On an IP-Centric Optical Control Plane

Chunsheng Xin, Yinghua Ye, Ti-Shiang Wang, and Sudhir Dixit,
Nokia Research Center
Chunming Qiao, SUNY Buffalo
Myungsik Yoo, Soongsil University

ABSTRACT

This article presents an experimental study on adopting and integrating the existing IP protocols and mechanisms into an optical network control plane. Although there has been much research effort on the conceptual and functional requirements for the control of optical networks, this article focuses on the design and implementation of an optical control plane. The proposed control plane implements the key functions such as routing, signaling, protection/restoration, and quality of service.

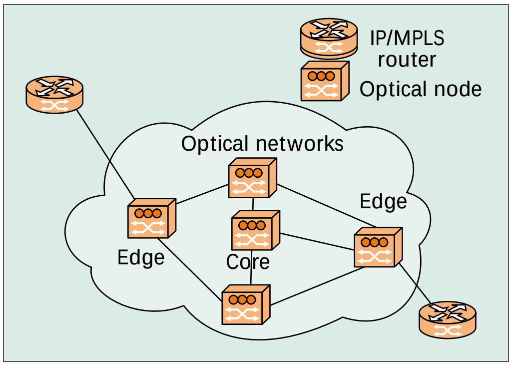
INTRODUCTION

The rapid growth of the Internet and new digital services are driving the demand for huge bandwidth. With technical and economic feasibility, the optical network is becoming an ideal Internet transport infrastructure in core and metro networks due to its potentially unlimited bandwidth (in this article, the term *optical network* refers to a wavelength-division multiplexed, WDM, optical network). Although optical networks are already in use to provide point-to-point connections for a multilayer architecture to transport Internet Protocol (IP) traffic, service providers have experienced high management cost and complexity. Therefore, such a multilayer model is undergoing a "delayering" and moving toward a two-layer architecture which transports IP traffic directly over the optical network. Nevertheless, issues such as rapid and effective bandwidth provisioning and protection/restoration remain quite challenging with this new architecture.

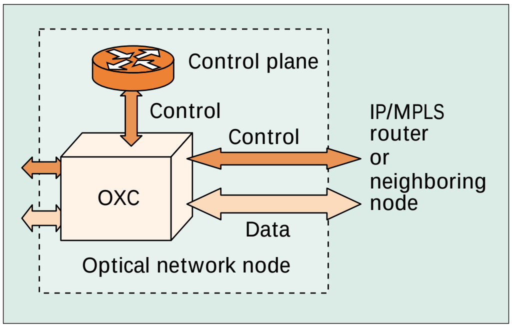
Reference [1] is one of the earliest papers (if not the very first) to propose the use of IP-centric control over optical networks. Since then, there have been many ongoing research activities in the area of IP over optical networks [2–6], and today there is a consensus that the IP routing and signaling protocols can be adapted for the optical network control. In particular, the multiprotocol lambda switching (MPλS) [3] control plane has been proposed for this purpose, which is essentially the multiprotocol label switching (MPLS) control plane with optical extensions. More recently, generalized MPLS (GMPLS) has also been proposed to further extend MPLS to support multiple switching types, for example, packet switching, time-division multiplexed (TDM) switching, lambda switching, and fiber switching [4]. In the MPλS/GMPLS control plane, Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS) is used as the interior gateway protocol (IGP), and Resource Reservation Protocol (RSVP) or Constraint-Based Routing Label Distribution Protocol (CR-LDP) is used as the signaling protocol. However, the components in the MPλS/GMPLS control plane primarily define the functionality and are not restricted to some specific protocols or algorithms, mainly to facilitate the parallel evolution of different components. For further information on the structure of the MPLS/GMPLS control plane, the reader is referred to [2–5].

