Chapter 1

Introduction and Overview

1.1 INTRODUCTION

Ethernet has been the dominant technology for local area networks (LANs) for many years. With the introduction of circuit concept over Ethernet by the IEEE and MEF specifications, Ethernet is becoming a key technology for Metropolitan Area Networks (MANs) and Wide Area Networks (WANs).

This book makes an attempt to describe various aspects of data networks and services based on carrier Ethernet, pseudowires (PWs), transport multiprotocol label switching (T-MPLS), multiprotocol label switching transport profile (MPLS-TP), and virtual private local area network service (VPLS).

Chapter 2 describes the basics of Ethernet, such as protocol stack, bridges, switches, and hubs. Chapters 3, 4, and 5 cover the key techniques that are being used in building carrier class Ethernet networks and services, namely, synchronization, PWs, and protection, respectively.

Chapter 6 begins describing Carrier Ethernet network architecture and services that are currently deployed in the industry and evolving. It is followed by Chapters 7 and 8, describing traffic management and Ethernet Operations, Administrations, and Maintenance (OAM) capabilities of Carrier Ethernet[1], respectively.

Chapter 9 is devoted to circuit emulation services (CES) because of its complexity. For the same reason, Chapter 10 is devoted to Ethernet local management interface (ELMI).

Addressing scalability of Carrier Ethernet is being questioned, despite of the availability of S-Tag and C-Tag combination. Provider Backbone Bridges (PBBs) and Provider Backbone Transport (PBT) resolve this scalability issue. Chapter 11 describes these technologies in detail.

Three technologies, namely, T-MPLS, MPLS-TP, and VPLS can compete or work with Carrier Ethernet in forming data networks. Chapters 12, 13, and 14 describe them in detail.

[1]In this book, Carrier Ethernet and Metro Ethernet are used interchangeably.

Networks and Services: Carrier Ethernet, PBT, MPLS-TP, and VPLS, First Edition. Mehmet Toy.

1.2 BASIC ETHERNET

Ethernet is physical layer LAN technology invented in 1973. Since its invention, Ethernet has become the most widely used LAN technology because of its speed, low cost, and relative ease of installation. Ethernet PHY supporting rates from 10 Mbps to 100 Gbps are available.

Chapter 2 describes CSMACD (Carrier Sense, Multiple Access, Collision Detect), protocol stack, bridges, switches, and hubs. Toward the end of the chapter, Ethernet frame formats defined by IEEE 802.3 and 802.2 are described.

1.3 SYNCHRONIZATION

The TDM (time-division multiplexing) networks transporting multiplexed voice and data traffic require highly accurate frequency information to be recoverable from their physical layers. Typically, a TDM circuit service provider (SP) will maintain a timing distribution network, providing synchronization traceable to a primary reference clock (PRC) that is compliant with clock quality. A reference timing signal of suitable quality is distributed to the network elements (NEs) processing the application. One approach is to follow a distributed PRC strategy. An alternative approach is based on a master–slave strategy.

Synchronization in Metro Ethernet Networks (MEN) is required mainly because of mobile backhaul and bandwidth explosion in mobile networks. Frequency, phase, and time synchronizations are needed. Chapter 3 describes two main synchronization techniques, Precision Time Protocol (1588v2) and Synchronous Ethernet (SyncE).

1.4 PSEUDOWIRES

Pseudowire Emulation Edge-to-Edge (PW3) is a mechanism used to emulate telecommunication services across packet-switched networks (PSNs), such as Ethernet, IP (internet protocol), or MPLS. The emulated services are mostly T1/T3 leased lines, frame relay, Ethernet, and minimize operating costs of service providers (SPs). PWs encapsulate cells, bit streams, and protocol data units (PDUs) and transport them across Carrier Ethernet Virtual Circuits (EVCs), IP, or MPLS tunnels. The transportation of encapsulated data usually require managing the sequence and timing of data units to emulate services such as T1/T3 leased lines and asynchronous transfer mode (ATM).

Chapter 4 describes various aspects of PWs, including its architecture, protocol stack, frame forwarding, control plane, resilience, and OAM.

