Efficient resource allocation with service guarantees in passive optical networks

T. Orphanoudakis,[1,*] H.-C. Leligou,[2] E. Kosmatos,[2] J. D. Angelopoulos,[2]
K. Kanonakis,[2] G. Prezerakos,[3] and I. Venieris[2]

[1] *University of Peloponnese, Karaiskaki Street, 22100, Tripoli, Greece*
[2] *National Technical University of Athens, Heroon Poloytechneiou 9, Zographou Campus, Athens, Greece*
[3] *Technological Education Institute (TEI) of Piraeus, P. Rali & Thivon 250, Aigaleo, Greece*
Corresponding author: fanis@uop.gr

Received December 18, 2006; revised March 9, 2007; accepted May 1, 2007;
published xx xx, xxxx (Doc. ID 78177)

We propose and evaluate a resource allocation scheme for time-division multiplexing passive optical networks (PONs), which supports multiple service classes; dynamic bandwidth allocation for services with varying (in time) capacity demand; and bounded quality of service parameters for services with real-time requirements. Although several algorithms have been proposed in the literature considering several of the above objectives in isolation, our work focuses on the fundamental problem of trading-off between PON upstream channel utilization and strict delay and jitter bounds when supporting a dynamically changing mix of services with different requirements. © 2007 Optical Society of America

OCIS codes: 060.0060, 060.2330, 060.4250.

1. Introduction

Passive optical networks (PONs) have emerged as an alternative access technology that enables the delivery of broadband services to residential users combining high bandwidth, increased flexibility, broad area coverage, and economically viable sharing of the expensive optical links. Due to their above inherent features, PONs have generated during the past decade substantial commercial activity also reflected in the work of several standardization bodies. Since the initial deployment of ATM-based PONs (APONs) newer standards support multigigabit rates and adapt better to the packet-based Internet applications. The full-service access networks (FSAN) group has produced its second generation standard for the so-called gigabit PON (GPON) supporting mixed time-division multiplexing (TDM), ATM, and packet-based services reaching symmetrical transmission rates of up to 1.244 or 2.488 Gbits/s, which was adopted by ITU-T and was included in the G.984.x series of ITU-T recommendations [1]. At the same time IEEE, through the activities of Ethernet in the First Mile (EFM) group, has standardized a gigabit Ethernet-friendly technology [2] called Ethernet PON (EPON), with the objective to leverage the great success of Ethernet as a LAN technology and exploit the economies of scale that the dominance of Ethernet has generated.

The fact that PONs can offer high capacity should not result in the misleading assumption that a bandwidth surplus can alleviate performance degradation due to delay and jitter, by employing simplistic access control schemes. To achieve both economical deployment and, most important, profitable operation of an EPON, the bandwidth allocation mechanism should be designed so as to optimally trade-off resource (i.e., bandwidth) consumption with performance guarantees in order to efficiently support applications with different requirements. The efficient support of different quality of service (QoS) levels is mandatory for the penetration of this technology, since it is tightly associated with the support of triple-play services (real-time multimedia content delivery, telephony, and data). Both delay-sensitive and best-effort applications should be simultaneously supported in the emerging PONs. In these tree-shaped systems, the performance in terms of delay, delay variation, and throughput depends on the upstream bandwidth allocation performed by the medium access controller

1536-5379/07/060001-0/$15.00

(MAC) residing at the optical line termination (OLT). While the IEEE 802.3ah describes the upstream and downstream transmission formats, it only defines the required operational procedures that can guarantee robust operation and interoperability between systems and components provided by independent vendors. The 802.3ah standard defines the so-called Multipoint Control Protocol (MPCP) and the type of messages that should be exchanged during operation; it does not specify algorithms that can be employed especially for bandwidth allocation, since this is considered an issue open to the specific vendors and network providers and should be dealt with according to their specific requirements.

In this paper we focus on a novel 802.3ah compliant dynamic bandwidth allocation (DBA) scheme realizing four different service policies to efficiently support any QoS requirement and dynamic service provisioning over EPONs. Emphasis has been placed on offering strict delay and jitter bounds maximizing upstream channel utilization efficiency. Finally, the impact of centralized versus distributed [intra-optical-network unit (ONU)] scheduling on both delay performance and efficiency is also investigated.

2. EPON Operational Parameters

The downstream direction in the tree-shaped topology of PONs operates in a broadcast fashion emulating point-to-point communication, while in the upstream channel an aggregate data flow is generated by means of burst transmissions from the active ONUs in a time-division multiplexing access (TDMA) fashion. The activation of each ONU's transmitter and window of operation is controlled by the OLT. To make dynamic arbitration of the upstream burst transmissions from multiple ONUs feasible, MPCP is deployed. MPCP uses two types of messages during normal operation for arbitration of packet transmissions: the REPORT message used by an ONU to report the status of its queues to the OLT (up to eight reported in a single message) and the GATE messages issued by the OLT and indicating to the ONUs when and for how long they are allowed to transmit in the upstream channel. Each GATE message can support up to four transmission grants targeting individual service entities within the same ONU (i.e., data queues). In the upstream, the granted ONU transmits (possibly) multiple Ethernet frames—as many integral packets can fit into the allocated transmission slot, since fragmentation is not allowed—from one or more queues preceded by the indispensable physical layer overhead. It also transmits REPORT messages in order to request additional grants. In EPONs, the traffic streams arriving at the ONUs from the customer premises are kept in queues. In compliance to the 802.1p prioritization scheme, it is possible to inject the traffic in up to eight logically separate, possibly prioritized, queues holding Ethernet frames, depending on QoS requirements, to allow for the enforcement of different service mechanisms. In this work, we consider four priority queues at the ONU side and show that this is an adequate requirement for the EPON multiplexing function to provide differentiated levels of service.

Both EPON and GPON have been designed to fit all fiber to the x (FTTx) solutions with expectations varying from country to country. For example, in Japan, the fiber to the home (FTTH) solution is already widely deployed while in most European countries the most prevalent case is the fiber to the business (FTTB) or to the curb. From the traffic engineering point of view, the difference among the various cases is the number of input streams sharing the upstream bandwidth and the QoS requirements of the users in combination to the tariff they are willing to pay. In the FTTH case, hundreds of customers share the upstream bandwidth and their requirements are those of triple-play services, i.e., the most demanding applications seem to be VoIP and video or interactive gaming with a one-way transmission time requirement of 1.5 ms while many studies have verified that a maximum of 2 ms transmission cycle time is an acceptable value [3,4]. In the FTTB case, however, the requirements can be stricter and the PON system should be capable of offering performance close to leased-line services. To reduce complexity in this cost-sensitive residential access system, services are grouped into behavior aggregates (classes) with a similar set of requirements providing scalability and flexibility (also following the 802.1P Q approach).

