INVITED PAPER *Special Section on the Architectures, Protocols, and Applications for the Future Internet*

New Directions for a Japanese Academic Backbone Network

Shigeo URUSHIDANI[†a)], **Shunji ABE**[†], **Kenjiro YAMANAKA**[†], **Kento AIDA**[†], **Shigetoshi YOKOYAMA**[†], **Hiroshi YAMADA**[†], **Motonori NAKAMURA**[†], **Kensuke FUKUDA**[†], **Michihiro KOIBUCHI**[†], ***and*** **Shigeki YAMADA**[†], ***Members***

SUMMARY This paper describes an architectural design and related services of a new Japanese academic backbone network, called SINET5, which will be launched in April 2016. The network will cover all 47 prefectures with 100-Gigabit Ethernet technology and connect each pair of prefectures with a minimized latency. This will enable users to leverage evolving cloud-computing powers as well as draw on a high-performance platform for data-intensive applications. The transmission layer will form a fully meshed, SDN-friendly, and reliable network. The services will evolve to be more dynamic and cloud-oriented in response to user demands. Cyber-security measures for the backbone network and tools for performance acceleration and visualization are also discussed.

key words: *100-Gigabit Ethernet, MPLS-TP, SDN, multi-layer network, security, cloud service, performance accelerator, network monitoring*

1. Introduction

National research and education networks (NRENs) [1]–[10] have expanded their backbone speeds and network service capabilities in parallel with the evolution of computer systems and experimental devices in many research fields. Starting with sharing super-computers, high-speed connectivity has brought about great changes in the style of research, especially in big-science fields. For example, the networks made it possible to distribute and share huge amounts of data generated from CERN's Large Hadron Collider (LHC) [11] among research centers around the world for analysis; enabled astrometry and geodesy to use very long baseline interferometry (VLBI) technology [12], [13] by freely selecting radio telescopes around the world; and urged visual researchers to produce non-compressed bidirectional ultra HDTV devices for highly realistic communication [14]. New projects, such as Belle II [15], ITER [16], VLBI2010 [17], will also need worldwide high-speed networks. Many NRENs therefore have begun to deploy 100-Gigabit Ethernet (100GE) technology since 2011, and a trans-Atlantic line was upgraded to 100 Gbps in June 2013 [18].

The NRENs also offer a range of network services from layer 1 to layer 3. In addition to ultra-high-speed IP services, virtual private network (VPN) services of layers 2 and 3 are offered for both domestic and international projects [19]. On-demand networking and software-defined-networking (SDN) technologies [20] are expected to enhance network service capabilities, so the NRENs and network testbeds are actively investigating them [21]–[23]. Security measures for the backbone networks have also been discussed and tested [24].

The NRENs are also involved in the development and acquisition of cloud services to meet the needs of academia. New programs for cloud services between industry and academia, such as Interenet2 NET+ [25], SURFconext [26], and JANET cloud service [27], have been expanding their service menu and the number of users. The programs not only define the frameworks for business models, authentication, security agreements, and service levels but also leverage the buying power of the academic community to reduce the costs.

Performance issues are of great concern for international research projects as well as domestic projects. High-performance protocols, such as GridFTP [28], and performance monitoring tools, such as PerfSONAR [29], have been discussed and developed in the NRENs in cooperation with their users in order to address end-to-end performance problems and offer secure network platforms. Bandwidth challenges using international lines [30] are also popular in the NRENs. Expanding 100-Gbps network areas, however, have brought about new challenges to enhance the performance.

The Science Information NETwork (SINET) [31], [32] operated by the National Institute of Informatics (NII) has offered a high-performance multi-layer network platform for Japanese academic communities, supported international projects with its international lines, and promoted secure private cloud services in collaboration with cloud service providers [33]. On-demand layer-1&2 services have dynamically offered end-to-end bandwidth for many projects and created international on-demand VPNs [21]. However, current 40-Gbps (STM-256) lines have become insufficient to handle the increasing traffic with high performance. SINET has also been asked to do more: offer attractive SDN services, promote a wider range of cloud services, address security issues, enhance and monitor end-to-end performance, and so on.

This paper describes an architectural design of a new SINET, called SINET5, which will be launched in April 2016. Please note that this design might be slightly changed due to budgetary conditions and as a result of a series of procurements. The remainder of this paper is organized

Manuscript received October 20, 2014.
Manuscript publicized December 11, 2014.

[†]The authors are with the National Institute of Informatics, Tokyo, 101–8430 Japan.
a) E-mail: urushi@nii.ac.jp

DOI: 10.1587/transinf.2014NTI0001

as follows. Section 2 briefly overviews the current version of SINET (SINET4) and clarifies the requirements for SINET5. Section 3 gives a new network design, which covers all 47 prefectures with 100-Gbps or more bandwidth and gives a minimized latency and high-reliability between each pair of prefectures. Section 4 describes on-demand and SDN service capabilities. Section 5 touches on security issues on the backbone network. Section 6 describes approaches to promote cloud services tailored to meet academic requirements. Section 7 introduces performance enhancement and monitoring tools. Finally, Sect. 8 presents the conclusion.

2. Requirements for Transformation

As of March 2014, the current SINET4 serves as a backbone network for more than 800 organizations: 86 national, 71 public, and 333 private universities; 60 junior colleges; 55 colleges of technology; 16 inter-university research institutions; and 184 other research institutions. The reach of SINET4 was gradually expanded from 34 to all 47 Japanese prefectures during 2011 and 2012. The network is comprised of 42 edge and eight core nodes, by which the users can access SINET4 with the bandwidth of 2.4 Gbps to 40 Gbps (Fig. 1). Each edge node is connected to the nearest core node, and the core nodes are interconnected with redundant routes. Every pair of the nodes is connected by a duplicated leased line, and every node is placed in a data center that is quake-resistant even to earthquakes of seismic intensity 7.0 and has a power supply capacity for at least 10 hours in case of blackouts. Thanks to this design, SINET4 survived the Great East Japan Earthquake in 2011 and continued to offer a range of multi-layer services [32]. However, this leased line approach did not allow SINET4 to inexpensively upgrade the line speed in order to deliver ever-increasing traffic but did increase the latency due to carrier-class hitless-switching capability. A network design using dark fibers and user-owned transmission devices for flexible line speed upgrade and smaller latency is therefore strongly desired for SINET5. Given the expecting traffic generated from existing and emerging applications and cloud services, each line speed of SINET5 needs to be 100 Gbps or more, and higher-speed interfaces, such as 400-Gigabit and 1-Terabit Ethernet interfaces, should be flexibly introduced when needed. Access line speeds of many user organizations to SINET5 nodes also need to be upgraded to 40 Gbps or 100 Gbps inexpensively. This will need a joint procurement in collaboration with these user organizations.

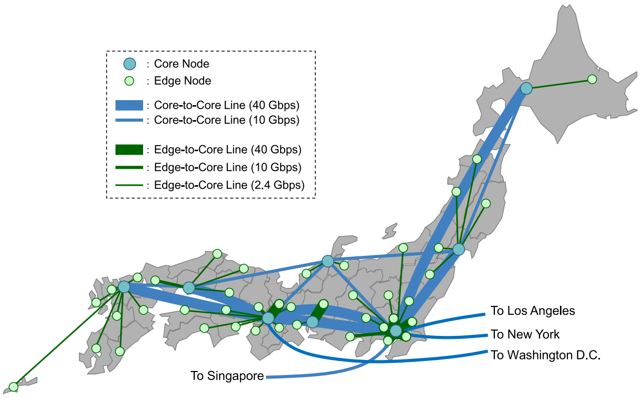

: Core Node
: Edge Node
: Core-to-Core Line (40 Gbps)
: Core-to-Core Line (10 Gbps)
: Edge-to-Core Line (40 Gbps)
: Edge-to-Core Line (10 Gbps)
: Edge-to-Core Line (2.4 Gbps)
To Los Angeles
To New York
To Washington D.C.
To Singapore

Fig. 1 Network topology of SINET4.

