

Toward Wireless Backhaul Using Circuit Emulation Over Optical Packet-Switched Metro WDM Ring Network

I-Fen Chao, *Member, IEEE*, and Maria C. Yuang, *Senior Member, IEEE*

Abstract—Optical WDM metro-ring networks have been considered to be promising solutions for wireless backhaul. In this paper, we propose an integrated traffic control scheme (ITCS), which facilitates circuit emulation service (CES) for wireless backhaul over a previously proposed experimental optical packet-switched WDM metro-ring network, HOPSMAN. The ITCS seamlessly integrates variable-bit-rate (VBR) CES wireless with connection-less best-effort data, satisfying stringent QoS requirements and retaining maximal network throughput. The ITCS ensures efficient setup of CES connections by employing a simple mean-rate-based distributed admission control followed by a novel slot-marking reservation. The total connection setup delay is comprised of the setup queueing delay and slot-marking delay. While the slot-marking delay is nearly a ring time, the mean setup queueing delay is formally computed through an accurate approximation based on an $M/G/m$ queueing analysis. We show simulation results to demonstrate that the ITCS accommodates remarkably high CES traffic loads while satisfying a wide range of delay requirements for wireless backhaul. Taking the background best-effort traffic into account, ITCS achieves exceedingly low delay, jitter, and delay bound for CES under various traffic loads and burstiness.

Index Terms—Circuit Emulation Service (CES), $M/G/m$ queueing analysis, Medium Access Control (MAC), Optical Packet Switching (OPS), Wavelength Division Multiplexing (WDM), Wireless Backhaul.

I. INTRODUCTION

THE ever-growing demand for wireless broadband services brings about the enormous capacity requirements for wireless backhaul that transports aggregate traffic from the base stations (BS) to the core network. Due to the fact that wireless backhaul traffic is of variable-bit-rate (VBR) nature, the increasingly stringent demand for meeting its quality-of-service (QoS) requirements [1]–[3] poses real challenges to the design of wireless backhaul infrastructures. Current backhaul infrastructures rely mostly on three transport media: copper, microwave radio links, and optical fiber [4], [5]. The legacy backhaul networks based on leased T1/E1 copper lines are not cost-efficient when

transmitting the rapidly increasing volume of bursty data traffic. While the wireless backhaul solutions based on microwave radio links are flexibly deployable, the broadband optical backhaul, which provides high bandwidth and spectral efficiency, has been considered to be the most practical and promising infrastructure.

There exist multiple variants of optical backhaul network architectures [6]–[9]. The work in [6] and [7] uses an optical ring either to interconnect BSs directly or interconnect central unit nodes, which are connected to multiple low-power and cost-effective remote antenna units. The architectures discussed in [6] and [8] also consist of a metro ring network with multiple passive optical networks (PONs) attached. The ONUs of each PON are either integrated with BSs [9] or are connected to wireless gateways of wireless mesh networks (WMN). While these studies propose bandwidth-efficient and wide-coverage optical backhaul network architectures, the critical issue of designing a backhaul network that fulfills a wide range of QoS requirements remains mostly unresolved.

Thanks to recent advances in optical device and networking technology, one prominent optical backhaul solution lies in the exploitation of optical packet switching (OPS) WDM ring networks [10], [11] in metropolitan areas. OPS WDM networks are capable of supporting a multitude of bandwidth-hungry multimedia services in a flexible and efficient manner. Pertaining to such OPS WDM networks, we have earlier proposed and prototyped an OPS WDM metro slotted-ring network, called HOPSMAN [10], [11]. HOPSMAN is governed by a novel medium access control (MAC) scheme, called probabilistic quota plus credit (PQOC). PQOC performs a highly efficient and fair bandwidth allocation based on a quota being exerted probabilistically and a time-controlled credit, achieving high delay and throughput performance particularly for traffic of high burstiness. In this study, our goal is to support wireless backhaul by designing and facilitating a new circuit emulation services (CES) [12], [13] over HOPSMAN.

In this paper, we propose a new MAC mechanism, called the integrated traffic control scheme (ITCS). The ITCS seamlessly integrates the CES for wireless backhaul and non-CES (nCES) for background best-effort data over HOPSMAN. While inheriting all the benefits from PQOC, the ITCS adds three new prominent features to best accommodate for wireless backhaul. First, ITCS allows a number of slots to be flexibly reserved at any wavelength and at any slot time, thereby accommodating high amounts of CES traffic. Second, through a simple, but effective mean-rate-based distributed call admission control (CAC) followed by a novel slot-marking reservation, ITCS achieves

Manuscript received February 3, 2013; revised June 2, 2013 and July 26, 2013; accepted August 6, 2013. Date of publication August 14, 2013; date of current version September 2, 2013.

I-F. Chao is with the Department of Electrical Engineering, Yuan Ze University, Zhongli 320, Taiwan (e-mail: ifchao@saturn.yzu.edu.tw).

M. C. Yuang is with the Department of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: mcyuang@csie.nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JLT.2013.2278403

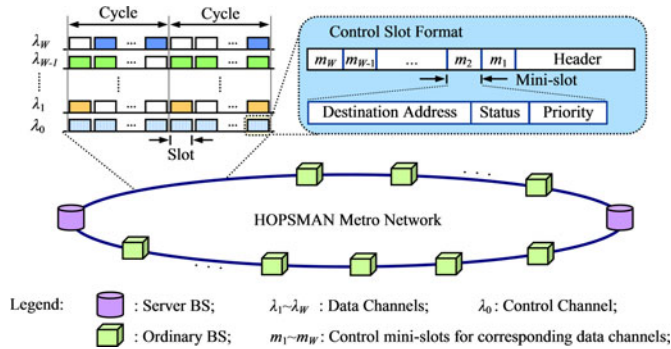


Fig. 1. HOPSMAN: Network architecture.

efficient setup of CES connections with exceedingly low delay and jitter. Third, nCES traffic takes advantages of all remaining and unused reserved bandwidth, optimizing the aggregate system throughput to nearly 100%.

Moreover, we conduct an analysis for the mean connection setup delay of the CES. The network is modeled as an $M/G/m$ queueing system. In existing work, accurate expected waiting time results for a multiple-server queueing system ($M/G/m$) have only been attained for a few special service distributions, such as deterministic, Erlang, and hyperexponential-2. Most works provides approximation formulas [14]–[19] that estimate the mean queueing delay for the $M/G/m$ queue from the first two or three moments of the service distribution. Their proposed approximation results maintain a less than 10% relative error only for certain properties of service distributions. The main contribution of our work lies in a novel approximation derivation that yields accurate results for a multiserver queueing system with a general service time that fits for our system.

The remainder of this paper is organized as follows. In Section II, we provide an overview of our previously designed network, HOPSMAN, and its basic MAC scheme (PQOC). In Section III, we propose the integrated MAC scheme, ITCS, and outline its detailed algorithm. In Section IV, we present the connection-setup delay analysis for the CES based on an $M/G/m$ queueing model. Simulation results are shown in Section V, followed by concluding remarks in Section VI.

II. OVERVIEW OF HOPSMAN AND PQOC

HOPSMAN [10], [11] is a unidirectional WDM ring network with multiple WDM data channels (λ_1 – λ_w , at 10 Gb/s) and a single control channel (λ_0 , at 2.5 Gb/s), as shown in Fig. 1. Channels are divided into synchronous time slots. In 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, which carry the status of the corresponding W data slots.

As an optical backhaul network, HOPSMAN can essentially adopt any type of ring-based backhaul solution. In this study, we consider an architecture where an optical ring interconnects BSs directly. HOPSMAN contains two BSs types: the ordinary-BS (O-BS), and the server-BS (S-BS). Both types of BSs are equipped with a fixed transmitter and receiver pair for accessing

the optical control channel, as well as a tunable transmitter and receiver pair for accessing optical data channels. For an S-BS, it is also equipped with an additional optical slot eraser, which employs a multiplexer/de-multiplexer pair and an array of W SOA ON/OFF gates to reinsert new null signals on the wavelengths. Unlike electrical data, the optical data can only be sent via null slots; otherwise, the receiver will fail to differentiate the received data. Therefore, HOPSMAN requires at least one S-BS to erase data. Moreover, since an optical slot eraser is expensive, HOPSMAN is shown to improve bandwidth efficiency cost-effectively by using only a small number of S-BSs [11].

