EVOLVING TO “END-TO-END MPLS” ARCHITECTURES

ALCATEL-LUCENT ENABLES SEAMLESS, SCALABLE, RESILIENT MPLS NETWORKS

TECHNICAL WHITE PAPER

Increasing demand for video content, Mobile broadband and Cloud services are pushing the limits of service provider networks. Service providers need to add network capacity at the lowest cost per bit and reduce their network operations cost. Traditionally, operators have implemented separate networks for separate services (e.g., fixed and mobile) which results in a sub optimal utilization of network resources. Many service providers are implementing fixed and mobile network convergence (FMC) to optimize network utilization and to reduce network capital and operational expenditures.

When offering services over a single or converged services network, the end-to-end network must be scalable, flexible to meet evolving service needs and support simple, rapid service provisioning. Multiprotocol Label Switching (MPLS) is the preferred choice for implementing end-to-end networks. Alcatel-Lucent, a market leader in MPLS development, offers a complete toolkit to implement “end-to-end MPLS” networks. Seamless MPLS is a key component of this toolkit.

This paper will review the motivation, architecture and building blocks required to implement a Seamless MPLS network. In addition it describes alternate interim options available for effectively deploying a scalable, reliable and manageable end-to-end MPLS network.

TABLE OF CONTENTS

MPLS – A PROVEN TECHNOLOGY

The proven ability of MPLS to work cohesively with Internet Protocol (IP) and Ethernet makes MPLS the preferred choice for implementing highly scalable networks. MPLS technology was initially implemented in service provider core networks. Many operators have already implemented architectures that combine business, residential and mobile services over a single converged IP/MPLS core network.

The success of MPLS in core networks and the benefits it brings has paved the way for the technology to be implemented in metro networks as an alternative to Asynchronous Transfer Mode (ATM) or legacy Ethernet-based aggregation. Within metro networks, in addition to increasing adoption of MPLS in aggregation networks, there is a strong trend towards extending MPLS into metro access networks.

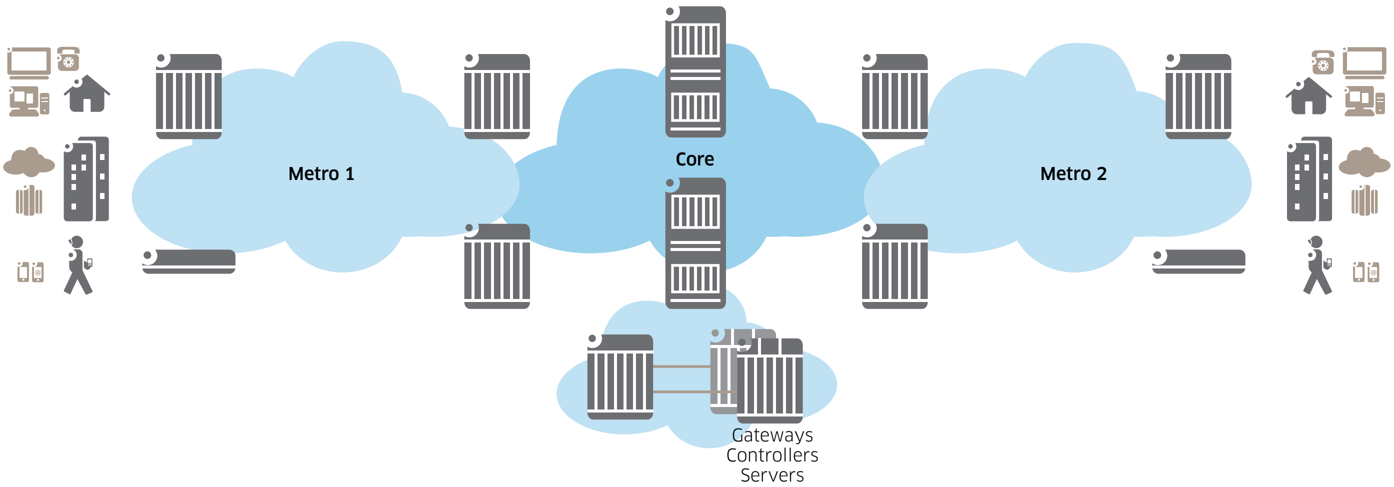
Heavy Reading's, Jan 2013-Ethernet Backhaul Market tracker report estimated 59% of cell sites served by packet backhaul will be served by MPLS in the aggregation network and 26% of the sites will be served by MPLS in the access network by 2015.

While MPLS adoption continues to gain momentum, implementing end-to-end MPLS networks presents a few important considerations for service providers.

WHY SEAMLESS MPLS?

Figure 1 depicts a multi-region network with two metro regions (metro 1 and metro 2) interconnected by a core region. Typically, each region is operated independently. Depending on the service deployed, the service end-points may be within the same metro region or across different regions. Deploying a service from one metro region to another requires provisioning at several intermediate points in the end-to-end network, making troubleshooting and fault recovery more complex.

Figure 1. Multi-region network for converged services



For instance, deploying a Layer 2 point-to-point service from one metro region to another may require provisioning of multiple segments within the metro-core regions, with appropriate mappings at the region boundaries. A preferred approach would be to deploy a single end-to-end service with minimal coordination between regions.

Additionally, service providers are constantly challenged by the evolving needs of their services mix:

- A business or residential service may initially be deployed in a centralized manner, but as the service grows, it may require a distributed model for provisioning purposes.
- A mobile service may need a particular base station to be reassigned to a different radio network controller (RNC) /mobile gateway node, or new technologies such as LTE may require any-to-any mobile backhaul connectivity.

The underlying network must support these evolving requirements flexibly, without impacting existing services.

