White Paper

Extending MPLS Across the End-to-End Network: Cisco Unified MPLS

What You Will Learn

Service providers have used Multiprotocol Label Switching (MPLS) for many years to efficiently manage and control traffic in core networks. Based on this success, MPLS is now being extended to aggregation and access networks to deliver consistent data, control, and operations, administration, and maintenance (OAM) planes in IP Next-Generation Networks (NGNs). However, for MPLS to scale to these new network domains while continuing to support simple operational procedures, changes must be made to the technology.

Cisco, one of the pioneers of MPLS and a longtime global leader in MPLS development, has embraced the challenge of delivering simple-to-operate and highly scalable MPLS technologies to promote this evolution. These new technologies include MPLS Transport Profile (MPLS TP), fast convergence with simplified 50-millisecond restoration schemes, and new OAM capabilities.

This paper examines the motivations, requirements, and solutions for operating MPLS across the end-to-end network. It discusses new Cisco® MPLS innovations in detail, and describes how network operators can employ them to realize the full benefit of end-to-end MPLS in today's IP NGNs.

Challenge: The Evolving Role of MPLS

MPLS has been widely adopted by carriers worldwide, initially in core networks, and now in aggregation networks as well. Carriers continue to realize substantial benefits from MPLS and its support for sophisticated traffic engineering, VPNs, and multiservice-transport-over-packet capabilities.

Since its introduction to service provider networks more than 10 years ago, MPLS technology has undergone continuous innovation and improvement in its operating characteristics, services delivered, and interoperability among equipment vendors. Most recently, users have seen dramatic improvements in the scaling properties and simplicity of operation for MPLS networks. These developments have opened up the potential to extend MPLS to access networks–ultimately providing a single end-to-end data and management plane for modern NGNs.

Two market trends are advancing the adoption of MPLS technologies in access networks. First, accelerating demand for data services on mobile handsets has led to more second-generation and third-generation (2G and 3G) cellular site deployments using packet-based Ethernet backhaul. These cell sites still use Time Division Multiplexing (TDM) and Asynchronous Transfer Mode (ATM) connections in the cell site. The technology of choice to transport those connections to the aggregation network is an MPLS pseudowire (PW) over the packet infrastructure. Conservative calculations suggest these packet-based backhaul solutions provide transport of packet data in the Radio Access Network (RAN) at one tenth the cost of previous technologies.

The second market adopting MPLS in access networks is the wireline service provider segment, which offers transparent Ethernet private-line services over copper and fiber access networks. While networking standards such as IEEE 802.1QinQ and 802.1ad have enabled transparent Ethernet private-line services for some time, they are limited in scale to the traditional limit of 4096 VLANs per trunk, limiting their ability to operate efficiently in access network implementations. These technologies also introduce complex operations into the provider network to

translate VLAN tags. MPLS pseudowires eliminate this scale limitation completely and simplify VLAN tag manipulation, as the VLAN tags are simply stripped off on entry to the pseudowire and reapplied at the egress of the pseudowire.

Both of these market trends are promoting the evolution of packet-based transport technologies that can meet the increasing demand for data traffic in a cost-effective and scalable manner. However, porting MPLS as implemented in today's core and aggregation networks to the access network is not a straightforward task. While the improvements to MPLS in core networks are incremental, the application of MPLS to access networks represents a significant change from existing practice. These changes are examined in the following section.

Solution: Cisco Unified MPLS Technology

To address the expanding role of MPLS within provider networks and further promote the adoption of MPLS among operators with accelerating traffic demands and diverse service requirements, Cisco has created Unified MPLS. Unified MPLS provides an architecture that combines all the latest developments within MPLS to support simplified and highly scalable MPLS deployments.

Considerations and Targets for Unified MPLS

Unified MPLS can be viewed as consisting of two domains that work transparently together to deliver on the promise of a simple-to-operate and resilient end-to-end packet transport network. The first domain includes the core and aggregation networks, and the second domain comprises the access network.

Core and Aggregation Networks

The following are principles for the deployment of Unified MPLS in the core and aggregation network domain.

