End-to-End Contention Resolution Schemes for an Optical Packet Switching Network With Enhanced Edge Routers

Fei Xue, *Member, IEEE*, Zhong Pan, Yash Bansal, Jing Cao, *Student Member, IEEE, Student Member, OSA*, Minyong Jeon, Katsunari Okamoto, *Senior Member, IEEE*, Shin Kamei, Venkatesh Akella, and S. J. Ben Yoo, *Senior Member, OSA*

***Abstract*—This paper investigates contention resolution schemes for optical packet switching networks from an end-to-end perspective, where the combined exploitation of both core routers and edge routers are highlighted. For the optical-core network, we present the architecture of an optical router to achieve contention resolution in wavelength, time, and space domains. Complementing the solution involving only the core router intelligences, we propose performance enhancement schemes at the network edge, including a traffic-shaping function at the ingress edge and a proper dimensioning of the drop port number at the egress edge. Both schemes prove effective in reducing networkwide packet-loss rates. In particular, scalability performance simulations demonstrate that a considerably low packet-loss rate (0.0001% at load 0.6) is achieved in a 16-wavelength network by incorporating the performance enhancement schemes at the edge with the contention resolution schemes in the core. Further, we develop an field-programmable gate-array (FPGA)-based switch controller and integrate it with enabling optical devices to demonstrate the packet-by-packet contention resolution. Proof-of-principle experiments involving the prototype core router achieve an error-free low-latency contention resolution.**

***Index Terms*—Contention resolution, edge router, optical packet switching (OPS), wavelength conversion, wavelength-division multiplexing (WDM).**

I. Introduction

In RECENT years, the phenomenal growth of Internet traffic and the rapid advance of optical technologies have consistently driven the evolution of Internet architecture. The exponential growth in Internet traffic has continued even during economic recessions, and data traffic has increasingly dominated the network bandwidth requirements worldwide. Meanwhile, wavelength-division-multiplexing (WDM) technology has been widely deployed, providing an attractive platform to exploit the bandwidth potential of fiber links. The current trend of converging data networking and telecommunications networking has influenced the next-generation Internet to be based on two main functional layers: the Internet protocol (IP) layer and the optical WDM layer. Within such an IP-over-WDM architecture, the need for bandwidth efficiency and access flexibility, together with scalability, has led to an emerging research interest in optical packet switching (OPS) technologies [1]–[8]. Changing the switching functionality from electronics to optics can resolve the electrical–optical–electrical conversion bottleneck in optical networks. Further, the OPS technology is envisioned to bridge the gap between the electrical Internet protocol/multiprotocol label switching (IP/MPLS) layer and the optical (WDM) layer and to provide transparency to data protocol and format.

OPS technologies can be classified in terms of fixed- versus variable-length packet, and in terms of synchronous versus asynchronous packet forwarding. Several approaches have been proposed to support fixed-length packet with either synchronous [4], [5] or asynchronous [6] operation. Unfortunately, these fixed-length solutions may cause severe network inefficiency, as they are not well tailored to variable-length IP traffic. The asynchronous variable-length packet switching approach has been discussed in the literature [7], [8] as a means to better support the diverse IP traffic. This scheme simplifies the optical hardware architecture by abolishing synchronization requirements, but it requires effective scheduling algorithms to attain a desirable performance. This work adopts an asynchronous variable-length approach called optical label switching [9], whose key feature is to employ a subcarrier-multiplexed (SCM) optical label carried in band within the same wavelength band and attached to each optical packet. The optical label contains information pertaining to packet forwarding. Based on the optical label, the router forwards an optical packet without converting the baseband data payload into the electrical domain.

Contention resolution schemes are key determinants of packet-loss performance in any packet switching paradigm. In an OPS network, contention arises if two or more packets compete for the same output fiber on the same wavelength at the same time. Especially in an asynchronous variable-length OPS network, how the contention is resolved has a significant effect on network performance. In an electrical packet network, the contention is typically resolved by the store-and-forward buffer queueing. Due to the lack of viable optical memories,

Manuscript received December 1, 2002; revised June 23, 2003. This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory under Agreement F30602-00-2-0543, by the National Science Foundation under Grant ANI-998665, by BellSouth, Cisco, Fitel, Fujitsu, Furukawa, New Focus, and Sprint, by the Optoelectronics Industry Development Association (OIDA) through equipment support, and by the California Microelectronic Innovation and Computer Research Opportunities (MICRO) program through matching support.

F. Xue, Z. Pan, Y. Bansal, J. Cao, M. Jeon, V. Akella, and S. J. B. Yoo are with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616 USA (e-mail: yoo@ece.ucdavis.edu).

K. Okamoto is with the Okamoto Laboratory, NTT Electronics Corporation, Ibaraki 311-0122, Japan.

S. Kamei is with NTT Photonics Laboratories, Kanagawa 243-0198, Japan.

Digital Object Identifier 10.1109/JLT.2003.819560

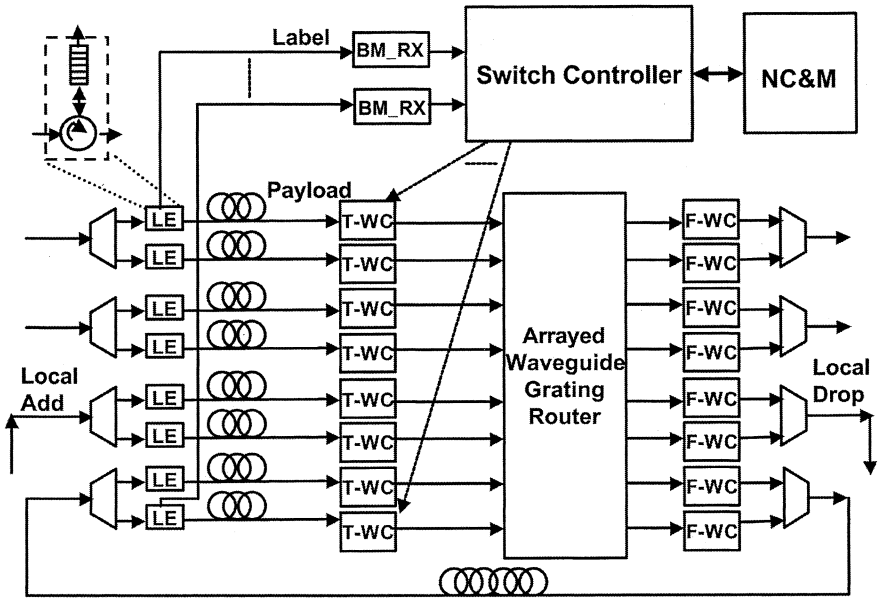

Fig. 1. Proposed optical router architecture (LE: label extractor; BM_RX: burst-mode receiver; T-WC: tunable wavelength converter; F-WC: fixed wavelength converter; NC&M: network control and management system).

an OPS network must take different approaches. Previous studies have proposed various opitical core router architectures and contention resolution algorithms, e.g., fiber delay line (FDL) [10], [11], wavelength conversion [12], and deflection routing [13]. Recent work [14] has compared various combinations of contention resolution schemes and demonstrated their performance. This paper will investigate an end-to-end contention resolution solution, which involves not only the contention resolution schemes in optical core routers, but also the functionalities of electrical edge routers. We will discuss the node architecture of the optical core router to take advantage of the wavelength domain afforded by the WDM technology. The core-oriented solution accomplishes contention resolution in the wavelength, time, and space domains. Furthermore, this paper proposes performance enhancement schemes implemented at network edges. We highlight a traffic-shaping function at the ingress edge and a proper dimensioning of the number of local drop ports at the egress edge through simulation experiments.