3. PON Dynamic Resource Allocation Framework

Several recent papers investigate both architectural issues and MAC protocols (a review of the most well known can be found in [5]). Initial attempts to efficiently implement DBA more or less depended on best effort polling of requests [6]. Most research attempts focused on the problem of fair or weighted sharing of bandwidth among users (e.g., [7]) or differentiated services through the discrimination of service classes during DBA (e.g., [8]). Only recently have there been proposals to strictly isolate real-time traffic from elastic, delay-tolerant traffic by means of specific bandwidth reservations in the EPON scheduling cycle (e.g., in [9–11]). The above-mentioned approaches have correctly identified the need to preallocate bandwidth for real-time traffic as the only means to provide acceptable access delay and combat the barrier that the large round-trip delays of EPONs raise in dynamically requesting bandwidth during load fluctuations.

The motivation for our work has been to investigate the operational parameters that affect the efficiency of PONs as multiservice broadband access systems and engineer solutions considering the following performance metrics as equally important: (i) average delay per class of service; (ii) delay variation for real-time services; (iii) bandwidth utilization of the shared upstream channel; and, last but not least, (iv) implementation and system operation and configuration cost. To this respect our contribution is focusing on the concept of four allocation strategies realized through appropriate queuing, OLT/ONU scheduling, and utilizing appropriate MPCP messaging. Our proposal adopts the approach investigated in [11] extending the protocol described therein to collectively handle four allocation strategies with enhanced bandwidth efficiency, as will be explained below. Additionally we propose what we believe to be the use of a novel DBA algorithm with optimized scheduling of granted upstream transmission windows initially presented in [12] in the context of operation for two classes of service (CoS).

4. Bandwidth Allocation Strategies

PONs can be considered as a mature access networking technology and have been attracting interest for far more applications than simple interactive Internet access, including their use as triple-play service delivery platforms up to TDM application concentrators and trunk networks, e.g., for mobile networks as an alternative to costly SDH infrastructure. Therefore, in order to support emerging services, QoS differentiation based only on delay criteria and support of just two discrete CoS, i.e., real-time versus delay-tolerant traffic actually is not adequate. Thus, from a traffic engineering point of view a number of services need to be mapped into distinct CoS, which should be appropriately supported by the EPON traffic multiplexing mechanisms. To reduce complexity in this cost-sensitive residential access system, services need to be grouped into behavior aggregates (classes) with a similar set of requirements providing scalability and flexibility. Given the fluidity of service class definition, we implement four priority classes in the MAC, leaving finer aggregations and more elaborate forwarding policies, to be implemented at the egress of the EPON system, as proposed in [13]. All that is required from the MAC is not to deny quality to any group of flows and achieve guaranteed QoS according to service level agreements during service configuration. The basis of our approach is the use of access priorities in the reservation system, which can be programmed to fit with required traffic descriptors by means of S/W programming and mapping of flows to EPON logical queues residing at the ONU side.

The behavior aggregates are mapped into distinct CoS, which should be appropriately supported by the EPON traffic multiplexing mechanisms (by means of appropriate queuing and scheduling algorithms). In this paper, we design and evaluate an algorithm for an EPON system targeting efficient support of all types of services (from delay sensitive to best-effort) without sacrificing utilization by means of four aggregation levels or priorities-which present the following features.

The high-priority class (CoS1) is devoted to delay-sensitive periodic constant bit rate (CBR) traffic. This class targets services with very strict delay requirements, which undergo strict traffic profile control (traffic conditioning). It aims to be the network provider's tool to offer virtual leased line service allocating the contracted peak rate R_{p1}.

The second class (CoS2) is devoted to real-time variable rate flows, such as video services or VoIP and it is provided with a guaranteed (sustained) rate (R_{s2}) and statistical upper delay bounds. Enough bandwidth to service this type of traffic (up to a contracted peak rate R_{p2}) is reserved but issued only upon request realizing DBA.

The third class (CoS3) is devoted to data services with higher requirements than best-effort. The traffic profile control assumed for this class aims at minimizing the loss of packets and the disturbance to other traffic. Its traffic parameters include a predefined minimum service rate (denoted as R_{g3}), which must be reserved during service configuration, while any request above that is treated as plain best-effort. The fourth priority (CoS4) is reserved for plain best-effort services as well as traffic that employs loss-based flow control at the *TCP* level and can be very disruptive to the other classes when sharing the same queue.

In the Internet Engineering Task Force (IETF) diffserv context, these four classes can be mapped to the expedited forwarding service, the top assured forwarding (AF) class, all four or the lower three AF classes, and best-effort class, respectively. It should be stressed that any action taken during the bandwidth allocation procedure in order to support delay-sensitive services does not affect the overall system cost since it only impacts the upstream bandwidth allocation algorithm implemented at the OLT.

5. Efficient Medium Access Arbitration

5.A. Achieving Statistical Performance Guarantees

The proposed algorithm aims at providing guaranteed statistical delay and jitter bounds to real-time traffic, while dynamically distributing unused bandwidth to bursty traffic with no strict QoS requirements. Each of the four behavior aggregates defined above is serviced based on a different mechanism implemented at the OLT. Starting with the high priority, as correctly identified in [10,11], the only means to combat the large round-trip delays introduced by reservation schemes is to preallocate bandwidth for real-time traffic. These approaches strongly resemble the so-called unsolicited grant service (UGS) of the DOCSIS 1.1 [14] protocol. The UGS mechanism is similarly used in cable [hybrid fiber coaxial (HFC)] networks, where the MAC controller at the HFC headend (CMTS) allocates a fixed number of minislots periodically to allow for a constant-bit-rate flow of information. Although our DBA mechanism also follows the approach of serving real-time traffic at its peak rate through periodic unsolicited grants, it additionally offers strict delay bounds and low-delay variation, due to an enhanced scheduling frame structure as will be explained later on.

For the high-quality class, (obviously associated with a higher tariff), the operator actually guarantees service to a contracted peak rate (R_{P1} expressed in bytes per second) and a strict delay bound $D_{\max}$. To achieve this, the relevant queues are granted upstream windows of fixed duration periodically, (once every D_m), which drives us to consider D_m as the scheduling period. Evidently this operational parameter is closely related to the desirable delay bound $D_{\max}$ as will be also shown by the performance evaluation results. The duration of the window allocated to each ONU is calculated as a function of the negotiated (peak) service rate and the grant period D_m while the exact value of D_m is selected as a near optimal trade-off between two basic factors: an acceptable delay bound for real-time traffic and reduction of scheduling and transmission overheads that stem from burst mode transmission of bursty traffic.