The objective of this article is to present an experimental study on adopting and integrating existing IP protocols to design a flexible, scalable, and resilient optical control plane. The focus of this article is on the design and implementation issues, not the conceptual and functional requirements and mechanisms studied in most of the existing literature. While the proposed control plane is similar to the MPλS/GMPLS control plane, several additional considerations are introduced in this article. Our motivation is that incremental extensions of existing protocols for reuse may not be the best choice in terms of software complexity and overhead in some cases. Therefore, during the design and implementation, a more careful examination is made of the reusable IP protocol software artifacts to determine necessary extensions as well as possible curtailments. The proposed control plane takes this principle into consideration and augments the ongoing MPλS/GMPLS research efforts.

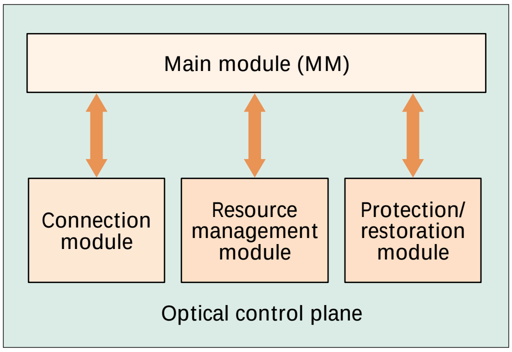
This article is organized as follows. First, we give an overview of the proposed optical control plane architecture and its four modules: the *resource management module* (RMM), the *connection module* (CM), the *protection/restoration module* (PRM), and the *main module* (MM).



■ **Figure 1.** *A mesh optical network.*



■ **Figure 2.** *The architecture of an optical node.*



■ **Figure 3.** *The proposed control plane structure.*

Considering the paramount importance of the survivability in optical networks, the proposed control plane contains an independent PRM for fault monitoring and fast protection/ restoration.

Then we describe, in more detail, the functions of the RMM and CM, respectively. To illustrate the importance of optical network survivability, we discuss the functions of PRM. Finally, we introduce the MM and conclude the current work while suggesting some future research topics.

ARCHITECTURE OVERVIEW

This article considers a mesh optical network used to connect high-speed IP/MPLS client routers in client networks, as shown in Fig. 1. Each optical node consists of an optical crossconnect (OXC) and a control plane, as shown in Fig. 2. The control plane is an IP-based controller that employs IP protocols to operate the underlying OXC. The control plane may be integrated into the same box as the OXC, or a separate router used to control the OXC. Between two neighboring OXCs, a dedicated out-of-band wavelength is preconfigured as the IP connectivity, and a reliable Transmission Control Protocol (TCP) socket connection is set up as the *control channel* to transmit the control messages. A control message is processed and relayed in hop-by-hop fashion.

In the control plane proposed in this article, the generic functions in the MPλS/GMPLS control plane are mapped into different functional modules. As shown in Fig. 3, the proposed control plane consists of four modules: the RMM, CM, PRM, and MM. The RMM is used for routing and wavelength assignment (RWA), topology and resource discovery, and quality of service (QoS) support. The routing protocol in RMM is OSPF-based, similar to that of MPλS/GMPLS, although the proposed control plane does not consider the complex routing hierarchy. CM is used for lightpath signaling and maintenance. In the MPλS/GMPLS control plane, CR-LDP or RSVP is used for signaling. However, in the proposed control plane, a more flexible and efficient signaling protocol based on TCP is implemented for CM. In the documented research efforts on the MPλS/GMPLS control plane, survivability has attracted less attention than other functions. Considering the paramount importance of survivability in optical networks, the proposed control plane contains an independent PRM for fault monitoring and fast protection/restoration. The objective of the MM is to receive the incoming messages and work closely with other modules to process the requests. With the above functional modules, the control plane enables the WDM layer to be extended into a network-layer technology that provides network-layer intelligence to enable the optical network to become a flexible, scalable, and resilient network.

THE RESOURCE MANAGEMENT MODULE

The main functions of the RMM are resource discovery and maintenance, QoS support, and RWA. In RMM, the topology and resource availability information is stored in a database consisting of some tables. The active neighbor information (two neighboring nodes are defined as active neighbors if the physical link between them is in working condition) between a given node and its neighbors is detected by the neighbor discovery mechanism and stored in a *local connectivity vector* (LCV) whose ith element represents the link cost of the active connectivity to the ith node. Link cost 0 indicates the link is in fault or there is no link between the two nodes. A *topology connectivity matrix* (TCM) is used to store the network topology (in OSPF this is stored in the link state database). Figure 4 shows a TCM of an optical network with N nodes. The ith row is the LCV of node i plus a timestamp (a sequence number) which is used to recognize the out-of-date LCV during the flooding procedure.

