Dynamic Bandwidth Allocation for Quality-of-Service Over Ethernet PONs

Chadi M. Assi, *Student Member, IEEE,* Yinghua Ye, *Member, IEEE,* Sudhir Dixit, *Member, IEEE,* and Mohamed A. Ali

***Abstract*—Ethernet-based passive optical network (EPON) technology is being considered as a promising solution for next-generation broadband access network due to the convergence of low-cost Ethernet equipment and low-cost fiber infrastructure. A major feature for this new architecture is the use of a shared transmission media between all users; hence, medium access control arbitration mechanisms are essential for the successful implementation of EPON: i.e., ensure a contention-free transmission and provide end users with an equal access to the shared media. In this paper, we propose to use the multipoint control protocol defined within the IEEE 802.3ah Task Force to arbitrate the transmission of different users, and we present different dynamic bandwidth allocation (DBA) algorithms to effectively and fairly allocate bandwidths between end users. These DBA algorithms are also augmented to support differentiated services: a crucial requirement for a converged broadband access network with heterogeneous traffic. We show that queueing delays under strict bandwidth allocation algorithms results in an unexpected behavior for certain traffic classes, and we suggest the use of DBA with appropriate local queue management to alleviate this inappropriate behavior. We conduct detailed simulation experiments to study the performance and validate the effectiveness of the proposed protocols.**

***Index Terms*—Dynamic bandwidth allocation (DBA), Ethernet-based passive optical network (EPON), quality-of-service (QoS), scheduling, simulation and modeling.**

I. Introduction

RAPID deployment of broadband services in the residential and small business area has played an important role in the evolution of access networks. Currently, Ethernet-based passive optical networks (EPONs) [1] are being considered as a promising solution for the next generation broadband access network (known also as the last mile access network) due to the convergence of low-cost Ethernet equipment and low-cost of fiber infrastructure. A passive optical network (PON) is a point-to-multipoint optical access network with no active elements in the signal path from source to destination. Here, all transmissions are performed between an optical line terminal (OLT) and optical network units (ONUs). The OLT resides in the central office (CO) and connects the optical access network to the metropolitan area network (MAN) or wide-area network (WAN). On the other hand, each ONU is usually located at either the curb [i.e., fiber-to-the-curb (FTTC) solution] or the end-user location [i.e., fiber-to-the-building (FTTB) and fiber-to-the-home (FTTH)], and provides broadband video, data, and voice services.

An EPON is a PON that carries all data encapsulated in Ethernet frames and is backward compatible with existing IEEE 802.3 Ethernet standards, as well as other relevant IEEE 802 standards. Moreover, Ethernet is an inexpensive technology that is ubiquitous and interoperable with a variety of legacy equipment; a step forward in making it most suitable for delivering Internet protocol (IP)-based applications and multimedia traffic over PON.

In the downstream direction, the OLT has the entire bandwidth of the channel to transmit data packets and control messages to the ONUs; in this broadcast and select architecture, all active ONUs listen to the channel and only the designated ONU will deliver the received traffic to its end users. On the other hand, in the upstream direction, a PON is a multipoint to point [1], [2] network, where multiple ONUs share the same transmission channel. Here, unless some kind of regulation is implemented, data streams transmitted simultaneously from different ONUs may still collide. Hence, access to the shared medium must be arbitrated by medium access control (MAC) protocols to prevent collisions between Ethernet frames of different ONUs transmitting simultaneously. In general, this is achieved by allocating a transmission window (or timeslot) to each ONU; each ONU should buffer data packets received from different subscribers until they are transmitted in the assigned time window. When the assigned time window arrives, the ONU will burst out frames at full channel speed.

One distinguishing feature that broadband EPON is expected to support is the ability to deliver services to emerging IP-based multimedia traffic with diverse quality-of-service (QoS) requirements [7]. A promising approach to support differentiated QoS is to employ a central controller that can dynamically allocate bandwidth to end users according to the traffic load. Thus, bandwidth management for fair bandwidth allocation among different ONUs will be a key requirement for the MAC protocols in the emerging EPON based networks. In this paper, we discuss an EPON architecture that supports differentiated services; we classify services into three priorities as defined in [5], namely the best effort (BE), the assured forwarding (AF), and expedited forwarding (EF). While EF services (such as voice and other delay sensitive applications) require bounded end-to-end delay and jitter specifications, AF is intended for services that are not delay sensitive but which

Manuscript received January 2, 2003; revised August 15, 2003.

C. M. Assi is with the Concordia Institute, Information Systems Engineering Department, Concordia University, Montreal, QC H3G 1M8, Canada (e-mail: assi@ciise.concordia.ca).

Y. Ye and S. Dixit are with the Nokia Research Center, Burlington, MA 01803 USA (e-mail: Yinghua.Ye@nokia.com; Sudir.Dixit@nokia.com).

M. A. Ali is with the Electrical Engineering Department, Graduate School of The City University of New York, New York, NY 10016-4309 USA (e-mail: eeali@ees1s0.engr.ccny.cuny.edu).

Digital Object Identifier 10.1109/JSAC.2003.818837

require bandwidth guarantees. Finally, BE applications (such as e-mail services) are neither delay sensitive nor do they require any jitter specifications.

We propose a dynamic bandwidth allocation (DBA) algorithm with QoS support over EPON-based access network. We investigate how gated transmission mechanisms [e.g., multipoint control protocol (MPCP)] [5] and DBA schemes can be combined with priority scheduling and queue management to implement a cost-effective EPON network with differentiated services support.

The rest of the paper is organized as follows. Section II presents a background to motivate our work. In Section III, we review the basic principles of MPCP. Different queue management and priority queueing mechanisms are presented in Section IV. Dynamic bandwidth allocation algorithms with QoS support are presented in Section V. Section VI presents the simulation results and Section VII concludes the work.

II. Background and Motivation

In EPON-based network, all ONUs share the same transmission channel while sending traffic in the upstream direction; thus, MAC arbitration mechanisms are required to avoid data collision and to fairly share the channel capacity. To achieve this, one needs to allocate a non overlapping transmission window (timeslot) to each ONU. The timeslot may be fixed (static) or variable (dynamic) based on the arbitration mechanism implemented at the OLT. In [2], the authors studied the performance of EPON using a fixed bandwidth assignment algorithm when all traffic belonged to a single class, i.e., no service differentiation. While this scheme is simple, it had a drawback that no statistical multiplexing between the ONUs was possible; in other words, since each ONU is allocated a fixed timeslot, light-loaded ONUs will probably under utilize their allocated slots, leading to increased delay to other ONUs and eventually deteriorate the throughput of the system. To cope with this problem, [3] proposed an OLT-based polling scheme, called interleaved polling with adaptive cycle time (IPACT). In principle, IPACT uses an interleaved polling approach, where the next ONU is polled before the transmission from the previous one has arrived. Different bandwidth allocation algorithms were studied, namely: fixed, limited, gated, constant credit, linear credit, and elastic. Amongst these algorithms, the limited (where the OLT grants an ONU the requested number of bytes but no more than a predefined value, W_{MAX}) exhibits the best performance. Although this scheme provides statistical multiplexing and results in efficient channel utilization, the algorithm is not suitable for delay and jitter sensitive services because of a variable polling cycle time. More recently, the authors of [4] studied how priority scheduling can be combined with dynamic bandwidth allocation. Unlike IPACT, here the arbitration mechanism is based on the MPCP [6] developed by the IEEE 802.3ah Task Force. The authors use a combination of limited service scheme [1], [4] (inter-ONU scheduling) and priority queuing (intra-ONU scheduling). They found that queuing delay for some traffic classes increases when the network load decreases, a phenomenon they termed light-load penalty. The authors pointed out the origin of this penalty and they proposed two different methods to eliminate it,—namely the two-stage buffers and the CBR credit (refer to [4] for detailed analysis on the light-load penalty). The drawback of the two-stage buffers is that the elimination of light-load penalty results in increased delay for higher priority classes. On the other hand, CBR credit only eliminates the penalty partially and requires external knowledge of the arrival process. On the other hand, the authors of [10] proposed a dynamic bandwidth allocation algorithm for multimedia services over EPON. They proposed to use strict priority queueing and presented control message formats that handle classified bandwidths using MPCP. However, no simulation results were reported to show the performance of their proposed DBA combined with the use of strict priority. In [11], the authors proposed a new bandwidth guaranteed polling (BGP) scheme that allows the upstream bandwidth to be shared based on the service level agreement between each subscriber and the operator. The algorithm is able to provide guaranteed bandwidth for premium subscribers according the SLAs while providing best effort services to other subscribers. The model considers dividing the ONUs in the network into two sets; one set contains the ONUs with bandwidth guaranteed services while the second set contains the ONUs with best effort services. Typically, this will not be the case in future emerging PON access networks, where one single ONU must be capable of provisioning different services for different users requirement. Moreover, the proposed BGP is not consistent, neither to be standardized, with the MPCP arbitration mechanism proposed for EPON by the IEEE 802.3ah Task Force.

