

# MPLS/VPLS Evolution: A Riverstone Perspective

## Introduction

Carrier networks are currently going through a significant evolution. Instead of maintaining separate transport networks to provide voice, video and data services, they are moving towards a common packet based network infrastructure, in order to minimize capital expenditure and operational costs while offering new services, and hence to stay competitive.

This paper describes how the combination of Ethernet, IP routing and MPLS provide the foundation to meet these convergence goals. The following topics will be discussed:

- Industry trends
- Carrier requirements
- VPN Services
- How Riverstone hardware and software solutions best address these market needs
- Current and future Riverstone product delivery phases
- The benefits for carriers to go with Riverstone

## Some History

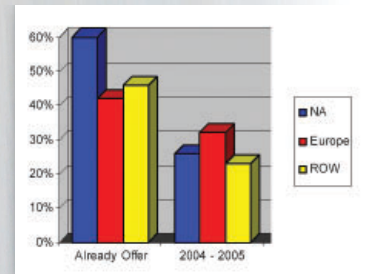
In the late 1990's, competitive carriers started to offer new transport services which were cheaper, faster and more flexible than the traditional leased lines or Frame Relay access services. These new services were based on Ethernet, serving as both a service UNI for end-users but also as a switching/transport technology.

Initially, regular Ethernet switches were used for this transport. Requirements were to support the full range of 802.1Q VLANs (4096) and the ability to support a large number of MAC addresses. New requirements quickly emerged in order to scale and cost-effectively operate service provider (SP) Ethernet backbones:

- The ability to support VLAN translation to handle customers with overlapping VLANs,
- The ability to transparently carry customer Spanning Tree BPDUs,
- The ability to handle a full range of VLANs per customer, independently of the SP VLANs, known as Q-in-Q.

These extensions have recently been standardized by the IEEE 802.1ad working group. Refinements to the Spanning Tree Protocol (STP), such as Rapid Spanning Tree (802.1w) for faster convergence and Multiple VLAN per Spanning Tree instance (802.1s) for basic traffic engineering were also added.

**Riverstone launched the first VPLS enabled routers in 2002.**



**Carrier Plans for VPLS Rollout**  
- Heavy Reading, Oct 2004

It quickly became necessary to provide a more scalable approach to operate such networks. This implied that switches used for transport had to become carrier-class. In other words, they had to provide the same reliability, scalability and security capabilities that traditional TDM or ATM switches offered. MPLS has most of the attributes required to meet such challenges by supporting strong tunneling, traffic engineering, QoS and fast protection capabilities. Complementary hardware and software capabilities deliver the additional reliability necessary.

## Phase 1: MPLS

MPLS came to the forefront when IP WAN routers could not perform longest prefix match lookups at wire-speed. By tagging an IP packet at the ingress point of an MPLS enabled network, further hops along the path only need to perform a label lookup instead of a longest prefix lookup, which is a simpler operation.

MPLS came to the forefront when IP WAN routers could not perform longest prefix match lookups at wire-speed. By tagging an IP packet at the ingress point of an MPLS enabled network, further hops along the path only need to perform a label lookup instead of a longest prefix lookup, which is a simpler operation. This operation is similar to the circuit identifier (VCI) switching found within ATM. These tags, known as MPLS labels, are actually used to create a circuit, and hence augment IP with a connection oriented approach (see [MPLS-ARCH]).

Based on these capabilities, traffic engineering (TE) was one of the first applications of MPLS. The ability to place IP traffic on designated paths offered much better control than traditional IP TE capabilities which often meant using an ATM overlay model at additional cost or adjusting IGP metrics with a greater potential for error.

In addition, the ability to run standard routing protocols such as OSPF and ISIS is a big improvement over the use of STP, in their ability to resolve loops within the SP network. These protocols received extensions to carry traffic related attributes making both traffic engineering and QoS possible. Dynamic signaling protocols such as the Label Distribution Protocol [LDP] and Resource Reservation Protocol [RSVP-TE] allow tunnels to be set-up over such a routed network. Such tunnels can be protected with the use of back up paths or RSVP-TE fast reroute to deliver sub-1 second restoration time.

These MPLS features paved the way for both traffic engineering and QoS support into IP. In 2001, such MPLS capabilities were added to Riverstone routers.