For the second class, which is characterized by a sustainable (R_{S2} expressed in bytes per second) and a peak information rate (R_{P2} expressed in bytes per second), a number of bytes is granted unsolicitedly, exactly as happens for high class, while surplus bandwidth up to the contracted peak rate (R_{P2}) is granted upon request. Both the unsolicited and the surplus amount of bytes are calculated as a function of the contracted rates for the scheduling period D_m. It is worth stressing that while for the high priority the unsolicited grants cover the peak rate, it can be chosen by the provider to cover either the sustainable rate or a lower rate than that for the second, trading-off delay performance for efficiency as will be shown in the simulation results section. So, considering the case where unsolicited grants cover the sustainable rate of the second class, the total number of unsolicited grants for the ith ONU (UG_i) in bytes is expressed as follows:

$$UG_i = (R_{P1i} + R_{S2i})D_m,$$

while the following relation should hold

$$NT_{pre} + \sum_{1}^{N} UG_i < 1\ \text{Gbps}\ \frac{D_m}{8},$$

where T_{pre} is the time consumed for the transmission of the preamble, the delimiter plus the required guard time, which is for convenience expressed in the equivalent number of bytes; while, N is the number of active ONUs. ONUs that have not agreed to pay for any of the two high-priority services, are only granted a report message (every D_m) that will allow the access controller at the OLT to become aware of the status of its queues; when a nonempty queue is reported (polling of requests), the relevant queue will be granted an upstream transmission window in the next scheduling period.

5.B. Efficient Dynamic Bandwidth Allocation

The concept behind the proposed algorithm is to serve all queues from the same ONU in contiguous transmission windows, scheduling a single grant (GATE message) per ONU, as a mean to economize on bandwidth wasted on physical layer overheads. The reason to opt for this scheduling is graphically explained in Fig. 1 where upstream burst allocations and actual data transmissions (covering a time window D_m) from different ONUs and CoS queues are shown. In Fig. 1(a), a possible scheduling of grants accommodating the requests collected in earlier polling cycles is shown. The order of the allocations (which target individual queues) affects the achieved efficiency. In Fig. 1(a), higher-priority allocations precede lower-priority allocations. Taking into account that the duration of the CoS 1 allocations is fixed, the position of the two higher-priority allocations are fixed in every scheduling cycle D_m, forming subframes per CoS (a scheduling discipline also selected in [10,11]). In EPONs, part of the allocations may be wasted since the exact ONU queue occupancy and packet delineation is not known by the grant scheduler at the OLT, so it is very likely that at the end of the allocated window the leftover time does not match the length of the next packet in the first-in–first-out (FIFO) queue, a phenomenon called unused slot remainder (USR), also shown in Fig. 1 only for the second ONU as an example. In Fig. 1(b), the alternative to serve all CoS queues from the same ONU before allocating slots to other ONUs is shown. This way subframes per ONU are formulated. Obviously, this schedule introduces fewer physical layer overheads. Further efficiency improvement is possible allowing the ONU to decide on the distribution of the allocated subframe duration (single burst in this case) among its CoS queues as shown in Fig. 1(c) (also called intra-ONU scheduling in [8]), since this increases the probability that enough time will remain to reduce USR by accommodating at least one more packet from any priority queue also depending on the packet size distribution. As will be shown Section 6, this approach can lead to improvement of delay for high-priority queues. Option (a) achieves the objective of controlling delay variation, trading-off upstream bandwidth efficiency. However, we will present below an efficient scheduling algorithm that combines the best features of both approaches (a) and (b).

AQ: #7

The efficiency improvement depends on the upstream time D_m as well as on the number of supported queues and ONUs. Most EPON bandwidth allocation algorithms assume a scheduling period that corresponds to the period of allocation decisions for an upstream transmission window of equal duration and propose to service each queue once in every scheduling cycle [11]. The longer the scheduling period, the higher is the efficiency achieved (minimizing physical layer overheads). However, assuming that this will also be the service period for all services (including delay sensitive ones), the scheduling period also directly affects (if not represents) the delay

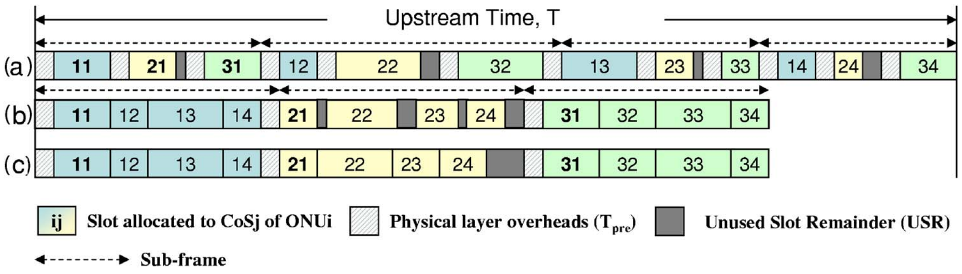

Fig. 1. Example of OLT burst allocation leading to the USR effect.

observed by CBR-like services. To achieve a delay guarantee of 1.5 ms for voice services [3,4], a scheduling period of equal duration should be selected. For an EPON supporting 16 ONUs each equipped, for example, with eight queues and a scheduling period of 1.5 ms, if the allocations to the queues are not contiguous, 16 μs$*$8 μs$*$1 μs is devoted to physical layer overheads, representing the 8.5% (128 μs/1500 μs) of the available upstream bandwidth. Thus, allocating contiguous transmission windows in all the queues of each ONU can decrease the bandwidth consumed to physical layer overheads up to eight times, or, in other words, from 8.5% to 1.2%.

For the preparation of upstream allocations, the MAC controller uses an allocation list, which contains precalculated grants (expressed in bytes per ONU and queue) and is scanned in a cyclic manner. The total number of bytes covers an upstream transmission window of duration D_m, i.e., it can schedule the transmission of up to 1 Gbits/s$*D_m$ of upstream traffic, called hereafter allocation list bytes (ALB). The allocation list consists of two consecutive entries per ONU, i.e, $2*N$ entries (where N is the number of registered ONUs). The first entry contains the number of bytes that will be granted without waiting for the ONU to place the relevant requests, i.e., scheduled as unsolicited grants, to service the two higher classes (UG_i) of the ONU at their guaranteed rate. The second entry contains the bytes that can be allocated dynamically if requested, as surplus bandwidth to service the second class up to its peak rate, the third, and the fourth [denoted as initially allocated slot (IAS_i)]. IAS_i is calculated per ONU based on its contracted service rate for the second and third priority as follows

$$IAS_i = \frac{w_i}{\sum_{i=1}^{N} w_i}\left(ALB - \sum_{i=1}^{N}\{UG_i + T_{pre} + T_{\mathrm{report}}\}\right),$$

where w_i is the service weight representing the share of the total upstream link capacity reserved for the surplus bandwidth to serve the second priority ($R_{p2}-R_{s2}$), guaranteed minimum rate for the third and potentially (left as an option) a weighted service of the fourth priority CoS queues of ONU i, and T_{report} is the time required for the transmission of a MPCP report message (64 bytes). The service weight w_i can be used to enforce proportional sharing of the upstream transmission window among ONUs. Since this represents an initial allocation of bandwidth based on the assumption that all ONUs appear backlogged during an interval D_m, the algorithm dynamically redistributes transmission intervals based on the actual requests from ONUs following the maximum–minimum fair-sharing algorithm.

