MPCP-ℓ: Look-Ahead Enhanced MPCP for EPON

Xiaomin Liu†, George N. Rouskas‡
†School of Electronic and Information Engineering, Beihang University, Beijing, 100191, China
‡Department of Computer Science, North Carolina State University, Raleigh, NC 27695-8206, USA

***Abstract*—We present a simple yet effective enhancement to the operation of the EPON multipoint control protocol (MPCP) that results in significant performance gains across the whole range of traffic loads. The enhancement, inspired by our earlier work in a related but different context, allows the OLT to perform look-ahead scheduling on the upstream channel. The look-ahead operation is fully compatible with the existing standard, and may be implemented via software updates to the OLT without affecting the operation of ONUs. In addition to improvements in delay and throughput performance, look-ahead enhanced MPCP also opens up new opportunities for the design of sophisticated DBA algorithms to support advanced quality of service (QoS) capabilities.**

I. Introduction

The growth in applications, services, and use of computer networking continues to increase at explosive rates, requiring network service providers to offer improved and faster Internet connectivity so as to keep up with user demand. Traditional access network architectures connecting subscribers to the central office (CO) represent a bottleneck as technology constraints place a limit in the data rates they can support. Therefore, optical networks using fiber communication have an important role to play in the access network. Passive optical networks (PONs) have long been considered attractive due to their longevity, low operational costs, and high capacity. As a matter of fact, PONs are already widely deployed in the first/last mile of today's operational access networks [1]. Various PON standards have been developed, including ATM-PON (APON), Broadband PON (BPON), Gigabit PON (GPON), and Ethernet PON (EPON).

In this work, we consider the EPON standard that represents the dominant trend of PON technology application in the access network [2]. EPON takes full advantage of the PON physical layer architecture in delivering high data rates, while also making use of Ethernet technology in the data link layer. Consequently, EPON makes it possible to achieve Ethernet-like economies of scale and provides simple, easy-to-manage connectivity to Ethernet-based, IP equipment, both at the customer premises and at the central office.

EPON is based on the point-to-multipoint architecture that is common to all PON technologies [3]. Specifically, EPON is deployed in a tree or tree-and-branch topology, that connects an optical line terminal (OLT) to multiple optical network units (ONUs), typically via a $1:N$ splitter and $N:1$ combiner in the downstream and upstream direction, respectively. The OLT, located at the CO, is the root of the tree topology. The ONUs, residing at or near the customer premises, are the leaves of the tree topology and connect to user equipment. Communication between the OLT and the ONUs is carried out in a different mode depending on the direction of the transmission. In the downstream direction, i.e., from the OLT to the ONUs, the EPON operates in a point-to-multipoint mode such that traffic from the OLT is broadcast to the ONUs. In the upstream direction, Ethernet packets from the ONUs to the OLT are time division multiplexed onto the single upstream wavelength. EPON is considered a shared media network in the upstream direction, and uses the multipoint control protocol (MPCP) to manage and coordinate access to the shared upstream channel [3].

In this paper, we present a simple yet effective enhancement to the operation of MPCP that results in significant decrease, up to 70%, of packet delay across the whole range of traffic loads. The enhancement, inspired by our earlier work in a related but different context [4], consists of a *look-ahead* operation that is managed by the OLT and allows the latter to coordinate access to the upstream channel in an efficient and effective manner. The look-ahead operation is fully compatible with, and requires no changes to, the standard; hence, it may be implemented via software updates to the OLT without affecting the operation of ONUs. With the proposed enhancement, the stable operation regime of EPON may be extended to high loads in a manner that benefits both network operators (who may offer higher-value bandwidth-demanding services) and subscribers (who may enjoy a better user experience).

The paper is organized as follows. In Section II, we discuss briefly the operation of MPCP at the MAC layer of EPON, and we review related work on dynamic bandwidth allocation (DBA) schemes. In Section III we introduce the look-ahead enhancement to MPCP, and in Section IV we present simulation results to demonstrate the performance benefits that can be achieved. We conclude the paper in Section V.

