

Layer 2 Services over MPLS: A Unified Core Architecture

A unified MPLS core for layer 2 services and TDM transport provides numerous benefits for service providers.

Executive Summary

As network traffic showed explosive growth in the 1990s, Layer 2 services became a significant revenue source for service providers. Unfortunately, since technology development did not progress in a timely and synchronized manner, service providers now need to manage several networks that are optimized for different services (e.g., voice, data with QoS, best effort data, etc.). In addition, these networks may share part of their resources in a layered manner, i.e., adding one network on top of the other one (e.g., IP over ATM), which results in an overly complicated operation and an inefficient use of network resources.

MPLS provides a scalable solution for evolving service providers multiple “service networks” into a unified “multiservice network.” To make this transition happen, one critical evolutionary process is necessary to carry the existing Layer 2 traffic across the newly built MPLS core networks. In this white paper, we present an MPLS architecture for Layer 2 services based on Martini’s and Malis’ Internet Drafts.

This white paper first addresses today’s methods of network operation for multiple services. Next it addresses how MPLS core networks bring service providers a simple and elegant solution for multiservice networks. The Martini and Malis architectures are then provided in detail, with a further look into the encapsulation for carrying Layer 2 protocol information across the MPLS core. This white paper concludes with an example of how existing ATM traffic can be transported across an MPLS core network.

The Evolution of Layer 2 Services

Layer 2 services are widely deployed by service providers. Services such as Frame Relay, ATM and managed LAN are contributing significant revenues to carriers’ businesses. The objective of Layer 2 services is to improve the resource utilization of the Layer 1 network, while providing the required QoS for point-to-point connectivity. In other words, Layer 2 technologies mainly focus on providing a point-to-point virtual transport with dynamic bandwidth allocation. Since the design philosophy of Layer 2 networks is very different from the Layer 3 IP networks, most Layer 2 technologies are not ultimately capable of scaling to the needs of today’s service-provider networks, thereby Layer 2 only networks present a significant limitation for future growth.

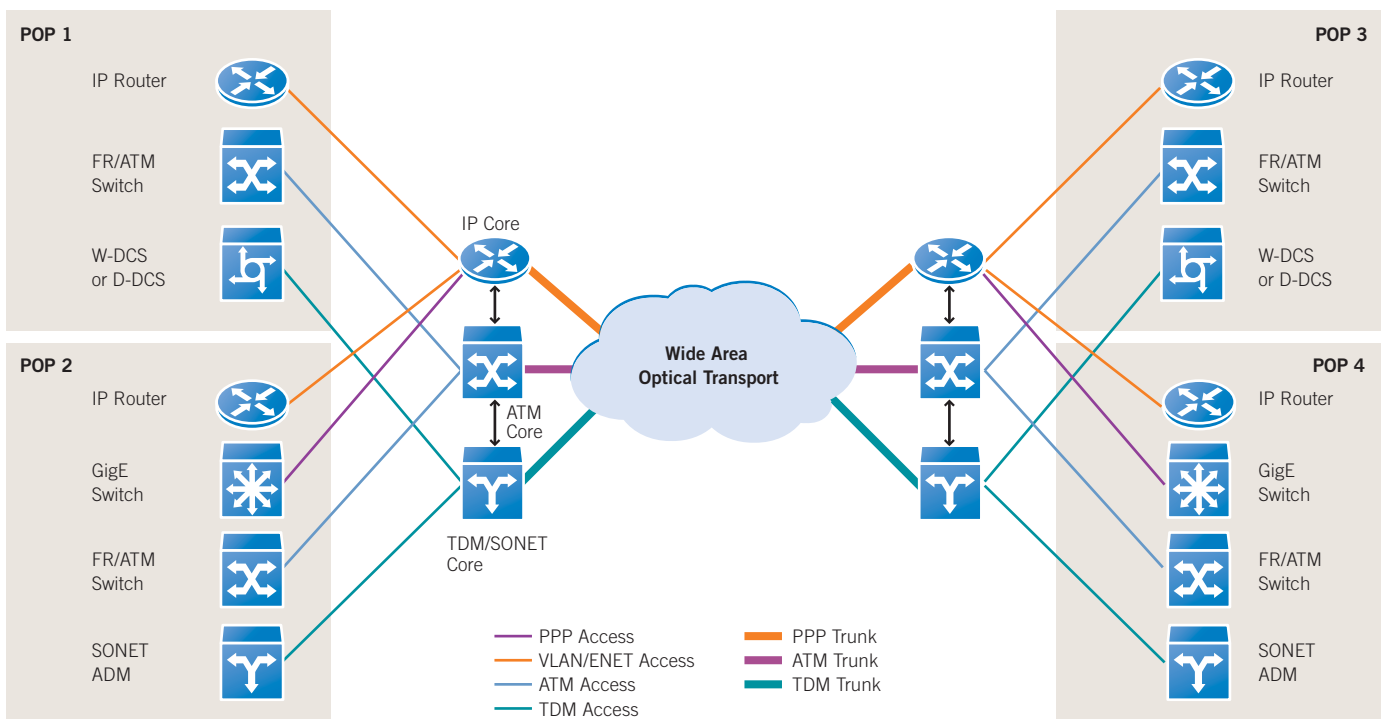


Figure 1. Present multiservice multi-networks.

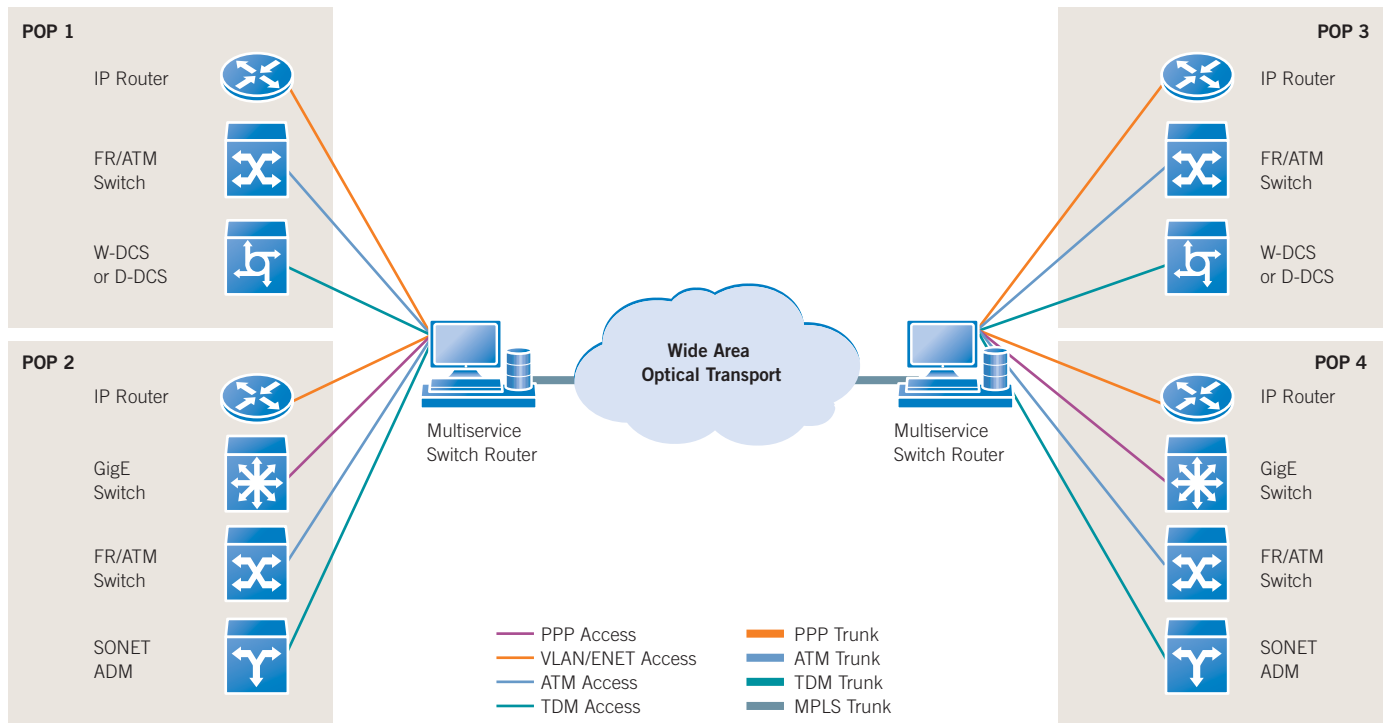


Figure 2. Tellabs multiservice networks.

