

SEAMLESS MPLS

Table of Contents

Abstract	3
Introduction	3
MPLS in the Access	3
Benefits	3
Scaling	4
Effectiveness	4
“Thus Far, but No Further”	4
Seamless MPLS	4
Decoupling	5
Network Architecture	5
Components	5
Connectivity Blueprint	5
Scaling	7
Control Plane	7
Putting It All Together	8
Service Restoration	9
OAM: Failure Detection and Isolation	9
Service Architecture	9
Rigid Service Edge	9
Flexible Service Delivery	10
New Service Rollout	10
Conclusion	10
References	11
About Juniper Networks	11

Abstract

Just when you thought that MPLS has peaked, that the pace of innovation has slowed, and that MPLS is getting boring, two promising new developments—namely “MPLS in the access” and “seamless MPLS”—bring fresh excitement to service provider networks. MPLS in the access is evolutionary, but a necessary prerequisite to seamless MPLS, which has the potential to revolutionize the life cycle of service offerings.

Introduction

It is hard to overstate the impact that MPLS has had on service provider networks. In half a decade (1999–2004), MPLS transformed the WAN portion of most service providers. MPLS’s fast-paced deployment can be attributed to two key qualities: its excellent synergy with IP (an almost universally deployed technology) and its versatility. This versatility was evidenced by the wide variety of reasons that MPLS was introduced into networks: for traffic engineering and enhanced quality-of-service (QoS) features; for fast restoration on network failures; for convergence of multiple networks to a single infrastructure; and for a new service, “BGP/MPLS IP VPNs,” which serves both as a technology for service provider-based VPNs for enterprise clients and as a technique for compartmentalization of network elements (see RFC3209, RFC4090, RFC4364, and MPLS Apps).

Over the past half decade, MPLS has made inroads both to the rest of the service provider world as well as to other parts of service provider networks, such as the metro area network (MAN) and access networks. The timing is fortuitous, supporting another sweeping change: the migration of TDM-based infrastructure to Ethernet. Ethernet, while an extremely successful LAN technology, needed help to meet the stringent requirements of service provider networks—MPLS filled in nicely. MPLS has also progressed functionally, with the emulation of point-to-point and multipoint-to-multipoint Layer 2 services (RFC4447, RFC4761, and RFC4762), and the addition of multicast capabilities, both natively in MPLS (RFC4875) and within VPNs (mVPN). These developments are of interest from two points of view: (a) a carrier-grade infrastructure for the metro network, and (b) a vehicle for offering new services.

This expansion outward from the core, however, has been opportunistic and somewhat haphazard. “MPLS in the access” asks the question: Why shouldn’t MPLS be used in all access networks in a systematic fashion? This has several benefits and several challenges. Juniper considers the benefits significant and the challenges solvable. The following sections contain a high-level architecture for MPLS in the access. “seamless MPLS” takes this one step further to an analysis that asks what fundamental change would occur if the entire network were based on MPLS. The result is startling, and offers a new view of MPLS—not just as a network technology, not as a service in itself, or as a service enabler—but also as a vehicle for flexible service delivery. This last aspect has the potential to dramatically change the nature of service offering.

MPLS in the Access

As stated previously, MPLS has proven its value in the WAN, so much so that most WANs are built around MPLS. At the same time, MANs are changing from TDM- and ATM-based networks to ones based on Ethernet. The first question that arises with such a change is what should the underlying infrastructure for an Ethernet-based metro and access network be? The success of MPLS in the WAN naturally suggests the use of MPLS here as well. This leads to the next set of questions. Will the benefits seen in the WAN play out in the metro and access as well? Will MPLS scale to the required extent? Can MPLS fill this role effectively? What leads to seamless MPLS? What else will make this possible? We’ll take these questions in order.

Benefits

The idea of using MPLS for access is not new. It is already being done for some applications such as mobile or DSL backhaul. The benefits of MPLS seen in the WAN are apparent in these applications as well. Thus, the question of how MPLS will improve metro and access networks is clear. Some standards organizations (such as the IP/MPLS Forum and the Broadband Forum) are attempting to formalize these approaches. However, these deployments are somewhat ad hoc. What is being suggested in this paper is the systematic use of MPLS for the entire access and metro network—whether mobile or fixed, residential or business, copper or fiber. Several service providers are looking at the implications of doing this.