II. MPCP and Related Work

A. Overview of MPCP

MPCP, developed and standardized by the IEEE 802.3ah task force [5], is the protocol used to arbitrate the upstream transmission among the ONUs. MPCP does not dictate a specific dynamic bandwidth allocation (DBA) scheme, but it facilitates the implementation of DBA schemes by enabling the exchange of information that the OLT needs to allocate bandwidth to each ONU. MPCP introduces two 64-byte MAC control messages, `GATE` and `REPORT`. The OLT grants bandwidth to each ONU by sending a `GATE` message that informs

This work was supported in part by the National Science Foundation under Grant CNS-1113191.

the ONU of the start time and duration of its transmission on the upstream channel. Each ONU requests bandwidth by sending a GATE message to the OLT that reports the current size of its transmission buffer. The two messages also carry timestamps that make it possible to determine the round-trip time (RTT) between the OLT and each ONU; the OLT uses the RTT information to ensure that the transmission windows of different ONUs do not overlap in time.

Transmission in the upstream channel proceeds in rounds such that each ONU is allocated one transmission window within each round. Upon receiving the REPORT messages from ONUs, the OLT executes a DBA to calculate the bandwidth grants for the next round; the DBA is outside the scope of MPCP and can be used as a differentiating feature in the offerings of EPON providers. In determining the bandwidth grant (i.e., size and start time of the transmission window) for each ONU, the OLT also considers the RTT information and necessary guard times between successive windows assigned to different ONUs. Upon receipt of a GATE message from the OLT, an ONU uses the specified window to transmit an amount of data that does not exceed the size of the grant it received; it also updates its local time based on the timestamp carried in the GATE message to maintain synchronization with the OLT. If the amount of data buffered for transmission at the ONU is larger than the size of the grant, the ONU defers tranmission of the excess data for the next transmission window. Along with user data, each ONU also transmits a REPORT message to the OLT within its transmission window; this REPORT message contains updated information about the current size of the transmission buffer at the ONU.

B. Related Work

Since all ONUs share the capacity of the common channel in the upstream direction, the development of efficient DBA algorithms that avoid collisions and attempt to optimize the utilization of the shared bandwidth resource has been a main focus of EPON related research. For a comprehensive survey of the literature that reviews and classifies a wide range of DBA algorithms for EPONs, the reader is referred to [6]. In this section we only summarize schemes that are most relevant to our work.

It was recognized early on [7] that bandwidth allocation schemes based on TDMA or basic polling would not be effective in an access network based on EPON technology: TDMA performs poorly under bursty IP traffic, while polling leads to high delays due to the accumulation of walk times. The IPACT (interleaved polling with adaptive cycle time) algorithm is an early scheme that improves upon basic polling thus achieving high utilization [8], [9]. According to IPACT, the OLT uses GATE messages to poll the ONUs in a round-robin fashion and grant each ONU a transmission window that reflects its backlog (as reported in the corresponding REPORT message). Two key ideas underlie the operation of IPACT. First, unlike basic polling schemes, the OLT does not poll each ONU sequentially; rather, it pipelines the GATE messages such that the walk times overlap and the idle time on the upstream channel is reduced significantly. Second, since the bandwidth grants reflect the instantaneous queue loads at the respective ONUs, the length of each transmission round adapts to the aggregate load on the upstream channel, and the bandwidth is allocated according to the requirements of each ONU leading to effective statistical multiplexing.

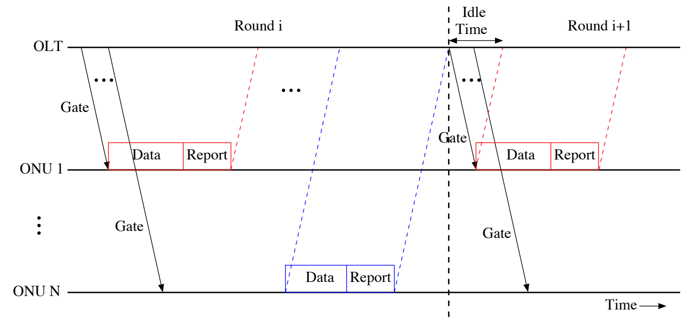

Fig. 1. The operation of the IPACT algorithm

The operation of IPACT is shown in Figure 1. As we can see, by pipelining the GATE messages, IPACT makes it possible to schedule the transmission windows of the various ONUs on the upstream channel so as to eliminate idle times (ignoring the guard bands between the transmission windows). The only idle time is between two consecutive rounds, as seen in the figure and as we discuss in more detail in the next section.