The entire WDM ring is divided into a number of *cycles*, each of which is composed of a predetermined fixed number of slots. The previously proposed MAC (PQOC) allows each BS to transmit a maximum number of packets (slots), or *quota*, within a cycle. Significantly, even though the total bandwidth is equally allocated to every BS through the quota, unfairness surprisingly appears at downstream BSs when the network load is high. To resolve the unfairness problem, the quota is exerted in a nondeterministic but probabilistic manner, based on a probability, called P_Q . As a result, empty or used slots are randomly dispersed over each cycle, thus resolving the aforementioned unfairness problem. Additionally, if a BS cannot consume its entire quota in a cycle, the BS yields the unused bandwidth to downstream BSs. In return, the BS earns the same number of credits as that of unused bandwidth. These credits allow the BS to transmit more packets beyond its original quota in a limited number of upcoming cycles (called *window*). Thus, PQOC not only efficiently and fairly allocates bandwidth by a quota that is exerted probabilistically within a fixed-length cycle, but it also regulates a fair use of unused bandwidth, a feature that is particularly effective in metro environments where traffic is naturally bursty.

The implementation of PQOC is described as follows. The quota (Q) for each BS is calculated according to the network configuration, $Q = (2S/(S + 2)) \cdot C \cdot W/N$ [11], where S is the number of S-BSs, C is the number of slots in a cycle, W is the number of wavelength, and N is the number of BSs. With the Q value determined, the probabilistic quota, P_Q , is basically designated as the ratio of the Q value to the total number of slots in a cycle, C . Each control slot contains a header (for synchronization), as well as W minislots that carry the destination addresses, statuses, and priority fields of the corresponding W data slots. There are four distinct statuses for each data slot- BUSY, BUSY/READ (BREAD), IDLE, and IDLE/MRKD (IMRKD). A BS is allowed to transmit data only when it receives the access permission on the basis of probabilistic quota and also finds an IDLE slot. If a BS has no traffic to transmit but attains access permission, the BS will earn the remaining number of slots as credits for future use. Immediately, the BS will update the statuses of the same number of data slots from IDLE to IMRKD. Afterwards, a BS will use its credits to transmit more packets beyond the probabilistic quota on any IMRKD data slots within the window. Additionally, the BS alters the state to BUSY when it transmits a packet. A destination BS that has successfully received (dropped) a packet modifies the slot state from BUSY to BREAD. This allows the next S-BS to erase the

Assumptions: $Quota=2, Cycle Length=4 Slots$

BS 5: At Cycle K , $Credit=2$, 3 arriving packets destined to BSs 1, 3, and 4

Slot	λ_1	λ_2	λ_3	λ_4	Operations
1	2	7	4		Send data (to BS 1) via λ_1 (one $Quota$ left)
2		4	3		Cannot send (Receiv. contention with BSs 3, 4)
3	2				Send data (to BS 4) via λ_3 (no $Quota$ left)
4			5		Send data (to BS 3) via λ_3 using credit, $Credit--=1$

BS 6: At Cycle K , $Credit=0$, 1 arriving packet destined to BS 3

Slot	λ_1	λ_2	λ_3	λ_4	Operations
1	1	7	4		
2		4	3		Cannot send (Receiv. contention with BS 3)
3	2	4			Send data (to BS 3) via λ_4 (one $Quota$ left)
4			3	5	Set λ_1 slot as IMRKD, $Credit++=1$

BS 7: At Cycle K , $Credit=0$, no arriving packet

Slot	λ_1	λ_2	λ_3	λ_4	Operations
1	1	7	4		
2		4	3		Set λ_1 slot as IMRKD, $Credit++=1$
3	2	4	3		
4			3	5	Set λ_2 slot as IMRKD, $Credit++=2$

Legend: : IDLE slot; : IMRKD slot; d : data slot destined to BS d ;

Fig. 2. PQOC: An example.

data slot by changing the status from BREAD back to IDLE, enabling slot reuse by downstream BSs.

The operation of PQOC can be explained via a simple example illustrated in Fig. 2. Recall that, since each BS has only one tunable receiver, the *receiver contention problem* occurs when there is more than one packet destined for the same receiver in a single time slot. In addition, because a BS has only one tunable transmitter, any BS is restricted to accessing at most one wavelength in a single time slot. Such a limitation is referred to as the *vertical-access constraint*.

As shown in Fig. 2, BS 5 has two *Credits*, which allows BS 5 to transmit more than $Quota (=2)$ number of packets in Cycle K . At time slots 1 and 3, BS 5 sends two packets, thus using up the entire $Quota$. Note that at time slot 2, BS 5 fails to send any packet, due to the receiver contention problem. While at time slot 4, BS 5 finds an IMRKD slot, and thus, sends the last packet (to BS 3) by using one *Credit*. Then, the K th Cycle propagates to downstream BS 6. BS 6 sends one data packet using one $Quota$. With one $Quota$ left, BS 6 thus alters one IDLE slot into IMRKD slot at time slot 4, λ_1 . Note that due to *vertical-access constraint*, all BSs can only send/receive at most one data packet at each time slot.

III. INTEGRATED TRAFFIC CONTROL SCHEME (ITCS)

The ITCS inherits the PQOC scheme and seamlessly integrates the CES for wireless backhaul and nCES for background best-effort data over HOPSMAN. In this section, we first describe the basic ITCS design concept, followed by the detailed three-phase access operations of ITCS. Note that, since metro area networks are aggregation networks, we consider the multiplexing of several connections over one CES connection via

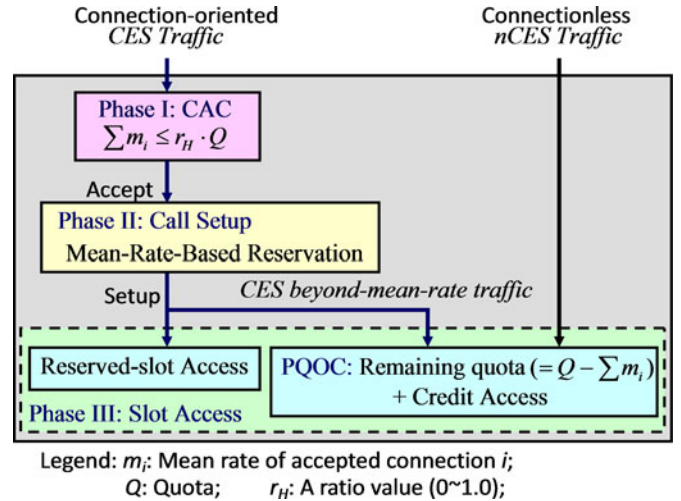


Fig. 3. Integrated traffic control scheme (ITCS).

network connection managers [20], which are beyond the scope of this paper.

A. Basic Design

Most MAC schemes for single-channel ring networks adopt rate-based reservation access control scheme. Unfortunately, for our multichannel HOPSMAN, the rate-based transmission may be hindered due to the receiver contention problem and the vertical-access constraint. Accordingly, the ITCS adopts a slot-marking-based reservation scheme, in which the slots reserved for the CES traffic are clearly specified and marked.

All slots are categorized as reserved high-priority slots or unreserved slots. As shown in Fig. 3, while best-effort nCES traffic always uses unreserved slots, CES traffic uses reserved slots within its mean rate and uses remaining unreserved slots beyond its mean rate. To support CES wireless-backhaul traffic, ITCS operates in three phases (see Fig. 3): call admission phase, call setup phase, and slot access phase. In the first phase, the CAC scheme simply admits the call for a BS if the mean transmission rate of its CES traffic (slots/cycle) does not exceed the bandwidth allocated to the CES traffic of the BS. Once the connection is permitted, in the second phase, the connection is established by marking (i.e., reserving) the required number of slots in every cycle of one entire ring. Note that the priority field in each control minislot is for the marking/reservations of a CES-connection. The field includes a High (H)/Low-priority (L) bit, a CES source address, and a CES destination address. With the source and destination address filled during the marking, the setup process allows slot marking/reservation to be free from the receiver contention problem and be compliant with the vertical-access constraint. In the last phase, the BS then transmits its below-mean-rate CES data on those reserved marked slots, and its beyond-mean-rate traffic on remaining slot bandwidth. If there still remain unused slots, nCES traffic will directly enter the phase III for accessing slots based on the PQOC protocol described earlier.