Scaling

While the question of the benefits of using MPLS in metro and access networks has been answered, the issue of scalability is harder. WANs generally consist of on the order of 100 to 1,000 Layer 3 devices. A metro network consists of about the same number of devices, albeit both Layer 2 and Layer 3. However, an entire network may contain several dozen metros. Enabling MPLS across this network means having 10 to 100,000 MPLS nodes, a degree of scale not yet seen in today's MPLS networks. Fortunately, we know we can build very large networks, in particular, the public IP network, which consists of several million devices. We will employ many of the same techniques used to build these IP networks to demonstrate an architecture for large MPLS networks.

Effectiveness

The third question, regarding MPLS's effectiveness in the metro and access, will be determined by three factors. The first is whether an MPLS-based metro can be cost-effective. Economies of scale, tighter integration of MPLS and transport, and MPLS's maturity all indicate that this is possible, even inevitable. The second is whether such a network is manageable. Recent efforts that focus on this issue, such as MPLS plug-and-play and improved OAM capabilities, signal a move in the right direction. The final factor is the value of operational convergence, the advantage of a single forwarding paradigm across the whole network. While this may seem obvious from a technical point of view, realizing this advantage may require a structural change in the organization of the service provider.

“Thus Far, but No Further”

Now, consider an ATM-based access network. DSLAMs connect to BRASs via ATM circuits. ATM virtual circuits are offered to enterprise customers as an on-ramp to the Internet or a corporate VPN. ATM even provides the technology for mobile (3G) backhaul. Thus, there is a single “converged” access technology. Unfortunately, this does not extend into the WAN, nor provide most of the services. The border between the ATM access network and the IP WAN network is “stiff,” even rigid—moving it requires redeploying physical devices, maybe even re-architecting the network, and reassigning responsibilities. The real drawback, however, is that this border defines where the majority of services are delivered. This lack of flexibility seriously hampers service providers' business. This is true whether the boundary is defined by a technology (ATM versus IP), geography (metro versus WAN), administration (mine versus yours), or other means. The advent of Ethernet-based metros and access networks does not in itself change this.

Seamless MPLS

This constraint on service delivery brings us to the fourth question. What else does seamless MPLS buy us? But first, a definition: a seamless MPLS network is one whereby all forwarding of packets within the network, from the time a packet enters the network until it leaves the network, is based on MPLS.

In a seamless MPLS network, there are effectively no boundaries (hence the word seamless). This allows very flexible models of service delivery. Some types of services may fare better with centralized delivery while others may work better distributed. Even within a given type of service, different instantiations may require different delivery models—service delivery may evolve over time as the service does. This also greatly simplifies new service offerings. A new service need not be rolled out at every “boundary node”—one can choose to deploy a single server for the service, and defer rolling the service out widely until it proves successful. Similarly, a once-popular service that is now winding down can be contracted down to a small number of servers until the last customer departs. Since delivering services is a service provider's primary business, an architecture that maximizes flexibility in offering and managing services is a key advantage.

Decoupling

There is a crucial point to be made here. Service provider networks exist to deliver services. Thus, there needs to be a tight coupling between network resources and services to optimize quality of experience, robustness, service-level agreements (SLAs), and other metrics. What seamless MPLS offers is a decoupling of network and service architectures—a means of making or changing a network element or design, or even a subnetwork, without affecting services (and vice versa). The choice of ATM here or Ethernet there, the definition of administrative boundaries, the use of different techniques for scaling or resilience—none of these should hinder the ability to deliver a service at a location and in a manner best suited to that service.

