IP-in-IPv6 Overlay Networking Technology for a Terabit-class Super-network

Yuuichi Naruse†, Takeshi Yagi, Kenichi Matsui, and Junichi Murayama

Abstract

This paper proposes IP-in-IPv6 overlay networking technology for a terabit-class super-network as a provider network to offer Internet protocol virtual private network (IP-VPN) services. In the conventional architecture based on BGP/MPLS IP-VPNs (BGP: border gateway protocol, MPLS: multiprotocol label switching), the number of provider edge routers that can be accommodated is limited to about one hundred. The scalability target of this network is to accommodate a thousand provider edge routers. In our solution, IPv6 is deployed as a provider network protocol and VPNs are laid over the IPv6 network. In addition, legacy connectionless forwarding and new address management techniques are applied to the network. As a result, our proposal economically achieves scalability and provides value-added functions such as VPN, load balancing, and multi-grade security.

1. Introduction

With the spread of the Internet, there has been a growing recognition of the importance of security. Because of its security features, an Internet protocol virtual private network (IP-VPN) is an attractive solution for enterprise networks, home networks, and Internet service provider (ISP) networks. Thus, good scalability is important for a provider network that offers an IP-VPN service, and it is especially important that a large-scale provider network can be built up from many small-scale routers. This is because, from the viewpoint of economy, the provider network should be able to grow gradually as the number of customers increases. Here, our specific objective is to let a provider network accommodate a thousand provider edge routers.

In the familiar VPN solution, an MPLS (multiprotocol label switching) network [1] is deployed as the provider network. In this network, logical paths called LSPs (label switched paths) are established in a mesh topology between provider edge routers. The use of these full-mesh paths means that an MPLS network cannot achieve scalability. In a practical implementation, the number of provider edge routers is limited to about a hundred because a provider router can manage only ten thousand paths. Although a large-scale MPLS network can be implemented using large-scale routers, this solution is not cost-effective at the initial service stage.

To solve this problem, we deploy an IPv6 network [2] as the provider network (IPv6: Internet protocol version 6). First, we apply legacy connectionless forwarding to the network. Thus, IPv6 packets are forwarded in a connectionless manner. Since full-mesh paths are not required, scalability is achieved. In a practical implementation, the number of provider edge routers can be more than a thousand, while the number of IP routes managed in a provider router can be less than a thousand. In this paper we call this IPv6-based provider network a terabit-class super-network.

Although an MPLS network is not scalable, it can offer many value-added functions such as IP-VPNs and load balancing. Thus, we also apply a new address management scheme to the network, which allows value-added information to be transferred using addresses instead of labels.

† NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Japan
E-mail: naruse.yuuichi@lab.ntt.co.jp

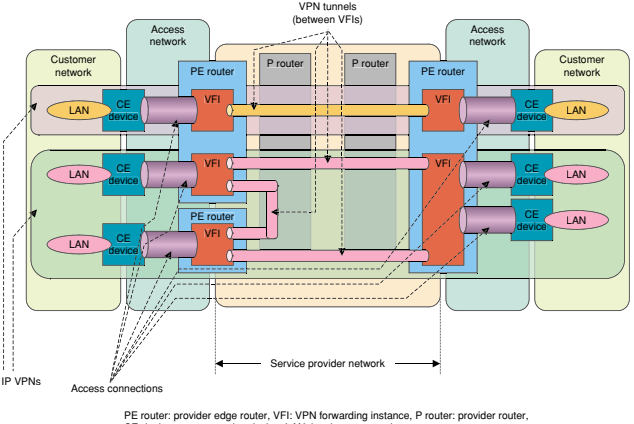
As a result, our proposed design economically achieves not only scalability but also value-added functions. Section 2 summarizes the design issues. Section 3 describes our proposed technology and section 4 evaluates its scalability. Section 5 describes node implementation and section 6 provides a summary and conclusions.

