Layer 3 VPN Services over IPv6 Backbone Networks: Requirements, Technology, and Standardization Efforts

Pradosh Mohapatra and Chris Metz, Cisco Systems
Yong Cui, Tsinghua University

ABSTRACT

Layer 3 virtual private networks (L3VPN) enable organizations to connect geographically dispersed sites to one another across the packet switched network of a service provider. The most popular form of L3VPN is based on BGP/MPLS (border gateway protocol/multiprotocol label switching) technology in which the service provider offers a network-based IP VPN routing and forwarding service to its customers across its own IPv4-based MPLS backbone network.

With the deployment of IPv6-based backbone networks underway, there is an emerging requirement to support these same L3VPN services across a native IPv6 backbone network. This introduces a requirement to provide routing and tunneling of IPv6 VPN (and IPv4 VPN) packets across an IPv6 backbone network. Softwires is an Internet Engineering Task Force (IETF) Working Group chartered to address the requirement of providing a generalized, network-based, multi-address family, IP routing and tunneling capability across native IP backbone networks pursuant to IPv6 transitions. Elements of the Softwires work can form the basis of an L3VPN over IPv6 solution.

After providing a brief overview of how L3VPN works in various topologies, this article presents the requirements for L3VPN services over an IPv6 backbone network and discusses a possible solution set that builds over the softwire technology and related IETF standards. Finally, we outline future directions and how the softwire technology can support new services and improved scalability.

INTRODUCTION

The rapid growth of the Internet and the inevitable expiration of available IPv4 network addresses forced the Internet Engineering Task Force (IETF) to develop a next generation Internet Protocol (called IP version 6). Although the existing IPv4-based technology was enhanced over the years to cope with the problem of IP address space depletion, this did not remove but rather delayed the need for a new network protocol as a long-term solution. In fact, it is quite clear that no IPv4 improvement can guarantee the end-to-end network transparency and cater to the growing demand of new services such as security, mobility, multimedia integration, quality of service (QoS), and end-to-end service level agreements (SLA).

Thus, most of the major service providers already are seriously contemplating the introduction of a capable IPv6-based transport function. The success of the 6bone test bed network which is where most of the Internet service providers (ISPs) experiment with IPv6 proved the seriousness and interest of providers to build large scale IPv6 backbones. Indeed, currently, there are operators who have deployed IPv6 backbone networks.

The versatility of service providers was enhanced over past years by their ability to provide routing services to attached constituent client networks. For example, [1] defines a method of providing provider network-based IPv4 VPN services over an IPv4-based MPLS backbone. An extension of this framework to provide IPv6 VPN services over an IPv4/MPLS backbone was proposed in [2]. These VPN technologies have gained much ground due to their capability of offering cost-effective, secure, and private network-like services. A comparable solution to provide the same set of services over an IPv6 network does not exist yet. This article aims to define and explore various schemes to map the same technology over a native IPv6 backbone network.

The rest of the article includes the following: we give an overview of the existing L3VPN technology. We describe the extension to the base L3VPN specification to support IPv6 VPN services over an IPv4-based MPLS backbone. This is followed by a discussion on the new requirements in the context of an IPv6 backbone. We

0163-6804/07/$20.00 © 2007 IEEE

then explore a framework built on existing techniques that can provide transparent IP VPN services over an IPV6 backbone. Finally, we conclude by summarizing and outlining future work.

BASICS OF L3VPN

The model of physical connectivity of L3VPN as defined in [1] is illustrated in Fig. 1. There are a number of *customer sites*, which are assumed to have connectivity within the site that does not use the backbone. Each site has a number of customer edge (CE) devices, connected to provider edge (PE) devices, in the backbone network. The backbone provides transit between PE devices, possibly using internal provider core routers, or P routers. These routers do not require a VPN related state. It is assumed that the backbone is provided by one or more network providers and that the sites are owned and managed by customers. The promise of L3VPN is that it provides each customer site in a corporate or campus network the capability to plug into the backbone network for connectivity with other sites across geographically dispersed areas. The routing and forwarding instance of a particular customer network is kept private and separate from other customer networks that also are plugged into the backbone. Thus, the network is considered as a *virtual private network*.