For joint research between different organizations, high-performance VPN services, such as layer-3 VPN (L3VPN) [34], layer-2 (L2VPN) [35], and virtual private LAN service (VPLS) [36], have become popular, and the number of VPN sites continues to rise and now exceeds 600. The layer-2-based VPN services are also used for secure private cloud services as well as campus LANs and need to be more dynamically established along with evolving on-demand cloud services. In addition, we are considering combining virtual extensible LANs (VXLANs) [37] with VLANs in order to support multi-provider cloud services.

For the Internet services, growing concerns about cyber-attacks, such as unauthorized access and DDoS attacks, have seriously affected research and educational activities. SINET4 is not a primary defense zone but has been asked to consider measures against these security issues with reference to Defense in Depth strategy [38].

While cloud services over SINET4 expanded academia's interest, their specifications have become quite complex as the number of service providers has increased. That makes it difficult for users to select appropriate cloud services depending on their purpose and policy. For example, while performance issues, such as computer input/output and network speeds, and security issues, such as share and encryption policies, are of great interest for cloud storage services, locations of data centers and laws applicable to their cloud data center operation must be considered for confidential data storage. A new framework similar to those of other NERNs [25]–[27] is therefore necessary for Japanese academia to comfortably use cloud services over SINET5.

SINET users always need high-performance data transfer protocols, because Japan is far from the world's central experimental sites and increasing network bandwidth does not directly result in a great improvement in end-to-end performance in such an environment. A new approach for performance acceleration has been desired that is scalable up to a data rate of 100 Gbps or more. In addition, SINET operators need to solve performance problems in cooperation with the users if they suffer from unexplained problems. Performance monitoring tools are therefore essential to see the problems and enable information about them to be shared between the users and SINET operators.

3. Network Design of SINET5

This section describes the network design of SINET5 to meet the requirements described above.

3.1 High-Level Network Architecture

The entire network architecture pictured between users and shared research and cloud resources is comprised of four layers: optical fiber, transmission, network service, and cooperation layers (Fig. 2). Access lines between user organizations and SINET5 are owned by users, but optical fibers for many access lines will be jointly procured under NII's initiative. The backbone network will flexibly expand its line capacity by using optical fibers and transmission devices; create small-latency connection patterns between SINET nodes by wavelength division multiplexing (WDM) and multi-protocol label switching - transport profile (MPLS-TP) paths; offer multi-layer network services by using SINET nodes and transmission devices; enable cooperation between users and SINET5 through an SDN controller and a monitoring tool; and cooperate with cloud and research facilities via application programming interfaces (APIs). SINET5 will also enhance users' data transfer performance by offering them a performance acceleration protocol that is installed in both user and shared facilities.

3.2 Optical Fiber and Transmission Layers

SINET5 will place SINET5 nodes, which receive access lines of user organizations and offer a range of layer-2&3 services, in 50 data centers and will connect them with dark fibers from Hokkaido to Kyushu (Fig. 3). Here SINET5 needs to use leased lines for the Okinawa line as well as international lines to USA, Europe, and East Asia. While SINET4 uses dark fibers for access lines, SINET5 will introduce inter-prefectural dark fibers between neighboring data centers nationwide. The dark fibers will connect data centers so as to form redundant routes between them, which will strengthen the entire network's reliability. Each SINET node will be connected by at least two dark fibers, and the Okinawa node will be connected by two leased lines, although they are not clearly shown in Fig. 3.

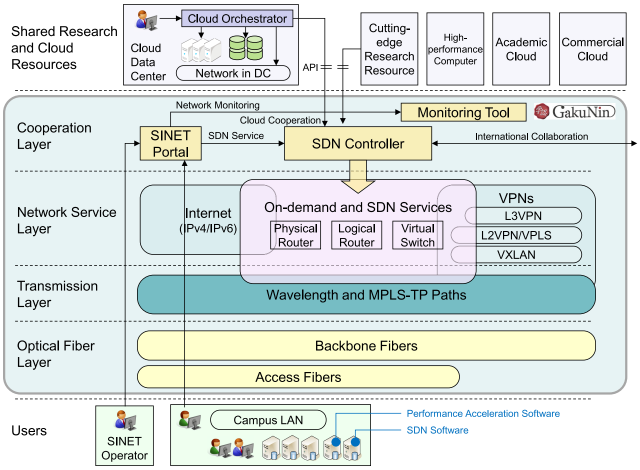


Fig. 2 High-level network architecture of SINET5.

A transmission device at each data center will be composed of a reconfigurable optical add/drop multiplexer (ROADM) and an MPLS-TP device (Fig. 4). The ROADMs will connect adjacent data centers with "adjacent" wavelength paths and distant data centers with "cut-through" wavelength paths. The bandwidth of each wavelength path will be 100 Gbps in the beginning of the operation, which will lead to a nationwide 100-Gbps backbone network covering all the prefectures. More bandwidth will be made available when needed by using new interfaces such as 400-Gigabit Ethernet interfaces. The MPLS-TP devices will establish two MPLS-TP paths between each pair of data centers. One will be a primary path, which will be established in principle on the smallest-latency route between them, and the other will be the secondary path, which will be set up on the disjoint route to the primary path. Each node will access every peer with the minimized latency and use the backup route in the case of a minimized-latency route failure. The users will therefore be able to usually obtain a maximum performance environment over SINET5.

SINET5 will consequently connect all the nodes in a fully meshed topology with minimized latency, while SINET4 connects nodes in a star-like topology (Fig. 5). This topology will also be SDN-friendly, because only edge nodes need to be configured for SDN services. Note that SINET5 enhances the network reliability by using sec-

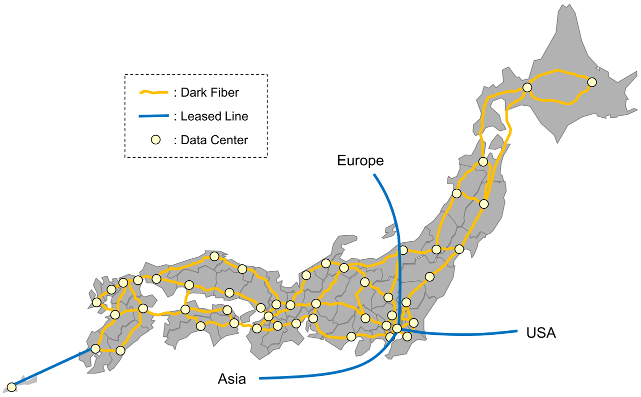


Fig. 3 Fiber-level topology image of SINET5.

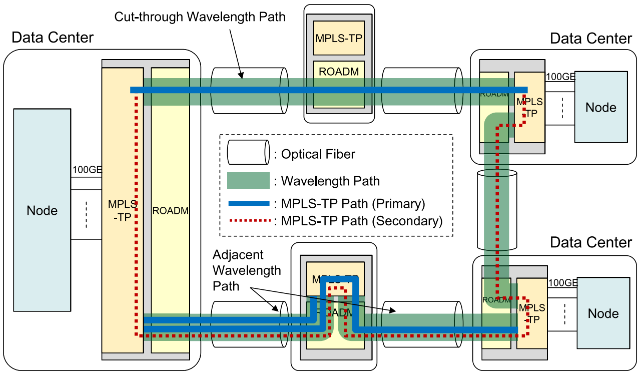
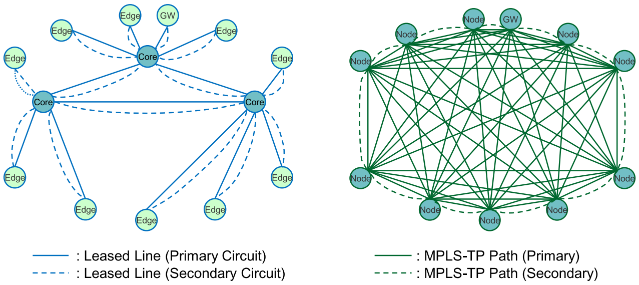


Fig. 4 Connections by wavelength and MPLS-TP paths.