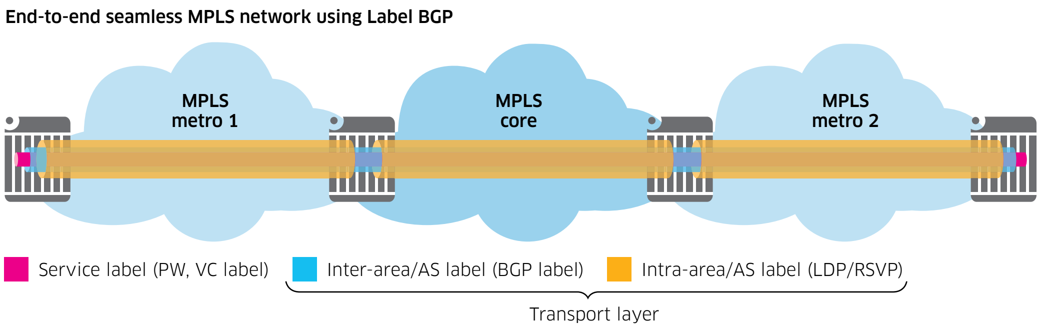
Based on a set of fairly common assumptions, the Seamless MPLS draft (draft-ietf-mpls-seamless-mpls) is one of the architecture options available to extend MPLS networks to integrate metro and core regions into a single MPLS domain. It provides a robust framework which enables service providers to deploy scalable and flexible end-to-end services.

Seamless MPLS architectures are being evaluated and implemented for LTE mobile backhaul architectures and evolution to converged (intra-AS or inter-AS) FMC architectures.

Seamless MPLS offers the following key attributes:

- **Scalability and resiliency (recognizing some nodes have limited capabilities)**: Access nodes can scale in number up to the thousands and are typically optimized for simplicity and lower cost. Seamless MPLS helps scale the end-to-end network to more than 100,000 MPLS devices, recognizing that some nodes (e.g., access) have limited capabilities.
- **A single end-to-end MPLS domain**: Seamless MPLS extends the core region and integrates metro regions into a single MPLS domain. This single domain makes managing and troubleshooting the transport and services layer more efficient.
- **A network without boundaries (seamless)**: Seamless MPLS allows MPLS-based services to be established between any two endpoints, without per-service configuration in intermediate nodes.
- **Rapid Fault detection and recovery**: Seamless MPLS supports end-to-end fault detection, fast protection (and end-to-end operations, administration and maintenance (OAM) functions.
- **Decoupling of transport layer from services**: Seamless MPLS allows services to be provisioned wherever they are needed, no matter how the underlying transport is laid out. This is achieved by implementing a three-layer hierarchy as shown in figure 2 consisting of a transport layer and a service layer. The subsequent sections will describe the three-layer hierarchy along with building blocks for Seamless MPLS.

Figure 2. Decouple transport layer from services


End-to-end seamless MPLS network using Label BGP
MPLS metro 1
MPLS core
MPLS metro 2
Service label (PW, VC label)
Inter-area/AS label (BGP label)
Intra-area/AS label (LDP/RSVP)
Transport layer

CREATING AN END-TO-END TRANSPORT LAYER

As shown in figure 3, a Seamless MPLS network consists of multiple regions: metro 1, core and metro 2. With other end-to-end MPLS options (e.g., end-to-end Label Distribution Protocol (LDP) in a flat network) Interior Gateway Protocol (IGP) or MPLS signaling information is not contained within the region and is exchanged across regions. This increases the size of routing/forwarding tables as well as the MPLS state within individual routers. The Seamless MPLS model addresses this challenge by introducing a hierarchy of transport and service layers. The Seamless MPLS transport layer consists of an inter-region tunnel and an intra-region tunnel.

Inter-region border gateway protocol (BGP) transport tunnel

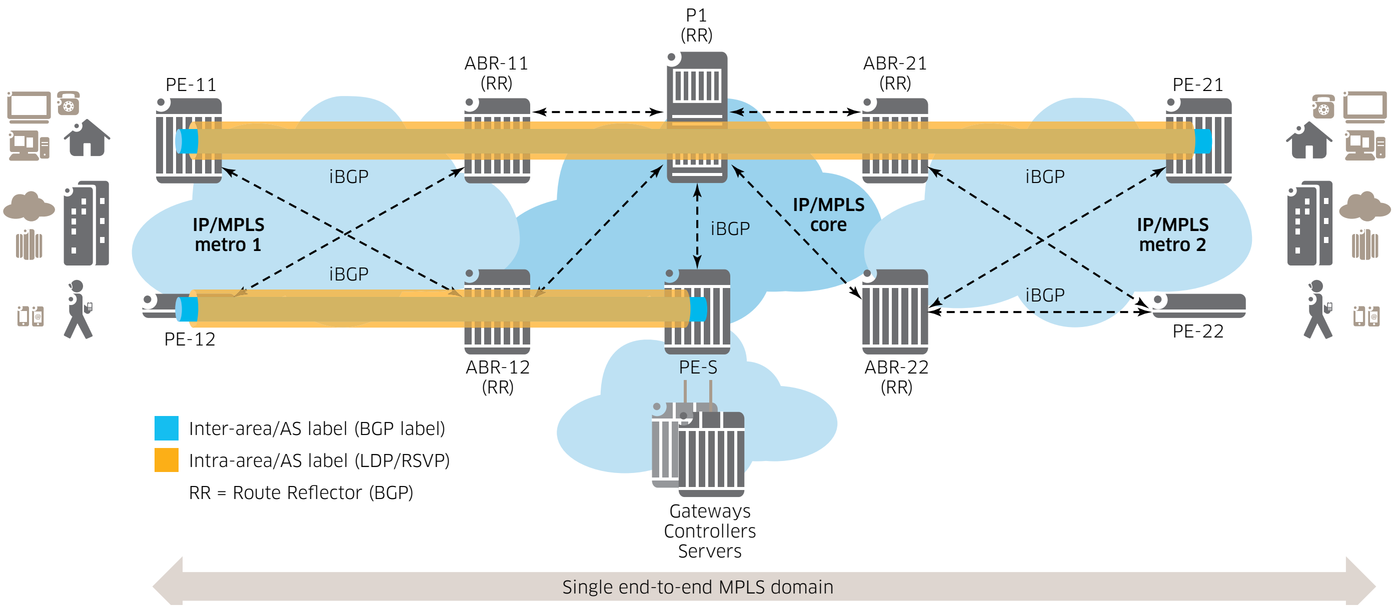
The inter-region transport tunnel must minimize scaling constraints on routing nodes within the end-to-end network. This requires controlling how reachability information is propagated across region boundaries. Each region communicates using end-to-end transport tunnels set up by RFC 3107 (carrying label information in BGP). BGP was built to handle boundary control and is therefore the best choice to create the inter-region transport tunnel.