In RMM, the local resource availability is stored in a *local resource table* (LRT). Table 1 shows the LRT of a node with M ports and W wavelengths per fiber. In Table 1, the node ID (a unique ID to address the optical node) represents the ID of the peering node connected to that port. λ_i status indicates the state of the ith wavelength in the fiber attached to the port. A wavelength can be in one of the following states: *used and preemptable, used*

	Node *I*	....	Node *j*	....	Node *N*	Timestamp
Node *I*	11	....	0	---	8	101
⋮	⋮		⋮		⋮	⋮
Node *i*	0	....	6	....	20	12
⋮	⋮		⋮		⋮	⋮
Node *N*	8	....	11	....	13	899

Figure 4. *A topology connectivity matrix.*

Port 1	Node ID	λ_I status λ_I SRLG list	...	λ_j status λ_j SRLG list	...	λ_w status λ_w SRLG list
...	...	...	...	...	...	...
Port *i*	Node ID	λ_I status λ_I SRLG list	...	λ_j status λ_j SRLG list	...	λ_w status λ_w SRLG list
...	...	...	...	...	...	...
Port *M*	Node ID	λ_I status λ_I SRLG list	...	λ_j status λ_j SRLG list	...	λ_w status λ_w SRLG list

Table 1. *A local resource table.*

and nonpreemptable, *reserved*, *available*, and *faulty*. The used and preemptable state indicates this wavelength is being used by a low-priority lightpath and can be preempted by a high-priority lightpath. The reserved state indicates this wavelength is used by one or multiple protection path(s). The λ_i SRLG list stores the shared risk link group (SRLG) information of the primary paths whose protection paths have reserved this wavelength. It is valid only if the λ_i status is "reserved." An SRLG is an administration group associated with some optical resources that likely share common vulnerability to a single failure [7]. For example, all fibers in the same conduit can be assigned one SRLG identifier because they will all likely fail under a conduit cut. The RMM also maintains a *global resource table* (GRT) consisting of LRTs of all nodes (in OSPF, this is stored in the link state database).

The RMM adopts a similar neighbor discovery mechanism as in OSPF [8]. To discover and maintain active neighbor information, each control plane sends *hello messages* periodically on all outgoing links. If a control plane does not receive a hello message from a neighboring control channel within some initially negotiated hold time, this link is considered failed. The PRM will be invoked to process the fault. On the other hand, if a control plane receives a hello message from a failed link, the link is considered restored. When the LCV is updated due to a link cost change, link failure, or restoration, it is broadcast to the network.

Resource discovery is also based on OSPF [8] and employs a similar flooding mechanism in OSPF to broadcast the LCV and LRT of a given node to the entire network. Each node builds their TCM and GRT using the received LCVs and LRTs. Generally, when there is a change or the update timer expires, the LCV is encapsulated in a *topology update message* and flooded in the network. For the LRT, the wavelength status needs to be broadcast. However, the LRT is generally large and impossible to encapsulate in a single message. In addition, it is not necessary to broadcast the entire LRT whenever there is a small change (e.g., a status change of one wavelength of a port). Hence, alternatively, only the wavelength status that has changed since the last broadcast need be flooded in the network. The status change of one wavelength is specified by the tuple <node ID, port ID, λ status>. Generally, to reduce the message processing overhead, multiple tuples (i.e., multiple wavelength status changes) are encapsulated in a *resource update message*. The λ_i SRLG list will not be encapsulated in the *resource update message* unless the λ_i status is reserved. As in the OSPF flooding mechanism [8], the control plane employs the sequence number (timestamp) from the *topology update message* and *resource update message* to resolve the looping problem.

