5-2 Studies on Architecture and Control Technology for Optical Packet Switches

HARAI Hiroaki

In this paper, we first describe overview and advanced technology of optical packet switches (OPS) and requirement for practical use of OPS of which target is the Internet and 10 Tbps throughput. Then, we report switch architecture, recent activities of integrated technology and electronic control systems for OPS.

Keywords
Optical packet switching, Switch architecture, Dedicated output buffer, Buffer management

1 Introduction

Studies on optical packet switches, mainly initiated in the 1990s for optical ATM (Asynchronous Transfer Mode) switches, are still underway, targeting applications in Internet technologies. The scope of applications for optical switches have gradually come to center on high-speed transmission of optical packets through the creation of a closed-domain optical label switch network under the IP (Internet Protocol) network, in a manner similar to MPLS (Multi-Protocol Label Switching). In itself, this architecture is not significantly different from systems using optical ATM switches in IP over ATM networks. The main difference lies in the fact that while ATM handles 53-byte “synchronous fixed-length packets”, the present switch handles 40-1,500-byte “asynchronous variable-length packets”, aimed at Internet applications such as MPLS and IP networks. Furthermore, the target throughput has also been changed from the initial >10 Gbps level to the present goal of >10 Tbps.

This paper will describe an overview and the architectural technology of optical packet switches having a target throughput of >10 Tbps and applications of these switches to the Internet environment described above. Performance requirements for the architectural technology and trends in the control technologies for optical packet switches will also be discussed. Part of this paper is introduced in more detail in[1]. For more details on the relevant optical technologies, readers are referred to[2].

2 Optical packet switches and performance requirements

2.1 Functions of optical packet switches

First, we will describe the functions of optical packet switches. As shown in Fig. 1, packet switches have five main functions: switching, buffering, forwarding (label lookup), buffer management (contention resolution), and routing. The focus of the present paper is on asynchronous variable-length packet processing, and so a discussion of synchronous functions is basically unnecessary.

Switching (the details are provided in Section **2.2**) and buffering (Section **2.3**) functions have been studied in implementation technologies in optical systems. This is due to the fact that the advantages of optical systems can be exploited to their fullest extent at higher payload speeds with optical packet switches, since these do not require access to payloads at intermediate nodes. In contrast, improving payload speeds for all-electronic packet switches require internal buses with higher speeds and faster memory access, resulting in increased cost.

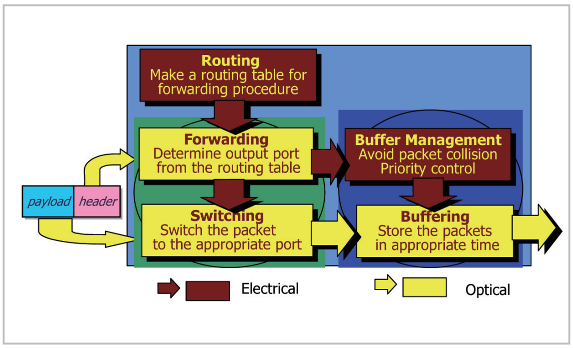
Routing
Make a routing table for forwarding procedure
Forwarding
Determine output port from the routing table
Buffer Management
Avoid packet collision
Priority control
payload
header
Switching
Switch the packet to the appropriate port
Buffering
Store the packets in appropriate time
Electrical
Optical

Fig. 1 *Internal functions of dedicated-output-buffer optical packet switches. Arrows show the flow of control information and data.*

The popularly adopted method of label lookup consists of conventional electronic processing, which checks the packet label against a table of labels retained in memory[3]-[5]. However, such methods require processing on multiple processors at channel rates of 40-160 Gbps. Since the number of packet arrivals is dependent on channel rates and not on the label-processing speed, a method is currently being investigated for instant label lookup using an optical system[6]-[8]. With this and similar techniques, label lookup is performed while the optical signal is in transmission, and so parallel processing is not required. However, at present, multiple optical systems are required to process large volumes of labels, raising the pressing issue of equipment size.

Since optical systems are not suited for buffer management and routing—the former requires logical arithmetic and the latter demands both logical arithmetic and memory—the adoption of electronic processing systems suited to the order of time required (in which buffer management is on the order of several tens of nanoseconds and routing involves scales of several seconds) is believed to offer a realistic solution. However, processing in this case must be designed to enable full exploitation of the characteristics of optical systems.