RFC 3107 (carrying label information in BGP) specifies how the label mapping information for a particular route is piggybacked in the same BGP update message used to distribute the route itself. When BGP is used to distribute a particular route, it can also be used to distribute an MPLS label mapped to that route.

A region may represent an Open Shortest Path First (OSPF) area, Intermediate System to Intermediate System (IS-IS) level, OSPF/IS-IS instance, or even an autonomous system (AS). The Area Border Router (ABR) nodes act as Route Reflectors (RRs) for the region and act as a RR client to the core RRs (P1 in Figure 3). The network represents an inter-area scenario and hence uses iBGP between RRs and RR clients. Provider Edge (PE) loopbacks and label bindings are advertised by Labeled BGP. Figure 3 depicts two inter-region transport tunnels:

1. Inter-region tunnel between PE-11 and PE-21
2. Inter-region tunnel between PE-12 and PE-S

Figure 3. Creating the end-to-end Transport layer using Seamless MPLS



These tunnels provide the PE to PE reachability across regions and provide the inner tunnel label of the transport layer hierarchy. For tunnel 1, the ABR nodes (ABR-21/ ABR-22) receive the loopback and advertise the loopback and a label with next hop self to ABR-11 and ABR-12. These ABRs, in turn, advertise the loopback with next hop self to PE-11. Note: The Seamless MPLS draft suggests that only the local ABRs change the next hop to self (e.g. for PE-21 loopback, ABR-21 changes the next hop to self but not ABR-11). While this approach has some scalability advantages, it requires that PE routers in metro 1 have RSVP or LDP reachability to all ABR nodes in the core area.

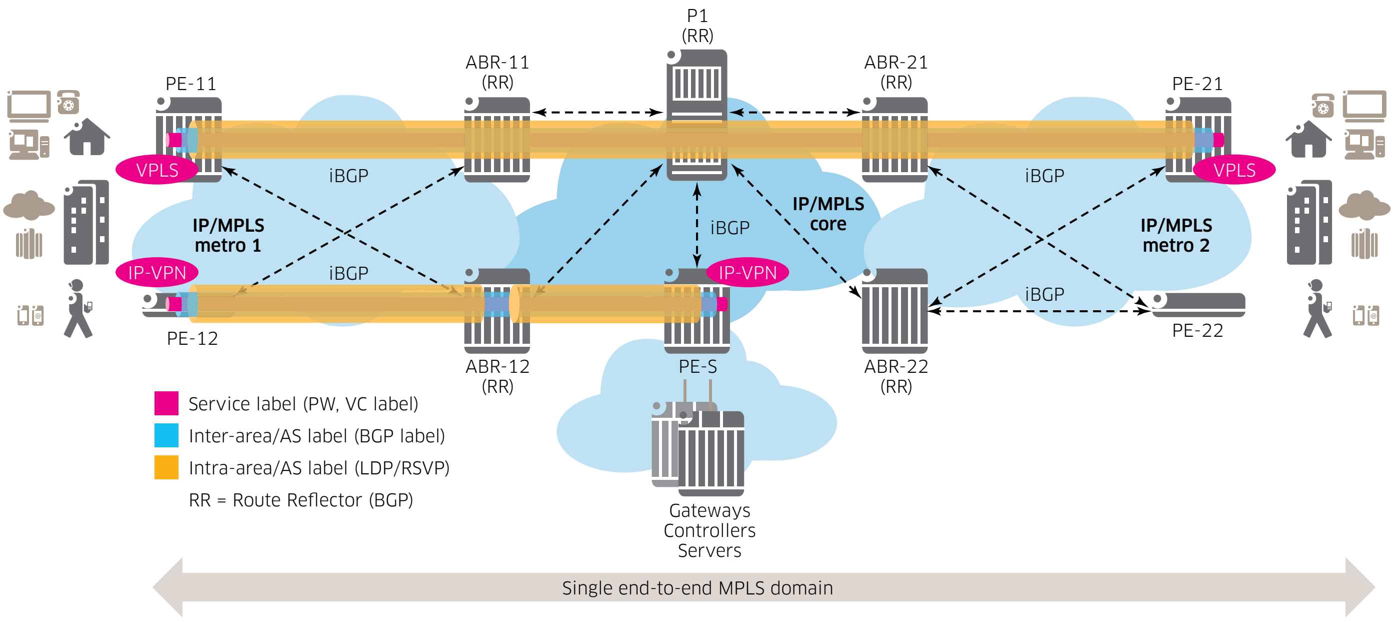
Upon initial configuration, BGP tunnels are established automatically between PE nodes in the end-to-end network using the mechanism described above. This includes tunnels between PE nodes within a metro as well as PE nodes between metros. Often, BGP tunnel connectivity may not be required between all PE nodes. A key benefit of the BGP-based approach is the ability to use BGP policies to limit (permit/deny) propagating loopback reachability to different parts of the network on an as-needed basis. BGP filtering policies based on IPv4 prefixes or BGP communities may be configured on specific nodes within the network to prevent loopback propagation (and hence BGP tunnel creation beyond that point).

Intra-region LDP/RSVP-TE transport tunnel

The intra-region transport tunnels provide transport for the inter-region BGP tunnel within each region. These tunnels provide the outer tunnel label of the transport layer hierarchy. This intra-region tunnel may use LDP or Resource Reservation Protocol with Traffic Engineering (RSVP-TE) and is used to switch the packet between BGP peers (i.e. routes point to the BGP next hop)

IMPLEMENTING SERVICES WITH SEAMLESS MPLS

Figure 4. Implementing services using Seamless MPLS



As described in the previous section, initial configuration creates the inter-region tunnels between PE nodes and intra-region tunnels between BGP peers. This forms the transport layer of the hierarchy. Once the transport layer is created, services can be provisioned end-to-end.

Figure shows two service examples,

1. A Layer 2 Virtual Private LAN Service (VPLS) between PE-11 and PE-21. This may be a business service offered across metro 1 and metro 2. The service label for the VPLS is created using targeted LDP (tLDP) between PE-11 and PE-21.
2. A Layer 3 IP-VPN service between PE-12 and PE-S. This may be a cell site to mobile gateway connection. The service label for the IP-VPN service is created using multiprotocol BGP (MP-BGP) for IP-VPN between PE-12 and PE-S.