1.5 PROTECTION

Services such as broadcast video, voice over IP, video on demand require five 9 availability. When failures occur, they are not supposed to be noticed by

the subscriber. Automatic protection switching guarantees the availability of resources in the event of failure and ensures that switchover is achieved in less than 50 ms. The 50 ms is currently being debated in the industry. In fact, Metro Ethernet Forum mandates 500 msec protection switching at UNI (user network interface) or ENNI (external network–network interface). The smaller switching time increases the cost of the equipment.

Chapter 5 describes various protection techniques, including Linear Protection, Ring Protection, and Link Aggregation Group (LAG)/Link Aggregation Control Protocol (LACP).

1.6 CARRIER ETHERNET ARCHITECTURE AND SERVICES

Chapter 6 begins describing Carrier Ethernet network architecture and services that are developed by Metro Ethernet Forum and currently deployed in the industry.

Ethernet transport network is a two-layer network consisting of Ethernet MAC (ETH) layer network and Ethernet PHY (ETY) layer network. The ETY layer is the physical layer defined in IEEE 802.3. The ETH layer is the pure packet layer.

The ETH layer network is divided into ETH subnetworks that are also called ETH flow domain (EFD). An EFD is a set of all ETH flow points transferring information within a given administrative portion of the ETH layer network. EFDs may be partitioned into sets of nonoverlapping EFDs that are interconnected by ETH links. An IEEE 802.1D bridge represents the smallest instance of an EFD.

The termination of a Link is called a flow point pool (FPP). The FPP describes configuration information associated with an interface, such as an UNI or ENNI. Ethernet frame is exchanged over ETH layer that consists of preamble, start frame delimiter (SFD), destination MAC address (DA), source MAC address (SA), (Optional) 802.1QTag, Ethernet Length/Type (EtherType), user data, padding if required, frame check sequence (FCS), and extension field, which is required only for 100 Mbp half-duplex operation. A service frame can be a unicast, multicast, broadcast, and Layer 2 Control Protocol (L2CP) frame.

Two interfaces are defined for Carrier Ethernet, UNI, and ENNI. A connection between UNIs is called Ethernet Virtual Connection (EVC), while a connection between an UNI and an ENNI is called Operator Virtual Connection (OVC). An EVC or an OVC can be point-to-point, point-to-multipoint, or multipoint-to-multipoint. The following services are built on top of these EVCs and OVCs:

- Ethernet line (E-Line) services consisting of Ethernet private line (EPL) and Ethernet virtual private line (EVPL) services
- Ethernet local area network (E-LAN) services consisting of EPL and EVPL services

- Ethernet Tree (E-Tree) services consisting of Ethernet Private Tree (EP-Tree) and Ethernet Private Virtual Tree (EVP-Tree) services
- Ethernet Access services offered over ENNI

The operator responsible from the end-to-end EVC may order a tunnel between its remote user that is connected to another operator's network and its ENNI gateway. This access of the remote user to the SP's network is called User Network Interface Tunnel Access (UTA). The remote user end of the tunnel is called Remote User Network Interface (RUNI), while the SP end of this tunnel is called Virtual User Network Interface (VUNI). Attributes of these services are described in detail in Chapter 6.

Chapter 6 is followed by Chapters 7 and 8, describing traffic management and OAM capabilities, respectively.

1.7 CARRIER ETHERNET TRAFFIC MANAGEMENT

Traffic Management, which is also called packet conditioning, is queuing, scheduling, and policing frames. The conditioning function includes the following:

1. Counting bytes and packets that pass and drop
2. Policing packet flow according to predetermined burst/flow rate (includes both color aware and color unaware)
3. Setting color
4. Setting 1P/Differentiated Services Code Point (DSCP)/type of service (TOS) priority value
5. Re-marking/remapping 1P/DSCP/TOS based on a predefined re-marking table
6. Shaping packets to conformance to predetermined flow rate
7. Sending packets to a particular queue

Once frames are classified into a Class of Service (CoS) flow, ingress frames are then policed according to the CoS bandwidth profile assigned to the flow, consisting of Committed Information Rate (CIR), Excess Information Rate (EIR), Committed Burst Size (CBS), Excess Burst Size (EBS), Coupling Flag (CF), and Color Mode (CM).

Frames that are marked "Green" by the policer are always queued, and frames that are marked "Yellow" are only queued if the fill level of the queue is less than a defined threshold. This ensures that "Green" frames are always forwarded.

The MEF defines three classes as H, M, and L. Their priorities are indicated by Priority Code Point (PCP), DSCP, or TOS bytes. Each CoS label has its own performance parameters.