The next application of MPLS was VPN services. It started with BGP VPNs as defined in RFC 2547, and evolved into a carrier's carrier model with RFC 2547-bis, followed by Martini and the Virtual Private LAN Service (VPLS). Riverstone's focus is primarily on these two latter technologies and hence this paper only discusses Martini and VPLS, although BGP VPNs are supported in existing RS routers and in a future release of the Riverstone 15008 Ethernet Edge Router.

## Phase 2: Ethernet Martini and VPLS

Martini pseudo-wires provide the additional capabilities of traffic separation between customers and multiplexing different customer flows onto a transport tunnel for point-to-point connectivity. Transport tunnels are set up with either RSVP-TE or LDP as explained in the previous section. Customer circuits are established via targeted LDP in order to provide traffic separation between two provider end points.

VPLS provides multipoint connectivity and enables LAN-to-LAN services, by providing a per-customer broadcast domain as if all sites (the Customer Edge, or CE) were attached to the same LAN, as described in the Internet Draft [VPLS-LDP], co-authored by Riverstone Networks (see Figure 1.). All provider edge (PE) routers are fully meshed in order to provide optimized site-to-site reachability without having to run a spanning tree protocol to avoid loops. Multicast traffic is treated as broadcast traffic, and hence gets replicated across all ports that belong to a specific customer instance. The full mesh and replication requirements require limitations on the

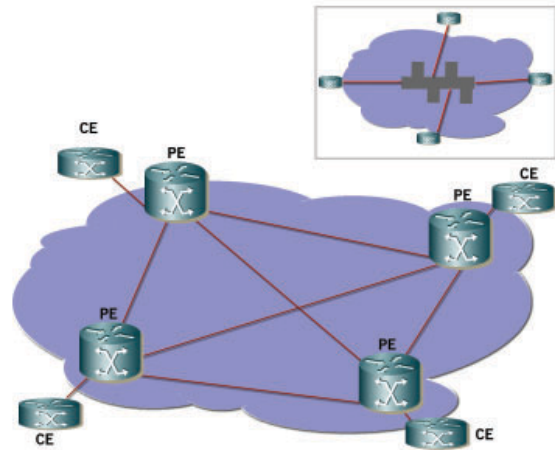
# MPLS/VPLS Evolution - A Riverstone Perspective

total number of VPLS PEs that can be deployed within a single VPLS domain. Once the total number of PEs reaches 40-60, it is recommended to create a multi-tier hierarchy in order to scale VPLS services (described next).

In 2002, Riverstone launched the first Martini and VPLS enabled routers by adding these capabilities to the RS product family.

Riverstone routers have been optimized to support the large number of LDP adjacencies required between PEs. The data path is entirely handled by hardware for wire-speed throughput.

Riverstone routers have been optimized to support the large number of LDP adjacencies required between PEs, along with a large number of hardware-based replications in order to minimize latency. The data path is entirely handled by hardware for wire-speed throughput. MAC and MPLS label lookups, as well as MPLS label manipulation (push/pop operations) are processed in dedicated FPGAs for the RS product line and in micro-coded engines on the 15000 product line. Hence, data plane changes and optimizations do not necessitate new line cards as would be the case with traditional ASIC based technology. The RS platform has been optimized to support up to 128 replications in hardware, even if the actual number of fully meshed VPLS PEs in typical deployments is much smaller. The maximum latency for the largest packet size (1522 bytes) is about 1.5 msec over a GigE link. Access ports/customer facing ports do not have to be MPLS enabled, and therefore do not need specific line cards to be mapped to MPLS circuits. Only the uplink into the MPLS core needs such capabilities. In fact, we see many deployments using VLANs at the edge and MPLS/VPLS in the metro core. The RS and Riverstone 15000 series hardware can support up to 20K VPLS instances, with provisions to support up to 40K already embedded within the hardware.



**Figure 1: VPLS - the network is a single broadcast domain as represented by the smaller image**

It is equally important to be able to terminate legacy interfaces, such as ATM or Frame Relay in addition to Ethernet ports since these technologies can still be used transparently as access interfaces into MPLS/VPLS. In the case of Ethernet, traffic conditioning (such as rate limiting and shaping) provides similar capabilities in terms of QoS and bandwidth partitioning to ATM and Frame Relay virtual circuits. Each circuit is then mapped to a Virtual Circuit (VC) LSP based on the incoming port, VLAN, or VLAN range. This process is known as FEC lookup. 802.1p tags can be used to classify traffic into the proper LSPs and to mark the corresponding MPLS EXP bits. The frame scheduler uses such markings to provide DiffServ QoS upon contention. The Metro Ethernet Forum, where Riverstone plays an active role, is in the process of standardizing these QoS features. The RS and 15000 families already support these features and also include further refinements which are not part of the current specifications. For instance, with the next version of the 15000 family OS, it will also be possible to lookup the DSCP codepoints specified in the IP header of

Riverstone routers have also been enhanced to cope with customers that require dual-homed CPEs. Not only can the Spanning Tree protocol be turned on between customer sites and PEs, but also a unique loop detection mechanism that does not require the use of STP can be enabled to prevent loops.

customer packets to identify which queue, scheduling algorithm and frame marking to apply.

