# Providing Delay Bounds for Real-time Traffic over EPONs

HELEN-C. LELIGOU, THEOFANIS ORPHANOUDAKIS, KONSTANTINOS KANONAKIS,
GEORGE PREZERAKOS, JOHN D. ANGELOPOULOS
Technological Educational Institute of Piraeus, P. Ralli & Thivon 250, Aigaleo GR12244 GREECE
(phone +30-210-5381338, fax: +30-210-5381260

*Abstract:* - Since the emergence of the Ethernet-based PON technology, an increasing interest is observed regarding the issue of Dynamic Bandwidth Allocation. Several mechanisms have been proposed aiming to enhance the multiplexing techniques and the Medium Access Control protocols that have been accepted in the recent IEEE standard. In this paper we propose a new mechanism which targets strict delay bounds for real time traffic, while efficiently multiplexing multiple classes of traffic and achieving high utilization of the EPON upstream channel.

*Key-Words:* - EPON, MAC, DBA, QoS

## 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 last decade substantial commercial activity also reflected in the ongoing work of several standardization bodies. Since the initial deployment of ATM-based PONs (APONs) newer standards support multi-gigabit rates and adapt better to the packet-based Internet applications. The Full Service Access Networks (FSAN) group has recently produced its second generation standard for the so-called Gigabit PON (GPON) supporting mixed ATM and packet based services reaching symmetrical transmission rates of up to 1.244Gbps or 2.488Gbps ([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.

In this paper we focus on a novel access control mechanism that could enhance the performance of EPONs in terms of both bandwidth management and Quality of Service (QoS) provided to the network subscribers. The proposed access mechanism can be implemented in compliance with the 802.11ah protocols ([2]) effectively providing Dynamic Bandwidth Allocation (DBA), which has been a concern to many researchers and system architects recently.

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. In order 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. If this can be achieved then a bundle of services can be available over EPONs at competitive prices, attract users and increase network utilization at acceptable levels. The efficient support of different quality of service 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 transmission, telephony and data). Both delay-sensitive and best effort applications should be simultaneously supported in the emerging PONs where -unlike APONs where provisioning is implemented per virtual connection ([3])- a signaling infrastructure is not present. 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 (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 802.11ah standard defines the so-called Multipoint Control Protocol (MPCP) and the type of messages that should be exchanged during operation; it doesn’t specify though specific 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.

Several recent articles investigate both architectural issues and MAC protocols (a review of the most well known can be found in [4]). Initial attempts to efficiently implement DBA more or less depended on best effort polling of requests [5]. Most research attempts focused on the problem of fair or weighted sharing of bandwidth among users (e.g.[6]) or differentiated service through the discrimination of service classes through DBA (e.g. [7]). 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 [8] and [9]). The above mentioned approaches have correctly identified the need to pre-allocate 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. These approaches strongly resemble the so called Unsolicited Grant Service (UGS) of the DOCSIS 1.1 ([10]) 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. Our DBA mechanism also follows the approach of statically preallocated bandwidth for real-time traffic, provided by means of unsolicited grants (a concept called GBR in [9]), while proposing an enhanced scheduling frame structure that can achieve both deterministic strict delay bounds and low delay variation for real-time traffic, as well as efficient multiplexing of delay tolerant traffic, scheduling transmission grants for multiple service queues in contiguous slots within a single burst whenever possible.

The remainder of the paper includes a short description of the system architecture and the description of the proposed algorithm which is evaluated using computer simulations in section 4.

## 2 System Architecture

PONs are based on a passive star fibre network, which connects a number of ONUs (Optical Network Units) at the subscriber side to one OLT in the local exchange, as shown in Fig. 1. The traffic streams arriving at the ONUs from the customer premises are kept in queues. In compliance to the 802.1Q prioritization scheme it is possible to inject the traffic in up to 8 logically separate, possibly prioritized, queues holding Ethernet frames, depending on QoS requirements, to allow for the enforcement of different service mechanisms.