The RMM utilizes different RWA algorithms to support QoS. Current RWA algorithms do not consider wavelength conversion. However, they can easily be extended to consider this situation. Explicit source routing is realized in the control plane, and the Shortest Path First (SPF) algorithm is used to compute the end-to-end routing path based on TCM. The computed routing path is cached for subsequent requests to save per-request routing computation overhead. The first-fit policy is used to assign an available wavelength for a lightpath based on GRT. In the control plane, three QoS classes have been defined: mission-critical, protection-sensitive, and best-effort. Best-effort service only requires restoration and may be preempted by mission-critical traffic. For a client request in this service class, the RMM computes an end-to-end path using SPF and assigns a wavelength in "available" status along the path. Mission-critical service needs a dedicated protection path, and the establishment of either the primary or protection lightpath may preempt an existing lightpath of best-effort service class. For a mission-critical client request, the primary path is computed first, and an edge-disjoint path is computed as the protection path. Then for each path, RMM assigns a wavelength of available or used and preemptible status along that path.

Protection-sensitive service requires only shared protection; once the primary and protection lightpaths of this service type have been established, neither can be preempted. On the other hand, the establishment of both primary and protection lightpaths cannot preempt any existing lightpaths. Path computation for the protection-sensitive service class is similar to that for the mission-critical service class. The wavelength assignment for the primary path is also similar, except it requires that the assigned wavelength be in available status. However, wavelength assignment for the protection path is more complex because wavelength sharing needs more consideration. With protection-sensitive traffic, two protection paths can share a wavelength if their primary paths are link disjoint. In order to maintain such information, the control plane assigns a unique SRLG ID to each link. Thus, there is an SRLG list for a primary path. For those protection paths that share a wavelength (in reserved status), the SRLG lists of their primary paths are stored in the λ_i SRLG list of this wavelength. When there is a change in the λ_i SRLG list (e.g., adding or deleting an SRLG ID), it is broadcast to the network in the *resource update message*. Such information is collected into the GRT by each node. Thus, during the wavelength assignment for a protection path, the RMM can check for conflict between the SRLG list of its primary path and the λ_i SRLG list of a reserved wavelength in order to determine if this reserved wavelength can be assigned. Therefore, by maintaining a λ_i SRLG list for each reserved wavelength, the RMM can assign a reserved wavelength to a protection path such that multiple protection paths can share the same wavelength.

THE CONNECTION MODULE

The main functions of the CM are lightpath signaling and maintenance. At each node, a *lightpath table* (LT) is maintained by the CM to manage all lightpaths (originating, passing through, and terminating) over the OXC. Table 2 shows an entry of an LT. The lightpath identifier (lightpath ID) includes the source node ID, destination node ID, and a sequence number. The sequence number distinguishes the lightpaths originating at a given node. Thus, a lightpath ID uniquely identifies a lightpath in the entire network. The status attribute indicates the state of a lightpath (creating, reserved, active, or deleted). The QoS type attribute indicates the service class of the traffic. The input/output port ID represents the ID of the incoming/outgoing port of the lightpath in the OXC (the port ID is only unique in each node). The λ ID indicates the assigned wavelength of this lightpath.

The signaling of a lightpath is performed in a hop-by-hop fashion. When the MM receives a client request, it transfers the request to the PRM, where the QoS information is extracted from the request. Then the PRM invokes the RMM for RWA with the QoS parameter. For each path, a *connection request message* is formed including the lightpath ID, lightpath type (primary or protection), routing path, assigned wavelength, and QoS type. If this is a protection path of protection-sensitive service, the SRLG list of its primary path is also included in the message. This message is processed at each hop and then forwarded to the next hop toward the destination node. The destination node then sends an acknowledgment (ACK) back to the source node along the path. If the lightpath setup is blocked at any hop due to resource conflict, a negative ACK (NAK) is sent back to the source node.