- Class H is intended for real-time applications with tight delay/jitter constraints such as VoIP
- Class M is intended for time critical
- Class L is intended for non-time-critical data such as e-mail.

Service-level agreements (SLAs) between an SP and subscriber for a given EVC are defined in terms of delay, jitter, loss, and availability performances. SLAs are further defined for the following four performance tiers (PTs):

1. PT1 (Metro PT) for typical Metro distances (<250 km, 2 ms propagation delay),
2. PT2 (Regional PT) for typical Regional distances (<1200 km, 8 ms propagation delay),
3. PT3 (Continental PT) for typical National/Continental distances (<7000 km, 44 ms propagation delay),
4. PT4 (Global PT) for typical Global/Intercontinental distances (<27,500 km, 172 ms propagation delay).

Chapter 7 also describes SLA-application mapping.

1.8 ETHERNET OPERATIONS, ADMINISTRATIONS, AND MAINTENANCE (OAM)

Ethernet operations, administrations, and maintenance (OAM) consists of fault management, performance management, testing, and service-monitoring capabilities to support Ethernet services described in Chapter 6. Measurements; events/alarms; and provisioning attributes for interfaces (i.e., UNI, VUNI, ENNI), EVC, and OVC are defined.

Ethernet OAM can be categorized into the following:

- Link layer OAM
- Service layer OAM
- OAM via ELMI.

Link OAM provides mechanisms to monitor link operation and health, and improves fault isolation. The major features covered by this protocol are as follows:

- Discovery of MAC address of the next hop
- Remote failure indication for link failures
- Power failure reporting via Dying Gasp
- Remote loopback.

Ethernet (Alarm Indication Signal) AIS/RDI (Remote Defect Indication) is used to suppress downstream alarms and eliminate alarm storms from a single failure. This is similar to AIS/RDI in legacy services such as SONET (synchronous optical network), TDM, and frame relay.

The service OAM consists of continuity check, loopback, link trace, delay/jitter measurements, loss measurements, and in-service and out-of-service testing protocols to monitor SLAs, identify service-level issues, and debug them.

Connectivity Fault Management (CFM) creates a maintenance hierarchy by defining maintenance domain and maintenance-domain levels, where maintenance end points (MEPs) determine domain boundaries. Maintenance points (MPs) that are between these two boundary points, MEPs, are called Maintenance Intermediate Points (MIPs).

Continuity check messages (CCM) are issued periodically by MEPs. They are used for proactive OAM to detect loss of continuity (LOC) among MEPs and discovery of each other in the same domain. MIPs will discover MEPs that are in the same domain using CCMs as well. In addition, CCM can be used for loss measurements and triggering protection switching.

Ethernet loopback (ETH-LB) function is an on-demand OAM function that is used to verify connectivity of an MEP with a peer MP(s). Loopback is transmitted by an MEP on the request of the administrator to verify connectivity to a particular MP.

Ethernet link trace (ETH-LT) function is an on-demand OAM function initiated in an MEP on the request of the administrator to track the path to a destination MEP. They allow the transmitting node to discover connectivity data about the path. The PDU used for ETH-LT request information is called the link trace message (LTM), and the PDU used for ETH-LT reply information is called the link trace reply (LTR).

Performance measurements can be periodic and on-demand. They are available in 15-min bins. They can be in-service measurements to monitor health of the network as well as user SLAs, and out-of-service measurements before turning up the service, for isolating troubles and identifying failed components during failures.

Delay measurement can be performed using the 1DM (one-way delay measurement) or DMM/DMR (delay measurement message/delay measurement reply) PDUs. Loss measurement can be performed by counting service frames using the LMM/LMR (loss measurement message/loss measurement reply) PDUs as well as by counting synthetic frames via the SLM/SLR (synthetic loss message/synthetic loss reply) PDUs.

In addition to the above, Chapter 8 describes availability and testing based on RFC2544 and ITU-T Y.1731.

1.9 CIRCUIT EMULATION

The CES allows Carrier Ethernet networks to be able to support legacy equipment of users by supporting legacy interfaces such as TDM. As a result, the users benefit from Carrier Ethernet capabilities without replacing their equipment.

In the CES, data streams are converted into frames for transmission over Ethernet. At the destination site, the original bit stream is reconstructed when the headers are removed, payload is concatenated, and clock is regenerated, while ensuring very low latency. The MEN behaves as a virtual wire.