The basic IPACT algorithm has been extended and analyzed extensively in the literature. Variants of IPACT implementing different service disciplines can be classified as fixed service, gated service or limited service [10]–[13]. Analytical models to compute the mean packet delay and mean queue length in an IPACT system using mean value analysis (MVA) have been developed in [10], [11]. DBA schemes that estimate the amount of new packets arriving between two consecutive polling instants and grant ONUs a larger window size based on this estimate were proposed in [12], [13] and were shown to improve performance over the basic IPACT scheme. A polling protocol for EPONs called transmission upon reception (TUR) was introduced in [14]; the protocol ensures collision-free transmission while taking into account fairness considerations in allocating bandwidth. Several DBA variants that use quality of service (QoS) and fairness criteria have been proposed in [15]–[19]; these schemes aim to support differentiated services and applications with heterogeneous requirements while making efficient use of the network resource. For an in-depth discussion of DBA algorithms for EPONs please refer to [6].

III. Look-Ahead Enhanced MPCP

A. Motivation

Our objective is not to introduce a new DBA scheme, but rather to enhance the performance of the underlying MPCP protocol. The motivation for our work is based on the observation that most DBA schemes build upon the basic

IPACT scheme whose operation is shown in Figure 1. As the figure illustrates, there is an idle time on the upstream channel between two consecutive transmission rounds. This idle time is mandated by the fact that information carried by `REPORT` messages transmitted in round i is used to make bandwidth allocation decisions for round $i+1$. Specifically, regardless of the specific DBA employed, the OLT has to wait until it has received a `REPORT` message from each ONU in round i before it can finalize the bandwidth grants and send the first `GRANT` message in round $i+1$. For instance, the OLT has to ensure that the sum of bandwidth requests does not exceed a certain threshold on the length of a transmission round determined either by the specification or by desired bounds on, e.g., packet delay.

Let T_{idle} be the idle time between two consecutive rounds, RTT_{min} be the smallest RTT between the OLT and any of the ONUs, T_{proc}^{OLT} be the time required by the OLT to process the `REPORT` messages and execute the DBA, and T_{proc}^{ONU} denote the time required by the ONU to process the `GATE` message. Then, we have that:

$$T_{idle} \geq RTT_{min} + T_{proc}^{OLT} + T_{proc}^{ONU} \tag{1}$$

The idle time T_{idle} increases packet delay and reduces the utilization of the upstream channel: if R is the average length of a transmission round, then channel utilization is $R/(R+T_{idle})$, not accounting for guard bands or other overhead that is independent of MPCP.

Next, we introduce a new look-ahead operation for MPCP that completely eliminates the idle time T_{idle} and, hence, will improve the performance of any DBA scheme that is based on the basic IPACT.

B. MPCP-ℓ: MPCP with Look-Ahead

We define MPCP-ℓ, an enhancement of MPCP that implements a look-ahead operation with parameter ℓ as follows:

Definition 3.1 (MPCP-ℓ): The MPCP protocol configured such that queue length information carried by `REPORT` messages transmitted in round i is used by the OLT to allocate bandwidth in round $i+\ell$.

Clearly, when the look-ahead parameter $l=1$, MPCP-1 is equivalent to the basic MPCP protocol.