The downstream direction 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 TDMA fashion. The activation of each ONUs’ transmitter and window of operation is controlled by the MAC controller in the OLT. In order 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.

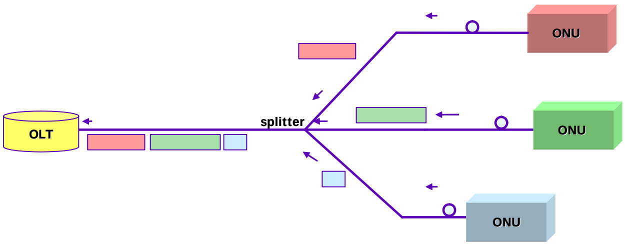

Fig. 1: The EPON architecture

The MAC controller at the OLT distributes the upstream bandwidth ensuring collision avoidance and bases its decisions on the report messages to match the fluctuating queue occupancy. Although it is feasible to schedule grants taking into account only the aggregate traffic queued at each ONU, this practice is not efficient, since it can lead to cases where lower priority traffic from one ONU can gain access to the upstream channel, while traffic from a higher priority queue of another ONU is waiting for its transmission grant. Thus it is preferable to take into account the status of individual queues of the ONUs. This does not necessarily imply that the ONU cannot decide how to allocate the granted bytes to its own queues, i.e. an ONU can decide to use bytes issued by the MAC controller for its low priority queue to service its high priority queue ([7]).

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 whenever the relevant GATE message indicates so. In this work, we consider just two priority queues at the ONU side, as a minimum requirement for the EPON multiplexing function to provide differentiated levels of service, while it is easy to extend the concept and the proposed algorithm to support multiple queues and service classes.

## 3 Mac Algorithm

The proposed algorithm aims at providing guaranteed delay bounds to real-time traffic, while dynamically distributing unused bandwidth to bursty traffic with no strict QoS guarantees. A first observation towards defining our bandwidth allocation algorithm is that for those applications that generate Constant Bit Rate (CBR) traffic, it would be feasible to compute a static distribution of the upstream bandwidth based on the contracted rates Si (i.e. a virtual leased line of service rate Si). Assuming that T is the granting cycle during which each ONU is assigned Wi bytes, Si=Wi/T. Although this way the maximum queuing delay that a packet may observe is limited by T, which is a very desirable effect for delay-sensitive applications such as voice, at the same time a static allocation would cause utilization degradation, for the rest of the applications due to the dominant effect of burstiness as witnessed for IP-based traffic. It should also be noted that during an active period (burst), an IP application typically generates packets of variable size. Since the amount of traffic to be served at each time interval cannot be known apriori and the EPON standard does not support segmentation and re-assembly (therefore an arbitrary byte allocation can be wasted if it cannot serve an integral packet), the enforcement of static allocations would unavoidably result in low utilization. To avoid this inefficiency, the REPORT messages are used in MPCP for the ONUs to announce -frequently- the queued traffic so that the MAC controller at the OLT will adjust the bandwidth allocation according to the queue occupancy, i.e. will effect DBA. However, employing this solution a packet may wait at the ONU queue for time equal to the round trip time –for the report of its generation to reach the OLT and the grant scheduled to enable its transmission to reach the ONU- augmented by the scheduling time, which incurs a high and variable delay. To this end, the traffic is kept in different queues at the ONU side depending on the QoS level agreed between user and operator (each queue length being reported independently) and the MAC controller at the OLT side implements two different service strategies to combine the advantages of static and dynamic allocations.

Hereafter we discuss the operation of the proposed algorithm assuming for reasons of simplicity that just two QoS levels are employed. For the high quality class, which can be analogous to the Expedited Forwarding (EF) class of the IETF DiffServ architecture (obviously associated with a higher tariff), an operator guarantees a strict delay bound Dm. In order to achieve this, a service rate has to be negotiated and respected between user and operator. For the lower quality class no guarantee is provided (i.e. Best Effort -BE- service), although a minimum rate can also be guaranteed for this type of traffic. In the following we denote these two guaranteed rate thresholds for the ith ONU as SHi and SLi resprectively.