Edge
GW
Core
Node
: Leased Line (Primary Circuit)
: Leased Line (Secondary Circuit)
: MPLS-TP Path (Primary)
: MPLS-TP Path (Secondary)

(a) SINET4

(b) SINET5

Fig. 5 Network topology comparison.

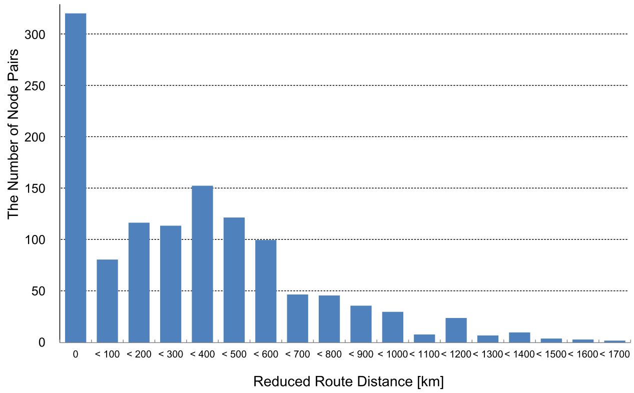


Fig. 6 Route distance reduction effect by minimized-latency routes.

ondary MPLS-TP paths that do not consume transmission resources, while SINET4 needs dedicated resources for secondary circuits of leased lines. Here, not all the secondary MPLS-TP paths are shown in the figure.

We evaluated how much the latency between SINET nodes will be reduced by the minimized-latency routes compared with those of SINET4 (Fig. 6). The total 1225 node pairs were evaluated in terms of the route distance. SINET5 will reduce the route distances between the nodes by an average of 333 km, which we believe will greatly enhance the end-to-end performance.

3.3 Network Service Functions

Each data center will have an IP router as the SINET5 node, which receives the access lines and offers a range of layer-2&3 services. Layer-1 services will also be able to be offered as wavelength services by transmission devices, although the users will have to pay for them. Each IP router will be connected to an MPLS-TP device with two or four 100GE interfaces for high availability.

An IP router at each data center will initially have five logical routers corresponding to five network services: the Internet (IPv4/IPv6), L3VPN, L2VPN and VPLS, layer-2 on-demand (L2OD), and SDN services (Fig. 7). Two logical routers for the same service between each pair of IP routers will be connected by two VLANs in order to use two interfaces of each IP router for load balancing. Note that only two VLANs are used for the two logical routers, and four interfaces of an IP router are used on a round-robin basis depending on the destinations. The two VLANs are set up through the same corresponding MPLS-TP path. Ten VLANs will therefore be necessary for each pair of IP routers and will be set up through the same corresponding MPLS-TP path. Here the total number of VLAN-IDs will be reduced by using VLAN-ID conversion capability of MPLS-TP devices. When a VLAN-ID is expressed as $L_3L_2L_1L_0$, L_1L_0 (≥ 1) is used for the IP router number, L_2 (≥ 1) is for the logical router number, and L_3 (≥ 0) is for load balancing. As SINET5 will have gateway routers at Tokyo and Osaka, the total number of VLAN-IDs will be 520 (= 52 (L_1L_0) × 5 (L_2) × 2 (L_3)). Note that these VLAN-IDs are assigned only in the backbone and do not affect the number of VLANs used in SINET layer-2 services. We call these VLANs backbone VLANs hereinafter.

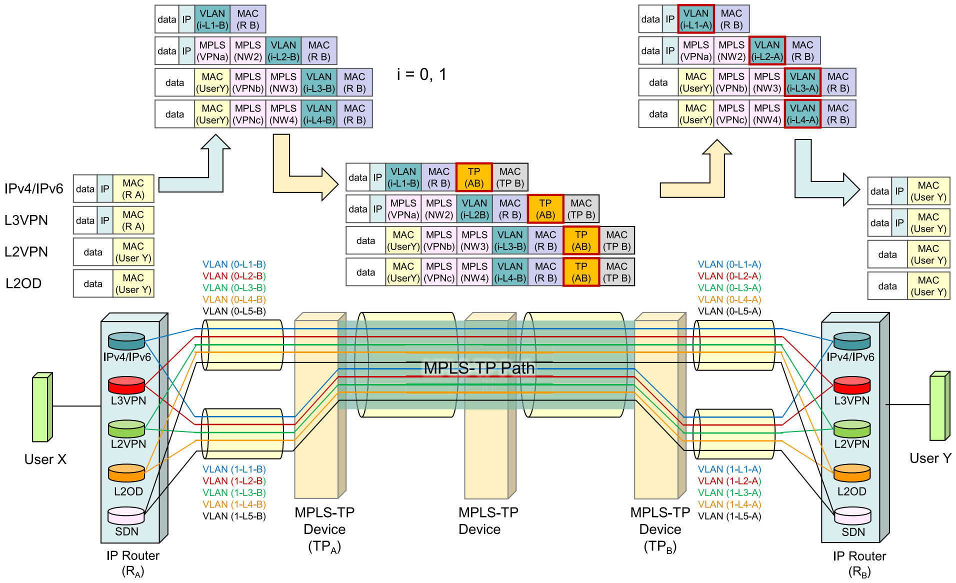
For further clarification of the end-to-end networking, the transitions of packet headers in four services are described as follows with reference to Fig. 7. That of the SDN service will be clarified in the near future.

Internet service: An IP packet received at IP router R_A is attached with the MAC address of the opposite IP router R_B and a backbone VLAN-ID whose L_1L_0 = the number of R_B and $L_2 = 1$, and loaded on one of the interfaces depending on whether $L_3 = 0$ or 1. The packet is received at the MPLS-TP device TP_A, attached with the MAC address of the opposite MPLS-TP device TP_B and MPLS-TP labels assigned for the MPLS-TP path between the TP_A and the TP_B, forwarded to the TP_B in accordance with the MPLS-TP labels, detached from MPLS-TP-related headers at the TP_B, and passed to the R_B with a converted VLAN-ID whose L_1L_0 = the number of R_A. The packet is detached from the VLAN-ID at the R_B and forwarded to the destination.

L3VPN service: An IP packet received at the R_A is encapsulated with MPLS labels [34], attached with the MAC address of R_B and a backbone VLAN-ID whose L_1L_0 = the number of R_B and $L_2 = 2$, and loaded on one of the interfaces depending on whether $L_3 = 0$ or 1. The packet is received at the TP_A, attached with the MAC address of TP_B and MPLS-TP labels that are the same as those of the Internet service, forwarded to the TP_B in accordance with the labels, detached from MPLS-TP-related headers, and passed to the R_B with the converted VLAN-ID. The packet is detached from the VLAN-ID and the MPLS labels at the R_B and forwarded to the destination.

L2VPN and VPLS services: An Ethernet frame received at the R_A is encapsulated with MPLS labels [35], [36], attached with the MAC address of R_B and a backbone VLAN-ID whose L_1L_0 = the number of R_B and $L_2 = 3$, and loaded on one of the interfaces depending on whether $L_3 = 0$ or 1. The difference between L2VPN and VPLS services is that the latter identifies the MAC addresses of original Ethernet frames while the former does not. The following treatments at MPLS-TP devices and IP routers are similar to those of the L3VPN services.

