The Internet Protocol Journal

June 2005 Volume 8, Number 2

A Quarterly Technical Publication for Internet and Intranet Professionals

In This Issue

You can download IPJ back issues and find subscription information at: **www.cisco.com/ipj**

From The Editor

The Internet is a constantly evolving environment which puts pressures on existing and evolving protocols. Any protocol changes must be carefully designed and even more carefully deployed to avoid any disruption to the running system. It is no longer possible to orchestrate a simple overnight switch, so engineers are considering various transition and evolution strategies. In this issue we bring you two examples of this kind of evolutionary protocol development.

Our first example relates to *IP Version 6* (IPv6). A great deal of effort is going into the deployment of IPv6, and good transition strategies can help. Tejas Suthar explains how *Multiprotocol Label Switching* (MPLS) can be used for a transition from IPv4 to IPv6.

Our second example looks at a possible enhancement to the *Border Gateway Protocol* (BGP). BGP in its current form is already nearly ten years old, and calls for its replacement can be heard from network operators. Russ White discusses some possible changes that would not require a wholesale protocol replacement.

It is not every day that a book on punctuation becomes an international best seller, and it is certainly not common for IPJ to review such a non-computer related book. But I think it is appropriate for several reasons. First, accurate punctuation is important not just for computer parsers, it is important for all professionals whether we are sending quick e-mails or writing project reports. Second, this is a really *fun* as well as informative book. And last, but not least, it gives me an opportunity to introduce you to Bonnie Hupton, who provides copy-editing services for this journal. Without her help, IPJ would be far less readable.

Our Website at **www.cisco.com/ipj** has a new look, but still contains links to our back issues, index files and the IPJ subscription system. Please take a moment to renew or update your subscription. If you have questions or comments, please send them to **ipj@cisco.com**.

—Ole J. Jacobsen, Editor and Publisher
ole@cisco.com

IPv6—A Service Provider View in Advancing MPLS Networks

by Tejas Suthar, TELUS Communications Inc.

We are all aware of the evolution of the *Internet Protocol* (IP) and its dominance on all aspects of our lives, either directly or indirectly. Currently IP Version 4 delivers critical business application traffic in a so-called new world of the Internet. As the evolution goes on, *IP Version 6* (IPv6)[5] is becoming a necessary element of the network. IPv6 will enable businesses to expand their capabilities exponentially without having any limitations or restrictions. As technologies evolve and the adoption of IP-enabled devices accelerates, IP will enter a new era as the protocol of choice for communications. Using globally unique IPv6 addresses increases the opportunity for service providers to create new business models and add revenue, and it increases the portfolio of services. However, the major demand for support of IPv6 will be mobile applications; the IT world will also tie in all the systems for transparent operation. The days are not far when permanent IPv6 addresses will be assigned to individuals for their communication purposes—either *Voice over IP* (VoIP), video over IP, video on demand, wireless Internet access, unified messaging, etc. Also, IP smart appliances are becoming more and more popular, and the result will be explosive usage and adoption of IPv6 addresses. Articles outlining the importance of IPv6 and limitations of IPv4 abound. This article is mainly geared toward highlighting the service provider networks that are built or currently being built to support IPv6 in a VPN fashion.

Multiprotocol Label Switching (MPLS)[4] is widely accepted as a core technology for the Next-Generation Internet that provides speed and functions in packet forwarding. Service providers that offer MPLS/VPN services to their customers are looking forward to adding IPv6 VPN services to their portfolio. Service providers that want to support IPv6 in traditional ways have few options, such as tunneling methods (for example, manual, *Tunnel Broker, Generic Routing Encapsulation* [GRE], or *Intrasite Automatic Tunnel Addressing Protocol* [ISATAP], which has scalability problems); or Native IPv6 with dual-stacked MPLS core. However, consider the following:

- For MPLS VPN services, service providers made a significant investment in building the IPv4/MPLS backbone. The return on investment thresholds are probably yet to be achieved.
- Backbone stability is another critical factor; service providers must offer reliable services, especially with regard to voice over MPLS. Most service providers have recently managed to stabilize their IPv4 infrastructure, and they are hesitant to make another significant move when it comes to supporting IPv6 unless the integration is smooth.

Standards bodies with help from vendors and leading service providers are addressing these concerns. Currently service providers have two approaches that they can deploy to support IPv6 without making any changes to the current IP (v4) MPLS backbones, namely *6PE*[1] and *6VPE*[2], originally defined in RFC 2547.

The 6PE approach lets IPv6 domains communicate with each other over an IPv4 cloud without explicit tunnel setup, requiring only one IPv4 address per IPv6 domain. The 6PE technique allows service providers to provide global IPv6 reachability over IPv4 MPLS. It allows one shared routing table for all other devices. Typical applications are IP toll voice traffic and Internet transit services over a common MPLS infrastructure. The 6PE technique does not provide any logical separation because it is for MPLS VPN.

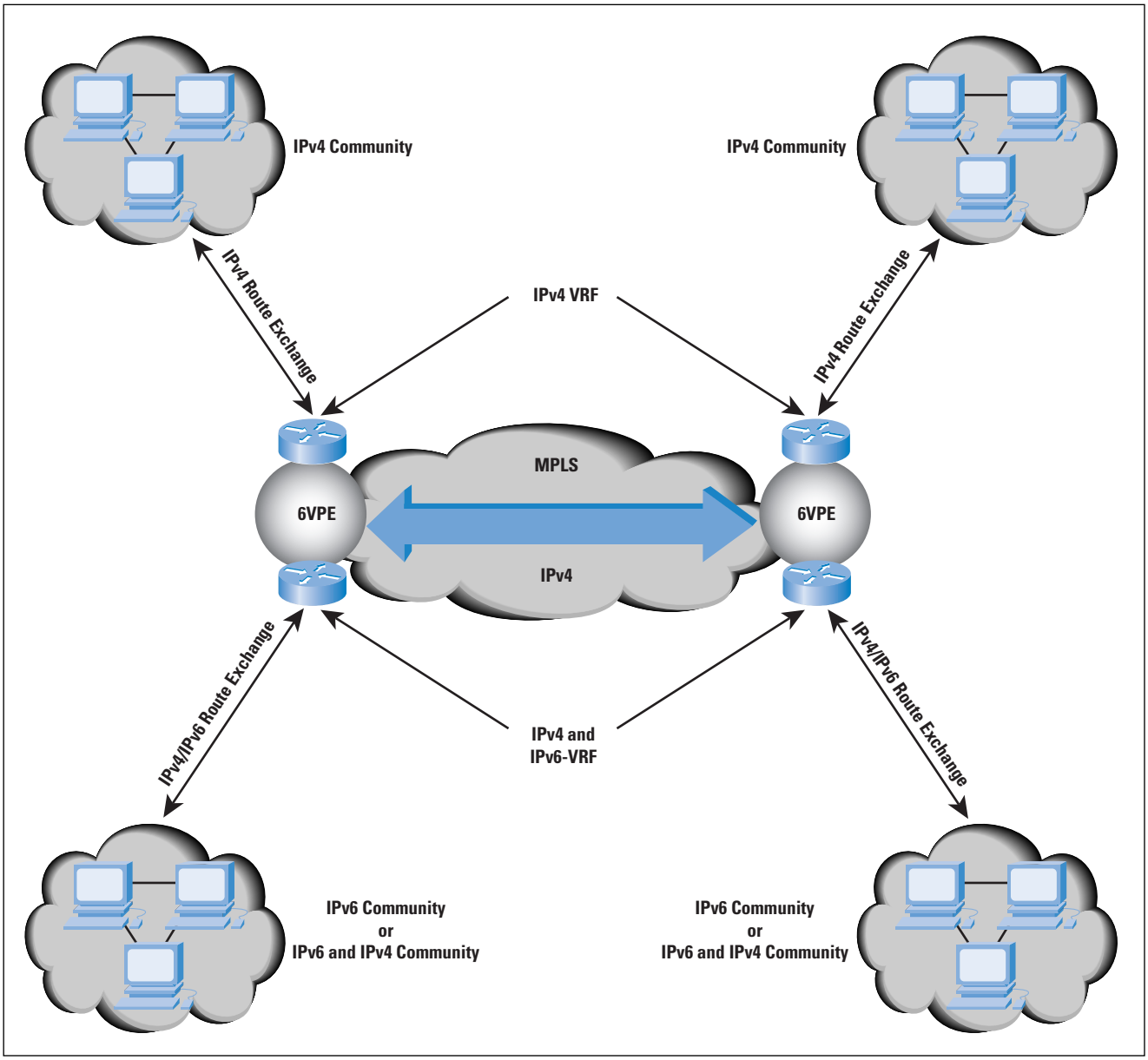
The newest feature to facilitate the RFC 2547bis-like VPN model for IPv6 networks is called 6VPE. It will save service providers from enabling a separate signaling plane, and it takes advantage of operational IPv4 MPLS backbones. Thus there is no need for dual-stacking within the MPLS core. This represents a huge cost savings from the operating expenses perspective and addresses the security limitations of the 6PE approach. 6VPE is more like a regular IPv4 MPLS-VPN provider edge, with an addition of IPv6 support within *Virtual Routing and Forwarding* (VRF). It provides logically separate routing table entries for VPN member devices. This article reviews this approach in more detail because it is the likely approach to succeed in the service provider network.

Under the Hood of 6VPE

Before we look into the 6VPE, it is important to clarify the definition of "dual stack," a technique that allows IPv4 and IPv6 to coexist on the same interfaces. Today, IPv4 has roots in most of the hosts that run applications. Moreover, stability as well as reliability of new applications over IPv6 is maturing. Therefore, coexistence of IPv4 and IPv6 is a requirement for initial deployment. With regard to supporting IPv6 on a MPLS network, two important aspects of the network should be examined:

- *Core:* The 6VPE technique allows carrying IPv6 in a VPN fashion over a non-IPv6-aware MPLS core. It also allows IPv4 or IPv6 communities to communicate with each other over an IPv4 MPLS backbone without modifying the core infrastructure. By avoiding dual-stacking on the core routers, the resources can be dedicated to their primary function to avoid any complexity on the operational side. The transition and integration with respect to the current state of networks is also transparent.
- *Access:* In order to support native IPv6, the access that connects to IPv4/IPv6 domains need to be IPv6-aware. Service provider edge elements (provider edge routers) can exchange routing information with end users. Hence dual stacking is a mandatory requirement on the access layer as shown in Figure 1.

