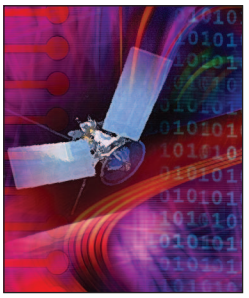
Editor: Chris Metz • chmetz@cisco.com



The Transition to IPv6, Part I

4over6 for the China Education and Research Network

Jianping Wu, Yong Cui, and Xing Li • *Tsinghua University*
Chris Metz • *Cisco Systems*

Researchers at the Tsinghua University in Beijing propose a mechanism for letting IPv4 networks communicate with each other across an IPv6 backbone via a Border-Gateway Protocol (BGP)-based control plane for advertising tunnels and IPv4 network prefixes. They've deployed a prototype implementation on the native-IPv6 China Education and Research Network 2 (CERNet2) backbone as the major part of the China Next-Generation Internet (CNGI) project, using current packet-encapsulation technology and an extension of BGP, and the IETF is currently considering their 4over6 mechanism as well.

To promote the transition from IPv4 and to propel IPv6 development, some countries are establishing large-scale pure IPv6 backbone networks. Yet, the large number of IPv4-based Internet applications and services presents the important challenge of how to let IPv4 networks communicate over IPv6 backbones.

To address this problem, we propose an IPv4 network-interconnection mechanism for use in IPv6 networks. Our *4over6* solution defines a Border-Gateway Protocol (BGP)-based control plane to advertise 4over6 tunnels and IPv4 network prefixes. The data plane uses standard IP encapsulation and decapsulation performed at IPv4–IPv6 dual-stack routers. We realize IPv4 network interconnection by use of routing transport on the control plane and packet transport on the data plane. Because it avoids explicit tunnels and manual configuration, the 4over6 mechanism is lightweight, adaptive to dynamic routes, and transparent for network end systems. Highly scalable and easy to deploy, it's designed for interconnecting large-scale networks.

We've deployed a prototype implementation of the 4over6 mechanism on a native IPv6 backbone in China, based on current packet-encapsulation technology and an extension of BGP. In addition to ongoing work in China, 4over6 is currently under consideration at the IETF.

IPv4 and IPv6 Coexistence

Supporting 20 million end users at 1,500 universities and institutions, the China Education and Research Network (CERNet) is the world's largest academic network (see Figure 1). In 2004, CERNet's operators won the bid to build its successor, dubbed CERNet2, and the China Next-Generation Internet (CNGI) exchange point (CNGI-6IX), located in Beijing. CERNet2 is the world's largest native IPv6 backbone network (see Figure 2, p. 82).

CERNet and CERNet2 have deployed both IPv4 and IPv6 networks and will continue to do so, which introduces a common and growing requirement for coexistence between the two IP address families. Different IPv4-to-IPv6 transition techniques come into play for communications between IPv6 networks over an IPv4 backbone, between IPv4 networks over an IPv6 backbone, and between IPv4 and IPv6 networks, hosts, and applications.

IPv6 packets can be transported across an IPv4 backbone using tunnels configured manually via IPv6-in-IPv4 or Generic Routing Encapsulation (GRE).[1,2] Automatic tunnel establishment, on the

 Published by the IEEE Computer Society 1089-7801/06/$20.00 © 2006 IEEE

Figure 1. The China Education and Research Network. CERNet's topology includes roughly 40 points-of-presence that form a backbone with bandwidths ranging from 10 Gbits down to 2.5 Gbits at the regional level, supporting many national network applications, including distance learning, digital libraries, and grid computing.

other hand, employs special IPv4-mapped IPv6 addresses as destinations, whereas 6to4 tunnels use the current IPv4 routing infrastructure and special IPv6 addresses to dynamically route encapsulated IPv6 packets to the nearest entry point into the IPv6 Internet.[3] Isolated IPv6 hosts can become interconnected functional IPv6 hosts through IPv4 multicast implementations.[4] Network operators can also use techniques such as Multiprotocol Label Switching (MPLS) to tunnel IPv6 or IPv4 packets, or they can outfit a host with a *dual stack*, so that it communicates with IPv6 hosts using IPv6 and with IPv4 hosts using IPv4. As is the case with CERNet2, however, deploying a native IPv6 backbone with attached IPv4 access networks requires an IPv4-over-IPv6 solution, so that IPv4 networks can communicate with each other across the backbone.