The following sub-sections describe the L3VPN architecture in brief. The basic idea behind the architecture is to use MPLS to transmit VPN packets between PE devices within the backbone network. This enables the P devices to be unaware of the VPN and have knowledge of how to forward packets with certain labels. These labels correspond to the PE addresses and are set up using an interior gateway protocol (IGP) and an MPLS signaling protocol, such as the label distribution protocol (LDP). VPN specific information is exchanged between PE devices (only) using the BGP [3] and employing the multiprotocol extension (MP-BGP) [4]. BGP is a routing protocol where two devices, called peers, maintain a transmission control protocol (TCP) session and exchange routes. There are two types of such sessions: an interior BGP (IBGP) session for peers within an autonomous system (AS) boundary and an exterior BGP (EBGP) session for peers across AS boundaries.

VPN ROUTING AND DISCOVERY

The control plane signaling required to set up L3VPN services is illustrated in Fig. 2 and includes the following steps:

- IPv4 routing protocol (static/dynamic) exchange between the CE device and the PE router so that all the customer prefixes are learned by the PE. The PE maintains a separate routing instance to store the customer prefixes for each VPN that is configured on the router.
- Multi-protocol BGP signaling between the PE devices to exchange the customer prefixes. The following information is sent with each prefix:

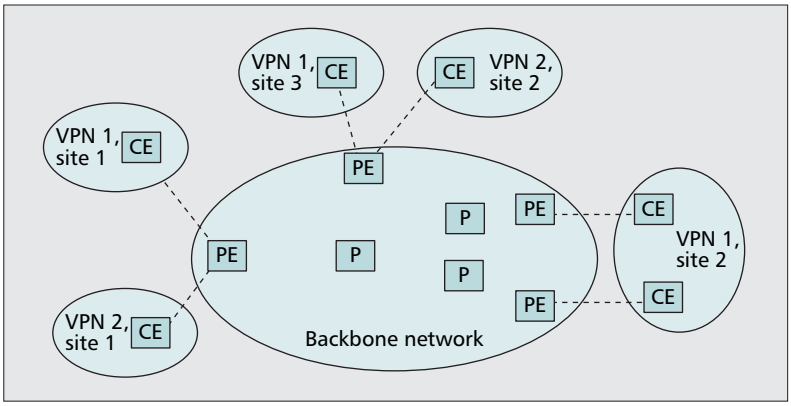

Figure 1. *Basic components of L3VPN service.*

  –As one customer's prefixes can overlap with prefixes from another customer, each customer prefix is converted to a VPNv4 address by prepending an 8-byte entity, called a route distinguisher (RD), to the prefix.
  –To control the distribution of prefixes to the PE devices, so that each PE stores prefixes only for the VPNs that it is attached to, the VPNv4 prefix advertisements contain another attribute called route targets (RT). Each customer VPN attached to a PE is configured with the RD and a set of RTs.
  –VPN label for the prefix. This is used by other PE devices to encapsulate the customer's IP packets with an MPLS header. At the advertising PE, the packet is forwarded to the appropriate CE by looking up this label in the forwarding table.
- When the BGP VPNv4 prefixes are received at a PE, the receiving PE filters them using the received and configured RTs. If the filtering succeeds, it installs the prefixes into the corresponding VPN routing instance. Then, these prefixes are advertised to the attached CE in each VPN context so that each site has connectivity to every other site.