Network Architecture

In this section, we will lay out the architecture underlying MPLS in the access and seamless MPLS. This requires new types of network nodes—the traditional partitioning of MPLS nodes into “provider edge” and “provider” devices (in the manner of RFC4364) is simplistic. It also needs new concepts—that of connectivity blueprints, and of transport pseudowires. In this framework, we will show how to scale to 100,000 nodes in a single MPLS network. We will describe the notion of service restoration (as opposed to connectivity restoration), and show what is required to achieve fast service restoration. Finally, we will talk about OAM.

Components

We begin by describing several types of “nodes” in a network, each with a different function. A physical device may combine several of these functions. Conversely, a single function may require multiple physical devices for its realization. There are four types of nodes in the network:

Access Node (AN)—These are the first (and last) nodes that process customer packets at Layer 2 or higher. Examples include DSLAMs, multi-tenant units (MTUs), PON termination devices (OLTs), and cell site gateways (CSGs).

Service Node (SN)—These nodes apply services to customer packets. Examples include L2PEs, L3PEs, Broadband Network Gateways (BNGs), peering routers, video servers, base station controllers, and media gateways.

Transport Node (TN)—TNs connect ANs to SNs, and SNs to SNs. Ideally, TNs have no customer or service state.

Service Helper (SH)—SHs enable or scale the service control plane. SHs do not forward customer data. Examples include service route reflectors, policy and control enforcers, RADIUS and AAA devices, and session border controllers.

Those familiar with the concept of “transport routers” may recognize that TNs are essentially transport routers.

A physical device may of course play multiple roles. For example, an AN may also be a SN—or, a SN may double as a TN. SHs may be embedded in SNs. It is often useful to “virtualize” a physical device that plays multiple roles (using the notion of logical routers) so as to minimize the impact of one role on another, both from a control plane and a management point of view.

Connectivity Blueprint

In addition to the aforementioned network components, there is the notion of an “end node” (EN) that lives “outside” the network, and represents a network customer—from individual subscribers to data center servers. With this, one can depict connectivity from the customer’s point of view as follows:

Basic connectivity: **EN1 <> “network” <> EN2**

Service view: **EN1 <> SN1 <> SN2 <> EN2**

Network view: **EN1 <> AN1 <> SN1 <> SN2 <> AN2 <> EN2**

This is a simple example, but serves to illustrate what’s needed for many point-to-point unicast applications, such as mobile or DSL backhaul.

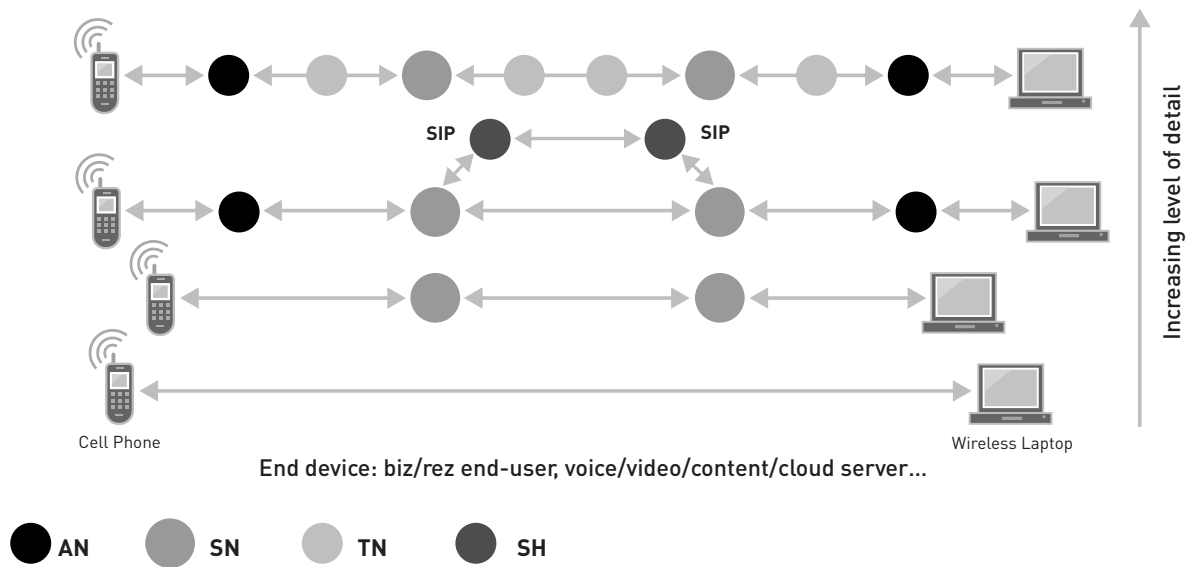


Figure 1: Connectivity Blueprint, from the most basic (bottom) to the most detailed (top)