- **Loop-free alternates provide fast link and node restoration.** Operations to achieve 50-millisecond restoration after a link or node failure can be simplified dramatically by introducing a new technology called loop-free alternates (LFA). LFA enhances the link-state routing protocols (Intermediate System-to-Intermediate System [IS-IS] and Open Shortest Path First [OSPF]) to find alternative routing paths in a loop-free manner. LFA allows each router to define and use a predetermined backup path if an adjacency (network node or link) fails. To deliver 50 millisecond restoration in case of link or node failures, MPLS Traffic Engineering Fast Reroute (MPLS TE FRR) can be deployed. However, this requires adding another protocol (Resource Reservation Protocol, or RSVP) for setup and management of TE tunnels. While this may be necessary for bandwidth management, the protection and restoration operation does not require bandwidth management. Hence, the overhead associated with adding RSVP TE is considered high for simple protection of links and nodes. LFA can provide a simple and easy technique without deploying RSVP TE in such scenarios. As a result of these techniques, today's interconnected routers in large-scale networks can deliver 50-millisecond restoration for link and node failures without requiring any configuration by the operator.
- **Hierarchy must be introduced into the MPLS design to scale aggregation and core networks.** For pure IP networks, operators accomplish this by introducing Border Gateway Protocol (BGP) to split the routing domains into a manageable size, so that the link-state databases of the routing protocols do not become too large. For MPLS networks, however, this approach would require that the Label Switch Paths (LSPs) be combined (or labels stacked) at the point where BGP joins the routing domains together. To eliminate this additional effort and provide end-to-end LSPs across multiple routing domains, RFC 3107 defines procedures for BGP to allocate labels, so that LSPs can be established across BGP boundaries. This technique allows operators to scale the routing design, as well as support end-to-end LSP operation.

- **BGP convergence must be improved to maintain end-to-end resiliency.** In order for BGP to support scaling of the MPLS network, faster convergence of BGP itself is needed. To support this, Cisco developed BGP Prefix Independent Convergence (PIC). Prior to this capability, BGP could take many seconds to converge after a link or node failure. With BGP PIC, convergence now occurs in less than 1 second, and in most cases achieves 50-millisecond restoration.

Access Networks

The second domain consists of access networks. The most significant factors that differentiate the environment for MPLS in access networks, compared to deployment in core and aggregation networks, can be summarized as follows.

- Access networks consist of many more devices than those found in core or aggregation networks. Typical access networks can span 100,000 devices or more, whereas core networks consist of hundreds of devices or fewer.
- Access networks have very simple topologies, either hub-and-spoke, as in the case of wireline central-office-based access, or ring topologies in the case of cell sites and Fiber-to-the-x (FTTx) implementations. This is very different from the much more comprehensive connectivity typically found in core networks.
- Devices in the access network must be optimized for cost, size, and power consumption. This tends to limit their control-plane processing capability when compared to core network devices.
- Due to the large number of devices in access networks, simple operation with cost-optimized network elements is an absolute necessity for operators to cost-effectively deliver service.
- Fast restoration (less than 1 second) is required, without adding protocol complexity to the design.

Each of these considerations must be addressed for MPLS to succeed in access network deployments.