The paper is structured as follows. Section II presents the architecture of the optical router and discusses its contention resolution scheme in the wavelength, time, and space domains. Section III discusses the enhancement schemes of the edge router and evaluates their effects on end-to-end network performances. Section IV describes the logic behavior and the functional design of an FPGA-based switch controller. Experimental demonstration of the packet-by-packet contention resolution is presented in Section V. The paper concludes with the summary in Section VI.

II. Optical Router Architecture and Contention Resolution

This section describes the architecture of an optical router and details the core-oriented contention resolution scheme. The architecture and schemes presented are used throughout the paper.

A. Optical Router Architecture

Fig. 1 depicts an overall architecture of the optical router discussed in this paper. The router has K fiber inputs/outputs (I/O) with each carrying W transport channels. We call them *external wavelengths* to distinguish them from *internal wavelengths*, which are used to forward packets through a switch fabric. The switch fabric consists of an array of tunable wavelength converters (T_WCs), an arrayed-waveguide grating router (AWGR), and an array of fixed-wavelength converters (F_WCs). The AWGR allows nonblocking wavelength-routed interconnection and supports wavelength-to-space mapping, where the individual output port can be addressed by proper selection of a wavelength at the input port. Based on its wavelength-dependent routing characteristics [15], the AWGR can forward incoming packets to their desired output ports by simply choosing appropriate internal wavelengths. As Fig. 1 shows, the T_WCs are manipulated by a switch controller to achieve this function. To avoid signal overlapping at the output fiber, the F_WC is necessary to convert the internal wavelength into the desired external wavelength. This AWGR-based switch fabric is strictly nonblocking, providing routing of any input wavelength of any input fiber to any output wavelength of any output fiber. In the optical router, a subset of the K I/O fiber ports is used to build the recirculating FDLs, which offers a fixed and finite amount of delay to provide sequential buffering. In addition, the optical router includes local add–drop ports to ingress/egress traffic from/to local client networks.

In the control plane, the optical router senses an asynchronous packet arrival and taps off an optical label from the attached packet by a label extractor (LE). A burst-mode receiver (BM_RX) recovers the optical label and converts it into the electrical label signal. The optical router makes the forwarding decision based on the label content and the forwarding table. The controller sends the control signal to the corresponding T_WC to forward the packet through the AWGR. A network

control and management system (NC&M) interfaces with the controller to update the forwarding table and to collect network statistics.

B. Core-Oriented Contention Resolution Schemes

The optical router exploits the wavelength, time, and space domains to resolve the contention by means of wavelength converters, FDLs, and optical switch fabric.

Wavelength conversion offers effective contention resolution without relying on buffer memory. In an OPS router, the tunable wavelength converter shifts the wavelength of a contending packet to an available internal wavelength, enabling conflict-free forwarding to its desired output fiber. Since wavelength conversion does not cause extra packet latency, jitter, or packet reordering problems, it is the preferred contention resolution scheme in an OPS network.

The time-domain contention resolution resorts to FDLs to offer a fixed and finite number of delays, partially imitating electronic random access memory. A contending packet enters the FDL at the output of a switch fabric and loops back to the input port after traversing the entire delay line. As Fig. 1 shows, the optical router combines the wavelength conversion with the FDL buffer so that the contending packet can be converted to any free wavelength within the FDL buffer.

Space deflection relies on the neighboring node to route the packet when contention occurs, with the expectation that other optical routers will eventually forward contending packets to their destinations. The effectiveness of deflection routing heavily depends on the network topology and the offered traffic matrix. A meshed topology with a high number of links may obtain a larger gain from deflection routing than is the case with a simpler network. Most deflection networks implement certain mechanisms to mitigate or prevent looping; one example is to set a maximum hop or deflection count for each packet.

To combine these schemes for an integrated solution, an optical router takes the following order of precedence to resolve contention: wavelength conversion, optical buffering, and space deflection. In this process, when contention occurs at one specific output fiber, the optical router will first attempt to resolve contention in the wavelength domain by seeking an alternative vacant wavelength at that output fiber. If none is available, the router will forward the packet to an available FDL for buffer. If there is no free FDL, the router will resort to the space domain by deflecting the contending packet to a secondary preferred output port. When all options fail, the packet will be discarded. In addition, the OPS network adopts a mechanism called *optical time to live* (OTTL) [16] to prevent an optical packet from being repeatedly deflected in the network or recirculated through the FDLs, by taking into account the physical impairments induced by optical devices when a packet travels through its all-optical data plane. The OTTL measures the remaining lifetime of a packet by directly monitoring its signal quality in the optical layer. An optical packet will also be discarded once its OTTL expires.

III. Performance of End-to-End Contention Resolution

In an end-to-end connectivity view, an edge router represents the access point between an OPS network and any legacy network domain. To validly assess the contention resolution schemes, we need to consider an OPS network integrated with the edge router support. In principle, an optical router performs three types of packet forwarding: forwarding transit packets, adding local packets from the ingress edge, and dropping local packets to the egress edge. A transit packet must cope with possible contention from local packets as well as other transit packets. As discussed hereafter, the enhanced edge router proposed in this work can effectively reduce the chance of contention between local packets and transit packets, thus significantly improving the effectiveness of contention resolution schemes in the core.

A. Traffic Shaping at Ingress Edge

Since an ingress edge router is the last stage to regulate the legacy network traffic before entering a core OPS network, its functional design should take into consideration the intrinsic characteristics of this traffic. Extensive traffic analyses [17] have revealed that Internet traffic is of a bursty nature and exhibits self-similarity. Meanwhile, statistical traffic data [18] show that nearly half of the IP packets are 40–52 B in length and that the IP packet length follows a characteristic distribution with peaks at 40, 576, and 1500 B. Some initial studies [19], [20] have observed that the packet-loss performance of an OPS network is degraded dramatically under self-similar traffic, mainly resulting from the frequent contention conditions occurring at the switching nodes. Due to the self-similar traffic pattern and the irregular packet-size distribution, the commonly used contention resolution schemes alone may not be sufficiently effective to maintain reasonable network performance under high loads. This observation suggests a desirable traffic-shaping function at the network edge to reshape the incoming traffic profile and to regulate the packet flow.