Most tunneling protocols focus on IPv6-over-IPv4 rather than IPv4-over-IPv6 scenarios. Although some encapsulation methods enable generic IPv4 packet tunneling in IPv6,[5] each such tunnel requires manual configuration — a burden that could hinder the build-out and deployment of native IPv6 networks supporting IPv4 connectivity. Given that the Internet and its application base are currently IPv4, operators could well resist transitioning their backbone networks to IPv6 without a scalable, automatic IPv4-over-IPv6 tunneling solution.

4over6 Transit Solution

Figure 3 (p. 83) illustrates the components of the 4over6 Transit solution. Developed by researchers at Tsinghua University at Beijing, it provides an automatic tunneling mechanism for scalable IPv4 packet transmission over an IPv6 backbone.[6] Provider (P) routers running the IPv6 protocol stack comprise the native IPv6 backbone, with

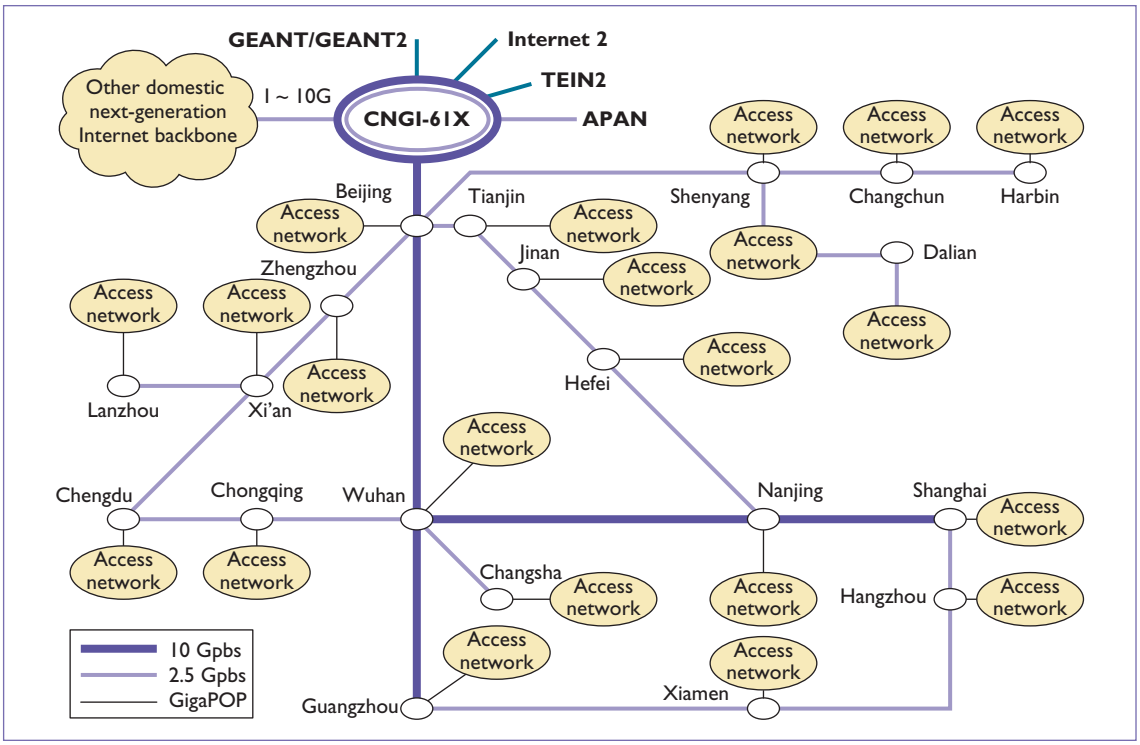


Figure 2. CERNet2. With interconnections to GÉANT, Internet2, and the Asia-Pacific Advanced Network (APAN) via the China Next-Generation Internet exchange point (CNGI-6IX), CERNet2 is the world's largest native IPv6 backbone network. It employs the 4over6 mechanism to support the coexistence of IPv4 and IPv6.