Extending MPLS to Access Networks

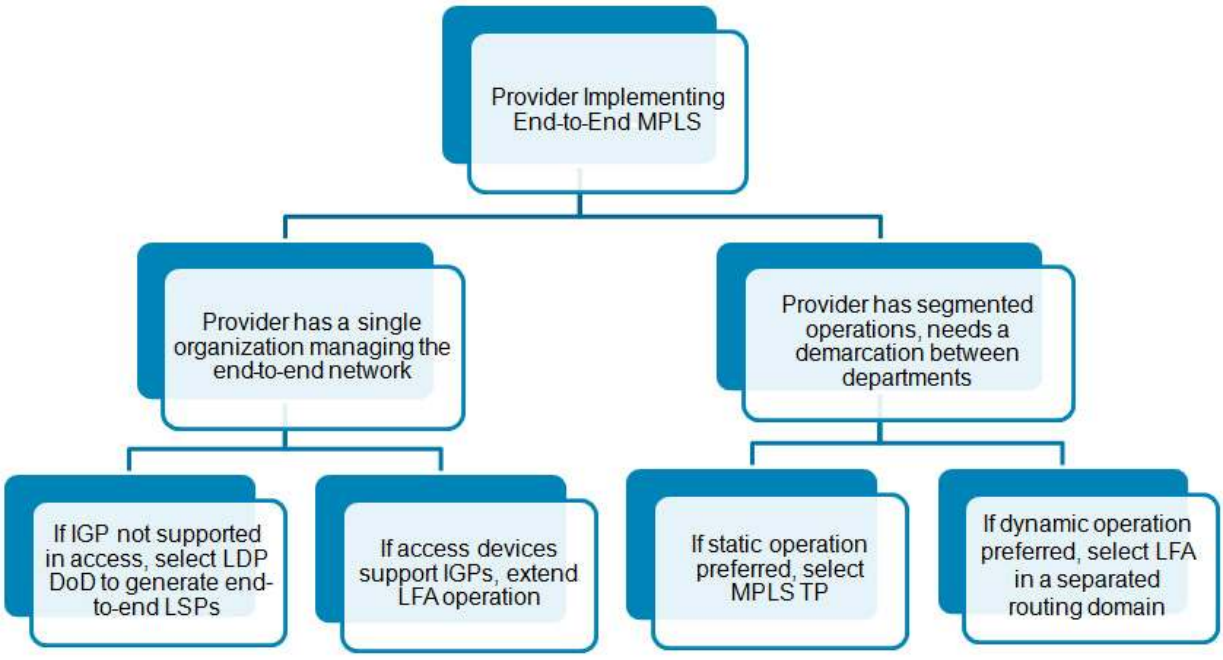
Operators worldwide recognize the following approaches for bringing MPLS to access networks. The first approach aligns with using a static control plane for MPLS (as represented by the initial phase of MPLS TP). The second approach uses developments in the dynamic control plane of MPLS and related routing and label distribution protocols.

The initial static phase of MPLS TP brings a number of benefits to MPLS operation in access networks. First, in-band OAM is introduced into MPLS OAM. This replicates the transport-centric operations familiar to TDM operators and provides a simple mechanism for validating that the data path is operational in the network. In addition, MPLS TP supports restoration mechanisms based on a backup path, rather than single link or node protection. While this approach is not typically feasible in core networks, it is a viable solution for access networks and simplifies operations.

Unified MPLS supports both the dynamic and static approach for access networks, so operators can choose the mode of operation for the access network that most closely aligns with their desired operational models. This choice is typically determined by organizational considerations. For operators using an integrated operations organization that manages the network from end to end, it makes sense for the LSPs to run end to end, as this provides the simplest design to operate. Operators using segmented operations (with separate groups managing access networks and aggregation and core networks) will require the ability to operate these network segments independently.

In the integrated operations case, MPLS services are configured on the edge node, and all nodes in between perform label-switch operations. In the separated operations case, services may be configured on the edge nodes; however, the provider must establish a point of demarcation for the different operational groups to manage their own entities. This demarcation is typically at a point of aggregation in the network where the transport from the access networks is terminated and mapped into an aggregation and core service (Figure 1).

Figure 1. Implementing End-to-End MPLS in Integrated and Segmented Operations Organizations



Overcoming Barriers to MPLS in Access Networks

Two challenges must be overcome before traditional IP/MPLS can be applied to any type of access network.

First, for traditional IP/MPLS, each endpoint requires a unique identifier within the network, which is usually a /32 loopback address that cannot be summarized within the network. As the application of this technology grows to tens or hundreds of thousands of endpoints in access networks, the burden on the routing protocol of having a link-state database containing a /32 address for each endpoint becomes too great. In fact, it becomes computationally infeasible for routers to run the Dykstra shortest-path-first algorithm in such circumstances.

The alternative to having the number of /32 identifiers overwhelm the link-state database is to break the network into separate routing domains to contain the expansion of /32 entries. This may or may not be feasible, based on the objectives of a given service. For example, in some RAN deployments, where the /32 entries are only needed up to the position of the radio network controller in the network, this method may be appropriate. However, in the case of transparent Ethernet private-line services, which typically run end-to-end on a network, this solution is likely infeasible. In either case, additional complexity is introduced into the design with the hierarchy.