In this context we propose a novel algorithm that can efficiently support DBA, while guaranteeing a strict upper bound on the delay of the high quality class (hereafter denoted as EF) traffic by pre-allocating transmission grants spaced in time. The spacing of these pre-allocated grants is performed in a deterministic way, which additionally results in low delay jitter for EF traffic. Our grant allocation mechanism is based on a fixed (and periodic) scheduling frame of duration Dm, which 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 preparation of upstream allocations the MAC controller uses an allocation list (denoted as AL), which is scanned in a cyclic manner, and contains pre-calculated grants expressed in bytes. The total number of bytes in this allocation matrix covers a transmission window of duration Dm, i.e. can schedule the transmission of up to 1Gbps*Dm of upstream traffic. The allocation list of length 2*N (where N is the number of registered ONUs) contains two consecutive entries for each ONU. The first entry contains the number of bytes that will be allocated to the EF class of the ONU. These are granted without waiting for the ONU to place the relevant requests, i.e. scheduled as unsolicited grants. Therefore we denote this byte allocation as UGi=SHi*Dm. The second entry contains the number of bytes that can be allocated -if requested- to the BE traffic dynamically (denoted as DABi). The logical organization of the allocation list and corresponding scheduled upstream transmissions is shown in Fig. 2. Note that while UGi bytes will always be allocated to the EF queue of ONU i, DABi bytes may be allocated to ONU i or another ONU as will be explained later on.

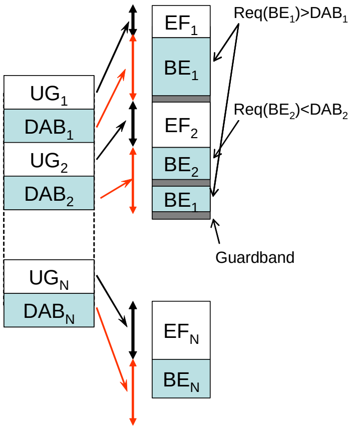

Fig. 2: Example of allocation list and related dynamically scheduled upstream transmissions

The proposed allocation mechanism strictly limits upstream allocations in Dm intervals (e.g. 2msec is deemed adequate for most real time services), which favors predictable performance. Additionally, in order to effect a fair (potentially weighted) bandwidth allocation for the BE traffic of each ONU, a service factor Q as a function of SLi is defined. Q expresses a service quota for each ONU i.e. the number of bytes that can be allocated to this ONU over a time interval Tq, multiple of Dm, assuming an average transmission rate SLi, i.e. Q= SLi*Tq. This factor can effect a weighted sharing of the available bandwidth among BE traffic of all ONUs, sharing it proportionally according to Q. The averaging window Tq is selected long enough -in contrast to the short Dm cycle- in order to favor queue length fluctuations in longer time intervals and higher average BE queue backlogs. The reason for that is that the BE class intends to support applications generating bursty traffic, which are delay-tolerant. Thus, it is preferable to allocate longer transmission windows less often, to avoid bandwidth waste on guard-bands and miss-filled allocations due to the absence of segmentation and re-assembly support.

The upstream allocations are decided every Dm based on the allocation list, the queue lengths reported by the ONUs stored in an array denoted as Req, and the quota Q of each ONU in two steps. First, each ONU is allocated UGi bytes for the service of its EF queue -as dictated in the allocation list- plus some additional bytes if its BE queue has placed requests, while in the second step further allocations to serve BE queues from all backlogged ONUs are computed, whenever possible. The first step guarantees that when both queues of an ONU are backlogged a single transmission burst to serve both can be scheduled reducing the physical layer overhead between burst transmissions from different ONUs. The second step actually effects dynamic and weighted bandwidth sharing, allocating any surplus bandwidth to BE traffic without disrupting the delay-sensitive EF traffic. Summarizing the notation we use in Table 1, the proposed grant scheduling algorithm can be described by means of the pseudo-code given in Fig. 3.