VPN FORWARDING OVERVIEW

Figure 2 shows an example of the forwarding table at PE1 that is connected to two separate VPNs: VPN 1 and VPN 2. For VPN 1, d1 and d2 are local prefixes that are learned from the CE, whereas d3 and d4 are remote prefixes learned from other PE. There are two scenarios with respect to VPN forwarding that are of interest:

- Ingress forwarding: CE1 attached to PE1 sends a packet destined to d3. Since the packet arrives at PE1 on an incoming interface that belongs to VPN 1, PE1 performs destination IP lookup in the VPN 1 forwarding table. The lookup returns:
  "*d3 → first-hop = P1, labels= [L1, L2], interface=if2*"
  Here, L2 is the VPN label for prefix d3 that was advertised from PE2, and L1 is the transport label that will be used in the core to reach PE2. PE1 thus pushes two labels on the packet and sends it out to the core.

The implicit assumption when discussing the L3VPN solution suite is that the attached client networks are IPv4 or IPv6 and that the backbone is native IPv4 or IPv4-based MPLS. Although there is no change in the makeup of such attached client networks, the same cannot be said for the backbone.

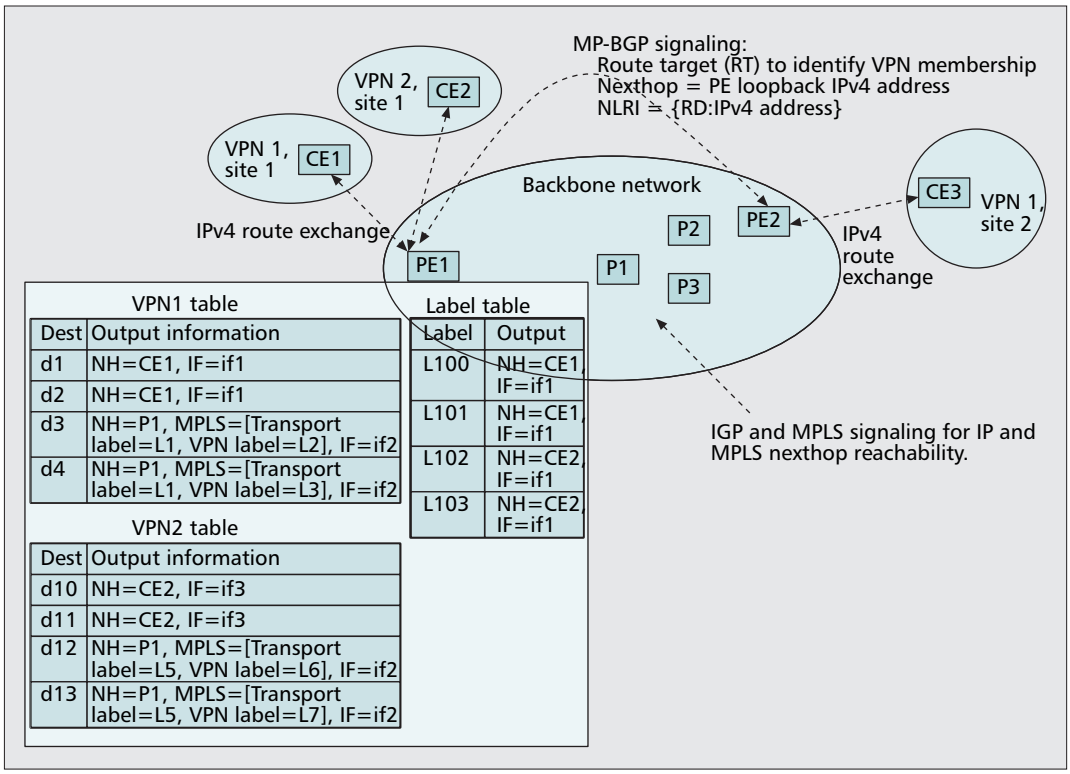

Figure 2. *L3VPN routing and forwarding.*

- Egress forwarding: CE3 sends a packet destined to d1. The behavior at PE2 is similar to what was described previously. When the packet reaches PE1, if there is a transport label, it is popped, and a lookup is performed in the label table for the VPN label. Figure 2 shows an example of the label table. The packet is forwarded to the corresponding CE, based on the result of the label lookup.