Figure 2 illustrates the operation of MPCP-2, i.e., when the look-ahead parameter $\ell=2$. Queue length information carried in `REPORT` messages in round i is used by the OLT at the end of the round to execute a DBA method and allocate bandwidth for round $i+2$. We make the reasonable assumption that DBA processing takes time less than the time for the ONUs to complete their transmissions in round $i+1$; if that is not the case, we can increase the look-ahead value, as we discuss shortly. Therefore, during round $i+1$, the OLT can start transmitting the `GATE` messages to inform ONUs of their transmission windows in round $i+2$, as shown in Figure 2. As long as the first such `GATE` message reaches the ONU before the end of round $i+1$, the idle time between rounds is eliminated, ensuring continuous transmission on the upstream channel (ignoring, of course, guard bands or other

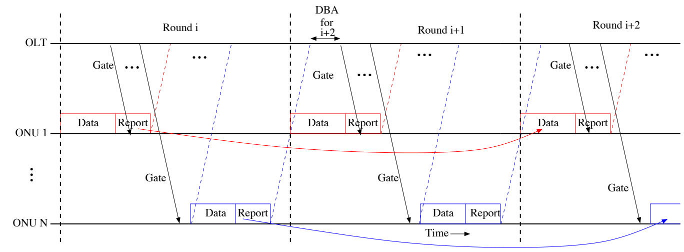

Fig. 2. The operation of MPCP-2 with look-ahead parameter $\ell=2$

gaps between frames mandated by Ethernet). In essence, the look-ahead operation masks the three components of the idle time in the right hand side of expression (1), namely minimum RTT between the first `GATE` message and receipt of first bit of data, processing time for DBA, and processing time at ONU, by overlapping them with data transmission during a round. As a result, this small change in the operation of MPCP completely removes the delays associated with the DBA and `GATE` messages.

This operation can be readily generalized to look-ahead values $\ell > 2$. Larger look-ahead values would be needed if the idle time from expression (1) is larger than the average length of a transmission round, e.g., due to long round-trip times between the OLT and ONUs and/or the processing requirements of the DBA. In this case, it may be necessary to use more than one round to completely mask the idle time.

The look-ahead feature of MPCP-ℓ, $\ell \geq 2$, improves the delay and throughput performance of the protocol by achieving better utilization of the upstream channel, as the numerical results we present in the next section indicate. The look-ahead operation affords further benefits. First, since the delay introduced by the RTT is masked, the allowable distance between the OLT and ONUs is limited only by transmission impairments, not performance issues due to the MPCP protocol. Second, whereas the execution time of DBA increases the idle time of MPCP, it does not affect the channel utilization of MPCP-ℓ, $\ell \geq 2$. Consequently, network operators may implement sophisticated bandwidth allocation algorithms that would not otherwise be possible to implement in basic MPCP due to running time constraints.

C. Look-Ahead Implementation Considerations

Implementing the look-ahead feature does not require any changes to the `GATE` messages. For `REPORT` messages, two relatively minor modifications are required at the OLT only, without any change to how ONUs operate. Specifically, the OLT uses the information in the `REPORT` messages to allocate bandwith not in the next transmission round, as with basic MPCP, but in a future round determined by the value of the look-ahead parameter ℓ. Second, the OLT must be careful in how to interpret the queue length information reported by the ONUs, since the latter record instantaneous queue lengths in their report. For instance, consider $\ell=2$ as shown in Figure 2. When, say, ONU 1 reports its queue length in round i, it includes all the packets currently in its queue. However, the

OLT has already allocated bandwidth to ONU 1 in a GATE message that has not been received or processed by the ONU. Therefore, the OLT must subtract this allocated bandwidth from the queue length information carried in the REPORT message from ONU 1 before it uses it to allocate bandwidth for round $i+2$. Finally, to jumpstart the look-ahead operation, for any value of ℓ, the OLT must initally transmit ℓ rounds of GATE messages that make bandwidth grants only large enough for the ONUs to transmit REPORT messages; from then on, the bandwidth grants will be determined from queue length information from ℓ rounds in the past.