Connectivity blueprints provide a clear picture of the connectivity and restoration capabilities needed for a particular service. The traditional view of MPLS assumes that full connectivity is needed across all nodes. This vastly increases the degree of scaling in the network. However, considering the previous connectivity blueprint, we see that ANs need not talk to ANs, and in fact only need to talk to their “own” SN.

There can be several other connectivity blueprints, to capture—for example, any-to-any connectivity or multicast.

Transport Pseudowires

An important element of seamless MPLS is the notion of “transport pseudowire.” A transport pseudowire is a pseudowire used within the network for the purpose of moving packets around, as opposed to a service pseudowire, which is used to instantiate a customer service. In the example connectivity blueprint, packets must go from the ingress AN to an SN—at this point, the desired end-to-end service may not be known. If a pseudowire is used between the AN and SN, this would be an example of a transport pseudowire. If on arrival at the SN, it is determined that the desired service is a pseudowire, then there would be a service pseudowire between the SNs.

Transport pseudowires underlie seamless MPLS. They span technologies, traverse boundaries, and negotiate administrative borders to get packets from where they happen to be (say an AN) to where they need to go (say an SN).

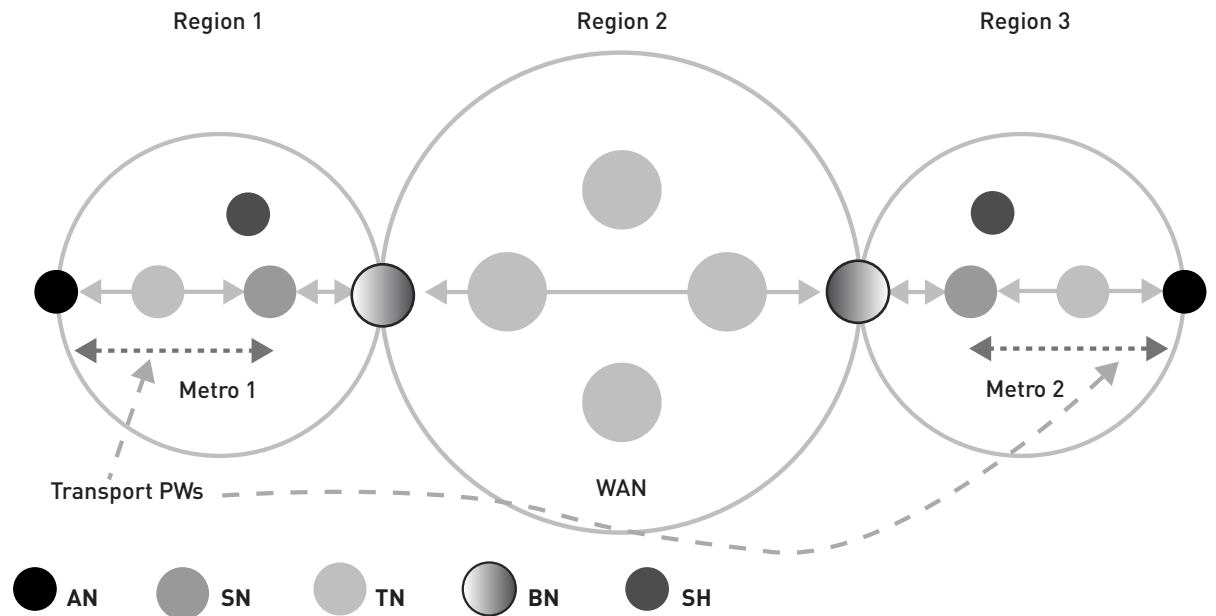


Figure 2: Components, Transport Pseudowires and Regions

Scaling

As indicated earlier, seamless MPLS requires scaling MPLS to up to 100,000 nodes. In this section, we will lay out an architecture to achieve this. One might consider building such a network as a single, large IGP area. However, while this is the simplest architecture, it is very hard to scale. The approach taken here is to divide the network into regions. Three alternatives are considered:

1. Each region is an area of an IGP (classical multi-area/multi-level partitioning).
2. Each region is an independent instance of an IGP.
3. Each region is a BGP Autonomous System running its own IGP.

These are all viable options, and each has its benefits and drawbacks—the choice is open to the provider. Fortunately, the mechanisms described here for scaling the network and services and for restoration apply across all of these options with little change. The notion of regions necessitates the introduction of another type of network node:

Border Node (BN)—BNs are simply TNs that interconnect two or more regions.

Control Plane

There is an infrastructure control plane and a service control plane— separating these contributes to the isolation between the network and service architectures. A single protocol can be used in multiple roles. BGP, in particular, can be used for signaling transport pseudowires (transport BGP or T-BGP)—to create hierarchical LSPs across regions (labeled BGP or L-BGP, see RFC3107)—and to provide services (service BGP or S-BGP). Note that the various names for the protocol indicate the application or role;—there are no changes to the protocol itself. The infrastructure control plane consists of an IGP (IS-IS or OSPF), protocols for label distribution (LDP [RFC5036], RSVP-TE [RFC3209], and ANCP [ANCP]), L-BGP, and signaling for transport pseudowires—using T-BGP and T-LDP.

The service control plane consists of S-BGP (for L2VPNs, L3VPNs, VPLS, and for Internet routing), S-LDP (for VPWS and VPLS), and other applications—including those provided by “service helpers” (see the following section). Examples are voice signaling and IPTV middleware.

The following describes the control plane requirements of each type of network node:

ANs participate in the infrastructure IGP. In addition, ANs need to participate in the creation of transport pseudowires.

SNs participate fully in the infrastructure control plane—that is, they participate in the IGP, label distribution, L-BGP, T-BGP, and/or T-LDP. In addition, SNs participate in the service control plane.

TNs participate in the infrastructure IGP and label distribution.

BNs participate in the infrastructure IGP, label distribution, and L-BGP.

SHs do not participate in the infrastructure control plane, except perhaps basic IGP connectivity. They participate in the service control plane. This architecture does not assume that SHs are MPLS-capable in the data plane.

Putting It All Together

Thus, the network is divided into multiple regions—a large, global network may have as many as 200-300 regions. Each region is an independent, manageable entity. The degree of autonomy depends on the particular type of region that is chosen. Each region has several types of nodes—ideally, these total no more than 500 nodes. All these nodes run an IGP and set up intra-region LSPs using LDP or RSVP-TE, and intra-region transport pseudowires. At the boundaries between regions, there are BNs that mediate control and data plane interactions between the regions, and are also responsible for creating inter-region LSPs over the intra-region LSPs using BGP-based hierarchy. The mechanisms for creating inter-region transport pseudowires depend on the number of such pseudowires needed. For a small number of such pseudowires, either T-LDP or T-BGP may be used, but if a large number of inter-region transport pseudowires is needed, T-BGP may prove a better choice. While the potential number of inter-region LSPs and transport pseudowires can be very large, the connectivity blueprints limit the LSPs and transport pseudowires actually needed to a much smaller, tractable number.

ANs will typically be at the very edge of the network, where they can connect to customers. These will also typically form the majority of network nodes. SNs can be in any region that makes sense—these will be comparatively few. TNs (and BNs) serve to connect up all other nodes. The number of these will depend on geographical spread, the number of metros, and the degree of aggregation.