Before further exploring the Layer 2 scalability issue, one must look back in the recent history of network evolution. In the early 1990s, SONET was proposed as the optical standard of inter-office transport for TDM voice circuits. As bursty data services, mainly carried by Frame Relay PVCs, emerged dramatically in service-provider networks, the 53-byte ATM cell standard was proposed as a solution optimized for both voice and data traffic. The connection-oriented ATM networks soon became the dominant technology to carry both Frame Relay (through AAL5 adaptation of the packet/frame data) and native ATM data traffic across the core network. ATM technology then further expanded into interoffice transport for voice traffic (through AAL1) and an option for integrated access (through AAL2 and AAL5) at the edge.

In the mid 1990s, service providers began their construction of IP networks and started offering Layer 3 services. IP packets could either be mapped into AAL5 across an ATM-based core network, or be mapped directly into TDM core transport through PPP protocol suites. This resulted in a very complex overlay network architecture in the core, see Figure 1.

The network traffic and the corresponding protocol layers in the current network are summarized as follows:

- For IP — IP/ATM/SONET or IP/POS
- For Frame Relay — FR/ATM/SONET
- For ATM — ATM/SONET
- For TDM — TDM/SONET or TDM/ATM/SONET

As illustrated in Figure 1, different networks need to be managed in the core for different services. One core network may or may not ride on the top of the other one, resulting in complex network management with reduced network reliability and increased operational costs. Service providers are now realizing that their ATM networks present a scalability issue that significantly limits future expansion in the core. The issues around ATM's lack of continued scalability are threefold:

- Increasing complexities as multiple control planes are placed on the core ATM network for each service, without any consolidated protocol to handle the architecture in a true consolidated, multiservice fashion
- A lack of hierarchy in its Layer 2 addressing space behind the two levels of VPI/VCI
- The lack of equipment vendor development progress for trunk interfaces higher than OC-12/STM-4 and OC-48/STM-16

To solve the current scalability issue of Layer 2 technologies, service providers need a new solution that provides unlimited scalability and superior QoS support. As a result, MPLS arises as the technology of choice for the future core networks. An example in Figure 2 shows how MPLS can provide a simplified solution for a scalable core network.

This white paper focuses on a proposed MPLS architecture that can provide a scalable, QoS-enabled, tunneling solution for all existing Layer 2 services. The point-to-point tunnel, established by MPLS signaling such as LDP and RSVP, can transport Layer 2 traffic across an MPLS network with an extended definition on encapsulation.

For a unified MPLS core, the network traffic and the corresponding protocol layers are summarized as follows:

- For IP — IP/MPLS/POS
- For Frame Relay — FR/MPLS/POS
- For ATM — ATM/MPLS/ POS
- For TDM — TDM/MPLS/ POS

The Benefits for Service Providers

A unified MPLS core for Layer 2 services and TDM transport provides numerous benefits for service providers.

- **Better Scalability** — MPLS provides a virtually unlimited scalability through a hierarchical label stack.
- **Same Privacy/Security** — MPLS forwards traffic across the core via pre-established LSP, which offers the same privacy/protection as ATM/FR PVCs.
- **Same QoS Support** — With CoS indicated in the EXP field of MPLS header and the advanced queuing solutions provided by MPLS LSRs, such as Per-Flow Queuing (PFQ) at each hop. MPLS enables a full spectrum of QoS support in IP backbone networks.
- **Same Reliability** — MPLS offers secondary LSP provisioning (and automatic re-routing) that protects the primary LSP in the event of a failure. In addition, fast re-route mechanisms can provide a sub-50ms switchover time.
- **A Converged Core** — MPLS is ready to concentrate all the traffic from various Layer 2 protocols at the edge, and move them transparently across the core.
- **Lower Operational Cost** — A unified MPLS core simplifies operation and maintenance and reduces the troubleshooting time under network outages.
- **Lower Transport Cost for TDM** — MPLS can multiplex TDM traffic with dynamic bandwidth allocation support, which translates into accommodating more TDM circuits on a SONET ring.
- **New Services and Revenues** — MPLS enables service providers to create new services. With more advanced and IP-based networking functions, service providers can generate more revenues from MPLS-based networks than from any other type.
- **Easy Migration to GMPLS** — Early adoption of MPLS eases the migration to the developing GMPLS optical core.

Architecture Overview

In a unified MPLS core architecture, all multiservice traffic will be converted into MPLS packets and tunneled through the MPLS core. To serve multiservice traffic from the edge, the MPLS core must provide:

- A point-to-point tunneling architecture that is transparent to Layer 2 services
- An encapsulation approach that carries important Layer 2 circuit parameters across the MPLS core

Tellabs' products support the MPLS-based core architecture defined in Martini's Internet Drafts (ref. [1], [2]) and Malis' Internet Draft (ref. [3]) for Layer 2 services.

Before Layer 2 traffic can be delivered across the MPLS core, an LSP tunnel needs to be established. The Layer 2 traffic first enters the MPLS core at an ingress LSR, is then transported through an LSP tunnel and is eventually converted back to the original Layer 2 protocols at the egress LSR. The whole process is totally transparent to Layer 2 traffic.

There are two forwarding decisions to be made in the forwarding process. The ingress LSR needs to make the first forwarding decision as to which LSP to use for reaching its destined egress LSR. The egress LSR then needs to make the second forwarding decision regarding on which circuit the traffic will continue its journey.

Therefore, a "two-label" approach was proposed in Martini's draft. At the ingress LSR, two labels are prepended to carry Layer 2 PDU across the MPLS core and are then removed at the egress LSR. The first label, called a Tunnel Label, decides which LSP will be used to get packets from the ingress LSR to the egress LSR. The second label, called a VC Label, provides Layer 2 forwarding information at egress LSR.

The labeling process for Layer 2 PDUs is illustrated in Figure 3. An optional Control Word may be inserted below the VC label to transport the Layer 2 control information to the receiver end and thus help ensure a sequential delivery across the MPLS core. We will review the functionalities of the optional Control Word in the section entitled "General Encapsulation."

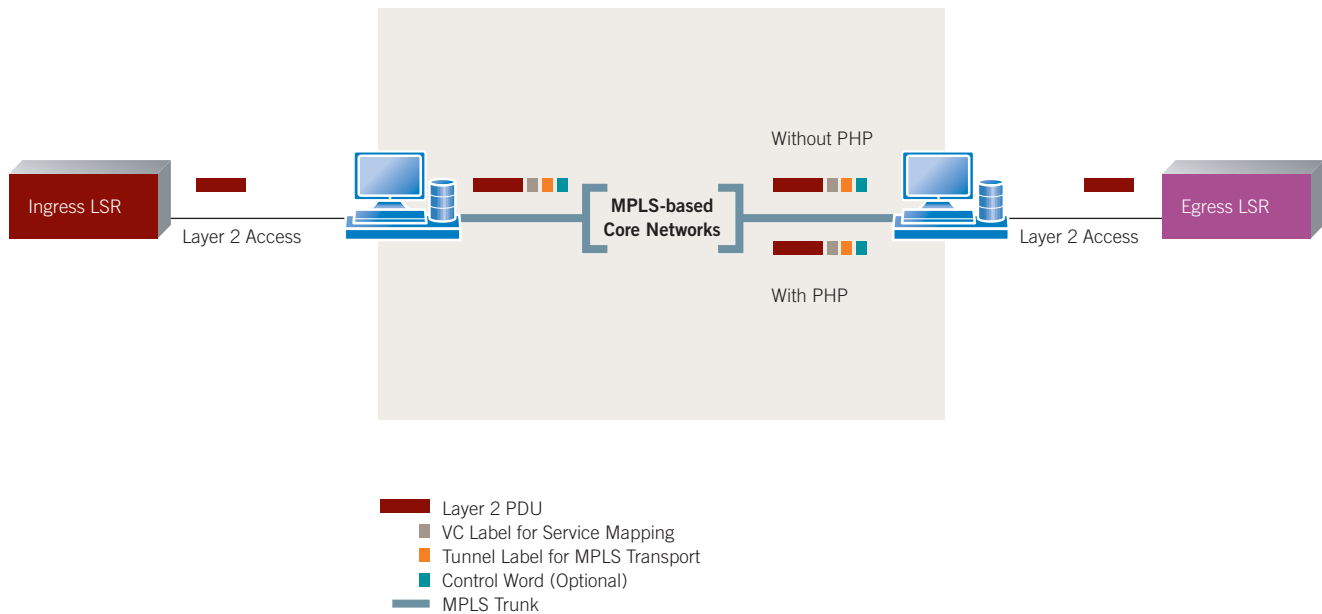


Figure 3. Appending additional labels for Layer 2 traffic across MPLS core networks.

