HOPSMAN: An Experimental Optical Packet-Switched Metro WDM Ring Network With High-Performance Medium Access Control

Maria C. Yuang, I-Fen Chao, and Bird C. Lo

***Abstract*—Future optical metropolitan area networks (MANs) have been expected to exploit advanced optical packet switching (OPS) technologies to cost-effectively satisfy a wide range of applications having time-varying and high bandwidth demands and stringent delay requirements. In this paper, we present the architecture and access control of our experimental high-performance OPS metro WDM slotted-ring network (HOPSMAN). HOPSMAN has a scalable architecture in which the node number is unconstrained by the wavelength number. It encompasses a handful of nodes (called server nodes) that are additionally equipped with optical slot erasers capable of erasing optical slots resulting in an increase in bandwidth efficiency. In essence, HOPSMAN is governed by a novel medium access control (MAC) scheme, called probabilistic quota plus credit (PQOC). PQOC embodies a highly efficient and fair bandwidth allocation in accordance with a quota being exerted probabilistically. The probabilistic quota is then analytically derived taking the server-node number and destination-traffic distribution into account. Besides, PQOC introduces a time-controlled credit for regulating a fair use of the remaining bandwidth, particularly in the metro environment with traffic of high burstiness. Extensive simulation results show that HOPSMAN with PQOC achieves exceptional delay-throughput performance under a wide range of traffic loads and burstiness.**

***Index Terms*—Metropolitan area network (MAN); Wavelength division multiplexing (WDM); Optical packet switching (OPS); Medium access control (MAC); Bandwidth allocation**.

Manuscript received December 3, 2009; accepted December 3, 2009; published January 15, 2010 (Doc. ID 120953).

M. C. Yuang (e-mail: mcyuang@csie.nctu.edu.tw) and I.-F. Chao are with the Department of Computer Science, National Chiao Tung University, Taiwan.

B. C. Lo is with the Department of Computer Science and Information Engineering, Mingshin University of Science and Technology, Taiwan.

Digital Object Identifier 10.1364/JOCN.2.000091

I. Introduction

For long-haul backbone networks, optical wavelength division multiplexing (WDM) [1] has been shown to be successful in providing virtually unlimited bandwidth to support a large amount of steady traffic based on the optical-circuit-switching (OCS) paradigm. Future optical metropolitan area networks (MANs) [2,3], on the other hand, are expected to cost-effectively satisfy a wide range of applications having time-varying and high bandwidth demands and stringent delay requirements. Such facts bring about the need of exploiting the optical-packet-switching (OPS) [3,4] paradigm that takes advantage of statistical multiplexing to efficiently share wavelength channels among multiple users and connections. Note that the OPS technique studied here excludes the use of optical signal processing and optical buffers, which are current technological limitations OPS faces. Numerous topologies and architectures [2–11] for OPS-based WDM metro networks have been proposed. Of these proposals, the structure of slotted rings [5–12] receives the most attention. Essentially, these slotted-ring networks offer high-performance access and efficient bandwidth allocation by means of medium access control (MAC) schemes [2].

Before assessing the OPS WDM slotted-ring networks, we first examine some formerly proposed MAC schemes for ring networks. These schemes can generally be categorized as quota based or rate based. In the quota-based schemes, each node is allocated a quota that is the maximum transmission bound within a variable-length cycle. Most of the research work focuses on the dynamic adjustment of the cycle length. Asynchronous transfer mode ring protocol (ATMR) [13] allows the last active node to initialize a reset signal rotating on the ring to inform all nodes to restart a new cycle. MetaRing [14] uses a token-based signal circulating around the ring. When a node receives the token, it either forwards the token and thus starts a new cycle immediately or holds the token un-

1943-0620/10/020091-11/$15.00 © 2010 Optical Society of America

til the node has no data to send or the quota of the previous cycle expires. These schemes were shown to achieve high network utilization and great fairness. However, they cause cycle lengths to prolong several ring times, resulting in a large maximum delay bound and delay jitter, and thus poor bursty-traffic adaptation. In the rate-based schemes, resilient packet ring (RPR) (IEEE 802.17) [15] is based on a predetermined leaky bucket rate to transmit data, in combination with a local-fairness algorithm to resolve the potential congestion problem. Compared with the quota-based schemes, the rate-based scheme was shown to reduce the maximum delay bound [16]. However, the leaky rate is modified only after receiving the feedback from the downstream nodes when congestion occurs. As a result, due to using the predetermined rate and the slow response to rate changes, the scheme yields poor statistical multiplexing gain and dissatisfying delay-throughput performance especially under the high-burstiness fluctuating traffic condition. The goal of this paper is to present a quota-based MAC scheme that tackles the performance problem from the perspective of the determination of the quota rather than the cycle length.

There have been numerous OPS WDM slotted-ring networks proposed in the literature [2]. In the following, we assess three well-known prototyping networks that are most relevant to our work. First, the hybrid optoelectronic ring network (HORNET) [5] is a bidirectional WDM slotted-ring network in which each node is equipped with a tunable transmitter and a fixed-tuned receiver. It employs a MAC protocol, called a distributed queue bidirectional ring (DQBR), which is a modified version of the IEEE 802.16 distributed queue dual bus (DQDB) protocol [17]. DQBR requires each node to maintain a distributed queue via a pair of counters per each wavelength to ensure that packets are sent in the order they arrive at the network. With DQBR, HORNET achieves acceptable utilization and fairness at the expense of high control complexity for maintaining the same number of counter pairs as that of wavelengths. Moreover, due to the use of a fixed-tuned receiver, HORNET statically assigns each node a wavelength as the home channel for receiving packets. Such static wavelength assignment results in poor statistical multiplexing gain and thus throughput deterioration.

The second prototyping network is called a ring optical network (RingO) [6], which is a unidirectional WDM slotted-ring network with N nodes, where N is equal to the number of wavelengths. Each node is equipped with an array of fixed-tuned transmitters and one fixed-tuned receiver operating on a given home wavelength that identifies the node. Such a design gives rise to a scalability problem. RingO employs a MAC protocol, called a synchronous round robin with reservations (SR^3) [7], which is a combination of the synchronous round-robin (SRR), token-control quota-based (Multi-MetaRing), and slot-reservation mechanisms. The scheme was shown to achieve high utilization and fairness. As for the fairness scheme, Multi-MetaRing, it inherits all the pros and cons from the MetaRing. Specifically, there are W numbers of tokens rotating on W wavelengths. The scheme encounters an additional problem in which a node may hold several tokens at the same time due to the fact that only one data packet can be sent per slot time. The problem results in an increase in access delay and throughput degradation.

