Application-aware Synergetic Network Control Technology Demonstrated in a Field Trial in US Nationwide Photonic Network

Yukio Tsukishima†*, Akira Hirano, and Masahiko Jinno*

Abstract

Lambda-on-demand functionality that autonomously adjusts the number of optical paths (i.e., wavelengths of light (lambdas)) between photonic cross connect (PXC) systems in response to application traffic between the PXCs has been demonstrated in a US nationwide photonic network for the first time. In the field trial, two lambdas were used from Chicago to San Diego. Two PXCs with generalized multi-protocol label switching (GMPLS) capability were installed in Chicago and one in San Diego to form a triangular topology. In addition, we developed control servers to monitor application traffic on the first lambda and automatically activate the second lambda when needed. The control servers with the PXC control plane automatically configured the number of lambdas and layer-2 connections to accommodate dynamic changes in application traffic transparently.

1. Introduction

One of the important roles of networks is to connect applications, so networks should be optimized to meet the bandwidth and quality-of-service (QoS) requirements of those applications. Such optimization of the network configuration requires technologies that can control a multilayer[*1] network. An IP-over-photonic network (IP: Internet protocol) with generalized multi-protocol label switching (GMPLS) capability can control multilayer network configurations and is a promising candidate for the design of next-generation high-performance networks. In the field of grid computing, grid networks have recently been designed to use GMPLS-enabled IP-over-photonic networks comprising photonic cross connects (PXCs)[*2] and optical paths, as shown in **Fig. 1** [1]. The technology used in optical networking to switch individual wavelengths of light onto separate paths for specific routing of information is called lambda switching (sometimes called photonic switching or wavelength switching). A lambda connects directly between routers, skipping intermediate routers. A PXC is responsible for routing packets on lambdas. A super scheduler reserves computer and network resources in response to user demands. A network resource manager and a computer resource manager control the network and computers according to the resource reservations. However, two problems arise when a network configuration is changed in response to application demands. One is that a user cannot correctly know the bandwidth and QoS that an application requires and adequately request them from the network. The other is that dynamic changes to the network configuration disrupt the IP routing. Little has been done to resolve these problems.

In this paper, we present evaluation results for the dynamic lambda-on-demand functionality for a real-time bandwidth-intensive application, the scalable adaptive graphics environment (SAGE) [2] developed by the Electronic Visualization Laboratory (EVL), over a US nationwide testbed. The lambda-

† NTT Network Innovation Laboratories
Yokosuka-shi, 239-0847 Japan
E-mail: tsukishima.yukio@lab.ntt.co.jp

*1 Multilayer network: The OSI (open system interconnection) model defines a networking framework for implementing protocols in seven layers. Of these, layers 1–3 are of interest to us here. Layer 1 is the physical layer, layer 2 is the data link layer, and layer 3 is the network layer.

*2 PXC: A photonic cross connect is a transparent optical cross connect (OXC); that is, it is a more specific type of device while an OXC is a generic one. In the networking context, a transparent device means one that transmits/propagates an optical signal without terminating it.