In this paper, we propose a light-load penalty-free bandwidth allocation algorithm that supports differentiated services in EPON-based access networks by employing a suitable priority queuing (intra-ONU scheduling). Our work differentiates itself from previous work by proposing to use a particular traffic priority queueing combined with a specific bandwidth allocation algorithm that is not confined to limited slot allocation. We propose that excessive bandwidth resulting from lightly loaded ONUs to be allocated to other highly loaded ONUs to achieve higher channel utilization. We also enhance the inter-ONU scheduling to provide efficient QoS-based DBA. Here, inter-ONU scheduling messages for allocating bandwidth to different ONUs are transmitted via MPCP messaging protocol defined within the IEEE Task Force.

III. Overview of the MPCP

MPCP arbitration mechanism is being developed by the IEEE 802.3ah Task Force [6] to support time slot allocation by the OLT. Although MPCP is not concerned with any particular bandwidth allocation, it is meant to facilitate the implementation of various allocation algorithms in EPON. MPCP is a two-way messaging protocol defined to arbitrate the simultaneous transmission of different ONUs and resides at the MAC control layer. The protocol relies on two Ethernet control messages (GATE and REPORT) in its regular operation and three other message frames (REGISTER_REQUEST, REGISTER, REGISTER_ACK) in the auto-discovery mode. Auto-discovery mode is used to detect a newly connected ONU and to learn the round-trip delay and MAC address of that ONU. In this particular work, we are only concerned about the regular (nondiscovery) operation of MPCP.

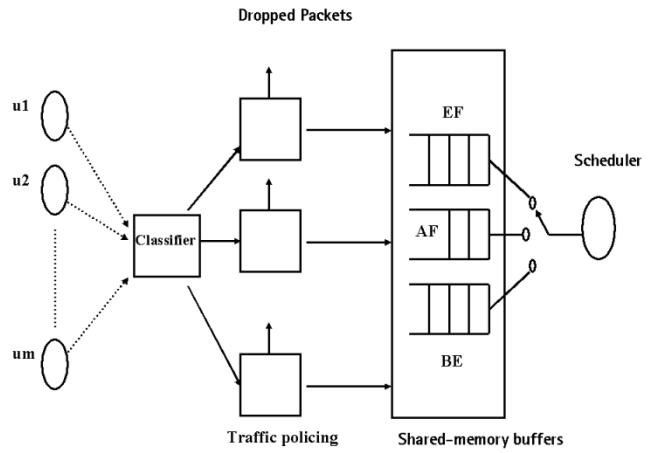


Fig. 1. Intra-ONU scheduling.

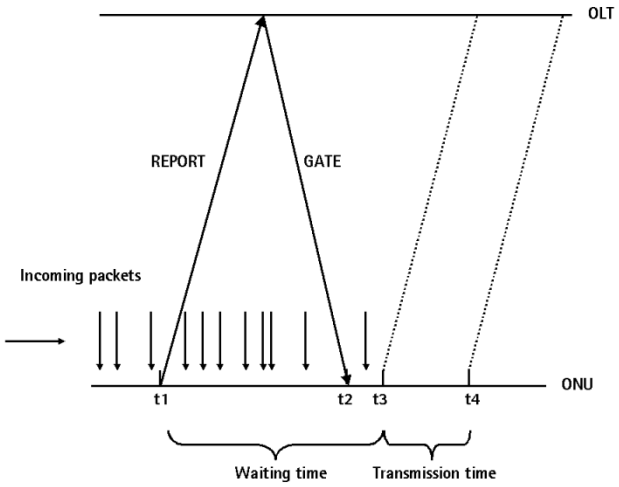


Fig. 2. Illustrative example.

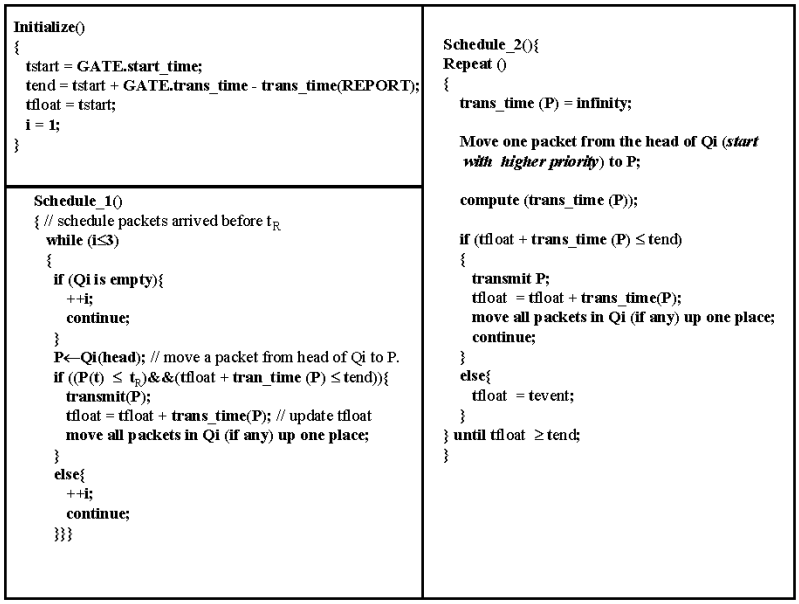
In its normal operation, MPCP gets a request from the higher MAC control client layer to transmit a GATE message to a particular ONU with the following information: time when the ONU should start transmission and the length of the transmission. Upon passing a message to the MAC layer, MPCP (in OLT and each ONU) timestamps the message with its local time. Upon receiving a GATE message matching its MAC address, each ONU will program its local registers with "*transmission start*" and "*transmission length*." Also, the ONU will update its local clock to that of the timestamp in the received control message, hence avoiding any potential clock drift and maintaining in SYNC with the OLT. When the transmission "*start timer*" expires, the ONU will start its contention-free transmission. The transmission may include multiple Ethernet frames, depending on the size of the allocated transmission window and the number of backlogged packets at the ONU. Note that, no packet fragmentation is allowed, i.e., if the next frame does not fit in the allocated time slot, it will be deferred to the next timeslot.

REPORT messages are sent by ONUs in the assigned transmission window together with data frames. A REPORT message can be either transmitted at the beginning of the timeslot, or at the end depending on the bandwidth request approach implemented by the ONU. It typically contains the desired size of the next timeslot based on the ONUs buffer occupancy. The ONU should also account for additional overhead when requesting the next time slot; this includes 64-bit frame preamble and 96-bit interframe gap (IFG) associated with each frame. Upon receiving a REPORT, the OLT passes the message to the DBA module responsible for bandwidth allocation decision and it will recalculate the round-trip time (RTT) to the source ONU. Note that when supporting differentiated services, each ONU has to report the status of its individual priority queues [10] and the OLT can choose to send one or multiple priority grants within the same GATE message depending on the bandwidth allocation algorithm implemented.