The metro network of the European IST Data and Voice Integration over DWDM (DAVID) [8,9] attempted to address the overall efficiency of ring-to-ring traffic and fairness and quality-of service (QoS) control inside a metro ring. DAVID is structured to be comprised of several independent fiber rings interconnected via a bufferless semiconductor-optical-amplifier- (SOA-) based packet switch, i.e., the hub node. The hub node is responsible for forwarding data packets among different rings of the network in the optical domain via an available wavelength. Due to having multiple rings, the hub requires each node to make a slot reservation prior to the transmissions and has to resolve a feasible wavelength-to-wavelength permutation [11] at all times. Within each ring, the Multi-MetaRing scheme is employed to ensure the fairness control. Each active node is equipped with a tunable transmitter, a tunable receiver, and an SOA-based slot eraser, enabling high slot reuse but at the expense of prohibitive system cost.

Our major goal has been to design and prototype a high-performance optical-packet-switched metro WDM ring network (HOPSMAN) [4]. In this paper, we present the architecture and the MAC scheme of HOPSMAN. HOPSMAN has a scalable architecture in which the node number is unconstrained by the wavelength number. Nodes are equipped with high-speed photonic hardware components that are capable of performing nanosecond-order OPS operations. HOPSMAN also encompasses a small number of server nodes that are additionally equipped with optical slot erasers capable of erasing optical slots resulting in an increase in bandwidth efficiency. In essence, HOPSMAN incorporates a MAC scheme, called probabilistic quota plus credit (PQOC). PQOC embodies a highly efficient and fair bandwidth allocation in accordance with a quota being exerted probabilistically. Unlike the existing quota-based schemes, our goal is to determine the quota rather than the cycle length. Taking the server-node number and destination-traffic distribution into account, we analytically derive the probabilistic quota. In addition, PQOC introduces a time-controlled credit for regulating a fair use of the

remaining bandwidth, particularly in the metro environment with traffic of high burstiness. Extensive simulation results show that HOPSMAN with PQOC achieves great fairness and exceptional delay-throughput performance under a wide range of traffic loads and burstiness.

The remainder of this paper is organized as follows. In Section II, we present the network and node architectures of HOPSMAN. In Section III, we describe the PQOC scheme and delineate the analysis for the determination of the probabilistic quota. Simulation results are then shown in Section IV. Finally, concluding remarks are given in Section V.

II. Network and Node Architectures

HOPSMAN is a unidirectional WDM slotted-ring network with multiple WDM data channels (λ_1–λ_W, at 10 Gb/s) and one control channel (λ_0, at 2.5 Gb/s), as shown in Fig. 1. Channels are further divided into synchronous time slots. Each data-channel slot contains a data packet in addition to some control fields to facilitate synchronization. Within each slot time, all data slots of W channels are fully aligned with the corresponding control slot. Each control slot is then subdivided into W minislots to carry the status of W data slots, respectively.

HOPSMAN contains two types of nodes: ordinary-node (O-node) and server-node (S-node). Each node of both types has a fixed transmitter and receiver pair for accessing the control channel. While an O-node is a regular node with only one tunable transmitter and receiver pair for accessing data channels, an S-node is equipped with multiple tunable transmitter and receiver pairs and an additional optical slot eraser. It is important to note that HOPSMAN requires at least one S-node, and as will be shown later, bandwidth efficiency improves cost-effectively by using only a small number of S-nodes.

The node architecture is shown in Fig. 2. It is best described as consisting of two building blocks for control-channel processing and data-channel accessing. For control-channel processing, a fixed optical

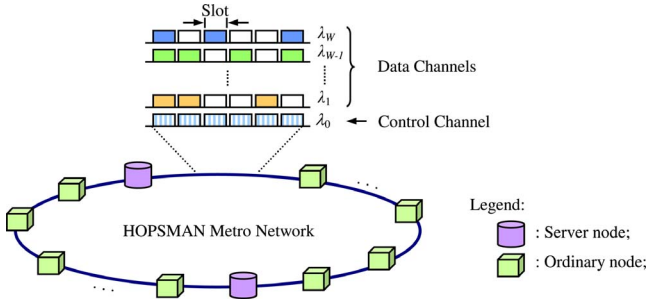

Fig. 1. (Color online) HOPSMAN: network architecture.

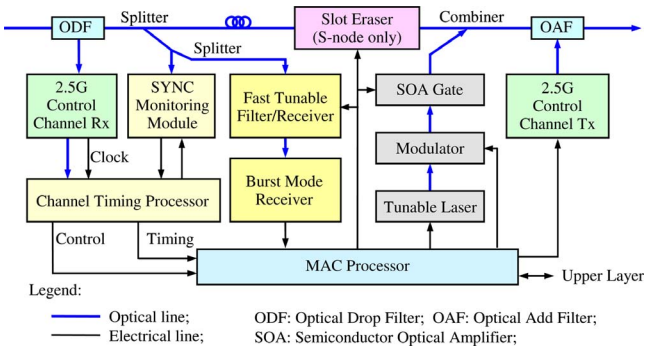

Fig. 2. (Color online) HOPSMAN node architecture.

drop filter (ODF) at the input port first extracts the optical signal from the control channel slot by slot. The control information is electrically received by a fixed-tuned receiver and processed by the MAC processor. While the control information is extracted and processed, data packets remain transported optically in a fixed-length fiber delay line. The channel timing processor, in coordination with the SYNC monitoring module, is responsible for extracting the slot boundary timing and for subsequently providing the activation timing for other modules. Having obtained the control information, namely, the status of W data channels, the MAC processor then executes the PQOC scheme to determine the add/drop/erase operations for all W channels and the status updates of the corresponding minislots in the control channel. Finally, a fixed-tuned transmitter inserts the newly updated control signal back in the fiber, which is, in turn, combined with the data channels' signal via the optical add filter (OAF).

Data-channel accessing corresponds to add and drop operations of data packets based on the broadcast-and-select configuration. Specifically, packets of all wavelengths are first tapped off ("broadcast") through wideband optical splitters. They are in turn received ("selected") via an optical tunable filter/receiver. The realization and use of such a tunable receiver makes HOPSMAN scalable, namely, the number of nodes is no longer constrained by the number of wavelengths. To transmit a packet onto a particular wavelength, the node simply tunes the tunable transmitter to the wavelength. Finally, to discontinue unneeded data packets on any wavelengths, the slot eraser (in an S-node only) employs a mux/demux pair and an array of W SOA on/off gates to reinsert new null signals on the wavelengths.