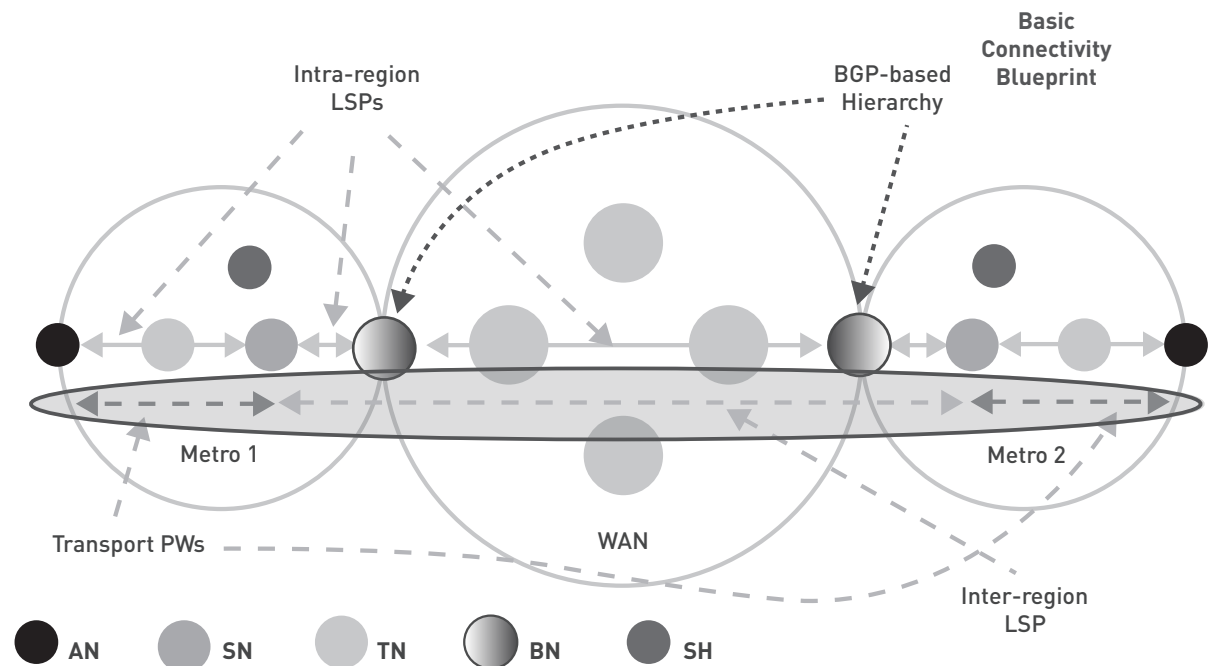


Figure 3: Intra- and Inter-Region LSPs; BGP-based Hierarchy

Service Restoration

Traditional MPLS “fast reroute” focuses on fast connectivity restoration when a link or node in the network fails. However, looking at the connectivity blueprint, it is clear that a failure of a TN has different consequences from a failure of an SN. What is of interest to the end customer is a fast restoration of service, no matter how it is achieved in the network. Thus, there is a need for a clear distinction between connectivity restoration and service restoration. The latter is the desired feature, while the former is a technique that can be used to achieve the former. However, connectivity restoration cannot be used in all cases, in particular, to protect SNs or SHs.

Protecting an SN when a failure occurs can range from an explicit disruption of service that is apparent to end customers (who may have to sign on again), to a very brief outage with a transparent failover to a backup SN. In the former case, the service disruption must be detected, and the AN must fall back to an available SN that can provide the service. In the latter case, one needs fast failure detection—such as made possible by Bidirectional Forwarding Detection (BFD) and Ethernet OAM. One also needs a designated backup SN that has all the required state to transparently take over the service when the primary fails.

Achieving resilience for SHs requires entirely different techniques. SHs typically have underlying databases that need to be resilient. SHs may work in clusters that simultaneously improve resilience and offer load balancing. Checkpointing and virtualization can be used to increase availability. SNs may be given multiple IP addresses for their SHs. Alternatively, SNs may access SHs via anycast IP addresses.