IV. ONU Queue Management and Priority Queuing

Bandwidth management and fair scheduling of different traffic classes [12] will play an important role in supporting QoS in the emerging EPON-based differentiated services (Diffserv)-capable access network. Priority queuing is considered a useful and relatively simple method for supporting differentiated service classes. Diffserv [5] is an IETF framework for classifying network traffic into classes, with different service level for each class. Fig. 1 shows the Queue management tasks carried out by each ONU. Each ONU maintains three separate priority queues that share the same buffering space. Packets are first segregated and classified (packet classification is done by checking the type-of-service (TOS) field of each IP packet encapsulated in the Ethernet frame) and then placed into their appropriate priority queues. The queuing discipline is as follows: if an arriving packet with higher priority finds the buffer full, then it can displace a lower priority packet. Alternatively, if a low-priority packet arrives and the buffer is full, then the packet is dropped. However, unless some kind of traffic policing is implemented at the ONU to regulate the flow of higher priority traffic and ensure that it conforms to its service level agreement (SLA), lower priority traffic may experience excessive delays and increased packet loss, resulting in a complete resource starvation. Thus, traffic policing [12] is required at the ONU to control the amount of traffic each user is allowed to send. After classifying the packets, they are checked for their conformance with the service level agreement and unnecessary traffic is dropped. The lower priority traffic is more likely to be dropped in favor of the higher priority traffic; however, control mechanisms are also necessary to control the flow of high-priority traffic if they exceed their agreed service contract.

Moreover, a priority-based scheduler is required for scheduling packet transmission. Strict priority scheduling mechanism (defined in P802.1D, clause 7.7.4) schedules packets from the head of a given queue only if all higher priority queues are empty. This situation will penalize traffic with lower priority at the expense of uncontrolled scheduling of higher priority traffic, resulting in increasing the level of unfairness (indefinite increase in packet delay, higher packet loss, uncontrolled access to the shared media, etc.). We illustrate the operation of such scheduler via a simple example shown in Fig. 2, where



```
Initialize()
{
  tstart = GATE.start_time;
  tend = tstart + GATE.trans_time - trans_time(REPORT);
  tfloat = tstart;
  i = 1;
}

Schedule_1()
{ // schedule packets arrived before t_R
  while (i≤3)
  {
    if (Qi is empty){
      ++i;
      continue;
    }
    P←Qi(head); // move a packet from head of Qi to P.
    if ((P(t) ≤ t_R)&&(tfloat + tran_time (P) ≤ tend)){
      transmit(P);
      tfloat = tfloat + trans_time(P); // update tfloat
      move all packets in Qi (if any) up one place;
    }
    else{
      ++i;
      continue;
    }}}

Schedule_2(){
Repeat ()
{
  trans_time (P) = infinity;

  Move one packet from the head of Qi (start
   with higher priority) to P;

  compute (trans_time (P));

  if (tfloat + trans_time (P) ≤ tend)
  {
    transmit P;
    tfloat = tfloat + trans_time(P);
    move all packets in Qi (if any) up one place;
    continue;
  }
  else{
    tfloat = tevent;
  }
} until tfloat ≥ tend;
}
```

Fig. 3. Pseudocode for priority scheduler.

only one ONU is requesting transmission. At time t_1, the ONU sends a REPORT to the OLT requesting bandwidth based on its current buffer occupancy. Upon receiving the message, the OLT allocates a timeslot to the ONU by sending a GATE message. Assume that this GATE message arrives to the ONU at time t_2 and the transmission is scheduled to start at a later time t_3. Now, the waiting time is $(t_3 - t_1)$, during which more packets may be arriving into the buffer and contending for transmission.

As mentioned previously, in strict priority scheduling the high-priority traffic arriving during this period (waiting period) will be scheduled ahead of the reported lower priority traffic. This will result in potentially deferring the transmission of lower-priority traffic for the next (or more) cycle(s), increasing indefinitely their queueing delay and prohibiting them from being transmitted in their allocated time window as specified by the bandwidth allocation algorithm. Hence, to alleviate this unfairness problem, we propose a priority-based scheduling algorithm. In priority scheduling, only those packets that arrive before t_1 are given high priority for transmission (given also that the bandwidth or size of the timeslot allows for the transmission). The order of the transmission is based on their priorities, i.e., round robin service discipline. If packets arriving before t_1 are all scheduled, and the current timeslot can still accommodate more traffic, it will be allocated for packets arriving during the waiting period (i.e., $t_3 - t_1$) based on their priorities. This scheme will ensure fairness in scheduling packets by allowing all traffic classes access to the channel as reported to the OLT, while adhering to their priority while being scheduled.

The pseudocode of the priority scheduler is shown in Fig. 3. It consists of two parts: initialization and schedule. The following are the parameters used: t_{start}, t_{end} represent the start time and end time of the transmission window allocated to the ONU; t_{float} represents a time indicator to show the progress in filling the time window (with data frames or empty packets), t_R is the time at which a REPORT message from the previous cycle was transmitted. trans_time(P) represents the time it takes to transmit a packet P over the PON, $trans\_time(p) = ((8^{*}size(p))/R)$, where R is the channel speed in bps and size(P) is the packet size in bytes. $P(t)$ represents the arrival time of packet P. Q_i $(i = 1, 2, 3)$ is a queue of priority i, where Q_1 represents the queue for high-priority traffic. Finally, t_{event} is the time at which any event occurs in the network.

Upon receiving a GATE message, the ONU will initialize (see Fig. 3) its transmission parameters by reading their associated values from the received message. Note that, we assume here that the ONU transmits its REPORT message for the next cycle at the end of the current allocated window. This is the reason for computing t_{end} the way it is shown in Fig. 3.

The function $Schedule\_1()$ will be called to schedule packets backlogged in the queue whose arrival time is less than t_R. If higher priority traffic arrive while reported lower priority traffic are being scheduled, they have to wait until the reported lower priority traffic are transmitted. If all reported traffic are being scheduled and the time window allocated to this ONU still can accommodate more traffic, $Schedule\_2()$ is invoked to schedule the transmission of frames arriving during the waiting time based solely on their priorities.

V. Dynamic Bandwidth Allocation With QoS Support

A critical issue in implementing efficient QoS-based EPON is the bandwidth allocation algorithm. The overall goal of bandwidth allocation is to effectively and efficiently perform fair scheduling of timeslots between ONUs in EPON networks. We mentioned the use of MPCP to arbitrate the ONUs' transmission; however, MPCP does not specify or require any particular allocation algorithm. Rather, MPCP provides a means of communication between the OLT and the different ONUs. Each

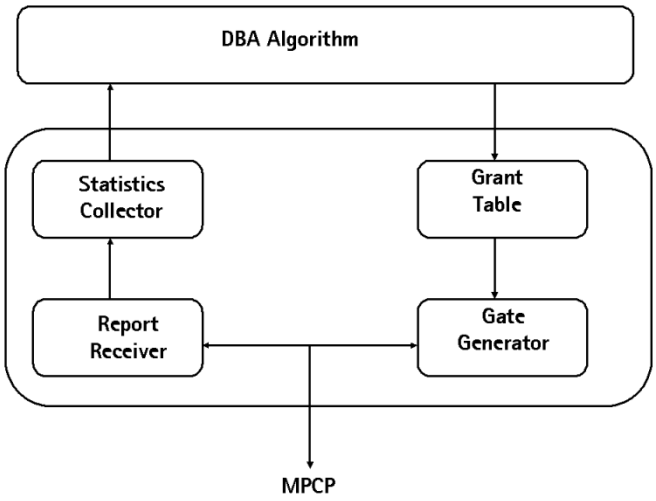


Fig. 4. Management block at the OLT.