2. Design issues

A provider network that offers an IP-VPN service should conform to the PPVPN (provider provisioned virtual private network) framework [3] shown in **Fig. 1**. This framework is composed of a service provider (SP) network and access networks. The SP network is composed of provider (P) routers and provider edge (PE) routers, which are connected to each other via physical links. Each PE router accommodates router functions called VFIs (VPN forwarding instances). VFIs belonging to the same VPN are connected to each other via virtual connections called VPN tunnels, which are established in a mesh topology between all VFIs belonging to the same VPN if reachability is to be retained between them. Since VFIs belonging to different VPNs are not connected, VPN security is achieved. In access networks, access connections are established to connect VFIs and customer edge (CE) devices to each other. In the familiar solution, an MPLS network is deployed as the SP network and VPN tunnels are implemented by LSPs.

The first issue arising with this approach is the number of LSPs that must be established between VFIs, as shown in **Fig. 2**. LSPs must be established in a mesh topology between VFIs belonging to the same VPN. Approximately ten thousand LSPs are required to connect a hundred VFIs in a VPN. Thus, if the SP network accommodates a hundred such VPNs, the P routers must be able to handle a million LSPs. In a practical implementation, P routers can handle only ten thousand LSPs at most.

The second issue is the number of LSPs established between PE routers. To reduce the switching processing load of P routers, two-layered LSPs, as shown in **Fig. 3**, are effective [4]. In this solution, the upper-layer LSPs are established in a full-mesh topology between all PE routers irrespective of the VPN



PE router: provider edge router, VFI: VPN forwarding instance, P router: provider router,
CE-device: customer edge device, LAN: local area network

Fig. 1. PPVPN framework.

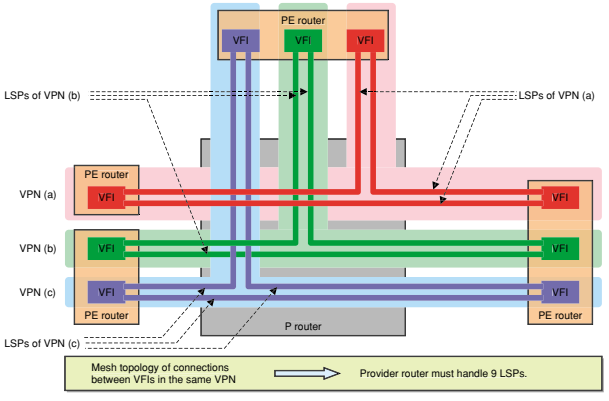


Fig. 2. Label switched paths (LSPs) for VPN tunnels.

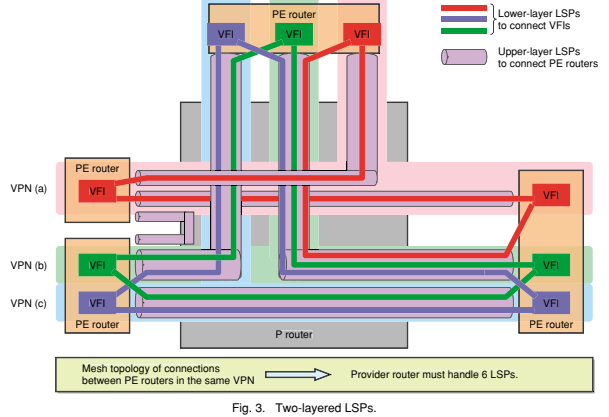


Fig. 3. Two-layered LSPs.

configurations and the lower-layer LSPs, which connect VFIs, are aggregated and tunneled into the former LSPs. At a rough estimation, the number of LSPs to be handled in P routers can be reduced to ten thousand when there are one hundred PE routers. However, it becomes a million when the number of PE routers increases to a thousand.

The third issue is traffic monitoring for traffic engineering. To reduce the number of LSPs established between PE routers, the upper-layer LSPs can be merged at P routers. Thus, a multipoint-to-point LSP is established from ingress PE routers to an egress router. In this solution, all ingress PE routers transmitting IP packets add the same label to IP packets if their destination is the same egress PE router. The P router checks the labels of IP packets received from each ingress PE router and decides the output port based on a label switching table. For the example shown in **Fig. 4**, all IP packets from ingress PE routers A, B, and D to egress router C are labeled with L-3, and the P router transmits these packets to port-3 after referring to a label switching table. That is, the number of entries in a P router's label switching table is only the same as the number of PE routers. However, in this approach, checking the label of IP packets does not allow P routers to classify the transit packets according to a pair of ingress-egress PE routers. This makes traffic engineering difficult, so the performance cannot be improved even if the scalability is improved.