2.2 Switching system and performance requirements

The three major architectures of switching systems are as follows.

(1) A method that switches between an input channel and an output channel using optical space switches[5]

(2) Architecture with optical couplers (splitters) and optical gates[4][7]: Signals at each input port are split into a number N of signals with optical gates allocated for each. For each output, only one gate is allowed to open, thereby creating a strict-sense non-blocking switch .

(3) Architecture with wavelength converters and AWGR (Arrayed Waveguide Grating Router)[3]: For example, a system consisting of N DEMUX/MUXs, NW tunable wavelength converters (TWC) and fixed wavelength converters, and $NW \times NW$ AWGRs, when N is the number of fibers connected to switches and W is the number of multiplexed wavelengths of the fiber. Using a TWC that can convert a signal into any of NW wavelengths, it is possible to create a strict-sense non-blocking switch.

Architectures formed from combinations of the above are also possible. At NICT (National Institute of Information and Communications Technology), development and validation experiments are underway for architectures of types (1) and (2)[7].

The creation of switches on increased scales require increased conversion bandwidths in wavelength converters, an increased number of demultiplexed wavelengths for the AWGR, reduced power loss in optical gates and switches, and increased scale of the optical switches. For example, for a strict-sense non-blocking switch architecture having 10 Tbps throughput connected to 64 channels (ports) with a channel rate of 160 Gbps, the scale of the optical switch required is 1×64 and the wavelength conversion bandwidth is

12.8 THz (102.4 nm). This wavelength conversion bandwidth is calculated on the assumption that 64 wavelengths are necessary for AWGR input/output when a single port consists of a wavelength of 160 Gbps (at 200 GHz).

To minimize the guard time (minimum packet interval) and increase channel utilization, it is also important to cut the switching times required for wavelength conversion and optical switching. For example, the channel utilization when transmitting a 46-byte IP packet over a 64B/66B-modulated 10-Gigabit Ethernet (physical speed of 10.3125 Gbps) is approximately 53% (calculated based on the assumption that the preamble, Ethernet frame, and IFG (inter-frame gap) are 8, 64, and 12 bytes, respectively). For a packet with 1,500 bytes, the utilization is 94%, and for the average length packet of 250 bytes derived from[9], utilization is approximately 84%. To achieve equivalent channel utilization when using optical packet switches with channel rate of 40 Gbps, the guard times must be reduced below 7.6, 16.7, and 8.9 nanoseconds, respectively (Fig. 2). Therefore, in order to take advantage of the increased channel rate, the guard time—in other words, the switching times required for wavelength conversion and optical switching—must be reduced to the nanosecond order or lower.

2.3 Optical-fiber delay line buffering and contention resolution control

The essential difference between conventional electronically controlled packet switches and optical packet switches is whether or not the switch is of the "store-and-forward" or the "progressive" type. Electronic packet switches store data in memory (RAM) before contention resolution and header processing. Since no optical memories are presently available, optical buffers consist of progressive optical fiber delay lines (FDL). Generally speaking there are three types of optical FDL buffers (referred to below simply as "optical buffers"), as described below, in addition to a number of additional, similar types.

(1) Optical switches and FDL[5] [7] [8]

(2) Optical couplers, FDL and optical gates

(3) Wavelength converters, AWGR and FDL

The piece on the left in Fig. 3 shows an example of a 4-input 1-output optical buffer consisting of an optical switch and FDL of different lengths (which are proportional to unit length *D*) that allows a selection from among four types of delays. While a rearrangeable non-blocking optical switch will suffice for synchronous fixed-length packet switching, asynchronous variable-length packet switching requires a strictly non-blocking switch.

Some discussion of buffer size is in order at this point. Presently, if costs are not a factor, it is easy to create buffers with retention capacities of nearly 10,000 packets using semiconductor memory. In contrast, optical packet switches that rely on optical fiber delay lines cannot be expected to offer similar capacities. This was the major obstacle to the realization and study of optical buffers. However, in recent years, some researchers have