A packet-aggregation mechanism provides an effective way to achieve this traffic-shaping function, which enables an ingress edge router to assemble jumbo optical packets from the client IP packets of the same egress destination and of common attributes. The edge router sorts the incoming client packets into their corresponding assembly queues. The creation of an optical packet is triggered by a parameter called maximum payload size (MPS, in bytes), which sets an upper boundary on the length of optical packets. The edge router will assemble an optical packet when the buffer occupancy of an assembly queue reaches the MPS. In addition, the aggregation mechanism adopts a time-out period to avoid excessive queuing delay, after which it will also generate an optical packet even if the MPS value is not reached. Once an optical packet departs the assembly queue, it enters the ingress transmission buffer, which is designated to a specific local add port of the optical router. An optical packet ingresses from the edge to the optical router only when there is a vacant wavelength on the preferred output fiber of the optical router. This buffering mechanism avoids possible contention between transit packets and local add packets.

B. Simulation Studies on Performance Evaluation

This paper presents simulation studies driven by self-similar traffic with a realistic IP packet-length distribution to investi-

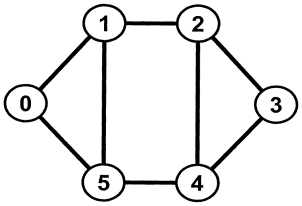

Fig. 2. Simulation network topology.

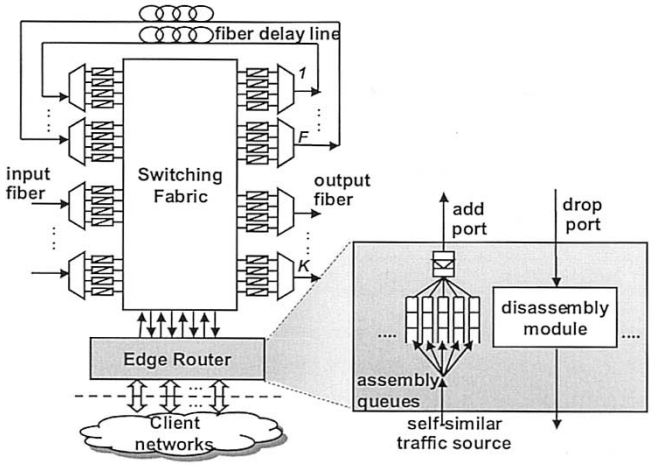

Fig. 3. Nodal architecture of optical packet router for simulations.

gate the effect of the traffic-shaping function. Rather than considering one single optical router, this paper simulates an OPS network with core-oriented contention resolution capabilities, where the networkwide packet-loss rate (PLR) is evaluated.

In the simulated network shown in Fig. 2, each WDM fiber link carries four wavelengths ($W = 4$) transmitting at 2.5 Gb/s. The length of each fiber link is 20 km, inducing a 100-μs propagation delay for each packet hop. As Fig. 3 depicts, each simulated node represents one optical router connecting with the edge router via four local add–drop ports; each local port has the line speed of 2.5 Gb/s. For a specific local add port, there is one dedicated traffic source to generate IP packets, which follow a realistic IP packet-length distribution (the average packet size $=$ 404.5 B, the maximum packet length $=$ 1500 B). This work adopts a self-similar model called "Sup_FRP" [21] to generate packet traces with Hurst parameter $H = 0.8$ and fractal onset time scale $\mathrm{FOTS} = 100$ μs. This model is capable of describing both short-time and long-term correlation structures of a self-similar traffic process. The other important statistical properties of generated traffic traces, such as the coefficient of variation and the autocorrelation function, will be presented later in this subsection, with a detailed discussion of the traffic analysis. For a specific ingress node, its generated IP traffic is equally dispersed into the other egress nodes, i.e., a uniform traffic matrix is used in this work. The IP packets go through the traffic shaper (packet assembly queues) within the edge router to be aggregated into optical packets, and the local transmitter sends the generated optical packets to the OPS network. The load ρ of the local transmitter, defined as the ratio between the total numbers of bits offered per unit time and the line speed, varies from 0.3 to 0.7. In the simulation, when a packet arrives at the optical router, the router will first determine its preferred output fiber and use the wavelength converter to choose an available wavelength on that preferred output fiber whenever one exists. If none is available, the switch will forward the packet to the available FDL buffer. The holding time of the FDL is equal to the maximum optical packet length, and the number of FDLs is set to $F = D - 1$ (here, D is the node degree of the optical router). If there is no free FDL, the packet will be deflected to a secondary output port predefined in the simulation. When all options fail, the packet will be discarded. To mimic the OTTL described earlier, we set a maximum hop count to limit how many hops an optical packet can travel. Note that each time the packet goes through the FDL or the packet is transmitted from one node to another, it is counted as one hop. By using this method, we take into account the physical impairment that a packet suffers when it travels through a physical device. The maximum hop count is set at five. Since the uniform traffic matrix is used, all assembly queues at each edge router have the same time-out period. We set this time-out period T to be $T = 2 \times d$, where d defined as average payload fill time is given by the MPS value divided by the traffic arrival rate for a particular destination.

To evaluate the effects of the packet-aggregation mechanism on the traffic patterns, this work simulates traffic processes on network links and analyzes their characteristics. As previous studies [22]–[24] have observed, for systems with limited buffers, the network performance is dominated by the short-term correlation structure of the traffic process. Thus, the traffic analyses conducted here focus on the short-term correlation structure. To estimate the burstiness of a traffic process, we adopt the coefficient of variation (CoV) and the first-lag autocorrelation function (ACF(1)) for inter-arrival time to capture both the first-order and second-order properties [25], [26]. Here, the CoV is defined as the ratio of the standard deviation (STD) to the mean value of the inter-arrival time, and the ACF(1) describes the correlation between a sample and its previous one in a process of inter-arrival time. The analysis results presented here are on the traffic processes measured from one wavelength channel of the link from node 1 to node 2 at the load 0.6. Table I summarizes the statistical properties of the traffic processes in terms of both the inter-arrival time and the packet length. We observe that the packet-aggregation mechanism results in lower values of CoV and ACF(1) than those in the cases of "no shaper," thus reducing the burstiness of the raw input traffic processes. As MPS values increases the aggregated traffic tends to be smoother. In addition, the packet-length analyses indicate tighter distributions for the jumbo optical packets than those for raw IP packets. All these results demonstrate that the packet-aggregation mechanism is capable of shaping the raw input traffic to a smoother arrival pattern with more regular sized packets.