At each hop, the lightpath signaling processing can be divided into two parts: resource reservation/release and lightpath state transfer processing. Resource reservation is responsible for reserving

Lightpath ID			Status	QoS type	Input port ID	Output port ID	λ ID
SRC node ID	DEST node ID	SEQ NUM					

Table 2. *A lightpath table entry.*

the wavelength resource; resource release is basically the inverse of resource reservation. The following procedure illustrates resource reservation:

- Determine the input/output ports by the path information.
- If the QoS type is best-effort and the assigned wavelength is available, set the wavelength status to used and preemptible.
- If the QoS type is mission-critical:
  –If the assigned wavelength is available, the wavelength status is set to used and nonpreemptible.
  –If the assigned wavelength is used and preemptible, abort the existing lightpath on this wavelength. The wavelength status is set to used and nonpreemptible.
- If the QoS is protection-sensitive:
  –If it is a primary lightpath and the assigned wavelength is available, set the wavelength status to used and nonpreemptible.
  –If it is a protection lightpath type:
  –If the assigned wavelength is available, set its status to reserved.
  –If the assigned wavelength is reserved, compare the SRLG list in the message with the SRLG list of the assigned wavelength λ_i SRLG list) if there is no SRLG conflict, Add the SRLG list in the message to the SRLG list of the assigned wavelength.

If resource reservation fails (in wavelength status checking or SRLG list checking), a NAK, including the information in the connection request message, is sent back to the upstream node. Otherwise, the CM begins to process the lightpath state transfer. Lightpath state transfer processing is also invoked when the CM receives an ACK or NAK for a lightpath from the downstream node. Figure 5 shows the state transfer of a lightpath. After resource reservation succeeds, the CM allocates an entry in the LT and properly sets all attributes in the entry with the status attribute set to creating. When CM receives an ACK for a lightpath from the downstream node, the status attribute is set to active or reserved depending on whether it is a primary or protection lightpath. A protection lightpath in reserved status becomes active when the PRM has detected a failure on the primary lightpath and invoked the protection process. When the CM receives a NAK for a lightpath from the downstream node during the setup procedure, or gets a teardown or abort message after the lightpath has been established, the resource will be released, and the associated entry in the LT will be cleared.

THE PROTECTION/ RESTORATION MODULE

The PRM provides the functions of the setup coordination of the primary and protection lightpaths, fault detection, and notification. We con-

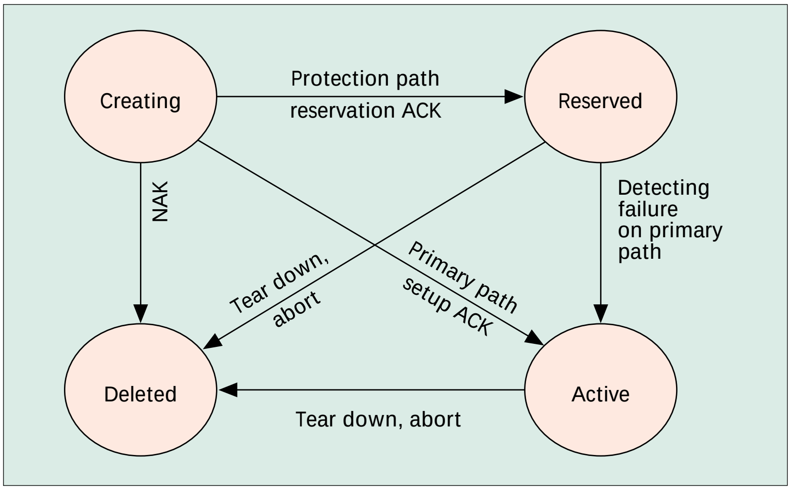


Figure 5. *A lightpath state transfer graph.*

sider end-to-end protection/restoration where a link disjointed path is selected for protection/restoration. Such end-to-end protection/restoration can provide fast recovery without locating the faulty link. Protection methods include fiber-level and channel-level protection. This article considers channel-level protection because it results in high network utilization. Because the mesh network is rich in connectivity, the spare resources can be shared among multiple protection paths as long as their corresponding working paths are link disjoint.