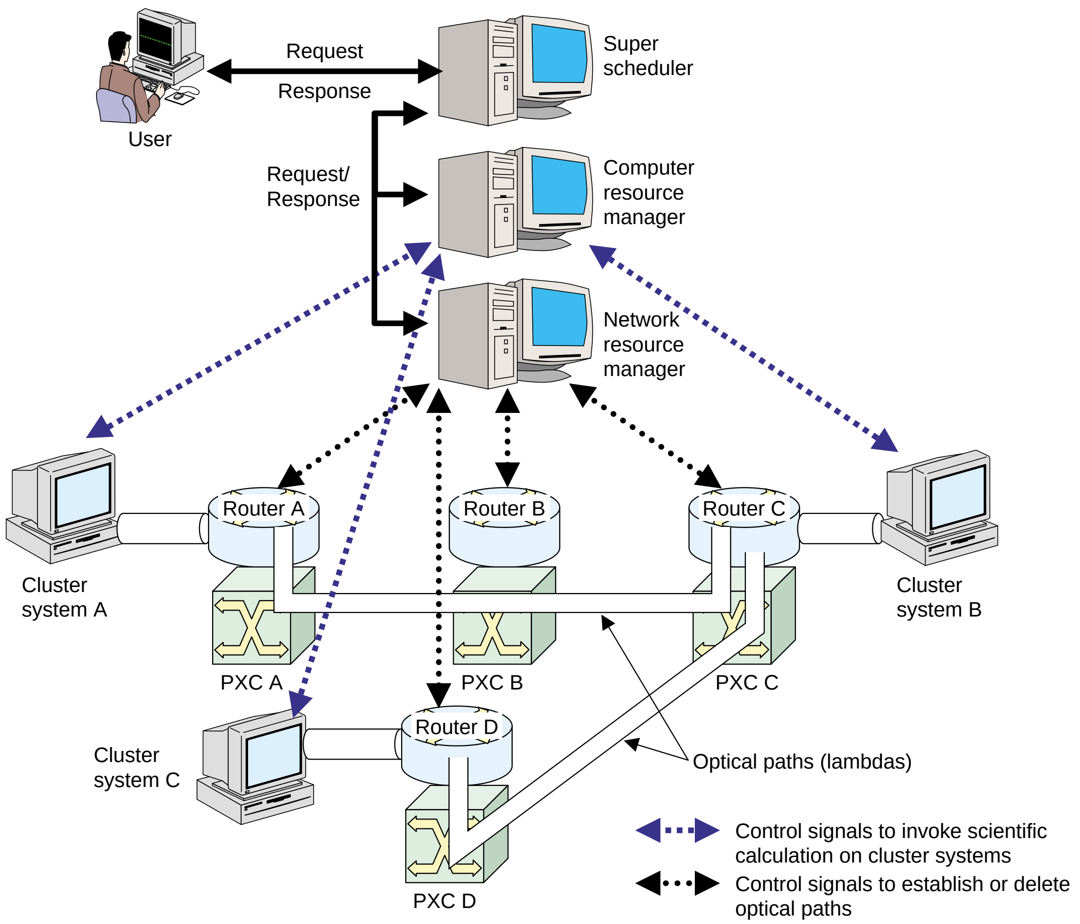

Fig. 1. Grid network.

on-demand technology [3]-[6], which we proposed in 2001, autonomously adjusts the multilayer network configuration without IP routing disruption according to the application demand. We used two 10-GbE (gigabit Ethernet) links from Chicago to San Diego. Two PXCs with GMPLS capability were installed in Chicago and one in San Diego to form a triangular topology. In addition, we developed control servers to monitor the application traffic on the first lambda and to invoke the automatic activation of the second lambda. The control server with the PXC control plane automatically configured the number of lambdas and layer-2 (L2) connections to accommodate the dynamic changes in SAGE traffic transparently.

2. Application-aware synergetic network

Since 2001, we have been developing application-aware synergetic technologies that can autonomously adjust multilayer networks without IP routing disruption in response to dynamic changes in the bandwidth and QoS that applications require. These include lambda-on-demand technology [3]-[6], optical virtual concatenation technology [7], lambda-coloring technology, GMPLS link restoration technology [8], and multiple-failure recovery technology [9]. As shown in **Fig. 2**, the lambda-on-demand technology monitors application traffic on the first lambda and establishes a second lambda if needed to handle the volume of application traffic. The optical virtual concatenation technology monitors the delay time on each lambda and unifies the delay times between lambdas so that packets arrive at their destination in the same sequence as at the origin. The lambda-coloring technology shown in **Fig. 3** distributes IP-routed packets to different lambdas based on the QoS level of the packet and assigns a priority to each lambda according to the QoS level of the accommodated packets. The high-priority lambda has a dedicated protection lambda, which transports the same packets as those streaming on the high-priority lambda. Packets on the dedicated protection lambda are