There are three main challenging issues about the hardware implementation of HOPSMAN [4]. They are the synchronization of the data and control channels, the design and implementation of high-speed photonic tunable receivers, and optical slot erasers. We now briefly describe our solutions to meeting these challenges. First, the channel timing synchronization is

ensured via two levels of alignment: coarse-grained and fine-grained synchronization. The first-level coarse-grained synchronization is achieved by inserting a fixed short-fiber delay line in the optical data-channel path to accommodate the basic control computation latency. The fine-grained synchronization is accomplished by matching a fixed-pattern preamble field (i.e., the SYNC field) at the beginning of each control slot. Second, the fast optical tunable filter/receiver is implemented based on a polarization-insensitive four-wave-mixing (FWM) method, which uses a sampled-grating distributed-Bragg-reflector (SGDBR) fast tunable pumping laser and an SOA. Due to the fact that the receiver's tuning delay solely depends on the tuning speed of the pumping laser, our FWM-based tunable filter/receiver achieves a tuning time of less than 25 ns. Finally, the optical slot eraser has been built with a mux/demux pair and an array of SOA gates, which can be turned on/off in 5 ns and achieve an on/off extinction ratio greater than 30 dB. More detailed information about the hardware design and implementation can be found in [4].

III. MAC Scheme—Probabilistic Quota Plus Credit

HOPSMAN is governed by a MAC scheme, called probabilistic quota plus credit (PQOC). In this section, we first describe the basic concepts of probabilistic quota and credit. We then present the analytic derivation for the determination of the probabilistic quota, which is followed by the detailed algorithm of the scheme.

A. Basic Design

Before presenting the MAC scheme, we first introduce a term that will be frequently used throughout the rest of the paper. Since each O-node has only one tunable receiver, receiver contention [2] occurs when there is more than one packet destined for the same receiver in a single slot time. Thus, two packets that are destined for the same node are prohibited to simultaneously occupy a single slot time via two different wavelengths. Likewise, because an O-node has only one tunable transmitter, any O-node is restricted to access at most one wavelength in a single slot time. Such a limitation is referred to as the *vertical-access constraint*.

The entire WDM ring is divided into a number of *cycles* (see Fig. 3), each of which is composed of a predetermined, fixed number of slots. Basically, PQOC allows each node to transmit a maximum number of packets (slots), or *quota*, within a cycle. Significantly, even though the total bandwidth is equally allocated

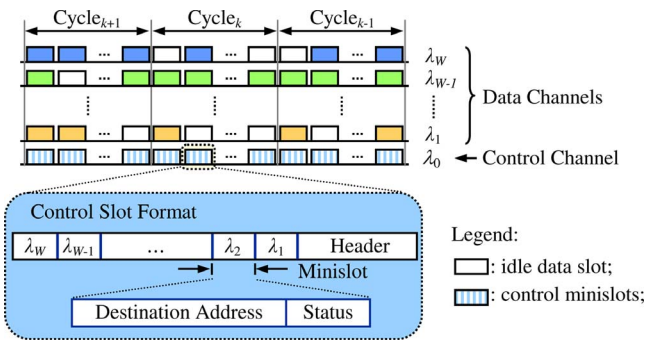

Fig. 3. (Color online) Cycle and slot structures.

to every node by means of the quota, unfairness surprisingly appears when the network load is high. This is because upstream nodes can access empty slots first, resulting in an increasing tendency for downstream nodes to encounter empty slots that are located *vertically* around the back of the cycle. This issue, as well as the vertical-access constraint, gives rise to poorer throughput and delay performance for downstream nodes. To resolve the unfairness problem, the quota is exerted in a probabilistic rather than a deterministic fashion, as "probabilistic quota" implies. In other words, rather than transmitting packets immediately if there remains quota (and idle slots of course), each node makes the transmission decision according to a probability. For example, the probability is set to be equal to the quota divided by the cycle length. The determination of the probabilistic quota will be detailed in the following subsection. Such an approach evenly distributes the idle slots within the entire cycle at all times and thus eliminates unfairness against downstream nodes. It is worth noting that, using the probabilistic quota a node may end up making fewer packet transmissions than its quota. This is caused by failing to find idle slots when the access is permitted according to the probability. The problem can be simply resolved by unconditionally granting a packet transmission in a subsequent slot time when there exists an idle slot.

Furthermore, if a node cannot use up its entire quota in a cycle, i.e., has fewer packets than its quota, the node yields the unused bandwidth (slots) to downstream nodes. In return, the node earns the same number of slots as credits. These credits allow the node to transmit more packets beyond its original quota in a limited number of upcoming cycles, called the *window*. That is, the credits are only valid when the number of elapsed cycles does not exceed the window. The rationale behind this design is to regulate a fair use of the unused remaining bandwidth particularly in the metro environment with traffic that is bursty in nature. Notice that there are system tradeoffs in PQOC involving cycle length and window size.

For example, the smaller the cycle length, the better the bandwidth sharing, and the larger the window size, the better the bursty-traffic adaptation, both at the cost of more frequent computation. The cycle length and window size can be dynamically adjusted in accordance with the monitored traffic load and burstiness via network management protocols. These issues go beyond the scope of this paper.

The implementation of PQOC is fairly simple. As shown in Fig. 3 each control slot contains a header (for synchronization purposes) and W minislots carrying the statuses of the corresponding W data slots. There are four distinct states for each data slot—BUSY, BUSY/READ (BREAD), IDLE, and IDLE/MRKD (IMRKD). A node wishing to transmit in a cycle and attaining access permission on the basis of probabilistic quota must first find an IDLE slot. If it succeeds, the node transmits the packet and alters the state from IDLE to BUSY. Otherwise, the node unconditionally transmits its packet on the next available slot without casting the probability again. A destination node that has successfully dropped a packet modifies the slot state from BUSY to BREAD. This allows the next S-node to erase the data slot by changing the status from BREAD back to IDLE, enabling slot reuse by downstream nodes. Furthermore, if a node has no packets to transmit but attains access permission on the basis of probabilistic quota, the node then earns the remaining number of slots as credits for future use, by altering the same number of data slots from IDLE to IMRKD. Finally, a node uses its credits to transmit more packets beyond the probabilistic quota on any IMRKD data slots within the window, and subsequently updates the state to BUSY.

B. Bandwidth Allocation—Probabilistic Quota Determination

Assume that there are S server nodes (S-node 1 to S-node S) in the network dividing itself into S sections (sections 1 to S). Each section contains more than one node including the server node of the section. To simplify the illustration, the S-node for a section is placed in the most downstream location in that section. Namely, section 1 is preceded by S-node S in section S, and section k is preceded by S-node $k-1$ in section $k-1$, for $k=2$ to S. More specifically, for a network with N nodes, we have $N=\Sigma_{k=1}^{S}N_k$, where N_k is the total number of nodes in section k.