IV. Numerical Results

A. System Model

For the simulation results we present in this section, we consider an EPON with one OLT and N ONUs, $N = 16, 32$. The distance between the OLT and the ONUs is within the range of $2-5$ Km. All nodes (OLT and ONU) employ one transmitter and one receiver. The upstream and downstream channels each operate at 1 Gbps and are on different wavelengths. The GATE and REPORT messages each have a length of 64 bytes. The length of each transmission round can be no larger than 2 ms, and we assume a 5 μs guard time between successive transmission windows from different ONUs. The buffer size of each ONU is limited to 10 KB.

Packet lengths L (in bytes) at each ONU are generated from a tri-modal distribution that is meant to reflect the distribution of packet lengths that has been observed in the Internet [20]:

$$P[L=x] = \begin{cases} 0.4, & x = 40 \\ 0.2, & 41 \le x \le 1449 \\ 0.4, & x = 1500 \end{cases} \tag{2}$$

Consequently, the average packet length is 770 B. We consider two types of traffic distribution, uniform and hot-spot. With uniform distribution, each ONU generates the same amount of traffic, whereas under the hot-spot distribution 25% (respectively, 75%) of the ONUs generate 80% (respectively, 20%) of the traffic.

Recall that the focus of our work is on the look-ahead enhancement to MPCP, not bandwidth allocation. Therefore, in our simulations we use a simple strategy that allocates each ONU a bandwidth grant sufficient to satisfy the corresponding request, as long as the length of the transmission round is no larger than 2 ms. If the sum of bandwidth requests for a given round exceeds 2 ms, then the OLT scales down all the requests by a constant factor so that their sum does not exceed 2 ms, and grants the corresponding amount to each ONU.

Figures 3-5 present the results of OPNET simulations comparing the performance of three protocols, MPCP-ℓ, $\ell = 1, 2, 3$, where MPCP-1 is equivalent to the original MPCP protocol and the other two implement two versions of look-ahead scheduling. These figures present results for a 16-node EPON; results for the 32-node network are very similar and are omitted due to page constraints. Specifically, Figures 3-4 plot the average packet delay as a function of the traffic load for the 16-node network under uniform and hot-spot traffic, respectively. The main observation from the figures is that the look-ahead operation reduces the delay considerably compared to the original MPCP, across the whole range of traffic loads. Despite the fact that all packets are kept in the queue for an additional amount of time equal to 1 or 2 transmission rounds under MPCP-2 and MPCP-3, respectively, look-ahead eliminates the idle time between successive rounds resulting in lower delay per round such that in steady state the average packet delay is significantly lower. We also note that the delay is higher for MPCP-3 than for MPCP-2. This result is due to the fact that, for the system parameters we used in the simulation scenarios, the idle time is smaller than the length of a transmission round, hence the extra delay that packets incur under MPCP-3 does not offer any extra benefit. As we mentioned earlier, a look-ahead parameter $\ell > 2$ would be of value when two or more transmission rounds are necessary to completely mask the idle time, i.e., for networks with long RTTs or when the DBA is computationally expensive. Another interesting observation from these figures is that delays under the hot-spot traffic scenario are lower than those under the uniform traffic scenario. This behavior can be explained by the fact that in the simulations we have the OLT schedule ONUs with large demands early in the transmission round. Since four out of sixteen ONUs generate 80% of the total traffic under the hot-spot scenario, scheduling this traffic early reduces the overall average delay.

Finally, Figure 5 plots the aggregate throughput on the upstream channel as a function of traffic load. As we can see, the throughput increases almost linearly until the load reaches 60%. After that point, the overhead due to idle time on the operation of the original MPCP is evident in the fact that the corresponding curve increases more slowly. On the other hand, the look-ahead feature also improves the throughput performance of the protocol, and, especially for $\ell = 2$, the throughput curve increases almost linearly until the load reaches 80%.

The results indicate that a relatively small modification to the MPCP protocol that allows the OLT to perform look-ahead scheduling of ONU requests for bandwidth so as to mask the idle time between transmission rounds. This look-ahead feature can be implemented via a software update to the OLT and can be very effective in (1) lowering the average packet delay, and (2) allowing the upstream channel to operate at high loads without a significant decrease in the traffic carrying capacity.