provider edge (PE) routers running dual stacks at the edge. IPv4-only access networks then connect to one or more PE routers via IPv4 customer edge (CE) routers. The PE routers communicate through IPv4 with the attached CE routers and IPv6 with the P routers. In addition, the PE routers communicate with each other to establish dynamic IPv4-over-IPv6 inter-PE tunnels and to advertise which IPv4 networks are reachable through a particular tunnel. With these tunnels in place, a source IPv4 host in one access network can send IPv4 packets, encapsulated in IPv6 tunnel headers, across the IPv6 backbone to a target IPv4 host in another access network. This is the functionality we call 4over6.

A key differentiator of the 4over6 solution is the fact that a PE router automatically builds the inter-PE tunnels bearing IPv6-encapsulated IPv4 packets. In the process, a PE also distributes the IPv4 network prefixes that are reachable through that IPv6 tunnel (from the local PE's perspective). In other words, remote PEs can use control-plane functions to tell local PE routers about IPv6 tunnels they can use to reach a set of IPv4 network prefixes downstream from the remote PE.

The control-plane function operating between the PE routers is based on the Multiprotocol Border Gateway Protocol. MP-BGP is a natural choice because it runs on PE routers and is easily extensible to transport new routing information between those routers in a reliable, scalable manner.[7]

Figure 4 illustrates the format of the MP-BGP attribute containing the 4over6 routing information.[8] PE routers supporting 4over6 first exchange BGP-capability messages in which

- the `address family identifier` (AFI) = 1, indicating that the field contains an IPv4 prefix;
- the `subsequent AFI` (SAFI) = 67, indicating that the field carries 4over6 information;
- the `next_hop` field includes the IPv6 address of a virtual interface on the advertising PE; and
- the `network length reachability information` (NLRI) field includes an IPv4 network prefix that's reachable through the IPv6 tunnel that terminates on the advertising PE router.

Remote PE routers (such as PE2 in Figure 3) advertise subsequent BGP update messages to the local PE routers (PE1 in Figure 3, for instance). The local