B. Phase I: Call Admission Control

For slot access, the quota (Q) (i.e., the number of slots in each cycle) of a BS is divided into two parts: reserved quota (which is equal to the total sum of the mean rates of the accepted CES connections) and the remaining quota (which is equal to Q minus the number of reserved quota). Note that the remaining quota along with the accumulated credits, which are derived in the same way as in PQOC, are used for transmitting the CES beyond-mean-rate traffic and nCES unreserved traffic.

The CAC in a BS locally admits a new CES connection only when the sum of the current reserved quota and the mean rate of the new connection are bounded under a predefined quota ratio, i.e., $r_H \cdot Q$:

$$\sum m_i \leq r_H \cdot Q \quad (1)$$

where m_i is the mean rate of each accepted connection, and r_H is a predetermined system parameter which is set between 0 and 1.0, and can be considered as the maximum accepted percentage for the CES traffic. The remaining proportion of quota, $(1 - r_H) \cdot Q$, is called the minimum remaining quota. It is worth pointing out that this CAC guideline is logically identical to the notion of equivalent bandwidth. For example, taking the equivalent bandwidth to be simply twice the mean rate of each connection yields an r_H being equal to 0.5.

C. Phase II: Call Setup

In this phase, the admitted BS marks the required number of high-priority slots in every cycle of one single ring time according to the mean rate and also based on the relative locations of the source and destination BSs of the connection. Namely, the BS has to resolve two issues: 1) how many slots, and 2) which slots to mark for the connection. Note that the markings are flexibly made at any wavelength and at any slot time.

For the first issue, the required number of slots to be marked depends on the relative locations of the source and destination BSs. Assume that there are S server-BSs (S-BS 1 to S-BS S) in the network, dividing the network into S sections (sections 1 to S), with each section containing more than one BS (including the server BS of the section). Due to the geometry of a single ring, when a downstream BS wishes to send data to an upstream BS that is located within the same section, the data will circulate around the ring for more than a single ring time. Therefore, the data in the reserved slot still occupies the slot (not erased by an S-BS) when it circulates a whole ring time before reaching the source BS the second time. This causes the BS to fail in transmitting data for the following ring time. Consequently, to guarantee the mean rate transmission, in the case when a downstream BS wishes to send data to an upstream BS within the same section, the BS marks twice the slots corresponding the mean rate of the connection in each cycle. Notice that in this case, although twice the mean-rate-associated slots are reserved, still only the slots corresponding to the mean rate are used at a time. Unused slots and those slots that are made available after being erased by an S-BS can still be used by all other traffic, resulting in no bandwidth waste. In all other cases, the

BS marks the number of slots corresponding to the mean rate of the connection in each cycle.

For the second issue, the BS chooses slots that are compliant with the vertical-access constraint, and free from the receiver-contention problem. Specifically, the BS first checks if any minislot—in the corresponding control slot—contains its own CES-connection reservation. And then, the BS further checks other BSs' reservations at this time slot (i.e., to check whether any destination address within the priority field of this control slot is the same as the destination BS of this CES connection). If the BS finds any, it is not allowed to perform the marking/reservation at the time slot (it will try to repeat the process at the following time slot within this cycle).

Based on the received marking/reservations information every BS can solve both above issues. Furthermore, the IDLE, IMRKD, and BREAD slots are all candidates for marking because these slots are either empty, or will be erased by the next encountered S-BS. Once the markings are made successfully in all cycles on one ring time, the connection is established and CES data are allowed to be transmitted. If the BS fails to mark during any of the cycle, the BS will attempt to mark in another 2 or 3 rounds of ring times. If these attempts are still unsuccessful, the BS fails to set up the connection and is responsible for unmarking previously marked slots.

D. Phase III: Slot Access

At each slot time, each BS that is ready to transmit first checks whether any minislot—in the corresponding control slot—contains its own CES-connection reservation. If it does, it then accesses the slot and updates the status from IDLE to BUSY. Otherwise, the BS will attempt to access unreserved slots for CES *beyond-mean-rate traffic* (higher priority), or nCES traffic, based on the PQOC scheme. Namely, the access is determined either by the probabilistic quota [11] and the BS's success in finding an IDLE slot, or by using credits to transmit more unreserved traffic beyond the probabilistic quota on any IMRKD data slots. To maximize the throughput, ITCS facilitates a flexible use of marked/reserved bandwidth. For example, if a BS finds a CES connection that has fewer arrivals than its mean rate in a cycle, it allows the unused marked slots to be used either for unreserved traffic of its own, or by other downstream BSs. For the latter case, its credit accumulates accordingly. However, to prohibit subsequent transmissions of the original connections from being obstructed, these packets must be erased by an S-BS before the source BSs of the original connections can use them. The detailed ITCS algorithm is shown in Fig. 4.

IV. CALL SETUP QUEUEING DELAY ANALYSIS

For a CES connection, the total call setup delay is comprised of the call setup queueing delay (Phase I) and the marking delay (Phase II). The call setup queueing delay is defined as the time from a new connection arriving at a BS until it is accepted by the CAC. As will be shown in the next section, the setup marking delay is nearly one ring time, and is thus considered to be a constant. Namely, the call setup queueing delay constitutes the major part of the total call setup delay. In this section, we

Variables

Q : quota;
 P_Q : probabilistic 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 probability quota;
 np_{queue} : number of packets in the queue;
 nh : number of reserved high priority data;
 c : a CES connection;
 $c.r$: mean rate of connection c ;
 $c.src$: source node of connection c ;
 $c.dest$: destination node of connection c ;

Slot type

$Header$: {CYCLE_BEGIN, NORMAL};
 $Status$: {BUSY, BREAD, IDLE, IMRKD};
 $Priority$: {L, (H, $c.src$, $c.dest$)};

Main Process() /*execute at each slot time*/

1. Read the control slot;
2. Add new connections arrivals to a WAIT queue;
3. **for** (each c in WAIT queue) /* CAC */
 - if** ($nh+c.r \leq r_H * Q$) $nh += c.r$;
 Add c to a SETUP queue; **endif endfor**
4. **for** (each c in SETUP queue) /*Setup connections*/
 - if** ($c.src$ and $c.dest$ are in the same network section **and** $c.src > c.dest$)
 Reserve $2 * c.r$ slots per cycle of the ring;
 - else** Reserve $c.r$ slots per cycle of the ring; **endif**
 - if** (c setup ok) Add c to a TRMIT queue; **endif**
- endifor**
5. **if** (slot's $Header$ is CYCLE_BEGIN)
 - Add the number of n_{CES} arrivals of previous cycle and CES arrivals in TRMIT queue to np_{queue} ;
 - Compute nc , P_Q according to [11] (based on np_{queue});
 $ns_{imrkd} = \max(0, Q - np_{queue})$; $np_{PQ} = 0$; **endif**

/* Receive packets */

6. Receive the packet destined to it,
 and update BUSY to BREAD;
- /* Server node erases packets */**
7. **if** (node is S-node) Erase BREAD packets;
 Update BREAD to IDLE; **endif**
- /* Transmission process */**
8. **if** (Randomize(P_Q))
 $np_{PQ}++$; **endif** /*Store transmission allowance*/
/* Transmit reserved data */
if (exist((H, $c.src$, $c.dest$), IDLE))
 Transmit a packet from c of TRMIT queue;
 Update IDLE to BUSY; $np_{PQ}--$;
if (c transmits all) Teardown c ; $nh -= c.r$; **endif**
else /* Transmit unreserved data */
if (exist a vertical-access-constraint-free unreserved data packet)
switch
 - case** ($nc > 0$ and exist(L, IMRKD)):
 /*transmit by credit */
 $nc--$; Transmit the packet; $np_{queue}--$;
 Update IMRKD to BUSY; **break**;
 - case** ($np_{PQ} > 0$ and exist(L, IDLE)):
 /*transmit by quota using IDLE */
 $np_{PQ}--$; Transmit the packet; $np_{queue}--$;
 Update IDLE to BUSY; **break**;
 - case** ($np_{PQ} > 0$ and exist(L, IMRKD)):
 /*transmit by quota using IMRKD */
 $np_{PQ}--$; $ns_{imrkd}++$;
 Transmit the packet; $np_{queue}--$;
 Update IMRKD to BUSY; **break**;
 - otherwise**
 No transmission; **endswitch endif**
9. **if** ($ns_{imrkd} > 0$ and exist(L, IDLE)) /* Mark slots */
 $ns_{imrkd}--$; Update IDLE to IMRKD; **endif**

Fig. 4. Detailed ITCS algorithm.

derive the mean call setup queueing delay based on an M/G/m queueing model. For simplicity, we consider the case where each connection has a mean rate of 1-slot per cycle. That is, each connection uses one quota (considered as a server) when it is set up, and releases one quota when a connection is torn down.