Tunnel LSP Setup

The tunnel LSP transports Layer 2 traffic from one end of the MPLS core to the other. It can be established by static LSP provisioning or by a signaling protocol. A number of signaling protocols can be used to establish the LSP tunnel such as LDP, CR-LDP and RSVP-TE. Among these protocols, CR-LDP and RSVP-TE also can provide further traffic engineering capability for QoS support.

QoS support — PFQ

The MPLS core must provide a rich set of traffic management functionalities to support QoS for Layer 2 services. The edge-to-edge QoS can be achieved by careful per-hop traffic management in the MPLS core. The essential traffic management functions are:

- **Traffic Policing** — for monitoring the incoming traffic at the edge to ensure that the traffic conforms to the service contract.
- **Congestion Management** — for packet dropping precedence in the event of buffer overflow. This will ensure that the switch resource is optimized for high-priority traffic and minimize the packet retransmission for higher layer protocols.
- **Traffic Shaping** — for regulating the incoming traffic and smoothing out the burstiness of the outgoing traffic.
- **Priority Queuing** — for providing QoS/CoS differentiation at each output queue.

PFQ is a required mechanism to embrace all traffic management solutions listed above. Tellabs' traffic management solutions are presented in another white paper — "Traffic Management in Packet-Switched Core Networks."

VC Setup

The VC is used to bind end-to-end Layer 2 circuits together over tunnel LSP. To establish a VC across an MPLS core, LDP is applied to carry circuit-binding information and other interface-specific parameters from one end to the other end.

The VC label bindings are distributed using the "Downstream Unsolicited" mode. In other words, the labels are downstream-assigned and the label-to-FEC bindings are distributed from downstream to upstream by an LSR without an explicit request from its next hop.

It is necessary for the MPLS core to identify a particular VC and to understand interface-specific parameters on both ends. Therefore, a new type of FEC element needs to be defined. The VC FEC element is shown as follows (see Figure 4):

- **The C Bit** — for indicating the presence of a Control Word.
- **VC Type (15 bits)** — for specifying the type of VC. Currently, values have been defined for Frame Relay, ATM AAL5 VCC transport, ATM transparent cell transport, Ethernet, Ethernet VLAN, HDLC, PPP, Circuit Emulation over MPLS (CEM), ATM VCC cell transport and ATM VPC cell transport.
- **VC Information Length (8 bits)** — for specifying the length of the VC ID field and the interface parameters field.
- **Group ID (32 bits)** — for specifying a group of VCs. The Group ID is intended to be used as a port index, or a virtual tunnel index. In the event of a physical port failure, the Group ID can be used to send a wild card label withdrawals to remote LSRs.
- **VC ID (32 bits)** — for identifying a particular VC.
- **Interface Parameters (Variable Length)** — for providing interface-specific parameters.

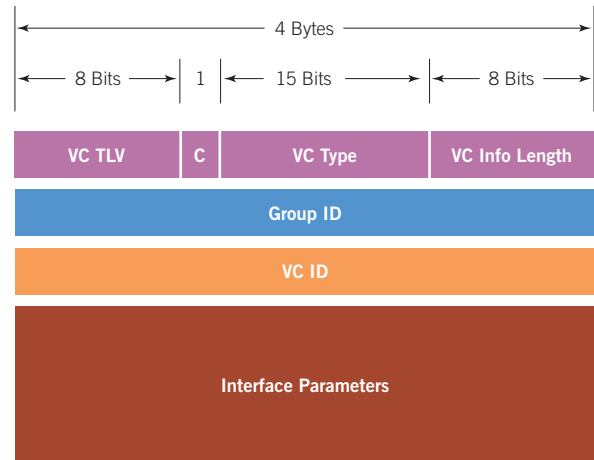


Figure 4. Format of virtual circuit FEC element.

Multiple Interface Parameters can be included in the VC FEC element. The structure of each Interface Parameters field is organized as follows (see Figure 5):

- **Parameter IDs (8 bits)** — for identifying the information type. Currently, the parameter IDs have been defined for the Interface MTU, the Maximum Number of Concatenated ATM Cells, the optional Interface Description string, the Circuit Emulation over MPLS (CEM) Payload Bytes and other CEM Options.
- **VC Info Length (8 Bits)** — for specifying the length of the interface parameter field.
- **Value (Variable Length)** — for carrying the value of the specific information.

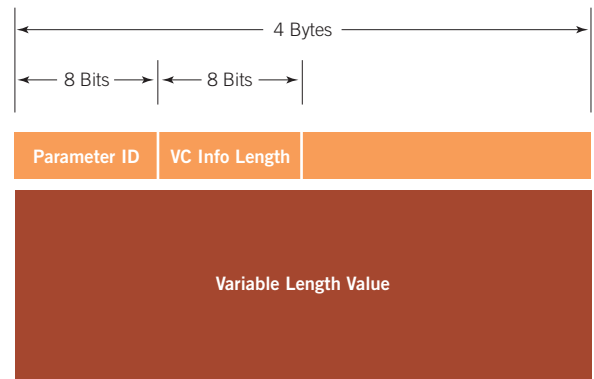


Figure 5. Format of interface parameters field.

The Interface Parameters field is used to validate that the ingress and egress ports at the edge have the necessary capabilities to interoperate with one another. The pre-defined information types include:

- **Interface MTU** — for indicating the interface MTU, which is the maximum packet size allowed to be transmitted at the egress interface, excluding encapsulation overhead. In case this parameter does not match in both directions of a specific VC, that VC cannot be enabled.
- **Maximum Number of Concatenated ATM Cells** — for specifying the maximum number of concatenated ATM cells that can be processed as a single PDU by the egress LSR. Grouping ATM cells in an MPLS packet is permitted. An ingress LSR transmitting concatenated cells on this VC can concatenate a number of cells up to the value of this parameter. This parameter does not need to match in both directions of a specific VC.
- **Interface Description String (Optional)** — for specifying an administrative description text string to the remote LSR.

- **CEM Payload Bytes** — for indicating the TDM payload size in all packets on the CEM stream. In a given CEM stream, all of the packets have the same number of payload bytes. Note that it is possible that the packet size may be larger than the SPE size in the case of an STS-1 SPE, which could cause two pointers to be specified in the CEM header, since the payload of MPLS packet may contain two J1* bytes for consecutive SPEs. To avoid this case, the number of payload bytes must be less than or equal to 783 for STS-1 SPEs.
- **CEM Options (Optional)** — for indicating optional CEM flags.

* J1 byte is the start of STS-1 SPE.

Label Withdrawal

In the event of a physical link failure, the LSPs need to withdraw all the VC labels on the failed interface. The traffic may be re-routed through a secondary LSP so that the impact on services can be minimized.

Label Settings and MTU Requirements

For the tunnel label, the EXP field can be used to identify the QoS/CoS of the Layer 2 services. However, in the case of Penultimate Hop Popping (PHP), the egress LSR will not see the tunnel label at all. Therefore, it is necessary to set the EXP field of the VC label to the same as the tunnel label so that its value is always visible to the egress LSR.

The S bit of the tunnel label must be set to 0 since there is always a VC label below it. Similarly, the S bit of VC label must be set to 1 to denote that the VC label is at the bottom of the stack.

The TTL field is not required to be set to a particular value. It may be set to a value of 2 for the VC label.

At the core of the multiservice MPLS network, support for segmentation and reassembly operations is typically not provided or necessary (in contrast to the edge/LER functions). Therefore, the MTU of the MPLS network must be sufficiently large in size for both MPLS add-on headers plus the largest Layer 2 frame size. If a packet, after being encapsulated by an ingress LSR, has a size larger than the MTU, it will be dropped. Moreover, if a packet, after being decapsulated back to Layer 2 PDU at an egress LSR, has a size larger than the MTU of the destined Layer 2 interface, it will be dropped as well.