ONU periodically reports its buffer occupancy status to the OLT and requests slot allocation. Upon receiving the message, the OLT passes this information to the DBA module (see Fig. 4). The DBA module in turn performs the bandwidth allocation computation and generates grant messages (note that grant messages are carried by MPCP GATE messages; each GATE might carry more than one grant message). Once the grant table is generated, the OLT transmits to the ONUs this information through MPCP GATE messages. The grant allocation table is updated by the output of the DBA algorithm (Fig. 4). Grant instructions are then compiled into MPCP GATE messages, and transmitted to the ONUs after performing RTT compensation.

We consider a PON access network with N ONUs. The transmission speed of the PON is R Mb/s (same for both upstream link and downstream link). We denote the granting cycle by T_{cycle}, which is the time during which all active ONUs can transmit and/or report to the OLT. Making T_{cycle} too large, under fixed slot allocation, will result in increased delay for all Ethernet frames, including those carrying high-priority traffic. The reason is that larger cycle time results in larger transmission window size; and, hence, at low loads the allocated slots will be underutilized: while one ONU is ineffectively holding the transmission channel, backlogged traffic at the next ONU will experience increased packet delays. Meanwhile, at high loads the situation is different; the transmission media exhibits higher utilization, which could result in lower average packet delays however maximum packet delays will be increased. On the other hand, making T_{cycle} too small will result in more bandwidth being wasted by guard intervals (note that timeslots allocated to ONUs are separated by guard times, T_g), will result in increased CPU processing load, and might potentially prevent large packets from being transmitted because no packet fragmentation is allowed. The guard intervals are necessary to provide protection for fluctuations in RTT of different ONUs. We also denote B_i^{MIN} as the minimum guaranteed bandwidth (in bytes) for ONU i, i.e., the minimum bandwidth OLT allocates under heavy load operation (i.e., peak times)

$$B_i^{\text{MIN}} = \frac{(T_{\text{cycle}} - N \times T_g) \times R}{8} w_i \tag{1.1}$$

where w_i is the weight assigned to each ONU based on its SLA, $\sum_{i=1}^{N} W_i = 1$.

Note that if all ONUs were not to be classified based on their SLA (i.e., $w_i = w = 1/N \ \forall i$, and $\sum_{i=1}^{N} w_i = 1$), then the minimum guaranteed bandwidth for each ONU will be

$$B_i^{\text{MIN}} = \frac{(T_{\text{cycle}} - N \times T_g) \times R}{8 \times N}. \tag{1.2}$$

Generally speaking, there are two categories of bandwidth allocation algorithms, fixed slot allocation (FSA) and DBA. In FSA, each ONU is allocated a minimal guaranteed bandwidth. If one ONU has less data to transmit, then other ONUs will have to wait until the granted transmission time for that particular ONU expires, thus resulting in inefficient channel utilization. Under dynamic allocation, however, the allocated timeslot will adapt to the requested bandwidth. Let R_i be the requested bandwidth for ONUi, and B_i^g be the granted bandwidth.

One way to allocate bandwidth to ONUi is as follows:

$$B_i^g = \begin{cases} R_i, & \text{if } R_i < B_i^{\text{MIN}} \\ B_i^{\text{MIN}}, & \text{if } R_i \geq B_i^{\text{MIN}} \end{cases}. \tag{2}$$

This approach is known as limited bandwidth allocation and has been studied in [1] and [4].

Due to the bursty nature of Ethernet traffic [8], [9], some ONUs might have less traffic to transmit while other ONUs require more than B^{MIN}. This results in a total excessive bandwidth ($B_{\text{total}}^{\text{excess}} = \sum_i^M (B_i^{\text{MIN}} - R_i)$, where $B_i^{\text{MIN}} > R_i$, and M is the set of light-loaded ONUs), which is not exploited under the previous approach (Limited Allocation). To improve the limited bandwidth allocation algorithm, one can exploit this excessive bandwidth by fairly distributing it amongst the highly loaded ONUs; for this reason, we develop the following method to allocate the excess bandwidth:

$$B_i^g = B_i^{\text{MIN}} + B_i^{\text{excess}} \tag{3}$$

$$B_i^{\text{excess}} = \frac{B_{\text{total}}^{\text{excess}} \times R_i}{\sum_{k \in K} R_k} \tag{4}$$

where B_i^{excess} is the excessive bandwidth allocated to ONUi and K is the set of heavily loaded ONUs. Thereafter, we refer to this algorithm as DBA1.

When providing services to different traffic classes with different QoS requirements, the requested bandwidth R_i consists of high-priority (H_i), medium-priority (M_i), and low-priority (L_i) bandwidth, and the ONU can request the OLT to assign, within the allocated timeslot, bandwidth for each class. This information is conveyed to the OLT, for bandwidth allocation, in the following message $\text{REPORT}(\text{H}_\text{i}, \text{M}_\text{i}, \text{L}_\text{i})$, where

$$R_i = H_i + M_i + L_i. \tag{5}$$

This information is made available through the use of MPCP REPORT message; note that MPCP specifies that each ONU can report up to eight queues (i.e., eight queue reports per ONU), where a report bitmap field [13], [14] specifies the queues (and their order) for which their REPORTs are transmitted.

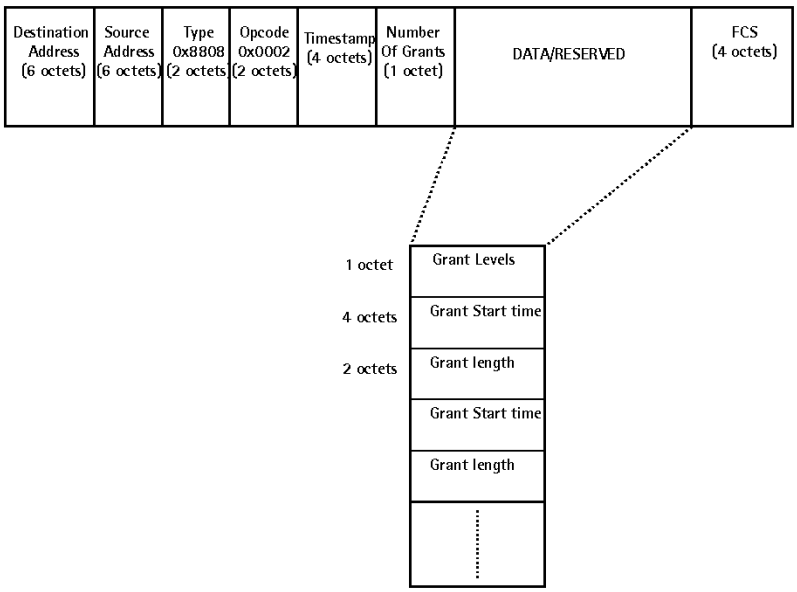


Fig. 5. MPCP GATE message format.

Note that under this scheme if there is no intra-ONU scheduling (e.g., the ONU prefers to shift the complexity of the queue management to the OLT or OLT might be capable of performing better per class bandwidth allocation), the OLT then can choose to generate multiple grants, each for a specific traffic class, to be transmitted to the ONU using a single GATE message: $B_i^g = H_i^g + M_i^g + L_i^g$, where H_i^g, M_i^g, L_i^g are the bandwidths granted to the three traffic classes, respectively. The 64-byte MPCP GATE message format is depicted in Fig. 5. Clearly, the OLT can specify to the ONU the number of grants carried by the GATE message. These grant instructions are carried by the GATE message in the designated DATA/RESERVED field (39 octets); each grant consists of a grant "start time" and a grant "length" (total of six octets), hence, a total of six grants (36 octets) can be carried by a GATE to a particular ONU. When the GATE message is received by the ONU, the latter should be able to classify grants to their particular queues; hence, the OLT will include an additional one byte of data "*grant level*" to identify the order of the queues to which grants are generated; e.g., in a system with eight queues per ONU, 10 110 000 indicates that three priority queues (Q0, Q2, and Q3) have been assigned grants and their grant information (i.e., start time and length) follows in the same order. Note that MPCP (as of now) does not specify any particular way to this per class allocation and its implementation is vendor specific.