OAM: Failure Detection and Isolation

Failure detection is a prerequisite for any service offering. This is needed at the network element level (link and/or node), transport level (LSP or transport pseudowire), and the service level. Failure detection at the network level is more scalable, can be leveraged across many services, and yields a faster reaction time. Typical techniques used here include BFD (mentioned previously) and LSP ping. However, service-level failure detection will discover several faults that might otherwise be missed, and provides a valuable safety net. Techniques here will be dependent on the service being offered. For MPLS services, LSP ping can be used. For Ethernet services, Ethernet OAM—such as defined by the IEEE or ITU—can be used.

Failure isolation typically uses LSP traceroute, which narrows to the failure to a node or a link in the case of network element failures. For service failures, the SN at fault must be identified.

Service Architecture

Now that we have defined a network architecture that scales and provides the required resilience and manageability, let’s see how services can be delivered in this environment. A simple view of the life cycle of a service offering is plan, provision, launch, expand, contract, and phase-out. The next few sections will illustrate how this is achieved more effectively in a seamless MPLS network. First, let’s define what we mean by service delivery. Packets arrive from the customer at an AN. The AN must then forward these packets to the appropriate SN—based on identifying the customer, the desired service, or both. If the AN identifies the customer (such as by the local loop over which access is affected), it must present this to the SN. Otherwise, the customer must identify himself to the SN—this may require the use of a SH. The SN must then deliver the service (a video stream, a voice call connection, portal access to a Web 2.0 service). Again, doing so may require an SH, and may need the cooperation of other SNs.

Rigid Service Edge

As we mentioned earlier, using different technologies across the network means creating boundaries at the technology junctions—these also form natural boundaries for service delivery. This can be very constraining—rolling out a new service may mean putting (or enabling) service delivery at all these junctions. Thus, the network architecture affects service architecture. Moving service delivery points to optimize bandwidth usage or quality may mean redefining network architecture, or at least redefining boxes.

This approach also requires coordinated provisioning in each technology domain. For example, provisioning a DSL subscriber means provisioning the access line (from subscriber to the DSLAM), provisioning the DSLAM with the appropriate ATM VC (or Ethernet VLAN) across the metro, and provisioning the BRAS at the WAN edge. Troubleshooting requires checking each segment and the crossover points.

Flexible Service Delivery

Contrast this with the seamless MPLS approach—all MPLS means no technology boundaries. Network (or region) boundaries are for scaling and manageability, and do not affect packet forwarding, because it doesn't matter how many hops or region boundaries the transport pseudowire—which carries packets from the AN to the SN—takes or crosses. This is the Holy Grail. The network architecture is about network scaling, network resilience, and network manageability. The service architecture is about optimal delivery—service scaling, service resilience (via replicated SNs), and service manageability. The two are decoupled—each can be managed separately and changed independently.

New Service Rollout

Let's see how this works. Suppose a service provider wants to define a new, experimental service. To keep costs down, it is desirable to roll this out very narrowly, perhaps with just a single SN. This would allow planning to be an easy, lightweight task. There is also a single point of provisioning—a database (such as a RADIUS server) that instructs all ANs how to identify customers for this service, and to which SN to go. ANs exchange signaling with the SN to set up a transport pseudowire to carry the traffic.

If this service proves successful, a wider rollout may be called for (to improve latency, to reduce overall bandwidth consumption, or other reasons). This can be achieved by deploying more SNs and updating the database to direct ANs to redirect their transport pseudowires to the nearest SN. No change of network architecture or movement of network nodes is needed to effect this, just a movement of transport pseudowires. Further, some regions can choose to go to a more distributed service delivery while others (with lower uptake) can stay with a more centralized model. Is service resilience needed? Deploy backup SNs for each existing SN, and update the database with information on primary and backup SNs for each AN. The ANs thus set up transport pseudowires in pairs.

Over time, newer services are offered, and most subscribers move on, away from this service—unfortunately, not all do. The time has come to reduce the investment in the old service. Update the database to point all ANs to a few (or one) residual SNs, de-provision and redeploy the rest of the SNs, and thus shrink the service down—allowing “legacy” users to continue but not burden the whole network.