General Encapsulation

There are additional pieces of protocol information that MPLS needs to carry across the core besides Layer 2 PDUs. To achieve this, additional headers need to be defined in the MPLS core. In Martini's and Malis' Drafts, two optional headers are defined — an optional Control Word and a CEM Header for SONET/SDH traffic.

The Control Word

The motivation of a Control Word is to provide the following functionalities for Layer 2 services:

- Provide transport Layer 2 control information
- Add appropriate padding for packets smaller than the minimum transport unit
- Ensure an ordered packet delivery

The option of sending a Control Word over a specific LSP must be synchronized between the ingress LSR and the egress LSR. This can be achieved by manual configuration of the LSRs or by LDP signaling.

To deliver the required information, the 4-byte Control Word is defined as follows (see Figure 6):

- **Rsvd (4 bits)** — reserved for future use.
- **Flags (4 bits)** — for carrying protocol-specific control information.
- **Length (8 bits)** — for specifying the packet length so that additional padding can be removed at egress LSR. Note that this field is for packets with less than 256 bytes in length only.
- **Sequence Number (16 bits)** — for tracking the packet sequence.

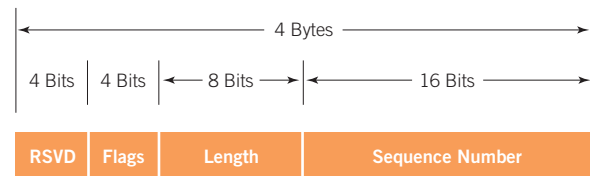


Figure 6. The format of optional control word.

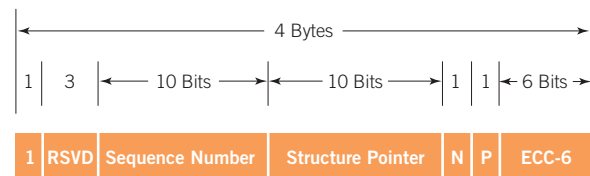


Figure 7. The format of CEM header.

To ensure an ordered packet delivery, the egress LSR needs to support “receive sequence number processing.” Packets that are received out of order may be either dropped or re-ordered at the egress LSR.

The CEM Header

In order to transport SONET/SDH SPEs across an MPLS core, the SPE must be broken into fragments. A 4-byte CEM header is prepended to each fragment.

The format of CEM header is defined as follows (see Figure 7):

- **The D bit** — for identifying Dynamic Bandwidth Allocation (DBA) mode. DBA is an optional mode during which trivial SPEs are not transmitted into the packet network.
- **RSVD (3 bits)** — reserved for future use.
- **Sequence Number (10 bits)** — for tracking the packet sequence.
- **Structure Pointer (10 bits)** — for specifying the offset of the J1 byte within the CEM payload.
- **The N and P bits** — for indicating negative and positive pointer adjustment events. They are also used to relay SONET/SDH maintenance signals such as AIS-P.
- **ECC-6 (6 bits)** — an Error Correction Code (ECC) to protect the CEM header. This offers the ability to correct single-bit errors and to detect up to two-bit errors.

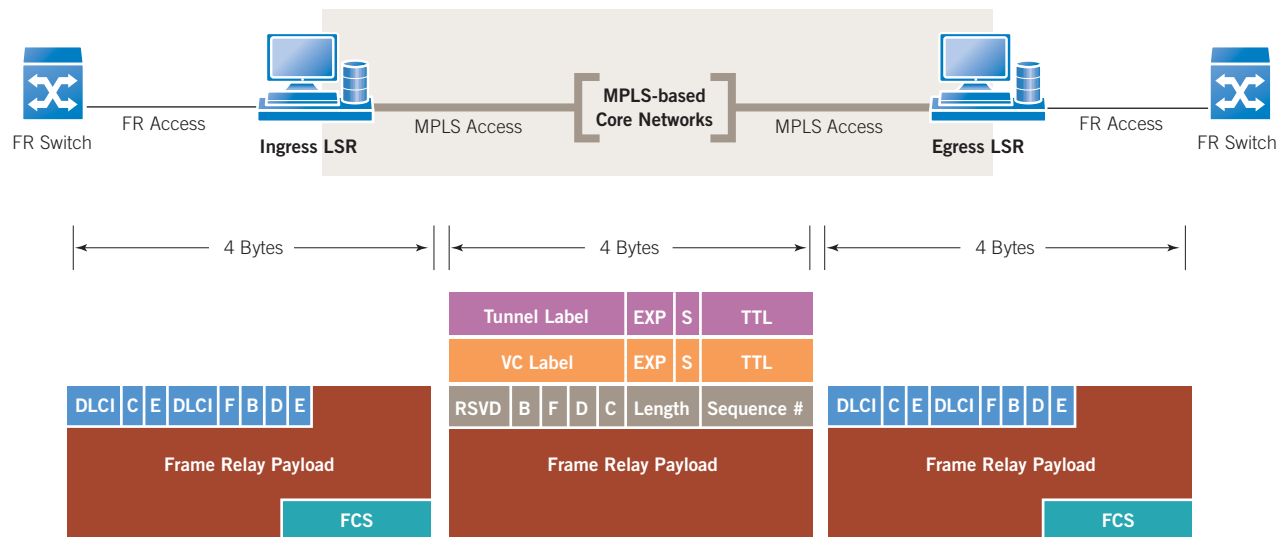


Figure 8. MPLS tunnels and encapsulation for FR packets.

Protocol-Specific Encapsulation

This section explores the encapsulation details that carry meaningful Layer 2 protocol information across the MPLS core. For some protocols, the Layer 2 header is not transported in the MPLS core. By doing this, the network usage for the control information can be minimized. For other protocols, the Layer 2 frames are simply transported as a whole in the MPLS core and may be concatenated using a single MPLS packet.

In general, no error correction field, such as HEC or FCS, is carried across the MPLS core. However, the CEM header does provide an error correction field for encapsulated control information.

At the receiving end, the egress LSR needs to know to which destination port and circuit ID (such as DLCI for Frame Relay, VPI/VCI for ATM and VLAN tag for Ethernet/VLAN) to forward the traffic. This is typically completed at the initial circuit provisioning stage.

Frame Relay

For Frame Relay traffic, only the payload is carried in the MPLS packet. The ordered delivery is necessary, and therefore a Control Word is required.

The detailed packet mapping is illustrated in Figure 8.

- **The B Bit** — for mapping Backward Explicit Congestion Notification (BECN) in the Frame Relay header.
- **The F Bit** — for mapping Forward Explicit Congestion Notification (FECN) in the Frame Relay header.
- **The D Bit** — for mapping Discard Eligibility (DE) in the Frame Relay header. The ingress LSR may consider the DE bit when

determining the value to be placed in the EXP field of the MPLS label stack. Similarly, the egress LSR may consider the EXP field of the VC label when queuing the packet for egress.

- **The C Bit** — for mapping Command/Response (C/R) in the Frame Relay header.

The edge LSRs may change the BECN and/or FECN bits from zero to one in order to reflect congestion in the MPLS network that is known to the edge LSRs, and the D/E bit from zero to one to reflect marking from edge policing of the Frame Relay CIR. The BECN, FECN and D/E bits cannot be changed from one to zero under any circumstances.

ATM

Currently for ATM traffic, two modes of encapsulation have been defined in Martini's draft. One is designed for AAL5 CPCS-PDUs; the other is simply for ATM cell transport.

ATM AAL5 VCC Transport

The objective of ATM AAL5 VCC Transport is to deliver AAL5 CPCS-PDUs on a particular ATM PVC across the MPLS core to another ATM PVC. A SAR process is required for this mode. The ingress LSR needs to reassemble AAL5 CPCS-PDUs from the incoming VC and transport each CPCS-PDU as a single MPLS packet. The AAL5 trailer is not transported. At the receiving end, the egress LSR needs to segment the AAL5 CPCS-PDUs into ATM cells.