In this model, the arrival process is assumed to be a Poisson process with the arrival rate, λ . The number of servers (m) is given as the maximum admissible quota ($r_H \cdot Q$) of CES traffic since the CAC accepts a new connection if the used reserved quota is under $r_H \cdot Q$. The general service time is the sum of the time spent in the call setup stage (i.e., setup marking delay) and the transmission time. Without loss of generality, we assume that the transmission time is exponentially distributed. Therefore, the service time is an exponential duration added over a constant time (i.e., the marking delay), which is expressed as the general distribution $f_{\bar{x}}(x) = \mu e^{-\mu(x-c)}$, $x \geq c$, where μ is the mean rate of the exponential distribution, and c is a constant time. As stated earlier, the contribution of the constant part c to the mean service time $E[x] = c + (1/m\mu)$ is much smaller than that of the exponentially distributed part being in average $1/m\mu$.

The derivation strategy begins with the following facts and observations regarding the determination of the remaining service times seen by arriving and departing customers (i.e., CES connections). The first fact states that the remaining times for the customers in service that are either found by the Poisson arrivals or are left by Poisson departures are mutually independent and identically distributed in the same way as the distribution of the residual service-life variables [21]. Second, in an M/G/m queueing system, the stationary departure process is a Poisson process if and only if the service distribution is exponentially distributed [22]. This second characteristic presents an obstacle when deriving the remaining service time at departure instants for other general service distributions. Fortunately, we observe that if the general service distribution is similar to an exponential distribution, then the departure process will be similar to a Poisson process. Following this logic, since the particular service distribution in our system is similar to an exponential distribution, we make the assumption in our approximation that the remaining service times left behind by a departing customer follow approximately the distribution of the residual service-life. Thus, a crucial feature of our approximation is that the

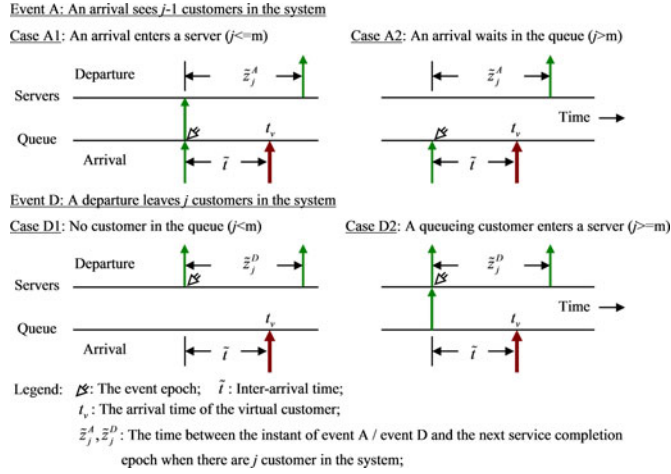


Fig. 5. Occupancy distribution analysis for M/G/m under FCFS.

remaining service times found by arrivals and departures are distributed in the same way as the residual service-life.

Since the service time is denoted by the random variable \tilde{x} , we define \tilde{x}_i^r , ($1 \leq i \leq m$) (m is the server's number), as the residual life for the service time. In other words, $\tilde{x}_1^r, \dots, \tilde{x}_{m-1}^r, \tilde{x}_m^r$ are independent random variables, each with the residual life distribution, $f_{\tilde{x}_i^r}(x) = (1 - F_{\tilde{x}}(x))/E[x]$. Here, index i indicates the number of the residual service-life variables and should not be mistaken for the index of any particular server. Because these random variables are all mutually independent and identically distributed, \tilde{x}_i^r does not necessarily belong to any specific server. Further, with the given service distribution, the remaining service distribution \tilde{x}_i^r can be derived as

$$f_{\tilde{x}_i^r}(x) = \frac{1 - F_{\tilde{x}}(x)}{E[x]} = \begin{cases} e^{-\mu(x-c)}/(c + 1/\mu), & x \geq c \\ 1/(c + 1/\mu), & 0 \leq x \leq c. \end{cases} \quad (2)$$

In the following, we first derive the occupancy distribution by examining the system's equilibrium probability. Let P_j^a be the equilibrium probability that an arriving customer finds j customers in the system; P_j^d denotes the equilibrium probability that a departing customer leaves j customers in the system; and P_j is the stochastic equilibrium occupancy probability. In the M/G/m queueing system, where stationary departure and arrival distributions are equal for systems where arrivals and departures occur one by one [23], $P_j^a = P_j^d$. Due to Poisson arrivals, the arrival, departure, and stochastic equilibrium occupancy distributions are all identical in an M/G/m queueing system (i.e., $P_j = P_j^a = P_j^d$).

Fig. 5 shows a visual outline of the occupancy distribution analysis for M/G/m, which will be referred throughout the rest of the section. First, we assume that a virtual arrival (an arrival in the system at time t_v that does not require any service) finds j ($j > 0$) customers in the system. Prior to this virtual arrival, two events may be possible: the previous event may be an arrival that sees $j-1$ customers in the system (see Fig. 5, Event A), or a departure that leaves j customers behind (see Fig. 5, Event D). While there are j customers in the system, the time elapsed between the arrival (Event A) or departure (Event D) instant and

the beginning of the next service completion epoch is denoted as \tilde{z}_j^A and \tilde{z}_j^D , respectively. Let \tilde{t} denote the interarrival time, an exponential random variable with parameter λ . Note that the durations $\tilde{z}_j^A, \tilde{z}_j^D$ are clearly greater than \tilde{t} ; otherwise, another event (a departure) would happen before t_v , which contradicts the aforementioned assumptions made, yielding

$$P_j = P_{j-1}^a \cdot P(\tilde{t} < \tilde{z}_j^A) + P_j^d \cdot P(\tilde{t} < \tilde{z}_j^D). \quad (3)$$

Since the arrival, departure, and stochastic equilibrium occupancy distributions are identical, we arrive at

$$P_j = P_{j-1} \cdot P(\tilde{t} < \tilde{z}_j^A) + P_j \cdot P(\tilde{t} < \tilde{z}_j^D). \quad (4)$$

Here

$$\begin{aligned} P(\tilde{t} \leq \tilde{z}) &= \int_0^\infty \int_0^{\tilde{z}} dF_{\tilde{t}}(t) dF_{\tilde{z}}(z) = \int_0^\infty (1 - e^{-\lambda z}) dF_{\tilde{z}}(z) \\ &= 1 - F_{\tilde{z}}^*(\lambda) \end{aligned} \quad (5)$$

where $F_{\tilde{z}}^*(\lambda)$ is the Laplace Transform of \tilde{z} with parameter λ , and \tilde{z} can be replaced by \tilde{z}_j^A or \tilde{z}_j^D .