As mentioned earlier and shown in Fig. 3, packets that arrive during the waiting time will have their transmission deferred to the cycle after the next one, posing additional delays; although some traffic might tolerate this, those that are delay-sensitive will not. To prevent the high-priority traffic from being penalized, we suggest that the ONU estimates (based on some statistical history from previous cycles) the bandwidth required by this type of traffic arriving during the waiting time t_w and we propose the following model:

$$R_i = \left(H_i + E_i^W(n)\right) + M_i + L_i \tag{6}$$

where $E_i^W(n)$ is the amount of high-priority traffic expected to arrive during the waiting time during cycle n and it can be estimated as follows:

$$E_i^W(n) = A_i^W(n-1) \tag{7}$$

where $A_i^W(n-1)$ is the actual amount of high-priority traffic arriving during the waiting period in cycle $n-1$. Note that because the traffic with high priority is not considered bursty, we can model its behavior by using a Poisson distribution, a simplified model to estimate the expected arrival rate during the waiting period.

The last issue the DBA is concerned with is the generation of the grant table. Upon receiving all REPORT messages from the active ONUs, the DBA module is invoked (see Fig. 4), to generate the table of grants. The DBA needs DBA_TIME to finish its computation and generate the grants table. As shown in Fig. 6, this mechanism results in some idle time where the PON channel is not utilized. This idle time is estimated as follows: $T_{\text{idle}} = \text{RTT} + \text{DBA\_TIME}$, where RTT is the round-trip time.

A straightforward method to account for this drawback is to use a gate-ahead mechanism; here the OLT will issue GATE messages for cycle "n" while receiving REPORT messages from cycle "$n-1$." This scheme might work well if the computation time "DBA_TIME" is considerably large, i.e., the OLT spends a large amount of time computing and updating the grant table; moreover, the transmission windows allocated to ONUs are based on the freshness of the received REPORT messages at the OLT, thus, this property could be lost if granting ONUs whose REPORTs were received at cycle "$n-1$" will happen at cycle "$n+1$" and potentially results in inaccurate window allocation and increased packet delay. However, if static slot allocation is to be used, then the gate-ahead will have its merits.

To address this deficiency, we propose a modified grant table generation algorithm (termed DBA2); here, the OLT needs to employ some early allocation mechanism in which an ONU

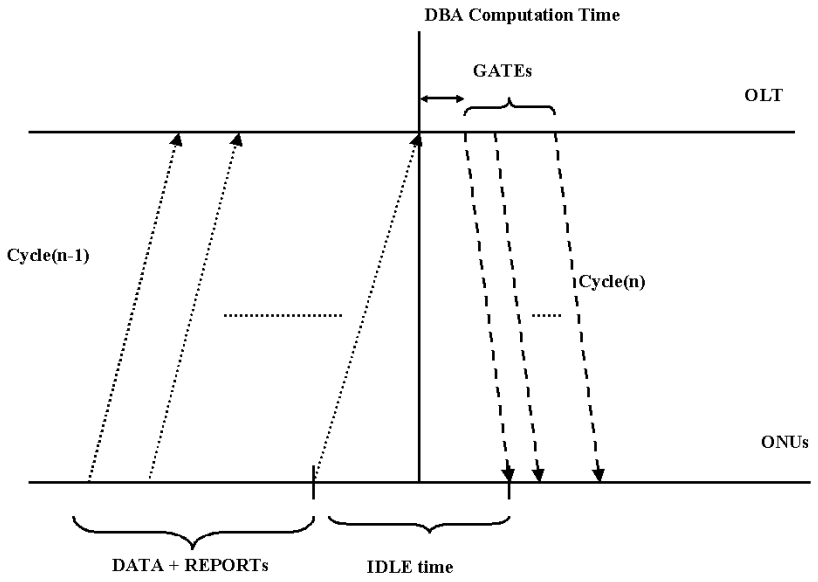


Fig. 6. Dynamic bandwidth allocation.

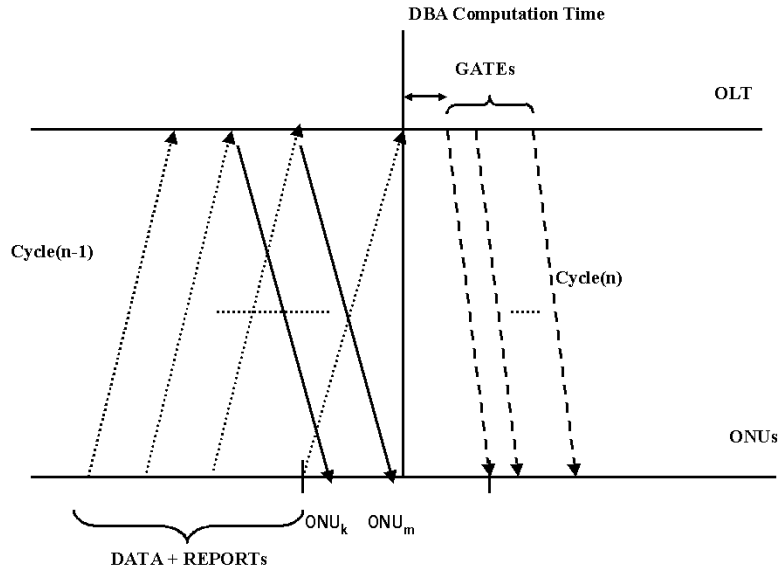


Fig. 7. Enhanced dynamic bandwidth allocation.

requesting bandwidth $R_i < B_i^{\mathrm{MIN}}$ can be scheduled instantaneously without waiting. Whereas, those who are requesting $R_i \geq B_i^{\mathrm{MIN}}$ will have to wait until all REPORT messages have been received and the DBA algorithm has computed their bandwidth allocation. Here, as shown in Fig. 7, this scheme will compensate for the idle time, and by allocating the lightly loaded ONUs early, we expect this modified algorithm to effectively increase the channel throughput and eliminate the waiting delay, which could penalize the delay sensitive traffic.

VI. Performance Evaluation

In this section, we compare the performance of the different bandwidth allocation algorithms presented in the previous sections and we study the impact of priority queueing on the overall performance of the network. For this reason, an event-driven packet-based simulation model is developed using C++. We consider a PON architecture with 16 ONUs connected in a tree topology. The distance between the OLT and the splitter is 20 km and between each ONU and the splitter is 5 km. The channel speed is considered to be 1 Gb/s and the maximum cycle time is set to 2 ms [1]. Each ONU supports three priority queues, sharing the same buffering space of size 10 Mb. The guard time separating two consecutive transmission windows is set to 1 μs and the IFG between Ethernet frames within the same slot is 96 bits.