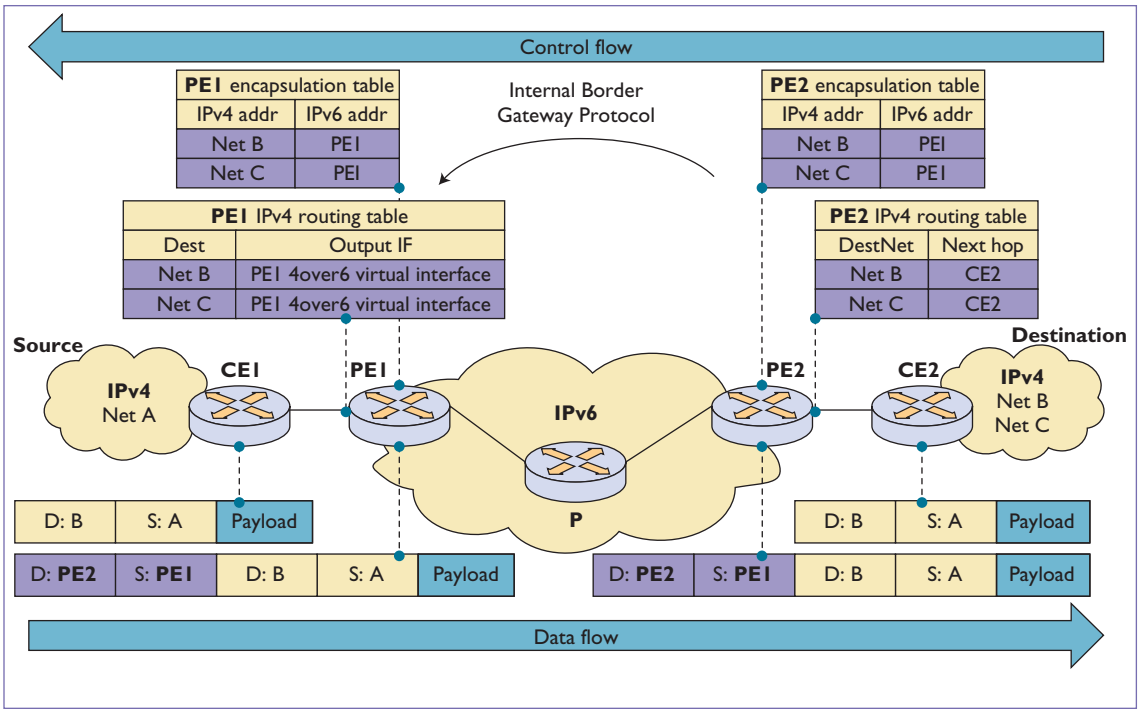


Figure 3. 4over6 Transit. This proposed solution defines a BGP-based control plane to advertise 4over6 tunnels and IPv4 network prefixes. The data plane uses standard IP tunnel encapsulation and decapsulation performed at the provider edge (PE) routers while custom edge (CE) routers remain unchanged from ordinary IPv4 routers.

PE routers then store this information and use it to forward IPv4 packets to the correct tunnel leading to the remote PE router.

4over6 Forwarding

In our 4over6 transition solution, IPv4 packets need to travel over an IPv6 backbone from the original IPv4 access network to the destination IPv4 access network. This process, which we call 4over6 forwarding, essentially comprises three parts:

- encapsulation of the incoming IPv4 packet with an IPv6 header;
- transmission of the encapsulated packet over the IPv6 transit backbone; and
- decapsulation of the IPv6 header and transmission of the original IPv4 packet.

Given that the IPv6 transit backbone is unaware of the IPv4 payload, transmitting the encapsulated packet across the backbone is business as usual. The 4over6 encapsulation and decapsulation processes occur exclusively on the PE routers.

Each 4over6 PE router maintains an encapsulation table with one or more entries composed of the destination IPv4 network address and the corresponding advertising remote PE router's IPv6 4over6 virtual interface (VIF) address.

When an IPv4 packet arrives at the ingress

Address family identifier (2 octets): AFI_IP = 1
Subsequent AFI (1 octet): SAFI_4 OVER6 = 67
Length of next hop (1 octet): 16
Next hop: IPv6 address of 4over6 virtual interface
Subnetwork point-of-attachment (SNPA) address
Length of first SNPA(1 octet)
First SNPA (variable)
Length of second SNPA (1 octet)
Second SNPA (variable)
. . .
Length of Last SNPA (1 octet)
Last SNPA (variable)
Network length reachability information NLRI (variable): Destination IPv4 network address

Figure 4. Multiprotocol Border-Gateway Protocol attribute format. Provider edge routers act as 4over6 functional entities to distribute 4over6 routing information to each other via the MP-BGP attribute field.