Figure 1: 6VPE Overview



The IPv6 VPN solution defined in this article offers many benefits. Especially where a coexistence of IPv4 and IPv6 is concerned, the same MPLS infrastructure can be used without putting additional stress on the provider router. Also the same set of *Multiprotocol Border Gateway Protocol* (MPBGP) peering relationships can be used. Because it is independent of whether the core runs IPv4 or IPv6, the IPv6 VPN service supported before and after a migration of the core to IPv6 can be done independent of the customer VPN.

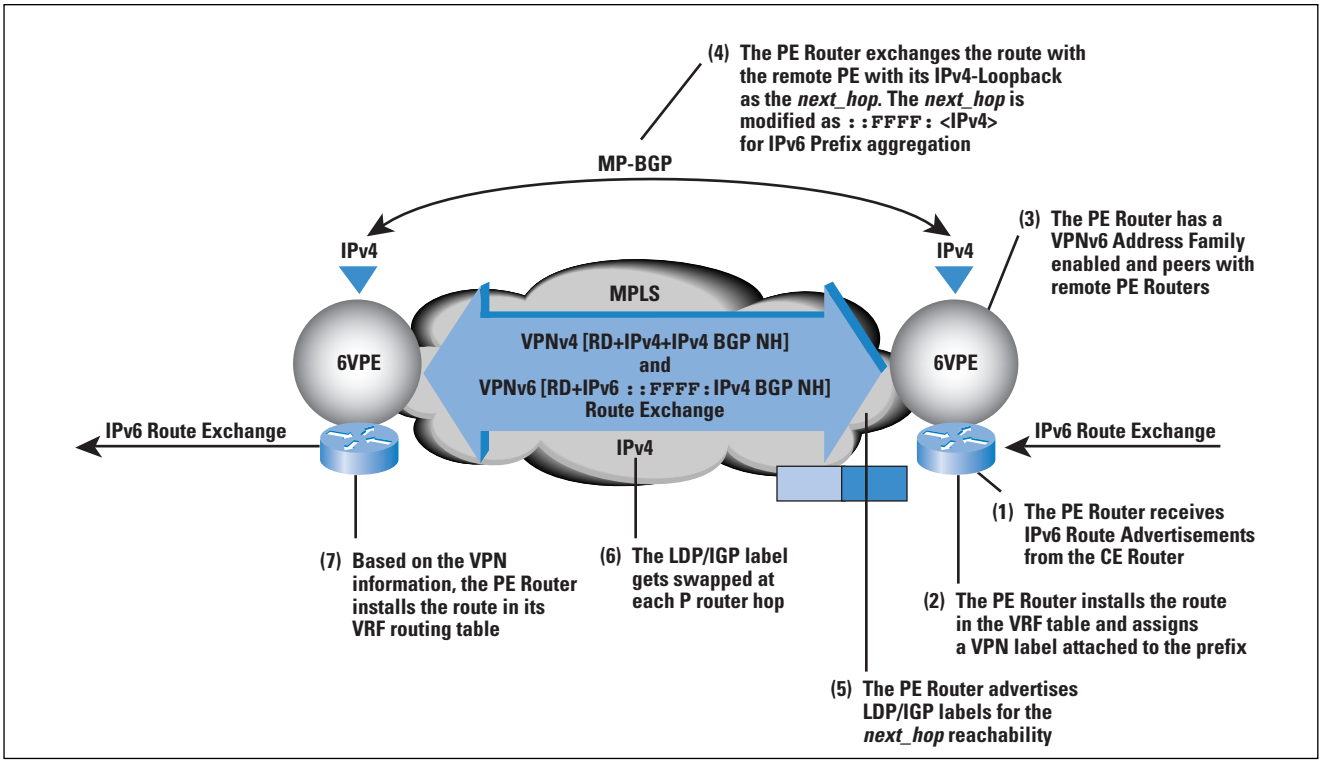
Within the MPLS core, the backbone *Interior Gateway Protocol* (IGP) (*Intermediate System-to-Intermediate System* [IS-IS] or *Open Shortest Path First* [OSPF]) populates the global routing table (v4) with all provider edge and provider routes. As outlined in the draft for IPv4 MPLS VPN (2547-bis), 6VPE routers maintain separate routing tables for logical separation. This allows the VPN to be private over a public infrastructure.

The VRF table associated with one or more directly connected sites (customer edge devices) form close IPv6 or IPv4 speaking communities. The VRFs are associated to physical or logical interfaces. Interfaces can share the same VRF if the connected sites share the same routing information. MPLS nodes forward packets based on the top label. IPv6 packets and IPv4 packets share the same common set of forwarding characteristics or attributes, also known as *Forwarding Equivalence Class* (FEC) within the MPLS core.

6VPE Operation

When IPv6 is enabled on the sub-interface that is participating in a VPN, it becomes an IPv6 VPN. The customer edge-provider edge link is running IPv6 or IPv4 natively. The addition of IPv6 on a provider edge router turns the provider edge into 6VPE, thereby enabling service providers to support IPv6 over the MPLS network.

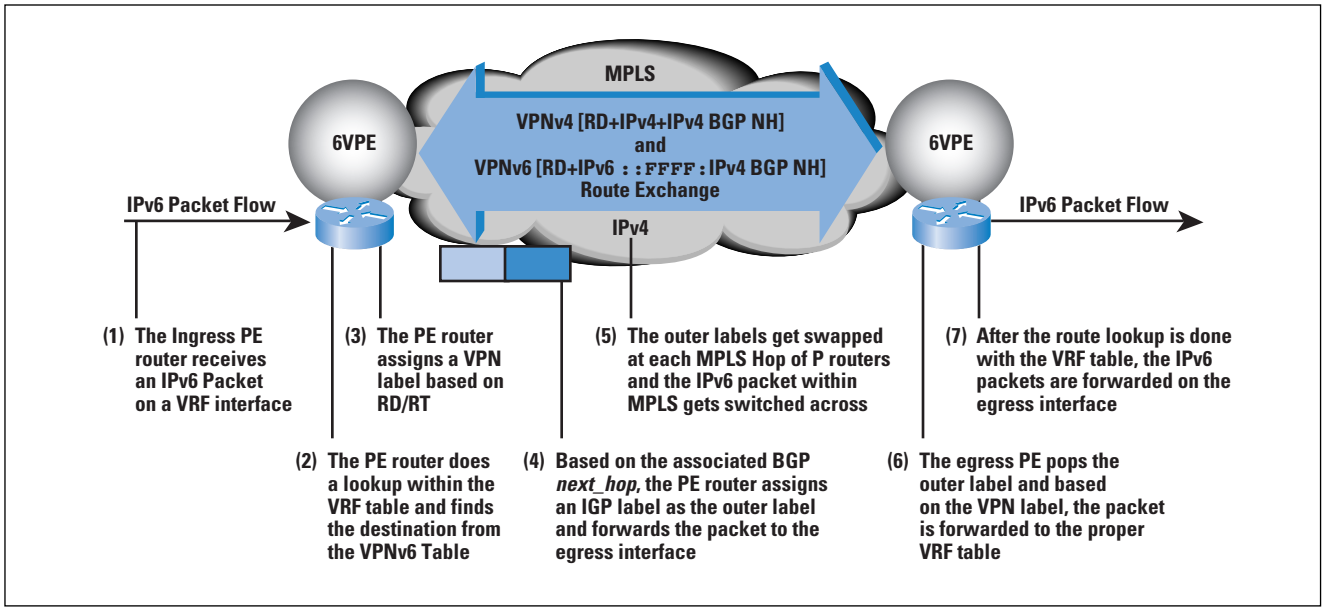
Figure 2: 6VPE Route Advertisement



As outlined in Figure 2, provider edge routers use VRF tables to maintain the segregated reachability and forwarding information of each IPv6 VPN. MPBGP with its IPv6 extensions distributes the routes from 6VPE to other 6VPEs through a direct *internal BGP* (iBGP) session or through VPNv6 route reflectors. The next hop of the advertising provider edge router still remains the IPv4 address (normally it is a loopback interface), but with the addition of IPv6, a value of **::FFFF:** gets prepended to the IPv4 *next_hop*. The technique can be best described as automatic tunneling of the IPv6 packets through the IPv4 backbone. The MP-BGP relationships remain the same as they are for VPNv4 traffic, with an additional capability of VPNv6. Where both IPv4 and IPv6 are supported, the same set of MPBGP peering relationships is used.

MPBGP is enhanced to carry IPv6 in a VPN fashion known as VPNv6, which uses a new VPNv6 address family. The VPNv6 address family consists of 8 bytes—a *Route Distinguisher* followed by a 16-byte IPv6 prefix. This combination forms a unique VPNv6 identifier of 24 bytes. The Route Distinguisher value has a local significance on the router, and the *Route Target* advertises the membership of the VPN to other provider edge routers.

Figure 3: 6VPE Packet Forwarding



In Figure 3, packet forwarding is explained showing end-to-end operation. When the ingress 6VPE router receives an IPv6 packet, destination lookup is done in the VRF table. This destination prefix is either local to the 6VPE (which is another interface participating in the VPN) or a remote ingress 6VPE router. For the prefix learned through the remote 6VPE router, the ingress router does a lookup in the VPNv6 forwarding table. The VPN-IPv6 route has an associated MPLS label and an associated BGP *next_hop* label. This MPLS label is imposed on the IPv6 packet. The ingress 6VPE router performs a PUSH action, which is a top label bind by the *Label Distribution Protocol* (LDP)/IGPv4 to the IPv4 address of the BGP *next_hop* to reach the egress 6VPE router through the MPLS cloud. This topmost-imposed label corresponds to the *Label Switched Path* (LSP). So, the bottom label is bound to the IPv6 VPN prefix through BGP and the top label is bound by the LDP/IGP. The IPv6 packet, now with two labels, gets label-switched through the IPv4/MPLS core router (provider routers) using the top label only (referred to as the *IGP label*). Because only the top label is of significance to the provider core, it is unaware of the IPv6 information in the bottom label.