The second barrier to overcome when deploying MPLS in access networks is that, in order for traditional IP/MPLS networks to deliver 50-millisecond restoration, traffic engineering is required. This increases protocol complexity due to the need to add RSVP to the network and design a fast reroute tunnel overlay. Neither RSVP nor the tunnel overlay design deliver the simplest possible option for access networks.

Choosing the Right Technology

The choice of technologies to overcome these two barriers is most naturally determined by the capabilities of the access nodes being deployed in the network, and the operator's preference for a dynamic or static control plane.

In the case of wireline operations, the access devices tend to be very simple Digital Subscriber Line Access Multiplexers (DSLAMs) or Passive Optical Network (PON) devices. Such devices typically do not have the capability to implement dynamic control planes but must still meet stringent performance targets. In this case, LFA operation, with its requirement for a link-state database routing protocol, is not appropriate. Thus the choice is between LDP Downstream on Demand (DoD) to support dynamic operation for end-to-end LSPs, or MPLS TP for static operations in segmented organizations. Both options offer simple operations and fast convergence, and eliminate the need to deploy any routing protocol in the access network.

In mobile RAN backhaul networks, the access devices tend to be Ethernet-switched devices that do have the ability to run link-state routing protocols. This gives operators greater flexibility in their choice of operational model. In this case, LFA provides very simple operation and supports the any-to-any requirements of LTE (Long Term Evolution) cellular technology.

The communication requirements of LTE networks are driving more MPLS functionality down to the cell site level. In cases where the new infrastructure buildouts for LTE require virtualization through VPNs to offer transport for multiple services (business services and in some cases residential services as well), we see cell site gateways acting as MPLS Provider Edge routers. This requires new scaling properties that are met with BGP 3107 operations and extend labeled BGP all the way to the cell site.

Alternatively, mobile operators can deploy MPLS TP in the RAN network to support point-to-point operations. This model aligns with the idea of segmented operations between access and aggregation networks, as it provides a natural demarcation point between access and aggregation domains. However, while MPLS TP does offer simple operations with a transport-oriented operational model, it will not readily support the any-to-any connectivity required as these RAN networks transition to LTE without creating a full mesh of MPLS TP static LSPs between the endpoints. The Automatic Neighbor Relation protocol that is part of the Self Organizing Network suite of LTE is used to set up cell-site-to-cell-site communications. Without direct any-to-any communications in the transport network to support this connectivity, additional latency that will likely exceed the 5 millisecond requirements of LTE for handover means that MPLS-TP will not deliver optimal transport for all topologies in an LTE network.