We then derive the occupancy distribution based on the assumption that the remaining service times found by arrivals and departures are distributed in the same way as the residual service-life, \tilde{x}_i^r . The detailed derivation is presented as follows. When an arrival comes seeing $j-1$ customers in the system, it either enters an idle server directly (causing its remaining service time to be \tilde{x}), or waits in the queue when all the servers are occupied. This causes the service time \tilde{z}_j^A to be distributed as $\min\{\tilde{x}, \tilde{x}_1^r, \tilde{x}_2^r, \dots, \tilde{x}_{j-1}^r\}$ when $j \leq m$, or $\min\{\tilde{x}_1^r, \tilde{x}_2^r, \tilde{x}_3^r, \dots, \tilde{x}_m^r\}$ when $j > m$ (see Fig. 5, cases A1 and A2, respectively). When a departure occurs, it either leaves j customers who are already in service or releases a server for the first customer in the queue. Thus, the service time \tilde{z}_j^D is distributed as $\min\{\tilde{x}_1^r, \tilde{x}_2^r, \tilde{x}_3^r, \dots, \tilde{x}_j^r\}$ when $j < m$, or $\min\{\tilde{x}, \tilde{x}_1^r, \tilde{x}_2^r, \dots, \tilde{x}_{m-1}^r\}$ when $j \geq m$ (see Fig. 5, cases D1 and D2, respectively). By taking this and Eq. (4) and (5) into account, the occupancy distribution P_j can be obtained, as shown in Eq. (6), as:

$$P_j = \begin{cases} P_0(1 - F_{\tilde{x}}^*(\lambda))/F_{\tilde{x}_1^r}^*(\lambda), & \text{if } j = 1 \\ \begin{aligned} &P_{j-1}(1 - F_{\min(\tilde{x}, \tilde{x}_1^r, \tilde{x}_2^r, \dots, \tilde{x}_{j-1}^r)}^*(\lambda)) \\ &/ F_{\min(\tilde{x}_1^r, \tilde{x}_2^r, \tilde{x}_3^r, \dots, \tilde{x}_j^r)}^*(\lambda), \end{aligned} & \text{if } 1 < j < m \\ \begin{aligned} &P_{j-1}(1 - F_{\min(\tilde{x}, \tilde{x}_1^r, \tilde{x}_2^r, \dots, \tilde{x}_{m-1}^r)}^*(\lambda)) \\ &/ F_{\min(\tilde{x}, \tilde{x}_1^r, \tilde{x}_2^r, \dots, \tilde{x}_{m-1}^r)}^*(\lambda), \end{aligned} & \text{if } j = m \\ \begin{aligned} &P_{j-1}(1 - F_{\min(\tilde{x}_1^r, \tilde{x}_2^r, \tilde{x}_3^r, \dots, \tilde{x}_m^r)}^*(\lambda)) \\ &/ F_{\min(\tilde{x}, \tilde{x}_1^r, \tilde{x}_2^r, \dots, \tilde{x}_{m-1}^r)}^*(\lambda), \end{aligned} & \text{if } j > m. \end{cases} \quad (6)$$

To proceed, we conduct a Laplace Transform of \tilde{x} and \tilde{x}_r with parameter λ :

$$\begin{aligned} F_{\tilde{x}}^*(\lambda) &= \int_{-\infty}^\infty e^{-\lambda x} f_{\tilde{x}}(x) dx = \int_c^\infty e^{-\lambda x} \mu e^{-\mu(x-c)} dx \\ &= \frac{\mu}{\lambda + \mu} e^{-\lambda c} \end{aligned} \quad (7)$$

$$\begin{aligned}
F_{\tilde{x}_i^*}^*(\lambda) &= \int_{-\infty}^{\infty} e^{-\lambda x} f_{\tilde{x}_i^*}(x) dx = \int_c^{\infty} \frac{e^{-\lambda x} e^{-\mu(x-c)}}{c+1/\mu} dx \\
&\quad + \int_0^c \frac{e^{-\lambda x}}{c+1/\mu} dx \\
&= \frac{1}{(\lambda+\mu)(c+1/\mu)} e^{-\lambda c} + \frac{1}{\lambda(c+1/\mu)} (1 - e^{-\lambda c}).
\end{aligned} \tag{8}$$

Assuming that $\tilde{b}_1 = \min(\tilde{x}, \tilde{x}_1^r, \tilde{x}_2^r, \dots, \tilde{x}_{k-1}^r)$, $k = 2, 3, \dots, m$, the PDF of \tilde{b}_1 is

$$\begin{aligned}
F_{\tilde{b}_1}(b) &= P(\tilde{b}_1 \leq b) = 1 - P(\tilde{b}_1 > b) \\
&= 1 - P(\tilde{x} > b, \tilde{x}_1^r > b, \tilde{x}_2^r > b, \dots, \tilde{x}_{k-1}^r > b) \\
&= 1 - P(\tilde{x} > b) P(\tilde{x}_i^r > b)^{k-1} \\
&= \begin{cases} 1 - e^{-\mu(b-c)} [e^{-\mu(b-c)}/(\mu c + 1)]^{k-1}, & b \geq c \\ 1 - [(\mu c + 1 - \mu b)/(\mu c + 1)]^{k-1}, & b < c \end{cases}
\end{aligned} \tag{9}$$

yielding

$$\begin{aligned}
F_{\tilde{b}_1}^*(\lambda) &= \int_0^{\infty} e^{-\lambda b} dF_{\tilde{b}_1}(b) = k e^{-\lambda c} / (\mu^{k-2} (c+1/\mu)^{k-1} (\lambda + k\mu)) \\
&\quad - (k-1) e^{-\lambda(c+1/\mu)} \int_1^{c\mu+1} w^{k-2} e^{\lambda w(c+1/\mu)} dw.
\end{aligned} \tag{10}$$

Assuming that $\tilde{b}_2 = \min(\tilde{x}_1^r, \tilde{x}_2^r, \tilde{x}_3^r, \dots, \tilde{x}_k^r)$, $k = 2, 3, \dots, m$, the PDF of \tilde{b}_2 is

$$\begin{aligned}
F_{\tilde{b}_2}(z) &= P(\tilde{b}_2 \leq b) = 1 - P(\tilde{b}_2 > b) \\
&= 1 - P(\tilde{x}_1^r > b, \tilde{x}_2^r > b, \tilde{x}_3^r > b, \dots, \tilde{x}_k^r > b) \\
&= 1 - p(\tilde{x}_i^r > b)^k \\
&= \begin{cases} 1 - [e^{-\mu(b-c)}/(\mu c + 1)]^k, & b \geq c \\ 1 - [(\mu c + 1 - \mu b)/(\mu c + 1)]^k, & b < c \end{cases}
\end{aligned} \tag{11}$$

which leads to

$$\begin{aligned}
F_{\tilde{b}_2}^*(\lambda) &= \int_0^{\infty} e^{-\lambda b} dF_{\tilde{b}_2}(b) = \frac{k e^{-\lambda c}}{\mu^{k-1} (c+1/\mu)^k (\lambda + k\mu)} \\
&\quad - k e^{-\lambda(c+1/\mu)} \int_1^{c\mu+1} w^{k-1} e^{\lambda w(c+1/\mu)} dw.
\end{aligned} \tag{12}$$

We can now substitute the numerical results of Eq. (7), (8), (10), and (12) into Eq. (6). From the requirement that the probabilities sum to unity ($\sum P_j = 1$), the occupancy distribution (P_j) can thus be derived. A final step involves letting \tilde{n}_q be the expected number of customers in the queue. By applying Little's formula, we have

$$E[\tilde{w}] = E[\tilde{n}_q]/\lambda = \sum_{j=m}^{\infty} (j-m) \cdot P_j/\lambda \tag{13}$$

where \tilde{w} is the call setup queueing delay. This concludes the derivation for the mean queueing delay of an M/G/m queue.

V. SIMULATION RESULTS

In this section, we evaluate the performance of the ITCS with respect to: 1) Throughput comparisons, 2) CES call setup delay, and 3) CES and nCES delay and jitter. The values used for the parameters in the simulation are listed below. The network has a total of 20 BSs ($N = 20$), where BS 1 is designated as an S-BS. There are 21 cycles on the ring ($R = 21$), with each cycle consisting of 100 slots per wavelength ($C = 100$). The optical fiber accommodates nine wavelengths ($W = 9$). Note that while the simulation parameters are chosen to clearly demonstrate our system's performance, they may be flexibly adapted to hold any other value. Traffic destinations are assumed to be uniformly distributed among all BSs. To focus on the CES traffic performance, there is only a single server BSs ($S = 1$). The quota for each BS is then calculated to be $Q = (2S/(S+2)) \cdot C \cdot W/N = 30$ [11]. The system load (L) represents the normalized load per slot per wavelength, including both a load of CES (high priority) data (L_H), and a load of nCES (low priority) data (L_L). Clearly, $L = L_H + L_L$.