Fig. 4 shows the networkwide PLRs plotted against the load of the local transmitter. It should be pointed out that all PLR calculations in this study are based on the IP packet loss. When an optical packet containing multiple IP packets is discarded or successfully transferred, the simulation will record the number of IP packets encapsulated in this optical packet. As a result, the

TABLE I
CHARACTERISTICS OF TRAFFIC PROCESS (LOAD $\rho = 0.60$)

		no shaper	MPS= 3000	MPS=9000
Inter-arrival Time (μs)	min	0.13	2.56	17.70
	max	51.10	93.50	239.93
	mean	2.12	13.13	45.18
	STD	2.04	4.82	14.17
	CoV	0.962	0.367	0.314
	ACF(1)	0.170	0.0406	-0.0997
Packet Length (Bytes)	min	40	277	4185
	max	1500	3000	9000
	mean	405.69	2462.2	8470.4
	STD	512.50	455.01	428.69
	CoV	1.26	0.19	0.051

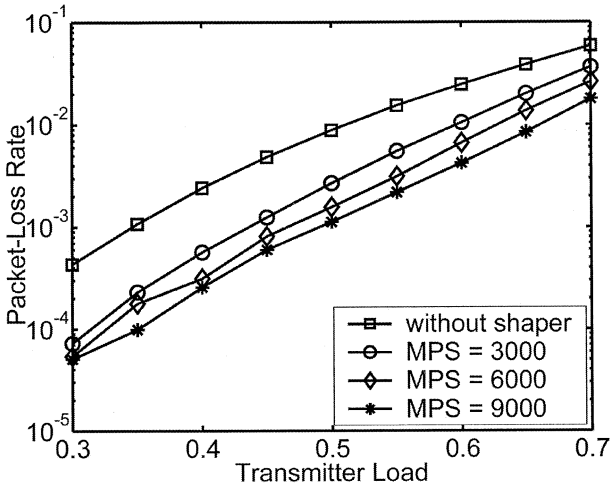

Fig. 4. PLRs in simulated OPS networks.

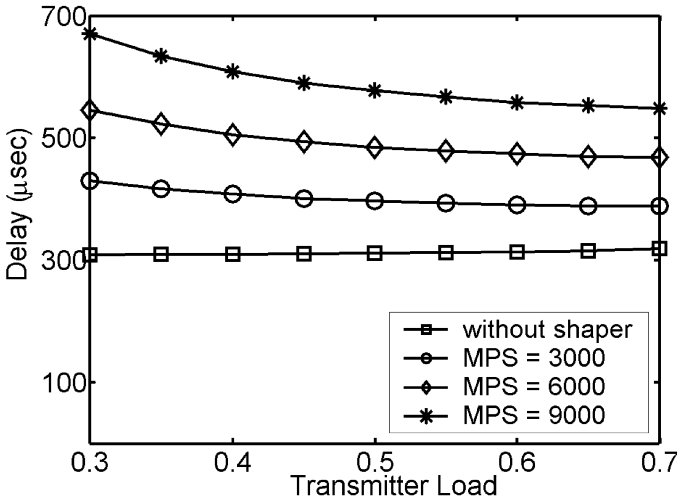

Fig. 5. Average end-to-end delays for flow from node 0 to node 3 in simulated networks.

simulation can measure the number of IP packets successfully transferred through the network (nt) and those discarded within the network (nd). The PLR is then calculated as the ratio of nd to the sum of nd and nt. As Fig. 4 shows, the traffic-shaping function results in noticeably smaller PLRs than in the baseline network without the traffic-shaping function, and this improvement becomes greater as the MPS value increases. For instance, the PLR of the baseline network without traffic shaper is about 0.88% at the load 0.5, but with the help of traffic shaping, it can decrease to 0.27% and 0.17% for $\mathrm{MPS} = 3000$ and 6000, respectively, and the PLR 0.12% can be achieved when $\mathrm{MPS} = 9000$, indicating a several-fold benefit of traffic-shaping in reducing the PLRs. This observation is well in line with the favorable effect of the packet-aggregation function on the traffic characteristics, as discussed previously.

The reduction of the networkwide PLR is achieved at the expense of an extra assembly delay at the edge router. We take a traffic flow from node 0 to 3 as an example to discuss the delay performance. Note that each network hop induces a propagation delay of 100 μs. Fig. 5 shows the mean values of the end-to-end delay for this flow. We observe that the induced extra assembly delays are less than 400 μs for all cases, primarily because the assembly delay is explicitly limited by the time-out period. Further, it is evident that a higher traffic load results in a smaller assembly delay. Thus, the extra delay imposed by the packet-aggregation mechanism is insignificant.

C. Local Drop Port Dimensioning at the Egress Edge

In an OPS network, optical packets are forwarded from optical routers to egress edge routers through local drop ports. In principle, the number of the local drop ports limits the transmission capacity from the optical core to the electrical edge. When no free drop port is available, a local packet will remain in the optical domain and continue utilizing optical resources, thereby reducing the effectiveness of the contention resolution schemes in the core. Intuitively, it is always beneficial to packet-loss performance to increase the number of local drop ports. However, this increase will introduce additional hardware requirements and a higher cost to routing systems. To balance this tradeoff, we present a simulation-based approach to achieving a proper dimensioning of the number of local drop ports based on their usage patterns.

The simulation setup is similar to the one described in Section III-B except that each core router is assumed to have $W \times D$ drop ports, instead of a fixed number (i.e., four). Recall that the variable W refers to the number of wavelengths in each fiber

link and that D is the node degree of a core router. This assumption guarantees that a core router will always have free ports available to drop local packets, even in an extreme case when $W \times D$ local packets simultaneously arrive from its neighboring nodes. In the simulation, each core router will continuously monitor how many local drop ports are in use (hereafter denominated as U) and take a statistical sample when a local packet arrives. Based on these samples, we obtain an estimated distribution of U for each router. Fig. 6 shows the cumulative distribution functions (CDF) obtained for node 1. Here the CDF $F(u) = Pr[U \leq u]$ measures the probability that the number of local drop ports in use is less than or equal to u. For instance, at load 0.5 and 0.6, the probability of four or fewer drop ports being in use at any time (i.e., $F(4)$) is nearly 0.86 and 0.78, respectively, and this value decreases to 0.69 when the load is 0.7, reflecting that a higher load requires a heavier usage of drop ports. For proper dimensioning, the number of local drop ports should be large enough to accommodate a load with high probability. Based on the distributions shown in Fig. 6, eight drop ports appears to be an appropriate choice for node 1 to balance the tradeoff between cost and performance, because the probability of more than eight drop ports being simultaneously used is negligibly small (e.g., less than 0.0008 at the load 0.5). For any given traffic matrix and network topology, we can adopt this approach to obtain a reasonable dimensioning by taking into account the usage distribution of local drop ports.

To demonstrate the benefit of the increased number of drop ports to packet-loss performance, we configure each core router with eight local drop ports and compare the simulation results with those obtained in Section III-B. As Fig. 7 shows, the achieved PLRs with eight drop ports are notably lower than those with four drop ports, in both the networks with traffic shapers and without traffic shapers. Since a larger number of drop ports provides a higher transmission capacity to dispatching the local packets to the egress edge, the optical resources in the core router can be more dedicated to resolving contention among transit packets, thus improving their effectiveness and enhancing network performance.