TABLE 1
NOTATION, VARIABLES & FUNCTIONS

| Abbreviation | Parameter/Function |
|---|---|
| $UG_i$ | Unsolicited Grant for ONU *i* (bytes) |
| $DAB_i$ | Unallocated grant(s) to be assigned dynamically (bytes) |
| $Req_i$ | ONU *i* unserved bandwidth Requests (bytes ) |
| $Q_i$ | ONU *i* Quota (bytes) |
| $G_{EF}$ | Grant targeting EF queue (bytes) |
| $G_{BE}$ | Grant targeting BE queue (bytes) |
| $GAP_i$ | Unallocated remainder of $DAB_i$ (bytes) |
| $T_i$ | ONU *i* UG transmission start time (expressed in bytes for convenience) |
| $Tg_i$ | $GAP_i$ transmission start time (bytes) |
| $T_{pre}$ | Physical overhead duration (bytes) |
| $GATE(i, g_{EF}, g_{BE}, T_{on}, T_{off})$ | GATE message generation, targeting ONU *i*, granting $g_{EF}$ and $g_{EF}$ (bytes) for the EF and BE queue respectively to be transmitted in contiguous burst starting at time $T_{on}$ and ending at $T_{off}$ |
| *MinAlloc* | Minimum grant size (e.g. equal to minimum packet size plus $T_{pre}$) |
| *LastSrvOnu* | Points to the last served ONU during previous scheduling frame |

Dynamic sharing of unallocated upstream time slots (represented by the values in the GAP matrix appearing in Fig. 3) is based on the execution of the second step iteration of the algorithm. ONUs are served in a round robin fashion (indicated by the auxiliary pointer LastSrvONU in Fig. 3) until exhaustion of either their request or their quota within an interval Tq. Obviously during ONU configuration quota values should be selected to satisfy the condition: ΣQi<Tq/Dm*Σ{DABi}). Depending on the policy for resetting Qi values to their initial values, the Q factors can be used to either enforce rate limiting in a non work conserving manner or weighted bandwidth sharing in a work conserving manner (if reset earlier than Tq is allowed).

It is worth stressing that the difference of the start pointers of EF grants for ONUs i, i+1 will always be equal to UGi+DABi, which contributes to the reduction of delay variation. GATE messages issued during the first step have always the reporting flag ON, i.e. they cause the ONU to transmit a REPORT message, which ensures an adequate ONU polling frequency (issued even in the case when an ONU has not subscribed for any EF class services). Finally the execution of the first step ensures that bandwidth for EF, REPORT and BE transmissions will be allocated in contiguous slots allowing for a single burst reducing physical layer overheads (due to the factor Tpre).

```
/* 1st STEP */
For $i$=1 to $N$
  $G_{EF} = UG_i$
  $G_{BE}$ = min$\{Req_i, Q_i, DAB_i\}$
  $Q_i = Q_i - G_{BE},$
  $GAP_i = DAB_i - G_{BE}$
  $Req_i = Req_i - G_{BE}$
  $Tg_i = T_i + G_{EF} + G_{BE} + T_{pre}$
  $GATE(i, G_{EF}, G_{BE}, T_i, T_i + G_{EF} + G_{BE})$
/* 2$^{nd}$ STEP */
$j$=$LastSrvOnu$+1
$end$=FALSE
For $i$=1 to $N$ /* GAP pointer */
  While $GAP_i$>$MinAlloc$ and not $end$
            $G_{BE}$ = min$\{Req_i, Q_i, GAP_i\}$
  $Q_i = Q_i - G_{BE}$
  $GAP_i = GAP_i - G_{BE}$
   If $G_{BE}$>0 $N$ $GATE(j, 0, G_{BE}, Tg_i, Tg_i + G_{BE})$
    $Tg_i = Tg_i + G_{BE} + T_{pre}$
    If $j \neq N$          $j=j+1$
         Else          $j$=1
    If $j$=$LastSrvOnu$+1 $end$=TRUE
  If $end$ break /* until all requests have been
served */
If not $end$ $LastSrvOnu$=$j$-1
   Else $LastSrvOnu$=$j$
```