L2OD service: The current networking process of this



data | IP | VLAN (i-L1-B) | MAC (R B)
data | IP | MPLS (VPNa) | MPLS (NW2) | VLAN (i-L2-B) | MAC (R B)
data | MAC (UserY) | MPLS (VPNb) | MPLS (NW3) | VLAN (i-L3-B) | MAC (R B)
data | MAC (UserY) | MPLS (VPNc) | MPLS (NW4) | VLAN (i-L4-B) | MAC (R B)
i = 0, 1
data | IP | VLAN (i-L1-A) | MAC (R B)
data | IP | MPLS (VPNa) | MPLS (NW2) | VLAN (i-L2-A) | MAC (R B)
data | MAC (UserY) | MPLS (VPNb) | MPLS (NW3) | VLAN (i-L3-A) | MAC (R B)
data | MAC (UserY) | MPLS (VPNc) | MPLS (NW4) | VLAN (i-L4-A) | MAC (R B)
IPv4/IPv6 | data | IP | MAC (R A)
L3VPN | data | IP | MAC (R A)
L2VPN | data | MAC (User Y)
L2OD | data | MAC (User Y)
data | IP | VLAN (i-L1-B) | MAC (R B) | TP (AB) | MAC (TP B)
data | IP | MPLS (VPNa) | MPLS (NW2) | VLAN (i-L2B) | MAC (R B) | TP (AB) | MAC (TP B)
data | MAC (UserY) | MPLS (VPNb) | MPLS (NW3) | VLAN (i-L3-B) | MAC (R B) | TP (AB) | MAC (TP B)
data | MAC (UserY) | MPLS (VPNc) | MPLS (NW4) | VLAN (i-L4-B) | MAC (R B) | TP (AB) | MAC (TP B)
data | IP | MAC (User Y)
data | IP | MAC (User Y)
data | MAC (User Y)
data | MAC (User Y)
VLAN (0-L1-B)
VLAN (0-L2-B)
VLAN (0-L3-B)
VLAN (0-L4-B)
VLAN (0-L5-B)
VLAN (0-L1-A)
VLAN (0-L2-A)
VLAN (0-L3-A)
VLAN (0-L4-A)
VLAN (0-L5-A)
MPLS-TP Path
IPv4/IPv6
L3VPN
L2VPN
L2OD
SDN
User X
User Y
VLAN (1-L1-B)
VLAN (1-L2-B)
VLAN (1-L3-B)
VLAN (1-L4-B)
VLAN (1-L5-B)
VLAN (1-L1-A)
VLAN (1-L2-A)
VLAN (1-L3-A)
VLAN (1-L4-A)
VLAN (1-L5-A)
MPLS-TP Device (TP_A)
MPLS-TP Device
MPLS-TP Device (TP_B)
IP Router (R_A)
IP Router (R_B)

Fig. 7 Transitions of packet headers in services.

service is basically the same as that of the VPLS service. The difference is that the configurations for L2OD service are dynamically set up by the SDN controller in response to user requests.

3.4 Reliability and QoS Control Functions

Lessons from the experiences of the Great East Japan Earthquake in March 2011 have prompted higher reliability toward SINET5. The optical fiber layer will have redundant routes for all the data centers, the transmission layer will form disjointed MPLS-TP paths between every pair of data centers, and each IP router will be connected to an MPLS-TP device with two or more interfaces at the same location. If one of the interfaces of an IP router or an MPLS-TP device is down, the IP router transfers all of the packets to the remaining interfaces. If the primary MPLS-TP path is down between a pair of IP routers, MPLS-TP devices switch from the primary MPLS-TP path to the secondary MPLS-TP path within 50 msec and receive all ten VLANs for five logical routers in the secondary MPLS-TP path without IP routers' awareness. If both the primary and secondary MPLS-TP paths between the pair are down, the IP routers detect the down by the bidirectional forwarding detection (BFD) protocol, find the other routes by using OSPF for the Internet service, and establish the other MPLS paths by using the fast reroute (FRR) [39] for VPN services.

As for quality of service (QoS) capabilities, IP routers and MPLS-TP devices will both have four queues and perform QoS control in accordance with the similar packet queueing and discarding algorithms to those of SINET4 [32].

We tested some combinations of real IP routers and MPLS-TP devices and confirmed that there were no problems in terms of general interoperability, failure recovery, or QoS control functions.

4. On-Demand and SDN Services

This section describes on-demand services that will be passed to SINET5 and new services that will be offered as SDN services.

4.1 Extended On-Demand Services

SINET4 has offered layer-1 and layer-2 on-demand services similar to SDN services [32]. The layer-1 on-demand service receives user requests regarding destinations, durations, and path bandwidths via the portal system; calculates the best routes on a minimal-latency or maximum-bandwidth basis; manages the entire bandwidth assignment; and establishes the end-to-end paths. The availability of an end-to-end path varies depending on the utilization of layer-2/3 services on candidate routes, and the bandwidth is sometimes gathered from two different routes. More than 1,000 layer-1 paths each whose duration was less than a week were set up and released so far. However, this layer-1 path setup needs more than two hours for the path bandwidth of 8.4 Gbps, which was increased from 2.4 Gbps in 2011 for VLBI projects, and has become inconvenient recently.

We therefore decided to offer only the L2OD service in SINET5, which has similar functions for bandwidth assignment on layer-2 paths, and expand the maximum available bandwidth for this service to several tens of Gbps. Although the layer-2 on-demand service cannot offer a complete communication environment, i.e. no packet loss or jitter, which the layer-1 service offers, it will offer virtually no packet loss by high-priority transfer control and very small jitters by utilizing end-to-end 100-Gbps routes available nationwide. Only two IP routers need to be dynamically configured for this purpose thanks to the fully meshed topology of SINET5, and this will lead to paths being easily and promptly established between any locations. The bursty traffic is moderately spaced within the specified bandwidth in the ingress IP router so as to reduce the jitter over the transit MPLS-TP devices. The on-demand service functions will be ported into SINET5's SDN controller, although this service uses the NETCONF interface [40] between the controller and SINET nodes.

4.2 New Services for Campus LANs in Cloud Era

Evolving cloud services have been accelerating the relocation of research and education resources from on-the-premise to remote cloud data centers over SINET. While the nationwide 100-Gbps backbone network will accelerate this trend more by riding regional disparities for the services, manually configuring the increasing number of VPNs will delay service delivery as well as increase the workload of SINET operators. We therefore plan to launch a virtual campus LAN service that allows campus LAN operators to freely expand their LAN areas over SINET5 in order to easily and flexibly utilize the cloud services (Fig. 8).

For this purpose, virtual switch instances, which work as virtually dedicated Ethernet switches for campus LANs, will be configured in the SINET5 nodes. These instances are separated from other service instances and do not limit the number of VLANs between the campuses and cloud data centers over SINET5, although each SINET node might have a scale limit. The virtual switch instances will also support VXLAN tunnel end point (VTEP) [37] functions for VXLANs and combine VXLANs with VLANs, which will enable the users to use a greater variety of cloud services between different data centers. The campus LAN operators require configuration changes via the SINET portal, and the SDN controller, which converts the requirements into the real configurations, controls the virtual switch instances though an SDN interface. We also plan to offer APIs of the SDN controller through which orchestrators of cloud services and other systems can control the dedicated virtual switch instances.