Riverstone routers have also been enhanced to cope with customers that require dual-homed CPEs. Not only can the Spanning Tree protocol be turned on between customer sites and PEs, but also a unique loop detection mechanism that does not require the use of STP can be enabled to prevent loops. In the latter case, MAC address moves are monitored on a per-port/per-VLAN instance, and when a specific threshold is reached, the corresponding ports are blocked. See [VPLS-APPLIC] for deployment examples of such features.

The ability to limit the total number of MAC addresses that can be learned on a port/VLAN prevents VPLS MAC tables from being exhausted by a single customer or by denial of service attacks.

## Operations

In terms of fault verification and isolation, the first features available are LSP ping and traceroute facilities. These tools are used on demand during troubleshooting times. When an LSP fails to deliver traffic, the failure cannot always be detected by the MPLS control plane. LSP ping (modeled after the ICMP echo request/reply) is used to verify that packets that belong to a particular Forwarding Equivalence Class (FEC) actually end their MPLS path on an LSR that is an egress for that FEC. A Traceroute packet is sent to the control plane of each transit LSR, which performs various checks that it is indeed a transit LSR for this path; this LSR also returns further information that helps check the control plane against the data plane, i.e., that forwarding matches what the routing protocols determined as the path.

## Provisioning

Setting up a VPLS Network requires the following tasks:

- Connecting CE device to the PE device
- Configuring the PE devices for the new service:
  - Creating IP Interfaces – On the PE node all the network facing ports need to be configured with proper IP addresses.
  - Creating loopback interfaces or node ID per PE.
  - Configuring an IGP protocol (such as OSPF) for route exchange with the rest of the PEs in the backbone
- Adding interfaces to LDP & RSVP
- Configuring tunnels interface among all PEs in order to set the LSPs
- Setting up mesh of tunnel LSPs
- Define the VPLS instances in each box:
  - Setting up customer profile or assigning FEC type which could be based on:
    - Port-VLAN
    - Port
    - Port-VLAN-range
    - VLAN range
    - VLAN
  - Connecting customer profile to its LDP peer – this will create virtual circuits from the customer facing interface to the far end LDP peer. (The same applies in the reverse direction)

Riverstone's Service Activator (RMC-SA) is an application that automates all the above tasks and also includes additional features such as automatic discovery of all VPLS-aware nodes in the service provider backbone and the ability to specify QoS profiles on a per-customer, per-site basis.

With hierarchical VPLS, it is possible to create a two-tier or three-tier hierarchical network. A new class of MPLS switches, also called Multi-Tenant Unit (MTU) switches, is used in the access portion of the network to aggregate customer traffic into point-to-point Martini or Q-in-Q circuits.

### Phase 3: Hierarchical VPLS and Enhanced Operations

With hierarchical VPLS, it is possible to create a two-tier or three-tier hierarchical network. A new class of MPLS switches, also called Multi-Tenant Unit (MTU) switches, is used in the access portion of the network to aggregate customer traffic into point-to-point Martini or Q-in-Q circuits (Figure 2). These circuits are terminated within VPLS PEs. Instead of creating a full mesh of MTUs, only "core" VPLS PEs

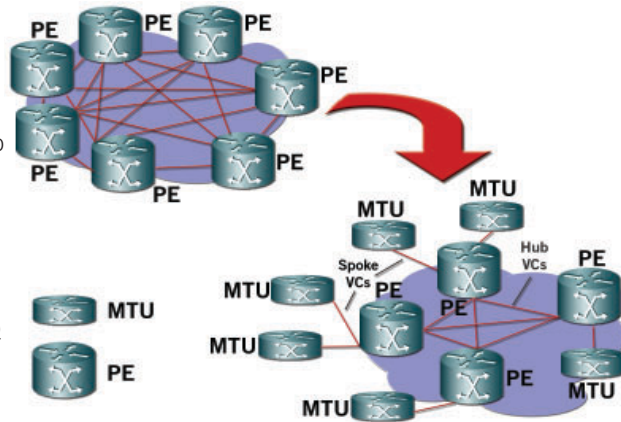


Figure 2: Hierarchical VPLS

need to be fully meshed. MTUs are designed such that a limited set of MPLS features are needed in order to provide cheap and easy-to-manage devices. RS routers can be configured by software to act as MTUs or PEs.