The egress provider edge router, receives the labeled IPv6 VPN packet and performs a lookup on the second label, a process that uniquely identifies the target VRF and the egress interface. A further Layer 3 lookup is performed in the target VRF, and the IPv6 packet is sent toward the proper customer edge router in IPv6 domain.

In summary, from the control plane perspective the prefixes are signaled across the backbone in the same way as for regular MPLS/VPN prefix advertisements. The top label represents the IGP information that remains the same as for IPv4 MPLS. The bottom label represents the VPN information that the packet belongs to. As described earlier, additionally the MPBGP *next_hop* is updated to make it IPv6-compliant. The forwarding or data plane function remains the same as it is deployed for the IPv4 MPLS VPN. The packet forwarding of IPv4 on the current MPLS VPN remains intact.

6VPE Design Recommendations and Considerations

The following sections identify general recommendations that should be considered when deploying IPv6 in a service provider network:

Working with Enterprise Implementations

Typically *Customer Metropolitan-Area Networks* (C-MANs), also known as *Campus Networks* or *Customer LAN* (C-LAN) elements, form the enterprise network, whereas the 6VPE and customer edge provide the entry point into network access. IPv6 can be supported partially or fully on an enterprise network. In situations where enterprise-wide IPv6 deployment does not exist, network administrators can elect to tunnel the IPv6 traffic toward the provider's customer edge or 6VPE. This can be done with 6-to-4 tunneling methods currently[7]. So, if a site router within a C-MAN or C-LAN aggregates all IPv6 traffic and tunnels to a provider-managed customer edge or 6VPE router, then integration as well as migration becomes smooth. Therefore, it is important for the vendor and the customer to work together in determining the best approach.

Dual VRF Membership per Interface

RFC 2547 for IPv4 recommends one VRF per interface. When running dual stack on a 6VPE, multiple VRF configurations on a single physical or logical interface are required (IPv4 and IPv6). Each VRF instance configuration on a dual-stacked interface forms IPv4 and IPv6 address families. Each address family within VRF runs a VRF-aware routing protocol—such as static routing (static IPv6 unicast routing for IPv6), BGP (BGP with IPv6 enhancements for IPv6), OSPF (OSPFv3 for IPv6), or *Routing Information Protocol* (RIP) (RIPng for IPv6).

MTU Requirements

One important piece of information within the network elements is the capacity of the interface to transfer the size of datagrams. This is known as the *Maximum Transmission Unit* (MTU). The minimum link MTU for IPv4 packets is 68 bytes, whereas for IPv6 the minimum MTU should be 1280 bytes. While designing and planning for IPv6 support, the network elements should be examined along with interfaces and underlying network technologies to ensure the MTU requirements.

Dealing with Link-Locals

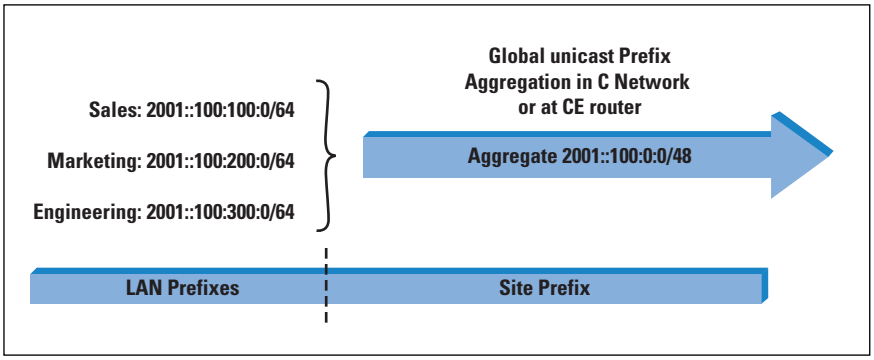
Because link-local scope addresses are defined as uniquely identifying interfaces within a single link, only those may be used on the provider edge-customer edge link.

However, they are not supported for reachability across IPv6 VPN sites and are never advertised with MPBGP to remote provider edges. As outlined in the RFC for IPv6 address assignments, the link locals (**`FE80::x`**) should not be advertised outside their local scope. Because the link-local addresses are embedded on the IPv6-enabled interface for certain local tasks, the link-local addresses are not and should not be advertised anywhere outside the local link scope, including the customer edge and 6VPE running IPv6. Globally unique aggregatable IPv6 prefixes are defined as uniquely identifying interfaces anywhere in the network. These addresses are expected for common use within and across IPv6 VPN sites. They are obviously supported by this IPv6 VPN solution for reachability across IPv6 VPN sites and advertised through MPBGP to remote provider edges.

Router Capacity Impact

Dual-stacking also introduces another task, namely hardware analysis to determine the resource capacity, that is, CPU and memory usage. Increased memory consumption may occur because of the dual-stack *Routing Information Base* (RIB). It also has implications for the *Interface Descriptor Block* (IDB) and *Routing Descriptor Block* (RDB) limits of hardware. The IDB limit is the capacity of particular equipment to support a number of physical and logical interfaces, whereas the RDB limit is the number of routing protocols and instances supported on such equipment. Typically these values (limits) are very high, but 6VPE is such an important element of the MPLS network that these facts must be considered. From a business case perspective, scalability, high aggregation, and rapid Return on Investment are expected, hence it is important to consider these factors in the design.

Figure 4: Route Aggregation



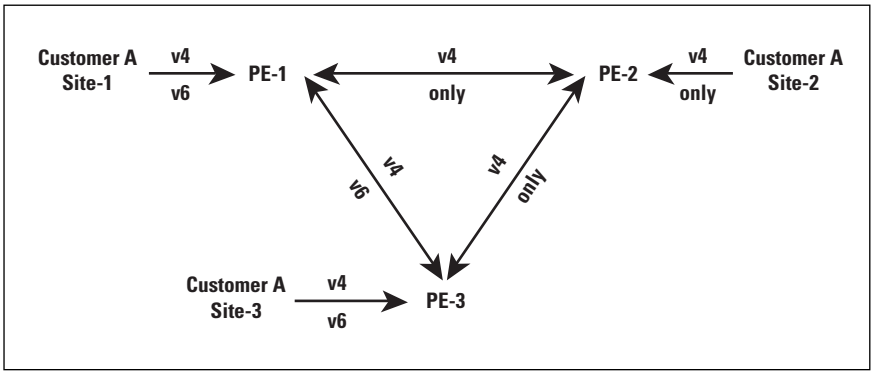
Router Memory Impact

The memory challenges can occur also when large numbers of IPv6 prefixes are advertising toward service provider network elements. In that event, the enterprise on the C-LAN or service provider on the customer edge router may elect to perform route aggregation. IPv6 prefixes can be aggregated to their higher-level significant boundary. Figure 4 shows an example of IPv6 prefix aggregation. Moreover, when a packet arrives on a dual-stacked interface (VRF-aware interface), the 6VPE router determines the packet version number by looking into the IP header. The per-packet header lookup is normally performed (it is a basic router function), but the extra work required by the router is to determine the version number. This additional task creates a longer processing cycle.

The Address Family Identifier and its Importance

All the elements referenced as dual-stacked, such as provider edge and customer edge routers, run IPv4 as well as IPv6 addressing and routing protocols. The 6VPE elements can also mix and match VPNv4 and VPNv6 peering sessions with other 6VPE routers or with route reflectors. What does the term "mix and match" mean here? It was an important enhancement to traditional BGP when MPBGP extensions were introduced. The address family within MPBGP is modular to facilitate distinct peering relationships, and is expressed using the *Address Family Identifier* (AFI). The regular BGP capabilities are exchanged after the peering sessions are turned on. In order for two provider edge routers to exchange labeled IPv6 VPN prefixes, they must use BGP capabilities negotiation to ensure that they both are capable of processing such information. When the service provider network is running VPNv4 peering sessions with other respective elements in the network, it exchanges the VPNv4 AFI capabilities with others. When the VPNv6 peering sessions are turned on, it renegotiates the capabilities and fresh peering sessions are established. The peering sessions established are based on common features if either of the peers does not agree on any of the capabilities.

Figure 5: VPNv4 and VPNv6 AFI



In Figure 5, three provider edge routers out of two need to exchange VPNv6 traffic, but all three provider edge routers need to maintain their existing VPNv4 capabilities. This is possible with the AFI configuration feature, which makes the migration steps very smooth. Service providers can mix and match VPNv4 and VPNv6 provider edge routers as required. Functions of 6VPE can be turned on when and where required. If the customer edge routers are dual-homed to different provider edge routers, the integration of customer IPv4 and IPv6 networks becomes painless. This scenario outlines hybrid environments, but it does not address the IPv4 and IPv6 communication. Consider techniques such as *Network Address Translation* (NAT) or application layer gateways for the IPv4 and IPv6 communication.

Route Reflectors for MP-IBGP

For advertising VPN membership, provider edge routers peer with VPNv4 route reflectors for scalability, thereby avoiding the need for full-mesh MP iBGP sessions among all provider edge routers. The same concept is supported for VPNv6. The same VPNv4 route reflectors can be upgraded to support VPNv6 address families.

Route reflectors can also make addition or removal of a provider edge router from a network simple and flexible. Alternatively, the BGP confederation option can also be deployed to provide MPBGP peering sessions among provider edge routers.