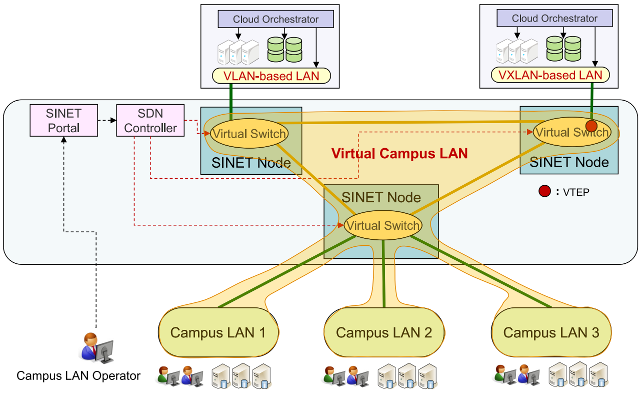


Fig. 8 New service using virtual switch instances in SINET5.

5. Security Measures

SINET5 plans to build a first-defense zone in the Internet services in cooperation with user organizations. The first-step measures should be DDoS detection and mitigation, detection of other possible threats, and prompt informing of them.

5.1 DDoS Attack Measures

SINET4 currently detects DDoS attacks or anomaly traffic behaviors by collecting and analyzing network flows, called NetFlows [41], from SINET routers. If we detect DDoS attacks, we mitigate them by black hole routing [42] with the approval of user organizations (Fig. 9). SINET5 must use more scalable collecting and analyzing tools because the data sampling points will be expanded due to the fully meshed topology and the sampling rate might be increased when possible for deeper analysis. In addition, we are also considering a measure that more effectively mitigates target anomaly traffic at all the SINET routers by propagating the attack source information and filtering parameters by BGP Flowspec [43], functions of which GÉANT has been verifying toward introduction [24].

5.2 Detection of Other Threats

Because major cyber-attacks come from the commercial Internet, the traffic flows of the border interfaces should be carefully observed. SINET4 is currently connected to the commercial Internet via major exchange points, JPIX and JPNAP, and tier-1 ISPs in Tokyo and Osaka, and the current total interface capacity is 61 Gbps. We monitored one of the interfaces as a trial by using a commercial intrusion detection device in July and August 2014, when we detected a maximum of over 4,000 pieces of zero-day malware per day at the interface. As the total interface capacity of SINET5 will be expanded by partly using 100GE interfaces, we must apply more scalable monitoring devices to identify security threats from huge amounts of traffic over the interfaces. The mirrored traffic from the interfaces might be sliced so as to be analyzable in multiple security devices on a real-time basis.

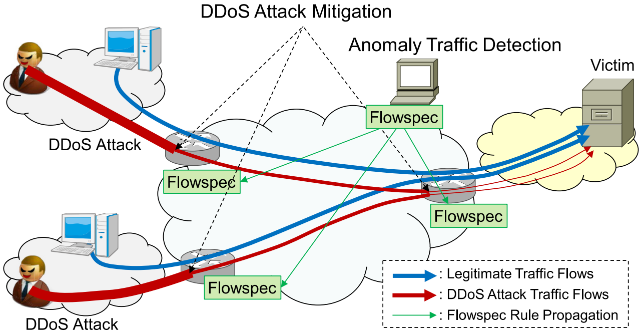

DDoS Attack Mitigation
Anomaly Traffic Detection
Victim
Flowspec
DDoS Attack
Flowspec
Flowspec
Flowspec
DDoS Attack
: Legitimate Traffic Flows
: DDoS Attack Traffic Flows
: Flowspec Rule Propagation

Fig. 9 Anomaly traffic detection and DDoS attack mitigation.

5.3 Informing and Sharing of Threats

Some NRENs have their frameworks for information sharing about cyber-security problems among trusted partners, called academic computer security incident response teams (CSIRTs). For example, REN-ISAC [44] plays an important role for Internet2 participants, as does NSHARP [45] for GÉANT's partners. A similar cooperation framework between SINET5 and its user organizations is necessary for prompt and exact informing and sharing of the security threat information.

6. Promoting Cloud-Based Services

This section describes a new framework to promote cloud services and an inter-cloud service concept for scalable and reliable cloud services.

6.1 Cloud Gateway

With increasing cloud services, we are formulating a checklist to select appropriate cloud services depending on users' purpose and policy with reference to existing documents [25], [46]. We also plan to release the evaluation results and classification on a wide range of available cloud services in accordance with the check list. This release will help the users to easily develop cloud service specifications and enable joint procurements for the same cloud services, which will lead to dramatic cost reduction in academia as a whole. In addition, we plan to develop a cloud portal system, called Cloud Gateway (Fig. 10), through which the users can use the cloud services with which their organizations have contracted. This portal system will show cloud services in a menu-driven style and enable the users to use the services on a single sign-on basis through the academic access management federation, called GakuNin [47].

6.2 Inter-Cloud Environment Service

Inter-cloud technology, which enables distributed cloud resources to be used as a virtually combined resource over networks, has been developed in order to enhance the availability and to ensure the computing performance [48]. As researchers involved in joint research often share computer resources and data between different organizations, this inter-cloud approach looks promising. However, it is not easy for the researchers to realize the inter-cloud environment, because they need to have high levels of operation skills in both computer and network technologies. We therefore plan to develop a new service through which users can create inter-cloud environments over SINET5 on demand. The inter-cloud environment, which comprises individual distributed cloud resources including related software and a virtual network to connect their resources, will be created on demand when a user selects the elements through the portal system (Fig. 10). The virtual network, based on VLANs or VXLANs, will be established through the APIs described in Sect. 4.2, and the software platform will be easily built by using container technology [49].

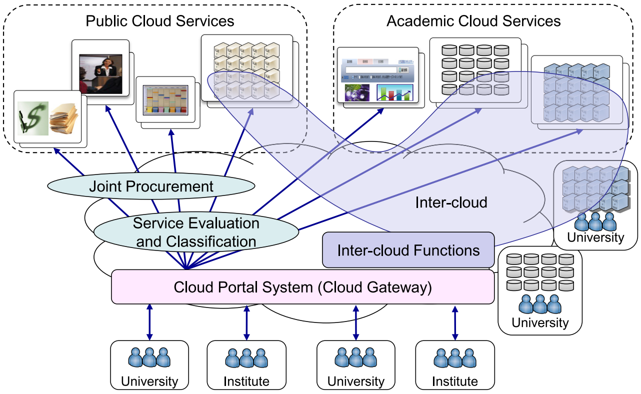


Fig. 10 Cloud service promotion using Cloud Gateway.

7. Enhancing End-to-End Performance

This section describes our original tools to enhance end-to-end performance in a long distance environment and to visualize the performance on a real-time basis.

7.1 Advanced File Transfer Protocol

As the transmission control protocol (TCP) suffers from performance degradation in a long distance environment, there are three types of approaches to address the issue: using substitute protocols [50], [51], improving the TCP protocol [52]–[54], and handling multiple TCP connections [55], [56]. The last type enhances the performance by increasing the number of connections and is primarily used in international projects. The existing approaches, however, gradually degrade the performance after attaining the highest performance with an increasing number of TCP connections. This is because increased TCP connections begin to interfere with each other and cause the congestion. The number of TCP connections for the highest performance depends on the bandwidth and the distance between two locations. As joint research projects usually permit the same number of TCP connections for each partner on a fair basis, not all the partners obtain sufficient performance under the circumstance.