Technologies for MPLS in the Access Network

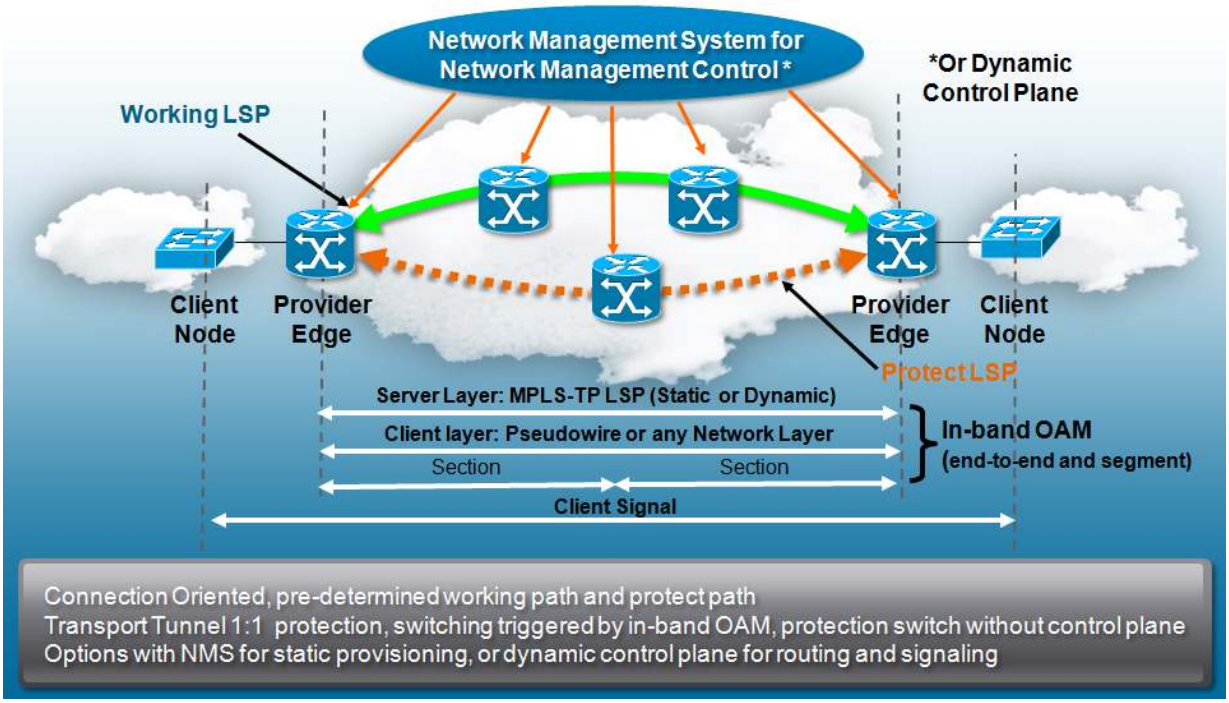
This section provides an overview of the technologies introduced in the preceding sections, and concludes with a summary of how they can be combined in different ways to meet varying operator demands.

MPLS TP in Access Networks

MPLS TP in access networks can address both of the concerns associated with bringing traditional IP/MPLS to access networks – the potential for overloading the router's link-state database, and the added complexity of having to define a separate tunnel overlay for restoration (Figure 2).

As the initial phase of MPLS TP makes use of manually provisioned label-switched paths, no /32 endpoint identifier is required. In fact, no routing protocol or label distribution protocol for point-to-point pseudowires is required at all.

Figure 2. MPLS TP Overview



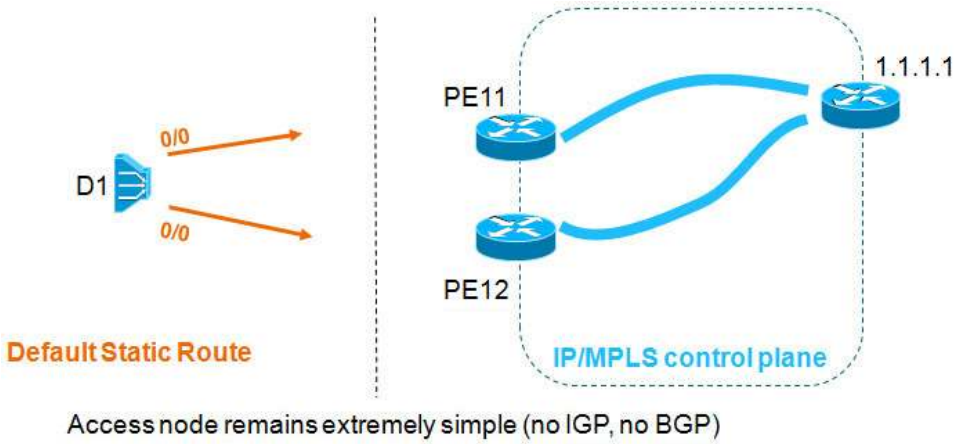
In addition, MPLS TP does not require a separate tunnel overlay to be defined for restoration purposes. Instead, MPLS TP works on the basis of defining an end-to-end primary and backup label-switched path. Using the Bidirectional Forwarding Detection (BFD) protocol, routers and switches continually send fast keep-alive messages down the primary label-switched path. Should three keep-alive messages not be received, the network declares the primary path down and switches traffic to the backup path.

Through these mechanisms, the MPLS TP approach eliminates the concerns about /32 address overload, hierarchical routing protocol design complexity, and fast reroute tunnel overlay design. MPLS TP also works for any topology in access networks. However, the approach is not without drawbacks. Manually provisioning all paths is, by definition, a labor-intensive task that could be done by programmatic means with less overhead. Having only a primary and backup path also eliminates the option to use alternative paths that may be available in the network should both primary and backup paths fail.