Decoupling transport and services layers within the Seamless MPLS framework allows services to be provisioned wherever they are needed, independent of the underlying transport layer. Further, services can be established between any two endpoints, without per-service configuration in intermediate nodes.

SEAMLESS MPLS – SCALING ACCESS NODES

Access nodes typically outnumber aggregation, edge and core nodes in a network by an order of magnitude. Access nodes may include digital subscriber line access multiplexers (DSLAMs), gigabit passive optical networks optical line termination (GPON OLTs), Metro Ethernet switches, customer-located equipment (CLE) and cell site routers (CSRs). Seamless MPLS supports an end-to-end architecture which extends IP/MPLS capabilities to access nodes, recognizing they may have limited capabilities. LDP Downstream on demand (DoD) and LDP forward equivalence class (FEC) to BGP stitching are two features that help extend Seamless MPLS to access nodes.

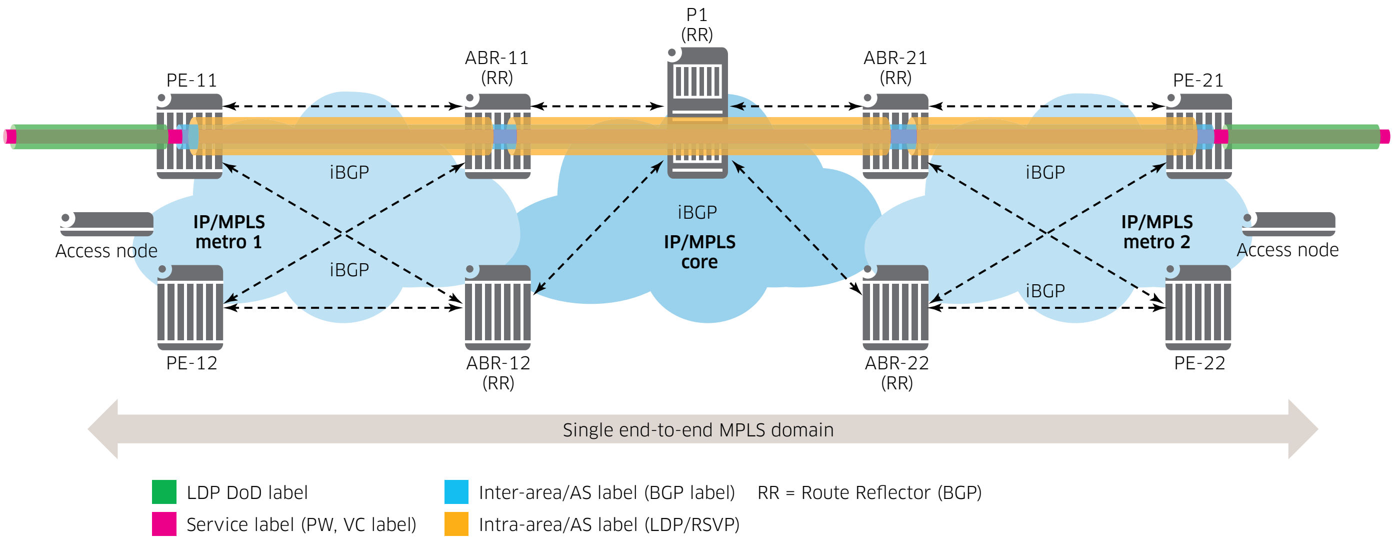
LDP Downstream on Demand (DoD)

Figure 5 depicts the use case for LDP DoD in end-to-end MPLS network design. One of the main goals of Seamless MPLS (defined in draft-ietf-mpls-ldp-dod) is to meet the specific requirements of access devices, based on their position in the network topology and their compute and memory constraints, which limit the amount of state they can hold.

This can be achieved with LDP-DoD, as specified in RFC 5036. In this topology, access nodes can implement a simple IP routing configuration with static routes, limiting number of IP Routing Information Base and IP Forwarding Information Base (IP RIB and IP FIB) entries required on the access node. In general, MPLS routers implement LDP Downstream Unsolicited (LDP DU), advertising MPLS labels for all the loopback routes in their RIB. LDP DoD enables on-request label distribution, ensuring only the required labels are requested, provided and installed.

The access node in figure 5 is configured with static default routes to the PE nodes (access nodes are typically dual-homed to the PE nodes). The access and PE nodes support LDP-DoD. The access node will request a label for a FEC from the PEs using LDP DoD. The aggregation PE nodes reply with the label information, which the access nodes can use to establish the label-switched path (LSP) to the destination.

Figure 5. Extending IP/MPLS to the access node with LDP DoD



Seamless MPLS with LDP-DoD therefore, enables on-request label distribution. This ensures only the required labels are requested, provided and installed, thereby supporting end-to-end architectures where the access nodes may have compute and resource constraints.

LDP FEC to BGP stitching

LDP FEC to BGP stitching may be used along with LDP DoD or LDP DU. The above example describes LDP FEC to BGP stitching with LDP DoD,which may be typical for supporting access nodes with limited capabilities .

The PE nodes in figure 5 perform a translation (LDP FEC to BGP stitching) function.