Inter-AS VPN

When a VPN contains two or more sites that are connected through different autonomous systems (that possibly belong to different service providers), normal routing and communication methods as described previously cannot be used. Three different options are described in [1] to provide site connectivity for these inter-AS or inter-provider VPN. They rely on exchanging VPN prefixes between border routers (termed as autonomous system border router or ASBR) and building an end-to-end MPLS label switched path (LSP).

Existing L3VPN Technology for IPv6

The feature to facilitate the RFC (request for comments) 4364-like VPN model for IPv6 networks is referred to as 6vPE [2]. It takes advantage of operational IPv4-based MPLS backbones and extends BGP signaling to carry VPN IPv6 prefixes across the core without a requirement for the core routers (P) to run dual-stack. The PE routers run in a dual-stack mode that:

- Exchanges IPv6 routing information with the CE.
- Advertises and receives VPNv6 prefixes across the IPv4/MPLS core.

6vPE is more like a regular IPv4 MPLS-VPN provider edge, with an addition of IPv6 support within the context of a VPN. This is illustrated in Fig. 3.

6vPE offers service providers a straightforward, incremental approach for adding IPv6 VPN services over an existing IPv4 MPLS backbone network.

New Requirements

The implicit assumption when discussing the L3VPN solution suite is that the attached client networks are IPv4 or IPv6 and that the backbone is native IPv4 or IPv4-based MPLS. Although there is no change in the makeup of such attached client networks, the same cannot be said for the backbone. Some network providers have deployed (or plan to deploy in the near future) native IPv6 backbone networks [5]. This introduces the requirement to support IPv6 VPN and IPv4 VPN connectivity across a native IPv6 backbone network.

At first glance, an obvious solution would be to augment the native IPv6 backbone with an MPLS control and forwarding plane. However, it is not quite clear if or when that may happen for several reasons. First, MPLS already works quite well in conjunction with IPv4 to build backbone LSP that handle different applications including IPv6. Second, implementing MPLS on IPv6 routers adds cost and complexity that providers wish to avoid while rolling out the next generation of Internet pro-

tocol in its early stages. Indeed, the first adopters of IPv6 backbone networks have been quite content to rely on basic IPv6 routing and forwarding rather than the addition of MPLS label switching and the attendant MPLS signaling protocols (operating over IPv6) that would be required. Moreover, IPv6 was designed to provide better QoS, security, mobility, and other functionality that is not sufficiently addressed by IPv4 and MPLS.

Assuming a native IPv6 backbone, one way to support IPv6 VPN services is to employ manually configured tunnels between IPv6-speaking CE or PE routers [6]. The payload of packets transported through these tunnels would be a customer's IPv6 or IPv4 packets and the tunnel header would be IPv6.

A mesh of inter-CE IPv6 tunnels imposes the configuration burden of *O (number of remote CE routers in the VPN per local CE)* on the CE router administrators but places no demands on the provider's IPv6 backbone beyond IPv6 routing and forwarding. An advantage, of course, is that multiple discreet customer VPN can be overlaid on top of the provider's network.

Manually configuring inter-PE IPv6 tunnels is more problematic. First, the providers must set up and maintain the mesh of IPv6 tunnels. Second, the providers must configure per-VPN routing policies on each ingress PE to ensure that inbound VPN packets are injected into the correct IPv6 tunnel and on each egress PE to direct the de-encapsulated customer VPN packet towards the correct destination CE. Both actions are configuration-intensive, and all customer VPNs must share the IPv6 (and IPv4) address space maintained on the PE routers.

A more desirable and scalable technique is to generalize the L3VPN protocol machinery so that IPv4 or IPv6 VPN packets can be tunneled across a native IPv6 packet switched network (PSN). MP-BGP addresses the problem of exchanging IPv4 VPN and IPv6 VPN routing information across the IPv6 backbone. The challenge comes in providing a scalable means by which those IP VPN packets can be transported across a native IPv6 backbone.