In the next section, we will describe how inter-domain circuits can also be used to break the core mesh between such VPLS border PEs in order to provide inter-domain or inter-carrier connectivity. MTUs can be multi-homed into different VPLS PEs for added reliability.

Once a hierarchical topology is built, it is possible to further optimize traffic replication for both broadcast and multicast traffic to enable applications such as video distribution efficiently.

Once a hierarchical topology is built, it is possible to further optimize traffic replication for both broadcast and multicast traffic to enable applications such as video distribution efficiently. The replication effort is spread among ingress/egress MTUs, PEs and border PEs (Figure 3). Only sites that listen to specific multicast streams receive the corresponding traffic. In order to do so, IGMP and PIM snooping is used to track which ports and which circuits belong to a specific multicast group, as described in [VPLS-MCAST]. It is also possible to minimize the number of LDP signaling adjacencies needed since a fewer number of PEs are meshed together. This also reduces the total number of LSPs required between PEs from  $O(N^2)$  to  $O(N)$ .

The ability to run MPLS over aggregated links allows VPLS services to scale from 1 Gbps to multiple Gbps while offering added resiliency to link failures.

Riverstone has recently introduced new OAM facilities in order to detect VPLS specific connectivity problems. Since there are no standards available yet to detect VPLS specific failures, Riverstone

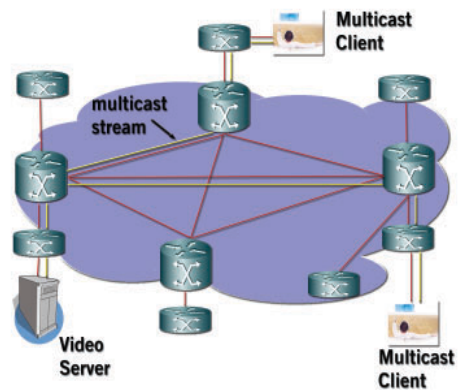


Figure 3: Multicasting with HVPLS

has implemented these new VPLS OAM facilities based on customer feedback. As soon as standards become available, these tools will be made compliant. While LSP ping and traceroute are used to detect specific MPLS transport issues, VPLS ping and traceroute are used to detect connectivity at the Ethernet MAC level of all VPLS aware nodes along the path, such as verifying that MAC addresses have been learned properly. Riverstone plans on submitting an IETF draft describing these OAM extensions.

Both RS and 15000 routers support redundant power supplies, switch fabrics and main processor cards. Riverstone also added hitless protection switching (HPS) capabilities to cope with software failures.

In terms of high availability, both RS and 15000 routers support redundant power supplies, switch fabrics and main processor cards. Riverstone also added hitless protection switching (HPS) capabilities to cope with software failures. Upon failure or crash of the main processor card, the backup card

OAM LAYERS	
OAM Tools	Failure Detection Level
VPLS Ping & Traceroute	MAC
VCCV	VC LSP (PW)
LSP Ping & Traceroute	Tunnel LSP
802.1ah & BFD	Link

\*Note: BFD can also be used in conjunction with LSP ping and VCCV for enhanced failure detection

which keeps a synchronized state of the main processor takes over without affecting the data traffic flows handled in hardware. This includes restart capabilities for most routing and MPLS protocols. Riverstone routers maintain a separation of control and forwarding functions. Control processors are dedicated to control and management functions while ASICs and micro-processors handle the data forwarding functions. This separation enables data to be forwarded even in the case of control software failures. Graceful restart capabilities are available for both routing protocols such as OSPF or BGP and also MPLS signaling protocols like LDP and RSVP. RSVP-TE fast reroute extensions can be used to establish backup LSP tunnels for local repair of LSP tunnels. Two fast-reroute options have been specified: the one-to-one backup and the facility backup methods. In phase 3, Riverstone routers support the former option, which creates detour LSPs for each protected LSP at each potential point of local repair.

These features enable new applications that require high availability and quality of service such as VoIP to run on MPLS enabled IP networks and DSL aggregation over Ethernet which uses VPLS as a core transport technology.