When a client request is received by the MM, it is transferred to the PRM. Then the PRM invokes the RMM for RWA with the QoS parameter extracted from the request. For each path, a *connection request message* is formed, and the CM is invoked for signaling by the PRM. If the client request requires protection, the signaling of the primary and protection paths will be invoked in parallel by the PRM. Hence the PRM (of the source control plane) has to take care of the asynchronous feedbacks (i.e., ACK or NAK) of the two lightpaths. Only if both the primary and protection lightpaths have been set up will the CM send an ACK to the client. Otherwise, when one of the two lightpaths fails (getting NAK) on setup, the CM sends a NAK to the client; if the other lightpath has been set up, that lightpath needs to be torn down.

The PRM is also responsible for detecting the failures and initiating protection/restoration. Fault detection can be done by hardware to detect low-layer impairments such as loss of signal, or at a higher layer via link-probing mechanisms. In the control plane, the neighbor discovery mechanism is used to detect the failure of a link via periodically exchanging hello messages, as discussed earlier. Node failure can be divided into OXC and control plane failure. OXC failure is detected by the control plane through the hardware mechanism. Control plane failure is detected indirectly through the neighbor discovery mechanism. If a control plane fails, every neighboring control plane will not be able to receive hello messages from the failed control plane. As a result, every neighboring control plane broadcasts a topology update message to the network. After receiving all of these topology update messages, every control plane observes that control plane has failed.

After detecting a failure, the control plane will send out a failure notification for each affected lightpath by transmitting a *failure indication signal* (FIS) toward the source node. This notification is relayed hop by hop to the upstream control plane. Once the source control plane receives the FIS, it will check the QoS attribute of the lightpath. If it is best-effort, the source node will calculate and signal a restoration path, and then invoke rerouting to recover the working traffic. If the QoS attribute is protection-sensitive, the source node immediately invokes the setup signaling for the previously reserved protection lightpath. For mission-critical traffic, the destination node can detect the failure of the primary lightpath and automatically turn to the protection path.

The Main Module

The main function of the MM consists of initializing the control plane, waiting for client requests or incoming messages (from neighboring control planes), and invoking other modules to process each message or request. When a control plane is started initially or rebooted (after a failure), the control plane establishes the *control channels* to neighboring control planes (the neighboring information is manually configured) and its clients. It also creates all tables and data structures (LRT, LCV, LT, GRT, TCM, etc.). Then the control plane queries neighboring control planes to update some of these tables such as TCM and GRT. After initialization, the MM will accept requests from clients or neighboring control planes and invoke other modules to process the requests.

Conclusions and Future Work

This article describes the design and implementation of an IP-centric control plane for optical networks. The control plane is capable of several key functions, such as service provisioning, routing, signaling, protection/restoration, and QoS support. In the future, more effective routing algorithms, with other desirable traffic engineering capabilities to improve resource utilization in optical networks, as well as more effective protection/restoration mechanisms will be studied.

References

[1] C. Qiao and M. Yoo, "Optical Burst Switching — A New Paradigm for an Optical Internet," *J. High Speed Net.*, Special Issue on Optical Networking, vol. 8, no. 1, 1999, pp. 69–84.
[2] B. Rajagopalan *et al.*, "IP over Optical Networks: A Framework," Internet draft, draft-many-ip-optical-framework-01.txt, 2000, work in progress.
[3] D. Awduche *et al.*, "Multi-Protocol Lambda Switching: Combining MPLS Traffic Engineering Control With Optical Crossconnects," *IEEE Commun. Mag.*, Mar. 2001, pp. 111–16.
[4] A. Banerjee *et al.*, "Generalized Multiprotocol Label Switching: An Overview of Routing and Management Enhancements," *IEEE Commun. Mag.*, Jan., 2001, pp. 144–50.
[5] Z. Tang *et al.*, "Extensions to CR-LDP for Path Establishment in Optical Networks," Internet draft, draft-tang-crldp-optical-00.txt, 2000, work in progress.

[6] Y. Ye *et al.*, "Design Hybrid Protection Scheme for Survivable Wavelength-Routed Optical Transport Networks," *Proc. Euro. Conf. Net. & Opt. Commun.*, June 2000, pp. 101–8.
[7] J. Strand, A. Chiu, and R. Tkach, "Issues For Routing In The Optical Layer," *IEEE Commun. Mag.*, Feb. 2001, pp. 81–87.
[8] J. Moy, "OSPF v. 2," IETF RFC 2328.