L3VPN over Native IPv6 Framework

The IETF formed a new working group called Softwires, chartered to identify and develop dynamic and scalable techniques that facilitate IPv6-over-IPv4 and IPv4-over-IPv6 tunneling. Two problem spaces are defined [7]: hub and spoke and mesh. Hub and spoke examines solutions for IP tunneling within an access network, and mesh looks at the dynamic routing and tunneling of one type of address family (e.g., IPv4) across a backbone network supporting a different address family (e.g., IPv6).

Most germane to our discussion of L3VPN across an IPv6 backbone is the softwire mesh framework [8]. It is built on the notion (subscribed to by L3VPN) that a backbone network supporting the routing and forwarding of one address family (called internal IP or I-IP) can offer transit service between attached client networks supporting a different address family (termed external IP or E-IP).

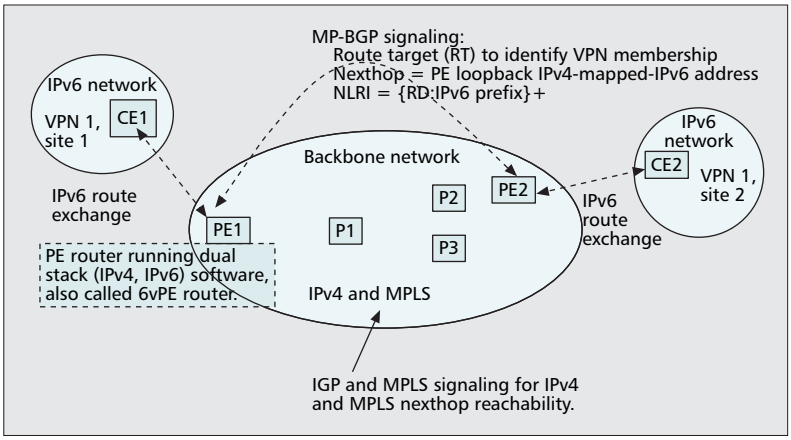

Figure 3. *6vPE operation.*

The P routers in the backbone network only deal with I-IP routing information sufficient to forward an I-IP addressed packet between PE routers across the backbone network. The PE routers, called address family border routers (AFBR), are dual-stack in that they maintain I-IP and E-IP routing information. The former is composed of internal backbone routes leading to other AFBR and P routers. The latter is composed of client E-IP routes and associated next hop, egress AFBR nodes. Source client E-IP packets traversing the I-IP backbone on their way to distant E-IP destinations must be tunneled between the ingress and egress AFBR nodes using I-IP headers.

An inter-AFBR tunnel is called a softwire, and it uses standard IP (e.g., [9, 10]) or MPLS encapsulation headers that are imposed and disposed by the ingress and egress AFBR nodes respectively. AFBR nodes peer with one another to learn each other's softwire encapsulation (e.g., I-IP) parameters (if required) and to exchange E-IP reachability information. Thus, the scope of the softwire effort is to create a generalized, routing and forwarding solution for tunneling E-IP packets across an I-IP backbone supporting two basic scenarios: a) E-IP is IPv6, and I-IP is IPv4 and b) the opposite, where E-IP is IPv4, and I-IP is IPv6.

Indeed, the softwire mesh solution reuses much of the L3VPN protocol suite including MP-BGP as a means to advertise E-IP reachability information between peering AFBR nodes. However, softwire mesh moves beyond L3VPN in the following aspects:

- E-IP prefixes maintained on the AFBR nodes can be treated as VPN and stored in multiple private routing tables or as global and stored in a single public routing table. In the latter case, global E-IP routing is enabled between peering AFBR nodes connected via softwires crossing the I-IP backbone. For example, a provider's IPv6 backbone could provide softwire mesh connectivity between islands of global IPv4 connectivity. Again, we note that the provider's IPv6 backbone routers do not see or process any IPv4 (E-IP) routes.

