLS-TO OAM

The G-ACh is used in both PWs and MPLS-TP LSPs and is very similar to the associated channel as defined by RFC 4385. The G-ACh acts as a container or channel that runs on the PW and carries OAM messages. For example, VCCV may be sent over an associated channel to monitor whether the PW is available. The associated channel is a generic function, such that it can also run over LSPs, and is capable of carrying user traffic, OAM traffic, and management traffic over either a PW or an LSP. It can also carry Automatic Protection Switching (APS) information and Data Communications Channel (DCC), Signaling Communication Channel (SCC), Management Communication Channel (MCC) management traffic, and others.

The G-Ach begins with the bits 0001 that indicates a non-IP/PW packet. The channel type lists the message type that is embedded in the appending bytes indicated as an OAM, APS, Embedded Communications Channel (ECC), or other message type. If the channel type refers to an OAM message, the following bytes will deliver the actual OAM message, such as continuity verification/check, and loopback. It is important to note that this generic construct defined for MPLS-TP will be reused by IP/MPLS, which will provide a very extensive set of OAM tools and support FCAPS functions for end-to-end management.

Provider (Backbone) Bridging

The initial Ethernet specification was very limited in terms of service differentiation and quality of service (QoS) features. The IEEE 802.1q delivered a VLAN tag that is used to provide different levels of QoS to up to 4,096 instances/services inside an Ethernet signal. While useful for smaller Ethernet networks, this enhancement was insufficient for providers' aggregation networks that encompass larger numbers of subscribers and services. IEEE 802.1ad added service providers' VLAN (S-VLAN) tags to the Ethernet frame, allowing providers to separate their VLANs from their customers' VLAN (C-VLAN), and offer differentiation in larger provider bridged networks. Despite these enhancements, provider bridging has one major limitation: it uses the customer MAC address for switching operations. If the customer MAC address is unknown, the provider bridges will flood their ports with the unknown address. This flooding mechanism will drastically rise in larger networks and reduce the efficiency of the transport network.

To scale the provider bridging for larger backbone networks, IEEE initiated the 802.1ah Provider Backbone Bridging (PBB) working group. Adding a backbone MAC address and backbone VLAN (B-TAG) enables isolating the provider network from the customer traffic. All switching and differentiation decisions are based on backbone MAC/TAG fields. For further service differentiation, a service identification tag (I-TAG) was added to the B-VLAN, C-VLAN, and S-VLAN, as shown in Figure 6.
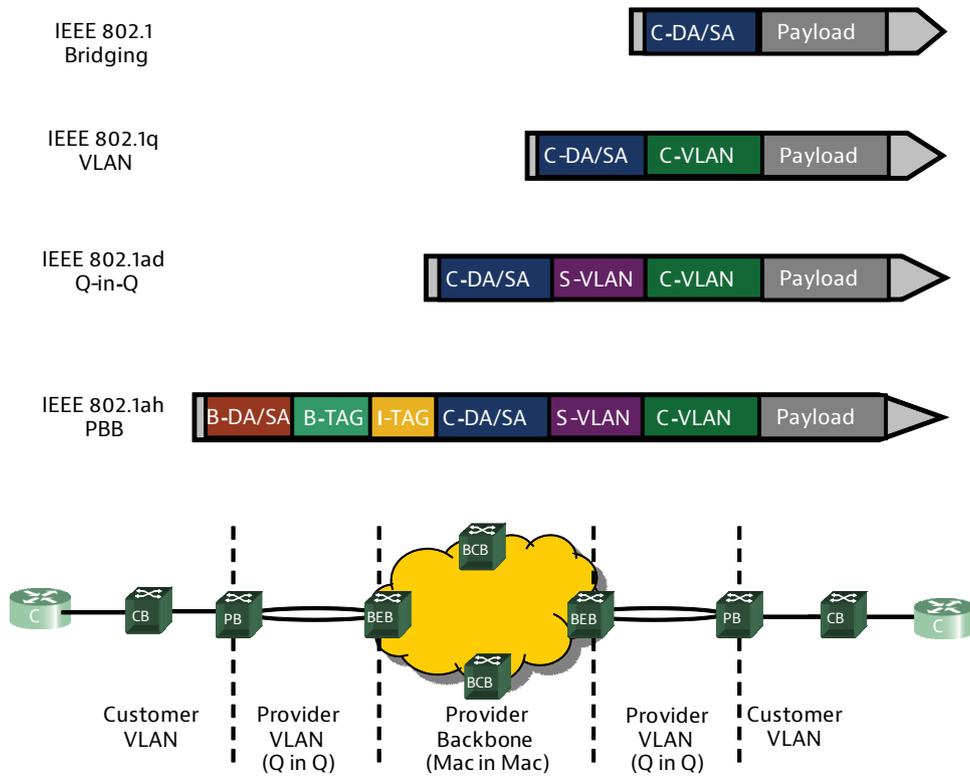

Figure 6. PBB/PBB-TE Network

Incoming customer traffic arrives at a provider backbone edge bridge (BEB) where a backbone MAC address, B-TAG, and optionally an I-TAG are attached to the customer frame. The provider backbone core bridges (BCB) switch and differentiate traffic based on the outer labels (backbone labels). The egress BEB removes the outer label and delivers the original customer traffic to the customer bridge (CB).