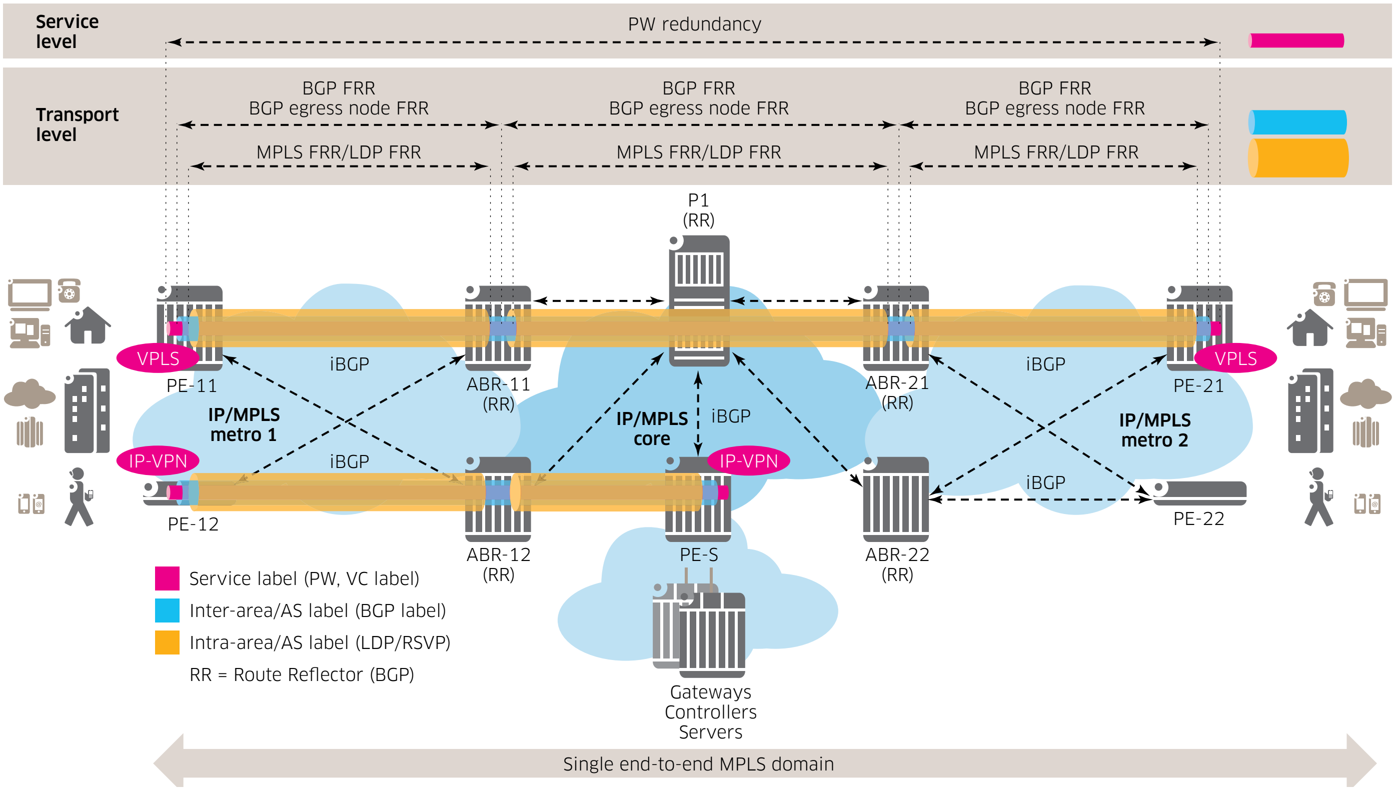
- PE-11 can export an access node LDP Forwarding Equivalence Class (LDP FEC) into BGP and advertise this as a label route using RFC 3107.
- PE-21 translates the /32 BGP labeled routes into LDP FEC and redistribute this FEC to LDP-DU peers and to LDP-DoD peers (access nodes), if requested.

The outermost label represents the LDP tunnels used to switch the packet between BGP peers within the region (intra-region tunnel).The middle label (inter-region BGP tunnel) is used to switch the packet to the destination PE (PE-11 or PE-21 depending on traffic direction). The innermost label is the MPLS service label between the access nodes.

SEAMLESS MPLS END-TO-END RESILIENCY

As shown in figure 6, Seamless MPLS provides end-to-end resiliency at the transport and service layers. The framework supports Pseudowire (PW) redundancy at the service layer. The transport layer supports protection of the inter-region transport tunnel (BGP tunnel), as well as the intra-region (LDP or RSVP tunnel) transport tunnel. During failures, this ensures local fast protection (i.e., LDP FRR, RSVP FRR or BGP anycast) is initiated while end-to-end protocol convergence occurs, which eventually results in new set of BGP transport tunnels being created end-to-end.

Figure 6. End-to-end resiliency with Seamless MPLS



Service layer redundancy

PW redundancy allows PWs to be protected with a pre-provisioned PW and switching traffic over to that standby PW in the event of failure. Normally, PWs are redundant because of the transport tunnel redundancy mechanism. For instance, if the tunnel is an RSVP LSP and is protected by a secondary standby path and/or by Fast-Reroute (FRR) paths, the PW is also protected.

There are a few applications in which tunnel redundancy does not protect the end-to-end PW path, such as when there are two different destination PE nodes for the same Virtual Leased Line (VLL) service. The main use case is to provision dual-homing customer premises equipment (CPE) or access node to two PE nodes located in different points of presence. The other use case is to provision a pair of active and standby broadband remote access server (BRAS) nodes, or active and standby links to the same BRAS node to provide service resiliency for broadband service subscribers. The Alcatel-Lucent end-to-end MPLS toolkit supports basic PW redundancy as well as unique methods to address extended PW redundancy scenarios.

For layer-3 IP-VPN services, BGP egress node FRR mechanisms for the IP-VPN address families (v4 and v6) are supported. These are described later in this document.

Transport layer redundancy – Inter-region tunnel

Inter-region transport layer redundancy is supported using BGP FRR and/or egress node BGP FRR mechanism.

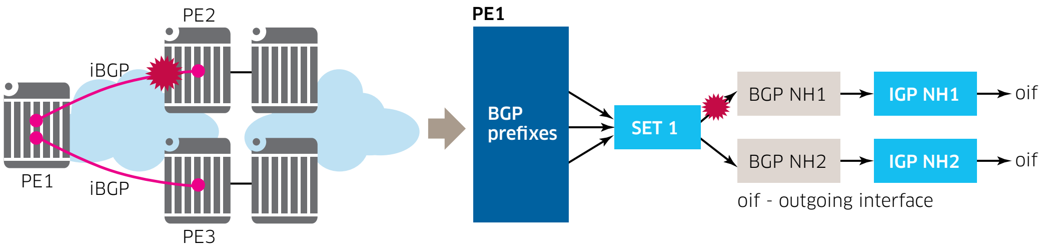
BGP FRR (or Edge PIC)

BGP fast reroute (FRR) or Edge PIC (Prefix Independent Convergence) is a feature that brings together indirection techniques in the forwarding plane and pre-computation of BGP backup paths in the control plane to support fast reroute of BGP traffic around unreachable/failed next-hops.