QoS Considerations

Service providers operating customers' MPLS VPN networks and also providing *Quality of Service* (QoS) should account for the new introduction of IPv6 and its impact. QoS and queuing of important application traffic requires distinct policies for IPv4 and IPv6, in turn possibly requiring additional operational tasks where IPv4 and IPv6 networks coexist. Other design considerations should be made to account for each individual network. Both IPv4 and IPv6 have a commonality, which is the 3-bit IP *Precedence* (or *Type-of-Service* [ToS]) field within the IP headers. Alternatively, the *Differentiated Services* (Diff-Serv)-compliant QoS models can also be employed. Irrespective of the technique, QoS is an important factor when low-speed links are concerned. However, there is no additional advantage of QoS on IPv6 versus IPv4. At some point in the future IPv6 can be different by using the flow label in the IPv6 header. QoS within the MPLS core remains *MPLS Experimental Value* (MPLS_EXP)-based and is untouched but still is effective with the addition of IPv6.

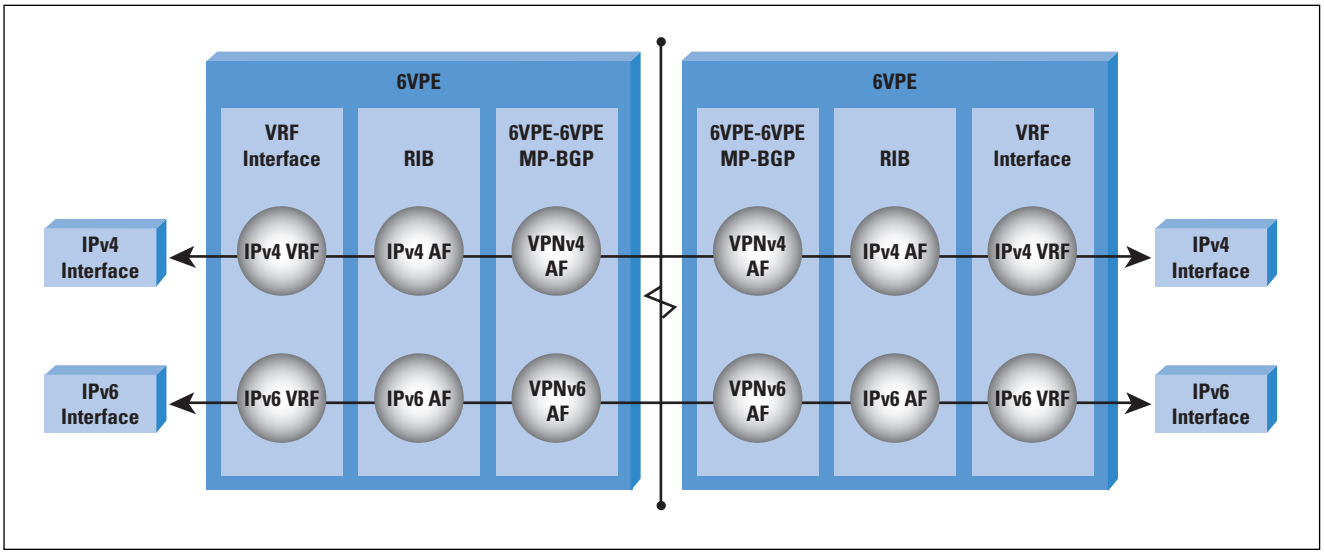
Device Management

Finally, device management is another important aspect that service providers must consider. Device management in a dual-stacked network can be done through an IPv4 or IPv6 address. Where the IPv6 VPN service is supported over an IPv4 backbone, and where the service provider manages the customer edge, the service provider can elect to use IPv6 for communication between the management tool and the customer edge for such management purposes. The management systems, including *Operations-Support-System* (OSS) servers, need to be aware of IPv6 and must run proper *Simple Network Management Protocol* (SNMP) stacks in order to perform IPv6-based management. From the VPN perspective it still remains transparent how the device and services are managed.

Enhancements to the Draft

The current MPLS VPN services that service providers have implemented are based on RFC 2547bis, the Internet Draft required to enhance the Layer 3 VPN approach further to address the IPv6 support. The "BGP-MPLS VPN extension for IPv6 VPN"[1] is the current Draft that addresses the need for IPv6 support over MPLS networks in a VPN environment. Also, to avoid an extra layer of signaling, the Draft addresses the scalable automatic tunneling of VPN-based IPv6 prefixes. The basic functions remain the same as outlined in RFC 2547. Some of the extensions outlined will require additional work in order to be effective in the service provider network.

Figure 6: Dual Mode 6VPE AFI Model



The standard RFC 2547bis introduces "address family" concepts, as well as MPBGP to carry VPN information across the MPLS network. This enables formation of a full mesh between customer sites. The provider edge routers advertise their VPN membership to other provider edge routers through direct iBGP or value(s). As shown in Figure 6, these new address families are introduced to support IPv6 within VPN, IPv6, and VPNv6. If configured for dual stacking, the interface belongs to multiple VRF instances, IPv4 and IPv6. Each instance maintains its own RIB. MPBGP is now capable of handling the VPNv6 address family to advertise the IPv6 prefix across the VPN.

Summary

"Staying abreast of the best" has always been challenging for service providers when it comes to technology deployment or support. Time to market is another challenge. This article provides a view of the service provider challenges. In this new era where explosive use of IPv6 is envisioned, it is extremely important for service providers to have a simplified, automated, fail-proof, and cost-effective network design. The Internet Draft discussed advances the capabilities to achieve this and allows service providers to take a practical approach in supporting IPv6 for customers' next-generation applications. The Draft brings service providers closer to the IPv4-to-IPv6 transition with a simple, cleaner, cheaper, and scalable solution.

For Further Reading

[1] Jeremy De Clercq, Dirk Ooms, Marco Carugi, Francois Le Faucheur, "BGP-MPLS VPN extension for IPv6 VPN," **`draft-ietf-l3vpn-bgp-ipv6.06.txt`**, February 2005.

[2] Eric Rosen and Yakov Rekhter, "BGP/MPLS VPNs," **`draft-ietf-l3vpn-rfc2547bis-03.txt`**, October 2004. (See also RFC 2547, March 1999, by the same authors.)

[3] Mallik Tatipamula, Patrick Grossetete and Hiroshi Esaki, "IPv6 Integration and Coexistence Strategies for Next-Generation Networks," *IEEE Communications Magazine,* Vol. 42, No. 1, January 2004.

[4] Bates, Chandra, Katz, and Rekhter, "Multiprotocol Extensions for BGP4," RFC 2858, June 2000.

[5] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," RFC 2460, December 1995.

[6] Rekhter and Rosen, "Carrying Label Information in BGP4," RFC 3107, May 2001.

[7] Carpenter, B. E., Moore, K., Fink, R., "Connecting IPv6 Routing Domains Over the IPv4 Internet," *The Internet Protocol Journal,* Volume 3, No. 1, March 2000.

TEJAS SUTHAR holds CCIE # 8423. He is working as a Service Architect at TELUS Communications Inc. in Toronto. He focuses on Converged Network designs for customers in various industry sectors. He is very active in IP-related deployments. E-mail: **`tejas.suthar@gmail.com`**

Graph Overlays on Path Vector: A Possible Next Step in BGP

by Russ White, Cisco Systems

Over the past several years, much research and thought has gone into a replacement for the current interdomain routing protocol, *Border Gateway Protocol* (BGP)[1]. For instance:

- In 2002, the *Internet Research Task Force* (IRTF) published a set of requirements for a next-generation interdomain routing protocol. In fact, several sets of requirements documents have been published in this area.
- In December 2001, *The Cook Report* noted that BGP needs to be replaced[2]:
- In October 2003, the *Workshop on Internet Routing Evolution and Design* (WIRED) presented papers arguing that BGP needs to be replaced[3].
- In December 2001, the IETF published RFC 3221[5], authored by Geoff Huston, which provided some background information toward finding a replacement for BGP.

There are probably thousands of references in magazine articles, conference proceedings, and research papers, all stating that BGP should be replaced. Of course, all these discussions wind up at the same place: It is almost impossible to replace BGP, wholesale, in the public Internet, or even in any of the private networks running BGP today.

The basic problem is you cannot take the network down, and you cannot replace the routing protocol without taking the network down. Many very clever ideas have been proposed to get around this problem—complex transition schemes, moving partitions, and all sorts of other concepts. But, in the end, the idea of transitioning from one routing protocol to another on something as large—and as distributed in both geography and ownership—as the Internet, has been a hard wall against which all the proposals for new interdomain routing protocols pile up. In an article[4] here in *The Internet Protocol Journal,* Geoff Huston states:

> "Another approach is to consider the feasibility of decoupling the requirements of inter-domain connectivity management with the applications of policy constraints and the issues of sender- and receiver-managed traffic-engineering requirements. Such an approach may use a link-state protocol as a means of maintaining a consistent view of the topology of inter-domain network, and then use some form of overlay protocol to negotiate policy requirements of each Autonomous System, and use a further overlay to support inter-domain traffic-engineering requirements."

In this article, we propose building on this concept, but in a novel way: rather than replacing BGP, or attempting to solve all the currently perceived problems with BGP at once, we attempt to address two problems in a way that does not heavily modify day-to-day BGP operation. Rather than replace BGP, enhance it to account for new re-quirements by providing new capabilities. If done right, this avoids the problem of deploying a new routing protocol altogether, because BGP is already deployed throughout the Internet.

Problems with BGP

No discussion of replacing BGP would be complete without a discussion of why so many people think BGP needs to be replaced. We need to consider three main points in this area: *convergence speed, policy,* and *security*. Each of these is covered in the following sections.

BGP Convergence Speed

Through various studies, and through examining the way in which BGP works, it has been shown that BGP, in an interdomain environment, always converges roughly in:

(Maximum AS_PATH – Minimum AS_PATH) × Minimum Advertisement Interval

To understand why this is so, let's examine the following small internetwork as it converges.