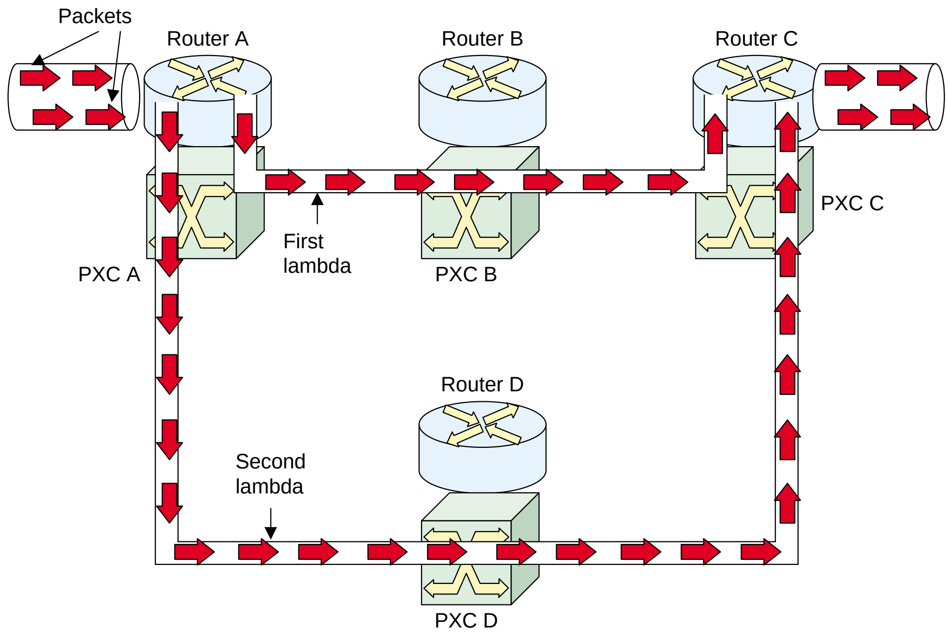

Fig. 2. Lambda-on-demand and optical virtual concatenation technologies.

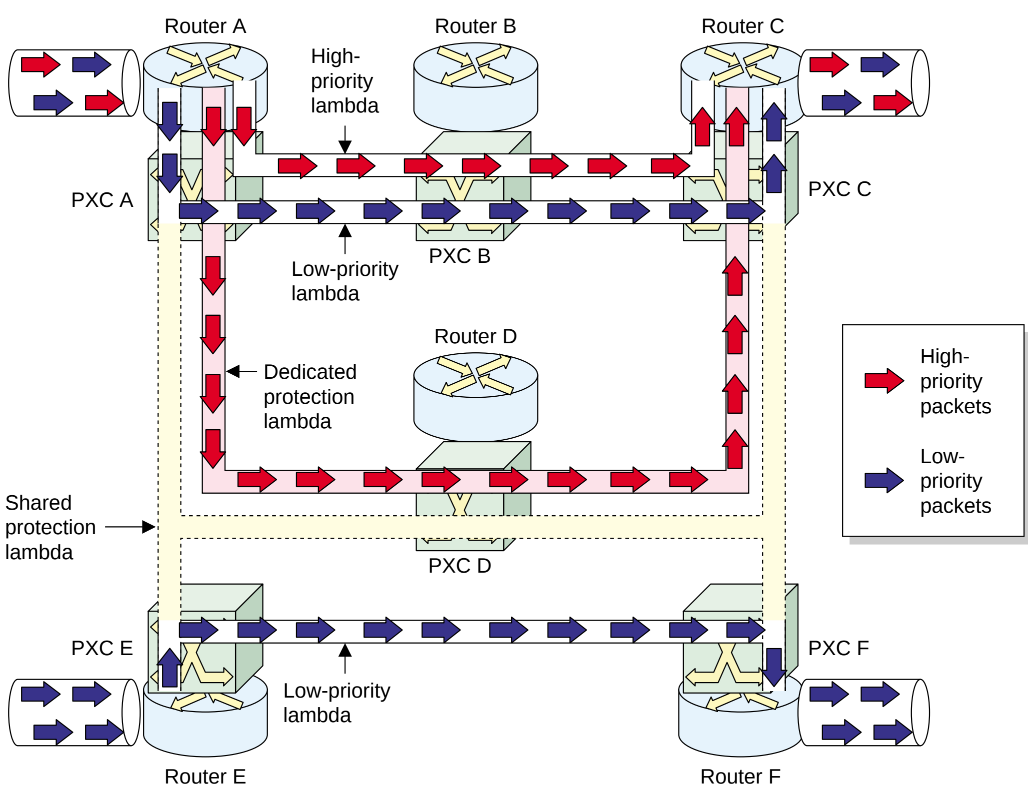

Fig. 3. Lambda-coloring and failure recovery technologies.

normally discarded at the egress side, but they start to pass through the egress side without impediment when a failure of the high-priority lambda occurs. However, the low-priority lambda has a shared back-up lambda which is set up only in the event of a failure. When failures of the low-priority lambda occur, the shared backup lambda is provided by the GMPLS link restoration technology, which can restore failures between ingress and egress IP routers without IP-routing disruption caused by the failures. If multiple failures occur on the lambdas, multiple-failure recovery technology dynamically sets up new lambdas as backup for the failed ones. We will install these technologies into PXCs and IP routers.

3. Network configuration and SAGE middleware system

The geographical network configuration for this field trial is shown in **Fig. 4**. It consists of two 10-GbE links that cross the USA. One link uses the National LambdaRail [10] via the StarLight [11] facility in Chicago, to Seattle, Los Angeles, and eventually San Diego. Using Cisco 15808 and ONS15454 gear, 10-GbE LAN-PHYs (local area network physical layers) are directly transported on dense wavelength division multiplexing (DWDM) links. The second 10-GbE link was mapped onto an OC-192 channel provided by TeraGrid [12] with the help of MPLS on a Juniper T-640. The 10-GbE LAN-PHY extends from Chicago, through Los Angeles, to San Diego.