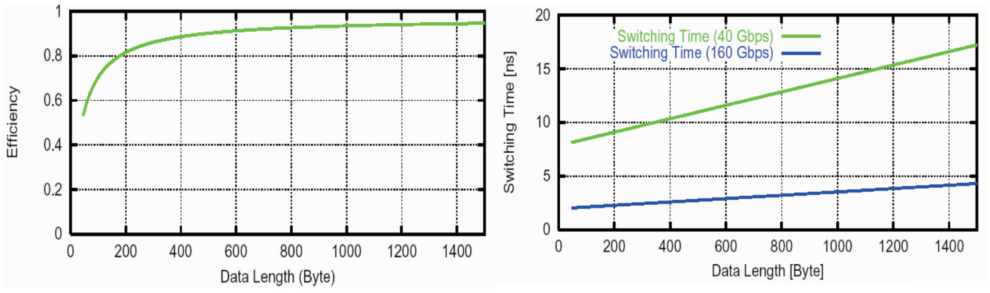

Fig.2 *Utilization of mapping IP packet on Ethernet frame (left) and switching time required for equivalent utilization under the same packet length conditions (right)*

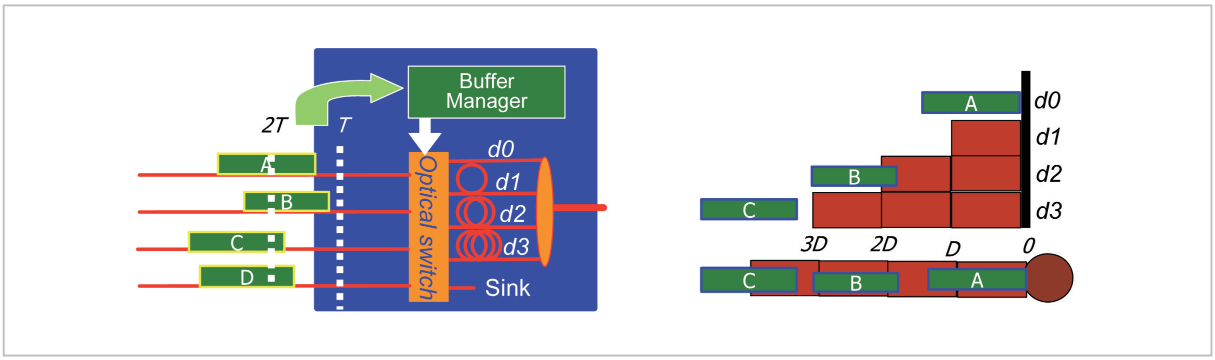

Fig.3 (Left) Optical-fiber delay line buffer; (Right) An example delay line allocation that avoids contention at switching or output after arrival of the packet shown on the left

come to the opinion that optical buffers may only need to retain 20 or so packets, provided that assistance is available from upper layers of the network[10]. The view has also emerged that only several tens of delay lines are necessary to handle packets in practice, based on the load and average packet loss probability obtained simply from ITU-T recommendations and statistical data[11]. Therefore, it was decided that optical buffers with such capacities could be developed as a start, in view of actual application.

It should be noted that there are two logical-level problems in optical buffer architecture. One is the deterioration in channel utilization and packet loss probability performance, caused by the provision of delay times solely for discrete values (0, D, $2D$, …); this is the problem of large granularity. The second problem lies in the difficulty of performing interrupt processing such as HOL (Head-of-Line) Priority Queuing.

Below is an actual example of the first problem. Appropriate delay lines are selected for packets arriving from different paths (Fig. 3, left) by buffer management units so that packets do not collide with one another at switching or output. The upper and lower images in the right side of Fig. 3 show the positions of packets A, B, and C relative to the output port when the packets are sent to delay lines d_0, d_2, and d_3. Since the optical buffer only provides discrete delay values, a void is created between two successive packets. This void results in deterioration in channel utilization and packet loss probability.

Furthermore, since the optical buffer has no memory, it will have to feature a function to process a number of packets equal to the number of channels (within the time required to process packets of the minimum packet lengths) in order to be able to handle packets arriving both simultaneously and successively from multiple ports. In other words, this will mean that the processing speed of this function will determine the channel rate and number of channels of the packet switch. Thus, to achieve a 64-port 10-Tbps optical-packet switch with a channel rate of 160 Gbps and the ability to handle a minimum packet length of 64 bytes, the switch must be able to process (i.e. to select delay lines for) 64 packets in 3.2 nanoseconds.

Past studies have shown that wavelength conversion is effective in contention control, and that packet loss by wavelength conversion may be improved by increasing the number of wavelengths. However, we must note that the number of wavelengths that may be used is dependent on the number of wavelengths handled by the adjacent node. As reported by numerous studies in the past, wavelength conversion becomes particularly effective when combined with the use of delay line buffers.