For the traffic model considered here, an extensive study shows that most network traffic (i.e., http, ftp, variable bit

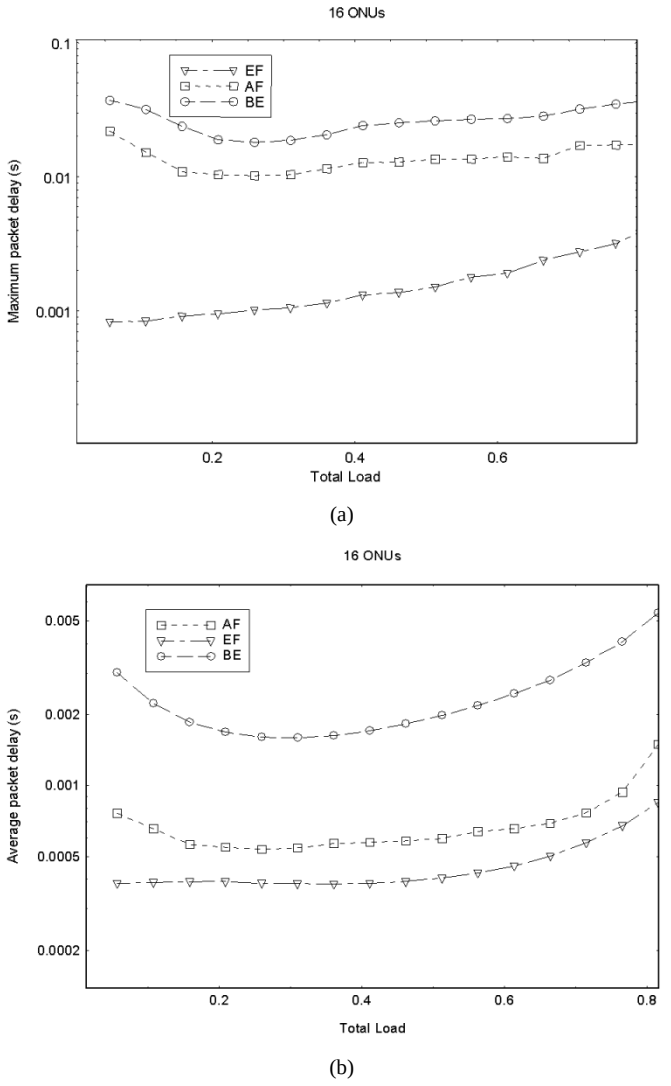


Fig. 8. (a) MD. (b) AD bandwidth allocation algorithm is DBA1 with strict priority queueing.

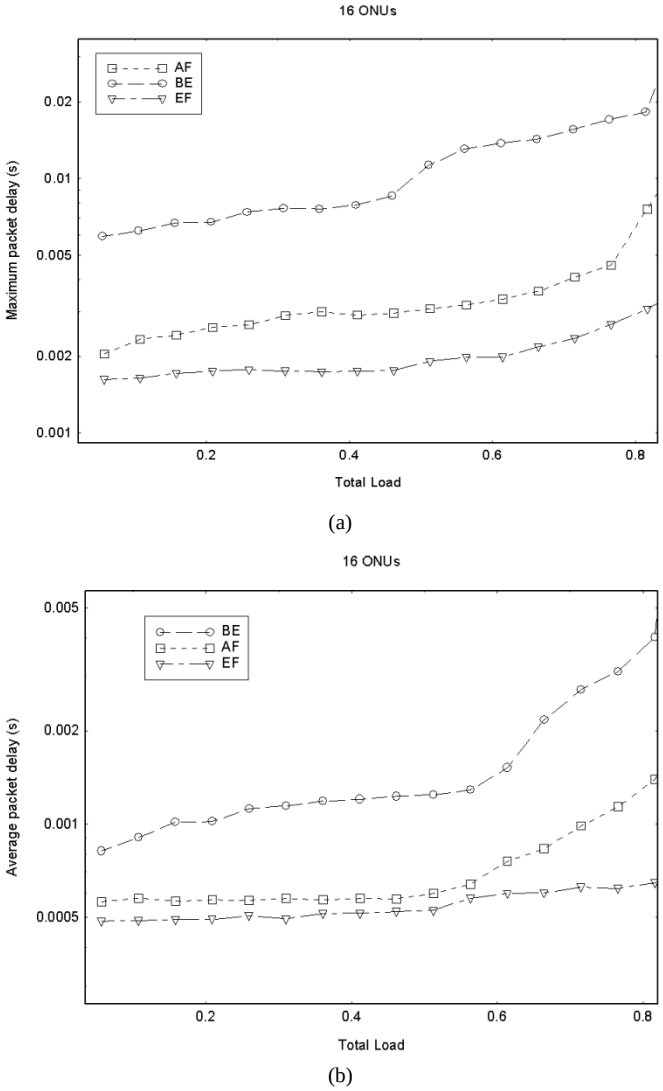
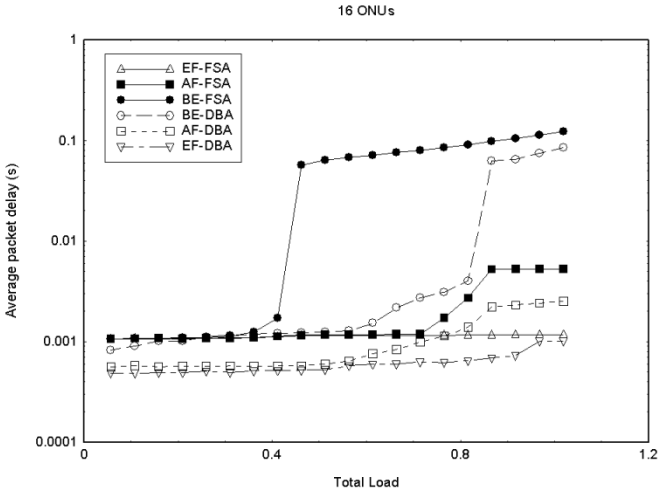


Fig. 9. (a) MD. (b) AD bandwidth allocation algorithm is DBA1 with priority queueing of Fig. 3.

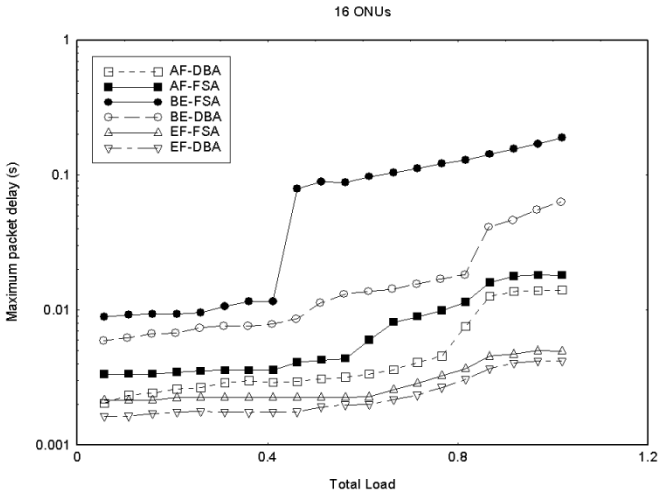
rate (VBR) video applications, etc.) can be characterized by self-similarity and long-range dependence (LRD) [8]. This model is used to generate highly bursty BE and AF traffic classes, and packet sizes are uniformly distributed between 64 and 1518 bytes. On the other hand, high-priority traffic (e.g., voice applications), is modeled using a Poisson distribution and packet size is fixed to 70 bytes [5]. The traffic profile is as follows: 20% of the total generated traffic is considered of high priority, and the remaining 80% equally distributed between low- and medium-priority traffic. Our simulator takes into account the queuing delay, transmission delay and the packet processing delay. The metrics of comparison are: average packet delay (AD), maximum packet delay (MD), and the throughput or channel utilization.

We first start by studying the impact of integrating DBA1 with strict priority scheduling at the ONU. In Fig. 8, we show the results of the AD and MD. Clearly, the limitation of this approach is the increased delays experienced by BE and AF traffic classes. Here, a higher priority packet always has the preference of being transmitted over other types of traffic and, hence, preventing other traffic classes from using their allocated bandwidth. This situation will result in increasing indefinitely the packet delay (average and maximum) of lower priority traffic. In addition, due to the fact that the transmission of each ONU is regulated by the OLT and the slot allocation is based on previously (from previous cycle) reported buffer occupancy, a waiting time (see Fig. 2) is experienced by each ONU until its transmission turn comes; thus, more traffic is likely to arrive during this time. The strict priority scheduler again will give preference for transmission to higher priority traffic arriving during the waiting time (unreported traffic). This situation will penalize other traffic classes by further increasing their average and maximum delays, and results in an interesting phenomenon: as the load decreases, the average (and maximum) packet delay increases [Fig. 8(a) and (b)]. There are two main reasons behind this behavior: unfair scheduling as mentioned before, and the fact that at very light loads, the OLT is more likely to assign smaller timeslots that are easily manipulated by the high-priority traffic.