LDP Downstream on Demand in the Access Network

Label allocation Downstream on Demand offers an alternative approach for bringing MPLS to access networks, using a simple label distribution protocol implementation without the need to increase routing protocol complexity. Figure 3 illustrates the operation of this approach.

Figure 3. LDP Downstream on Demand Overview



In this figure, the access node (D1) is configured with static default routes to PE11 and PE12, the two points through which the rest of the network can be reached. Once D1 is configured to establish a connection with device 1.1.1.1, D1 requests labels from PE11 and PE12 to reach this destination using LDP Download on Demand. PE11 and PE 12 reply, and D1 can then establish the LSP to its destination.

Through this approach, LDP Download on Demand keeps the access node extremely simple and eliminates propagation of /32 host routes within the network. This simple mechanism requires very little processing capability within the access node, with no routing protocol requirements. Currently, this technology works only for hub-and-spoke topologies; however, it is now being augmented to extend to rings as well. Restoration with LDP Download on Demand is fast, but depending on the capabilities of the access nodes, it may not reach the 50-millisecond threshold.

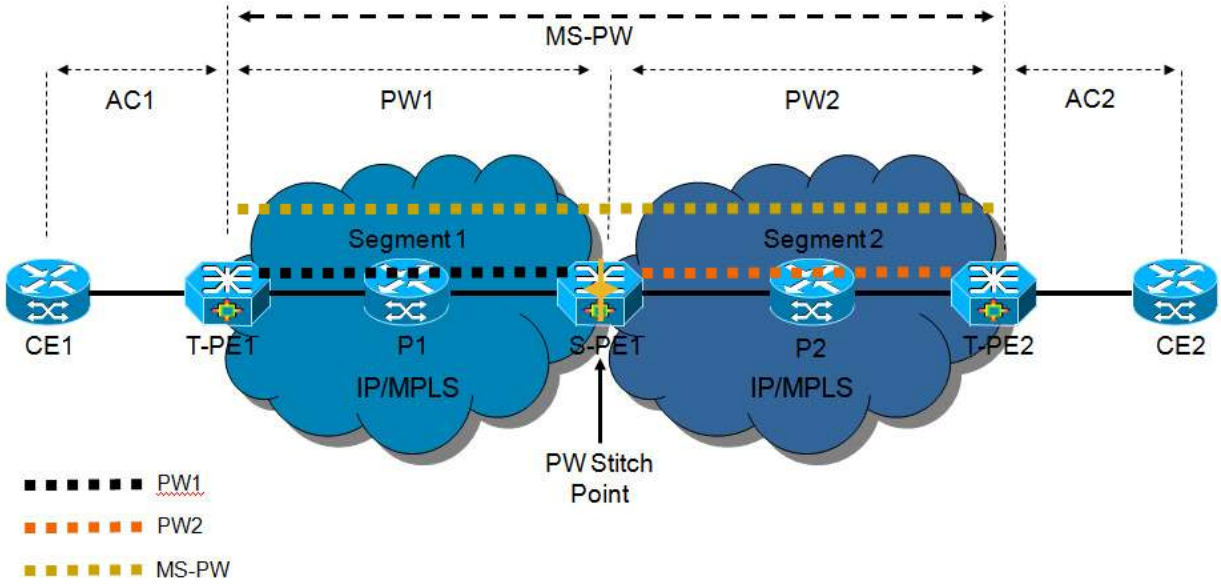
The main point to understand here is that LDP Download on Demand allows very simple devices with limited memory and CPU resources to participate in end-to-end MPLS with acceptable operational characteristics. It is not suggested that LDP Download on Demand is equivalent to or more attractive than the more usual Downstream Unsolicited mode of label allocation. In cases where the network element has enough resources to participate in LDP Downstream Unsolicited procedures, superior operations result.

LFA in the Access Network

If the access devices support Interior Gateway Protocol (IGP) and per-prefix label allocation, LFA for IP/MPLS can offer 50-millisecond restoration with no additional configuration required on the access device.