3 Trends in optical packet switching technology

In this chapter, we will first describe the technology behind optical packet switch architecture and present an example of prototype development; we will then summarize present trends in electronic processing technologies. The optical technologies of individual elements are described in more detail in [1] and [2] (in the present issue).

3.1 Switch architecture

The target of research and development of optical packet switches at NICT consists of the dedicated-output-buffer type $N \times N$ packet switches shown in the left part of Fig. 4 (corresponding to the case of $N = 4$). In the figure, the label switch includes the forwarding and switching functions that were presented in Fig. 1, and the buffer includes the buffer management and buffering functions also shown in Fig. 1. The routing function in Fig. 1 was omitted since it did not affect our decision in selecting the present architecture. Other existing types of switch architectures include the input-buffer, shared-feedback-buffer (Fig. 4, right), and shared-output-buffer architectures. Below, we will explain why NICT selected the dedicated-output-buffer architecture.

Compared to input-buffer architecture, output-buffer architecture displays superior delay and throughput characteristics due to the absence of HOL (head of line) blocking. However, this architecture has the disadvantage of requiring a switch with a bus speed N times higher than that of the input-buffer architecture, rendering implementation difficult. In order to avoid the HOL blocking problem, an MIQ (multiple input queue) is under consideration, which can yield logical performance equivalent to that of output-buffer architecture. However, as can be seen from the image on the left-hand side of Fig. 4, the architecture of the optical packet switch in our study is basically a bundle of N $1 \times N$ switches, meaning that there are N channels for a single output port. The resulting performance is equivalent to that achieved using a switch having a bus speed that is N times greater. To avoid contention at the output port, the MIQ requires an (output) arbitration function. The arbitration function here assumes the use of a memory buffer on the input side, and so may not be applicable to cases in which optical fiber delay line buffers are used.

The dedicated-output-buffer architecture is considered superior to shared-output-buffer and shared-feedback-buffer architectures for a number of reasons. First, the dedicated-output-buffer architecture allows fully independent control of the switch and the buffer parts. To be more precise, the process for determining the output port needs only to be performed once at the switch part, allowing full use of the advantages of high-speed optical label processing. Second, the output port will be uniquely determined once the data is sent to the buffer, thus eliminating any possibility of switching to multiple ports from the buffer. This means that buffer control can be performed independently for each output port, allowing for high-speed buffer control. When

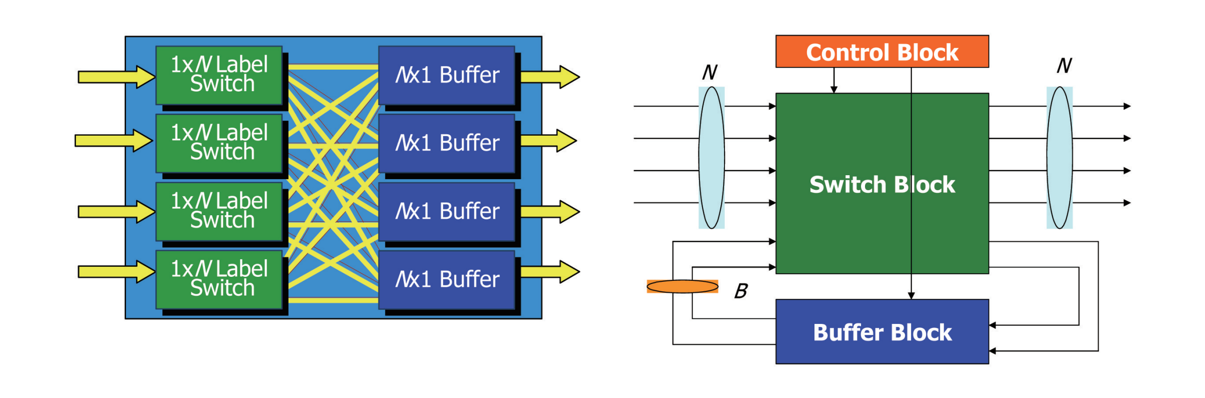

Fig.4 (Left) Architecture of a dedicated-output-buffer optical packet switch (control blocks are present independently within the label switch and buffer); (Right) Architecture of a shared-feedback-buffer optical packet switch

buffer control speed can be completely ignored (i.e., when control time = 0), the shared buffer architecture can yield equivalent performance with fewer buffer resources. However, under the present circumstances, the dedicated-output-buffer architecture with N ports and B delay lines for the buffer produces a buffer management method NB times faster than the shared buffer architecture[12]. Additionally, the dedicated-output-buffer architecture only requires several small-scale switches (such as $1 \times N$, $N \times B$, or $1 \times B$), while the shared buffer architecture requires relatively large switches [such as $(N+B) \times (N+B)$], as seen in the right part of Fig. 4.