The TDM data is delivered at a constant rate over a dedicated channel. The TDM PW operates in various modes, Circuit Emulation over PSN CESoPSN, Structure-Agnostic TDM over Packet (SAToP), TDM over IP (TDMoIP), and HDLCoPSN (High-Level Data Link Control (HDLC) emulation over PSN).

In SAToP, the TDM is typically unframed. When it is framed or even channelized, the framing and channelization structure are completely disregarded by the transport mechanisms. In such cases, all structural overhead must be transparently transported along with the payload data, and the encapsulation method employed provides no mechanisms for its location or utilization. On the other hand, structure-aware TDM transport may explicitly safeguard TDM structure. The unstructured emulation mode is suitable for leased line.

Ethernet CES provides emulation of TDM services, such as N $\times$ 64 kbps, T1, T3, and OC-n, across a MEN, but transfers the data across MEN. From the customer perspective, this TDM service is the same as any other TDM service.

Circuit emulation applications interconnected over a Circuit Emulation Services over Ethernet (CESoETH) service may exchange signaling in addition to TDM data. With structure-agnostic emulation, it is not required to intercept or process CE (customer-edge) signaling. Signaling is embedded in the TDM data stream, and hence it is carried end-to-end across the emulated circuit. With structure-aware emulation, transport of common channel signaling (CCS) may be achieved by carrying the signaling channel with the emulated service, such as channel 23 for DS1.

In addition, Chapter 9 describes performance monitoring, provisioning, and fault management of CES over Ethernet.

1.10 ETHERNET LOCAL MANAGEMENT INTERFACE (ELMI)

Chapter 10 describes ELMI protocol that operates between the CE device and network element (NE) of SP.

The ELMI protocol includes the following procedures:

- Notification to the CE device of the addition of an EVC
- Notification to the CE device of the deletion of an EVC
- Notification to the CE device of the availability (active/partially active) or unavailability (inactive) state of a configured EVC
- Notification to the CE device of the availability of the RUNI
- Communication of UNI and EVC attributes to the CE device.

In order to transfer ELMI messages between the UNI-C and UNI-N, a framing or encapsulation mechanism is needed. The ELMI frame structure is based on the IEEE 802.3 untagged MAC-frame format, where the ELMI messages are encapsulated inside Ethernet frames.

At ELMI, STATUS Message is sent by the UNI-N to UNI-C in response to a STATUS ENQUIRY message to indicate the status of EVCs or for the exchange

of sequence numbers. STATUS ENQUIRY is sent by the UNI-C to request status or to verify sequence numbers.

The ELMI procedures are characterized by a set of ELMI messages that will be exchanged at the UNI. These message exchanges can be asynchronous or periodic. Periodic message exchanges are governed by timers, status counters, and sequence numbers.

1.11 PBT

Addressing scalability of Carrier Ethernet is being questioned. PBBs and PBT described in Chapter 11 attempts to resolve the scalability issue. These extensions to the Ethernet protocols are developed to transform Ethernet to a technology ready for use in MANs/WANs.

Services supported in LAN/MEN such as E-LAN and E-LINE will be supported end-to-end. This results in no changes to the customer's LAN equipment, providing end-to-end usage of the technology, contributing to wider interoperability and low cost. SLAs provide end-to-end performance, based on rate, frame loss, delay, and jitter, and enable traffic engineering (TE) to fine-tune the network flows.

Underlying protocols for the PBT are as follows:

- IEEE 802.1AB link layer discovery protocol, which is used to discover the network layout and forwarding this information to the control plane or management layer
- The IEEE 802.1ag protocol to monitor the links and trunks in the PBT layout
- Expansion of the PBB protocol defined in IEEE 802.1ah
- IEEE 802.1Qay, PBBs with TE or PBT protocol.

The PBB, namely, MAC-in-MAC encapsulation, supports complete isolation of individual client-addressing fields and isolation from address fields used in the operator backbone. Client provider bridge (PB) frames are encapsulated and forwarded in the backbone network, based on new B-DA (backbone destination address), B-SA (backbone source address), and B-VID (backbone VLAN-ID).