3. Network design

Our design uses connectionless forwarding and an address management scheme to improve scalability and extend value-added functions, respectively.

3.1 Connectionless forwarding

To solve the problem described in Section 2 (the increasing number of LSPs), we apply connectionless forwarding to the SP network of the PPVPN framework, as shown in **Figs. 5** and **6**. In our solution, an IPv6 network is deployed as the SP network and IP-VPNs are laid over the IPv6 network. Each VFI is assigned its own IPv6 address and behaves as a terminal of the IPv6 network. VPN tunnels are mapped to IPv6 routes at ingress VFIs. A customer IP packet is transferred in the following three-step procedure.

1) Encapsulation: First, the ingress VFI encapsulates customer IP packets into provider IPv6 packets [5].
2) Connectionless forwarding: Next, P routers

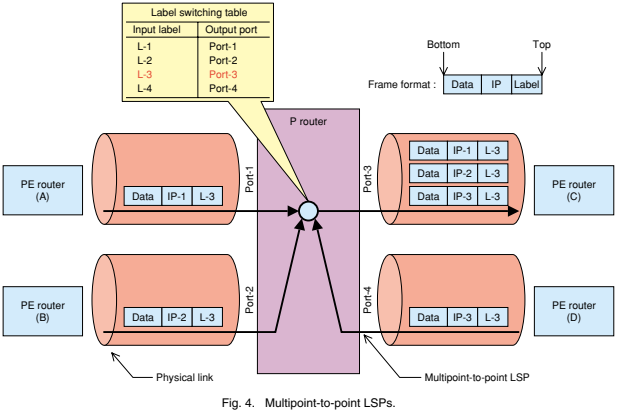


Fig. 4. Multipoint-to-point LSPs.

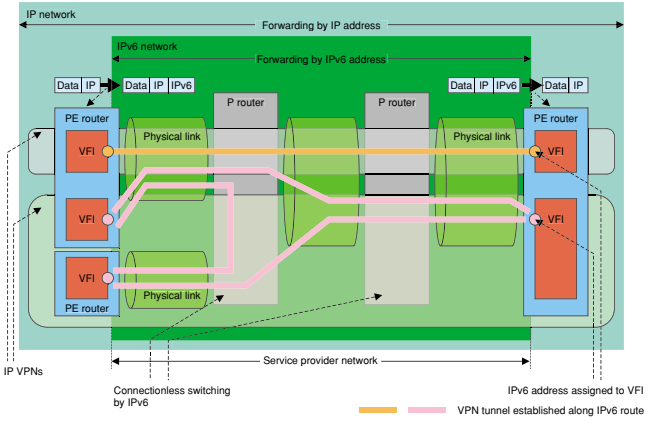


Fig. 5. Overview of the IP-in-IPv6 overlay networking technology.

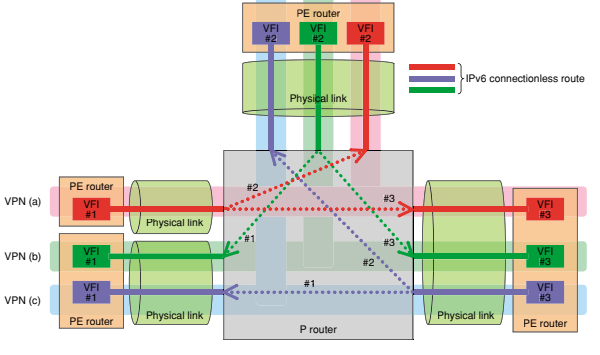


Fig. 6. IPv6 connectionless forwarding in provider router.

forward the IPv6 packets in a connectionless manner.

3) De-capsulation: Finally, the egress VFI de-capsulates the customer IP packets from the provider IPv6 packets and forwards them to the destination CEs.