To support pseudowire operations, LFA will need to be configured with knowledge of all /32 host identifiers in the routing domain. However, a multisegment pseudowire approach provides a way to limit propagation of /32 addresses while still offering end-to-end label-switched paths. Figure 4 illustrates the principles of multisegment pseudowire operation.

Figure 4. Multisegment Pseudowire Operation



In Figure 4, the pseudowire stitch point is a manual connection of the two pseudowire legs, joined through a point-to-point virtual forwarding instance (VFI). (This is considered a point-to-point VFI, as only two segments are attached.)

The mechanism LFA uses to deliver simple 50-millisecond restoration is similar to the Enhanced Interior Gateway Routing Protocol (EIGRP) concept of a feasible successor. An LFA-enabled routing protocol (either OSPF or IS-IS) will predetermine a backup path and, should the primary path fail, start using the backup path immediately when a failure is recognized in the primary path. LFA uses a very simple approach to determine a loop-free path: it is any path that does not point back through itself. Because this logic is implemented within the router as part of the routing computation process, it presents no interoperability issues, as all communications between network elements remain the same. The only potential issue is that a non-LFA-capable router or switch in the network will not offer the same convergence performance. The device will still work in the LFA network, but it will be slower to converge.

RFC 3107 Operation

RFC 3107 defines procedures for having BGP allocate labels to routes between BGP peers. This technique is useful in cases where MPLS networks must scale. RFC 3107 operation can be used to isolate much of the routing data that exists in an MPLS access domain from the core network. By implementing RFC 3107 at the aggregation point, where access networks are aggregated toward the core, BGP label allocation eliminates the need for core devices to learn all of the prefixes in the access domains as routes are summarized. This approach is illustrated in Figure 5.

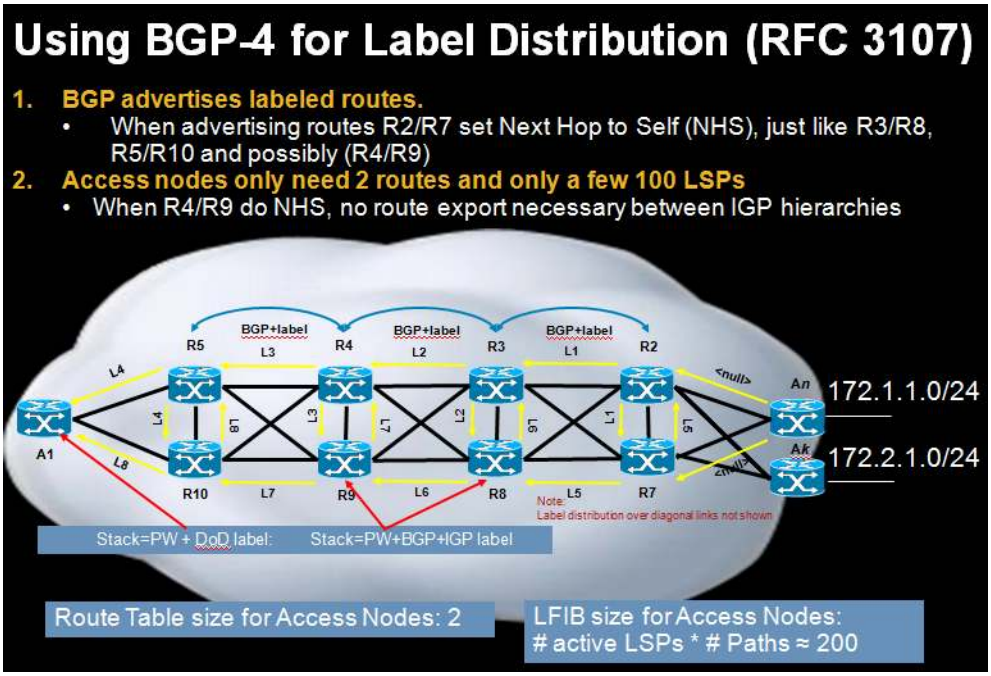
The concept of RFC 3107 in some ways parallels the operation of Layer 3 MPLS VPNs. In Layer 3 MPLS VPNs, the Provider Edge (PE) device allocates two labels to an incoming (unlabeled) packet. The first (outer) label is used to switch the packet on an LSP to a destination Provider Edge. The second (inner) label is used by the destination Provider Edge to identify the interface on which the packet should be sent out.