The experimental configuration is shown in more detail in **Fig. 5**. We developed three PXC prototypes, each equipped with an 8×8 planar lightwave circuit (PLC) optical switch, as shown in **Fig. 6**. It supports various kinds of interfaces including both a network-node interface (NNI) and optical user-network interface (OUNI) and has a supervisory and control unit for management and signaling communication with GMPLS capability within the control plane. The interface cards can accommodate both 10-GbE LAN-PHY and OC-192 (10-GbE WAN-PHY (WAN: wide area network)). We installed three PXC systems: two in Chicago and one at the California Institute for Telecommunications and Information Technology at the University of California San Diego. Each PXC has a network element manager (NE-Mgr), which signals other NE-Mgrs via GMPLS to establish or delete a lambda. Each of the NE-Mgrs controls PXC switches using the extended generic switch management protocol (GSMP). The L2 switches at each site have control servers that control their respective NE-Mgr via a GMPLS/OUNI and the L2 switch via SNMP (simple network management protocol) over telnet, and a remote control server via a proprietary protocol. The L2 switches accept two GbE traffic flows from the SAGE transmitter (Tx) or to the receiver (Rx). These NE-Mgrs and control servers are connected to each other on a routed control-plane network that is independent of the 10-GbE data-plane network.

SAGE [9], [13] is a middleware system for managing visualization and high-definition video streams intended to be viewed on ultrahigh-resolution displays such as EVL's 100-megapixel LambdaVision, which is a tiled liquid crystal display wall. SAGE consists of a free space manager, SAGE application interface libraries (SAILs), SAGE receivers, user interfaces (UI clients), and a display, as shown in **Fig. 7**. The free space manager receives user commands from the UI clients and controls pixel streams between SAILs and SAGE receivers. A SAIL captures output pixels from an application and streams them to the appropriate SAGE receivers. A SAGE receiver can retrieve multiple pixel streams from dif-

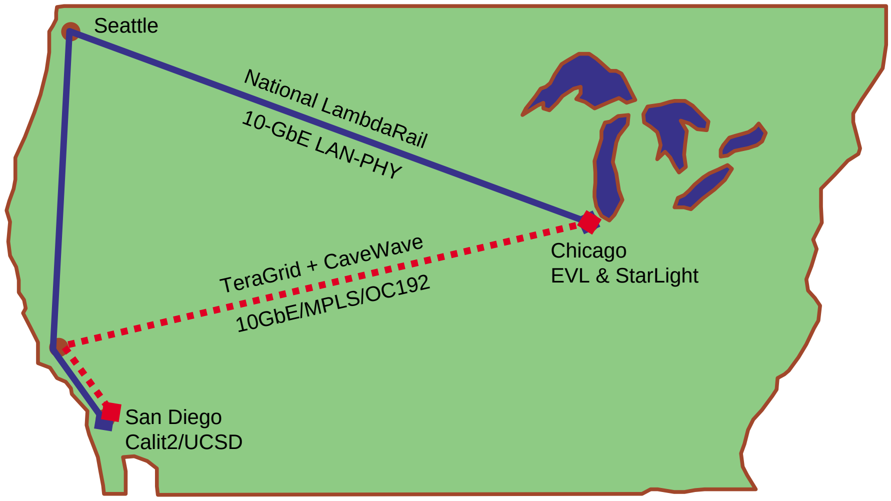

Fig. 4. Network configuration.