Basically, P routers should identify IPv6 routes towards VFIs. However, they are only required to identify ones towards PE routers if the IPv6 address field is hierarchically partitioned into a PE router identifier and VFI identifier as shown in **Fig. 7**. Thus, the number of IP routes to be handled in P routers can be about the same as the number of PE routers. Since, in this approach, the P routers can classify the transit packets according to a pair of ingress-egress PE routers by referring to the PE router identifiers in the source and destination IPv6 addresses, traffic engineering can be performed as described in ref. [6]. Here, in traffic engineering, IPv6 routes are not distributed individually for each VPN, but collectively using the route identifier. Therefore, the number of IP routes to be handled in a P router does not become too large. Moreover, P routers do not have a filtering table, so the ability to provide scalability is not affected. When a large number of PE routers are to be accommodated, the IPv6 address field should be further partitioned to include an area identifier.

3.2 Address management

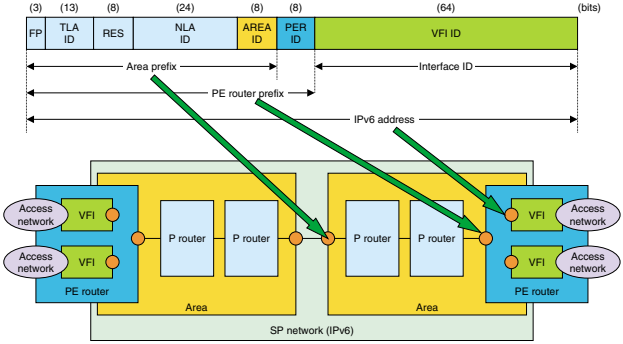
Although high scalability is achieved by the connectionless forwarding, extending value-added functions is also an important issue. In our approach, most functions are implemented by address management. Here, schemes for virtual private networking, load balancing, and multi-grade security are described. Note that, in the SP network, these schemes are not required in the packet forwarding procedure, but are required in the address management procedure.

(1) Virtual private networking

An IP-VPN is created by assigning an IPv6 address to a VFI and mapping a customer IP route to a provider IPv6 route corresponding to a VPN tunnel. To maintain VPN security, this procedure should be implemented by the network operator. However, configuration errors may occur, so to solve this issue, a unique VFI identifier in the IPv6 address is used for each IP-VPN. This means that a VFI identifier should be treated as a VPN identifier, as shown in **Fig. 8**.

(2) Load balancing

Load balancing is important to improve both reliability and performance economically. To improve reliability, multiple physical routes should be estab-



FP: format prefix for aggregatable global unicast addresses, TLA ID: top-level aggregation Identifier, RES: reserved for future use, NLA ID: next-level aggregation identifier, AREA ID: area identifier, PER ID: provider edge router identifier, VFI ID: VPN identifier

Fig. 7. Address hierarchy for SP network.

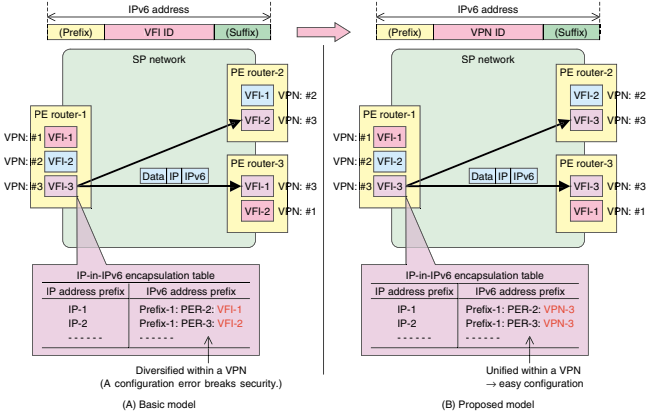


Fig. 8. Virtual private networking scheme.

lished and packets should be rerouted to bypass faults. Furthermore, to improve performance still further, all these routes should be used to balance loads even when there are no faults in the SP network.