References

1. Reza Vaez-Ghaemi, "Timing and Synchronization in Packet Networks," June 2010.

2. Reza Vaez-Ghaemi, "Ethernet OAM Test Applications,"
http://www.jdsu.com/product-literature/ethoam_wp_acc_tm_ae.pdf,January 2009.

3. Reza Vaez-Ghaemi, "Next-Generation Optical Transport Networks," October 2010.

4. M. Bocci et al., "A Framework for MPLS in Transport Networks," Work in Progress, October 2009.

5. H. van Helvoort et al., "A Thesaurus for the Terminology used in Multiprotocol Label Switching Transport Profile (MPLS-TP) Drafts/RFCs and ITU-T's Transport Network Recommendations," Work in Progress, June 2009.

6. Busi et al., "MPLS-TP OAM Framework," Work in Progress, December 2009.

7. M. Vigoureux et al., "Requirements for OAM in MPLS Transport Networks," Work in Progress, August 2009.

8. N. Sprecher et al., "MPLS-TP OAM Analysis," Work in Progress, November 2009.

9. Busi et al., "MPLS-TP OAM based on Y.1731," Work in Progress, March 2010.

10. IETF "Application of PWE3 to MPLS Transport Networks," June 2009.

11. IEEE 802.1D, *Media Access Control Bridges*, June 2004.

12. IEEE 802.1Q, *Virtual Bridged Local Area Networks*, November 2006.

13. IEEE 802.1ad, *Virtual Bridged Local Area Networks Amendment 4: Provider Bridges,* December 2005.

14. IEEE 802.1ah, *Virtual Bridged Local Area Networks Amendment 6: Provider Backbone Bridges,* June 2008.

15. IEEE 802.1Qay, *Draft Standard for Local and Metropolitan Area Networks—Virtual Bridged Local Area Networks—Amendment: Provider Backbone Bridge Traffic Engineering,* May 2007.

16. IEEE 802.1ag, *Draft 8.1 Virtual Bridged Local Area Networks—Amendment 5: Connectivity Fault Management,* September 2007.

17. IEEE 802.3ah, Part 3: *Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications Amendment: Media Access Control Parameters, Physical Layers, and Management Parameters for Subscriber Access Networks,* 2004.

18. IETF RFC 3031, "Multiprotocol Label Switching Architecture," January 2001.

19. IETF RFC 3032, "MPLS Label Stack Encoding," January 2001.

20. IETF RFC 3270, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services," May 2002.

21. IETF RFC 3429, "Assignment of the 'OAM Alert Label' for Multiprotocol Label Switching Architecture (MPLS) Operation and Maintenance (OAM) Functions," November 2002.

22. IETF RFC 3916, "Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)," September 2004.

23. IETF RFC 3985, "Pseudo-Wire Emulation Edge-to-Edge (PWE3) Architecture," RFC 3985, March 2005.

24. IETF RFC 4448, "Encapsulation Methods for Transport of Ethernet over MPLS Networks," April 2006.

25. IETF RFC 4553, "Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)," June 2006.

26. IETF RFC 5085, "Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires," December 2007.

27. IETF RFC 5317, "Joint Working Team (JWT) Report on MPLS Architectural Considerations for a Transport Profile," February 2009.

28. IETF RFC 5586, "MPLS Generic Associated Channel," June 2009.

29. IETF RFC 5654, "Requirements of an MPLS Transport Profile," September 2009.

30. ITU-T Recommendation G.7712, "Architecture and Specification of Data Communication Network," June 2008.

31. ITU-T Recommendation G.8101, "Terms and Definitions for MPLS-TP," December 2006.

32. ITU-T Recommendation G.8110.1, "Architecture of MPLS-TP Layer Network," November 2006.

33. ITU-T Recommendation G.8112, "Interfaces for the MPLS-TP Hierarchy," October 2006.

34. ITU-T Recommendation G.8121, "Characteristics of MPLS-TP Network Equipment Functional Blocks." March 2006.

35. ITU-T Recommendation G.8131, "MPLS-TP Linear Protection," February 2007.

36. ITU-T Recommendation G.8132, "MPLS-TP Ring Protection," November 2007.

37. ITU-T Recommendation G.8151, "Management Aspects of the MPLS-TP Network Element," October 2007.

VIAVI

viavisolutions.com