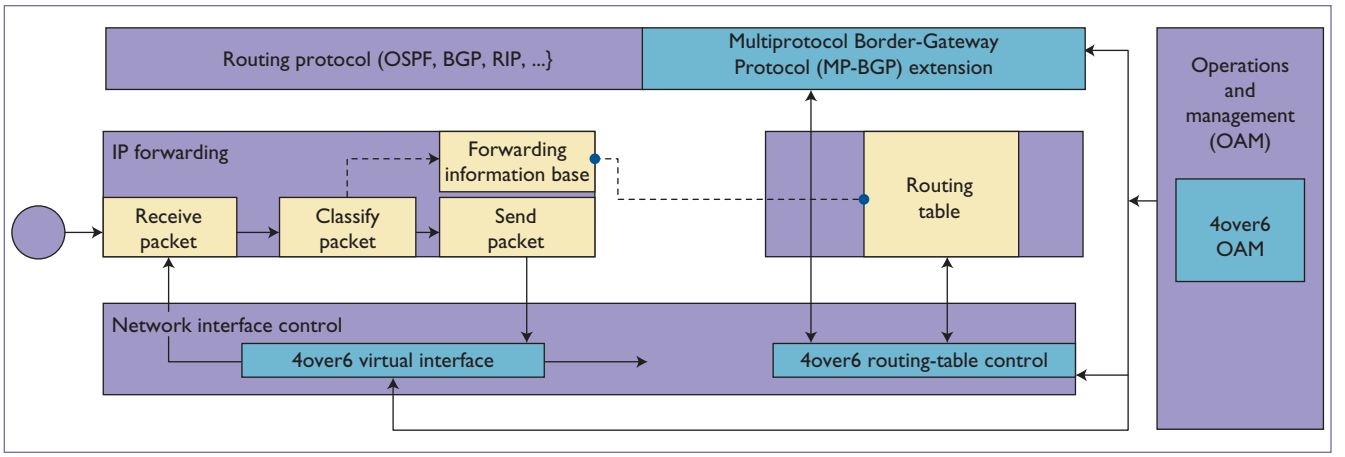


Figure 5. 4over6 modules in the prototype solution developed at the Tsinghua University Lab. In addition to the original IP forwarding subsystem, the prototype includes four 4over6 modules: operations and management (OAM), Multiprotocol Border-Gateway Protocol (MP-BGP) extension, routing table (RT) control, and virtual interface (VIF), which are located in the OAM, BGP, and network interface control subsystems, respectively.

4over6 PE router, a lookup of the destination address yields a pointer to an entry in the encapsulation table. From there, the 4over6 PE router constructs a new IPv6 header in which to encapsulate the IPv4 packet. The source address in this header is the IPv6 address of the VIF on the *ingress* 4over6 PE router, whereas the destination address is the IPv6 address of the VIF on the *egress* 4over6 PE router. The ingress 4over6 PE router then forwards the packet across the native IPv6 backbone network based only on the destination address contained in the IPv6 header. When the egress 4over6 PE router receives the packet, it removes the IPv6 header and forwards the original IPv4 packet to the downstream CE router and on to the target IPv4 host.

4over6 Prototype

We developed a Linux-based prototype of our solution at the Tsinghua University Lab in Beijing, and deployed it on CERNet2. As Figure 5 illustrates, our 4over6 implementation on the routers has four functional modules:

- 4over6 operations and management (OAM);
- MP-BGP extension;
- 4over6 routing table (RT) control; and
- 4over6 VIF.

The 4over6 OAM module located in the OAM system area processes configuration commands that define the PE routers' tunnel attributes (IP version number, interface address, tunnel identifier, tunnel type, and so on), the 4over6 address prefix, the startup of 4over6 function and route-redistribution commands.

The MP-BGP extension module supports the distribution of the 4over6 attribute, which in turn populates the PE encapsulation table containing the destination IPv4 prefix and IPv6 header address entries. It will also listen to and respond to signals generated by the 4over6 RT control module when the PE encapsulation's contents change.

The 4over6 RT control module receives the routing information from both the MP-BGP and routing management (RTM) system to construct and maintain the encapsulation table. This module also notifies the MP-BGP extension module to generate update messages when items change and updates the items received by the MP-BGP module for storage in the PE encapsulation table.

The ingress and egress 4over6 PE routers use the VIF to point to the appropriate PE encapsulation tables, which in turn yield the appropriate IPv6 header or IPv4 next-hop address. Because a router will set a corresponding entry in the forwarding table for each interface, including the VIF, the original IP-forwarding subsystem remains unchanged from a general dual-stack router, which looks up forwarding tables for received packets and forwards them to 4over6 VIFs as the output interfaces. The VIF module then encapsulates or decapsulates the packet according to the encapsulation table maintained by the 4over6 RT control module.

IETF Softwires

Recognizing the need to build and extend the

existing set of IPv6 transition mechanisms, the IETF established the Softwires working group (www.ietf.org/html.charters/softwire-charter.html). Simply put, a softwire is a tunnel that supports IPv6-over-IPv4 or IPv4-over-IPv6 connectivity. The working group is chartered to define "discovery, control, and encapsulation methods for connecting IPv4 networks across IPv6 networks and IPv6 networks across IPv4 networks in a way that will encourage multiple, interoperable implementations."