V. Concluding Remarks

In this paper, we have presented MPCP-ℓ, a variant of the MPCP protocol for EPON that allows for look-ahead scheduling of the upstream channel. Although this work focused on demonstrating the benefits of the look-ahead operation using a simple bandwidth allocation algorithm, the new look-ahead feature makes it possible to design new, sophisticated DBA schemes that can take advantage of the additional information at the OLT to support advanced QoS capabilities; this is an area of ongoing research in our group. Another direction of research is to design a look-ahead variant for next-generation, multiwavelength Ethernet PONs.

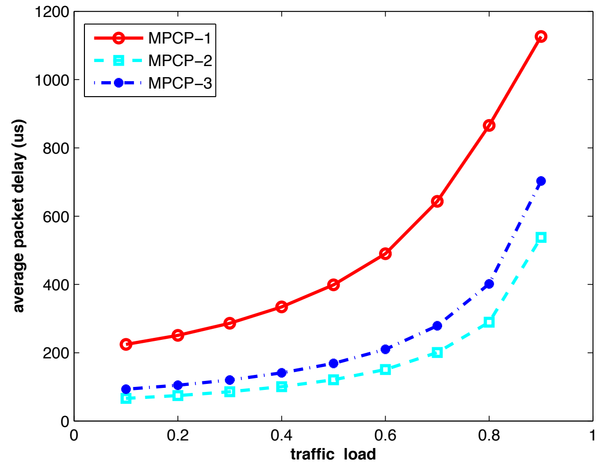

Fig. 3. Packet delay vs. traffic load, $N = 16$ and uniform traffic

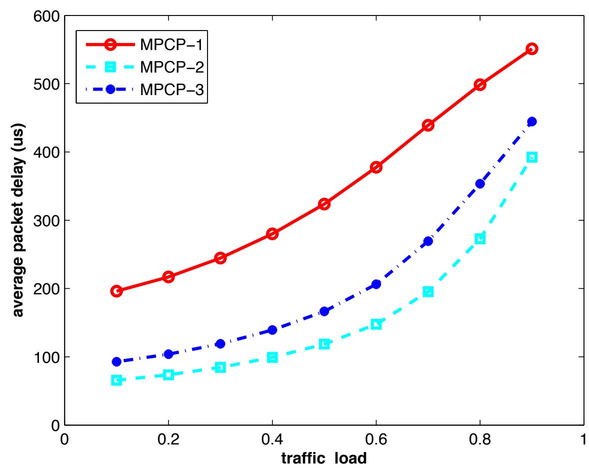

Fig. 4. Packet delay vs. traffic load, $N = 16$ and hot-spot traffic

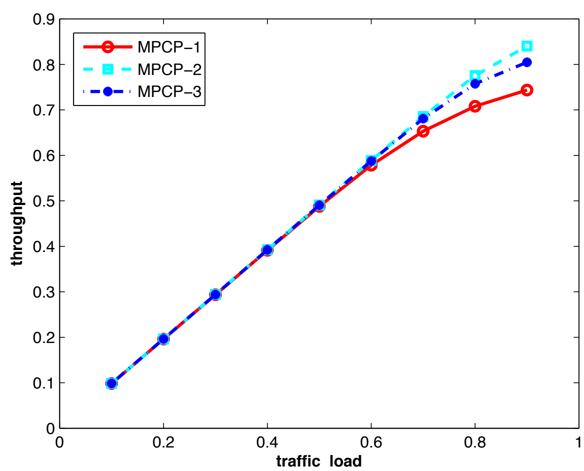

Fig. 5. Throughput vs. traffic load, $N = 16$ and hot-spot traffic

REFERENCES

[1] M. McGarry, M. Reisslein, and M. Maier, “WDM ethernet passive optical networks,” *Communications Magazine, IEEE*, vol. 44, no. 2, pp. 15–22, 2006.