Note that while UG_i bytes will always be allocated to the high-priority queue of ONU i, a grant between 0 and IAS_i bytes may be allocated to ONU i in a “use it or lose it” policy. Any unused portion of IAS_i can be made available to other ONUs in a dynamic fashion. Thus, according to actual demand per scheduling frame per ONU, the final scheduled transmission slot denoted as dynamically allocated slot (DAS_i), accounts for surplus bandwidth to serve excess requests of the second CoS queue up to its R_{p2}, bandwidth to serve upon request the third CoS queue up to its R_{g3} plus any surplus bandwidth to serve any excess requests from the third CoS queue and the total number of requests from the fourth CoS queue.

Upstream allocations are decided based on the above and the actual requests reported by ONUs in the previous scheduling frame in two steps: (i) first the number of bytes that will be allocated to each ONU for each queue are decided and then (ii) the exact position of the start and end transmission pointers are defined, shifting the allocations to maintain limited jitter and avoid disturbance of real-time services. For the calculation of the bytes to be allocated to the queues of the ONUs, the algorithm described in pseudocode in Fig. 2 is used.

5.C. Jitter Reduction

Following the above algorithm for the computation of the exact allocation of transmission slots to ONUs within the complete D_m interval, the only remaining issue is the computation of start transmission times for each ONU. For this computation, we take into account the following requirements: maintain low upper bounds on delay and jit-

```
/* allocate unsolicited grants to CoS1 & CoS2 queues of ONU i */
For $i$=1 to $N$
  $G_{i1} = UG_{i1}, G_{i2} = UG_{i2}, Req_{i2} = \max\{0, Req_{i2} - UG_{i2}\}$
$BW_{avail} = \Sigma_{(i=1 \text{ to } N)} \{IASi\}$
/* allocate surplus bandwidth to CoS2 up to their peak rate */
For $i$=1 to $N$
  $G_{i2} = G_{i2} + \min\{R_{p2i}*D_m, Req_{i2}\}$
  $BW_{avail} = BW_{avail} - \min\{R_{p2i}*D_m, Req_{i2}\}$
  $Req_{i2} = \max\{0, R_{p2i}*D_m - Req_{i2}\}$
/* allocate initial bandwidth ($R_{g2}$) to 3nd priority queues */
For $i$:=1 to $N$
  $G_{i3} = \min\{R_{g3i}*D_m, Req_{i3}\}$
  $BW_{avail} = BW_{avail} - \min\{R_{g3i}*D_m, Req_{i3}\}$
  $Req_{i3} = \max\{0, R_{g3i}*D_m - Req_{i3}\}$
/* Max-Min subroutine – Fair sharing of excess bandwidth */
/* init */
For $i$=1 to $N$
  $contendingi = 1, R_3 = Req_{i3}, R_4 = Req_{i4}, Req_i = R_3 + R_4$
$end = 0$
while not $end$  /* Recalculate shares */
  $sum_w = \Sigma_{(i \in Contending)} \{wi\}$      /* i ∈ Contending | $contending_i$ = 1 */
  $end = 1$        /* If all requests are below $FairShare$ then end */
  For $i$=1 to $N$
      $FairShare = (BW_{avail} * w_i) / sum_w$
      If $Req_i < FairShare$) and ($contending_i$)
        $GBE_i = Req_i$
        $contending_i = 0$
        $BW_{avail} = BW_{avail} - GBE_i$
        $end = 0$
      Else if ($contending_i$)
        $GBE_i = FairShare$
  $G_{i3} = G_{i3} + GBE_i * R_3 / Req_i$
  $G_{i4} = GBE_i * R_4 / Req_i$
  $Req_{i3} = \max\{0, Req_{i3} - (GBE_i * R_3 / Req_i)\}$
  $Req_{i4} = max\{0, Req_{i3} - (GBE_i * R_3 / Req_i)\}$
```

Abbreviation	Parameter/Function
IAS_i	Initially Allocated Slot for ONU *i* (bytes)
DAS_i	Dynamically Adjusted Slot for ONU *i* (bytes)
UG_{ij}	Unsolicited Grant for ONU *i*,Queue *j* (bytes)
Req_{ij}	ONU *i*,CoS Queue *j* unserved bandwidth Requests (bytes)
G_{ij}	Grant targeting CoS queue *j* of ONU *i* (bytes)
EXC_i	Excessive (compared to IAS_i) requests of ONU *i* (bytes)
GAPi	Unallocated slot (remainder of IAS_i for ONU *i*) (bytes)
TN	Next Transmission Start Time
TS_i	ONU *i* transmission start time (expressed in bytes for convenience)
T_{pre}	Layer 1 overhead (guard time, preamble, delimiter) duration in bytes

Fig. 2. Guaranteed allocation and fair sharing of surplus bandwidth following the maximum–minimum criterion and notations.

ter of real-time services and schedule transmissions from all CoS queues of an ONU in contiguous slots so as to avoid unnecessary physical layer overheads incurred due to burst-mode transmission.

Therefore in *each* scheduling round, the start time of the upstream transmission window of each ONU is initialized to a fixed precomputed value ($\Sigma_1^k\{UG_i+IAS_i\}$ for ONU k). As long as (UG_i+IAS_i) bytes suffice to service every active ONU every D_m, which is usually the case in low- and medium-offered loads, the start times are kept equal to the initially computed value, and each ONU ends up transmitting periodically every D_m. Thus, high-priority traffic is served in a perfectly periodic way, and any observed delay variation stems from the fact that the grant period does not coincide with the packet interarrival time. The latter coefficient of jitter is inevitable in any EPON MAC protocol, since the access latency introduced by the distributed nature of the access control scheme introduces burstiness in the traffic profiles. Even if no other multiplexing occurs, due to flow aggregation in FIFO queues, the aggregate granting per D_m introduces the possibility that multiple packets arriving in a CoS queue, even if they conform to a specific source traffic profile, have to wait until they are transmitted as a single burst over the upstream link.