Moreover, a slot passed by an S-node is considered as either *available bandwidth (AB)* if the slot is empty or erased or *used bandwidth (UB)* if the slot is nonempty and cannot be erased (has not been read). Thus, the quota for a node can be computed as the mean value of the total amount of AB observed by a section divided by the total number of nodes in the section. For instance, in an observed section (referred to as section b), we attain $Q_b=\overline{AB_b}/N_b$, where Q_b denotes the quota to be allocated to any node in section b, $\overline{AB_b}$ is the mean value of the total amount of AB passed down by S-node $b-1$, and N_b is the total number of nodes in section b. It is worth noting that the value of $\overline{AB_b}$ is relevant to the destination traffic distribution. S-nodes are also expected to receive more traffic than O-nodes. Thus, in the following, we derive a closed form for $\overline{AB_b}$ under an assumption that a probability p_A of destination traffic is uniformly distributed among all nodes (including the S-nodes), while the remaining probability $1-p_A$ $(=p_S)$ of traffic is additionally destined to all S-nodes. Clearly in the case of $p_S=0$, destination traffic is uniformly distributed among all nodes in the network. Regarding the value of p_A, it can be obtained through periodic traffic monitoring via network management protocols, which are beyond the scope of this paper.

The value of $\overline{AB_b}$ can be computed as the mean total bandwidth (total number of slots in a cycle) minus the mean UB. Thus, in the following we analyze the mean UB passed through S-node $b-1$. The analysis is presented in two parts: one is to consider the transmissions from any section to itself, and the other is to consider the transmissions from a section to the other sections. For the first part, since the total amount of traffic (slots) generated from any section (take section k as an example) is Q_kN_k, the traffic amount from section k to section k itself is $(Q_kN_kp_S/S)+(Q_kN_kp_AN_k/N)$. Of this amount of traffic, the fraction $(Q_kN_kp_S/S)+(Q_kN_kp_AN_k/2N)$ will be erased by section k's most downstream node, i.e., S-node k. Notice that the second term corresponds to the traffic sent from upstream to downstream nodes within section k. Therefore, the remaining traffic, $Q_kN_kp_AN_k/2N$, which is sent from downstream to upstream nodes within section k, will be passed through the entire ring and seen by section b as UB.

For the second part, let's take section $b+2$ as an example. The traffic that is sent from section $b+2$, passed through the entire ring, and seen by S-node $b-1$ and section b as UB, is the total amount of traffic destined to sections b and $b+1$. Thus, the mean UB becomes $Q_{b+2}N_{b+2}[(p_s/S)+(p_AN_b/N)]+Q_{b+2}N_{b+2}\times[(p_s/S)+(p_AN_{b+1}/N)]$. Finally, with all sections taken into account by summing all UB that passes through S-node $b-1$, and given that the total bandwidth in a cycle is CW (where C is the total number of slots in a cycle and W is the number of data channels), we obtain $\overline{AB_b}$ as the following equation:

$$\overline{AB_b}=CW-\sum_{k=1}^{S}\left\{Q_kN_k\left(\frac{p_AN_k}{2N}+U_k\right)\right\},$$

where

$$U_k = \begin{cases} \sum_{m=b}^{k-1} \left(\frac{p_S}{S} + \frac{p_A N_m}{N} \right), & \text{if } b \le k \le S \\ \sum_{m=b}^{S} \left(\frac{p_S}{S} + \frac{p_A N_m}{N} \right) + \sum_{n=1}^{k-1} \left(\frac{p_S}{S} + \frac{p_A N_n}{N} \right), & \text{if } 1 \le k < b \end{cases} . \quad (1)$$

Notice that Eq. (1) cannot be solved because there is more than one unknown variable ($\overline{AB}_b$ and Q_k's) in the equation. In the following, we solve the equation under a prevailing case in which S-nodes are evenly located in the network, namely, $N_1=N_2=N_3=\ldots=N_S=N/S$. In this case, due to the same behavior of S-nodes, we obtain the same quota Q, where $Q=\overline{AB}/(N/S)$, for all nodes in the network. With this additional equation and the simplified version of Eq. (1) as

$$\overline{AB} = CW - \sum_{k=1}^{S} \left\{ Q \frac{N}{S} \left(\frac{p_A}{2S} + \frac{k-1}{S} \right) \right\}, \quad (2)$$

we can attain the closed-form solution for Q, as

$$Q = \frac{CW}{N} \left(\frac{2S}{S - p_S + 2} \right). \quad (3)$$

With the quota determined, we are now ready to obtain the probabilistic quota, denoted as P_Q. Given the total number of packets currently in the queue as np_{queue}, P_Q can simply be expressed as

$$P_Q = \min\left(\frac{Q}{C}, \frac{np_{queue}}{C} \right). \quad (4)$$

C. Detailed Algorithm

The detailed PQOC algorithm is given in Fig. 4. An np_{queue} number of packets that have newly arrived or failed to transmit in the previous cycle are scheduled to transmit in the current cycle. The algorithm is executed on a per-slot basis. If the slot is marked as the beginning of a cycle, the algorithm determines three variables: P_Q, the number of idle slots to be marked (ns_{imrkd}), and the number of credits available up to this cycle (nc). Basically, P_Q is computed with Eq. (4), ns_{imrkd} takes the greater value between 0 and ($Q - np_{queue}$), and nc is calculated on a sliding window basis. nc can also be given based on a fixed window rather than the sliding window strategy. Usually, a sliding window strategy is superior to a fixed window strategy with respect to bursty-traffic adaptation, albeit at the cost of requiring greater computing effort and storage needs. Finally, the primary work of the algorithm is to determine whether transmission of packets is allowed or not, as clearly depicted in the transmission process (Steps 5 and 6) in Fig. 4.