Based on the above considerations, NICT has, as mentioned earlier, decided to focus its efforts on developing an architecture embodying the output-buffer method. A number of attempts have been made to develop optical packet switches with 2.4-Gbps or 10-Gbps-based optical buffers[3][4]. A more recent example of such development efforts involved a 40-Gbps-based optical buffer[5]. In 2002, NICT succeeded in developing an optical packet switch featuring optical label processing and an optical buffer for a channel rate of 40 Gbps[7], and in 2005, further succeeded in the development of an optical packet switch for a channel rate of 160-Gbps by modifying the optical-label and optical-buffer processing methods[8].

3.2 Trends in buffer control systems

It is believed that conventional IP and ATM address/label processing techniques may be directly applied to electronic label processing of optical packet switches, and to the best of the authors' knowledge, there have been no reports concerning unique electronic address/label processing techniques proposed specifically for optical packet processing.

Accordingly, from this point on in this section we will mainly deal with trends in buffer management (scheduling).

Generally, studies on buffer management can be grouped broadly into three categories.

(1) Increasing the buffer utilization, which may be further divided into three types.

  (a) Increasing link utilization by reducing voids (Example:[13])

  (b) Increasing fiber utilization through the introduction of wavelength conversion [Example: [14]. [15] is an example of both (a) and (b)]

  (c) Optimization of optical buffer length (granularity) (Example [16])

(2) Development of high-speed processing techniques

(3) Priority queuing

The section below will summarize the current trends in high-speed processing techniques and priority queuing.

• **High-speed processing techniques**

As discussed in the case of electronic routers, increased speeds in contention resolutions are essential to the improvement of throughput performance. The authors have introduced a parallel pipeline architecture consisting of multiple, regularly aligned processors (Fig. 5) for buffer management. In this architecture, a contention resolution method only uses a single step per processor [time complexity O(1) per processor], even when packets arrive simultaneously from all the N ports[12][17]. Pipeline processing in this case consists of a finite ($\log_2 N + 1$) step [time complexity for algorithm for N packets = O(log N)]. In Fig. 5, processors and registers are represented by circles and squares, respectively. Since queue updating is reduced to once per unit processing time (corresponding to the minimum packet length), this method can yield a throughput N times larger than simple round-robin processing of O(N) times per unit processing time. The effectiveness of this method is at once evident when one remembers that the time complexity for Void Filling is given by an increasing function of the number of Voids V [time complexity O(V)], and that the buffer-control algorithm for a shared-feedback-buffer architecture is given by an increasing function of the number of delay lines B within the buffer [time complexity O(B)] to handle a packet. On the other hand, this method requires the use of multiple