[2] G. Kramer and G. Pesavento, “Ethernet passive optical network (epon): Building a next-generation optical access network,” *Communications magazine, IEEE*, vol. 40, no. 2, pp. 66–73, 2002.

[3] G. Pesavento, “Ethernet passive optical network (EPON)’,” *IEEE Optical Network Magazine*, vol. 4, no. 1, pp. 107–113, 2003.

[4] V. Sivaraman and G. N. Rouskas, “A reservation protocol for broadcast WDM networks and stability analysis,” *Computer Networks*, vol. 32, pp. 211–227, February 2000.

[5] G. Keiser and J. Wiley, *FTTX concepts and applications*. Wiley Online Library, 2006.

[6] M. McGarry, M. Reisslein, and M. Maier, “Ethernet passive optical network architectures and dynamic bandwidth allocation algorithms,” *Communications Surveys & Tutorials, IEEE*, vol. 10, no. 3, pp. 46–60, 2008.

[7] B. Mukherjee, *Optical WDM networks*. Springer-Verlag New York Inc, 2006.

[8] G. Kramer, B. Mukherjee, and G. Pesavento, “IPACT a dynamic protocol for an ethernet pon (EPON),” *Communications Magazine, IEEE*, vol. 40, no. 2, pp. 74–80, 2002.

[9] G. Kramer, B. Mukherjee, and G. Pesavento, “Interleaved polling with adaptive cycle time (IPACT): a dynamic bandwidth distribution scheme in an optical access network,” *Photonic Network Communications*, vol. 4, no. 1, pp. 89–107, 2002.

[10] B. Lannoo, L. Verslegers, D. Colle, M. Pickavet, M. Gagnaire, and P. Demeester, “Analytical model for the IPACT dynamic bandwidth allocation algorithm for EPONs,” *Journal of optical networking*, vol. 6, no. 6, pp. 677–688, 2007.

[11] M. Thanh Ngo, A. Gravey, and D. Bhadauria, “A mean value analysis approach for evaluating the performance of EPON with gated IPACT,” in *Optical Network Design and Modeling, 2008. ONDM 2008. International Conference on*, pp. 1–6, IEEE, 2008.

[12] Y. Zhu and M. Ma, “IPACT with grant estimation (IPACT-GE) scheme for ethernet passive optical networks,” *Journal of Lightwave Technology*, vol. 26, no. 14, pp. 2055–2063, 2008.

[13] H. Byun, J. Nho, and J. Lim, “Dynamic bandwidth allocation algorithm in ethernet passive optical networks,” *Electronics Letters*, vol. 39, no. 13, pp. 1001–1002, 2003.

[14] A. Gumaste and I. Chlamtac, “A protocol to implement ethernet over pon,” in *Communications, 2003. ICC’03. IEEE International Conference on*, vol. 2, pp. 1345–1349, IEEE, 2003.

[15] Y. Luo and N. Ansari, “Bandwidth allocation for multi-service access on EPONs,” *Communication Magazine*, vol. 43, no. 2, pp. 516–521, 2005.

[16] C. Assi, Y. Ye, S. Dixit, and M. Ali, “Dynamic bandwidth allocation for quality-of-service over ethernet pons,” *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 9, pp. 1467–1477, 2003.

[17] J. Xie, S. Jiang, and Y. Jiang, “A dynamic bandwidth allocation scheme for differentiated services in epons,” *Communications Magazine, IEEE*, vol. 42, no. 8, pp. S32–S39, 2004.

[18] R. Salles and J. Barria, “Fair and efficient dynamic bandwidth allocation for multi-application networks,” *Computer Networks*, vol. 49, no. 6, pp. 856–877, 2005.

[19] J. Zheng, “Efficient bandwidth allocation algorithm for ethernet passive optical networks,” *Communications, IEE Proceedings-*, vol. 153, no. 3, pp. 464–468, 2006.

[20] R. Sinha, C. Papadopoulos, and J. Heidemann, “Internet packet size distributions: Some observations.” http://netweb.usc.edu/˜rsinha/pkt-sizes/, October 2005.