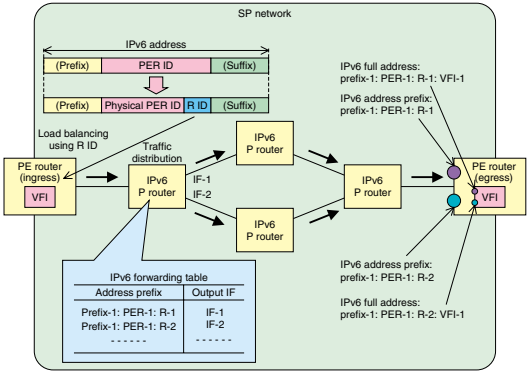
However, load balancing may cause IP packets to arrive in the incorrect order at the egress PE router. This is fatal for realtime streaming applications such as videoconferencing. The simple solution to this problem is for multiple IPv6 addresses to be assigned to an egress VFI and for P routers to manage multiple IP routes towards the same VFI. This means that an ingress VFI can map an application flow to a particular IPv6 address of the egress VFI.

In this simple solution, in each P router, IP routes towards each VFI should be specified, but this may conflict with scalability. In order to solve this problem, the PE router identifier field in the IPv6 address is partitioned into a physical PE router identifier and a route identifier, as shown in **Fig. 9**. Although the route identifier can be changed, the VFI or VPN identifier for the same VFI is not changed. This means that a single physical PE router is also assigned multiple IPv6 address prefixes. Thus, in each P router, only IP routes towards each PE router are specified and scalability is achieved by the address hierarchy.

When load balancing is required between areas, the area identifier field is also partitioned into a physical area identifier and a route identifier. In this case, the route identifier defined in the area identifier and PE router identifier are called a higher-layer route identifier and lower-layer route identifier, respectively.

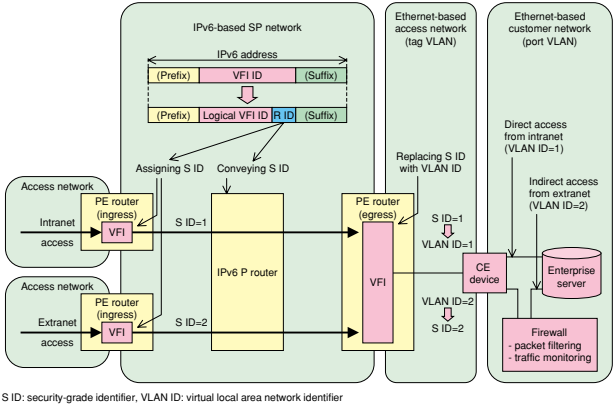
(3) Multi-grade security

The security of an IP-VPN is basically high. However, it is degraded when remote access or extranet access to the IP-VPN is allowed. To solve this problem, it is useful to have various grades of security. Multi-grade security is specified and assigned at an ingress PE router, according to the method of access, and carried to the destination customer network. In the SP network, the field of the VFI identifier in the IPv6 address is partitioned into the logical VFI identifier and the security-grade identifier, as shown in **Fig. 10**. This carries the security grade between PE routers. In an Ethernet-based access network or a customer network, the security grade is carried by using a VLAN (virtual local area network) identifier. The egress PE router converts the security grade identifier into a VLAN identifier.



R ID: route identifier, SP: service provider

Fig. 9. Load balancing scheme.



S ID: security-grade identifier, VLAN ID: virtual local area network identifier

Fig. 10. Multi-grade security scheme.

In a destination customer network, security procedures such as packet monitoring and packet filtering are applied packet-by-packet according to the notified security grade. Thus, the security of an IP-VPN can be kept high in the presence of various access methods.

4. Node design

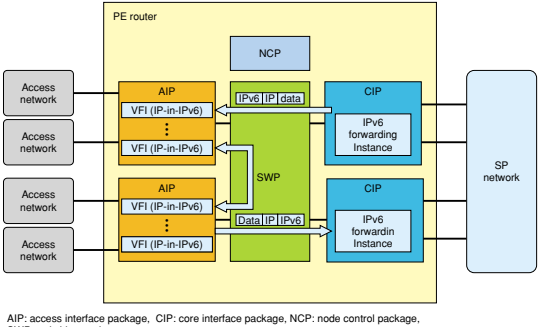
A PE router is the key device in implementing our network design, so in this section, we describe the design of a PE router that we implemented experimentally.