IV. Simulation Results

In this section, we present simulation results to demonstrate the performance of HOPSMAN with respect to throughput, access delay, and fairness. The settings of parameters for simulation are given in the following. The network has a total of 20 nodes ($N=20$), in which node 1 is always designated as an S-node. There are 20 cycles on the ring. Each cycle consists of 100 slots per wavelength. Without specific indication, traffic destinations are assumed to be uniformly distributed among all nodes ($p_S=0$). The credit window size is 10. Traffic is generated following either a Poisson distribution or a two-state (H and L) Markov modulated Poisson process (MMPP) [18] for modeling smooth and bursty traffic, respectively. Specifically, the MMPP is characterized by four parameters (α, β, λ_H, and λ_L), where α (β) is the probability of changing from state H (L) to L (H) in a slot, and λ_H (λ_L) represents the probability of arrivals at state H (L). Accordingly, given $\lambda_L=0$, the mean arrival rate can be expressed as $\beta\times\lambda_H/(\alpha+\beta)$, and traffic burstiness (B) can be given by $B=(\alpha+\beta)/\beta$. Finally, simulation is terminated after reaching a 95% confidence interval. Due to having multiple S-nodes, the *traffic intensity (TI)* to be generated per slot per wavelength

Variables
Q : quota;
nc : number of credits to be used;
ns_{imrkd} : number of idle slots to be marked;
np_{PQ} : number of packets allowed to be transmitted based on prob. quota;
np_{queue} : number of packets in the queue;
ws : credit window size;
$q[i]$: number of remaining quota in cycle i;
$c[i]$: number of credits used in cycle i;

Slot type
Header: {CYCLE_BEGIN, NORMAL};
Status : {BUSY, BREAD, IDLE, IMRKD};

Main Process() /*execute at each slot time*/
1. Read the control slot;
/* Computation at the cycle begin */
2. **if** (slot's *Header* is CYCLE_BEGIN)
/*enter the m^{th} cycle*/
Add the number of arrivals of pre-cycle to np_{queue};
Determine P_Q according to Eq. (4);
$ns_{imrkd} = \max(0, Q - np_{queue})$; $np_{PQ}=0$;
$nc=\max(0, \min(np_{queue}-Q, \sum_{i=m-ws}^{m-1}(q[i]-c[i])))$;
$q[m]=Q$; $c[m]=0$;
endif
/* Receive packets */
3. Receive the packet destined to it, and update BUSY to BREAD;
/* Server node erases packets */
4. **if** (node is S-node)
Erase BREAD packets;
Update BREAD to IDLE; **endif**
/* Transmission process */
5. **if** (Randomize(P_Q))
np_{PQ}++; **endif** /*store transmission allowance */
6. **if** (exist a vertical-access-constraint-free packet)
switch
case (nc>0 **and** exist(IMRKD)):
/*transmit by credit */
nc --; $c[m]$ ++;
Transmit the packet; np_{queue} --;
Update IMRKD to BUSY; **break**;
case (np_{PQ} >0 **and** exist(IDLE)):
/*transmit by quota using IDLE */
np_{PQ} --; $q[m]$ --;
Transmit the packet; np_{queue} --;
Update IDLE to BUSY; **break**;
case (np_{PQ} >0 **and** exist(IMRKD)):
/*transmit by quota using IMRKD */
np_{PQ} --; ns_{imrkd}++; $q[m]$ --;
Transmit the packet; np_{queue} --;
Update IMRKD to BUSY; **break**;
otherwise
No transmission; **endswitch endif**
/* Mark slots */
7. **if** (ns_{imrkd} > 0 **and** exist(IDLE))
ns_{imrkd} --; Update IDLE to IMRKD; **endif**

Fig. 4. (Color online) Detailed PQOC algorithm.

is unequal to the normalized load (L) per slot per wavelength. They can be related, however, according to Eq. (3), as

$$TI = L\frac{Q}{CW/N} = L\left(\frac{2S}{S - p_S + 2}\right). \tag{5}$$

From Eq. (5), given S S-nodes in the network, the maximum value of TI [defined as the *maximum throughput* (T^S_{max})] occurs at the normalized load L being equal to one. That is,

$$T^S_{max} = \frac{2S}{S - p_S + 2}. \tag{6}$$

From Eq. (6), we observe that T^S_{max} increases when p_S is increased. This is because more traffic destined to S-nodes will enable more data slots being erased, i.e., more usable bandwidth after S-nodes. Consequently, if the network inspects that an extra percentage of traffic is additionally and constantly being sent to S-nodes (p_S), it can recompute and extend the value of quota for each user achieving greater throughput. As previously described, the determination of the p_S value is a network management issue and out of the scope of the paper.

In Fig. 5, we first draw comparisons of the analytic and simulation results on system throughput under different S-node numbers and two different tunable-transceiver-pair settings. Analytic results are obtained based on Eqs. (5) and (6). In the simulation, we assume there are 60 nodes in the network. The results plotted in Figs. 5(a) and 5(b) are obtained from the system with one and two pairs of optical tunable transceivers at each node, respectively. First, we observe from Fig. 5(a) that due to the vertical-access constraint when each node is equipped with only one pair of tunable transceivers, the resulted throughput performance is lower than the theoretical maximal throughput that is derived by the above analysis. To justify this fact and to demonstrate the validity of the analysis, we use the second setting in which each node is equipped with two pairs of tunable transceivers, with the result that the vertical-access constraint is lifted. Results are shown in Fig. 5(b). From the figure, analytic results are shown to be in profound agreement with the simulation results. Moreover, it is clear that increasing the S-node number results in an improvement in throughput, but at a declining rate as the number of S-nodes grows. For example, the throughput improves most dramatically when the network changes from having one S-node to two S-nodes. The result explains the economy and efficiency of HOPSMAN behind the scarce use of S-nodes.

Focusing on the network with one S-node (Node 1), we next examine the throughput and delay performance of HOPSMAN under various loads and burstiness. Simulation results are displayed in Fig. 6. As depicted in Fig. 6(a), despite the vertical-access constraint, the probabilistic-quota design helps HOPSMAN achieve 100% throughput and fairness under all loads that are less than or equal to 0.9. However, as the network becomes highly saturated when the load reaches 0.99, we observe inevitable throughput deterioration for downstream nodes as a result of the vertical-access constraint. An intensive comparison of delay with and without the probabilistic-quota design will be given shortly. We show in Figs. 6(b)–6(d) that HOPSMAN guarantees delay fairness under all non-saturated loads. As expected, delay increases with the traffic burstiness. Most importantly, as shown in Fig. 6(b), HOPSMAN achieves remarkably low delay under $L = 0.7$. Taking this result, along with other results in Fig. 6(a) taken into consideration, it is clear that HOPSMAN achieves superior bandwidth allocation under heavier loads while being able to provide random access under lighter loads.