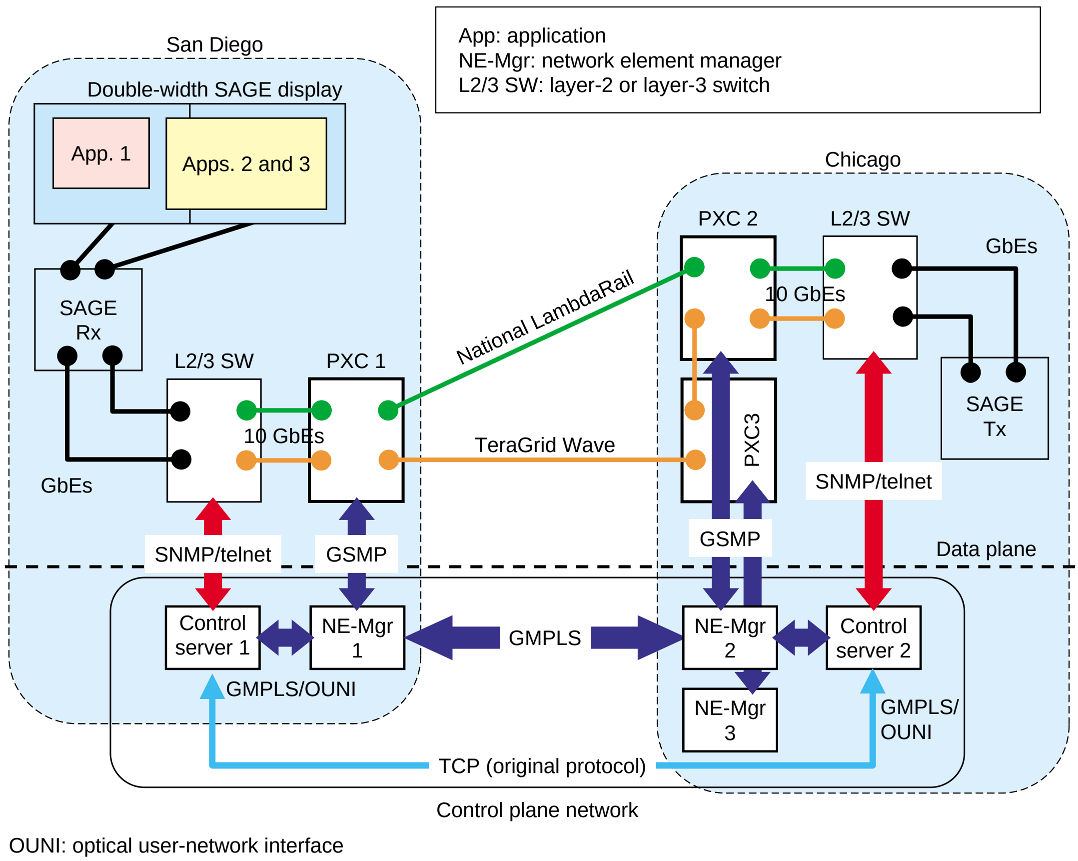

Fig. 5. Experimental configuration.

ferent applications and display streamed pixels on multiple tiles. Remote visualization applications, such as 3D rendering, remote desktop, video streams, and large 2D maps, stream their rendered pixels (or graphics primitives) to SAGE, allowing any given layout on the displays, e.g., the output of arbitrary M × N pixel rendering cluster nodes can be streamed to X × Y pixel display screens.

4. Lambda-on-demand with GMPLS, SNMP, and telnet

A schematic of the lambda-on-demand scheme with GMPLS, SNMP, and telnet is shown in **Fig. 8**. Function #1 periodically monitors the amount of application traffic on the first lambda using SNMP polling between the ingress control server and ingress L2 or L3 system. Function #2 judges when to add or delete a second lambda or when to acquire the status via the ingress control server. Function #3 checks the status of the network resources such as that of the L2/L3 system interfaces and lambdas, using the ingress and egress control servers. The cycle of Func-

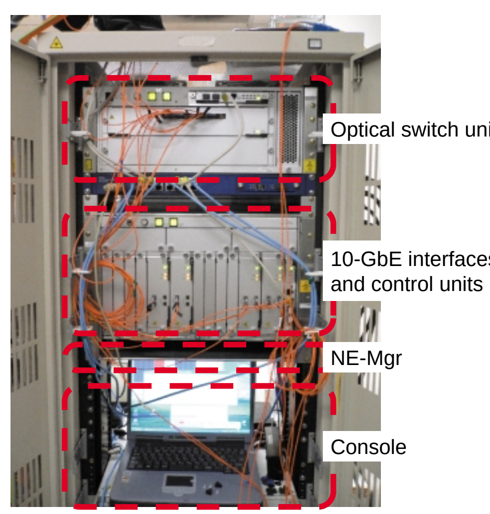

Fig. 6. PXC system with GMPLS control.

tions #1–#3 is repeated regularly independent of Functions #4–#8. Functions #4–#8 are initiated when the decision to add or delete a second lambda has been made. Function #4 establishes the second lambda via GMPLS. Function #5 checks if the establish-