Fig. 3: Scheduling algorithm

## 4 Performance Evaluation

To evaluate the proposed algorithm, a simulation model was developed. It includes 16 ONUs, each equipped with 2 different queues. The offered load is shared uniformly among all ONUs, Dm was set to 2ms while the duration of the guard band and the Physical layer overhead transmission (i.e. Tpre) were assumed equal to 1μs. The EF traffic was generated by CBR sources with short (64 Bytes) packets representing voice traffic, while the BE traffic was generated by on-off sources and the packet length followed the trimodal distribution i.e. 64, 500, 1500 bytes with probability 0.6, 0.2 and 0.2 respectively.

In Fig. 4 we show the queuing delay as a function of load for a traffic mix, where the EF traffic represents 10% of the total offered load. Both the average and maximum queuing delay values observed by the EF class do not depend on the total offered load as expected. It is worth stressing that the maximum delay never exceeds (even when the overall network load exceeds the upstream capacity) the 2ms bound, which was the selected operational parameter. Furthermore, in case lower delay bounds are required to satisfy specific services, with appropriate selection of Dm the proposed design would achieve even lower delay bounds. As regards the BE traffic, the delay observed is always higher than that experienced by the EF but remains limited as long as the offered load is below 90%. Above this, the effects of EPON physical and MAC layer overheads present a significant impact on BE delay but even in this case perfect isolation for the EF traffic is achieved and only the BE class suffers the congestion.

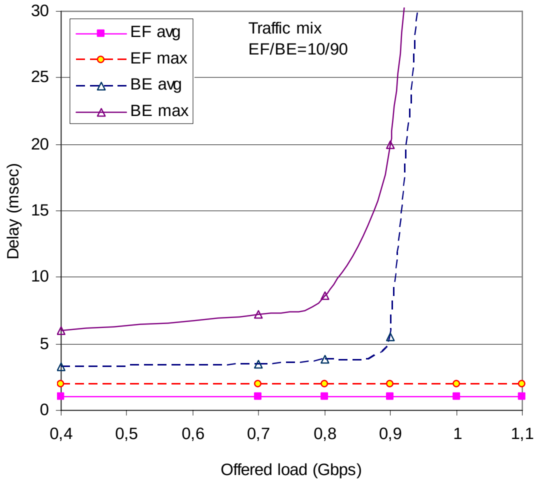

Fig. 4: Average and maximum queuing delay for BE and EF traffic vs. load

Since the EF class would typically be used for voice and video applications the average queuing delay is not the only performance metric of interest. In Fig. 5 we also show the Probability Density Function of access delay for an offered load of 70%. Due to the deterministic service offered to the EF class we note a uniform distribution of the experienced delay values around the average value of 1ms, tightly bounded by the worst case delay of 2ms. As expected, BE traffic observes delay values spread in a larger interval, which is considered acceptable for the delay tolerant nature of this class.

To illustrate the use of Q as a policing tool, a scenario where a single ONU is loaded at a rate higher than the sustained rate configured for this ONU at the OLT (i.e. out of profile traffic) is shown next. Quota for this ONU have been set assuming a service rate of 33Mbps, while its sources inject traffic at 45Mbps. For the other ONUs enough quota to satisfy their traffic load (also 45 Mbps) are assigned. As shown in Fig. 6, while the total offered load is below 0.8, the delay observed by the out of profile ONU is constantly increasing indicating heavy congestion (in practice buffer overflow conditions) while the BE traffic of other ONUs enjoys good performance (limited delay, slightly higher than EF traffic).