We now study the impact of credit window size on

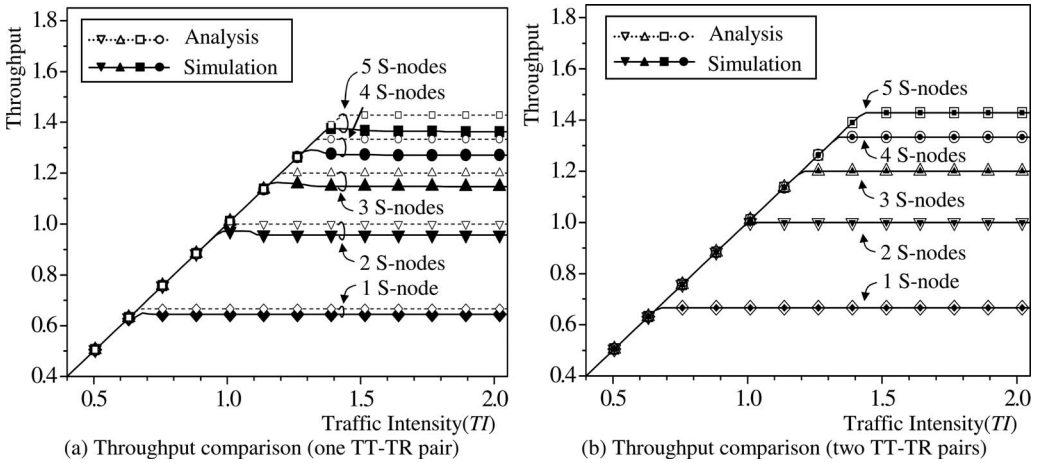

Fig. 5. Analytic and simulation results on system throughput under different S-node numbers.

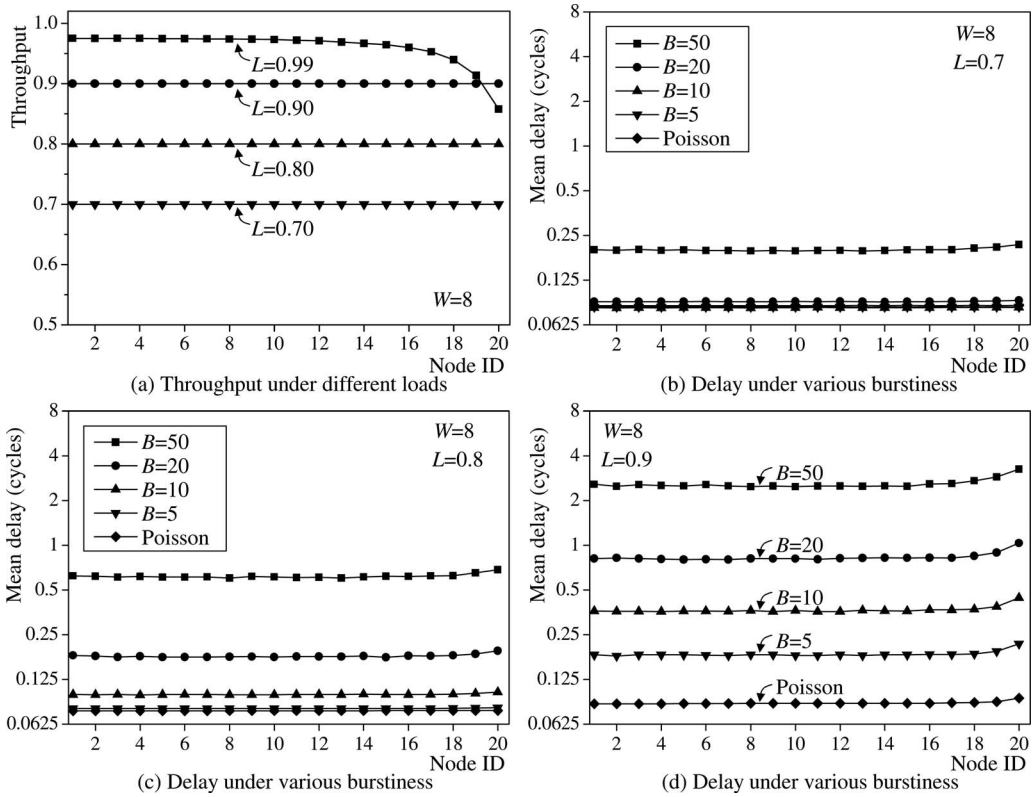

Fig. 6. Throughput and delay performance of HOPSMAN.

mean access delay under various traffic burstiness. Simulation results are plotted in Fig. 7. As shown in the figure, mean delay declines with increasing window size under all burstiness. However, as was mentioned previously, larger window sizes result in higher computational complexity. Fortunately, we have discovered that such delay reduction occurs most effectively around smaller window sizes (less than ten). Thus, the results can serve as a guideline on the determination of an appropriate and small window size satisfying an acceptable grade of delay.

Recall that in PQOC one uses credit to access slots beyond the quota in order to exert fair access control

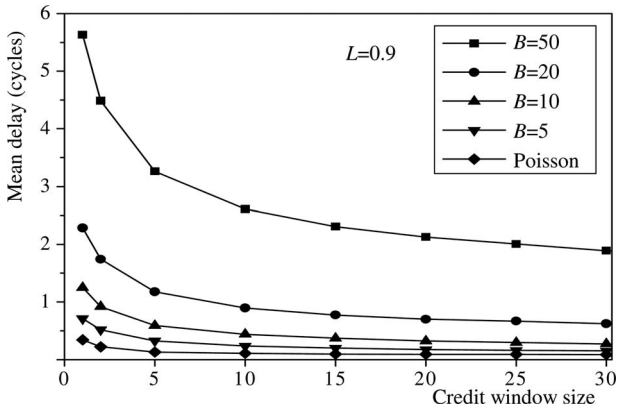

Fig. 7. Credit window size under various burstiness.

of the remaining unused bandwidth on all nodes. This is demonstrated in the following via simulation for a network with malicious nodes. In the simulation, nodes 5 and 15 are malicious nodes. While each malicious node generates excessive loads of 0.09, each other node generates a load of 0.045, rendering the network highly saturated, namely, with a total load of 0.99. Simulation results are displayed in Fig. 8. In the figure, we draw a comparison of delay between the PQOC scheme (with credit) and PQO (without the credit) under two burstiness levels of traffic. As shown in Fig. 8(a), under low-burstiness traffic, PQOC makes the two malicious nodes suffer severe delay while leaving other normal nodes completely unaffected. On the contrary, the PQO scheme without credit gives rise to delay deterioration (and thus unfairness) to the neighboring nodes of the two malicious nodes. As the traffic burstiness increases, the delay unfairness problem worsens as shown in Fig. 8(b). In this case, PQOC can still guarantee a high grade of fairness among all nodes, except for several downstream nodes due to network saturation. Thus, the PQOC scheme is evidently robust and fair even when under attack by malevolent nodes.