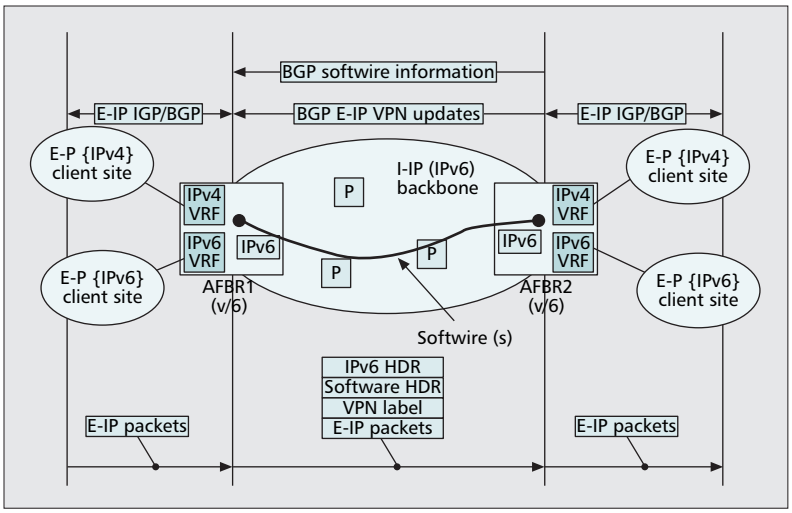

Figure 4. *L3VPN over IPv6 based on the softwire mesh framework.*

- IPv6 is supported as an I-IP type. This is not explicitly excluded when discussing L3VPN but is called out here naturally as part of a particular IPv6 transition scenario. Thus, in theory, currently and likely in future practice, the providers deploying a native IPv6 backbone will offer the same diverse E-IP connectivity (e.g., IPv4 VPN, IPv6 VPN) to client sites attached to their backbones as do their counterparts that are currently running IPv4 or MPLS in their backbones.
- Defines a uniform (and optional) method for automating the distribution of softwire encapsulation information between AFBR nodes. This does not supersede existing tunnel set-up methods including manual configuration or MPLS signaling. Rather it employs BGP as the delivery transport enabling an egress AFBR node to inform all ingress AFBR nodes of encapsulation types and associated parameters supported on that egress AFBR [11]. This type of information would be of interest to the ingress AFBR set that could be required to forward E-IP packets across the I-IP backbone to the egress AFBR.

For example, an egress AFBR, configured with L2TPv3 session id and cookie header parameters, could supply this information along with its own I-IP IPv6 address to the ingress AFBR set. The ingress AFBR nodes have sufficient knowledge of the encapsulation information to apply, if and when they are required to forward packets, across the I-IP IPv6 backbone to that egress AFBR.

This BGP distribution technique reduces the amount of encapsulation configuration work from $O(N)$ to $O(1)$ for each egress AFBR. It also is advantageous, for example, if the attributes of the tunnel must be updated in real-time (e.g., L2TPv3 cookie rollover [12]).

Figure 4 illustrates the softwire mesh framework applied to the L3VPN over the IPv6 backbone case. The AFBR nodes attach to E-IP IPv4 and E-IP IPv6 client sites and store those routes in the respective VPN tables. MP-BGP is employed to communicate AFBR softwire encapsulation information that results in the creation of the softwire mesh. E-IP reachability information in the form of E-IP VPN prefixes, labels, route targets, and so on is exchanged between AFBR nodes using MP-BGP. E-IP packets arriving at the ingress AFBR are resolved to a destination E-IP prefix, an I-IP next hop, corresponding VPN label, and softwire encapsulation action and then, emitted into the network. At the egress AFBR, the E-IP packets are de-encapsulated and sent to the attached customer network.