We therefore started developing a new high-performance protocol based on TCP, called massively multi-channel file transfer protocol (MMCFTP), which specifies the speed instead of the number of TCP connections and enhances the performance irrespective of the distance between two locations [57]. This protocol dynamically changes the number of TCP connections depending on the observed latency and packet loss and keeps trying to obtain the specified speed. We verified that this protocol works very well even in an international environment, such as between an Amazon Web Services (AWS) data center in Dublin and

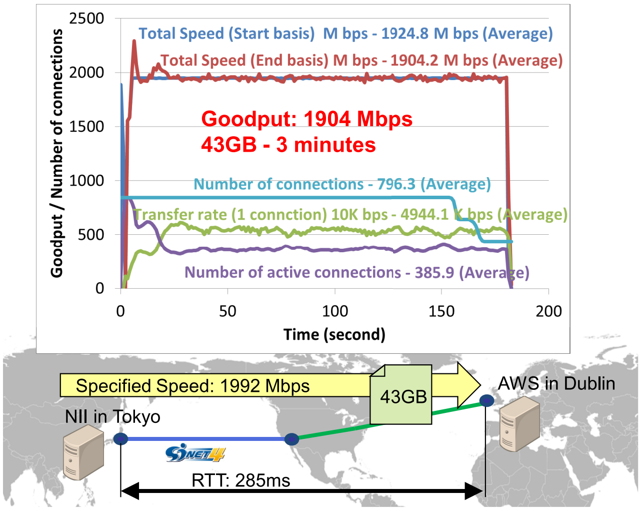
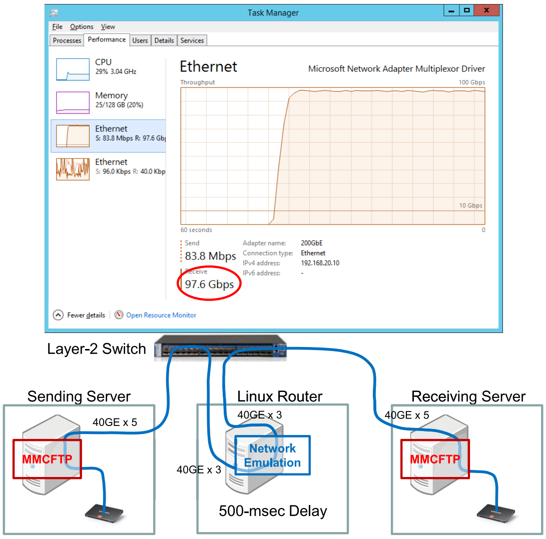


Fig. 11 Performance evaluation of MMDFTP between AWS and NII.



Send
83.8 Mbps
Receive
97.6 Gbps

Adapter name: 200GbE
Connection type: Ethernet
IPv4 address: 192.168.20.10
IPv6 address: -

Fig. 12 Experiment toward 100-Gbps throughput.

NII in Tokyo, whose RTT is about 285 msec (Fig. 11). A 43-GB file was transferred between them by specifying the speed of 1.992 Gbps, because we found the access speed to the AWS data center was actually limited to 2 Gbps. We confirmed the file was transferred in about 3 minutes via SINET4 and the commercial Internet, and the average goodput (i.e. the average application-level throughput excluding protocol overhead) was 1.904 Gbps. As requested, we started the trial distribution of this protocol software to users with the limited maximum speed of 5 Gbps.

We are now improving this protocol to be adaptable to 100-Gbps networks. As this protocol has another advantage (it can work very well even under link aggregation thanks to using many TCP connections), we first tested its performance through multiple 40-Gigabit Ethernet (40GE) interfaces (Fig. 12). A large-delay pseudo network was built by a layer-2 switch and a Linux router. Sending and receiving servers were connected to the switch with five 40GE interfaces by taking into account the performance limit of network interface cards. The Linux router was connected to the switch with two three-40GE-interfaces and gave 500-msec delay by using network emulation. When we specified the transfer speed of 98 Gbps, the gained goodput and throughput were 93.7 Gbps and 97.6 Gbps, respectively. As this result was obtained in a memory-to-memory transfer environment, the performance for a large file should be evaluated along with high-speed disk devices.

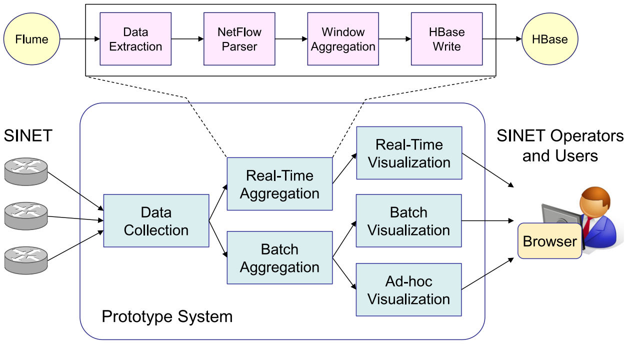


Fig. 13 Prototype system for performance monitoring.

7.2 Performance Monitoring

SINET operators need to solve performance problems in cooperation with users if they suffer from unexplained problems. Performance monitoring tools are therefore essential to see the problems and share information about them between the users and SINET operators. We learned though experience that the operators can deeply understand the problems and take effective actions to resolve them by seeing the same information about performance issues as the users. We therefore started developing a performance visualization tool for real-time monitoring as well as time-series monitoring.

A developed prototype system collects NetFlow data from SINET routers, aggregates and calculates on a real-time or a batch basis, and visualizes the results in response to user requests (Fig. 13). The first version focuses on the real-time aggregation and visualization, which the users want the most, and uses a series of open-source software. Real-time aggregation of NetFlow data is realized by Spark Streaming [58] combined with Flume [59] and HBase [60]. Each process set of Spark Streaming (which is composed of data extraction, NetFlow parser, window aggregation, and HBase write processes) is executed as a short batch at one-second intervals. Our visualization software receives the stored data in HBase through the API and enables both SINET operators and users to monitor the network performance in real time. It has zooming interfaces to see more detailed flow behaviors, enables visual programming for their customized viewing, and offers shared comment spaces for better mutual communication. HPCI project [61], which has a K computer and frequently transfers huge storage data between related locations, has started using this monitoring tool (Fig. 14).

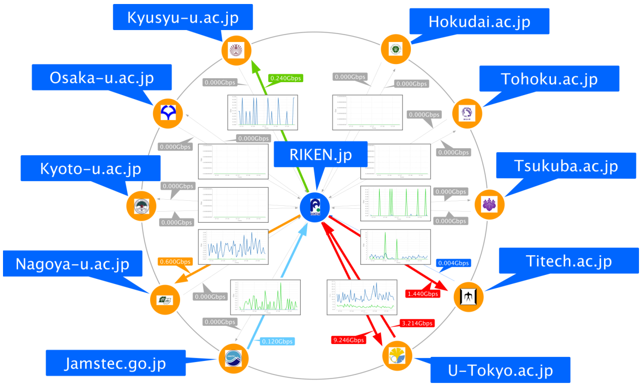


Fig. 14 Visualized performance in HPCI project.

We plan to enhance visualization capabilities more in terms of performance, security, and disaster recovery toward the 100-Gbps backbone network monitoring.

8. Conclusion

This paper described the entire network architecture and related services of SINET5. The nationwide ultra-high-speed, small-latency, reliable, SDN-friendly, and secure backbone network will greatly enhance the research and educational environment and flexibly grow in response to user demands. The advanced services will dynamically create and expand users' communication environment along with evolving cloud services. The users will be able to fully utilize the 100-Gbps backbone network with the high-performance protocol and improve the performance with the monitoring tool in cooperation with SINET operators. More detailed evaluation over the real field will be reported in the near future.