We next examine in Fig. 9 the impact of the probabilistic quota design on access delay under a variety of loads and burstiness. In the simulation, we draw a delay comparison between the PQOC scheme and the

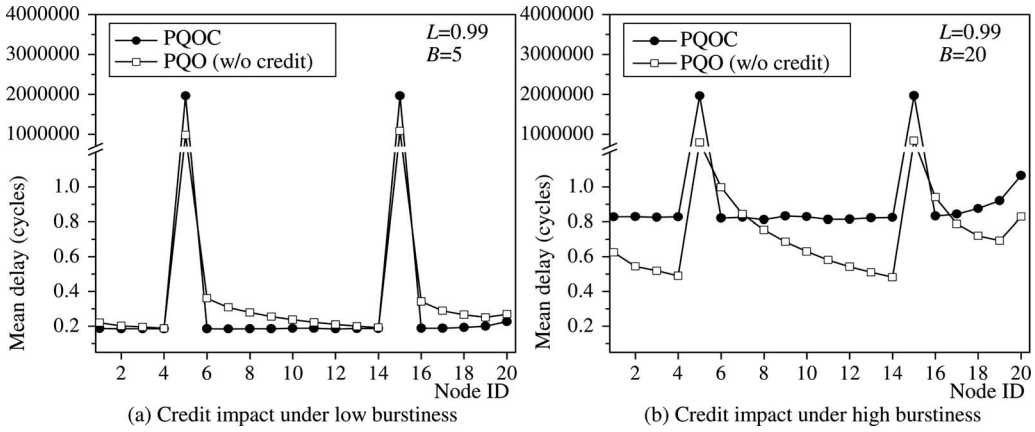

Fig. 8. Credit impact on delay for network with malicious nodes (nodes 5 and 15).

QOC scheme where the quota is exerted in a deterministic manner. As shown in Fig. 9(a), under medium loads of traffic, both schemes yield superlatively low delay. However, as the burstiness increases, despite the fact that QOC attains slightly better delay due to deterministic transmissions, the scheme begins showing signs of delay unfairness for downstream nodes, as shown in Fig. 9(b). Worst of all, as the load increases [see Figs. 9(c) and 9(d)], as a result of the vertical-access constraint, the QOC scheme not only undergoes severe unfairness toward downstream nodes but also incurs deteriorating delay for all nodes. In other words, PQOC is superior to QOC at a negligible cost. (It is worth noting that PQOC still incurs minor unfairness under exceedingly high loads due to the vertical-access constraint. Such a problem can be mitigated by applying unequal probabilistic quotas to nodes of different locations.) Consequently, the PQOC scheme invariably achieves superior delay and fairness irrespective of traffic loads and burstiness.

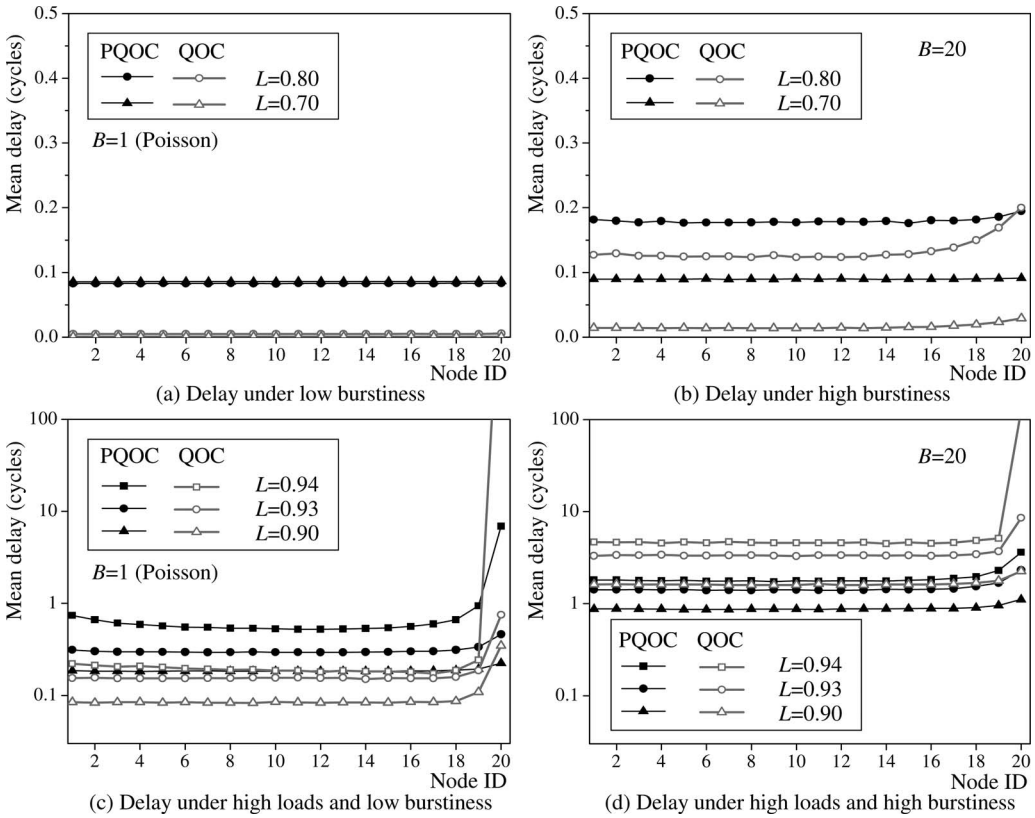

Fig. 9. Impact of probabilistic exertion of quota under various loads and burstiness.

V. CONCLUSIONS

In this paper, we have presented the architectural design and access control, i.e., PQOC, of our experimental optical packet-switched metro WDM slotted-ring network, HOPSMAN. In addition to ordinary nodes (O-nodes), HOPSMAN encompasses a few server nodes (S-nodes) that are equipped with optical slot erasers, resulting in a significant increase in system throughput. Essentially, the PQOC scheme employs a novel probabilistic-quota-based method to achieve fair and efficient bandwidth allocation. Given the number of S-nodes and destination-traffic distribution, we derived a closed-form formula for the determination of the probabilistic quota. PQOC also uses a window-constrained credit-based approach to facilitate versatile allocation of the remaining bandwidth under highly bursty and fluctuating traffic environments. Simulation results delineated that HOPSMAN achieves 100% throughput when there are only two S-nodes in the network. Furthermore, HOPSMAN with PQOC was shown to achieve highly efficient and fair bandwidth allocation under various traffic loads and burstiness. Finally, HOPSMAN was justified robust and fair when under attack by malevolent nodes.