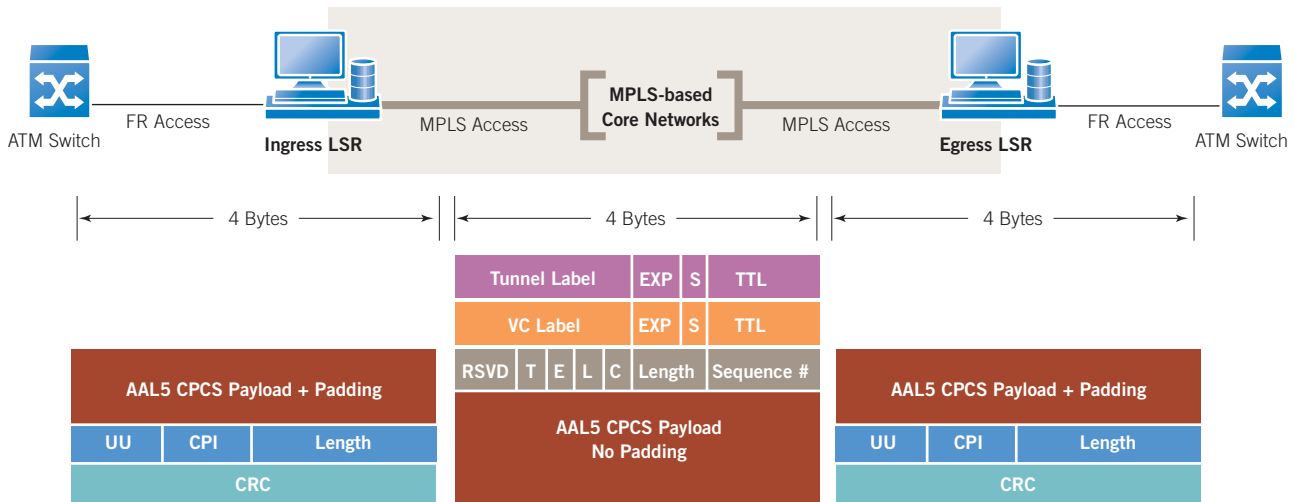


Figure 9. MPLS tunnels and encapsulation for ATM AAL5 CPCS PDUs.

To carry the AAL5 control bits and also ensure an ordered delivery, a Control Word is required.

The detailed packet mapping is illustrated in Figure 9.

- **The T bit** — for specifying the transport type. This bit indicates whether the MPLS packet contains an ATM cell or an AAL5 CPCS-PDU. The ability to transport an ATM cell in the AAL5 mode is intended to provide a means of enabling OAM functionality over the AAL5 VC.
- **The E bit** — for mapping Explicit Forward Congestion Indicator (EFCI) in the AAL5 CPCS-PDU trailer.
- **The L Bit** — for mapping Cell Loss Priority (CLP) bit in the ATM cells.
- **The C Bit** — for mapping C/R bit in the AAL5 CPCS-PDU trailer. When FRF.8.1 Frame Relay/ATM PVC Service Interworking traffic is being transported, the CPCS-UU Least Significant Bit (LSB) of the AAL5 CPCS-PDU may contain the Frame Relay C/R bit.

The edge LSRs may change the EFCI bit from zero to one in order to reflect congestion in the MPLS network that is known to the edge LSRs, and the CLP bit from zero to one to reflect marking from edge policing of the ATM SCR. The EFCI and CLP bits cannot be changed from one to zero under any circumstances.

ATM Transparent Cell Transport

In this mode, ATM cells are transported individually without a SAR process. The error correction field HEC is not transported. In other words, only 52 bytes of information are transported for each ATM cell. Grouping of ATM cells is possible in a single MPLS packet; however, both the ingress and egress LSRs need to support it. The length of each MPLS packet, without the MPLS header and the Control Word,

is a multiple of 52 bytes. The maximum number of ATM cells that can be concatenated in an MPLS packet is limited by the MTU of the ingress/egress interfaces. This is specified in the “Maximum Number of Concatenated ATM Cells” parameter of the FEC element.

The Control Word is optional for ATM Transparent Cell Transport.

The detailed packet mapping is illustrated in Figure 10.

OAM Cell Support

OAM cells may be transported on the VC LSP if the ATM Transparent Cell Transport is supported. Otherwise the OAM support on ATM PVCs can be achieved using OAM Cell Emulation as follows:

- **OAM Loopback** — If an LSR receives an F5 end-to-end loopback OAM cell from a ATM VC, and the LSR has a label mapping for the ATM VC, then the LSR will decrement the loopback indication value and loop the cell back on the ATM VC. The ingress LSR may also periodically generate F5 end-to-end loopback OAM cells on a VC. If the LSR fails to receive a response for a period of time, it will withdraw the label mapping for the VC.
- **Alarm Handling** — If an ingress LSR receives an AIS F5 OAM cell, or if a physical interface goes down, it needs to withdraw the label mappings for all VCs associated with the failure. When a VC label mapping is withdrawn, the egress LSR needs to generate AIS F5 OAM cells on the VC associated with the withdrawn label mapping.

As previously mentioned for the label withdrawal process, it is very useful to have a unique Group ID for each interface. This can significantly reduce the signaling response time of handling network outages.

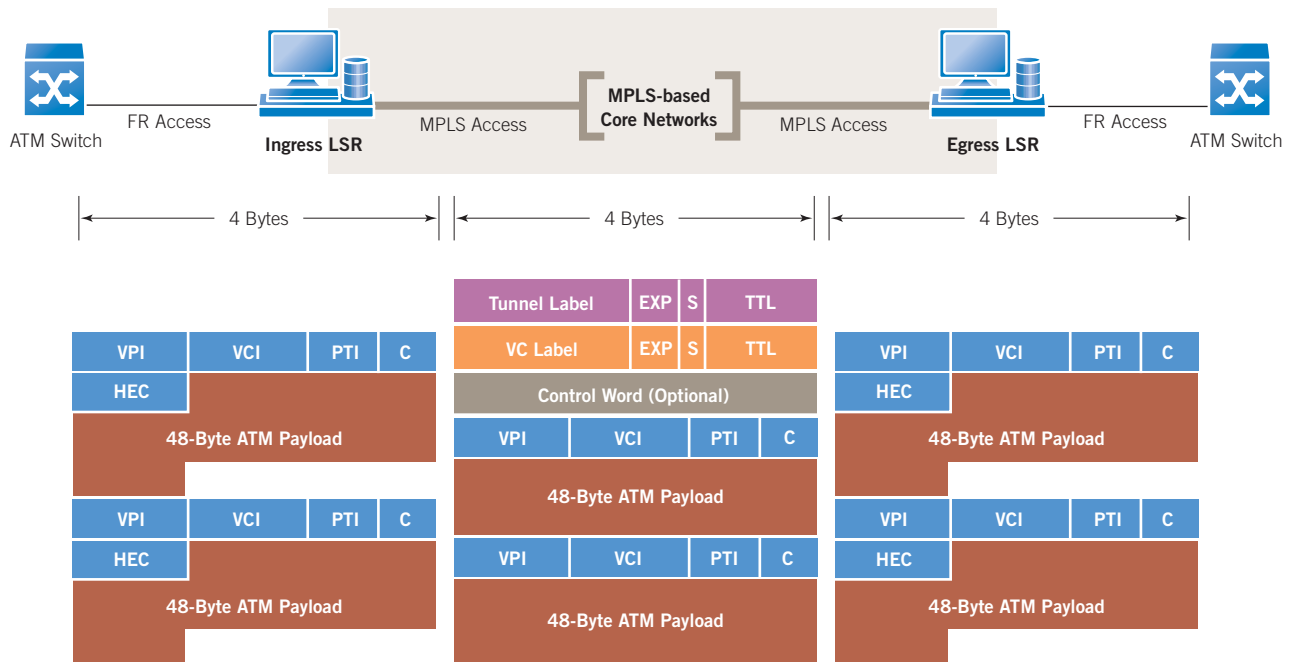


Figure 10. MPLS tunnels and encapsulation for ATM cells.

CLP Bit to MPLS Label Stack EXP Bit Mapping

For the purpose of congestion management in the MPLS core, the ingress LSR may consider the CLP bit when determining the value to be placed in the EXP fields of the MPLS label stack. This will extend the packet dropping policy in the ATM network to the MPLS network. Similarly, the egress LSR may consider the EXP fields when determining CLP bit setting for outgoing ATM cells.

VLAN/Ethernet

For Ethernet 802.1q VLAN, the entire Ethernet frame without the preamble or FCS is transported as a single frame.

For the extended QoS support in the MPLS core, the ingress LSR may consider the user priority field of the VLAN tag header when determining the value to be placed in the EXP fields of the MPLS label stack. Similarly, the egress LSR may consider the EXP field of the VC label when queuing the packet for egress.