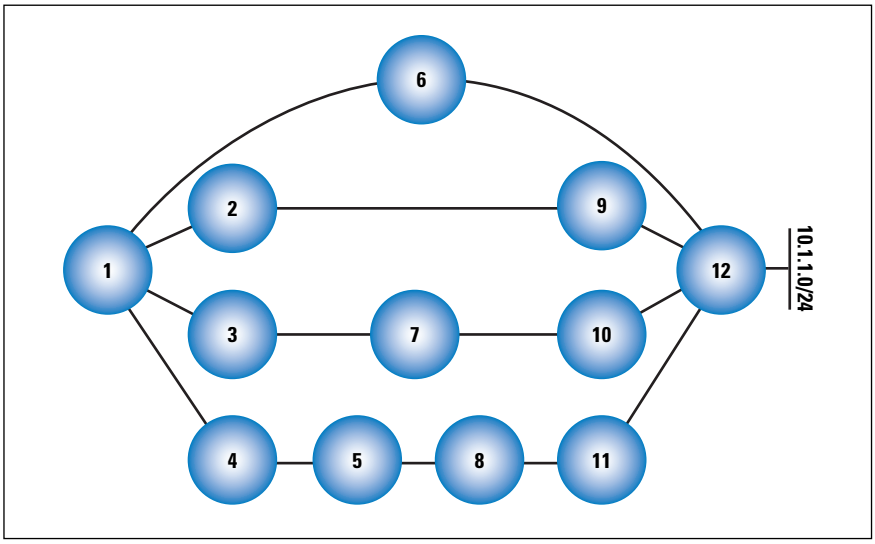


Figure 1: An Example Internetwork Using a Path Vector Protocol

Let's assume autonomous system (AS) 12 is advertising some destination, 10.1.1.0/24, and that every other autonomous system in the internetwork chooses the path to the right to reach that destination. So, for instance, AS4 chooses the path {5,8,11,12} to reach 10.1.1.0/24, AS3 chooses the path {7,10,12} to reach 10.1.1.0/24, AS2 chooses the path {9,12} to reach 10.1.1.0/24, and AS6 chooses the path {12} to reach 10.1.1.0/24.

At this point, let's examine what happens if AS12 loses its connection to 10.1.1.0/24. AS12 sends out a withdraw, which reaches AS6, 9, 10, and 11 at about the same time. These autonomous systems then send out withdraws, with the second set of withdraws reaching AS1, 7, and 8 at about the same time.

When AS1 receives this first withdraw, it examines its local table, and finds the next best path to reach 10.1.1.0/24 is through AS2, with the path {2,9,12}. AS1 does not realize that AS2 has received a withdraw for 10.1.1.0/24 at the same time it received the first withdraw for this destination from AS6. So, AS1 switches over to its next best path, and continues forwarding traffic to 10.1.1.0/24.

AS2, 7, and 8 now also send withdraws to each of their peers, including AS1, 3, and 5. AS1 now receives another withdraw, again for the path it is currently using to reach 10.1.1.0/24. AS1 examines its local tables and finds it has another path, through {3,7,10}, to 10.1.1.0/24, so it switches to that path, without knowing AS3 has just received a withdraw for this same path. AS3 and 5 now send withdraws to each of their peers, AS1 and 4. AS1 has again received a withdraw from the peer it is using to reach 10.1.1.0/24, so it examines its local tables, and finds it still has a path through {4,5,8,11,12} to reach this destination. It switches to this path, without realizing AS4 has just received a withdraw as well.

AS4 now sends the final withdraw to AS1, removing AS1's final path from its local tables. AS1 now removes all reachability information for 10.1.1.0/24, and the network is converged. Note that the actual convergence in this situation would be a bit more complicated, with AS1 sending updates at each stage, and all the other autonomous systems reconverging at each step along the way, but we have used only the simplest set of messages through the network, to illustrate the basic procedure BGP follows when converging.

This short example illustrates why BGP has the convergence characteristics described previously. BGP "hunts" through each possible autonomous-system path, from shorter ones to longer ones, until it finally converges. The rate at which it can hunt through each possible autonomous-system path is determined by the minimum advertisement interval, the rate at which new routing information is allowed to flow through the system.

This problem has several obvious solutions. The first is to simply increase the rate at which routing information flows through the system, by reducing the minimum advertisement interval. But, this plays against route flap dampening, and network stability in general, so, beyond some lower possible bound, reducing the minimum advertisement interval is not possible (without further modifications to BGP).

Another obvious solution is to simply add a "reason code" to the original withdraw. If AS12 originally stated it was withdrawing reachability to 10.1.1.0/24 because it had lost local connectivity to it, then all paths with AS12 in the path could have been discarded immediately, at the first step. The problem here is making certain the original withdraw message actually makes it through the network, from AS12 all the way to AS1. Because BGP is a very efficient protocol, many control messages of this type are actually removed from the network, through implicit withdraws, aggregation, and other mechanisms.

Policy

The second problem we encounter with BGP is its rather rough sense of policy. For instance, let's examine the following small network, and look at one specific example of where policy transmission and enforcement are problematic in BGP.

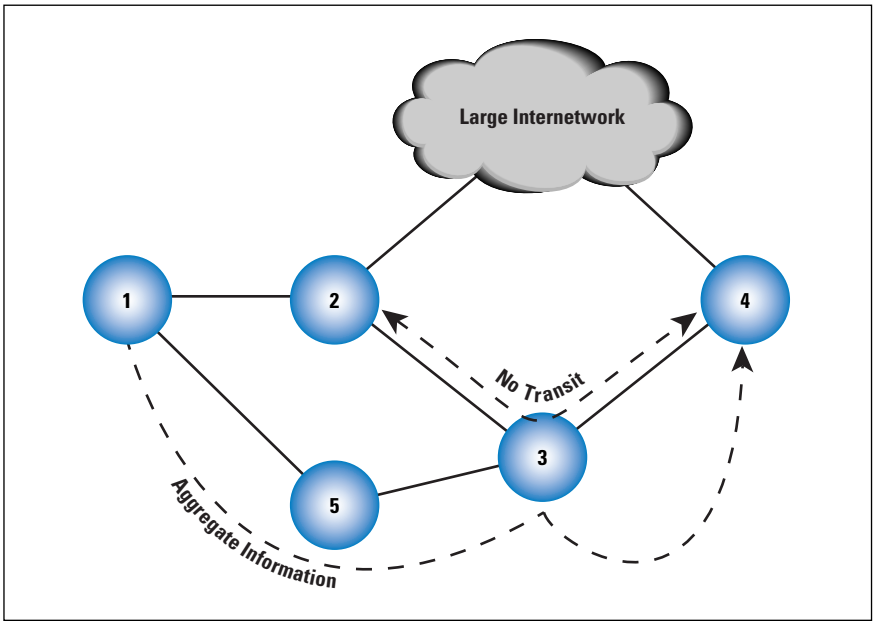


Figure 2: Issues with Policy Transmission in a Path Vector Protocol

Here AS2 has a policy that AS3 should never be used for transit. In other words, traffic originated in AS4 should always pass through the large internetwork rather than through AS3 to reach AS1. This type of situation is very common in the public Internet, such as when AS3 is actually AS2's customer. How can AS2 communicate this policy to AS4, however?

AS2 could simply mark the routing information it sends to AS3 so AS3 cannot readvertise it to AS4, but this is problematic. Simple mechanisms, such as marking the routes with the NO_EXPORT community, are easy for AS3 to simply strip off the routing information it receives. We could conceive of some way to cryptographically sign the included policy, so AS3 cannot disturb the policy and AS4 can see the policy when it receives the information from AS3, but this is problematic as well.

Suppose AS3 is receiving aggregated routing information directly from AS5, which includes some of the same destinations AS2 has advertised to AS3, but has blocked AS3 from advertising to AS4. AS3 could, conceivably, readvertise this routing information to AS4, and AS4 could prefer this shorter prefix aggregate to reach the destinations in AS1, rather than the paths through the large internetwork. AS4 would then forward traffic to AS3, which would then rely on its longer prefix routes, received from AS2, to forward this traffic to these destinations in AS1. AS3 is, contrary to AS2's policy, transiting traffic through AS2 to AS1. There is no simple answer to this problem.

Security

It has been widely acknowledged that BGP is an insecure protocol, with many areas where attackers can hijack, inject false routing information, and perform other attacks. The IETF's *Routing Protocols Security* (RPsec) working group is working on a set of documents describing vulnerabilities of BGP, and creating recommendations for systems to secure BGP. For the latest information about these Drafts, refer to the RPsec homepage at: **`http://www.rpsec.org`**

What sort of requirements are likely to come out of such an undertaking?

- Any proposed mechanism must be able to show that a specific autonomous system is authorized to originate specific routing information.
- Any proposed mechanism must be able to show that the AS Path carried in received routing information corresponds to a real path in the internetwork, beginning with the origin AS and ending in the advertising peer.

There will be many other requirements that proposed mechanisms for providing security for BGP will need to, or should, meet, but these two will be the largest areas of concern for our purposes.

Solving the Problems

Now that we have an idea of the three areas we want to solve problems in, how can we actually solve them? The most elegant solution would be a single mechanism that does not change the current semantics of BGP itself too greatly, would provide greater benefits as it is deployed throughout a large-scale internetwork, and would rely on existing—and understood—techniques within routing.

One perfect example of such a mechanism would be to simply overlay a link state-like graph of interconnectivity over the BGP protocol. This graph would provide information about the interconnections between autonomous systems, rather than between routers, and would be used to convey information about the topology and policies in the internetwork, rather than to find loop-free paths through the internetwork.

Let's go back through our three examples, and see how overlaying an internetwork connection graph would be able to solve some of the problems currently facing BGP.