ACKNOWLEDGMENTS

This work was supported in part by the NCTU/CCL Joint Research Center and in part by the National Science Council (NSC), Taiwan, under grant NSC98-2221-E-009-064-MY3.

REFERERNCES

[1] B. Mukherjee, "WDM optical communication networks: progess and challenges," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 10, pp. 1810–1824, Oct. 2000.

[2] M. Herzog, M. Maier, and M. Reisslein, "Metropolitan area packet-switched WDM networks: a survey on ring systems," *IEEE Commun. Surv. Tutorials*, vol. 6, no. 2, pp. 2–20, 2004.

[3] S. Yao, S. Yoo, B. Mukherjee, and S. Dixit, "All-optical packet switching for metropolitan area networks: opportunities and challenges," *IEEE Commun. Mag.*, vol. 39, no. 3, pp. 142–148, Mar. 2001.

[4] M. Yuang, Y. Lin, S. S. W. Lee, I. Chao, B. Lo, P. Tien, C. Chien, J. Chen, and C. Wei, "HOPSMAN: an experimental testbed system for a 10-Gb/s optical packet-switched WDM metro ring network," *IEEE Commun. Mag.*, vol. 46, no. 7, pp. 158–166, July 2008.

[5] I. White, M. Rogge, K. Shrikhande, and L. Kazovsky, "A summary of the HORNET project: a next-generation metropolitan area network," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 9, pp. 1478–1494, Nov. 2003.

[6] A. Carena, V. Feo, J. Finochietto, R. Gaudino, F. Neri, C. Piglione, and P. Poggiolini, "RingO: an experimental WDM optical packet network for metro applications," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 8, pp. 1561–1571, Oct. 2004.

[7] M. Marsan, A. Bianco, E. Leonardi, A. Morabito, and F. Neri, "All-optical WDM multi-rings with differentiated QoS," *IEEE Commun. Mag.*, vol. 37, no. 2, pp. 58–66, Feb. 1999.

[8] C. Develder, A. Stavdas, A. Bianco, D. Careglio, H. Lønsethagen, J. P. Fernández-Palacios Giménez, R. Van Caenegem, S. Sygletos, F. Neri, J. Solé-Pareta, N. Pickavet, N. Le Sauze, and P. Demeester, "Benchmarking and viability assessment of optical packet switching for metro networks," *J. Lightwave Technol.*, vol. 22, no. 11, pp. 2435–2451, Nov. 2004.

[9] C. Linardakis, H. Leligou, A. Stavdas, and J. Angelopoulos, "Implementation of medium access control for interconnecting slotted rings to form a WDM metropolitan area network," *J. Opt. Netw.*, vol. 3, no. 11, pp. 826–836, Nov. 2004.

[10] C. Jelger and J. Elmirghani, "A slotted MAC protocol for efficient bandwidth utilization in WDM metropolitan access ring networks," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 8, pp. 1295–1305, Oct. 2003.

[11] C. Linardakis, H. Leligou, A. Stavdas, and J. Angelopoulos, "Using explicit reservations to arbitrate access to a metropolitan system of slotted interconnected rings combining TDMA and WDMA," *J. Lightwave Technol.*, vol. 23, no. 4, pp. 1576–1585, Apr. 2005.

[12] J. Kim, J. Cho, S. Das, D. Gutierrez, M. Jain, C. Su, R. Rabbat, T. Hamada, and L. Kazovsky, "Optical burst transport: a technology for the WDM metro ring networks," *J. Lightwave Technol.*, vol. 25, no. 1, pp. 93–102, Jan. 2007.

[13] K. Imai, T. Ito, H. Kasahara, and N. Morita, "ATMR: asynchronous transfer mode ring protocol," *Comput. Networks ISDN Syst*, vol. 26, no. 6–8, pp. 785–798, March 1994.

[14] I. Cidon and Y. Ofek, "MetaRing—a full-duplex ring with fairness and spatial reuse," *IEEE Trans. Commun.*, vol. 41, no. 1, pp. 969–981, Jan. 1993.

[15] F. Davik, M. Yilmaz, S. Gjessing, and N. Uzun, "IEEE 802.17 Resilient Packet Ring Tutorial," *IEEE Commun. Mag.*, vol. 42, no. 3, pp. 112–118, Mar. 2004.

[16] C. Huang, H. Peng, and F. Yuan, "A deterministic bound for the access delay of resilient packet rings," *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 87–89, Jan. 2005.

[17] Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN), IEEE Standard 802.6, Dec. 1990.

[18] W. Fischer and K. Meier-Hellstern, "The Markov-modulated Poisson process (MMPP) cookbook," *Perform. Eval.*, vol. 18, no. 2, pp. 149–171, Sept. 1993.

Maria C. Yuang received her Ph.D. degree in electrical engineering and computer science from the Polytechnic University, Brooklyn, New York, in 1989. From 1981 to 1990, she was with AT&T Bell Laboratories and Bell Communications Research (Bellcore), where she was a member of the technical staff working on broadband networks and protocol engineering. In 1990, she joined National Chiao Tung University (NCTU), Taiwan, where she is currently a Professor in the Department of Computer Science. Prof. Yuang served as a Guest Editor for a Special Issue of the *IEEE Journal on Selected Areas in Communications* on Next-Generation Broadband Optical Access Network Technologies in 2009. She has served on the technical program committee of many technical conferences including IEEE ICC and GLOBECOM, and has been invited to give invited talks at many technical conferences. Her main research interests include broadband optical networks, wireless networks, multimedia communications, and performance modeling and analysis. She is a senior member of IEEE, and a member of OSA. She holds 16 patents in the field of broadband networking and has over 100 publications, including a book chapter.

I-Fen Chao received her B.S. and M.S. degrees in computer and information engineering from National Chiao Tung University, Taiwan, in 1992 and 1994, respectively. From 1995 to 1998 she was at CCL/ITRI, working on personal communications systems. From 1998 to 2003 she was with Faraday Technology Corporation, Hsinchu Science Park, Taiwan, as a Technical Manager working on an embedded OS/system. In 2003, she joined the Department of Computer Science, National Chiao Tung University, where she is currently pursuing a Ph.D. degree. Her current research interests include high-speed networking, optical networking, and performance modeling and analysis.

Bird C. Lo (S'98-M'01) received his Ph.D. degree in computer science and information engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2002. He was with National Chiao Tung University as a Research Assistant Professor, from 2003. He is currently an Assistant Professor in the Department of Computer Science and Information Engineering at Mingshin University of Science and Technology, Hsinchu, Taiwan. His main research interests include optical networking, multimedia communications, and e-learning.