When BGP fast reroute is enabled, the control plane attempts to find an eligible backup path for every received IPv4 and/or IPv6 prefix, depending on configuration. When BGP decides that a primary path is no longer usable (detected for example using next hop tracking or BFD), the affected traffic is immediately switched to the backup path. Traffic immediately fails over to backup path without the need to wait for the BGP decision process and FIB updates – i.e. the failover time and convergence is Prefix Independent. BGP fast reroute is supported with IPv4, labeled-IPv4, IPv6, 6PE, VPN-IPv4 and VPN-IPv6 routes.

In figure 7, BGP FRR (PIC) capability is enabled on PE1. An alternate BGP next-hop (NH2/on PE3) is provided to the FIB in PE1 along with the primary (best) path (NH1/on PE2). PE1 groups routes with the same next-hops in the FIB so that the time to switch many routes to the backup path is independent of the number of destination prefixes (Prefix Independent). If node PE2 fails, traffic is switched to the backup path (NH2) via PE3. NH1 unavailability may be detected via IGP or BFD mechanisms.

Figure 7. BGP FRR (EDGE PIC)

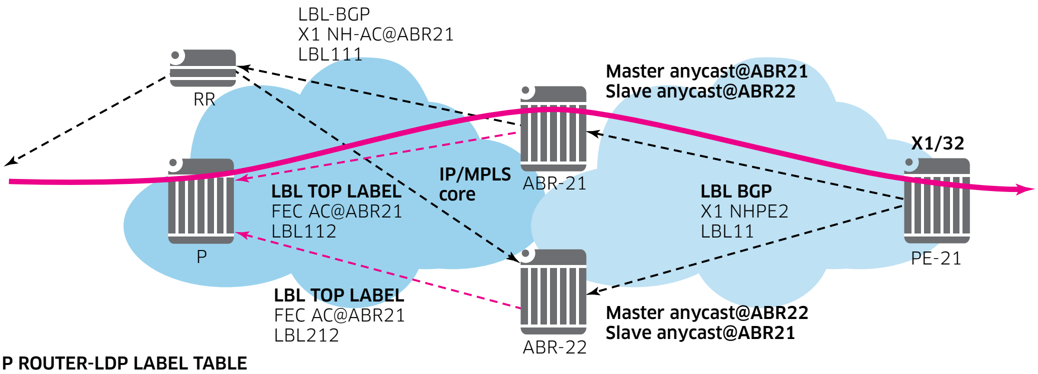


Depending on the topology and the fault detection mechanism, BGP FRR helps reduce BGP convergence to milliseconds.

BGP Egress node FRR (BGP Anycast)

Figure 8 depicts the topology model for BGP Egress node FRR protection. It provides ABR node failure as well as P router to ABR link failure protection within the end-to-end MPLS architecture that supports a RFC3107-based MPLS hierarchy. LDP must be used for the intra-region tunnel (the top or outermost label in the hierarchy).

Figure 8. BGP Anycast for ABR redundancy



P ROUTER-LDP LABEL TABLE

FEC	LBL
FEC: AC@ABR21	LBL 112
FEC: AC@ABR21	LBL212 (A)

SVC LBL
111
112

SVC LBL
11
X

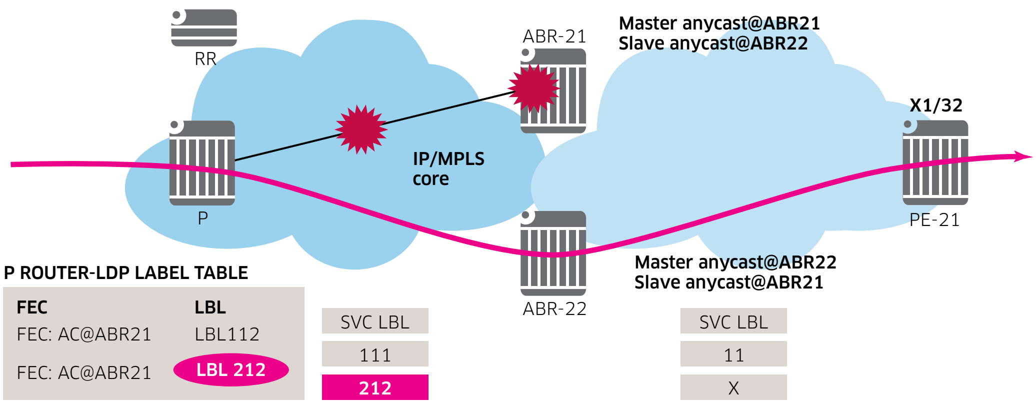
In this model, BGP uses an anycast address in the BGP NH, which means that both ABRs (ABR-21 and ABR-22) in the region can be addressed using this address. This allows for load balancing between ABRs.

PE-21 advertises its own system address (X1/32) with a label using LBL BGP (RFC 3107) to ABR-21 and ABR-22. Both ABRs advertise the network layer reachability information (NLRI), changing the next hop to the master anycast address. The figure shows the NLRI information for address X1, which is reachable via next hop master anycast@ABR21. Both ABRs receive the BGP route from the neighboring ABR, which will be stored in the RIB-in.

Although it is against “classic” BGP rules (that only best routes will be installed), the label received from the neighboring ABR will be installed in context for that neighbor ABR. Interface LDP (ILDP) provides intra-region connectivity and connectivity to the remote ABRs. ABR-21 and ABR-22 both advertise the same master anycast@ABR21 FEC to the P router, so the P router has two labels to reach the same FEC. This makes it possible for the slave to process any packet with the destination that matches the master ABR. Under normal conditions, the outer LDP tunnel (intra-region) uses label 112.

Figure 9 depicts a link or ABR node failure scenario.

Figure 9. Link or node protection with BGP anycast



FEC	LBL
FEC: AC@ABR21	LBL112
FEC: AC@ABR21	LBL 212

SVC LBL
111
212

SVC LBL
11
X

When the link between the P router and ABR fails or if the master ABR fails, the P router immediately switches to the alternate label (label 212) in its LDP label table. When ABR-22 receives the packet, it knows it has to swap the BGP label on behalf of the master. This BGP anycast-based option provides a fast restoration mechanism in the event of an ABR or P router to ABR link failure.

Transport layer redundancy – Intra-region tunnel

Intra-region transport layer redundancy is supported using LDP FRR or MPLS FRR.