The mechanics of the softwire mesh framework can certainly be used to support a solution for L3VPN over native IPv6 backbones. However, one subtle and very important detail must be addressed, and it pertains to the cases when MP-BGP advertises IPv4 VPN or global IPv4 E-IP reachability across an IPv6 backbone network. The detail in question is how to present and encode the BGP next hop information. Before discussing solution alternatives, we note the following:

- MP-BGP [4] has been interpreted to mean that both the routes being advertised, called network layer reachability information (NLRI) and the next hop address belong to the same network layer protocol, denoted by address family identifier (AFI) and subsequent address family identifier (SAFI) fields in BGP messages.
- Existing L3VPN solutions in operation over IPv4 or IPv4-based MPLS networks have cleverly circumvented this constraint by prepending bits to an IPv4 address in the next hop address fields such that it appears to share the same AFI/SAFI as that of the NLRI. This is doable because the length of the NLRI fields in the VPN cases (96 bits for VPNv4; 192 bits for VPNv6) are much larger than a 32-bit IPv4 address that fits into the next hop address field.

The situation now exists where the E-IP IPv4 next hop AFBR is not reachable across an I-IP IPv6 network. We no longer have the luxury of prepending bits to the next hop address to meet the AFI/SAFI NLRI match constraint. (You cannot fit a 128-bit IPv6 address into a 32-bit next hop address field.) It seems reasonable to relax this constraint so that a next hop address can enjoy its own AFI/SAFI designation decoupled from the AFI/SAFI designation of the NLRI.

The Softwires Working Group defined a solution [13] that allows a BGP update message to carry NLRI of a different AFI/SAFI than that of the next hop. It involves exchanging capability code points between BGP-speaking AFBR nodes at the time of session establishment that announce their ability to advertise NLRI of a particular AFI/SAFI pair with a next hop whose network protocol is determined by the value of the length of the next hop field. Thus, if the next hop length is 32 bits, then it is IPv4; if the length is either 128 bits or some multiple, it is treated as an IPv6 address. This option fits well into the framework for carrying E-IP IPv4 prefixes across a native I-IP IPv6 backbone by encoding an IPv6 next hop in the BGP updates.

Another option involves definition of a new SAFI to indicate a particular NLRI and next hop combination. For example, the AFI/SAFI of 1/128 identifies the IPv4 VPN address family, and a new SAFI value of X would designate the IPv4 VPN with an IPv6 next hop address. When factoring in the requirement for a unicast and multicast SAFI for each and examining the number of different NLRI and next hop combinations under consideration, one sees that this could result in SAFI explosion. In addition, this would highly complicate operator configuration and BGP processing when attempting to keep track of the different SAFI. This does not seem like a viable solution in the long term.

Conclusions and Future Work

It is obvious that the majority of the provider backbone networks are IPv4 or IPv4-based MPLS and likely will remain so for the foreseeable future. The existing L3VPN solutions based on RFC 4364 have proven to be quite flexible and mature in their operation and can be easily extended to accommodate IPv6 VPN unicast and multicast.

However, that is not to say that native IPv6 provider backbones will not appear. CERNET2, for one, is quite active on this front. Its requirement to support a dynamic and scalable BGP-based IPv4-over-IPv6 tunneling solution partially inspired the creation of the Softwires effort in the IETF [5]. Although the scope of Softwires is IPv6-over-IPv4 and IPv4-over-IPv6 routing and tunneling, we showed that the fundamental underpinnings of this work, (that is, BGP distribution of multi-AF reachability and encapsulation information coupled with IP tunneling of client E-IP packets across a provider backbone) can address the L3VPN over IPv6 solution requirement. With respect to multicast, it is envisioned that client E-IP multicast traffic can be accommodated across a multicast-enabled IPv6 backbone, employing mechanisms defined in [14].

How might the notion of L3VPN over IPv6 evolve? It is safe to say that any IPv6 backbone must tunnel legacy IPv4 networks together. That could involve IPv4 VPN, depending on the operational and service deployment choices made by the operator. In fact, we could say that viable IPv6 backbone operators will demand legacy global IPv4 and IPv4 VPN capabilities from day one for the basic reason that IPv4 traffic dwarfs IPv6 traffic, and single-purpose (e.g., IPv6 only) network infrastructure is not easy to cost-justify in this age of multi-service network convergence.