Acknowledgements

We wish to thank all of the members of the Organization of Science Network Operations and Coordination for their support of SINET. We are also grateful to Prof. Hideaki Sone of Tohoku University, Prof. Yasuo Okabe of Kyoto University, Prof. Jun Adachi, Prof. Yusheng Ji, Prof. Nobukazu Yoshioka, Mr. Koichi Ojiro, Mr. Asaji Sakamoto, Mr. Suguru Sato, Mr. Shinji Takano, Mr. Akitoshi Morishima, Mr. Hikaru Matsumura, Mr. Takahiro Kozono, Mr. Tadato Yamamoto, Ms. Mayuko Saito, Mr. Yoshinobu Masatani, Mr. Masaru Nagaku, Mr. Koji Sakaguchi, and Mr. Akira Kubo of NII for their continuous cooperation and support.

This work was partly supported by a grant from the Japan Society for the Promotion of Science (KAKENHI (23240011)).

References

[1] E. Boyd, "Internet2 update: Innovation through advanced network services," TIP2013, Jan. 2013.
[2] M. Enrico, "GÉANT Reloaded! (A 500G R&E Network is coming soon...)," APAN 35th Meeting, Jan. 2013.
[3] J. Metzger, "ESnet5 deployment: Lessons learned," TIP2013, Jan. 2013.
[4] R. Evans, "(almost) two years of 100GE on SuperJANET5... and what it means for Janet6," APAN 35th Meeting, Jan. 2013.
[5] SURFnet: http://www.surf.nl/en/services-and-products/surfinternet/index.html
[6] NORDUnet: http://www.nordu.net/ndnweb/home.html
[7] CANARIE: http://www.canarie.ca/templates/about/publications/docs/AnnualReport2014_en.pdf
[8] X. Li, "CERNET 100G updates," APAN 35th Meeting, Jan. 2013.
[9] B. Cho, "New infrastructure with 100G and services of KREONET/KRLight," APAN 35th Meeting, Jan. 2013.
[10] D. Wilde, "AARNet: 100G coast-to-coast," TNC2014, May 2014.
[11] LHC: http://home.web.cern.ch/about/computing
[12] N. Kawaguchi, "Trial on the efficient use of trunk communication lines for VLBI in Japan," 7th International eVLBI Workshop, June 2008.
[13] GSI VLBI: http://www.spacegeodesy.go.jp/vlbi/en/index.html
[14] M. Maruyama, H. Kimiyama, A. Yutani, M. Kakiuchi, H. Otsuki, K. Kobayashi, M. Sakai, M. Kobayashi, M. Sano, and A. Inoie, "World's first successful experiment of transmission, storing, and distribution of uncompressed 8K ultra-high-definition live-streaming video," IEICE Technical Report, NS2014-34, May 2014 (in Japanese).
[15] Belle II: http://belle2.kek.jp/
[16] ITER: http://www.iter.org/
[17] VLBI2010: Y Fukuzaki, M. Ishihara, J. Kuroda, S. Kurihara, K. Kokado, and R. Kawabata, "New project for constructing a VLBI2010 antenna in Japan," IVS 2012 General Meeting, March 2012.
[18] E.-J. Bos, "100G intercontinental: The next network frontier," TNC2013, June 2013.
[19] LHCONE: http://lhcone.net/
[20] SDN: https://www.opennetworking.org/sdn-resources/sdn-definition
[21] NSI: "GLIF automated GOLE proves NSI connection service v2.0 interoperability," http://www.glif.is/publications/press/20131203.html
[22] H.P. Dempsey, "Testbeds as a service," 2014 Internet2 Global Summit, April 2014.
[23] Y. Kanaumi, S. Saito, E. Kawai, S. Ishii, K. Kobayashi, and S. Shimojo, "RISE: A wide-area hybrid OpenFlow network testbed," IEICE Trans. Commun., vol.E96-B, no.1, pp.108–118, Jan. 2013.
[24] L. Poulopoulos, W. Routly, and J. Haas, "Firewall on demand multidomain: Security via BGP flowspec & a web platform," 2014 Internet2 Global Summit, March 2014.
[25] Internet2 NET+: http://www.internet2.edu/vision-initiatives/initiatives/internet2-netplus/
[26] SURFconext: http://www.surf.nl/en/services-and-products/surfconext/index.html
[27] Janet cloud services: https://www.ja.net/products-services/janet-cloud-services
[28] GridFTP: http://toolkit.globus.org/toolkit/docs/latest-stable/gridftp/
[29] PerfSONAR: https://fasterdata.es.net/performance-testing/perfsonar/
[30] For example, http://sc09.supercomputing.org/?pg=bandwidth.html
[31] S. Urushidani, S. Abe, Y. Ji, K. Fukuda, M. Koibuchi, M. Nakamura, S. Yamada, R. Hayashi, I. Inoue, and K. Shiomoto, "Design of versatile academic infrastructure for multilayer network services," IEEE JSAC, vol.27, no.3, pp.253–267, April 2009.
[32] S. Urushidani, M. Aoki, K. Fukuda, S. Abe, M. Nakamura, M. Koibuchi, Y. Ji, and S. Yamada, "Highly available network design and resource management of SINET4," Telecommunication Systems, vol.56, no.1, pp.33–47, May 2014.
[33] SINET usage examples: http://www.sinet.ad.jp/case-examples/
[34] E. Rosen and Y. Rekhter, "BGP/MPLS IP virtual private networks (VPNs)," RFC4364, Feb. 2006.

[35] L. Martini, E. Rosen, N. El-Aawar, and G. Heron, "Encapsulation methods for transport of Ethernet over MPLS networks," RFC4448, April 2006.
[36] K. Kompella and Y. Rekhter, "Virtual private LAN services using BGP," draft-ietf-l2vpn-vpls-bgp-08, June 2006.
[37] M. Mahalingam, D. Dutt, K. Duda, P. Agarwal, L. Kreeger, T. Sridhar, M. Bursell, and C. Wright, "Virtual eXtensible local area network (VXLAN): A framework for overlaying virtualized layer 2 networks over layer 3 networks," RFC7348, Aug. 2014.
[38] National Security Agency, "Defense in depth - A practical strategy for achieving information assurance in today's highly networked environments," https://www.nsa.gov/ia/_files/support/defenseindepth.pdf
[39] P. Pan, G. Swallow, and A. Atlas, "Fast reroute extensions to RSVP-TE for LSP tunnels," RFC4090, May 2005.
[40] R. Enns, "NETCONF configuration protocol," RFC4741, Dec. 2006.
[41] B. Claise, G. Sadasivan, V. Valluri, and M. Djeraes, "Cisco Systems NetFlow services export version 9," RFC3954, Oct. 2004.
[42] J.M. Soricelli and W. Gustavus, "Tutorial: Options for blackhole and discarding routing," NANOG 32, Oct. 2004.
[43] P. Marques, N. Sheth, R. Raszuk, B. Greene, J. Mauch, and D. McPherson, "Dissemination of flow specification rules," RFC5575, Aug. 2009.
[44] REN-ISAC: http://www.ren-isac.net/
[45] NSHARP: http://www.geant.net/Network/NetworkOperations/Pages/Network_Security.aspx
[46] Usage Guidelines and Checklist for Cloud Services, http://www.icer.kyushu-u.ac.jp/docs/ac/ac_guideline.pdf, March 2014 (in Japanese).
[47] GakuNin: http://www.gakunin.jp/
[48] Use Cases and Functional Requirements for Inter-Cloud Computing, White Paper, Global Inter-Cloud Technology Forum, Aug. 2010.
[49] Docker: https://www.docker.com/
[50] Y. Gu and R.L. Grossman, "UDT: UDP-based data transfer for high-speed wide area networks," Comput. Netw., vol.51, no.7, pp.1777–1799, May 2007.
[51] R. Stewart, "Stream control transmission protocol," IETF RFC4960, Sept. 2007.
[52] S. Ha, I. Rhee, and L. Xu, "CUBIC: A new TCP-friendly high-speed TCP variant," ACM SIGOPS Operating Systems Review, vol.42, no.5, pp.64–74, July 2008.
[53] D.X. Wei, C. Jin, S.H. Low, and S. Hegde, "FAST TCP: Motivation, architecture, algorithms, performance," IEEE/ACM Trans. Netw., vol.14, no.6, pp.1246–1259, Dec. 2006.
[54] A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar, "Architectural guidelines for multipath TCP development," IETF RFC6182, March 2011.
[55] B. Allcock, J. Bester, J. Bresnahan, A.L. Chervenak, I. Foster, C. Kesselman, S. Meder, V. Nefedova, D. Quesnel, and S. Tuecke, "Data management and transfer in high-performance computational grid environments," Parallel Computing, vol.28, no.5, pp.749–771, May 2002.
[56] bbFTP: http://doc.in2p3.fr/bbftp/
[57] K. Yamanaka, S. Urushidani, H. Nakanishi, T. Yamamoto, and Y. Nagayama, "A TCP/IP based constant-bit-rate file transfer protocol and its extension to multipoint data delivery," Fusion Engineering and Design, vol.89, no.5, pp.770–774, May 2014.
[58] Spark Streaming: https://spark.apache.org/streaming/
[59] Flume: http://flume.apache.org/
[60] HBase: http://hbase.apache.org/
[61] HPCI: https://www.hpci-office.jp/folders/english