LDP FRR using Loop Free Alternate (LFA)

When a local link fails without LFAs, a router must signal the event to its neighbors via the IGP, recompute new primary next-hops for all affected prefixes and only then install those new primary next-hops into the forwarding plane. This is a time consuming process. FRR using LFA reduces the reaction time to milliseconds (tens of milliseconds). LDP/FRR enables IP/LDP packets to be forwarded without waiting for IGP convergence.

For each destination in the network, a backup (alternate) loop-free next-hop is calculated using the IGP LFA calculation specified in RFC 5286. Traffic is sent via the alternate next hop when a link or node failure is detected. LFA coverage is dependent on topology. While dual-homed or full mesh topologies offer good LFA coverage, ring topologies do not. LFA coverage can be improved in ring topologies with shortcuts or remote LFA (draft-ietf-rtgwg-remote-lfa).

LDP FRR using LFA provides a rapid restoration option for architectures that do not require traffic engineering (TE) capabilities.

RSVP FRR (MPLS FRR)

RSVP FRR (defined in RFC 4090) provides the ability to establish backup LSP tunnels for local LSP tunnel repair. This mechanism enables the redirection of traffic onto backup LSP tunnels in 10s of milliseconds, in the event of a failure.

RSVP FRR can be accomplished using two methods, both of which can be used to protect links and nodes during network failure:

1. The “one-to-one backup” method, which creates detour LSPs for each protected LSP at each potential point of local repair.
2. The “facility backup” method creates a bypass tunnel to protect a potential failure point; by taking advantage of MPLS label stacking, this bypass tunnel can protect a set of LSPs that have similar backup constraints

RSVP based FRR (MPLS FRR) may be deployed in segments of the network where traffic engineering is required.

SEAMLESS MPLS OAM REQUIREMENTS

The Alcatel-Lucent “End-to-End” MPLS toolkit supports a comprehensive OAM toolkit for MPLS tunnels and services. This includes LSP-level diagnostics, Ethernet Connectivity Fault Management (CFM), service diagnostics for Layer 2 and Layer 3 MPLS services as well as a Service Assurance Agent (SAA), which allows service providers to configure a number of different tests to provide performance information, such as delay, jitter and loss for services or network segments. SAA functionality is used in conjunction with the OAM tools for performance monitoring.

RFC 6424 – LSP ping and trace route extensions for LSP hierarchy and stitching

The popular IP ping and trace route tools fall short, and therefore need to be supplemented with diagnostics specialized for the different levels within the service delivery model.

In accordance with RFC 6424 (Mechanism for performing LSP ping over MPLS tunnels) the OAM suite must support methods for performing LSP ping and trace over a variety of regular and stitched or hierarchical LSP types: RSVP P2P, RSVP P2MP, LDP unicast and LDP multicast FEC, LDP over RSVP, BGP label route, and LDP FEC stitched to a BGP label route).

LSP ping and trace are extended to support hop-by-hop and end-to-end validation of:

- A BGP RFC 3107 label route when resolved to an LDP FEC or RSVP LSP which represents an LSP hierarchy.
- An LDP FEC stitched to a BGP label route which represents LSP stitching followed by LSP hierarchy.
- A BGP label route stitched to an LDP FEC which represents LSP hierarchy followed by LSP stitching.
- An LDP FEC which is stitched to a BGP LSP and stitched back into an LDP FEC which combines a stitching of two LSP hierarchies.
- An LDP FEC tunneled over an RSVP LSP which represents an LSP hierarchy.

The end-to-end MPLS tool kit supports in-band, packet-based Operation, Administration, and Maintenance (OAM) tools with the ability to test the transport and service layers within a seamless MPLS network.

END-TO-END MPLS DEPLOYMENT – OTHER ALTERNATIVES

Having reviewed the Seamless MPLS model, a brief review of alternative approaches for deploying end-to-end MPLS services along with the pros and cons of each option are described in the following sections.

Flat network using LDP

One way to extend MPLS end-to-end is to implement a flat IP/MPLS multi-area network using LDP. This may be implemented with IGP aggregate route leaking, as defined in RFC 5283 (LDP extensions for inter area LSPs)[1]. RFC 5283 defines a new LDP label mapping procedure to support setting up contiguous inter-area LSPs while maintaining IP prefix aggregation on the ABR nodes. This procedure is similar to the one defined in the LDP specification (RFC 5036), but performs an IP longest match lookup when searching the FEC element in the RIB.

Pros:

- LSP transport tunnels can be created end–to-end.
- Services may be deployed without provisioning intermediate points.
- With prefix aggregation, leaking all the /32s into the IGP area/level is not required, reducing routing table size.

Cons:

- If prefix aggregation (summarized entries) is not supported, the router table sizes can become quite large, burdening routers. In addition, large tables take longer to converge and make troubleshooting complex.
- Even if prefix aggregation is supported and IP RIB/FIB tables are reduced, LFIBs are still flooded with all the /32s FECs of the whole network, reducing the overall scalability and increasing the complexity.
- Reducing FEC distribution requires complex policies.

MPLS network using RSVP

This solution is relevant when RSVP-TE is used to set up intra-area LSPs and inter-area traffic engineering features are required. Conversely, this solution is not ideal when LDP is used to set up intra-area LSPs, and inter-area traffic engineering features are not required.

Pros:

- LSP transport tunnels can be created end–to-end.
- Services may be deployed without provisioning intermediate points.

Cons:

- Router table sizes can get very large and the RSVP state needs to be maintained, affecting performance as the network needs to be scaled. Large routing tables burden routers, take longer to converge and make troubleshooting complex.
- Fine tuning and controlling FEC distribution is quite complex.

LDPoRSVP

LDPoRSVP may be used to improve RSVP scaling and it can even be used for cases in which RSVP is only supported in the network core. It is still essentially LDP end-to-end.

[1] RFC 5283-LDP extension for inter-area LSPs

While a label edge router (LER) may not have many tunnels, any transit node has thousands of potential LSPs, and if each transit node also has to manage detours tunnels or bypass tunnels, the total number can overwhelm the LSR.

With LDPoRSVP, the LDP can now benefit from two RSVP-TE features: TE and FRR convergence below 50 ms. Only user packets are tunneled over the RSVP LSPs, tLDP control messages are sent unlabeled using the IGP shortest path.