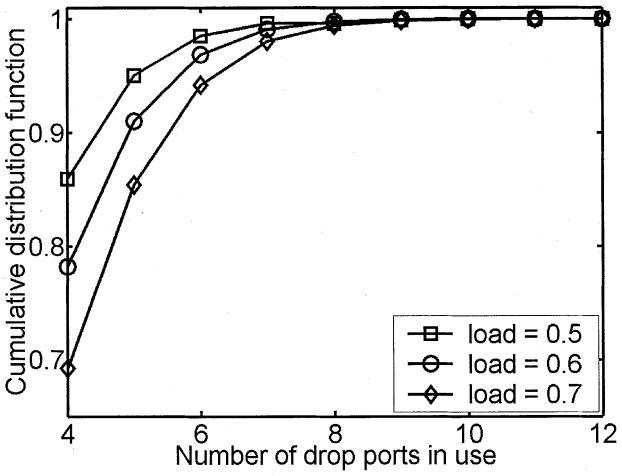

Fig. 6. CDF of the number of drop ports in use.

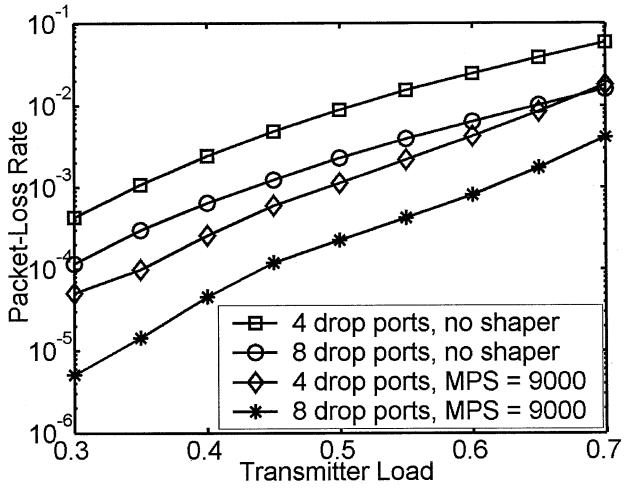

Fig. 7. PLR for networks with various configurations.

D. Scaling Efficiency of OPS Networks

OPS technologies must pursue a highly scalable solution to meet the long-term demands of the next-generation Internet. As discussed previously, optical routers can take advantage of the wavelength domain for contention resolution by using the wavelength conversion, and such an exploitation of natural parallelism in the wavelength domain promises an impressive potential for the scalability of the OPS networks.

To demonstrate the scaling efficiency of the OPS networks, this work compares four simulation scenarios by varying the number of wavelengths in each fiber link W, from 4, 8, 12, to 16. Simulations are based on the six-node network with the uniform traffic matrix described in Section III-B. Note that we scale the number of the local traffic sources in accordance with the wavelength number W, i.e., each optical router will connect with W local traffic sources through the ingress edge in the case of W wavelengths. The load is normalized for each transmitter, and the number of transmitters at each optical router increases in proportion to an increased number of wavelengths. The optical routers are capable of resolving contention in wavelength, time, and space domains, as outlined in Section II. The ingress edge implements the traffic shaping function (with $\mathrm{MPS} = 9000$ B), and the egress edge has a sufficient number of drop ports to accommodate local packets with high probability.

Fig. 8 shows the simulation results. It is evident that the scaled-up OPS network achieves a significant improvement in packet-loss performance, especially in the cases with light and medium loads. For instance, the achieved PLR is below 10^{-4} with the load 0.6 when $W = 8$, and this value decreases to be close to 10^{-6} when $W = 16$. As simulation results indicate, in certain cases, a one-or-two-orders-of-magnitude reduction of PLRs appears achievable by doubling the network resource dimension, even when the total traffic volume of the network is also doubled. In particular, a larger number of wavelengths results in more opportunities to relieve the contention through wavelength conversion, implying a scalable contention resolution scheme. These results also reveal that an OPS network with a large wavelength domain can achieve a considerably low PLR by combining the enhanced edge routers with the contention resolution scheme in the core.

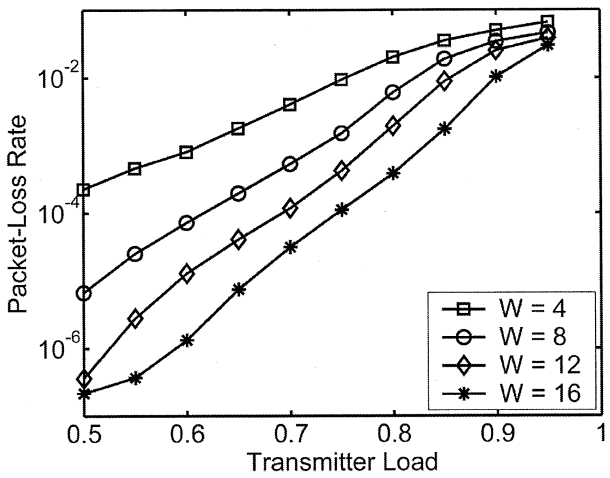

Fig. 8. PLR for networks with various numbers of the wavelengths.

IV. The Optical Switch Controller

This section discusses the design and implementation of an optical switch controller to support the aforementioned contention resolution schemes. For an optical label switching packet router, the packet forwarding decision is solely based on the label content and is independent of the data-payload format and bit rate, indicating a significant design advantage compared with conventional electrical routers, which make forwarding decisions at payload bit rates. In addition, the asynchronous and unslotted packet forwarding considered in this paper requires no complex segmentation and reassembly functions at high payload rates, thus facilitating controller implementation.

This work develops a centralized controller based on a commercial off-the-shelf FPGA to support the OC-3 (Optical Carrier-3: 155 Mb/s) label bit rate and a 16×16 switch fabric ($K = W = 4$). Within the controller, three major data structures are involved in the forwarding decision.

- *Preferred routing table* defines which output fiber is primarily used to transmit an optical packet based on its label content.
- *Deflection routing table* defines which output fiber is assigned to deflect a packet when requiring a space domain contention resolution.
- One dedicated *scoreboard* for each output fiber continuously monitors the status of the wavelength usage.

We describe the functional behaviors of the controller in the context of the optical router architecture (see Fig. 1) and discuss its main functional blocks, including *label-processing modules* and *arbiters*. When an optical packet arrives, the BM_RX converts the extracted label into an electrical signal and sends it to the controller. Within the controller, the label-processing module determines the preferred output fiber by looking up the preferred routing table and sends a forwarding request to the arbiter. Each output fiber has one dedicated arbiter, which is responsible for deciding whether to accept or reject the forwarding request from the label-processing module, based on the wavelength usage information in the scoreboard. If the request is accepted, the controller selects a proper wavelength on the preferred output fiber and then sends a control signal to the correct tunable wavelength converter (T_WC). If the preferred output arbiter rejects the request due to lack of a free wavelength, the label-processing module sends a similar request to the next arbiter based on the time- and space-domain contention resolution schemes. In the implementation, the control logic is pipelined for label processing, output fiber arbitration, and wavelength selection. The experiments described in Section V demonstrate that this FPGA-based switch controller achieves low forwarding latencies of 260 ns.