Convergence Speed

Looking at our small sample internetwork again:

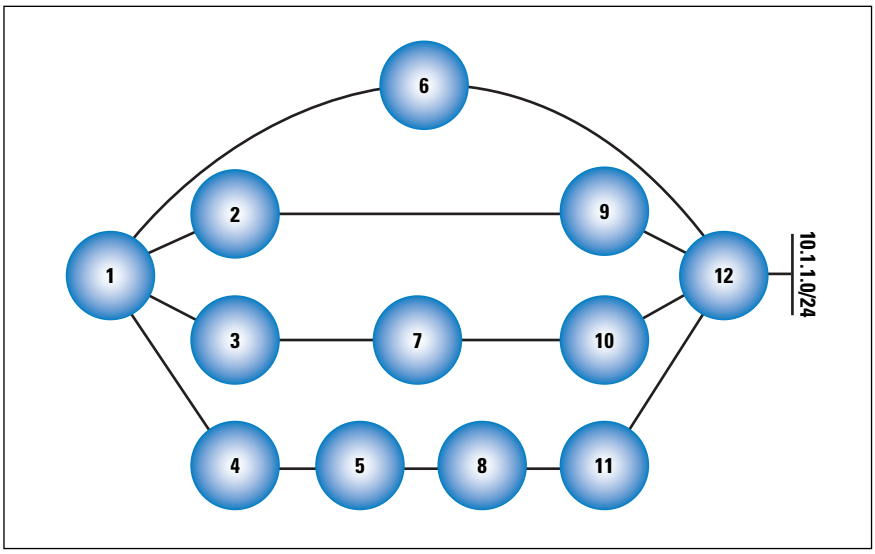


Figure 3: An Example Internetwork Using a Path Vector Protocol

What is the one thing we said would resolve the problems with BGP hunting through every possible longer autonomous-system path alternative to finally converge around loss of reachability to 10.1.1.0/24? Could AS12, somehow, communicate directly to every autonomous system in the internetwork that it has directly lost this connection, rather than waiting for AS1 to try every possible path to 10.1.1.0/24, and discover each one, in turn, withdrawn?

If we had a topological graph of the network, AS12 could simply remove 10.1.1.0/24 from its connectivity information. AS12 would then flood this information, on an interdomain basis, to all the other autonomous systems in the internetwork at roughly the same time. Thus, in the worst case, AS1 would receive this information at about the same time it received the first withdraw for 10.1.1.0/24, from AS6.

When AS1 receives this updated topology information from AS6, it will discover that AS12 is no longer connected to 10.1.1.0/24, and, therefore, it can remove every possible path to 10.1.1.0/24 containing AS12. This would allow AS1 to remove the paths {2,9,12}, {3,7,10,12}, and {4,5,8,11,12} at the same time. The internetwork now converges as soon as AS1 computes the new connectivity graph, and acts on it by examining each entry in its local tables and discarding the ones with AS12 in the autonomous-system path.

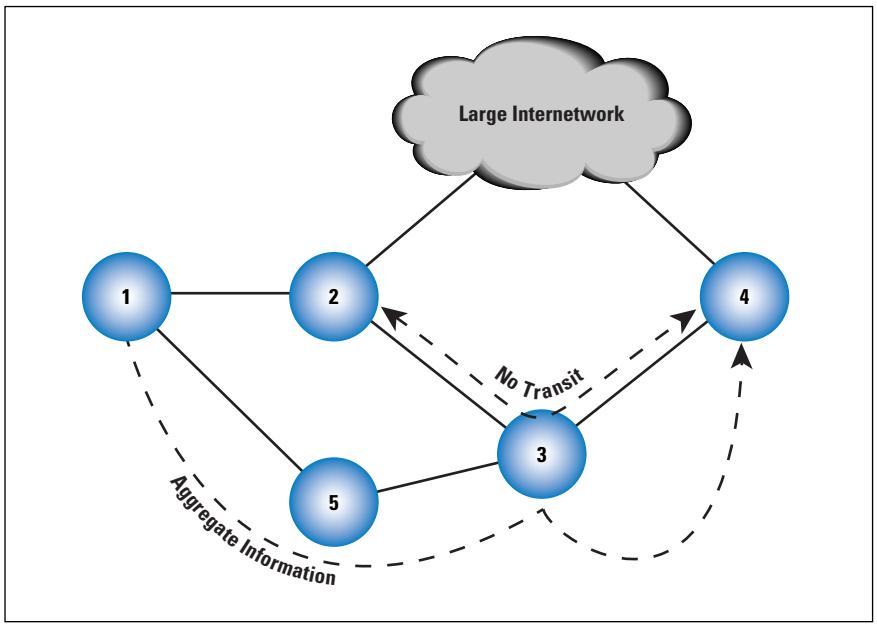
We have not changed the way BGP finds paths through the network—the path still is not valid unless we receive an advertisement from our connected peers. We have also not changed the format of any BGP updates, any peering state machines, or anything else. We have simply overlayed an interconnection graph on top of the current protocol mechanisms, which we can use to our advantage to speed up network convergence.

What about partial deployments in this situation? Suppose only autonomous systems 6, 7, 8, 9, 10, 11, and 12 are running this new extension. Would it still help us to speed up network convergence? When AS12 withdraws 10.1.1.0/24, AS6, 7, 8, 9, 10, and 11 would immediately discard any routes passing through AS12 to reach 10.1.1.0/24. At this point, they could each withdraw those routes, meaning AS1, 2, 3, and 5 would all receive a withdraw at about the same time. This short-circuits the number of possible paths for AS1 to hunt through, decreasing the amount of time the internetwork takes to converge. Even without a full deployment, we see some positive impact from this new technique.

Policy

Let's examine our policy problem after placing our interconnection graph on top of the internetwork.

Figure 4: Injecting Policy on an Interconnection Graph



Here, we see that AS2 could actually place its policy for AS3 not to transit traffic in the interconnection draft. AS4 would then be able to independently verify what AS2's policy toward AS3 transiting traffic is. AS4 could then examine the routing information it receives from AS3, and determine if it should install—or not install—routing information received from AS3, based on this policy.

Objections to an Interconnection Graph

When a link-state protocol has been proposed as a possible replacement for BGP in the past, two primary objections have been raised:

- Providers are reluctant to accept the wholesale replacement of a known working system with a new one.
- Many providers wish to hide their policies and connectivity to other providers or customers for policy reasons.

This article does not propose replacing BGP, just augmenting it, so the first argument is, to some degree, not valid against this approach. The second objection, that of using a link-state protocol for interdomain routing specifically, also does not apply, because we are not proposing changing the way BGP finds loop-free paths through the network. The proposed interconnection graph is not used for finding paths through the network, it is used only for faster signaling of path failure (by short-circuiting the slower withdraw mechanisms), and for providing a place to hang policy and security information.

Concentrating on a few smaller spaces allows us to design a smaller solution set that can be incrementally deployed in a simple way.

The second objection is harder to meet, simply because the concepts of policy within a routing system are hard to define and understand in all possible cases or respects. In fact, there are policy requirements not met by BGP today, but rather are met through contracts, packet filters, and other mechanisms (even sometimes by violating the BGP specification).

Consider two facts about this proposal that work around many of the specific objections we have heard in this area:

- The interconnection graph can be partial, in different parts of the internetwork. For instance, a given service provider might provide different views of who they are connected to to different peers, depending on their policy of revealing this information.
- The interconnection graph only contains autonomous system-level connectivity information, not specific peering-point information. For instance, two autonomous systems may be connected in a large number of places, or as few as one. The interconnectivity graph does not care about such details, only whether at least one connection exists. Such an interconnectivity graph would not reveal actual connection points between peering autonomous systems, how rich that connectivity is, nor any other information about the business relationship between the two peers.

  In fact, the types of interconnectivity information an interconnection graph could provide is already available by examining the autonomous-system paths of routes retrievable from various route view servers. Some mechanism would be required to collate this information into a usable graph, but a good deal of current research on the scaling and convergence properties of large-scale internetworks actually depends on the ability to build an interconnection graph before beginning any other work, so mechanisms to collate this data already exist, and are in use today.

Security

The internetwork interconnection graph can actually show whether a path exists from the origin to the advertising peer, through *signed certificates*. For example, soBGP[6] (**`ftp://ftp-eng.cisco.com/sobgp/index.html`**) uses this specific mechanism to validate the autonomous-system path carried in received routing information. Other research is currently being pursued in this area as well.

Summary

We have proposed a single step forward that could be used to resolve some of the problems facing BGP in the near term, and possibly provide the networking community with a path forward on other fronts as well. The concept of simply making incrementally deployable changes to BGP to solve pressing problems can provide us with options outside the normal lines of thinking: either making very small changes to BGP, making BGP more and more complicated, or simply replacing the BGP protocol, with all the deployment problems this would entail.

References

[1] Yakov Rekhter, Tony Li, "A Border Gateway Protocol 4 (BGP-4)," RFC 1771, March 1995.

[2] **http://www.cookreport.com/10.09.shtml**

[3] **http://www.net.informatik.tu-muenchen.de/wired/position/bruce.html**

[4] Geoff Huston, "Scaling Inter-Domain Routing—A View Forward," *The Internet Protocol Journal,* Volume 4, No. 4, December 2001.

[5] Geoff Huston, "Commentary on Inter-Domain Routing in the Internet," RFC 3221, December 2001.

[6] Russ White, "Securing BGP Through Secure Origin BGP," *The Internet Protocol Journal,* Volume 6, No. 3, September 2003.

RUSS WHITE works for Cisco Systems in the Routing Protocols Deployment and Architecture (DNA) team in Research Triangle Park, North Carolina. He has worked in the Cisco Technical Assistance Center (TAC) and Escalation Team in the past, has coauthored several books on routing protocols, including *Advanced IP Network Design, IS–IS for IP Networks,* and *Inside Cisco IOS Software Architecture*. He is currently in the process of publishing a book on BGP deployment, and is the co-chair of the Routing Protocols Security Working Group within the IETF. E-mail: **riw@cisco.com**

Book Reviews

A Brief History of the Future

A Brief History of the Future—The Origins of the Internet, by John Naughton, ISBN 0-75381-093X, 2000, Published by Phoenix, **http://www.orionbooks.co.uk**

This is a well-written book by a well-known Irish academic and journalist, which charts the growth of the Internet from a 1950s military project to the pervasive networking infrastructure that dominates the IT world today. It is relevant to the readership of this journal because it charts the growth of the technology that underpins the IP world—and it gives a sound understanding of the culture and approach that led to the development of the Internet as we know it.