BIOGRAPHIES

CHUNSHENG XIN (xin@cse.buffalo.edu) received his B.S. degree from Wuhan University in 1994 and his M.S. degree from the Institute of Software, Chinese Academy of Sciences, in 1997. He is currently pursuing his Ph.D. at the State University of New York in Buffalo. His research interests include control and provisioning in optical networks.

YINGHUA YE (Yinghua.Ye@nokia.com) received her Ph.D. in electrical engineering in 2000 from City University of New York. Currently she is a senior research engineer in broadband networks at Nokia Research Center. Her major interests include architecture design, network survivability, and real-time provisioning in IP/WDM.

TI-SHIANG WANG [M] (ti-shiang.wang@nokia.com) received his B.S. and M.S. degrees from Automatic Control Engineering, Feng Chia University, Taiwan, in 1986 and 1988. He received his Ph.D. in electrical engineering, Polytechnic University, New York, in 1998. He served as a lecturer at Nan-Kai Institute of Technology, Taichung, Taiwan, in 1991. From 1991 to 1998 his work was focused on control theory, detection and estimation, digital signal processing, and optical packet switching systems areas. Since 1999 he has been working as a senior research engineer and project manager in Nokia Research Center, Boston, Massachusetts. His areas of work are currently in the research and development of IP and WDM networks, optical access networks, terabit optical packet switches, and wireless networks. He has published more than 20 papers in conferences and journals, and also has one patent granted in the optical networking area.

SUDHIR DIXIT (sudhir.dixit@ieee.org) received his B.E. degree from Maulana Azad College of Technology, Bhopal, India, his M.E. degree from Birla Institute of Technology and Science, Pilani, India, and his Ph.D. from the University of Strathclyde, Glasgow, Scotland, all in electrical engineering. He also received his M.B.A. degree from Florida Institute of Technology, Melbourne. He is currently a senior R&D manager and a site manager at Nokia Research Center in Burlington, Massachusetts. His areas of interest are ATM, Internet, optical networks, and mobile/wireless networks. From 1991 to 1996 he was a broadband network architect at NYNEX Science and Technology (now Verizon Communications). Prior to that he held various engineering and management positions at other major companies, such as GTE, Motorola, Wang, Harris, and STL (now Nortel Europe). He has published extensively, given invited talks/panels, and has 20 patents either granted or pending.

CHUNMING QIAO [M] (qiao@computer.org) is currently a tenured associate professor in the Computer and Science Engineering Department of State University of New York at Buffalo. He is also an adjunct professor of the Electrical Engineering Department and director of the Lab for Advanced Network Design, Evaluation and Research (LANDER), which conducts cutting-edge research related to optical networks, wireless/mobile networks, and the Internet. His research has been supported by a number of NSF grants including a Research Initiation Award (IRA) and Information Technology Research (ITR) award, and by Alcatel Corporate Research Center, Nokia Research Center, Nortel Networks, and Telcordia (formerly Bellcore). He has over 10 years of research experience in optical networks, covering the areas of photonic switching devices, WDM communications, and optical packet switching. He pioneered research on the next-generation optical Internet, in particular, optical burst switching (OBS). He has published more than six dozen papers in leading technical journals and conferences, and given several keynote speeches, tutorials and invited talks.

MYUNGSIK YOO [M] (myoo@e.ssu.ac.kr) received his B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, in 1989 and 1991, respectively, and his Ph.D. in electrical engineering from State University of New York at Buffalo in 2000. He was a senior research engineer at Nokia Research Center, Burlington, Massachusetts. He is currently an assistant professor in the school of electronic engineering, Soongsil University, Seoul, Korea. His research interests include optical Internet, WDM optical networks, optical burst switching, QoS, and control and management issues.

In the future, more effective routing algorithms, with other desirable traffic engineering capabilities to improve the resource utilization in optical networks, as well as more effective protection/restoration mechanisms will be studied.