V. Experimental Demonstration of Packet-by-Packet Contention Resolution

This section presents a prototype optical routing system by integrating the FPGA-based switch controller with enabling optical subsystems and devices. Based on this prototype system, we conduct experiments to demonstrate the contention resolution schemes described in the preceding sections.

The experiment uses an optical label switching scheme, where each packet consists of a baseband payload (1536 b at 2.5 Gb/s) and a subcarrier label (32 b at 155 Mb/s) shifted to 14 GHz. Fig. 9 illustrates the experimental setup. The SCM transmitter employs a radio frequency (RF) circuit and a modulator to generate the packets. The label extractor uses a combination of narrow-band fiber Bragg grating and circulator to strip the label [27]. The BM_RX converts the extracted label into an electrical signal and sends it to the switch controller, which makes a switching decision and instructs the fast tunable laser diode to tune its wavelength to the target value. The semiconductor optical amplifier (SOA) converts the payload to the target wavelength by cross-gain modulation (XGM). Wavelength switching is mapped to space switching by the AWGR. The Mach–Zehnder interferometer wavelength converter (MZI WC) performs fixed-wavelength conversion on payloads after the AWGR by SOA cross-phase modulation (XPM), converting the payload back to one of the WDM wavelengths.

The experiments test three scenarios to demonstrate the wavelength-, time-, and space-domain contention resolution, respectively. For the following discussion, $(m, n)_{\text{in}}$ represents wavelength n on input fiber m, and $(m, n)_{\text{out}}$ represents wavelength n on output fiber m. Two packet streams arrive at $(1, 1)_{\text{in}}$ and $(2, 1)_{\text{in}}$, with the former arriving 50 ns earlier than the latter. At zero delay (0T), packet P1 with label L1 (L1 means that the destination is any wavelength on output fiber 1, denoted by $(1,^*)_{\text{out}}$) arrives at $(1, 1)_{\text{in}}$ and obtains its preferred path $(1, 1)_{\text{out}}$. After 50 ns, P1' with L1 arrives at $(2, 1)_{\text{in}}$. The controller finds $(1, 1)_{\text{out}}$ occupied, and thus contention arises. The controller resolves the contention either in the wavelength domain by converting the packet to another wavelength on the same output fiber $(1, 2)_{\text{out}}$, or in the time domain by sending the packet to a fiber delay line $(4, 1)_{\text{out}}$, or in the space domain by sending the packet to a deflecting route $(3, 1)_{\text{out}}$. At one unit delay (1T), P2 with L2 (destination $(2,^*)_{\text{out}}$) arrives at $(1, 1)_{\text{in}}$ and gets its preferred path $(2, 1)_{\text{out}}$. 50 ns later, P2' with L1 arrives at $(2, 1)_{\text{in}}$ and gets its preferred path $(1, 1)_{\text{out}}$ without contention. For the wavelength- and space-domain scenarios, packet patterns repeat from here. For the time-domain scenario, at two unit delay (2T), the packet previously sent to the delay

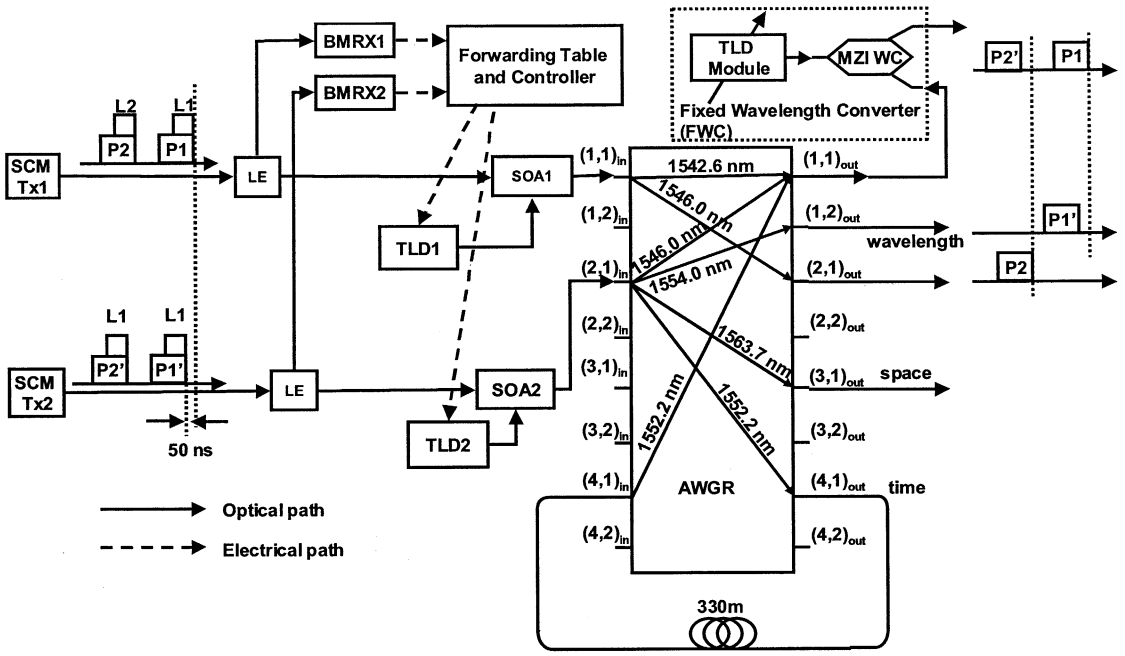

Fig. 9. Setup of the wavelength-domain contention resolution. SCM Tx = Subcarrier transmitter. LE = Label extractor. BMRX = Burst-mode receiver. TLD = Tunable laser diode. SOA = Semiconductor optical amplifier. AWGR = Arrayed-waveguide grating router. MZI WC = Mach–Zehnder interferometer wavelength converter.