Pros:

- A full mesh of intra-area or inter-area RSVP LSPs between PEs is not required.
- LDP transport tunnels can be created end–to-end.

Cons:

- LDP stitching points (ABRs) can take long to converge in case of failures.
- Router table sizes can grow quite large and RSVP state needs to be maintained, affecting performance as the network needs to be scaled. Large routing tables burden routers, take longer to converge and make troubleshooting complex.
- Fine tuning and controlling FEC distribution is quite complex.

PW switching (Multisegment PW)

PW switching allows VLL services to be scaled over a multi-area network by making a full mesh of targeted LDP sessions between PE nodes unnecessary. The end-to-end segment is split into multiple segments that are switched at switching points. PE nodes that terminate the end-to-end service are referred to as T-PEs and the intermediate PEs at the junctions of each segment are referred to as S-PEs.

The T-PE node acts as a master and S-PE nodes act as slaves for PW signaling. The S-PE waits for an LDP-mapping message from T-PEs. The PW is signaled using T-LDP. PW switching limits the propagation of /32s, however, MS-PW requires provisioning at multiple points (T-PE and S-PEs).

Pros:

- Full mesh of targeted LDP sessions between PE nodes is not required.
- Propagation of /32s and router table size are limited.

Cons:

- Need to provision intermediate points, unless dynamic multi-segment PW (draft-ietf-pwe3-dynamic-ms-pw) is deployed in every T-PE and S-PE. This might not be supported on low-end routers or access nodes.
- End-to-end debugging is more complex.
- Only VLL services are supported.

Inter-AS options

RFC 4364, BGP/MPLS IP VPNs describes three options for supporting inter-AS IP-VPNs.

Option A uses back-to-back connections between the autonomous system boundary router (ASBR) nodes. This option does not support end-to-end MPLS and is only suitable when number of IP-VPNs is very small, since it requires per-VPN configuration on ASBRs (i.e., a sub-interface and eBGP session is required for each IP-VPN).

Option B eliminates the need for per-VPN configuration on the ASBRs. The ASBRs receive IP-VPN information from PEs within the local autonomous system (AS) and forward this information to their eBGP peer ASBRs. The peer ASBR, in turn, forwards the IP-VPN information to its local BGP peers within the remote AS. This option is suitable for Inter-AS IP-VPNs between different service providers, as all routes advertised between ASs can be controlled by route policies at the ASBR.

Option C is essentially the seamless MPLS approach to implementing Inter-AS VPRNs and is described mainly for reference. With Option C, VPN prefixes are neither held nor re-advertised by the ASBR. PEs in different ASs can establish multi-hop multiprotocol-eBGP sessions to each other to exchange customer VPN prefixes over these connections. This is achieved by imposing a three-level label stack, which is the the Seamless MPLS architecture model. The bottom-level label is assigned by the egress PE (advertised in multi-hop MP-eBGP without next-hop override) and is commonly referred to as the VPN-label or service label. The middle label is assigned by the local ASBR-PE and corresponds to the /32 route of the egress PE (in a different AS) using BGP-LBL (RFC 3107, Carrying Label Information in BGP-4). The top level label is assigned by the local ASBR-PE(s)/32 loop-back address, which would be assigned by the IGP next-hop of the ingress PE.

Option-C allows for a higher scale of Virtual Private Routed Networks across AS boundaries and also expands the trust model between autonomous systems. As a result, this model is typically used within a single company that may have multiple autonomous systems.

The Seamless MPLS model helps address concerns with alternative approaches and is therefore being evaluated by service providers for implementing end-to-end MPLS networks and services.

CONCLUSION

MPLS is the preferred technology to implement scalable networks for business, residential and mobile services. Several options and technologies may be used to implement end-to-end MPLS networks. Seamless MPLS is preferred, as it provides maximum scalability, flexibility and ease of provisioning and maintenance. Seamless MPLS architectures are being evaluated and implemented for LTE mobile backhaul architectures and evolution to converged (intra-AS or inter-AS) FMC architectures. The technical considerations for implementing a Seamless MPLS network may include evaluating existing network equipment for seamless MPLS features support, as well as interoperability testing across multi-vendor equipment within the metro and core regions.

Some service provider networks may not be ready to implement Seamless MPLS architectures immediately. Alternative approaches described in this paper may be implemented as interim solutions to deploy end-to-end MPLS networks and services. For future scalability and flexibility, however, the seamless MPLS model must be considered.

Alcatel-Lucent, a leader in MPLS development and IP Service Routing offers a complete, comprehensive and industry validated feature set, which enables service providers to migrate to (or implement new) end-to-end MPLS network architectures.

ABBREVIATIONS

ABR	Area Border Router
ATM	Asynchronous Transfer Mode
AS	Autonomous system
ASBR	Autonomous system boundary router
BFD	Bidirectional Forwarding Detection
BGP	Border Gateway Protocol
BRAS	Broadband remote access server
CSR	Cell site router
CFM	Connectivity Fault Management
CLE	Customer-located equipment
DSLAM	Digital subscriber line access multiplexer
DoD	Downstream on Demand
FEC	Forward Equivalence Class
FRR	Fast-Reroute
FMC	Fixed and mobile network convergence
GPON OLT	Gigabit passive optical networks optical line termination
ILDP	Interface Label Distribution Protocol
IGP	Interior Gateway Protocol
IP	Internet Protocol
IP RIB	IP Routing Information Base
IP FIB	IP Forwarding Information Base
LDP	Label Distribution Protocol
LDP DU	LDP Downstream Unsolicited
LDP FEC	LDP Forwarding Equivalence Class
LER	Label edge router
LFIB	Label Forwarding Information Base
LSP	Label-switched path
LFA	Loop Free Alternate
MP-BGP	Multiprotocol BGP
MPLS	Multiprotocol Label Switching
NLRI	Network layer reachability information
OSPF	Open Shortest Path First
OAM	Operations, administration and maintenance
PIC	Prefix-Independent Convergence
PE	Provider Edge
PW	Pseudowire
RNC	Radio network controller
RSVP-TE	Resource Reservation Protocol with Traffic Engineering
RR	Route Reflector
SAA	Service Assurance Agent
tLDP	Targeted LDP
VLL	Virtual leased line
VPLS	Virtual Private LAN Service
VPRN	Virtual Private Routed Network (a.k.a IP-VPN)