The interarrival time between generated packets, for both CES and nCES traffic, is exponentially distributed. The CES traffic is generated with the following parameters. The mean rate of a connection is 1 slot per cycle, while the mean length of a connection is $\bar{l} + 2R$ slots ($\bar{l} = 2000$), the number of new connections arrivals follows a Poisson distribution with parameter $\lambda = Q \cdot L_H / (\bar{l} + 2R)$, and the aggregate burstiness of CES connections is denoted as B . These parameters are set according to the following principles. First, the mean rates of all the connections are assumed to simply remain at 1 slot per cycle, since a connection with a mean rate of x slots per cycle can be equally regarded as x number of connections with a mean rate of 1 slot per cycle. Second, we observe that if the length of a connection is shorter than a single ring time, we need extra management to tear down connections. Considering the high burstiness of CES connections and the fact that the length of a connection is longer than a few number of ring times, the length of a connection is set to be at least 2 ring times ($= 2R$). Finally, the bursty arrival of CES traffic is generated with a two-state (H and L) Markov Modulated Poisson Process (MMPP) [24]. More 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, if $\lambda_L = 0$, the mean arrival rate can be expressed as $\beta \times \lambda_H / (\alpha + \beta)$, while the aggregate burstiness (B) of CES connections can be expressed as $B = (\alpha + \beta) / \beta$. Finally, nCES traffic is generated according to load of nCES (low priority) data (L_L).

As described in the ITCS algorithm, a maximum allocation ratio r_H , ($0 < r_H \leq 1.0$) is also proposed to constrain the maximum ratio bound of accepted high-priority CES data. Therefore, we will only examine the nonoverloaded simulation results, where $L_H \leq r_H$. Simulations are terminated after reaching a

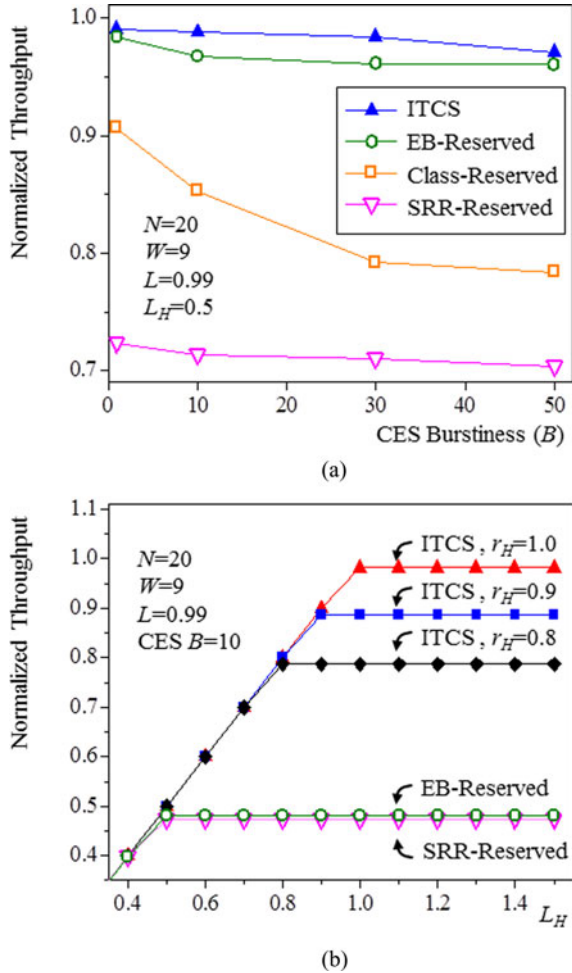


Fig. 6. Throughput comparisons. (a) System throughput comparison. (b) CES traffic throughput comparison.

95% confidence interval with a minimal sample size of 10^7 cycles.

A. Throughput Performance Comparisons

We draw comparisons of the throughput performance between ITCS and three other schemes: EB-Reserved, SRR-Reserved [2], and Class-Reserved [3] schemes. The EB-Reserved scheme is based on the same methodology as that of our scheme, except that the *equivalent bandwidth* to be reserved is twice the mean rate of each connection. The SRR-Reserved scheme makes reservation at their corresponding preferential frame-based slots, which are preassigned on a per-destination basis. The Class-Reserved scheme makes high-priority marks at the control channel, and makes such information known to all nodes whenever a node fails to transmit any high priority data.

Fig. 6 shows the results of throughput comparisons. As shown in Fig. 6(a), the SRR-reserved scheme exhibits the lowest system throughput. This is because each node receives data through its home channel, resulting in poor statistical multiplexing gain. For the class-reserved scheme, its throughput deteriorates with the increase of CES traffic burstiness. This is because exceedingly redundant CES reservations hinder the transmissions of low-priority data. Fig. 6(b) shows that the SRR-Reserved scheme

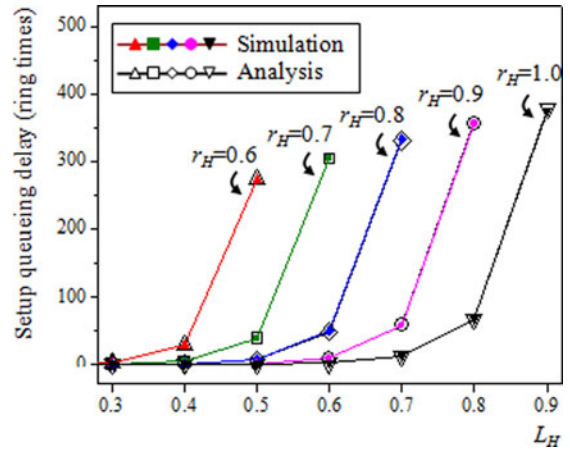


Fig. 7. Analytic and simulation results on mean setup queuing delay.

supports only medium loads of CES traffic due to reservations on fixed slots per each frame. Since the class-reserved scheme does not bring in the CAC function, its CES-traffic throughput performance is unobtainable.

As in our scheme and the EB-Reserved scheme, they almost fully utilize the bandwidth (the normalized throughput is nearly 1.0). However, due to using twice the mean rate for the reservation, the EB-Reserved scheme only achieves a maximal normalized CES throughput of 0.5. The results demonstrate that, the ITCS supports the maximum throughput of CES traffic that is nearly the same as the values of r_H , and also achieves an exceptional system throughput performance.

B. Call Setup Delay Performance

Recall that the total setup delay consists of setup queuing delay and setup marking delay. We first draw a comparison between analytic and simulation results. For the analytic results, the parameters of a multiple-server queueing system (M/G/m) are given as follows. The arrival process for new connections follows a Poisson process with parameter $\lambda = Q \cdot L_H / (\bar{l} + 2R)$, while the number of servers (m) is set as the quota for CES traffic, i.e., $m = \lfloor r_H \cdot Q \rfloor$. The service distribution (the service time includes constant setup marking delay and the connection duration) follows $f_{\tilde{x}}(x)$, where \tilde{x} represents the service time, $\mu = 1/\bar{l}$, and $c = 3R$. By varying r_H and L_H and by referring to the proposed analytic computation of an M/G/m queue from the Section IV, we can obtain the mean setup queuing delay, as plotted in Fig. 7.

In Fig. 7, we clearly show that analytical results are in complete agreement with simulation results. Further, the results also indicate that the mean setup queuing delay rises as L_H increases with any specific r_H , as expected. On the other hand, under the same load (L_H remains the same), the mean setup queuing delay decreases when r_H increases. This is because when r_H increases, more CES connections can be accommodated, thus reducing the setup queuing delay.

We next examine the setup marking delay via simulation results. As our study shows, there exists a sufficient number of empty slots for the connection setup under $L_H \leq 0.9$. The results are verified via the simulation results that are shown in