In the case of an RFC 3107 edge device used to scale deployment of MPLS services, the edge device receives a packet that has already had two labels applied. (These were appended by the device originating the MPLS

service, such as a pseudowire.) The outer label again identifies the LSP, and the inner label identifies the MPLS service. In this case, the RFC 3107 edge device replaces the outer label with two labels, generating a three-label stack. The now outermost label is used to switch the packet across the core between RFC 3107 BGP peers. The second (middle) label is used to direct the packet towards the final edge device in the LSP (once it exits the core network), and the third (now innermost) label is the MPLS service label.

In Figure 5, the blue label is the MPLS service label that remains constant throughout the length of the LSP. The green label is applied by BGP and remains constant in the core of the network, and the red label is used by LDP to switch the packet at each hop. The traffic flows from left to right in this figure.

Figure 5. RFC 3107 in Operation



BGP Prefix Independent Convergence Operation

BGP Prefix Independent Convergence (PIC) is the technology that enables RFC 3107 procedures to be implemented with dramatically improved reconvergence characteristics. Prior to BGP PIC, BGP convergence was slow, potentially resulting in minutes of outage. BGP PIC brings convergence into the range of 50 to 300 milliseconds, depending on topology, with no additional configuration required. BGP PIC is an algorithm enhancement implemented entirely within one routing device, so there are no interoperability issues with non-BGP PIC devices, just improved performance.

The basis of operation for BGP PIC is that the BGP routing process is modified to calculate not only the primary (best) path, but also a repair path in case this primary path to the BGP next hop becomes unavailable. Once the route to a primary next hop fails, the forwarding mechanism of the router points all next hops to the new repair path by updating just a single pointer. This is quicker than doing a prefix-by-prefix calculation, as with the new mechanism, only one pointer must be updated for all the paths that will use that new next-hop address. It is this function of updating a single pointer that is shared by all prefixes using the same next hop that makes this feature prefix-independent.

Conclusion

Cisco Unified MPLS describes the complete end-to-end MPLS architecture that positions MPLS in the data and control plane for every network element performing packet switching in a provider network. Within core networks, the extensive connectivity, relatively low number of devices, and installed base argue for traditional IP/MPLS deployment. Traditional IP/MPLS in this domain supports continued bandwidth growth, multipoint connectivity, and sophisticated quality-of-experience features. Moving toward the access network, however, new challenges must be overcome. This document has described several technologies that can be used to address these challenges. In the access domains, operators can choose from static or dynamic control plane options to suit their organizational structure and operational preferences.

A recent IETF draft proposed by Deutsche Telekom, France Telecom, and Cisco outlines how one combination of Unified MPLS technologies can be applied to deliver simple-to-operate, scalable end-to-end MPLS services. This draft can be reviewed at http://tools.ietf.org/html/draft-leymann-mpls-seamless-mpls-02. The essence of this draft is to use LDP Download on Demand on access DSLAMs, complemented with core operation of LFA, using RFC 3107 at the aggregation routers to scale the MPLS deployment and BGP PIC to help ensure overall convergence characteristics. The Cisco Unified MPLS architecture, however, is not restricted to this model. It allows operators to implement the same network services with static MPLS TP deployed in the access domain as well. Having this flexibility to select the most appropriate operational model provides operators with the ability to maintain internal procedures and gracefully migrate over time toward new operational models.

Due to the continuous innovation of MPLS technologies exemplified by Unified MPLS, MPLS is now a valid choice for supporting unified data, control, and OAM plane deployment across all domains within IP NGNs. The primary innovations described in this document deliver scale properties that make MPLS suitable for deployment in access networks, as well as dramatically simplifying configuration and operational aspects of MPLS to achieve 50-millisecond convergence for failures within the network.

For More Information

To find out more about the Cisco Unified MPLS architecture, visit: www.cisco.com/en/US/products/ps6557/products_ios_technology_home.html or www.cisco.com/go/cpt.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-656286-01 12/11