When the offered load increases (e.g., beyond 80% as will be shown in the performance evaluation section), the temporal fluctuations are more prominent and the silence (off periods) of some ONUs is exploited to service the peak of others, realizing DBA. Although on average the offered load can be serviced, the requests of some ONUs cannot be completely satisfied using only the initially calculated (UG_i+IAS_i) slot. In this case, the bandwidth, which is left unused by the silent queues, is proportionally allocated to backlogged queues, which are thus granted more than (UG_i+IAS_i) amount of bytes (this is described in detail in the pseudocode of Fig. 2). To keep the allocations targeting the queues of the same ONU contiguous, the start times of the upstream transmission window have to be shifted from their initial values, introducing variation in the high-priority periodic grant schedule (a problem also identified in [9]). To minimize the introduced delay variation, the required space is pursued in the neighboring allocations. This is actually achieved by the iterative shifting of transmission times by eliminating idle times (denoted as “GAPs” in the schedule in Fig. 3) in an alternating fashion (left–right). This is depicted in Fig. 3 for ONUs 3 and 5 that request more bytes than IAS_3 and IAS_5, respectively, for a case where six active ONUs are being scheduled and ONUs 1, 2, 4, and 6 have requested fewer bytes than their corresponding *IAS* estimations in the depicted scheduling round. The allocation of ONU 3 is accommodated by shifting only the allocation of ONU 4 and exploiting the GAP caused by the bandwidth leftover of ONU 2. It is worth stressing that ONU 1 and 2 allocations are not moved from their initial positions. A rigorous description of this mechanism can be found in Fig. 4. The end result is that this shifting in time, although it represents a deviation from an ideal CBR service, is performed in a uniform and smooth fashion so as to incur the smallest possible jitter in the highest-priority service.

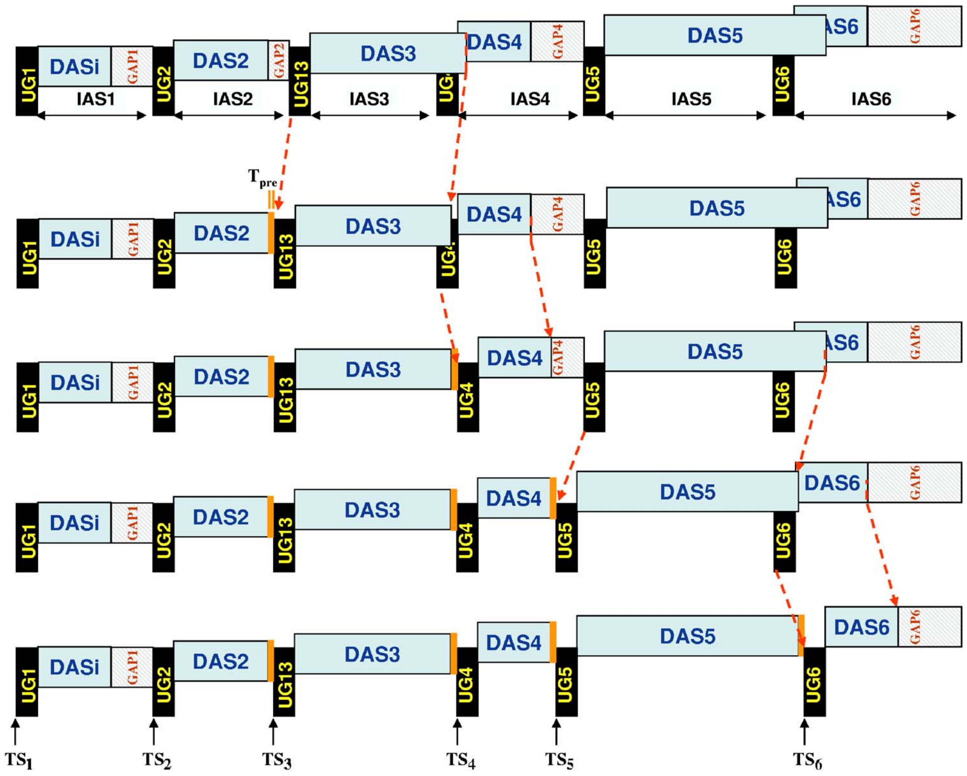

Fig. 3. Example of allocation list and related dynamically scheduled upstream transmissions.

```
MAX-MIN(Req)
For i=1 to N
  $DAS_i = IAS_i$, $GAP_i = 0$ $EXC_i = 0$
  If $IAS_i > G_{i3} + G_{i4}$
    $GAP_i = IAS_i - (G_{i3} + G_{i4}) + T_{pre}$
  Else
    $EXC_i = G_{i3} + G_{i4} - IAS_i$
k=0
For i=1 to N
  While $G_{BEi} > DAS_i$
    If (k mod 2)        /* ShiftLeft */
      $j = i - (\lfloor k/2 \rfloor + 1)$
      If i > 0 Then ReduceGap(i, j)
    Else    /* ShiftRight */
            $j = i + (\lfloor k/2 \rfloor + 1)$
            If j < N Then ReduceGap(i, j)
k = k+1

 /* Update final schedule… */
 TN=0
For i=1 to N
  $TS_i = TN$
  $TN = TN + G_{i1} + G_{i2} + G_{i3} + G_{i4} + GAP_i + T_{pre}$
/* ReduceGap(i, j) subroutine
$prevEXC_i = EXC_i$
If $prevEXC_i > GAP_j$ and $GAP_j > T_{pre}$
            $EXC_i = prevEXC_i - (GAP_j - T_{pre})$
            $GAP_j = T_{pre}$
Else If $GAP_j - prevEXC_i > T_{pre}$
            $GAP_j = GAP_j - prevEXC_i$
            $EXC_i = 0$
$DAS_i = DAS_i + (prevEXC_i - EXC_i)$
```

Fig. 4. Allocation shifting and pointer computation.