Although Q-in-Q supports a tiered hierarchy (i.e., no tag, C-Tag/C-VLAN ID, and S-Tag/S-VLAN ID), the SP can create 4094 customer VLANs, which is insufficient for large metropolitan and regional networks. The 802.1 ah introduces a new 24-bit tag field (I-SID), service instance identifier, to overcome 12-bit S-VID (S-VLAN ID) defined in PB. This 24-bit tag field is proposed as a solution to the scalability limitations encountered with the 12-bit S-VID defined in PBs.

PBBs operate the same way as traditional Ethernet bridges. CFM addresses the end-to-end OAM, such as loopback at specific MAC, link trace, and continuity check.

The PBB located at the backbone of the PBT network is called backbone core bridge (BCB). The bridge located at the edge of PBT network is called

backbone edge bridge (BEB). The BCB is an S-VLAN bridge used within the core of a PBBN. The BEB is a system that encapsulates customer frames for transmission across a Provider Backbone Bridge Network (PBBN).

The BEB is of three types: I type Backbone Edge Bridge (I-BEB), B type Backbone Edge Bridge (B-BEB), and IB type Backbone Edge Bridge (IB-BEB).

I-component is responsible for encapsulating frames received from customers, assigning each to a backbone service instance and destination identified by a backbone destination address, backbone source address, and a service instance identifier (I-SID).

B-component is responsible for relaying encapsulated customer frames to and from I-components or other B-components when multiple domains interconnect, either within the same BEB or externally connected, checking that ingress/egress is permitted for frames with that I-SID, translating the I-SID (if necessary) and using it to assign the supporting connection parameters (backbone addresses if necessary and VLAN identifiers) for the PBBN, and relaying the frame to and from the Provider Network Port(s).

The PBBN provides multipoint tunnels between provider bridged networks (PBNs), where each B-VLAN carries many S-VLANs.

Traffic engineered provider backbone bridging (PBB-TE), PBT, is intended to bring connection-oriented characteristics and deterministic behavior to Ethernet. It turns off Ethernet's spanning tree and media-access-control address-flooding and learning characteristics. That lets Ethernet behave more similar to a traditional carrier transport technology.

The frame format of PBT is exactly as the format used for implementing the PBB. The difference is in the meaning of the frame's fields. The VID and the B-DA fields together form a 60-bit globally unique identifier.

The control plane is used to manage the forwarding tables of the switches. To create PBT tunnels, all switches need to be controlled from one (PBT) domain. This technique enables circuit switching on an Ethernet

Chapter 11 further describes PBT networks and PBT-MPLS interworking.

1.12 T-MPLS AND MPLS-TP

Chapters 12, 13, and 14 describe three technologies that are competing or working with Carrier Ethernet in forming data networks, T-MPLS, MPLS-TP, and VPLS. Although MPLS-TP is supposed to replace T-MPLS, T-MPLS is already deployed. Therefore, Chapter 12 is devoted to it.

T-MPLS offers packet-based alternatives to SONET circuits and promises much greater flexibility in how packet traffic is transported through their metro and core optical networks. In T-MPLS, a new profile for MPLS is created, so that MPLS label switched paths (LSPs) and PWs can be engineered to behave similar to TDM circuits or Layer 2 virtual connections.

The T-MPLS is intended to be a separate layer network with respect to MPLS. However, the T-MPLS will use the same data-link protocol ID (e.g., EtherType),

frame format, and forwarding semantics as defined for MPLS frames. Unlike MPLS, it does not support a connectionless mode and is intended to be simpler in scope, less complex in operation, and more easily managed. Layer 3 features have been eliminated and the control plane uses a minimum of IP to lead to low-cost equipment implementations.

As an MPLS subset, T-MPLS abandons the control protocol family, which the Internet Engineering Task Force (IETF) defines for MPLS. It simplifies the data plane, removes unnecessary forwarding processes, and adds ITU-T transport style protection switching and OAM functions (e.g., connectivity verification, alarm suppression, RDI).

The key differences of T-MPLS when compared with MPLS include the following:

- Use of bidirectional LSPs
- No PHP (Penultimate Hop Popping) option
- No ECMP (Equal Cost Multiple Path) option.

The T-MPLS, similar to MPLS, defines UNI interface, which is the interface between a client and service node, and NNI (Network−Network Interface), which is between two service nodes.