Naughton takes the reader from the inception of the *Advanced Research Projects Agency Network* (ARPANET) through most of the major developments such as packet switching, mail, TCP/IP, and the Web, not only covering the technology, but also providing insights into the background of the Internet pioneers and the political environment.

Organization

The book is divided into three major sections, the first of which is largely concerned with scene setting and is aimed at bringing those less familiar with the subject area up to speed. In the first chapter, Naughton likens the evolution of the "Net" to that of amateur radio, moving on in succeeding chapters to cover basic technology and to provide some perception of scale and rate of growth.

The second part of the book covers the growth of the Internet up to the early 1990s. This starts by looking at the origins of the ARPA project, noting the influence of MIT and important figures such as Vannevar Bush, Norbert Weiner, and J.C.R Licklider. Naughton describes how ARPA was initiated and its relationship with NASA and academia, highlighting the desire to provide time-sharing systems and the breakthrough concept of the *Interface Message Processor* (IMP) as a solution to the "n-squared" problem. This is followed by two chapters that discuss the adoption of packet switching as the underlying technology, following its initial proposal by Paul Baran and further development by Donald Davies' team in the UK.

Naughton next examines how e-mail became the first "killer application" that drove up Internet usage, even telling the reader where the use of the ubiquitous "@" symbol comes from. He then considers the maturing network during the 1970s, discussing the formulation of the first *Request For Comments* (RFCs), the development of the gateway concept, and the evolution of TCP/IP. The discussion leaves the network area, concentrating on the evolution of UNIX and its impact, stressing the role of AT&T's regulatory situation. Then Naughton considers how this accelerated the development of USENET.

In a chapter called "The Great Unwashed," Naughton discusses the popularization of computing and networking, through the availability of the PC and the evolution of readily available file transfer tools such as X-Modem and the creation of bulletin board systems such as fidonet. He then considers the development of Open Source, telling the story of Linux and its derivation from MINIX.

The third section of the book deals with the emergence of the World Wide Web, tracing it back through the original ideas of Vannevar Bush and Ted Nelson, to its ultimate development by Berners-Lee at CERN. He links this to the subsequent development of Mosaic at NCSA and shows the dramatic impact this had on Internet growth.

Naughton concludes his book by looking at the prognosis for the "Net." Here he refuses to try to predict the future; instead he analyzes the forces that will drive the future of the Internet and discusses their impact in the past and hence their potential impact. At the end of the book, he provides notes and references for each chapter, a short section on the sources he consulted, and a comprehensive glossary.

Synopsis

I found this book provided excellent insights into the development of the Internet, adding a lot of perspective to the engineering field I currently work in. Naughton places appropriate emphasis on the technical, personal, commercial, and political factors that have steered its evolution. He is not afraid to disturb the reader's preconceptions by looking at things from unusual angles, and he emphasises the importance of *timing*. This is apparent when he points out that according to many sources, most of the important inventions around the Internet have come from graduate students, rather than the professors they work for. He similarly recounts the story that AT&T turned down the opportunity to run the "Net" in the early 1970s and reflects the view that if the Internet had not existed we could not invent it now.

This is an excellent read (it was nominated for the *Aventis Prize* in 2000), which helps the reader understand the How, When, Where, and Why of the Internet's development. It covers most of the major milestones in the evolution of our discipline and is very well-written.

The Author

John Naughton is Professor of Public Understanding of Technology at the Open University, and he writes a weekly column in *The Observer* Business Section, covering important developments and trends in the IT industry. He describes himself as a "Control Engineer with a strong interest in systems analysis and computer networks" and is a Fellow of Wolfson College, Cambridge.

—Edward Smith, BT, UK
edward.a.smith@btinternet.com

Eats, Shoots & Leaves

Eats, Shoots & Leaves, by Lynne Truss, ISBN 1-592-40087-6, Gotham Books, 2003.

Eats, Shoots and Leaves is a book about punctuation, but boring it is not. Informative and delightful it is. Lynne Truss includes in the book—which she says is not about grammar—wonderful examples of misused and misplaced punctuation marks. She claims to have written the book to unite us sticklers who do care about the written word, and how we communicate through it. We sticklers cringe with many misuses of punctuation, and we are cringing more and more often it seems.

Truss defines punctuation as a tool to clarify the written word, and who can argue with helps for clarification? She suggests that punctuation is dying, but then asks what would happen without it? Just imagine all the words in the first paragraph with no punctuation marks and no capital letters. You might be able to figure out its meaning with some work, but it would not be easy. Also consider, she suggests, the following:

> A woman, without her man, is nothing.
> A woman: without her, man is nothing.

Punctuation makes all the difference!

The book begins with a discussion of the apostrophe. Meaning "omission," the apostrophe was first used in the 16th century. The most common egregious misuse of this tool is found in the word "it's." It's translates "it is," but it is often used as a possessive word, as in "The keyboard is useless; some of it's keys are missing," when it should be "The keyboard is useless; some of *its* keys are missing." As a test, if you cannot substitute the words "it is" or "it has," it should be "its;" if you can, it is correctly "it's." And the same is true for you're and your. You're translates "you are," and your is the possessive ("It's your turn").

Another amusing example Truss gives is: Member's May Ball. Of course it should be Members' May Ball, because who would just one member dance with? Truss asks.

In her discussion of the comma, we learn that commas were first used 2000 years ago by Greek dramatists to show the actors where to pause or breathe. Then when printing was invented and used increasingly in the 14th and 15th centuries, a Mr. Aldus Manutius (1450–1515) developed italics, the semicolon, the comma, the colon, and full stops (we call them periods in the U.S.).

Truss is a master of the metaphor. She calls the comma the "sheepdog" of words. The comma organizes words, phrases, and groups of words that fit together. Consider one of her comma examples, a properly placed comma: No dogs, please.

Now think about that sentence without the comma: No dogs please. Now consider this: But many dogs *do* please. Thus the importance of the properly placed comma.

Truss addresses all the other marks, including semicolons, quotation marks, brackets, hyphens, parentheses, the four attention-grabbers: *italics*, the exclamation point, the dash —, and the question mark, and finally the ellipsis (the three dots ...). She tells us that, amazingly, someone actually did a PhD thesis on the ellipsis!

One chapter discusses the fact that proper use of punctuation steadily declined in the 20th century, many blaming the decline on television; and that it will continue to decline in the 21st century because of the Internet. E-mail messages cry for brevity, and brevity they get. For example, "**CU B4 8.**" "Netspeak" is, no doubt, here to stay. Language usage also is trending toward the deletion of spaces between words, so that now we say healthcare, chatroom, and the like.

And finally, Truss discusses the newest job that punctuation marks have assumed: emoticons. Examples include the smiley face :–), the sad face :–(, and many others, all made with common punctuation marks.

I thoroughly enjoyed this book, and recommend it to anyone who wants to learn while being entertained. It is a wonderful read.

—Bonnie E. Hupton, Editor
bhupton@sbcglobal.net

Read Any Good Books Lately?

Then why not share your thoughts with the readers of IPJ? We accept reviews of new titles, as well as some of the "networking classics." In some cases, we may be able to get a publisher to send you a book for review if you don't have access to it. Contact us at **ipj@cisco.com** for more information.

Fragments

Paul V. Mockapetris Wins 2005 ACM SIGCOMM Award

Paul V. Mockapetris, Chairman and Chief Scientist at Nominum Inc., is the winner of the 2005 *ACM SIGCOMM Award*. The SIGCOMM Award is widely recognized as the higheset honor in computer networking. The Award recognizes lifetime achievement in and contributions to the field. It is awarded annually to a person whose work, over the course of his or her career, represents a significant contribution to the field and a substantial influence on the work and perceptions of others in the field. The SIGCOMM Award is presented to Dr. Mockapetris "in recognition of his foundational work in designing, developing and deploying the *Domain Name System* (DNS), and his sustained leadership in overall Internet architecture development."

Paul Mockapetris created the original DNS protocol, wrote its first implementation, and worked with others to spread the DNS across the Internet. The design of DNS, which was the first major datagram protocol of the Internet, established a number of principles for key Internet infrastructural services. Its simplicity of design and fitness for purpose have stood the test of time. The strength of its design lies in a novel combination of hierarchy and caching that gives each organization absolute control over part of the namespace while simultaneously relying on caching to make the entire system efficient. Its success can be seen from the fact that DNS now handles many orders of magnitude more names and traffic than when it was first deployed, and yet the design and structure have remained intact. As a result the DNS design and caching mechanisms are often cited as two of the cornerstones on which the success of the Internet is built.

In addition to his work on DNS, Dr. Mockapetris' career has included pioneering work on multiprocessor operating systems, virtual machines, and ring LAN technology. Further, Dr. Mockapetris played an important role in the deployment of networking technologies internationally. Starting during 1990–1993 as a program manager at ARPA, Dr. Mockapetris fostered the international deployment of multimedia conferencing, multicast, and QoS. His strong leadership in development of Internet architecture continued as Chair of the Internet Engineering Task Force during 1994–1996, as member of the Internet Architecture Board during 1994–1996, and then as member of the Federal Networking Council. Dr. Mockapetris is also a recipient of the *IEEE Internet Award* and is an ACM Fellow.

In summary, through his sustained effort in support of the Internet archicture, beginning with DNS and continuing through work at ARPA, IETF, and industry, Dr. Mockapetris has made far-reaching and influential contributions to computer networking. The 2005 SIGCOMM award recognizes Dr. Mockapetris for this lifetime record of achievement.