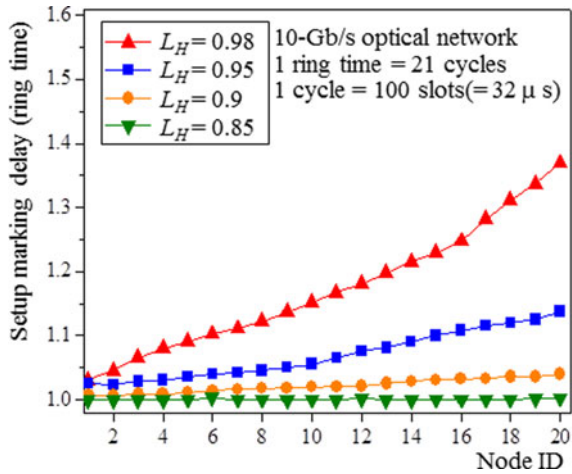
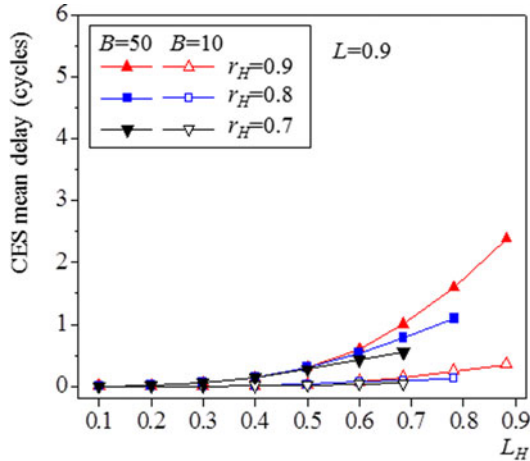
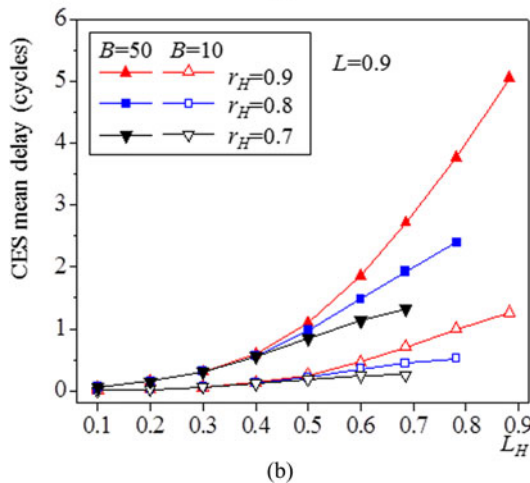


Fig. 8. Setup marking delay.



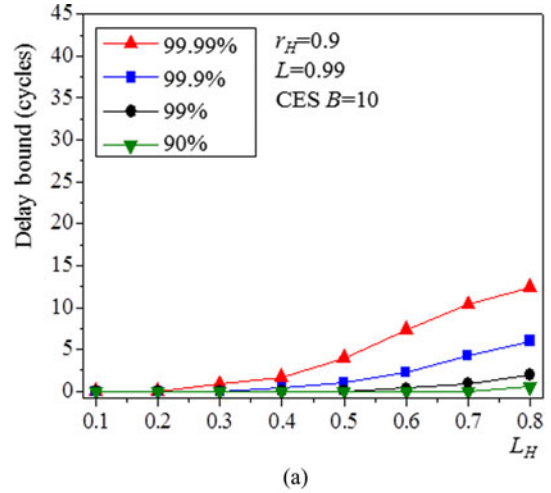
(a)



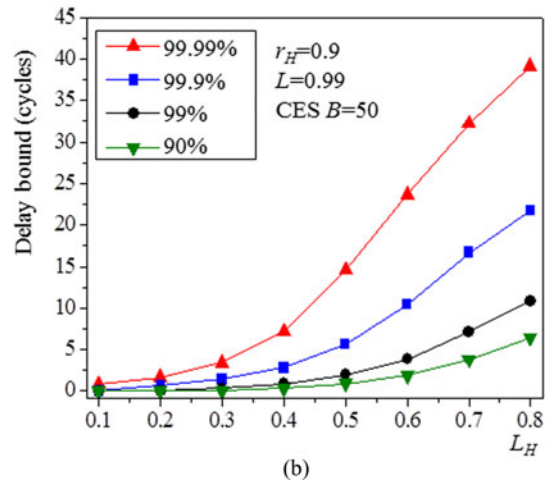
(b)

Fig. 9. Mean delay and jitter performance of CES traffic. (a) CES traffic delay under various r_H and burstiness. (b) CES traffic jitter under various r_H and burstiness.

Fig. 8. We depict in the figure that the setup marking delay is close to one ring time under $L_H < 0.95$. When $L_H = 0.98$, the marking delay slightly increases especially for the downstream BSs, due to the deteriorating receiver-contention problem under exceedingly high loads. At the first glance, there exists an unfairness problem of increasing marking delay for downstream BSs



(a)



(b)

Fig. 10. Delay bound of CES traffic. (a) Delay bound under low burstiness of CES traffic. (b) Delay bound under high burstiness of CES traffic.

under exceeding high loads. However, it is worth noting that, since the marking delay is much smaller than the setup queuing delay in this environment, such less-than-half of a ring time can be ignored.

C. CES and nCES Delay/Jitter Performance

Fig. 9 shows the results of the mean CES delay and jitter with various settings of r_H applied, under all admissible values of L_H (where the maximum value of L_H for each curve is nearly equal to the value of r_H and the same as the maximum achievable throughput obtained from the simulation results). The results demonstrate almost negligible delay and minimal jitter under a low burstiness ($B = 10$) condition, and maintain reasonably low delay and jitter (just a few number of cycles) under a higher burstiness ($B = 50$) situation. Moreover, under high-bursty fluctuated CES traffic, since a smaller value of r_H will increase the minimum guaranteed remaining bandwidth, it is expected that both the delay and jitter performance improve provided with a smaller value of r_H . The results can be used to determine proper values of r_H under various traffic burstiness.

Another key measurement is the delay bound of CES traffic, as shown in Fig. 10. Under a saturated system load ($L = 0.99$), we draw 99.99%, 99.9%, 99%, and 90% delay bounds. These

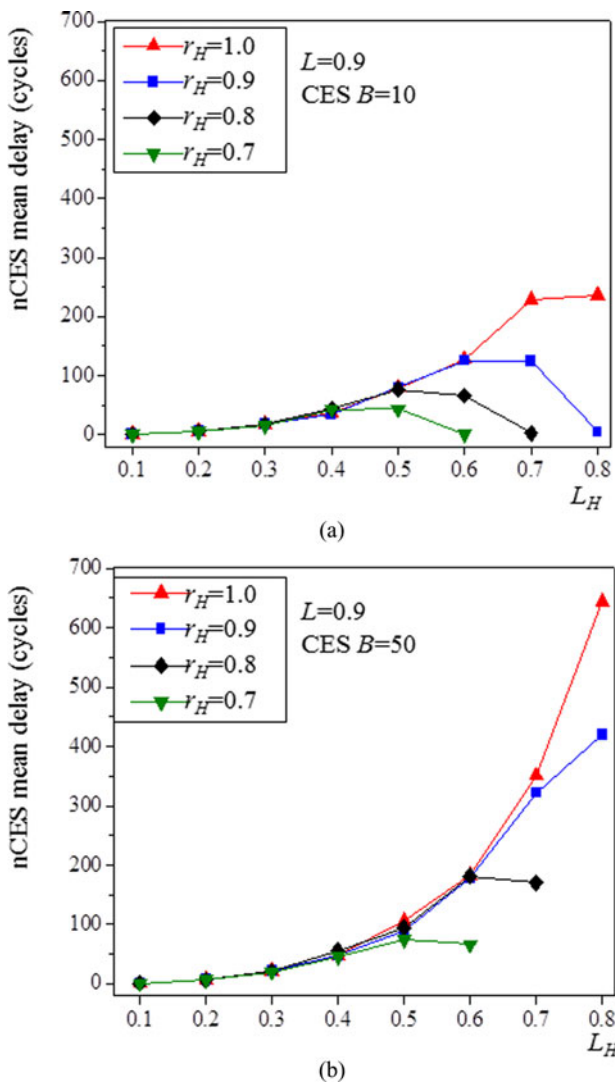


Fig. 11. nCES delay under various CES burstiness and loads. (a) nCES delay under low burstiness of CES traffic. (b) nCES delay under high burstiness of CES traffic.

results demonstrate that the ITCS assures of low bounds under various loads and burstiness of CES traffic for wireless backhaul. For example, under a medium-to-heavy load ($L_H = 0.8$) and burstiness $B = 10$, the ITCS assures of 99% and 99.9% delay bounds of 0.06 and 0.19 ms, respectively. For traffic with even higher burstiness, $B = 50$, the ITCS guarantees 99% and 99.9% delay bounds still being as low as 0.34 and 0.69 ms, respectively. Moreover, by strategically setting the value of r_H to constrain the maximum value of L_H , this system is capable of meeting the network's various delay bound requirements. While determining the exact value of r_H is network management problem that lies beyond the scope of this paper, our simulation results still serve as important guidelines for determining the optimal value of r_H that maximizes the system performance.