To cope with these limitations, we investigate the benefits of combining DBA1 with the intra-ONU scheduler (priority scheduling) presented in Fig. 3. Here, only those reported packets by

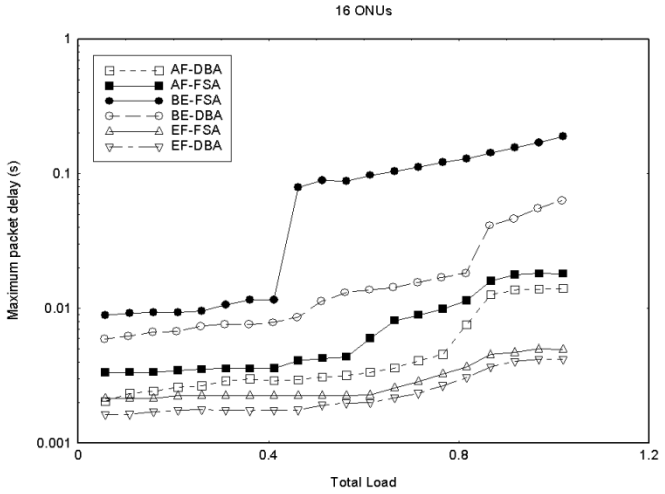


(a)

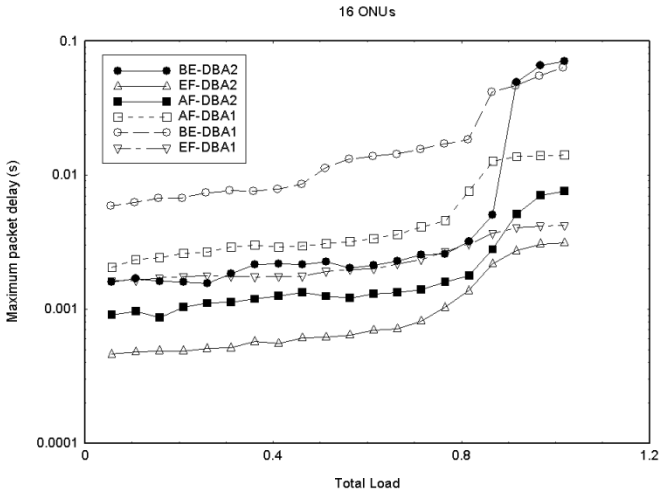


(b)

Fig. 10. Comparison between FSA and DBA1, with priority queueing of Fig. 3. (a) AD. (b) MD.



(a)



(b)

Fig. 11. Comparison between DBA1 and DBA2, with priority queueing and unequal share of excess bandwidth. (a) AD. (b) MD.

the ONU are scheduled for transmission in the current cycle. Packet scheduling is done in a round robin fashion, from high-priority to low priority. Fig. 9 shows that the light-load penalty, as experienced previously, is now eliminated because all packets from different classes are allowed to access their designated time slot as scheduled by the OLT. However, delays for high-priority traffic are now increased since more lower priority traffic is given the chance for transmission forcing higher priority packets arriving during the waiting period to wait for the next cycle transmission.

Another approach that can also be implemented to achieve global fairness (fair share of the transmission window amongst classes on a single ONU, as well as with traffic classes of other ONUs) between different traffic classes is as follows: ONU reporting its buffer occupancy to the OLT will demand from the OLT individual grants within the same GATE message (as described in Section V). Thus, the DBA will have to allocate per-class bandwidth to each class of service, refer to Section V for more on this analysis. Here, the ONU will leverage the functionality of the scheduler by pushing the responsibility and complexity further to the DBA. We have also simulated the behavior of this scheduling and similar results as in the previous algorithm were found.

Fig. 10(a) and (b) shows that under fixed slot allocation, packet delays for BE and AF traffic classes are substantially higher than delays incurred by the same traffic classes under DBA1. The figure shows the at load of 0.4, average and maximum delays for BE packets under FSA increase to almost 100 ms, whereas the delay increase under DBA1 picks up at a total network load of 0.8, while being still lower than the delay under FSA. The reason is that the DBA1 algorithm allows statistical multiplexing between the different ONUs competing for bandwidth allocation, a property that could not be exploited under FSA.

In Fig. 10, we compare the performance of EPON under both FSA and DBA1, both with priority scheduling discussed in Section IV.

Now, although the DBA1 presented here achieves better efficiency than the fixed slot allocation, it still has its limitations, because the OLT has to wait until all ONUs have transmitted their REPORT messages before it can do bandwidth allocation, as specified in Section V. Thus, there is an idle time where

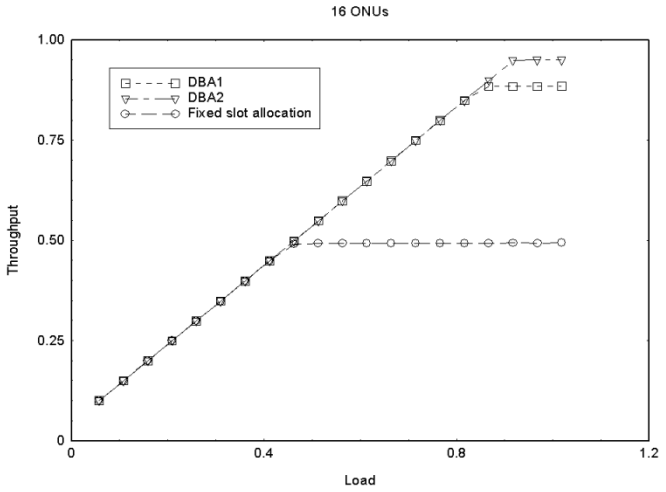


Fig. 12. EPON throughput.

the transmission channel is not utilized (see Fig. 6). This idle time is equal to a round trip propagation delay, plus the DBA computation time. DBA2 was proposed in Section V to improve upon this by allowing the OLT to schedule "on-the-fly" ONUs that are requesting bandwidth less than the minimum bandwidth guaranteed by EPON. This way the ONUs that are requesting more than B^{MIN} are deferred until all the REPORTs have been received. The OLT will also keep track of the excessive bandwidth from the set of lightly loaded ONUs and will distribute this excess bandwidth to other heavily loaded ONUs based on their requested bandwidth, i.e., two ONUs requesting bandwidths B1 and B2 more than B^{MIN} will be assigned excess bandwidths proportional to B1 and B2. Note also that one could also distribute this excess bandwidth fairly amongst those heavily loaded ONUs, i.e., all of those ONUs get the same share of excess bandwidth. We consider the former case in this study. Fig. 11 shows the simulation results for these two algorithms.

Fig. 11 shows that DBA2 outperforms DBA1 in terms of average packet delay and maximum packet delay. This is due to the early allocation property of DBA2, whereby a round trip delay is eliminated and the lightly loaded ONUs are scheduled in the same cycle in which they report to the OLT. Note also that the performance of all traffic classes under DBA2 is improved.

Finally, we compare the improvement in throughput when DBA2 is used. Fig. 12 shows the throughput of FSA, DBA1, and DBA2. As expected, FSA has the lowest throughput (less than 50%) due to the lack of statistical multiplexing between ONUs, whereas, DBA1 and DBA2 exploit this property to improve the upstream channel utilization. Finally, DBA2 achieves a throughput of 95% (compared with DBA1 achieving 88%), which is attributed to the early allocation property of the algorithm.