#### Phase 4: Inter-domain/Inter-provider HVPLS

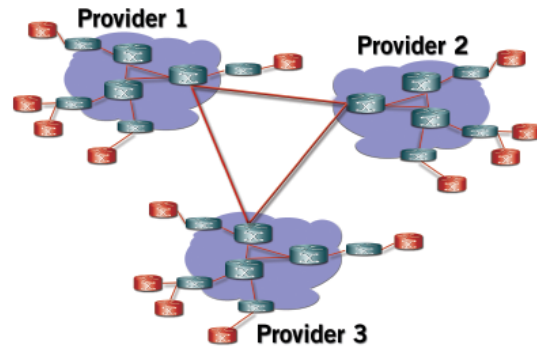
The current VPLS specification will be extended to provide inter-domain and inter-provider connectivity. The VPLS specification mentions that inter-domain spokes can be set up between border PEs (BPEs). Riverstone is currently specifying how redundant BPEs can be supported and dynamically elected, as well as how multiple inter-domain connections can be used between BPEs, in order to avoid single points of failure in the core network (Figure 4).

It is also possible to use BGP VPN carrier's carrier model to interconnect VPLS domains by terminating pseudo-wires directly into RFC 2547 Virtual Routing and Forwarding Instances (VRFs) or by mapping 802.1q VLANs into VRFs, as currently deployed by some carriers.

While Riverstone provides VPLS provisioning tools which automate how to configure VPLS services, automatic discovery of VPLS PEs can be achieved either with BGP or with Radius. There is still a debate about which protocol is best suited to perform

auto-discovery. Both models are valid and depend upon how the SP network is built. The IETF has actually progressed the two approaches as working group documents – showing the community interest in both. Riverstone intends to support both discovery methods.

In terms of OAM, there are several standards organizations defining extensions to Ethernet and VPLS OAM capabilities. Of interest are both the IEEE 802.3ah and 802.1ag and the IETF BFD and VCCV proposals. All these different methods are complementary. For instance, 802.3ah is more suitable for last mile connectivity checks while BFD is more practical for link connectivity check within provider IP/MPLS backbones, and VCCV allows individual pseudo-wires within MPLS tunnels to be checked.



**Figure 4: Inter-domain Hierarchical VPLS**

The use of point-to-multipoint LSPs will allow multicast enhancements to be added to the existing HVPLS multicast optimizations. With HVPLS combined with IGMP and PIM snooping, VPLS aware nodes participate intelligently in the replication effort of multicast traffic as described in previous sections. With p-to-mp LSPs, non-VPLS aware nodes, such as P routers, also get involved in traffic replication.

The interworking of Ethernet with ATM and Frame Relay, which is specified in the MPLS ATM/FR Alliance (where Riverstone participates actively), will also be enhanced such that OAM and QoS can be provided end-to-end. Thus, end-to-end paths can be treated as logical connections, by matching traffic profiles and mapping of OAM traffic. This includes the ability to map PW failures to access circuit failure notifications and vice versa. While bridged encapsulation over ATM/FR does not require a specific adaptation layer for VPLS, routed encapsulation mandates a function like ARP mediation in order to map different address resolution protocols between heterogeneous access technologies.

The need to provide dynamic interworking between such networks will become more significant over time. Since Riverstone routers have the capability to terminate both technologies, they are the right place to perform the necessary adaptations functions such as:

- Mapping of ATM control protocols to MPLS control protocols
  - Signaling of PWs upon ATM UNI signaling requests
- Dedicated PWs for ATM signaling and routing
- PWs as virtual trunks for ATM VPs
- Mediation functions for soft PVCs

LSP stitching, the ability to cross-connect two LSPs together, will allow LSPs to cross provider boundaries or TE domains. Inter-area and inter-AS TE will provide similar capabilities. MPLS NNI (Network-to-Network Interface) will provide tools to set up inter-carrier agreements, which entails translating and mapping different QoS levels as well as being able to provide intelligent routing capabilities between different carriers to meet the required SLAs. Riverstone routers will have the ability to rewrite

EXP markings based on the SLAs negotiated between providers.

In this phase, the facility backup option of fast reroute will be supported. Bypass tunnels are created in order to protect potential failure points. Such tunnels can protect a set of protected LSPs that have similar backup constraints.