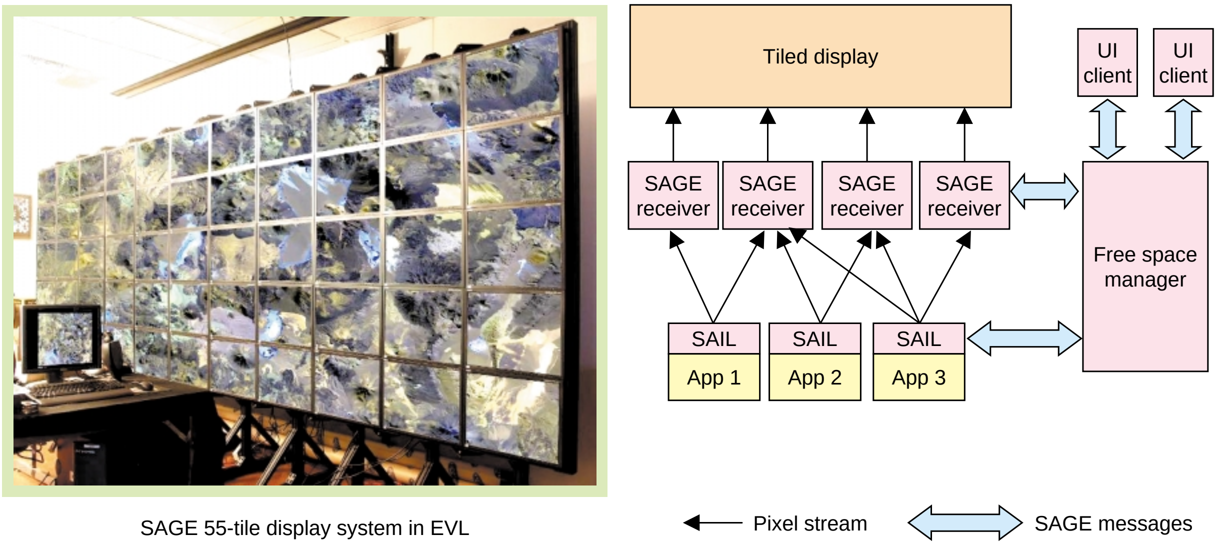

Fig. 7. Scalable adaptive graphics environment.

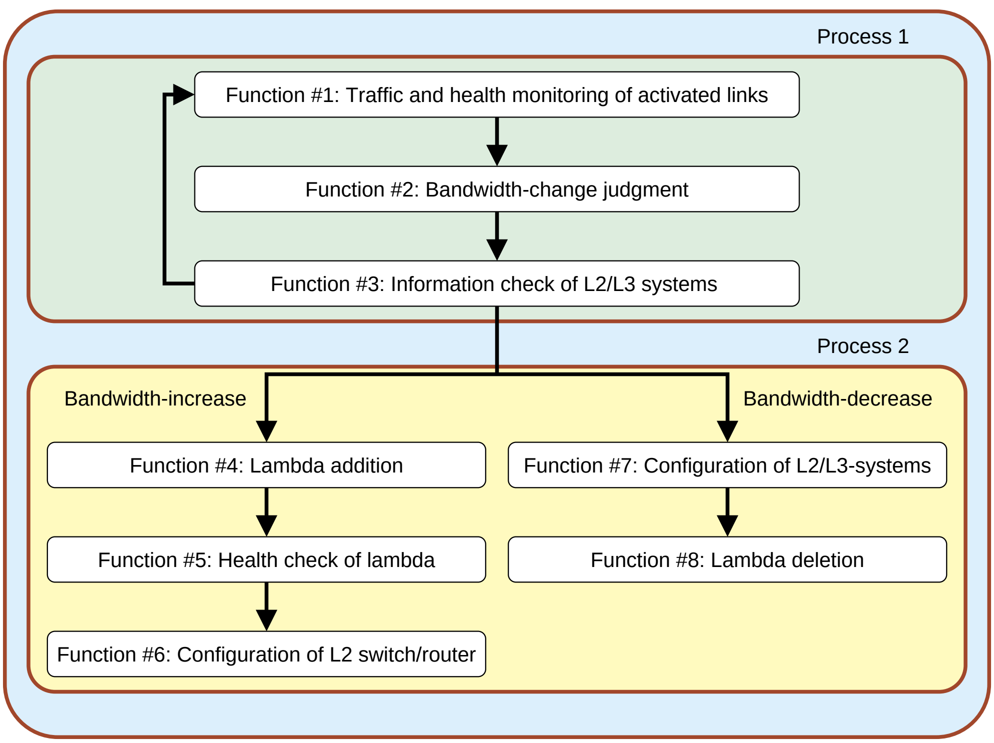

Fig. 8. Schematic of lambda-on-demand function.

ment of the second lambda was successful. Function #6 is cancelled if the second lambda is disabled. Function #6 activates the link on the second lambda and bundles links on the first and second lambdas via link aggregation (IEEE802.3ad). Function #7 releases the link on the second lambda from the bundling status and deactivates the link. Function #8 deletes the second lambda. The parameters in this field trial were as follows. The second lambda was designed to be engaged if a series of two traffic measurements, monitored for four seconds on the first lambda, exceeded 800 Mbit/s. The second lambda was disengaged if the same measurements yielded a bandwidth below 350 Mbit/s. IEEE802.3ad maintained the link status independently of the number of L2 connections when more than one L2 connection was established.