The Control Word is optional for VLAN/Ethernet.

The detailed packet mapping is illustrated in Figure 11.

HDLC

For HDLC, the PDU is transported in its entirety, including the HDLC address, control and protocol fields, but excluding HDLC flags and the FCS. Bit stuffing is removed in the MPLS packet.

The Control Word is optional for HDLC.

The detailed packet mapping is illustrated in Figure 12.

PPP

For PPP, the PDU is transported as a single MPLS packet, including the Protocol field, but excluding any media-specific framing information, such as HDLC Address, Control Fields and FCS. The egress LSR needs to know the destination address to forward the decapsulated HDLC frame. Currently, the VC LSP Interface MTU negotiation is not affected by PPP MRU advertisement. That is, if a PPP peer sends a PDU with a length in excess of that negotiated for the VC LSP, the ingress LSR will discard it.

The Control Word is optional for PPP.

The detailed packet mapping is illustrated in Figure 13.

TDM

TDM traffic can be transported through Circuit Emulation over MPLS (CEM). In order to transport SONET/SDH SPEs through an MPLS network, the SPE needs to be broken into fragments. A 4-byte CEM header and a label stack are prepended to each fragment (see Figure 14). The Control Word is not required for CEM. All CEM packets associated with a specific SONET/SDH channel will have the same length.

The detailed packet mapping is illustrated in Figure 14.

When a SONET/SDH fragment is transported across an MPLS network, the delay associated with each packet may vary. To ensure that the packet stream is played out at a fixed rate onto the corresponding SONET/SDH channel at the egress LSR, a buffering mechanism is required to account for delay variation in the CEM

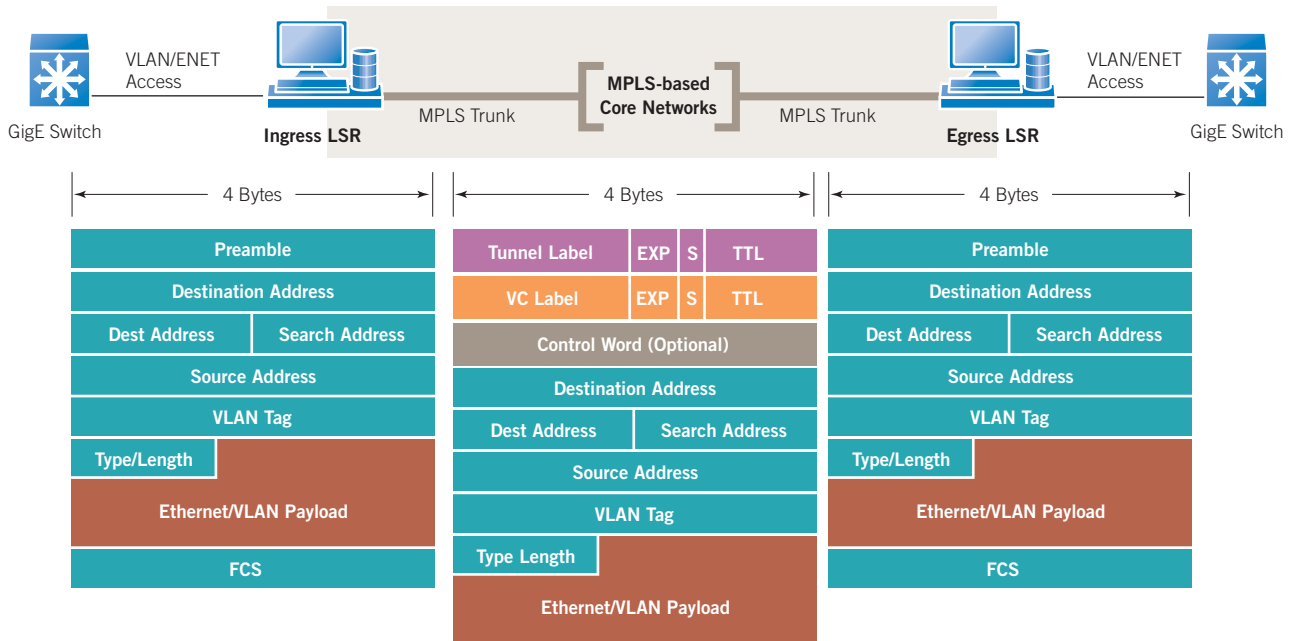


Figure 11. MPLS tunnels and encapsulation for VLAN/Ethernet packets.

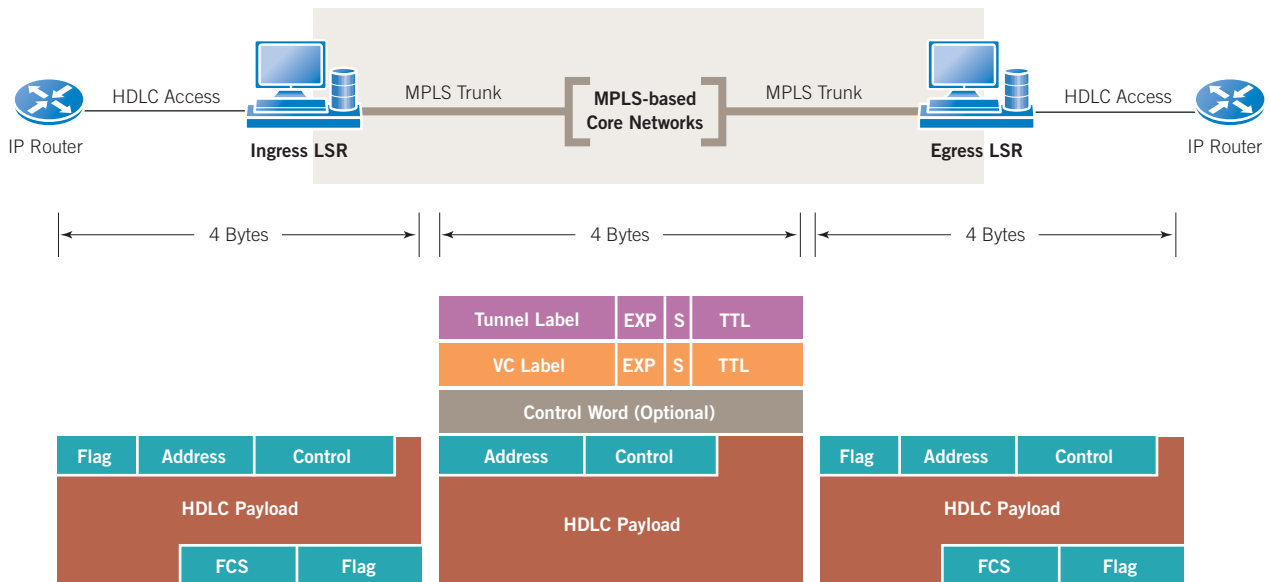


Figure 12. MPLS tunnels and encapsulation for HDLC frames.

packet stream. This buffering mechanism will be generically referred to as the CEM Jitter Buffer. The depth of the jitter buffer is adjustable and is a function of the average delay variation for a specific CEM channel.

The major objectives of CEM are to provide:

- DBA Support
- Alarm Handling
- Pointer Adjustment Indication

The property of a CEM packet and its associated processing are indicated by the D, N and P bits in the CEM header. A CEM packet is operated under either Normal Mode or DBA Mode. During Normal Mode, pointer adjustments need to be supported. During AIS-P, there is no valid payload pointer, so that pointer adjustments cannot occur. If DBA is supported, the determination of AIS-P and STS Unequipped must be based on the state of SONET/SDH Section, Line and Path Overhead bytes. During STS Unequipped, the SONET/SDH payload pointer is valid, and therefore pointer adjustments must be supported even during DBA.