Another area is the added value that an IPv6 backbone provider can offer to the attached client networks. Managing per-customer VPN route distribution and then, forwarding over special-purpose tunnels traversing the backbone, could constitute a unique service offering above and beyond just standard default routing. This introduces the requirement on the part of the network administrators to configure and manage prefix-to-tunnel routing policies on the ingress AFBR nodes, such that packets destined for a particular prefix are directed into the appropriate IP tunnel.

As L3VPN over IPv6 services mature, the requirements for fast convergence and minimal traffic loss on network failures also will become imperative. As such, native IP networks offer faster convergence properties compared to MPLS networks, as the forwarding modification is prefix independent for IP networks. We also will see more emphasis on IP fast reroute technology on native IPv6 networks to provide minimal traffic loss on link failures.

As L3VPN over IPv6 services mature, the requirements for fast convergence and minimal traffic loss on network failures also will become imperative. As such, native IP networks offer faster convergence properties compared to MPLS networks, as the forwarding modification is prefix independent for IP networks.

References

[1] E. Rosen and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPN)," RFC 4364, 2006.
[2] J. De Clercq *et al.*, "BGP-MPLS IP VPN Extension for IPv6 VPN," RFC 4659, Sept. 2006.
[3] Y. Rekhter, T. Li, and S. Hares, "A Border Gateway Protocol 4 (BGP-4)," RFC 4271, 2006.
[4] T. Bates *et al.*, "Multiprotocol Extensions for BGP-4," work in progress.
[5] J. Wu *et al.*, "The Transition to IPv6, Part I: 4over6 for the China Education and Research Network," *IEEE Internet Comp.*, May 2005, pp. 54–59.
[6] A. Conta and S. Deering, "Generic Packet Tunneling in IPv6 Specification," RFC 2473, 1998.
[7] S. Dawkins, Ed., "Softwire Problem Statement," work in progress.
[8] J. Wu *et al.*, "Softwire Mesh Framework," work in progress.
[9] D. Farinacci *et al.*, "Generic Routing Encapsulation (GRE)," RFC 2784, Mar. 2000.
[10] M. Townsley *et al.*, Eds., "Layer Two Tunneling Protocol — Version 3 (L2TPv3)," RFC 3931, Mar. 2005.
[11] P. Mohapatra and E. Rosen, "BGP Encapsulation SAFI," work in progress.
[12] M. Townsley *et al.*, "Encapsulation of MPLS over Layer 2 Tunneling Protocol Version 3," work in progress.
[13] F. Le Faucheur and E. Rosen, "Advertising an IPv4 NLRI with an IPv6 Nexthop," work in progress.
[14] E. Rosen, Ed., "Multicast in MPLS/BGP IP VPN," work in progress.

Additional Reading

[1] M. Tatipamula, P. Grossetete, and H. Esaki, "IPv6 Integration and Coexistence Strategies for Next Generation Networks," *IEEE Commun. Mag.*, vol. 42, no. 1, Jan. 2004, pp. 88–96.

Biographies

Pradosh Mohapatra (pmohapat@cisco.com) is a technical leader in the Service Provider Routing Group of Cisco Systems, San Jose, California. His current areas of interest include fast routing convergence, IP/MPLS, and layer 3 virtual private networks.

Chris Metz (chmetz@comcast.net) is a technical leader in the Routing Technology Group of Cisco Systems, San Jose, California. His current areas of interest include Internet architectures and services, IP/MPLS, transport layer protocols, and layer 2/layer 3 virtual private networks.

Yong Cui (cy@csnet1.cs.tsinghua.edu.cn) received a Ph.D. in computer science from Tsinghua University, Beijing. He is an associate professor in the Computer Science Department at Tsinghua University. His current research interests include Internet architectures, IPv6 transition, and quality of service. He is supported partly by the National Natural Science Foundation of China (no. 60403035) and the National High Technology Development Program of China (no. 2006AA01Z205).