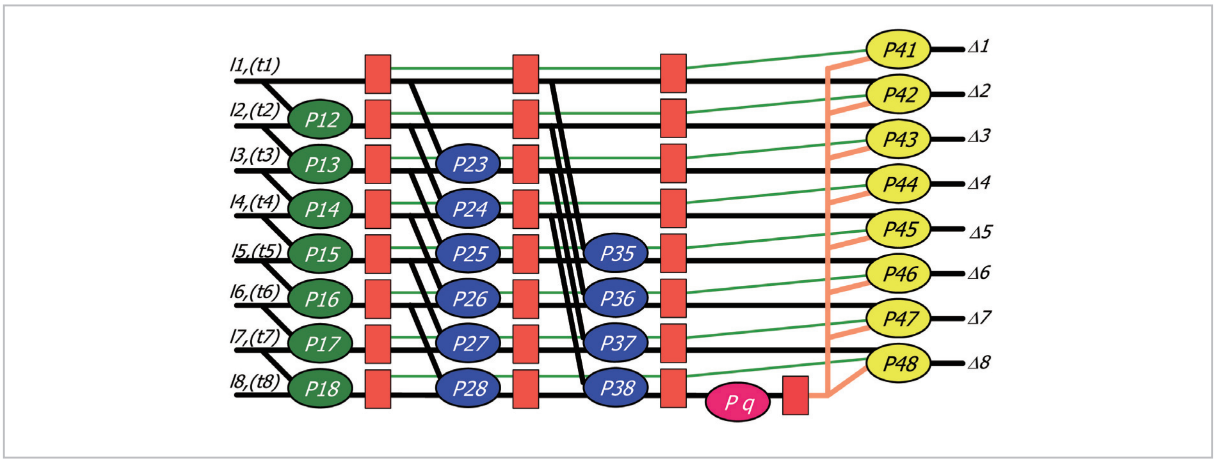

Fig.5 Overview of parallel and pipeline processing architecture for high-speed buffer management

processors, and so it is essential to examine the hardware, particularly with respect to circuit scale. It has been confirmed that the present method is capable of supporting buffer management of a 40-Gbps-based eight-port optical packet switch having a minimum optical packet processing size of 64 bytes, based on gate-level simulation using a 0.22 μm FPGA (Field Programmable Gate Array). Further, this method is also capable of supporting buffer management of 40-Gbps-based 128-port (5.12 Tbps) packet switching, based on logic-cell level estimation of the 0.13 μm FPGA. Hardware that may be implemented in a 160-Gbps-based NICT optical packet switch prototype[8] is also under development, although designed for a synchronous fixed-length switch targeting a different packet length. The present method is capable of providing priority queuing[18] and fairness control[12] with extended architecture.

• Priority queuing

For cases in which sufficient packet loss performance cannot be provided (due, for example, to restrictions in the number of delay lines for the optical buffer), it is possible to enhance performance with packets of a certain class by carrying out priority queuing (i.e., quality differentiation). In[19], the Threshold Dropping (TD) method is applied to the optical buffers, using as threshold values the wavelength conversion value and number of delay lines available for use (this method is synonymous with PBS, or Partial Buffer Sharing). NICT has confirmed that the PBSO (PBS with Overwriting) method, an extended form of PBS that performs priority queuing using only buffers, offers higher performance relative to the simple PBS method[20][21]. Parallel PBS, an extended form of the above parallel pipeline method, can perform compatible high-speed control and priority queuing[18].

4 Concluding remarks

This paper has presented an overview and discussed the architectural technology of optical packet switches targeting >10 Tbps throughput, and also addressed the application of these switches to the Internet environment. We described the performance requirements of each element, as well as trends in the development of control technologies.

In the future, it will be necessary to develop more scalable, more compact, and lower-power models of each component technology for the optical packet switch, in order to provide for actual implementation in an Internet environment. Furthermore, a number of points concerning network architecture must also be examined, as follows.

Interface for optical packets and upper-layer packets (such as IP packets): For example, methods for forming 160-Gbps optical packets from the 10-Gbps IP packets. In[22], a method has been proposed to create optical packets by adding the channel rates to total 160 Gbps using DWDM technology.

Routing: Although optical packet switching in independent networks such as the MPLS is fairly easy, such networks are difficult to grow. In order to ensure the widespread implementation of optical packet switching, it is important to develop systems in which optical packet switches may be used in connection with other types of switching nodes—for example, to develop control technologies such as routing methods[1] that allow for the use of optical packet switch with IP routers, one of the most representative types of switching nodes.

References

1 H. Harai, “Recent R&D activities in optical packet switching and its deployment to networking”, IEICE Technical Report (PN2004-67), pp. 35-40, Dec. 2004(Invited). (in Japanese)

2 N. Wada, “5-1 Research and Development of 160 Gbit/s/port Optical Packet Switch Prototype and Related Technologies”, This Special Issue of NICT Journal.

3 S. J. B.Yoo, F. Xue, Y. Bansal, J. Taylor, Z. Pan, J. Cao, M. Jeon, T. Nady, G. Goncher, K. Boyer, K. Okamoto, S. Kamei, and V. Akella, “High-performance optical-label switching packet routers and smart edge routers for the next-generation Internet”, IEEE Journal on Selected Areas in Communications, Vol. 21, No. 9, pp. 1041-1051, Sep. 2003.

4 K. Habara, H. Sanjo, H. Nishizawa, Y. Yamada, S. Hino, I. Ogawa, and Y. Suzaki, “Large-capacity photonic packet switch prototype using wavelength routing techniques”, IEICE Transactions on Communications, Vol. E83-B, pp. 2304-2311, Oct. 2000.