We next examine how nCES delay is impacted by various CES traffic loads. As Fig. 11 details, the mean access delay of nCES traffic first rises with increasing CES traffic loads. This result should be expected since heavier CES traffic will

hinder the transmission of nCES traffic, thus increasing the delay. Rather surprisingly, however, nCES delay actually drops at the highest value of L_H for each curve under CES traffic with $B = 10$, as plotted in Fig. 11(a). This occurs because, as the CES load increases, the nCES load decreases, and the minimum remaining quota, $(1 - r_H) \cdot Q$, just happens to be enough for nCES traffic when $L_L \leq 1 - r_H$. On the other hand, when the system is placed under high burstiness of CES traffic ($B = 50$) [see Fig. 11(b)], the impact on nCES delay increases and the drops for each curve disappear. Putting these special cases aside, the nCES delay is still mostly affected by loads imposed by CES traffic. Note that the cycle time, for a 10-Gb/s WDM metro ring network with 100 slots per cycle and 100 bytes per slot, is 8 μ s. Further, even in the worst case, the mean delay of nCES traffic is merely up to a few milliseconds, still highly acceptable for nCES traffic.

VI. CONCLUSION

We have proposed a new MAC mechanism, ITCS, which seamlessly integrates the CES wireless backhaul and non-CES (nCES) best-effort data over HOPSMAN. ITCS supports CES by a novel and flexible mean-rate-based call admission and slot-marking reservation scheme. The mean call setup delay is derived via an accurate approximation based on an M/G/m queueing analysis. Simulation results demonstrate that the ITCS accommodates a remarkably high CES traffic amount, while satisfying a wide range of delay/jitter requirements. Under a medium-to-heavy load ($L_H = 0.8$) and high burstiness $B = 50$, for example, the ITCS guarantees 99% and 99.9% delay bounds of being as low as 0.34 and 0.69 ms, respectively. The results show that the ITCS successfully meets stringent delay requirements for supporting wireless backhaul.

REFERENCES

- [1] L. Wang, M. Hamdi, R. Manivasakan, and D. Tsang, "Multimedia-MAC protocol: Its performance analysis and applications for WDM networks," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 518–531, Mar. 2006.
- [2] K. Bengi and H. Van As, "Efficient QoS support in a slotted multihop WDM metro ring," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 1, pp. 216–227, Jan. 2002.
- [3] H. Leligou, J. Angelopoulos, C. Linardakis, and A. Stavdas, "A MAC protocol for efficient multiplexing QoS-sensitive and best-effort traffic in dynamically configurable WDM rings," *J. Comput. Netw.*, vol. 44, no. 3, pp. 305–317, Feb. 2004.
- [4] O. Tipmongkolsilp, S. Zaghoul, and A. Jukan, "The evolution of cellular backhaul technologies: Current issues and future trends," *IEEE Commun. Surveys Tutorials*, vol. 13, no. 1, pp. 97–113, First Quarter 2011.
- [5] M. Cao, X. Wang, S. Kim, and M. Madhian, "Multi-Hop wireless backhaul networks: A cross-layer design paradigm," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 738–748, 2007.
- [6] N. Ghazisaidi, M. Maier, and C. Assi, "Fiber-Wireless (FiWi) access networks: A survey," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 160–167, Feb. 2009.
- [7] A. Hekkala, M. Lasanen, I. Harjula, L. Vieira, N. Gomes, A. Nkansah, S. Bittner, F. Diehm, and V. Kotsch, "Analysis of and compensation for non-ideal Rof links in DAS," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 52–59, Jun. 2010.
- [8] W. Shaw, S. Wong, N. Cheng, K. Balasubramanian, X. Zhu, M. Maier, and G. Kazovsky, "Hybrid architecture and integrated routing in a scalable optical-wireless access network," *J. Lightw. Technol.*, vol. 25, no. 11, pp. 3443–3451, Nov. 2007.

- [9] M. Milosavljevic, M. Thakur, P. Kourtessis, J. Mitchell, and J. Senior, "Demonstration of wireless backhauling over long-reach PONs," *J. Lightw. Technol.*, vol. 30, no. 5, pp. 811–817, Mar. 2012.
- [10] M. Yuang, I. Chao, B. Lo, P. Tien, J. Chen, C. Wei, Y. Lin, S. Lee, and C. Chien, "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, Jul. 2008.
- [11] M. Yuang, I. Chao, and B. Lo, "HOPSMAN: An experimental optical packet-switched metro WDM ring network with high-performance medium access control," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 2, no. 2, pp. 91–101, Feb. 2010.
- [12] V. Nguyen and T. Atmaca, "Metropolitan optical packet bus-based networks: Packet bursting and emulation of TDM services," *Computer Comm.*, vol. 33, pp. S110–S121, Nov. 2010.
- [13] Z. Ghebretensae, J. Harmatos, and K. Gustafsson, "Mobile broadband backhaul network migration from TDM to carrier ethernet," *IEEE Commun. Mag.*, vol. 48, no. 10, pp. 102–109, Oct. 2010.
- [14] Y. Takahashi, "An approximation formula for the mean waiting time of an M/G/c queue," *J. Oper. Res. Soc. Japan*, vol. 20, no. 3, pp. 150–163, Sep. 1977.
- [15] C. Wang and R. Wolff, "The M/G/c queue in light traffic," *Queueing Syst.*, vol. 29, no. 1, pp. 17–34, Aug. 1998.
- [16] C. Wang and R. Wolff, "Systems with multiple servers under heavy-tailed workloads," *Perf. Eval.*, vol. 62, no. 4, pp. 456–474, Aug. 2005.
- [17] O. Boxma, J. Cohen, and N. Huffels, "Approximations of the mean waiting time in an M/G/s queueing system," *Oper. Res.*, vol. 27, no. 6, pp. 1115–1127, Nov. 1979.
- [18] J. De Smit, "A numerical solution for the multi-server queue with hyper-exponential service times," *Oper. Res. Lett.*, vol. 2, no. 5, pp. 217–224, 1983.
- [19] F. Barceló and J. Paradells, "The M/H²/s queue in mobile communications: Approximation of the mean waiting time," presented at the IEEE 14th U.K. Teletraffic Symp., London, U.K., 1997.
- [20] S. Giordano, J.-Y. Le Boudec, P. Oechslin, and S. Robert, "VBR over VBR: The homogeneous, loss-free case," in *Proc. IEEE INFOCOM*, 1997, pp. 168–176.
- [21] R. Cooper and S. Niu, "Benes's formula for M/G/1-FIFO 'Explained' by preemptive-resume LIFO," *J. Appl. Probability*, vol. 23, no. 2, pp. 550–554, Jun. 1986.
- [22] F. Kelly, "The departure process from a queueing system," *Math. Proc. Cambridge Philosophical Soc.*, vol. 80, no. 2, pp. 283–285, Sep. 1976.
- [23] R. Cooper, *Introduction to Queueing Theory*. Washington, DC, USA: CEE Press, 1990.
- [24] W. Fischer and K. Meier-Hellstern, "The Markov-modulated poisson process (MMPP) cookbook," *Perf. Eval.*, vol. 18, no. 2, pp. 149–171, Sep. 1993.

I-Fen Chao received the M.S. and Ph.D. degrees in computer science and information engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1994 and 2010, 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. She is currently an Assistant Professor in the Department of Electrical Engineering at Yuan Ze University, Zhongli, Taiwan. Her current research interests include high-speed networking, optical networking, and performance modeling and analysis.

Maria C. Yuang received the B.S. degree in applied mathematics from the National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1978, the M.S. degree in computer science from the University of Maryland, College Park, MD, USA, in 1981, and the Ph.D. degree in electrical engineering and computer science from the Polytechnic Institute of New York University (NYU), New York, NY, USA, in 1989. From 1981 to 1990, she was with AT&T Bell Laboratories and Bell Communications Research (Bellcore), where she was a member of technical staff working on Broadband networks and protocol engineering. In 1990, she joined NCTU, where she is currently a Professor in the Department of Computer Science and Information Engineering. She served as a Guest Editor for the Special Issue of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on Next-Generation Broadband Optical Access Network Technologies in 2010. She has served on the Technical Program Committee of many international conferences including IEEE ICC and GLOBECOM, and has been invited to give invited talks at numerous technical conferences. Her current research interests include optical networks, optical data center networking, wireless access networks, and performance modeling and analysis. She is a senior member of IEEE, and a member of OSA. She holds 32 patents in the field of optical and wireless networking, and has more than 110 publications, and a book chapter.