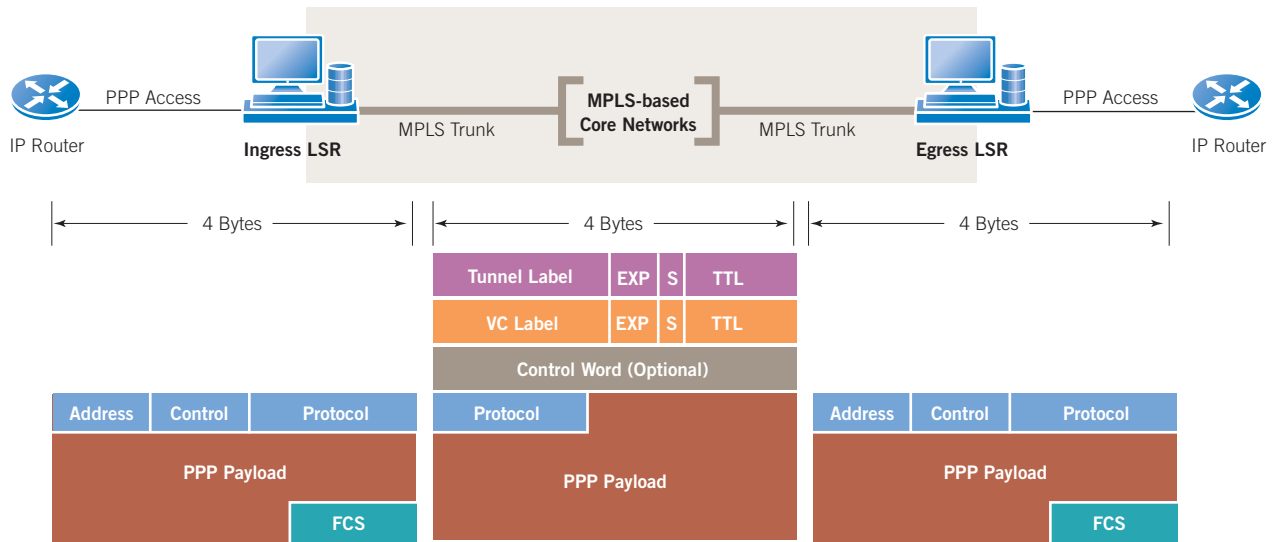


Figure 13. MPLS tunnels and encapsulation for PPP packets.

D	N	P	Interpretation
0	0	0	Normal Mode, No Ptr Adjustment
0	0	1	Normal Mode, Positive Ptr Adjustment
0	1	0	Normal Mode, Negative Ptr Adjustment
0	1	1	Normal Mode, AIS-P
1	0	0	DBA Mode, STS SPE Unequipped, No Ptr Adj
1	0	1	DBA Mode, STS SPE Unequipped, Pos Ptr Adj
1	1	0	DBA Mode, STS SPE Unequipped, Neg Ptr Adj
1	1	1	DBA Mode, AIS-P

Table 1. Interpretation of D, N and P Bits.

The interpretation of the D, N and P bits are summarized in Table 1.

Dynamic Bandwidth Allocation

When meaningful user data is not present in the SPE, such as during AIS-P or STS SPE Unequipped, the SONET/SDH SPE fragments need not be transported. In other words, only the CEM header and MPLS label stack are transmitted. This is done to conserve bandwidth and take advantage of statistical multiplexing gain. The leftover bandwidth can be used to deliver best effort traffic that shares the same physical link.

The operation during DBA Mode should be virtually identical to the Normal Mode operation. This can be accomplished by transmitting each DBA packet only after a complete packet of data has been received from the SONET/SDH channel. In addition, sending out CEM packets at a constant rate can minimize the risk of under or over running the jitter buffer during the transition in and out of DBA.

SONET/SDH Maintenance Signals

- AIS-P** — In a SONET/SDH network, circuit outages are signaled using maintenance alarms such as Path AIS (AIS-P). In particular, AIS-P indicates that the SONET/SDH Path is not currently transmitting valid end-user data, and the SPE contains all ones. If an ingress LSR receives an AIS-P code, it must set the N and P bits of CEM packets to 11 binary to signal AIS-P explicitly through the MPLS network. At the receiving end, if an egress LSR receives CEM packets with the N and P bits set to one, it must set outgoing SPE fragment to all ones and configure the SONET/SDH Overhead to signal AIS-P.
- STS SPE Unequipped** — The STS SPE Unequipped Indication is a slightly different case than AIS-P. When byte C2 of the Path Overhead (STS path signal label) is 00h and byte B3 (STS Path BIP-8) is valid, it indicates that the SPE is unequipped. The entire SPE is typically set to zeros. During Normal Mode, receiving an STS SPE Unequipped makes no difference. During DBA Mode, if an ingress LSR receives an STS SPE Unequipped, it must set the N and P bits to 00 binary, 01 binary or 10 binary to indicate SPE Unequipped with or without pointer adjustments. In addition, if an egress LSR receives a CEM packet with the D-bit set to one to indicate DBA active and the N and P bits set to 00 binary, 01 binary, or 10 binary to indicate SPE Unequipped with or without pointer adjustments, it must transmit a packet of all zeros onto the SONET/SDH interface and adjust the payload pointer to signal STS SPE Unequipped.
- RDI-P** — If the ingress LSRs receive an AIS-L code or detect a line error on an incoming SONET/SDH channel, an RDI-P needs to be sent towards the SONET/SDH channel. In the event of network errors, such as loss of packet synchronization, the egress LSRs need to send an RDI-P towards the packet network. This can be accomplished by modifying the SONET/SDH Path Overhead within the CEM packets. Specifically, the G1 byte must be updated to signal RDI-P and the B3 (Path BIP-8) must be re-computed.

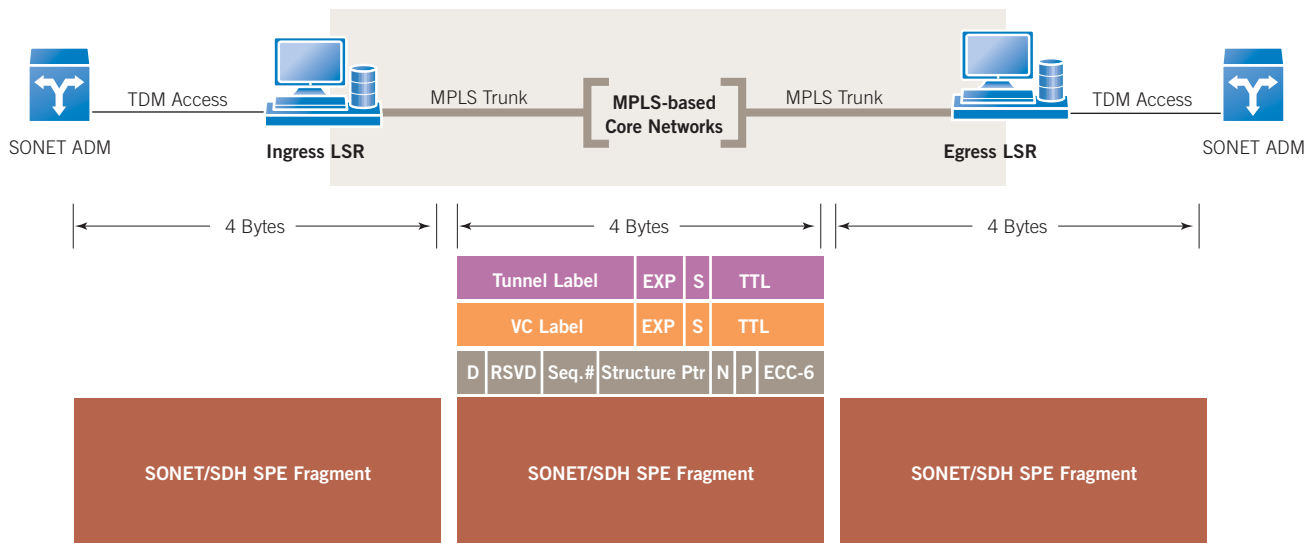


Figure 14. MPLS tunnels and encapsulation for SONET circuits.

		The Control Word	The CEM Header	Interworking Parameters
	Frame Relay	Required	Not Required	BECN, FECN, DE, C/R
ATM	AAL5 VCC Transport	Required	Not Required	Transport Type, EFCI, CLP, C/R
	Transparent Cell Transport	Optional	Not Required	
	VLAN/Ethernet	Optional	Not Required	
	HDLC	Optional	Not Required	
	PPP	Optional	Not Required	
	TDM	Not Required	Required	DBA Mode, AIS-P, STS SPE Unequipped, +/- Pointer Adjustment

Table 2. Encapsulation methods for Layer 2 services over MPLS.

Clocking Mode

If there is a frequency offset between the frame rate of the transport Overhead and that of the SONET/SDH SPE, then the alignment of the SPE will periodically slip back or advance in time through positive or negative stuffing. It is necessary to be able to regenerate the input service clock at the output interface. Two clocking modes can be used to regenerate the clock: Synchronous and Asynchronous.