SIGCOMM is the *Special Interest Group* (SIG) on *Data Communication* of the *Association for Computing Machinery* (ACM). SIGCOMM is a professional forum for the discussion of topics in the field of communications and computer networks, including technical design and engi-neering, regulation and operations, and the social implications of computer networking. The SIG's members are particularly interested in the systems engineering and architectural questions of communication. For more information please visit: **http://www.acm.org/sigcomm/**

Voice over IP (VoIP) And Government Policy

Voice over IP technology has the potential to provide much cheaper telephone service, particularly internationally. More importantly, it can enable exciting new services, such as voice-enabled Web pages and integrated phone, voice-mail, and e-mail. Unfortunately, some national governments are trying to limit its use. In late April, 2005, the *Advisory Committee on International Communications and Information Policy* (ACICIP) of the U.S. Department of State issued a very useful paper describing how VoIP works, the benefits it can provide, and what governments around the world are doing to promote or hinder its development.

Michael Nelson, the Internet Society's Vice President for Policy, represents ISOC on the Committee, and is helping draft "Version 2.0" of the paper, which will report on recent developments in additional countries. If you would like to make suggestions about the paper, please submit them to Michael Nelson at **mnelson@isoc.org**

For more information, see:
http://isoc.org/pubpolpillar/voip-paper.shtml

ISOC Commentary on the Status of the Work of WGIG, April 2005

When the first phase of the *World Summit on the Information Society* (WSIS) called on the UN Secretary General to set up the *Working Group on Internet Governance* (WGIG), it was in the context of supporting the WSIS *Action Plan*. The Plan calls for concrete actions to advance the achievement of internationally agreed development goals by promoting the use of ICT-based products, networks, services and applications, and to help countries overcome the digital divide. This is, by the way, something the Internet community has worked hard to achieve since the very first days of the Internet.

These goals include those described in the *Millennium Declaration*. The 8th goal of that document is to develop a global partnership for development, which would make available the benefits of new technologies—especially information and communications technologies—in cooperation with the private sector for the benefit of all. This is the context (making the benefits of ICT available to everyone) in which we initially engaged in the WSIS and WGIG efforts. The Internet has a huge potential as an enabler bringing these benefits to people everywhere and we remain excited about the WSIS mission. However, it is not clear how WGIG's actions to date have helped support achieving such goals.

The *Internet Society* (ISOC) believes that the best way to extend the reach of the Internet is to build on those aspects that have worked well, for example, the long established open, distributed, consensus-based processes and many regional forums for the development and administration of the Internet infrastructure. Decision-making about issues such as resource allocation or IP Address Policy has always been in the hands of the Internet community, in order to be as close to those who require and use the resources as possible. It is this participative model, close to the end users, that led to the phenomenal, stable growth of the Internet. The Internet community and its bottom-up processes are constantly evolving in response to changes in needs and availability. For example, in response to moves by the African Internet community, the African countries now have their own *Regional Internet Registry* [RIR] (AfriNIC) that helps coordinate users' needs and IP Policy in that region. Latin America has the same story to tell. Support for the development of both these RIRs (educational, financial and boot-strapping of various processes) came from the global Internet community and primarily came from the other RIRs.

Developing and maintaining the Internet infrastructure are just two aspects of what has come to be referred to as *Internet Governance*. WGIG has pointed out that there are many others, and has recognized the fact that Internet Governance encompasses a much wider range of topics than IP address and domain name administration. However, much of WGIG's focus has been on Internet infrastructure, thereby missing an opportunity to focus on those aspects of the Internet's development that are less developed and that could benefit from improved, lightweight mechanisms facilitating an exchange of information between policymakers and the Internet community. Examples here are issues con-cerning inappropriate usage of the Internet—cybercrime and spam being just two examples. Much work has already been done on technical solutions to these issues, and many legal frameworks already exist for handling criminal activity such as fraud. The challenge today is to bring the lawmakers and policymakers together with the Internet community to discuss the most appropriate mechanisms to ensure the continued development of the Internet.

Many players have a role, and this clearly includes governments and intergovernmental organizations. WGIG had a clear mandate to not only develop a working definition of Internet Governance, but also to develop a common understanding of the respective roles and responsibilities of governments, existing intergovernmental and international organizations and other forums, as well as the private sector and civil society encompassing both developing and developed countries. Unfortunately an inordinate amount of time has been spent focusing on challenging current structures (those that brought us the Internet and its rapid, stable growth), rather than looking forward to the potential benefits of extended cooperation with (and based on the proven success of) existing models and structures. WGIG seems to have lost sight of this larger goal.

Also, many of WGIG's premises seem to start with an assumption that the Internet needs a hierarchical top-down governance model, thereby ignoring the decentralized, distributed structure on which the Internet was so successfully built. Not only does this "governance hierarchy" model prevent an accurate understanding of the Internet's infrastructure and development (forcing key organizations to be classed in prescribed categories that do not fit with the reality of their actions or their role in developing and supporting the Internet) but it also will very likely lead to conclusions that will harm the Internet's development and growth.

While WGIG appears to ascribe the growth of the Internet to deliberate regulatory decisions to liberalize telecommunications, in reality regulatory measures have been a relatively small factor. A more significant factor in the growth of the Internet has been the fact that the Internet architecture has enabled many tens of thousands of users to develop their own applications independent of the underlying architecture, thereby empowering people to add true value to the global Internet network. The continued expansion of the Internet to developing countries though will be greatly aided in the future by a more competitive telecommunications environment. We urge WGIG to recommend more concrete and aggressive action in this direction.

Further, WGIG has put great focus on comparing the relative merits of established treaty bodies and intergovernmental organizations to undertake a central role in the development of Internet infrastructure while very largely overlooking areas where attention and support are required and where national governments more naturally have a role to play, areas such as misuse of the Internet (cybercrime and spam to name a few). The limited perspective of this approach displays an obvious bias in the characterization of the issues and seems to pre-suppose a solution. In conclusion, we would urge WGIG to spend more time looking at what is actually being done to enable more people around the world to take greater advantage of the power of the Internet. This includes a focus on the many regional and global education activities that different Internet-related organizations are undertaking to "connect the unconnected."

These same organizations are also working to make the Internet more secure, more accessible, more reliable, more affordable, and more versatile. The development of the Internet, as well as many well-established capacity-building efforts could be jeopardized by applying a too heavy-handed approach to the operation and administration of this unique network of networks. Decentralized, lightweight governance has clearly proven itself to be a positive feature not a weakness. We want to encourage WGIG and WSIS to work with the Internet community within the already well-established Internet model to improve co-operation between policy makers and the Internet community.

In the spirit of meeting the international development goals highlighted by WSIS, any review of today's Internet model or structures must be carried out in the context of how well they have worked in the past, how well they meet the needs of the people who depend upon them today, and how well they will adapt to changing requirements in the future; and not simply focus on a comparison to other historical telecommunications or governance models. These historical models have not been demonstrated to be well suited to the Internet. For more information, see:

http://isoc.org/

http://wgig.org/

http://www.itu.int/wsis/

An interview with the new IETF Chair

IBM Distinguished Engineer and former ISOC Chairman Dr. Brian Carpenter has just taken over the role of IETF Chair. In a recent interview, Brian describes the future challenges facing the IETF and the Internet in general. The full interview is available here:
http://resources.isoc.org/20503

Upcoming Events

The *Internet Corporation for Assigned Names and Numbers* (ICANN) will meet in Luxembourg City, Luxembourg, July 11–15, 2005 and in Vancouver, Canada November 30–December 4, 2005. For more information see: **http://www.icann.org**

The *South Asian Network Operators Group* (SANOG) will meet in Thimpu, Bhutan, July 16–23, 2005. More info at:
http://www.sanog.org

The *Internet Engineering Task Force* (m nn) will meet in Paris, France, July 30–August 5, 2005 and in Vancouver, Canada, November 6–11, 2005. For more information, visit: **http://ietf.org**

ACM's *SIGCOMM 2005* will be held in Philadelphia, PA, August 22–26, 2005. For more information visit:
http://www.acm.org/sigs/sigcomm/sigcomm2005

The *North American Network Operators' Group* (NANOG) will meet in Los Angeles, October 23–25, 2005. For more information see:
http://nanog.org

Call for Papers

The Internet Protocol Journal (IPJ) is published quarterly by Cisco Systems. The journal is not intended to promote any specific products or services, but rather is intended to serve as an informational and educational resource for engineering professionals involved in the design, development, and operation of public and private internets and intranets. The journal carries tutorial articles ("What is…?"), as well as implementation/operation articles ("How to…"). It provides readers with technology and standardization updates for all levels of the protocol stack and serves as a forum for discussion of all aspects of internetworking.

Topics include, but are not limited to:

- Access and infrastructure technologies such as: ISDN, Gigabit Ethernet, SONET, ATM, xDSL, cable, fiber optics, satellite, wireless, and dial systems
- Transport and interconnection functions such as: switching, routing, tunneling, protocol transition, multicast, and performance
- Network management, administration, and security issues, including: authentication, privacy, encryption, monitoring, firewalls, trouble-shooting, and mapping
- Value-added systems and services such as: Virtual Private Networks, resource location, caching, client/server systems, distributed systems, network computing, and Quality of Service
- Application and end-user issues such as: e-mail, Web authoring, server technologies and systems, electronic commerce, and application management
- Legal, policy, and regulatory topics such as: copyright, content control, content liability, settlement charges, "modem tax," and trademark disputes in the context of internetworking

In addition to feature-length articles, IPJ will contain standardization updates, overviews of leading and bleeding-edge technologies, book reviews, announcements, opinion columns, and letters to the Editor.

Cisco will pay a stipend of US$1000 for published, feature-length articles. Author guidelines are available from Ole Jacobsen, the Editor and Publisher of IPJ, reachable via e-mail at **ole@cisco.com**

The Internet Protocol Journal

Ole J. Jacobsen, Editor and Publisher

The Internet Protocol Journal is published quarterly by the Chief Technology Office, Cisco Systems, Inc.
www.cisco.com
Tel: +1 408 526-4000
E-mail: ipj@cisco.com

The Internet Protocol Journal, Cisco Systems
170 West Tasman Drive, M/S SJ-7/3
San Jose, CA 95134-1706
USA

ADDRESS SERVICE REQUESTED

PRSRT STD
U.S. Postage
PAID
PERMIT No. 5187
SAN JOSE, CA