Finally, with MPLS sitting at the boundaries of switching (data plane) and routing (control plane), Riverstone plans on providing seamless L2 and L3 capabilities to its product line:

- The ability to look into L3 header information such as DSCP marking while making forwarding decisions at L2 is an important QoS component. Instead of relying on 802.1p markings, which might not always be available or trusted, VPLS PEs can examine the ToS/DSCP codepoints and incoming port/VLAN combination to determine the best path towards a destination and to mark the MPLS EXP bits accordingly.
- The ability to terminate L2 domains into L3 domains within the same platform eliminates the need to use a two-box solution to provide internet access for instance. This means that Martini and VPLS PWs need to be terminated and mapped to an IP routing interface (or a VRF). PWs are then treated as virtual ports and hence can be used as any other physical or logical ports. A mix of Ethernet, ATM/FR virtual circuits and MPLS PWs can be part of the same L2 domain with a routed interface to the Internet for instance.
- The ability for IP to handle a large number of MAC addresses which requires a large IP-ARP table or DHCP snooping capabilities.
- The ability to balance the space used for IP routes vs MAC addresses allows specialized, high performance memory such as Content Addressable Memory (CAM) to be efficiently shared depending upon the type of services offered on a platform.

## Riverstone MPLS/VPLS Phased Implementation

	Features	Benefits to Carriers
<b>Phase 1</b>	<ul style="list-style-type: none"> <li>· MPLS tunneling</li> <li>· Dynamic signaling</li> <li>· Backup LSPs</li> </ul>	<ul style="list-style-type: none"> <li>· Traffic Engineering</li> <li>· QoS</li> <li>· Fast restoration</li> </ul>
<b>Phase 2</b>	<ul style="list-style-type: none"> <li>· Martini PWs</li> <li>· VPLS</li> <li>· Termination of FR, ATM, Ethernet into VPLS</li> </ul>	<ul style="list-style-type: none"> <li>· Leased line replacement</li> <li>· LAN-to-LAN services</li> <li>· Smooth migration of customer base</li> </ul>
<b>Phase 3</b>	<ul style="list-style-type: none"> <li>· Hierarchical VPLS</li> <li>· Dual-homing</li> <li>· HPS, Restart, Fast reroute</li> <li>· OAM</li> <li>· Provisioning</li> </ul>	<ul style="list-style-type: none"> <li>· Scalability</li> <li>· Enhanced multicasting</li> <li>· High availability</li> <li>· End-to-end troubleshooting</li> <li>· Fast service delivery</li> </ul>
<b>Phase 4</b>	<ul style="list-style-type: none"> <li>· Inter-domain VPLS</li> <li>· MPLS NNI, LSP stitching, inter-area TE</li> <li>· End-to-end management</li> <li>· IGMP/PIM snooping</li> <li>· L2/L3 seamless integration</li> </ul>	<ul style="list-style-type: none"> <li>· Large geographical services with inter-carrier QoS</li> <li>· Ease of provisioning, troubleshooting</li> <li>· Multiple services from one platform</li> </ul>

## Conclusion

With the launch of its new 15000 router platform, field proven RS platform, software and hardware roadmap and participation in standard organizations, Riverstone is at the forefront of Ethernet and VPLS technology. Combined with a complete solution package, from MTU, to PE, to border PE, end-to-end provisioning and management tools, Riverstone customers across all continents have been deploying successful large scale VPLS services. With the support of various interface types, such as Ethernet, ATM, Frame Relay, TDM and SONET/SDH, carriers can adopt a phased approach when migrating their customers to MPLS/VPLS services.

## Acronyms

CE	Customer Edge device
FEC	Forward Equivalence Class
LDP	Label Distribution Protocol
LSP	Label Switched Path
MPLS	Multi Protocol Label Switching
P	Provider switch (a.k.a Label Switch Router or LSR)
PE	Provider Edge router (a.k.a. Label Edge Router or LER)
PW	Pseudo-Wire
RSVP	Resource Reservation Protocol
TE	Traffic Engineering

## References

[LDP]	LDP specification (RFC3036), LDP State Machine (RFC3215)
[MPLS-ARCH]	MPLS Architecture (RFC3031), MPLS Label Stack Encoding (RFC3032)
[RSVP-TE]	RSVP-TE extensions for LSP tunnels (RFC3209)
[VPLS-APPLIC]	draft-ietf-l2vpn-vpls-ldp-applic (Internet Draft)
[VPLS-LDP]	draft-ietf-l2vpn-vpls-ldp (Internet Draft)
[VPLS-MCAST]	draft-serbest-l2vpn-vpls-mcast (Internet Draft)