To date, the Softwires working group has met at the past two IETF meetings and completed a problem statement, which presents the general softwire problems in which island networks of a particular address family (IPv4, for example) need to communicate with one another across a backbone of the other address family. Two typical scenarios address this problem: hubs and spokes, as characterized by one connection and an associated static default route, and mesh, which is characterized by multiple connections and routing prefixes.[9]

We presented our 4over6 solution to the Softwires working group as an Internet draft and are now working with industry colleagues to develop a more general softwire mesh solution that will employ MP-BGP as a means to advertise softwire tunnel encapsulation types, header information, and reachability to different IP network prefixes through a particular softwire tunnel.

In the next installment in this series, we'll describe the softwire mesh solution framework in greater detail. The working group envisages that this effort will form the basis for large-scale backbone deployments of native IPv6 and IPv6-tunnelled traffic, easing the transition to IPv6 in the long run.

Acknowledgments

CERNet is funded by the Chinese Government and managed by the Ministry of Education. The CNGI–CERNet2 project is supported by the China Next-Generation Internet Project, approved by the State Council, and organized jointly by eight ministries and commissions, led by the National Development and Reform Commission. This work is supported by the National Major Basic Research Program of China (no. 2003CB314801) and the National Natural Science Foundation of China (no. 60303006, 90604024).

References

1. R. Gilligan and E. Nordmark, *Transition Mechanisms for IPv6 Hosts and Routers*, IETF RFC 2893, Aug. 2000; www.ietf.org/rfc/rfc2893.txt.
2. S. Hanks, T. Li, and P. Traina, *Generic Routing Encapsulation over IPv4 Networks*, IETF RFC 1702, Oct. 1994; www.ietf.org/rfc/rfc1702.txt.
3. B. Carpenter et al., *Connection of IPv6 Domains via IPv4 Clouds*, IETF RFC 3056, Feb. 2001; www.ietf.org/rfc/rfc3065.txt.
4. B. Carpenter and C. Jung, *Transmission of IPv6 over IPv4 Domains without Explicit Tunnels*, IETF RFC 2529, Mar. 1999; www.ietf.org/rfc/rfc2529.txt.
5. A. Conta and S. Deering, *Generic Packet Tunneling in IPv6 Specification*, IETF RFC 2473, Dec. 1998; www.ietf.org/rfc/rfc2473.txt.
6. J. Wu, Y. Cui, and X. Li, "4over6 Transit using Encapsulation and BGP-MP Extension," IETF Internet draft, work in progress, Feb. 2006.
7. T. Bates et al., *Multiprotocol Extensions for BGP-4*, IETF RFC 2858, June 2000; www.ietf.org/rfc/rfc2858.txt.
8. R. Chandra and J. Scudder, *Capabilities Advertisement with BGP-4*, IETF RFC 2842, May 2000; www.ietf.org/rfc/rfc2842.txt.
9. X. Li et al., "Softwire Problem Statement," IETF Internet draft, work in progress, Feb. 2006.

Jianping Wu is a full professor in the computer science department at Tsinghua University, Beijing. He is also the director of the China Education and Research Network. His current research interests include computer network architectures, next-generation Internet, and formal methods. Wu has a PhD in computer science from Tsinghua University. Contact him at jianping@cernet.edu.cn.

Yong Cui is an assistant professor in the computer science department at Tsinghua University, Beijing. His current research interests include Internet architectures, IPv6 transition, and quality of service. Cui has a PhD in computer science from Tsinghua University. Contact him at cy@csnet1.cs.tsinghua.edu.cn.

Xing Li is a full professor in the electronic engineering department at Tsinghua University, Beijing. He is also a vice director of the China Education and Research Network. Li's current focus is on Internet architectures, IP multicast, and routing architectures. Li has a PhD in electrical engineering from Drexel University. Contact him at xing@cernet.edu.cn.

Chris Metz is a technical leader in the Routing Technology Group for Cisco Systems, based in San Jose, Calif. His current areas of interest include Internet architectures and services, IP/MPLS, transport-layer protocols, and layer-2/layer-3 virtual private networks. Contact him at chmetz@cisco.com.