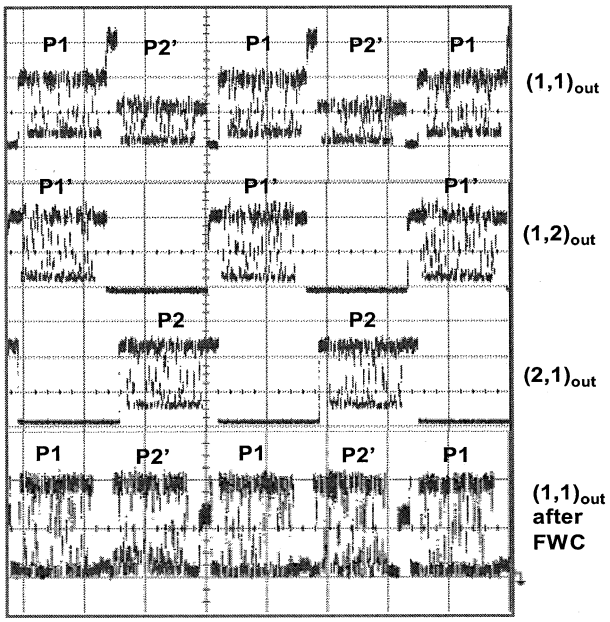

Fig. 10. Results for wavelength-domain contention resolution: traces at output ports. (FWC = Fixed Wavelength Converter.)

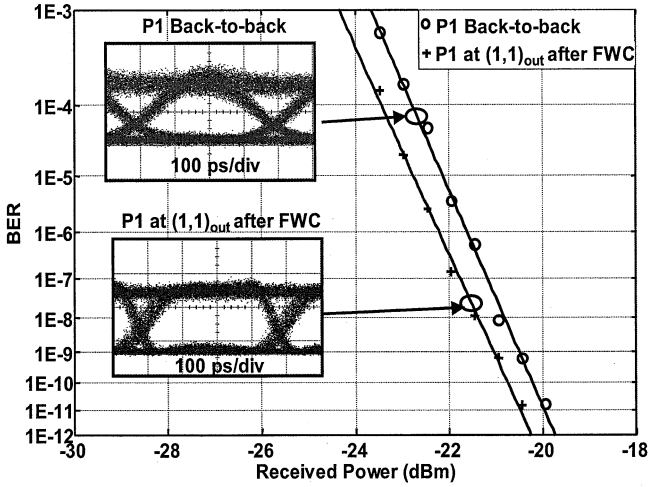

Fig. 11. Results for wavelength-domain contention resolution: BER and eye diagrams. (FWC = Fixed Wavelength Converter.)

line comes back at $(4,1)_{\text{in}}$ and goes to $(1,1)_{\text{out}}$, while P3 with L2 arrives at $(1,1)_{\text{in}}$ and goes to $(2,1)_{\text{out}}$, and P3' with L3 arrives at $(2,1)_{\text{in}}$ and goes to $(3,1)_{\text{out}}$. Then, the packet pattern repeats. The experiment provides a proof-of-principle demonstration of packet-by-packet contention resolution in wavelength, time, and space domains without using the full switching fabric architecture of Fig. 1, which requires a full set of tunable- and fixed-wavelength converters.

We present the results from the wavelength-domain contention resolution as an example. The payloads contain $2^{31}-1$ pseudorandom binary sequence (PRBS) data. Fig. 10 shows the traces at output ports to prove the successful operation. Comparing the bottom trace with the top one, the extinction ratio is improved, and the logic inversion due to cross-gain modulation is reversed. Fig. 11 shows the bit-error-rate (BER) measurement and eye diagrams. The power penalty at 10^{-9} BER is about -0.5 dB. The negative power penalty is achieved by 2R regeneration from XPM in the fixed-wavelength conversion. The "overshoots" between packets are caused by overlapping of the leading and trailing portions of packets. They can be eliminated

by careful timing adjustment. The processing delay caused by the forwarding table lookup and scheduling algorithm in the FPGA controller is 260 ns, ensuring a very low forwarding latency. In the experiment, the packet length is fixed to about 614 ns, and the guard time between two packets is about 205 ns. Supporting of variable length packets has very recently been implemented in the controller and will be discussed in a separate publication [28]. No fundamental difficulty is involved in this process, because the main extension required is to make the controller logic aware of packet-length information. We also emphasize that the packet switching demonstrated here is asynchronous, since the 50-ns delay between two contending packets is not a "slot" time.

VI. Conclusion

This paper presented a combined discussion on the end-to-end contention resolution schemes, incorporating wavelength-, time-, and space-domain contention resolution in the core network with the performance enhancement schemes at the network edge. The enhanced edge routers was proven effective in improving the network performance and in reducing the PLR. Simulation results also demonstrated the scaling efficiency of OPS networks, indicating that a considerable PLR is achievable with the proposed end-to-end contention resolution schemes. Experimental demonstration involving the prototype packet routing system achieved an error-free contention resolution with the forwarding latency of 260 ns. Currently, we are investigating the coordination mechanism among the core router, the edge router, and the NC&M systems to achieve an active contention resolution. The implementation of the enhanced edge router and the extended switch controller is also in progress.

References

[1] B. Meagher *et al.*, "Design and implementation of ultra-low latency optical label switching for packet-switched WDM networks," *J. Lightwave Technol.*, vol. 18, pp. 1978–87, Dec. 2000.

[2] I. Chlamtac *et al.*, "Scalable WDM access network architecture based on photonic slot routing," *IEEE/ACM Trans. Networking*, vol. 7, pp. 1–9, Feb. 1999.

[3] S. Yao, B. Mukherjee, and S. Dixit, "Advances in photonic packet switching: An overview," *IEEE Commun. Mag.*, vol. 38, pp. 84–94, Feb. 2000.

[4] C. Guillemot *et al.*, "Transparent optical packet switching: The European ACTS KEOPS project approach," *J. Lightwave Technol.*, vol. 16, pp. 2117–2133, Dec. 1998.

[5] D. K. Hunter *et al.*, "WASPNET: A wavelength switched packet network," *IEEE Commun. Mag.*, vol. 37, pp. 120–129, Mar. 1999.

[6] R. E. Wagner, R. C. Alferness, A. M. Saleh, and M. S. Goodman, "MONET: Multiwavelength optical networking," *J. Lightwave Technol.*, vol. 14, pp. 1349–1355, June 1996.

[7] L. Tancevski *et al.*, "Optical routing of asynchronous, variable length packets," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2804–2093, Oct. 2000.

[8] F. Callegati *et al.*, "Exploitation of DWDM for optical packet switching with quality of service guarantees," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 190–201, Jan. 2002.

[9] S. J. B. Yoo and G. K. Chang, "High-throughput, low-latency next generation Internet using optical-tag switching," U.S. Patent 6 111 673, Aug. 29, 2000.

[10] I. Chlamtac *et al.*, "CORD: Contention resolution by delay lines," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 1014–1029, June 1996.

[11] D. K. Hunter, M. C. Chia, and I. Andonovic, "Buffering in optical packet switches," *J. Lightwave Technol.*, vol. 16, pp. 2081–2094, Dec. 1998.

[12] S. L. Danielsen, P. B. Hansen, and K. E. Stubkjear, "Wavelength conversion in optical packet switching," *J. Lightwave Technol.*, vol. 16, pp. 2095–2108, Dec. 1998.