5.D. Impact of Intra-ONU Scheduling on Performance

To provide service precedence to second or third priority traffic, it is necessary that the ONUs report the length of the relevant queues independently and that the bandwidth allocation controller at the OLT decides the allocations per ONU and CoS. However, leaving the ONU to decide the final allocation of upstream granted time to its queues, improves performance both in terms of efficiency and access delay. This, on the other hand, comes at the cost of adding the complexity of per CoS queue management and scheduling at the ONUs, which we consider affordable if the ONU scheduler's complexity is kept low. This means that although the OLT has computed the start pointers and the length of the transmission *per queue*, the ONU respects only the start pointer for the high-priority queue and the total time allocated to this ONU. To decide which queue to service, strict priority scheduling is employed, which is shown in Section 6 to be an adequate solution easy to implement in the cost-sensitive access era. This leads to delay reduction because the allocation included in the GATE message was decided based on the queue status reported one round-trip time earlier, and it is very likely that the queuing situation has changed, e.g., a packet has arrived in the meantime in the second priority queue. In this case, this packet will exploit time allocated to the third priority of the same ONU for its transmission. The same may happen for third priority packets. The delayed (unserviced) lower-priority traffic will be reported again and will be serviced with the next GATE message. Allowing the ONU to decide the distribution of upstream time to its queues results in higher utilization, since no time allocated to a specific queue will be wasted due to a mismatch of time left over and head-of-line packet length. Also, drop policies can be independently applied at the ONU queuing points, but this has been left as a topic for further research and the results presented below have been collected under the infinite buffer assumption.

6. Performance Evaluation

To evaluate the proposed algorithm, a simulation model was developed using the OPNET simulator. It includes 16 ONUs, each equipped with four different queues. The offered load is shared uniformly among all ONUs, D_m was set to 2 ms while the duration of the guard band and the physical layer overhead transmission (i.e., T_{pre}) were assumed equivalent to 1 μs. Several scenario sets were carried out in order to assess the performance of the proposed scheme according to the metrics listed Subsections 5.A–5.D. A realistic traffic mix has been used in which high priority represented the 10% of the total offered load, while second, third, and fourth priority were injecting 15%, 20%, and 55% of the total load, respectively. High-priority sources were of CBR generating short fixed-size packets periodically (a model mostly applicable for voice traffic) while the source used for the rest of the types of traffic were of ON—OFF type (which is best applicable for self-similar Internet traffic), but with different burstiness factors. Namely, the burstiness was chosen 2, 5, and 5 for second, third, and fourth priority, respectively. Apart from the first class sources, the rest generate packets with a size following the trimodal distribution characterizing traffic generated from IP-based applications (packet sizes of 64, 500, 1500 bytes appear with probability 0.6, 0.2, and 0.2, respectively, according to [15]).

6.A. Access Delay and Throughput

Our first goal was to measure the average access delay as a function of the offered load and investigate the limitations in the achievable throughput. The results are included in Fig. 5, where the average delay values for each of the four classes (cumulative for all ONUs) is depicted. A first observation is that the system can handle offered load up to more than 90% of the link capacity (1 Gbits/s). Concerning the service differentiation capabilities we note that above 90%, only the best-effort traffic suffers the congestion, while the other three achieve 100% throughput with respect to their contribution to the total offered load. The best-effort traffic starts observing high-maximum-delay values when the offered load is higher than 70%. Most important, the higher-priority classes observe limited and totally acceptable average delays (even when the total offered load equals 100%, which satisfies our objectives for guaranteed delay bounds). Thus, perfect isolation is achieved. The third priority starts experiencing congestion only above 90%, when its maximum delay values start increasing, while first and second priority remains stable since they represent 10% and 15% of the offered load, respectively, and their service has been guaranteed through preprogrammed grants.

An interesting parameter that affects both the delay and the utilization is the portion of the sustainable rate of the second CoS serviced through unsolicited grants. To assess the impact of this parameter, two different scenario sets were tested. In the first, the second CoS queues were serviced unsolicitedly at their sustainable rate (i.e., $UG_{i2}=R_{S2}*D_m$) while in the second set they were serviced at half the sustainable rate ($UG_{i2}=R_{S2}/2*D_m$). The observed delay is shown in Figs. 1(a) and 5(b), respectively, while the utilization is depicted in Fig. 5(d). The later option achieves higher utilization without significant impact on the delay and jitter (as will be shown in Subsection 6.B) of the second CoS. This can be easily explained since a higher number

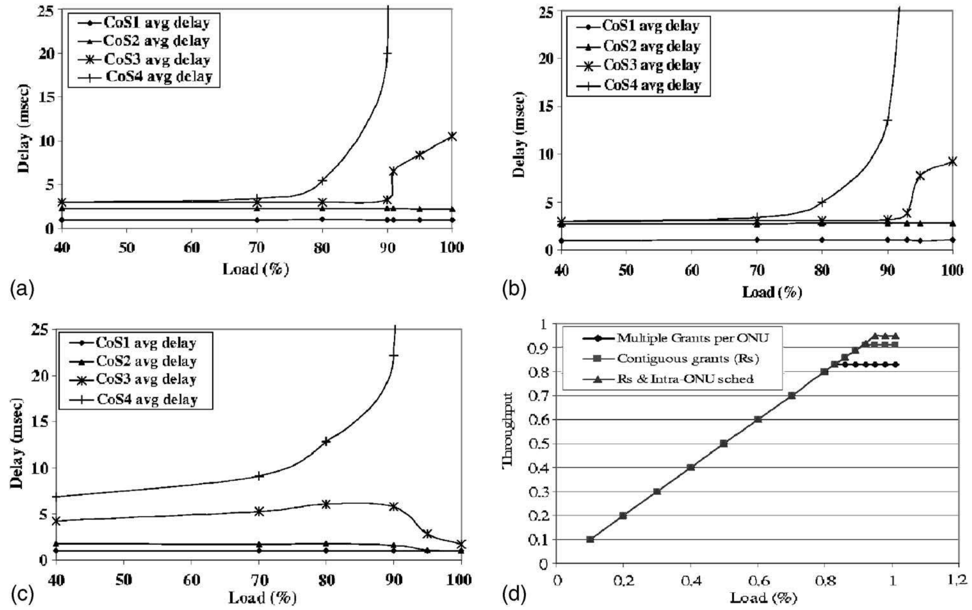

Fig. 5. Average queuing delay per CoS (a) $UG_{i2}=R_{S2}*D_m$, (b) $UG_{i2}=R_{S2}*D_m/2\ R_{S2}$, and (c) $UG_{i2}=R_{S2}*D_m$, and strict priority intra-ONU scheduling (d) upstream bandwidth utilization.

of unsolicited grants for the second CoS traffic of bursty nature increases the probability of scheduling allocations larger than the actually arriving traffic at the ONU during some of the scheduling intervals.