5 K. Ikezawa,etc., “Development of elemental technologies for an optical packet network”, IEICE Technical Report (PN2005-5), pp. 25-30, April 2005. (in Japanese)

6 K. Kitayama and N. Wada, “Photonic IP routing”, IEEE Photonic Technology Letters, Vol. 11, No. 12, pp. 1689-1691, Dec. 1999.

7 N. Wada, H. Harai, and F. Kubota, “Optical packet switching network based on ultra-fast optical code label processing”, IEICE Transactions on Electronics, Vol. E87-C, No. 7, pp. 1090-1096, July 2004. (Invited)

8 H. Furukawa, N. Wada, and T. Miyazaki, “Demonstration of 160 Gbit/s optical packet switching and buffering based on all-optical code label processing”, IEEE LEOS 18th Annual Meeting, No. MG5, pp. 89-90, Oct. 2005.

9 “WAN packet size distribution”, available from “http://www.nlanr.net/NA/Learn/packetsizes.html”.

10 D. Wischik and N. McKeown, “Part I: Buffer sizes for core routers”, ACM/SIGCOMM Computer Communication Review, Vol. 35, No. 3, July 2005.

11 H. Harai, “Optical packet switching technology, (in Japanese)” 2006 IEICE-PN/PIF Tutorial, Feb. 2006.

12 H. Harai and M. Murata, “High-speed buffer management for 40 Gb/s-based photonic packet switches”, IEEE/ACM Transactions on Networking, Vol. 14, No. 1, pp. 191-204, Feb. 2006.

13 L. Tancevski, S. Yegnanarayanan, G. Castanon, L. Tamil, F. Masetti, and T. McDermott, “Optical routing of asynchronous, variable length packets”, IEEE Journal on Selected Areas in Communications, Vol. 18, pp. 2084-2093, Oct. 2000.

14 A. Ge, L. Tancevski, G. Castanon, and L. S. Tamil, “WDM fiber delay line buffer control for optical packet switching”, in Proceedings of SPIE, Vol.4233 (OptiComm 2000), pp. 247-256, Oct. 2000.

15 T. Yamaguchi, K. Baba, M. Murata, and K. Kitayama, “Scheduling algorithm with consideration to void space reduction in photonic packet switch”, IEICE Transactions on Communications, Vol. E86-B, pp. 2350-2357, Aug. 2003.
16 F. Callegati, “Optical buffers for variable length packets”, IEEE Communications Letters, Vol. 4, pp. 292-294, Sep, 2000.
17 H. Harai, M. Murata, “Buffer management based on a parallel and pipeline mechanism to support 128×128 photonic packet switches with 40 Gbps ports”, IEICE Technical Report (PN2003-7), pp. 31-36, Sep. 2003. (in Japanese)
18 H. Harai, “Parallel and pipeline processing for prioritized buffer management in photonic packet switches”, in Proceedings of HPSR 2004 (The 2004 IEEE Workshop on High-Performance Switching and Routing), pp. 156-161, April 2004.
19 F. Callegati, G. Corazza, and C. Raffaelli, “Exploitation of DWDM for optical packet switching with quality of service guarantees”, IEEE Journal on Selected Areas in Communications, Vol. 20, pp. 190-201, Jan. 2002.
20 H. Harai and M. Murata, “Photonic buffer architecture to support prioritized buffer management for asynchronously arriving variable-length packets”, in Proceedings of ONDM 2003 (The 7th IFIP Working Conference on Optical Network Design and Modelling), pp. 1103-1118, Feb. 2003.
21 H. Harai and M. Murata, “Prioritaized buffer management in photonic packet switches for differentiated services (in Japanese)”, IEICE Technical Report (PS2001-59), pp. 139-144, Dec. 2001.
22 H. Harai, “Skew compensation for multi-wavelength optical packets in photonic packet-switched networks”, in OECC/COIN 2004 Technical Digest (9th Optoelectronics and Communications Conference), pp. 588-589, July 2004.

***HARAI Hiroaki**, Ph.D.*

Reseaerch Manager, Network Architecture Group, New Generation Network Research Center (former: Senior Researcher, Ultrafast Photonic Network Group, Information and Network Systems Department)

Research and Development of Network Architecture, Optical Path Network, and Optical Packet Switching