- **Synchronous** — When synchronous SONET/SDH timing is available at both ends of the circuit, the N and P bits are used to signal negative or positive pointer justification events and eliminate transport jitter.
- **Asynchronous** — If synchronous timing is not available, the N and P bits are not used for frequency justification and adaptive methods can be used to recover the timing. For example, the depth of the Jitter Buffer could be monitored and the frequency of the service clock could be adjusted accordingly.

Summary

We summarize the header options and the control parameters for Layer 2 services over MPLS in Table 2.

Sample Configuration: ATM over MPLS Core

To put everything together and show how MPLS tunnels can be built to transport Layer 2 services across the MPLS core, this white paper presents an example of connecting ATM PVCs over MPLS LSP tunnel using RSVP and LDP, see Figure 15.

First of all, a QoS-enabled tunnel needs to be established between the ingress LSR and the egress LSR. The MPLS core uses RSVP-TE to create a transport tunnel across the network with sufficient bandwidth allocated. Four tunnel labels, L1, L2, L3 and L4, have been assigned to this tunnel. On top of this tunnel, the

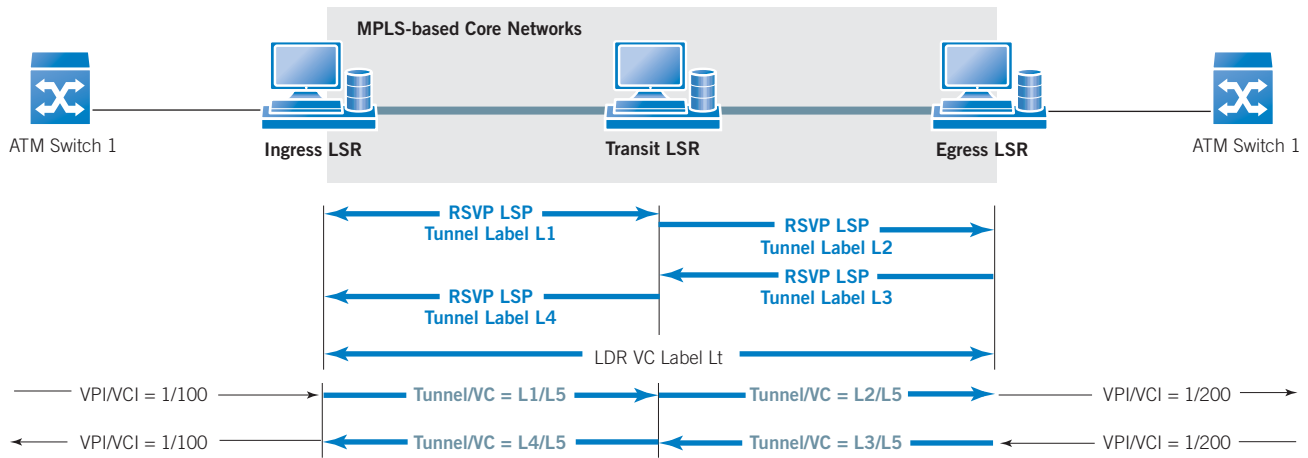


Figure 15. ATM VC cross-connect through MPLS core.

MPLS core uses LDP to create an MPLS network-wide VC label L5 that identifies the virtual circuit for ATM traffic between the ingress LSR and the egress LSR. The next step is to provision the cross-connects between ATM PVCs and MPLS LSPs at the ingress and egress LSRs. This completes the connectivity between ATM switch #1 and ATM switch #2.

If the ingress LSR receives ATM cells with VPI/VCIs = 1/100, the ATM cells will be mapped to the appropriate FEC, prepended with a VC label L5 and a tunnel label L1, and forwarded to the corresponding network-facing interface. When the transit LSR receives the MPLS packet with a tunnel label L1, the label L1 is popped out of the label stack, a new label L2 is pushed in, and the packet is forwarded to the appropriate interface. Note that the VC label L5 is totally invisible during MPLS transport process.

When the MPLS packet reaches the egress LSR, the tunnel label L2 and the VC label L5 are popped out of the label stack, assuming that PHP is not enabled. The egress LSR will then determine the decapsulation and the outgoing interface according to the VC label L5 and the optional Control Word. As a result, the MPLS packet is converted back to ATM cells with VPI/VCIs = 1/200 and forwarded to the appropriate interface. Similarly, on the reverse direction the ATM cells travel from ATM switch #2 to ATM switch #1 using the MPLS LSP with the tunnel labels L3 and L4 and the VC label L5.

Conclusion

This white paper presented how existing Layer 2 services can be integrated on a scalable MPLS core network. The overall architecture is based on Martini's and Malis' Internet Drafts. The tunneling architecture and the encapsulation details were the focus of this paper. We expect this architecture will soon become the standard for scalable network expansion and unified network operation in multiservice service provider networks.

Acronyms

AAL	ATM Adaptation Layer
ADM	Add-Drop Multiplexer
AIS	Alarm Indication Signal
B-DCS	Broadband Digital Crossconnect System
BECN	Backward Explicit Congestion Notification
BIP	Bit Interleaved Parity
C/R	Command/Response
CEM	Circuit Emulation over MPLS
CIR	Committed Information Rate
CLP	Cell Loss Priority
CPCS	Common Part Convergence Sublayer
CR-LDP	Constraint-based Routed Label Distribution Protocol
DBA	Dynamic Bandwidth Allocation
DE	Discard Eligibility
DLCI	Data Link Connection Identifier
EFCI	Explicit Forward Congestion Indicator
FCS	Frame Check Sequence
FEC	Forwarding Equivalence Class
FECN	Forward Explicit Congestion Notification
GMPLS	Generalized Multiprotocol Label Switching
HDLC	High-level Data Link Control
HEC	Header Error Control
LDP	Label Distribution Protocol
LSB	Least Significant Bit
LSP	Label Switched Path
LSR	Label Switched Router
MPLS	Multiprotocol Label Switching
MRU	Maximum Receive Unit
MTU	Maximum Transmission Unit
OAM	Operations And Maintenance
PFQ	Per-Flow Queuing
PDU	Protocol Data Unit
PHP	Penultimate Hop Popping
POS	Packet Over SONET
PPP	Point-to-Point Protocol
PVC	Permanent Virtual Circuit
RDI	Remote Defect Indication
RSVP	Resource reSerVation Protocol
RSVP-TE	Resource Reservation Protocol with Traffic Engineering extensions
SAR	Segmentation And Reassembly
SCR	Sustained Cell Rate
SDH	Synchronous Digital Hierarchy
SONET	Synchronous Optical NETwork
SPE	Synchronous Payload Envelope
STS	Synchronous Transport Signal
TTL	Time-To-Live
VCI	Virtual Circuit Identification
VLAN	Virtual Local Area Network
VPI	Virtual Path Identification
W-DCS	Wideband Digital Crossconnect System

References

- [1] L. Martini, et. al., Transport of Layer 2 Frames Over MPLS, July 2001.
- [2] L. Martini, et. al., Encapsulation Methods for Transport of Layer 2 Frames Over MPLS, July 2001.
- [3] A. Malis, et. al., SONET/SDH Circuit Emulation Service Over MPLS (CEM) Encapsulation, January 2001.
- [4] RFC 3031, E. Rosen, et. al., Multiprotocol Label Switching Architecture, January 2001.
- [5] RFC 3032, E. Rosen, et. al., MPLS Label Stack Encoding, January 2001.

North America

Tellabs
One Tellabs Center
1415 West Diehl Road
Naperville, IL 60563
U.S.A.
+1 630 798 8800
Fax: +1 630 798 2000

Asia Pacific

Tellabs
3 Anson Road
#14-01 Springleaf Tower
Singapore 079909
Republic of Singapore
+65 6215 6411
Fax: +65 6215 6422

Europe, Middle East & Africa

Tellabs
Abbey Place
24-28 Easton Street
High Wycombe, Bucks
HP11 1NT
United Kingdom
+44 871 574 7000
Fax: +44 871 574 7151

Latin America & Caribbean

Tellabs
Rua James Joule No. 92
EDIFÍCIO PLAZA I
São Paulo – SP
04576-080
Brasil
+55 11 3572 6200
Fax: +55 11 3572 6225