4.1 PE router architecture

A photograph of the router is shown in **Fig. 11** and its architecture is schematically shown in **Fig. 12**. In this architecture, a PE router is functionally composed of interface packages for access networks called AIPs (access interface packages), those for the SP network called CIPs (core interface packages), a node control package (NCP), and an internal switching package called SWP (switching package). The AIPs provide VFI functions and perform IP-in-IPv6 encapsulation or de-capsulation. The CIPs provide IPv6 forwarding functions and forward IPv6 packets in a connectionless manner. The NCP is used for node operation and control protocol processing such as generating ICMP (Internet control message protocol) echo replies [7]. The SWP is used to transfer data between these different packages.

Height: 795 mm, width: 482.6 mm, depth: 450 mm

Fig. 11. Prototype PE router.



AIP: access interface package, CIP: core interface package, NCP: node control package, SWP: switching package

Fig. 12. Functional composition of the prototype PE router.

4.2 Forwarding procedure

The forwarding procedure at an ingress PE router is as follows:

1) IP-in-IPv6 encapsulation
   When an AIP receives an IP packet sent from a CE device, it delivers the packet to the VFI associated with the VLAN identification in the Ethernet header. The VFI searches its IP-in-IPv6 encapsulation table using the destination IP address as a key and resolves the IPv6 address identifying the destination VFI. Then it encapsulates the IP packet into an IPv6 packet and searches its IPv6 forwarding table, using the resolved IPv6 address as a key, and resolves the internal tag identifying the output CIP or AIP.
2) Internal forwarding
   The SWP transfers the IPv6 packet from the AIP to a CIP or another AIP using the internal tag.
3) IPv6 forwarding
   When a CIP receives an IPv6 packet sent from an AIP, it searches its IPv6 forwarding table using the destination IPv6 address as a key and resolves the internal tag identifying the output interface. Then, it forwards the IPv6 packet to the next-hop P or PE router.

The forwarding procedure at an egress PE router is roughly the opposite:

1) IPv6 forwarding
   When a CIP receives an IPv6 packet sent from a P or PE router, it searches its IPv6 forwarding table using the destination IPv6 address as a key and resolves the internal tag identifying the output AIP.
2) Internal forwarding
   The SWP transfers the IPv6 packet from the CIP to an AIP using the internal tag.
3) IP-in-IPv6 de-capsulation
   When an AIP receives an IPv6 packet sent from a CIP or another AIP, it de-capsulates the IP packet from the IPv6 packet and searches its IP forwarding table using the destination IP address as a key, and resolves the internal tag identifying the output interface and MAC address identifying the CE device. In addition, the VLAN identifier is also resolved from the VPN or VFI identifier. Then, it forwards the IP packet to the CE device.

5. Evaluation of scalability

In a practical commercial network, it is important that a large-scale provider network can be composed of a large number of small-scale routers. However, increasing the number of PE routers also increases the number of forwarding routes such as LSP or IP routes. This may make a P router a bottleneck in route management. Thus, the number of PE routers to be accommodated by the SP network (N_{pe}) is limited. Here, we evaluate this number.

In the two-layered MPLS architecture described in Section 2, the upper-layer LSPs are established in a full-mesh topology between all PE routers. Thus, the number of uni-directional LSPs to be managed by P routers (N_{rc}) is approximately expressed by

$$N_{rc} = N_{pe} \times N_{pe}. \quad (1)$$

On the other hand, in the IPv6-based architecture proposed in Section 3, P routers forward IPv6 packets in a connectionless manner. This means that P routers can manage only IPv6 routes towards egress PE routers, without taking account of the specific ingress routers concerned. Thus, the number of IPv6 routes to be managed by P routers (N_{rp}) is

$$N_{rp} = N_{pe}. \quad (2)$$

In practical medium-scale routers, the maximum number of forwarding routes that can be managed (N_{rc} or N_{rp}) is limited to about ten thousand.

$$N_{rc} < 10{,}000 \quad (3)$$
$$N_{rp} < 10{,}000 \quad (4)$$

Therefore, in the MPLS architecture, from Eqs. (1) and (3), N_{pe} may be expressed as

$$N_{pe} < 100. \quad (5)$$

On the other hand, in the IPv6-based architecture, from Eqs. (2) and (4), N_{pe} is expressed as

$$N_{pe} < 10{,}000. \quad (6)$$

These results show that the IPv6-based architecture is superior to the MPLS-based architecture from the viewpoint of scalability, and only the IPv6-based architecture achieves our design objectives of accommodating a thousand PE routers in an SP network. In addition, the difference in the number of PE routers that can be accommodated by the SP network (N_{pe}) between the MPLS-based and IPv6-based architectures further increases as the number of forwarding routes that can be managed by P routers increases.