Conclusion

One goal of any technology is its eventual widespread deployment. As Yakov Rekhter—whom some call the “father of MPLS”—likes to say, “The proof of the pudding is in the eating.” MPLS has fared well by this yardstick, with most features successfully deployed in live networks. These deployments serve two purposes—to validate the technology and to provide an incubation lab for new ideas that build upon the ones being proven.

Over the next few years, we hope to see deployments of MPLS in the access and seamless MPLS. There is interest from vendors, operators, and standards bodies to see this happen. A real-life demonstration that MPLS can really scale to very large networks may bring up issues that weren't anticipated. The implementation of a new service offering life cycle will point the way to further innovation in services. The hope is that these developments—interwoven with others taking place now—will take communication to the next level of ubiquity, availability, and utility.

References

- "Protocol for Access Node Control Mechanism in Broadband Networks,"
S. Wadhwa, J. Moisand, S. Subramanian, T. Haag, N. Voigt, R. Maglione, draft-ietf-ancp-protocol, 2008.
- "Carrying Label Information in BGP-4," Y. Rekhter, E. Rosen, May 2001.
- "RSVP-TE: Extensions to RSVP for LSP Tunnels," D. Awduche,
L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow, December 2001.
- "Fast Reroute Extensions to RSVP-TE for LSP Tunnels,"
P. Pan, Ed., G. Swallow, Ed., A. Atlas, Ed., May 2005.
- "BGP/MPLS IP Virtual Private Networks (VPNs),"
E. Rosen, Y. Rekhter, February 2006.
- "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP),"
L. Martini, Ed., E. Rosen, N. El-Aawar, T. Smith, G. Heron, April 2006.
- "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling,"
. Kompella, Ed., Y. Rekhter, Ed., January 2007.
- "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling,"
M. Lasserre, Ed., V. Kompella, Ed., January 2007.
- "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs),"
R. Aggarwal, Ed., D. Papadimitriou, Ed., S. Yasukawa, Ed., May 2007.
- "LDP Specification,"
L. Andersson, Ed., I. Minei, Ed., B. Thomas, Ed., October 2007.
- "Multicast in MPLS/BGP IP VPNs,"
E. Rosen, R. Aggarwal, draft-ietf-l3vpn-2547bis-mcast , 2008.
- "MPLS-Enabled Applications: Emerging Developments and New Technologies,"
I. Minei, J. Lucek, Wiley, 2008.
- "MPLS: Technology and Applications,"
B. Davie, Y. Rekhter, Morgan Kaufmann, 2000.

About Juniper Networks

Juniper Networks, Inc. is the leader in high-performance networking. Juniper offers a high-performance network infrastructure that creates a responsive and trusted environment for accelerating the deployment of services and applications over a single network. This fuels high-performance businesses. Additional information can be found at www.juniper.net.

Corporate and Sales Headquarters

Juniper Networks, Inc.
1194 North Mathilda Avenue
Sunnyvale, CA 94089 USA
Phone: 888.JUNIPER
(888.586.4737)
or 408.745.2000
Fax: 408.745.2100

APAC Headquarters

Juniper Networks (Hong Kong)
26/F, Cityplaza One
1111 King's Road
Taikoo Shing, Hong Kong
Phone: 852.2332.3636
Fax: 852.2574.7803

EMEA Headquarters

Juniper Networks Ireland
Airside Business Park
Swords, County Dublin,
Ireland
Phone: 35.31.8903.600
Fax: 35.31.8903.601

Copyright 2009 Juniper Networks, Inc. All rights reserved. Juniper Networks, the Juniper Networks logo, JUNOS, NetScreen, and ScreenOS are registered trademarks of Juniper Networks, Inc. in the United States and other countries. JUNOSe is a trademark of Juniper Networks, Inc. All other trademarks, service marks, registered marks, or registered service marks are the property of their respective owners. Juniper Networks assumes no responsibility for any inaccuracies in this document. Juniper Networks reserves the right to change, modify, transfer, or otherwise revise this publication without notice.

To purchase Juniper Networks solutions, please contact your Juniper Networks representative at 1-866-298-6428 or authorized reseller.