Shigeo Urushidani is a professor and director in the Information System Architecture Research Division of the National Institute of Informatics (NII). He received B.E. and M.E. degrees from Kobe University in 1983 and 1985 and a Ph.D. from the University of Tokyo in 2002. He worked for NTT from 1985 to 2006, where he was engaged in the research, development, and deployment of high-performance network service systems including ATM, AIN, high-speed IP/MPLS, and GMPLS-based optical systems. His work also included the design of real networks, such as the KOCHI Information Superhighway, Japan Gigabit Network, and DoCoMo Data Backbone Network. He moved to NII in 2006 and is currently involved in the design and implementation of SINET. His current research interests include network architecture and system architecture for ultra-high-speed green networks. He is a member of IEICE, IPSJ, and IEEE.

Shunji Abe received B.E. and M.E. degrees from Toyohashi University of Technology, Japan, in 1980 and 1982. He received a Ph.D. from the University of Tokyo in 1996. In 1982, he joined Fujitsu Laboratories Ltd., where he engaged in research on broadband circuit switching systems, ATM switching systems, ATM traffic control, and network performance evaluation. He worked at the National Center for Science Information Systems, Japan (NACSIS) from 1995 to 1999. Since 2000, he has worked at the National Institute of Informatics (NII) as an associate professor, and he is now a director of SINET Promotion Office. He is also an associate professor of the Graduate University for Advanced Studies (SOKENDAI). His current research interests are Internet traffic analysis, network performance evaluation, and mobile IP system architecture. He is a member of IEICE, IPSJ, and IEEE.

Kenjiro Yamanaka received B.E. and M.E. degrees from University of Tsukuba, Japan, in 1986 and 1988. He joined NTT (Nippon Telegraph and Telephone Corporation) in 1988 and was engaged in the research of protocol synthesis and verification. He moved to NTT Data Corporation in 2002 and was engaged in a research project to build Inter-Cloud technology after several years' experience in business development. He is currently an associate research professor in the Advanced ICT Center, National Institute of Informatics (NII), Tokyo, and operates and develops NII's in-house cloud service. His research interests include high-speed networking, network performance evaluation, and cloud computing. He is a member of IEICE, IPSJ, ACM, and IEEE.

Kento Aida received Dr. Eng. in electrical engineering in 1997 from Waseda University, where he had been a research since 1992. He joined Tokyo Institute of Technology and became a research scientist at the Department of Mathematical and Computing Sciences in 1997, an assistant professor at the Department of Computational Intelligence and Systems Science in 1999, and an associate professor at the Department of Information Processing in 2003. He is now a professor at the National Institute of Informatics (NII) and has been a visiting professor at the Department of Information Processing in Tokyo Institute of Technology since 2007. He was also a researcher at PRESTO in Japan Science and Technology Agency (JST) from 2001 through 2005 and a research scholar at the Information and Computer Sciences Department in the University of Hawaii in 2007.

Shigetoshi Yokoyama is a project professor in the Information System Architecture Research Division of the National Institute of Informatics (NII). He received B.E. and M.E. degrees from Osaka University in 1979 and 1981 and a Ph.D. from the Graduate University for Advanced Studies in 2013. He worked for NTT from 1981 to 1988, where he was engaged in the research, development, and deployment of operating systems and distributed systems. His works also included the design of real workstations based on UNIX. He moved to NII in 2008 and is currently involved in the design and implementation of NII Cloud and academic inter-cloud hub. From 1989 to 1991, he was a visiting scientist in the Intelligent Engineering Systems Lab (IESL), Massachusetts Institute of Technology. His current research interests include inter-cloud computing architecture and autonomous service operations. He is a member of IEICE and IPSJ.

Hiroshi Yamada received a B.S. degree in mathematics from Nagoya University in 1985 and his Ph.D. in network engineering from Tokyo Institute of Technology in 1996. He worked in NTT laboratories from 1985 to 2014, where he was engaged in research and development on teletraffic theory, performance evaluation, computer network and protocol simulation using OPNET, and network security. He moved to NII in 2014 and is currently involved in the design and implementation of backbone network and its network security of SINET. He is a member of IEICE.

Motonori Nakamura graduated from Kyoto University, Japan, where he received B.E., M.E., and Ph.D. degrees in engineering in 1989, 1991, and 1996. From 1995, he was an associate professor at Kyoto University. Currently he is a professor at National Institute of Informatics, Japan (NII) and the Graduate University for Advanced Studies (SOKENDAI). His research interests are message transport networks, network communications, next generation Internet, and identity & access management. He is a member of IEEE, IEICE, IPSJ, and JSSST.

Kensuke Fukuda is an associate professor at the National Institute of Informatics (NII). He received his Ph.D. degree in computer science from Keio University in 1999. He worked in NTT laboratories from 1999 to 2005, and joined NII in 2006. He was a visiting scholar at Boston University in 2002 and a visiting scholar at the University of Southern California/Information Sciences Institute in 2014–2015. He was also a researcher of PRESTO JST (Sakigake) in 2008–2012. His current research interests are Internet traffic measurement and analysis, intelligent network control architectures, and scientific aspects of networks.

Michihiro Koibuchi received B.E., M.E., and Ph.D. degrees from Keio University, Yokohama, Japan, in 2000, 2002, and 2003. He was a visiting researcher at the Technical University of Valencia, Spain in 2004 and a visiting scholar at the University of Southern California in 2006. He is currently an associate professor in the Information Systems Architecture Research Division, National Institute of Informatics, Tokyo, and the Graduate University for Advanced Studies, Japan. His research interests include high-performance computing and interconnection networks. He is a member of the IEEE, IPSJ, and IEICE.

Shigeki Yamada is currently a professor and director with the Principles of Informatics Research Division, National Institute of Informatics, Japan. He received B.E., M.E., and Ph.D. degrees in electronic engineering from Hokkaido University, Japan in 1972, 1974, and 1991. He worked in the NTT (Nippon Telegraph and Telephone Corporation) laboratories from 1974 to 1999, where he was involved in research and development on digital switching systems and information and communication networks. He moved to NII in 2000. From 1981 to 1982, he was a visiting scientist in the Computer Science Department, University of California, Los Angeles. His current research interest includes future Internet architectures, software-defined networking, mobile and wireless networks. He is a senior member of IEEE and a member of IEICE and IPSJ.