6. Conclusion

Our IP-in-IPv6 overlay networking for a terabit-class super-network was designed not only to improve scalability, but also to extend the value-added functions of high-performance networks. While conventional MPLS-based technology can handle only one hundred provider-edge routers because reachability is retained through the use of full-mesh paths, our IPv6-based technology can handle a thousand provider-edge routers because reachability is retained in a connectionless manner. Moreover, as value-added functions, our technology enables virtual private networking, load balancing, and multi-grade security by address management. All these functions play important roles in cost-effectively improving the reliability and security of IP-VPN services. Therefore, our technology is attractive for commercial service providers who want to offer value-added IP-VPN services.

7. Acknowledgment

This research was supported by a grant from the Telecommunications Advancement Organization of Japan (TAO).

References

[1] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol label switching architecture," IETF RFC3031, Jan. 2001.
[2] S. Deering and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," IETF RFC2460, Dec. 1998.
[3] R. Callon and M. Suzuki, "A framework for layer 3 provider provisioned virtual private networks," IETF Internet-draft <draft-ietf-l3vpn-framework-00.txt>, Mar. 2003.
[4] E. Rosen and Y. Rekhter, "BGP/MPLS IP VPNs," IETF Internet-draft <draft-ietf-l3vpn-rfc2547bis-01.txt>, May 2003.
[5] A. Conta and S. Deering, "Generic packet tunneling in IPv6 specification," IETF RFC 2473, Dec. 1998.
[6] K. Matsui, T. Yagi, Y. Naruse, and J. Murayama, "Cut-through Optical Path Control Technology for a Terabit-class Super-network," NTT Technical Review, Vol. 2, No. 3, pp. 32-40, 2004.
[7] A. Conta and S. Deering, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification," IETF RFC2463, Dec. 1998.

Yuuichi Naruse
Research Engineer, Secure Communication Project, NTT Information Sharing Platform Laboratories.
He received the B.S. and M.S. degrees in physics from Tohoku University, Sendai, Miyagi in 1990 and 1992, respectively. He joined NTT in 1992. He has been engaged in R&D of ATM-LAN systems, Fibre-Channel network systems, and IP-VPN service platforms.

Takeshi Yagi
Secure Communication Project, NTT Information Sharing Platform Laboratories.
He received the B.E. degree in electrical and electronic engineering and the M.S. degree in science and technology from Chiba University, Chiba in 2000 and 2002, respectively. He joined NTT in 2002. His studies focus on IP-VPN architecture and his current research interests include IP routing and forwarding technology, traffic monitoring technology, and layer cooperation technology. He is a member of the Institute of Electrical Engineers of Japan (IEEJ) and the Institute of Electronics, Information and Communication Engineers of Japan (IEICE).

Kenichi Matsui
Secure Communication Project, NTT Information Sharing Platform Laboratories.
He received the B.E. degree in information engineering and the M.S. degree in information sciences from Tohoku University, Sendai, Miyagi in 1995 and 1997, respectively. He joined NTT in 1997. His work focuses on IP networking and his research interests include traffic engineering for optical IP networks and MPLS, on-demand QoS management, and managed IP multicast platforms. He is a member of IEICE, the Information Processing Society of Japan, and the IEEE Computer Society.

Junichi Murayama
Senior Research Engineer, Secure Communication Project, NTT Information Sharing Platform Laboratories.
He received the B.E. and M. E. degrees in electronics and communication engineering from Waseda University, Tokyo in 1989 and 1991, respectively. Since joining NTT in 1991, he has been engaged in R&D of ATM networks, large-scale IP networks, and IP-VPN service platforms. He is a member of IEICE.