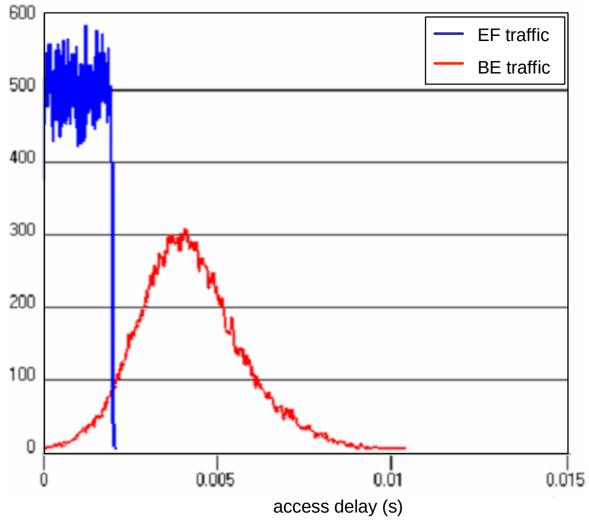

Fig. 5: Probability Density Function of access delay at 70% offered load

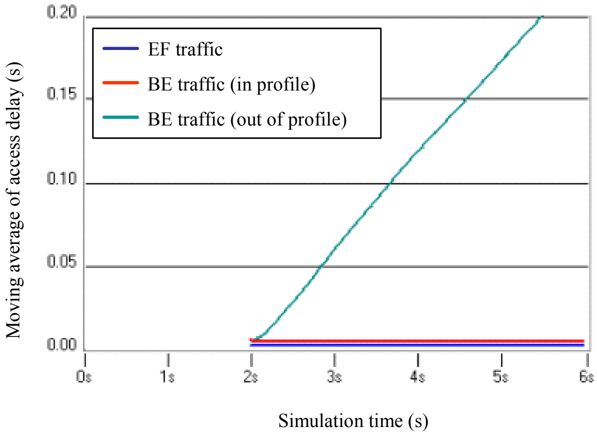

Fig. 6: Impact of rate quota on bandwidth sharing among BE queues

## 4 Conclusion

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 multiplex 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.

*Acknowldgement*

This research has been conducted within the framework of the “Archimedes: funding of research groups of TEI Piraeus” project co-funded by the Eurpean Union (75%) and the Greek ministry of education (25%).

*References:*

[1] J. D. Angelopoulos, et. Al. , “Efficient transport of packets with QoS in an FSAN-aligned GPON,” *IEEE Commun. Mag.*, vol. 42, issue 2, pp.92-98, Feb. 2004.

[2] IEEE Std 802.3ah-2004

[3] J. D. Angelopoulos, I.S. Venieris, G.I. Stassinopoulos, “A TDMA based Access Control Scheme for APON's,” *IEEE/OSA J. of Lightwave Technology*, vol. 11, No. 5/6, pp. 1095-1103, May/June 1993

[4] M. P. McGarry, M. Maier, and M. Reisslein, “Ethernet PONs: A Survey of Dynamic Bandwidth Allocation (DBA) Algorithms,” *IEEE Commun. Mag.*, vol. 42, no. 8, pp. S8–S15, Aug, 2004

[5] 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.*, vol. 40, pp. 74–80, Feb. 2002

[6] X. Bai, A. Shami, C Assi, “Statistical Bandwidth Multiplexing in Ethernet Passive Optical Networks,” in *Proc.Globecom 2005*, St. Luis, U.S.A., Nov. 2005

[7] H. Naser, Hussein T. Mouftah, “A Joint-ONU Interval-based Dynamic Scheduling Algorithm for Ethernet Passive Optical Networks,” *IEEE/ACM Trans. on Networking*, accepted for publication

[8] F. An et al., “A new dynamic bandwidth allocation protocol with quality of service in ethernet-based passive optical networks,” in *Proc. Int. Conf. on Wireless & Optical Comm.* (WOC03), Banff, Canada, Jul. 2003

[9] A. Shami et al., “Jitter Performance in Ethernet Passive Optical Networks,” *IEEE J. of Lightwave Technology*, vol. 23, No. 4, pp. 1745-1753, April 2005

[10] Data-Over-Cable Service Interface Specifications, Operations Support System Interface Specification, SP-OSSIv1.1-I02-000714, available at www.cablemodem.com