Consequently, there were no changes in the IP routing. A round-robin algorithm was used to distribute packets into the links.

5. Dynamic lambda-on-demand with L2 harmonization

An example of packet loss records as a function of elapsed time observed during our field trial is shown in **Fig. 9**. As the initial condition, one 10-GbE connection was working between the SAGE Tx and Rx. Then we started a small streaming application on SAGE that pushed the traffic below a threshold, as shown in the figure. Then we launched two additional streaming applications that required enough total bandwidth to invoke the activation of another lambda. To avoid undesirable oscillation of the additional lambda activation, we set a decision guard time of eight seconds. During setup, L1 (lambda) reconfiguration preceded L2 reconfiguration to avoid smooth transition in combination with L2 reconfiguration. As a result, the additional lambda completely eliminated packet loss, as shown in Fig. 9. We confirmed excellent streaming during most of the test time. In the transition when the new lambda was set up or torn down, we observed a slight increase in the packet loss; however, it was recovered within a few seconds. All the L1/L2 reconfigurations were completed in a few seconds in our field trial. The number of lambdas was controlled automatically depending on the changes in traffic volume of these applications without requiring any explicit messaging between the application and the control plane of the photonic network.

6. Conclusion

Dynamic lambda-on-demand functionality with harmonized re-configuration of the corresponding layer-2 connections with link aggregation was achieved for the first time using SAGE visualizing applications over a US nationwide network testbed. Two 10-GbE lambdas were automatically configured to provide end-to-end L2 connections using a newly developed control server that successfully negotiated L1/L2 connections. We confirmed excellent streaming for both configurations—with one or two lambdas—meeting the capacity demand of the SAGE applications without any messaging between the application and the control plane of the photonic network. We found that the lambda-on-demand technology was able to work smoothly with high-end applications supporting the idea that it can be applied to grid networks and virtual private networks. In fact, the grid network architecture with lambda-on-demand functionality is being discussed in the Global Grid Forum, a standards body. In the near future, we will integrate lambda-on-demand, optical virtual concatenation, and other functionalities and install them in grid network systems.

References

[1] G-lambda, http://www.g-lambda.net/

[2] OptIPuter, http://www.optiputer.net/index.html

[3] Y. Okazaki, A. Watanabe, T. Takahashi, and Y. Tsukishima, "Network Resource Evaluations of Photonic Network Employing Capacity Adjustment Capability," ECOC-IOOC2003, We2.4.3.

[4] I. Shake, M. Fukutoku, O. Morikawa, T. Kurahashi, Y. Tarui, Y.

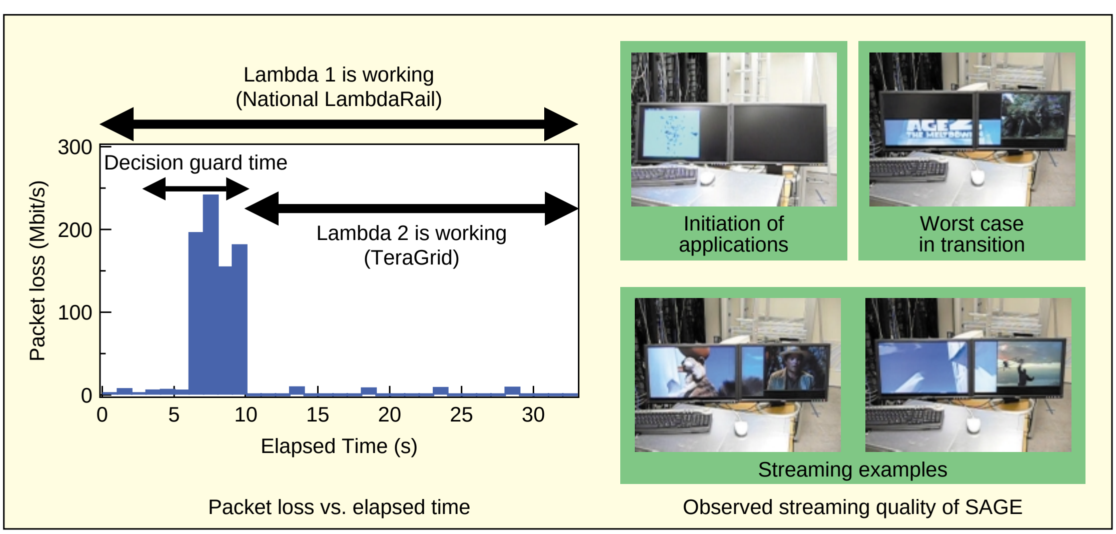

Fig. 9. Packet loss and streaming quality of SAGE.