In a typical T-MPLS network, a primary LSP and backup LSP are provided. The switching between the primary and secondary LSP tunnels can take place within 50 ms. These T-MPLS tunnels can support both Layer 3 IP/MPLS traffic flows and Layer 2 traffic flows via PWs. The T-MPLS protection can be linear or ring.

The MPLS-TP is a continuation of T-MPLS. A Joint Working Group (JWT) was formed between the IETF and ITU-T to achieve mutual alignment of requirements and protocols and come up with another approach. The T-MPLS is renamed as MPLS-TP to produce a converged set of standards for MPLS-TP. The MPLS-TP is a packet-based transport technology based on the multiprotocol label switching traffic engineering (MPLS-TE) and PW data plane architecture.

The objective is to achieve the transport characteristics of SONET/SDH (synchronous digital hierarchy) that are connection oriented; a high level of availability; quality of service; and extensive operations, administration, and maintenance (OAM) capabilities.

With the MPLS-TP, network provisioning can be achieved via a centralized Network Management System (NMS) and/or a distributed control plane. The generalized multiprotocol label switching (GMPLS) can be used as a control plane that provides a common approach for management and control of multilayer transport networks.

Networks are typically operated from a network operation center (NOC) using an NMS that communicates with the network elements (NEs). The NMS provides FCAPS management functions (i.e., fault, configuration, accounting, performance, and security management).

For MPLS-TP, the NMS can be used for static provisioning while the GMPLS can be used for dynamic provisioning of transport paths. The control

plane is mainly used to provide restoration functions for improved network survivability in the presence of failures and facilitates end-to-end path provisioning across network or operator domains.

Similar to T-MPLS, MPLS-TP uses a subset of IP/MPLS standards, where features that are not required in transport networks such as IP forwarding, PHP, and ECMP are not supported or made optional. On the other hand, MPLS-TP defines extensions to existing IP/MPLS standards and introduces established requirements from transport networks. Among the key new features are comprehensive OAM capable of fast detection, localization, troubleshooting, and end-to-end SLA verification; linear and ring protection with sub-50 ms recovery; separation of control and data plane; and fully automated operation without control plane using NMS.

Static and dynamic provisioning models are possible. The static provisioning model is the simplified version commonly known as static MPLS-TP. This version does not implement even the basic MPLS functions, such as label distribution protocol (LDP) and Resource Reservation Protocol–Traffic Engineering (RSVP-TE), since the signaling is static. It does, however, implement support for GAL (Generic Associated Channel Label) and G-ACh (Generic Associated Channel), which is used in supporting OAM functions.

An MPLS-TP label switching router (LSR) is either an MPLS-TP provider-edge (PE) router or an MPLS-TP provider (P) router for a given LSP. An MPLS-TP PE router is an MPLS-TP LSR that adapts client traffic and encapsulates it to be transported over an MPLS-TP LSP by pushing a label or using a PW. An MPLS-TP PE exists at the interface between a pair of layer networks. An MPLS-TP label edge router (LER) is an LSR that exists at the endpoints of an LSP and therefore pushes or pops the LSP label.

An MPLS-TP PE node can support UNI providing the interface between a CE and the MPLS-TP network, and NNI providing the interface between two MPLS-TP PEs in different administrative domains.

The details of the MPLS-TP architecture, OAM, and security are described in Chapter 13.

1.13 VIRTUAL PRIVATE LAN SERVICES (VPLS)

The MPLS facilitates the deployment and management of Virtual Private Networks (VPNs). The MPLS-based VPN can be classified as follows:

- Layer 3 multipoint VPNs or IP VPNs that are often referred to as Virtual Private Routed Networks (VPRNs)
- Layer 2 point-to-point VPNs, which basically consist of a collection of separate Virtual Leased Lines (VLL) or PW
- Layer 2 multipoint VPNs or VPLS.

The VPLS is a multipoint service, but unlike IP VPNs, it can transport non-IP traffic and leverages advantages of Ethernet.

Two VPLS solutions are proposed as follows:

1. VPLS using Border Gateway Protocol (BGP) that uses BGP for signaling and discovery
2. VPLS using label distribution that uses LDP signaling and basically an extension to the Martini draft.

Both approaches assume tunnel LSPs between PEs. PWs (PWE3s) are set up over tunnel LSPs (i.e., virtual connection (VC) LSPs).