This mismatch between allocations and queue status can be compensated, if intra-ONU scheduling is employed as shown in Fig. 5(c), simply by employing strict priority queuing. This scheme is responsible for the dip in delay for CoS3 for loads between 90% and 100%. In more detail, as the load increases, the amount of requested bytes per ONU increases as well (mainly due to CoS4 traffic, which represents the higher percentage) causing the uniform distribution of almost all the upstream bandwidth to the ONUs. When congestion occurs this has an impact only on the fourth priority queue, which above that point remains always congested (overflowing in an actual implementation) but actually enhancing the performance of the higher-priority queues since it constantly places requests to the OLT for service. The resulting allocations suffice to completely service even the newly arrived (and not reported yet) packets of CoS3. Thus, CoS3 packets "steal" the allocations triggered by CoS4 reports, leaving CoS4 to suffer the congestion. Employing intra-ONU scheduling actually results in trading-off CoS4 throughput with overall system utilization and CoS3 performance, which is not only acceptable but also an inherent objective of the proposed mechanism, since performance guarantees per CoS are tightly coupled with associated tariffs, which should be part of the service level agreements entered into during service initialization.

Finally, the overall throughput achieved in cases (a) and (b) is shown in Fig. 5(d) and is compared with the results reported in [11] (copied here to obtain the "multiple grants per ONU" curve). It is evident that adopting the proposed algorithm, higher bandwidth utilization (91% versus 83.5% in [11]) can be achieved as a consequence of relaxing the requirements for strictly fixed periodic scheduling of grants for CoS1, with insignificant impact on jitter as will be discussed next. This performance gain is the consequence of allocating contiguous transmission windows to queues of the same ONU. Further efficiency improvement is observed (up to 95%) when intra-ONU scheduling is adopted.

6.B. Jitter Performance

As mentioned above the proposed scheduling process may lead to a deviation from an ideal CBR service for CoS1 but the shifting in the time algorithm described in Figs. 3 and 4 contributes to a uniform and smooth distribution of delay in time so as to minimize jitter. This is shown in the curves of the probability density function (PDF) for the (cumulative among all ONUs) delay per CoS depicted in Figs. 6(a)–6(d) at 80% load, which is a value near network saturation.

As observed in Fig. 6(a), the maximum delay rarely exceeds the $D_m = 2$ ms bound, which was the selected operational parameter as a means to achieve statistical delay-bounds for delay sensitive traffic while enhancing efficiency. (This has been observed at all loads even when the overall network load exceeds the upstream capacity not shown due to space limitations.) The delay of the high priority is almost uniformly distributed between 0 and 2 ms as expected, while a very small, almost negligible percentage of packets observe delays between 2 and 2.4 ms. Note that the distribution of delay values across the whole spectrum of values is caused by the inherent bursty nature of the EPON MAC, since in most cases the scheduling period does not match the actual source period as mentioned in Subsection 5.C. This is also reported in the results included in [11] as shown by the periodic peaks of the curve of the autocorrelation function of the delay included therein. Implementing strict priority intra-ONU scheduling improves even more the jitter performance for CoS1, although the average delay remains almost unaffected as shown in Figs. 5(a) and 5(b).

For the second priority traffic, values higher than 2 ms are observed due to the reservation-based bandwidth allocation but is still bounded by $2*D_m$ for second priority (since enough bandwidth to service this type of traffic is always available due to prior acceptance). As shown in Fig. 6(b), when the unsolicited grants service the second CoS queue at its sustainable rate, the probability for the delay to be between D_m and $2*D_m$ is higher than the probability for the delay to be between 0 and D_m. This effect is reversed when simple priority intra-ONU scheduling is employed. In the same figure, it can be observed that when the unsolicited grants service the second Cos queues at half their sustainable rate, the PDF curve is biased toward the D_m and $2*D_m$ delay range.

The difference between the second and the third CoS [depicted in Fig. 6(c)] is that due to the provisioned service rate for the second, values below 2 ms are possible while for the third there is no possibility to achieve delay lower than D_m since the requests cannot be serviced before the next scheduling round. Only when intra-ONU scheduling is employed, CoS3 steals allocations caused by fourth priority requests, thus CoS3 is favored over CoS4. As in any other queuing situation, a bell-shaped curve represents the delay of best effort traffic.

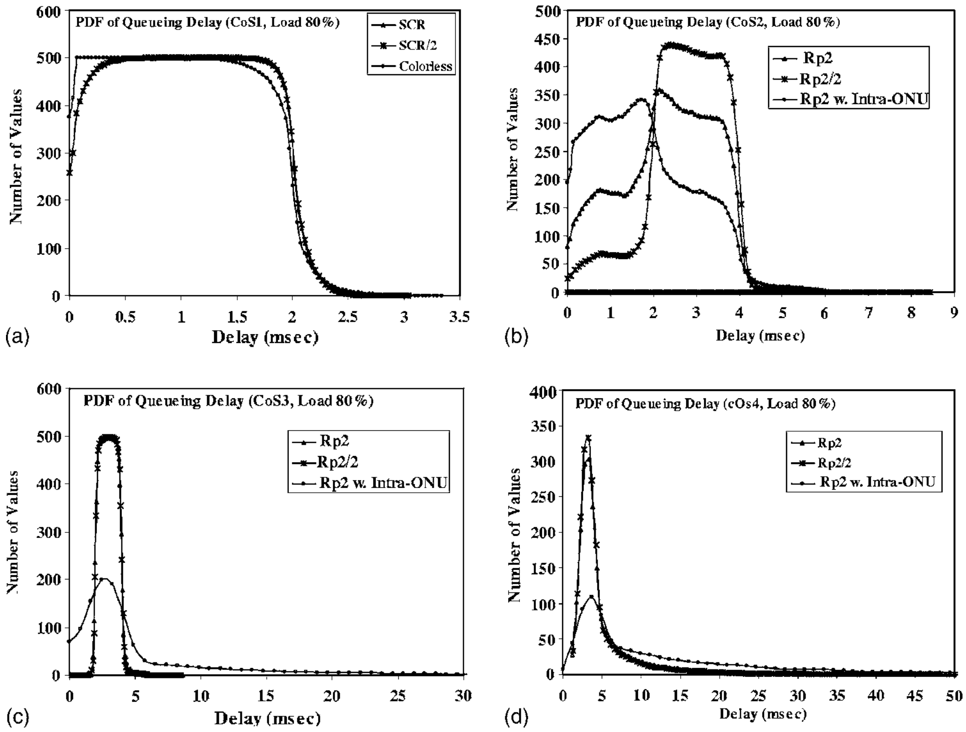

Fig. 6. PDF of delay per CoS for $UG_{i2}=R_{S2}*D_m$, $UG_{i2}=R_{S2}*D_m/2$ R_{S2}, and $UG_{i2}=R_{S2}*D_m$ and strict priority intra-ONU scheduling.