Tsukishima, and W. Imajuku, “Optical link capacity adjustment experiments for photonic IX,” ECOC2005, Tu3.4.3.

[5] M. Fukutoku, I. Shake, S. Yamamoto, W. Imajuku, T. Kurahashi, and Y. Tarui, “Router-OXC Cooperation for Fault Recovery Employing GMPLS Control Plane for Photonic IX,” OFC2006, OWQ7.

[6] Y. Tsukishima, A. Hirano, N. Nagatsu, T. Ohara, W. Imajuku, M. Jinno, Y. Takigawa, K. Hagimoto, L. Renambot, B. Jeong, J. Leigh, T. DeFanti, A. Verlo, and L. Winkler, “The First Application-driven Lambda-on-Demand Field Trial over a US Nationwide Network,” T. Digest of OFC2006 PDP4.

[7] A. Hirano, L. Renambot, B. Jeong, J. Leigh, A. Verlo, V. Vishwanath, R. Singh, J. Aguilera, A. Johnson, T. DeFanti, L. Long, N. Schwarz, M. Brown, N. Nagatsu, Y. Tsukishima, M. Tomizawa, Y. Miyamoto, M. Jinno, Y. Takigawa and O. Ishida “The first functional demonstration of optical virtual concatenation as a technique for achieving Terabit networking,” Future Generation Computer Systems, Vol. 22, Issue 8, pp. 876-883, 2006.

[8] Y. Tsukishima, A. Hirano, N. Nagatsu, T. Ohara, W. Imajuku, M. Jinno, Y. Takigawa, K. Hagimoto, L. Renambot, B. Jeong, J. Leigh, T. DeFanti, A. Verlo, and L. Winkler, “Stable IP-Routing Link Restoration: GUNI Restoration for Data Link Failure Between Routers in a Nationwide Photonic Network,” ECOC2006, We4.1.4.

[9] Y. Sone, N. Nagatsu, W. Imajuku, M. Jinno, and Y. Takigawa, Optical Path Restoration Scheme Escalation Achieving Enhanced Operation and High Survivability in Multiple Failure Scenarios,” WTC2006, OCN NW1.

[10] National LambdaRail, http://www.nlr.net/

[11] StarLight, http://www.startap.net/starlight/

[12] TeraGrid, http://www.teragrid.org/

[13] L. Renambot, A. Rao, R. Singh, J. Byungil, N. Krishnaprasad, V. Vishwanath, V. Chandrasekhar, N. Schwarz, A. Spale, C. Zhang, G. Goldman, J. Leigh, and A. Johnson, “SAGE: the Scalable Adaptive Graphics Environment,” Proceedings of WACE 2004, Nice, France, 2004.

Yukio Tsukishima
Engineer, Photonic Transport Network Project, NTT Network Innovation Laboratories.
He received the B.E. degree in electrical and electronic engineering and M.E. degree in physical electronics from Tokyo Institute of Technology, Tokyo, in 1999 and 2001, respectively. Since joining NTT Network Innovation Laboratories in 2001, he has been engaged in R&D of IP and optical layer interworking technologies using GMPLS functions.

Akira Hirano
Senior Research Engineer, NTT Network Innovation Laboratories.
He received the B.S. degree in physical chemistry and the M.S. degree in astrophysics from Osaka University, Osaka, in 1991 and 1993, respectively. He joined NTT in 1993 and worked on high-speed optical communications. He was a visiting researcher at the University of Illinois at Chicago, USA (2005–2006), where he engaged in research on photonic networking. His main interests are in high-speed optical transport systems, optical signal processing, and photonic networking. He is a member of IEEE and Institute of Electronics, Information, and Communication Engineers (IEICE) of Japan.

Masahiko Jinno
Senior Research Engineer, Supervisor, Photonic Transport Network Laboratory, NTT Network Innovation Laboratories.
He received the B.E. and M.E. degrees in electronics engineering from Kanazawa University, Kanazawa, and the Ph.D. degree in engineering from Osaka University, Osaka, in 1984, 1986, and 1995, respectively. He joined NTT in 1986. He was a guest scientist at the National Institute of Standards and Technology (NIST), Boulder, Colorado, USA from 1993–1994. He received the Young Engineer Award from IEICE in 1993, Best Paper Awards in 1997 and 1998 from the Second and Third Optoelectronics and Communications Conferences, and the NTT President’s Award in 2000. He is a member of IEICE and IEEE.