In order to establish MPLS LSPs, Open Shortest Path First (OSPF-TE) and RSVP-TE can be used, where OSPF-TE will take bandwidth availability into account when calculating the shortest path, while RSVP-TE allows reservation of bandwidth.

There are two key components of VPLS, PE discovery and signaling. PE discovery can be via Provisioning Application, BGP, and RADIUS. Signaling can be targeted via LDP and BGP.

In order to offer different classes of service within a VPLS, IEEE 802.1P bits in a customer Ethernet frame with a VLAN tag is mapped to EXP bits in the PW and/or tunnel label.

VPLS is a multipoint service, therefore, the entire SP network appears as a single logical learning bridge for each VPLS. The logical ports of this SP bridge are the customer ports as well as the PWs on a virtual private local area network service edge (VE). The SP bridge learns MAC addresses, at its VEs while a learning bridge learns MAC addresses on its ports. Source MAC addresses of packets with the logical ports on which they arrive are associated in the Forwarding Information Base (FIB) to forward packets.

In LDP-based VPLS, an interface participating in a VPLS must be able to flood, forward, and filter Ethernet frames. Each PE will form remote MAC address to PW associations and associate directly attached MAC addresses to local customer facing ports. Connectivity between PEs can be via MPLS transport tunnels as well as other tunnels over PWs such as GRE, L2TP, and IPSec. The PE runs the LDP signaling protocol and/or routing protocols to set up PWs, setting up transport tunnels to other PEs and delivering traffic over PWs.

A full mesh of LDP sessions is used to establish the mesh of PWs. Once an LDP session has been formed between two PEs, all PWs between these two PEs are signaled over this session. A hierarchical topology can be used in order to minimize the size of the VPLS full mesh when there is a large number of PWs.

Hierarchical virtual private local area network service (H-VPLS) was designed to address scalability issues in VPLS. In VPLS, all PE nodes are interconnected in a full mesh to ensure that all destinations can be reached. In H-VPLS, a new type of node is introduced called the multitenant unit (MTU), which aggregates multiple CE connections into a single PE, to reduce the number of PE-to-PE connections.

In BGP approach, VPLS control plane functions mainly autodiscovery and provisioning of PWs are accomplished with a single-BGP update advertisement. In the autodiscovery, each PE discovers other PEs that are part of a given VPLS

instance via BGP. When a PE joins or leaves a VPLS instance, only the affected PE's configuration changes, while other PEs automatically find out about the change and adapt.

The BGP route target community (or extended communities) is used to identify members of a VPLS. A PE announces usually via I-BGP that it belongs to a specific VPLS instance by annotating its Network Layer Reach-ability Information (NLRI) for that VPLS instance with route target RT and acts on this by accepting NLRIs from other PEs that have route target RT.

When a new PE is added by the SP, a single-BGP session is established between the new PE and a route reflector. The new PE then joins a VPLS domain when the VPLS instance is configured on that PE. Once discovery is done, each pair of PEs in a VPLS establishes PWs to each other, transmit certain characteristics of the PWs that a PE sets up for a given VPLS, and tear down the PWs when they are no longer needed. This mechanism is called signaling.

Autodiscovery and signaling functions are typically announced via I-BGP. This assumes that all the sites in a VPLS are connected to PEs in a single-autonomous system (AS). However, sites in a VPLS may connect to PEs in different ASs. In this case, I-BGP connection between PEs and PE-to-PE tunnels between ASs are established.

Hierarchical BGP VPLS is used to scale the VPLS control plane when using the BGP.

The advantages of VPLS may be summarized as follows:

- Complete customer control over their routing, where there is a clear demarcation of functionality between the SP and customer that makes troubleshooting easier
- Ability to add a new site without configuration of the SP's equipment or the customer equipment at existing sites
  - — Minimize MAC address exposure, improving scaling by having one MAC address per site (i.e., one MAC per router) or per service
  - — Improve customer separation by having CE router to block unnecessary broadcast or multicast traffic from customer LANs
  - — MPLS core network emulates a flat LAN segment that overcomes distance limitations of Ethernet-switched networks and extends Ethernet broadcast capability across WAN
  - — Point-to-multipoint connectivity connects each customer site to many customer sites
- A single CE–PE link transmits Ethernet packets to multiple remote CE routers
- Fewer connections required to get full connectivity among customer sites.

Adding, removing, or relocating a CE router requires configuring only the directly attached PE router. This results in substantial OpEx savings.