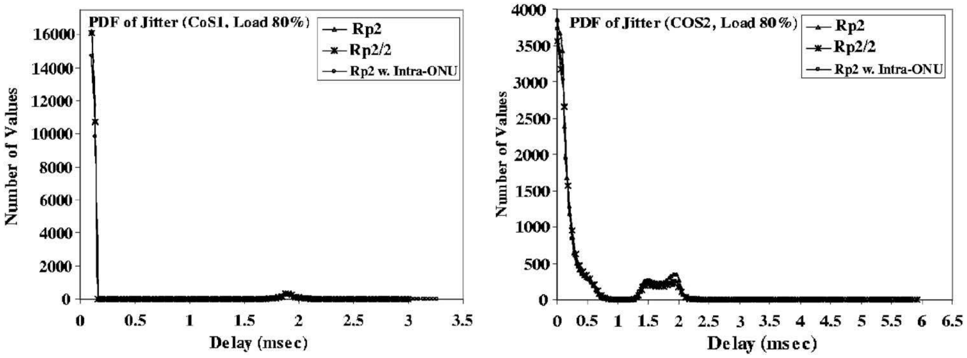

Fig. 7. PDF of packet delay variation for CoS1 and CoS2.

To illustrate the impact of the proposed algorithm on the jitter performance we show in Fig. 7 the PDF of the cumulative (across all ONUs) packet delay variation for the real-time classes (CoS1 and CoS2, in which cases jitter minimization is required) expressed as the one-way interpacket-delay-variation (ipdv) metric as defined in [16]. One can easily note the differentiation between CoS1 and CoS2 (evident by the slope of each curve), which was expected. A second observation is that when intra-ONU scheduling is employed jitter is slightly affected and this is caused by the temporal randomization of the periodic service of high-priority queues caused by the bandwidth-stealing effect discussed above. Finally the impact of the allocation shifting and pointer computation algorithm described in Fig. 4, implemented in order to improve the system bandwidth efficiency, is related to the second lower peak of the jitter curve (hardly visible for the CoS1 case) in the range between 1.5 and 2 ms in Fig. 7. The observed values around this point are caused by the sporadic shifting of high priority grants earlier or later than their original schedule infrequently causing packets to miss the scheduled transmission slot. For CoS1, which has strict performance requirements, we consider this effect negligible, and a potential way to reduce this even further would be to reduce the scheduling interval D_m The latter option though, would in turn have a negative effect in bandwidth utilization. Such an option could be handled better by a protocol with less overhead and more flexible polling mechanisms such as GPON, a comparison, which we intend to investigate in our future work [17].

7. Conclusions

To efficiently support all kinds of services, the proposed MAC algorithm assumes traffic segregation at the ONU side and allocates bandwidth based on discrete classes of service requirements. The algorithm can guarantee strict delay bounds for delay-sensitive traffic and efficiently multiplexes delay-tolerant traffic in a dynamic fashion, also enforcing proportional bandwidth sharing. As demonstrated by simulation results, service discrimination among classes can achieve very good performance even for real-time applications with stringent requirements, while also supporting different rate shares per ONU and per class of service.

References and Links

1. J. Angelopoulos, H. Leligou, T. Argyriou, S. Zontos, E. Ringoot, and T. Van Caenegem, “Efficient transport of packets with QoS in an FSAN-aligned GPON,” IEEE Commun. Mag. **42**, 92–98 (2004).
2. IEEE Std 802.3ah-2004 (IEEE, 2004).
3. ITU-T, Study Goup 15, “Gigabit-capable passive optical networks (GPON): General characteristics,” Recommendation 6.984.1 (ITU, 2003).
4. A. Shami, X. Bai, N. Ghani, C. M. Assi, and H. T. Mouftah, “QoS control schemes for two-stage Ethernet passive optical access networks,” IEEE J. Sel. Areas Commun. **23**, 1467–1478 (2005).
5. M. P. McGarry, M. Maier, and M. Reisslein, “Ethernet PONs: a survey of dynamic bandwidth allocation (DBA) algorithms,” IEEE Commun. Mag. **42**, S8–S15 (2004).
6. G. Kramer, B. Mukherjee, and G. Pesavento, “Interleaved polling with adaptive cycle time (IPACT): a dynamic bandwidth distribution scheme in an optical access network,” IEEE Commun. Mag. **40**, 74–80 (2002).
7. X. Bai, A. Shami, and C. Assi, “Statistical bandwidth multiplexing in Ethernet passive

optical networks," in *Proceedings of the Global Telecommunications Conference (GLOBECOM '05)*, (IEEE, 2005).

8. H. N. Ghani, A. Shami, C. Assi, and M. Raja, "Intra-ONU bandwidth scheduling in Ethernet passive optical networks," IEEE Commun. Lett. **8**, 683–685 (2004).
9. M. Ma, Y. Zhu, and T. H. Cheng, "A bandwidth guaranteed polling MAC protocol for Ethernet passive optical networks," in *Proceedings of the Twenty Second Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003)* (IEEE, 2003).
10. F. A, Y.-L. Hsueh, K. S. Kim, I. M. White, and L. G. Kazovsky, "A new dynamic bandwidth allocation protocol with quality of service in Ethernet-based passive optical networks," in *Proceedings of the International Conference on Wireless and Optical Communications (WOC03)*, Banff, Canada, 14–16 July 2003.
11. A. Shami, X. Bai, C. M. Assi, and N. Ghani, "Jitter performance in Ethernet passive optical networks," J. Lightwave Technol. **23**, 1745–1753 (2005).
12. T. Orphanoudakis, H. C. Leligou, E. Kosmatos, and G. Prezerakos, "Dynamic traffic mulitplexing over EPONs achieving guranteed QoS and optimal bandwidth utilization," presented at the *International Conference on Photonics in Switching 2006*, Herakleion (Crete), Greece, 16–18 October 2006.
13. J. D. Angelopoulos, H. C. Leligou, T. G. Orphanoudakis, and J. Pikramenos, "Using a multiple priority reservation MAC to support differentiated services over HFC systems," Int. J. of Commun. Syst. **15**, 325–340 (2002).
14. Data-Over-Cable Service Interface Specifications, Operations Support System Interface Specification, SP-OSSIv1.1-I02–000714, available at www.cablemodem. com.
15. K. Claffy, G. Miller, and K. Thompson, "The nature of the beast: recent traffic measurements from an Internet backbone," in *Proceedings of Internet Society (ISOC) INET'98*, Washington, D.C., 21–24 July 1998.
16. C. Demichelis and P. Chimento, "IP packet delay variation metric for IP performance metrics (IPPM)," Internet Engineering Task Force (IETF) RFC 3393 (IETF, 2002).
17. S. Lallukka and P. Raatikainen, "Link utilization and comparison of EPON and GPON access network cost," in *Proceedings of the Global Telecommunications Conference (GLOBECOM '05)* (IEEE, 2005).