VII. Conclusion

In this paper, we addressed the problem of dynamic bandwidth allocation in Ethernet-based PONs. We augmented the bandwidth allocation algorithms to support QoS in a differentiated services framework. It was shown that strict priority-based bandwidth allocation, under our assumptions for traffic behavior, will result in an unexpected behavior for certain traffic classes (light-load penalty, as reported in [4]) and we suggested the use of appropriate queue management with priority scheduling to alleviate this problem. Moreover, we showed that DBA algorithms that perform early bandwidth allocation for lightly loaded ONUs result in better performance in terms of average and maximum packet delay, as well as network throughput compared with some other dynamic allocation algorithms. We used simulation experiments to validate the effectiveness of the proposed algorithms.

References

[1] G. Kramer, B. Mukherjee, and A. Maislos, "Ethernet Passive Optical Network (EPON): a missing link in an end-to-end optical internet," in *Multi-Protocol Over WDM: Building the Next Generation Internet*, S. Dixit, Ed. New York: Wiley, Mar. 2003.

[2] G. Kramer and B. Mukherjee, "Ethernet PON: design and analysis of an optical access network," *Photonic Network Commun.*, vol. 3, no. 3, pp. 307–319, July 2001.

[3] ——, "Interleaved polling with adaptive cycle time (IPACT): a dynamic bandwidth distribution scheme in an optical access network," *Photonic Network Commun.*, vol. 4, no. 1, pp. 89–107, 2002.

[4] G. Kramer, B. Mukherjee, S. Dixit, Y. Ye, and R. Hirth, "On supporting differentiated classes of service in EPON-based access network," *J. Opt. Networks*, pp. 280–298, 2002.

[5] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, An Architecture for Differentiated Services, IETF, RFC 2475, Dec. 1998.

[6] IEEE 802.3ah task force home page [Online]. Available: http://www.ieee802.org/3/efm

[7] K. Rege *et al.*, "QoS management in trunk-and-branch switched ethernet networks," *IEEE Commun. Mag.*, vol. 40, pp. 30–36, Dec. 2002.

[8] W. Willinger, M. S. Taqqu, and A. Erramilli, "A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks," in *Stochastic Networks*. Oxford, U.K.: Oxford Univ. Press, 1996, pp. 339–366.

[9] M. S. Taqqu, W. Willinger, and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling," *ACM/SIGCOMM Comput. Commun. Rev.*, vol. 27, pp. 5–23, 1997.

[10] S. Choi and J. Huh, "Dynamic bandwidth allocation algorithm for multimedia services over Ethernet PONs," *ETRI J.*, vol. 24, no. 6, pp. 465–468, Dec. 2002.

[11] M. Ma, Y. Zhu, and T. H. Cheng, "A bandwidth guaranteed polling MAC protocol for Ethernet passive optical networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, Mar.–Apr. 2003, pp. 22–31.

[12] H. Shimonishi, I. Maki, T. Murase, and M. Murata, "Dynamic fair bandwidth allocation for diffserv classes," in *Proc. IEEE ICC*, vol. 4, Apr.–May 2002, pp. 2348–2352.

[13] *Virtual Bridged Local Area Networks*, IEEE Standard 802.1Q, 1998.

[14] Media Access Control Parameters, Physical Layers and Management Parameters for Subscriber Access Networks, IEEE Draft P802.3ah/D1.0TM, Aug. 2002.

Chadi M. Assi (S'02) received the B.S. degree in electrical and computer engineering from the Lebanese University, Beirut, Lebanon, in 1997, and the M.S. and Ph.D. degree in electrical engineering from the Graduate Center, City University of New York (CUNY), in 2000 and 2003, respectively, where he held the Mina Rees Dissertation Fellowship.

He spent one year at Nokia Research Center, Boston, MA, as a Visiting Scientist working in the area of broadband optical access networks from 2002 to 2003, and subsequently, he joined the Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC, Canada, in August 2003, as an Assistant Professor. His current research interests are in the general area of optical networking and more specifically provisioning and restoration, and network control and management.

Yinghua Ye (S'98–M'00) received the Ph.D. degree in electrical engineering in 2000 from City University of New York (CUNY).

She joined Nokia Research Center, Burlington, MA, in June 2000. Currently, she is a Senior Research Engineer in the Mobile Internet Performance Group, Nokia Research Center. She has published more than 30 papers in conferences and journals, and currently has two U.S. patents pending and one provisional application in the field of optical networking. Her research interests include service discovery, architecture design, network survivability, traffic engineering, and real time provisioning in optical networks.

Dr. Ye has served as Technical Committee Member for Opticom 2002, OptimCom 2003. She was actively involved with IEEE 802.3 Standardization activities in 2002 and made some contributions to MPCP.

Sudhir Dixit (S'75–A'80–M'80–SM'94) received the B.E. degree from Maulana Azad College of Technology (MACT), Bhopal, India, the M.E. degree from Birla Institute of Technology and Science (BITS), Pilani, India, and the Ph.D. degree from the University of Strathclyde, Glasgow, Scotland, all in electrical engineering. He also received the M.B.A. degree from Florida Institute of Technology, Melbourne, FL.

He is a Senior Research Manager at Nokia Research Center, Burlington, MA. His main areas of interest are pervasive communications, content delivery networks, and optical networking. From 1991 to 1996, he was a broadband network architect at NYNEX Science and Technology, (now Verizon Communications). Prior to that he held various engineering and management positions at other major companies, e.g., GTE, Motorola, Wang, Harris, and STL (now Nortel Europe Labortories). He has published or presented over 150 papers and has 28 patents either granted or pending. He has coedited *Wireless IP and Building the Wireless Internet* (Norwood, MA: Artech House, 2002), and edited *IP over WDM* (New York: Wiley, 2003). Presently, he is coediting *Content Delivery in the Mobile Internet* (New York: Wiley, 2004).

Dr. Dixit has been a Technical Co-Chair and a General Chair of the IEEE International Conference on Computer, Communications, and Networks, in 1999 and 2000, respectively, a Technical Co-Chair of the SPIE Conference Terabit Optical Networking in 1999, 2000, and 2001, respectively, a General Chair of the Broadband Networking in the Next Millennium Conference in 2001, a General Co-Chair of the OptiComm 2002 Conference, and a General Chair of International Conference on Communication and Broadband Networking 2003. He has also been an ATM Forum Ambassador since 1996. He has served as a Guest Editor of *IEEE Network*, *IEEE Communications Magazine*, and *Optical Networks Magazine* published by SPIE/Kluwer. Currently, he is on the Editorial Boards of the *Wireless Personal Communications Journal*, *Journal of Communications and Networks*, *IEEE Optical Communications*, and the *International Journal on Wireless and Optical Communications*.

Mohamed A. Ali received the M.S. and Ph.D. degrees in electrical engineering from the City College of the City University of New York (CUNY), in 1985 and 1989, respectively.

He joined the faculty of the Department of Electrical Engineering, City College of New York, in 1989, where he is currently a Professor. His research interest is in the general area of broadband information networking. His current research activities involve fiber optic communication systems, networking and architecture; IP/ATM/SONET-based DWDM and TDM multiple-access broadband networks; high performance IP/MPLS routers, next-generation networking paradigm and NGI, traffic engineering/provisioning and restoring in a hybrid IP/MPLS-centric DWDM-based optical networks; optical amplifiers and components; CATV distribution over fiber-based local access ATM networks. His major interests are in the areas of computer simulation and modeling of high-speed telecommunication systems. He has consulted for several major carriers on issues related to the pros and cons of a router-centric architecture deployed on a thin optical layer versus a hybrid router/OXC-centric architecture deployed on a rich and intelligent optical transport layer. He has published over 70 papers in professional journals and international conferences.

Dr. Ali received the NSF Faculty Career Development Award.