[13] F. Forghierri, A. Bononi, and P. R. Prucnal, "Analysis and comparison of hot-potato and single-buffer deflection routing in very high bit rate optical mesh networks," *IEEE Trans. Commun.*, vol. 43, pp. 88–98, Jan. 1995.

[14] S. Yao, B. Mukherjee, S. J. B. Yoo, and S. Dixit, "A unified study of contention-resolution schemes in optical packet-switched network," *J. Lightwave Technol.*, vol. 21, pp. 672–683, Mar. 2003.

[15] K. Okamoto *et al.*, "32 × 32 arrayed-waveguide grating multiplexer with uniform loss and cyclic frequency characteristics," *Electron. Lett.*, vol. 33, no. 22, pp. 1865–1866, Oct. 1997.

[16] J. Yang, M. Y. Jeon, J. Cao, Z. Pan, and S. J. B. Yoo, "Performance monitoring by sub-carrier multiplexing in optical label switching network," in CLEO/QELS 2003, Baltimore, MD, June 2003.

[17] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, Feb. 1994.

[18] K. Claffy, G. Miller, and K. Thopson. The Nature of Beast: Recent Traffic Measurements from an Internet Backbone. [Online]. Available: http:// ipn.nlanr.net/Papers/ Inet98/ index.html

[19] S. Yao *et al.*, "All-optical packet-switched networks: A study of contention resolution schemes in an irregular mesh network with variable-sized packets," in SPIE OptiComm'00, Richardson, TX, Oct. 2000.

[20] J. J. He, D. Simeonidouo, and S. Chaudhry, "Contention resolution in optical packet-switching networks: Under long-range dependent traffic," in *Proc. OFC'2000*, Thu4-1/295, pp. 295–297.

[21] B. Ryu and S. Lowen, "Point-process approaches to the modeling and analysis of self- similar traffic: Part I—Model construction," *Proc. IEEE INFOCOM '96*, vol. 3, pp. 1468–1475, Mar. 1996.

[22] B. Ryu and A. Elwaid, "The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities," in *Proc. ACM SIGCOMM'96*, 1996, pp. 3–14.

[23] M. Grossglauser and J. Bolot, "On the relevance of long-range dependence in network traffic," *IEEE/ACM Trans. Networking*, vol. 7, pp. 629–640, Oct. 1999.

[24] R. G. Garroppo, S. Giordano, S. Lucetti, and F. Russo, "Comparison of LRD and SRD traffic models for the performance evaluation of finite buffer system," presented at the Proc. ICC'2001, Helsinki, Finland, June 2001.

[25] S. Molnar and G. Miklos, "Peakedness characterization in teletraffic," in *IFIP Int. Conf. Performance Inform. Commun. Syst.*, Lund, Sweden, May 1998, pp. 25–28.

[26] M. K. Girish and J. Hu, "Modeling of correlated arrival processes in the Internet," in *Proc. 38th Conf. Decision Control*, Phoenix, AZ, Dec. 1999, pp. 4454–4459.

[27] H. J. Lee, S. J. B. Yoo, V. K. Tsui, and S. K. H. Fong, "A simple all-optical label detection and swapping technique incorporating a fiber Bragg grating filter," *IEEE Photon. Technol. Lett.*, vol. 13, pp. 635–637, June 2001.

[28] Z. Pan, H. Yang, Z. Zhu, J. Cao, V. Akella, S. Butt, and S. J. Ben Yoo, "Experimental demonstration of variable-size packet contention resolution and switching in an optical-label switching reouter," in Optical Fiber Communications Conf. (OFC 2004), Los Angeles, CA, Feb. 2004, submitted for publication.

Fei Xue (M'01) received the B.E, M.E., and Ph.D. degrees in computer engineering from Tianjin University, Tianjin, China, in 1992, 1995 and 1998, respectively.

Previously, he worked as a Research Associate with the Chinese University of Hong Kong, Shatin, New Territories, Hong Kong. From 1999 to 2000, he was a Postdoctoral Fellow with Simon Fraser University, Burnaby, BC, Canada. He is currently a Research Scientist with the University of California, Davis. His research interests include optical network architecture and protocol, performance evaluation of high-speed networks, and Internet traffic engineering.

Zhong Pan, photograph and biography not available at the time of publication.

Yash Bansal, photograph and biography not available at the time of publication.

Jing Cao (S'03) received the B.S. and M.S. degrees from the Department of Electronics Engineering, Tsinghua University, Beijing, China, in 1997 and 2000, respectively, and is working toward the Ph.D. degree at the Electrical and Computer Engineering Department, University of California, Davis.

His research focuses on optical integrated devices and system integration for next-generation optical networks.

Mr. Cao is a Student Member of the Optical Society of America (OSA).

Minyong Jeon, photograph and biography not available at the time of publication.

Katsunari Okamoto (M'85–SM'98), photograph and biography not available at the time of publication.

Shin Kamei, photograph and biography not available at the time of publication.

Venkatesh Akella, photograph and biography not available at the time of publication.

S. J. Ben Yoo received the B.S. degree in electrical engineering with distinction in 1984, the M.S. degree in electrical engineering in 1986, and the Ph.D. degree in electrical engineering with a minor degree in physics from Stanford University, Stanford, CA, in 1991. His Ph.D. dissertation was on linear and nonlinear optical spectroscopy of quantum-well intersubband transitions.

Prior to joining Bellcore in 1991, he conducted research on nonlinear optical processes in quantum wells, a four-wave-mixing study of relaxation mechanisms in dye molecules, and ultrafast diffusion driven photodetectors. During this period, he also conducted research on lifetime measurements of intersubband transitions and on nonlinear optical storage mechanisms at Bell Laboratories and IBM Research Laboratories, respectively. He was then a Senior Scientist at Bellcore, leading technical efforts in optical networking research and systems integration. His research activities at Bellcore included optical-label switching for the Next Generation Internet, power transients in reconfigurable optical networks, wavelength interchanging cross-connects, wavelength converters, vertical-cavity lasers, and high-speed modulators. He also participated in the Advanced Technology Demonstration Network and Multiwavelength Optical Networking (ATD/MONET) systems integration, the OC-192 synchronous optical network (SONET) ring studies, and a number of standardization activities. He joined the University of California, Davis (UC Davis), as Associate Professor of Electrical and Computer Engineering in March 1999. He is currently a Professor and the Branch Director of the Center for Information Technology Research in the Interest of Society (CITRIS). His current research involves advanced switching techniques and optical communications systems for the Next Generation Internet. In particular, he is conducting research on architectures, systems integration, and network experiments of all-optical label switching routers.

Prof. Yoo is an Associate Editor for IEEE PHOTONICS TECHNOLOGY LETTERS, a Senior Member of IEEE Lasers & Electro-Optics Society (LEOS), and a Member of the Optical Society of America (OSA) and Tau Beta Pi. He received the Bellcore CEO Award in 1998 and DARPA Award for Sustained Excellence in 1997 for his work at Bellcore.