An Analytical Model of the IEEE 802.3ah MAC Protocol for EPON-based Access Systems

Un modèle analytique du protocole MAC des arbres optiques passifs Ethernet

Sawsan AL Zahr
Maurice Gagnaire

2006C001

mai 2006

Département Informatique et Réseaux
Groupe Mathématiques de l'Information, des Communications et du Calcul

Dépôt légal : 2006 – 2ème trimestre
Imprimé à l’Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 0751-1337 ENST C (Paris) (France 1983-9999)

Ecole Nationale Supérieure des Télécommunications

Groupe des Ecoles des Télécommunications - membre de ParisTech

46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.enst.fr

Département INFRES

Titre

Un modèle analytique du protocole MAC des arbres optiques passifs Ethernet

Auteurs

Sawsan AL ZAHR, Maurice GAGNAIRE

GET/Télécom Paris (ENST) - LTCI-UMR 5141 CNRS
Département INFRES
46, rue Barrault
75634 PARIS CEDEX 13

Résumé

Au cours des dernières années, l'accès à haut débit est devenu un enjeu majeur. Cependant, une tendance lourde est l'émergence des arbres optiques passifs Ethernet (*Ethernet over Passive Optical Networks* ou EPON) en concurrence d'ADSL ce qui permet de fournir des services Ethernet évolutifs, fiables et prédictibles. Cette nouvelle technologie représente la convergence des équipements peu coûteux d'Ethernet et l'infrastructure peu coûteuse de la fibre. Le protocole MPCP (*Multi-Point Control Protocol*) a été standardisé par « Ethernet First Mile Alliance » pour gérer l'accès multiple sur les arbres optiques passifs Ethernet. Par ailleurs, plusieurs algorithmes d'allocation de ressources ont été proposés dans la littérature pour gérer, en combinaison avec le protocole MPCP, l'accès multiple au système de transmission. Dans ce rapport, nous étudions en particulier les performances de l'algorithme IPACT (*Interleaved Polling with Adaptive Cycle Time*) qui définit un schéma dynamique d'allocation de ressources. Nous proposons un modèle analytique basé sur la théorie des réseaux des files d'attente. Ce modèle permet de calculer les paramètres de performances des systèmes EPON sous le schéma d'allocation de ressources défini par IPACT.

Mots-clés : *Réseaux EPON, Protocole MAC, Algorithme d'allocation dynamique de ressources, IPACT, Qualité de Service, Réseaux des files d'attente.*

Cet article a été soumis à *LANMAN'05*

An Analytical Model of the IEEE 802.3ah MAC Protocol for EPON-based Access Systems

Sawsan Al Zahr and Maurice Gagnaire
Department of Computer Sciences and Networks
Telecom Paris - LTCI - UMR 5141 CNRS
46, rue Barrault F 75634 Paris - France
Email: {sawsan.alzahr,maurice.gagnaire}@enst.fr

***Abstract*— Ethernet Passive Optical Networks (EPON) access systems are considered as an alternative to ADSL for high speed access to the Internet. Supported by the Ethernet First Mile Alliance and the IEEE, the MPCP protocol has been adopted as the standardized MAC protocol for EPONs. Moreover, the IPACT dynamic bandwidth allocation mechanism has been proposed for hierarchical multiplexing. Today, IPACT is considered as interesting complement to MPCP for dynamic bandwidth allocation over EPONs. In this paper, we propose an original analytical model of the MPCP+IPACT protocol. Numerical results obtained from this model are commented and compared to computer simulations.**

***Index Terms*— Ethernet over Passive Optical Network (EPON), Multipoint Control Protocol (MPCP), Dynamic Bandwidth Allocation (DBA), Interleaved Polling with Adaptive Cycle Time (IPACT), Quality-of-Service (QoS), Queueing Networks.**

I. Introduction

Many investigations have been carried these last fifteen years to evaluate the feasibility of fiber-in-the-loop (FITL) access systems. During the years 1990, several major telecommunication carriers associated in the FSAN (Full Service Access Network) initiative have specified the concept of PON (Passive Optical Network). A PON is a passive point-to-multipoint optical access system. Two types of modems are used in a PON: the Optical Line Termination (OLT) at the head of the optical tree, and the Optical Network Units (ONUs) at the customer premises or close to these premises. For economy purposes, only two optical wavelengths are used in the system. A common upstream wavelength at 1300 nm is used at the ONUs whereas the OLT uses another wavelength at 1550 nm for downstream traffic. Standardized by the ITU-T (G.983), such systems are designed for the transport of ATM connections. The principle of APONs is now extended for Wavelength Division Multiplexing (WDM) and protection and restoration with the BPONs (Broadband PON) and GPONs (Gigabit PON) concepts. With the emergence of Ethernet switching in the metropolitan area since the year 2000, the IEEE and the Ethernet First Mile Alliance are promoting the concept of Ethernet PON (EPON) also known as the IEEE 802.3ah standard. Physical layer functionalities are roughly the same in APONs and EPONs. Both APONs and EPONs require a Medium Access Control (MAC) protocol in order to prevent collisions when MAC-PDUs are transmitted from the ONUs. The MPCP MAC protocol has been developed by the IEEE 802.3ah Task Force [1]. It is based on the concept of non-overlapping upstream transmission windows (timeslots) allocated to each ONU by the OLT. In addition to MPCP, multiple Dynamic Bandwidth Allocation (DBA) schemes have been proposed these last three years as complement of MPCP [2], [**?**]. These DBA schemes aim either to efficient statistical multiplexing or to Quality of Service (QoS) provisioning in EPON systems [3]. In this paper, we propose the first (at our knowledge) analytical model of the MPCP+IPACT protocol. In section II, we recall the main characteristics of the IEEE 802.3ah MAC protocol and of the IPACT dynamic bandwidth allocation scheme. In section III, we describe our analytical model of MPCP+IPACT. Finally, we present and comment in section IV numerical results obtained by means of this analytical model.

II. MAC Protocol and Bandwidth Allocation

In this section, we describe successively the principle of the IEEE 802.3ah MAC protocol and of the IPACT DBA mechanism.

A. *Mutipoint Control Protocol (MPCP) Overview*

Although MPCP is not concerned with any particular bandwidth allocation (QoS dependant), it has been designed in order to facilitate the implementation of various DBA algorithms. MPCP relies on two Ethernet control messages (GATE and REPORT) to provide the signalling infrastructure (control plane) for coordinating upstream data transmission. Three other signalling messages are used for automatic ONUs discovery (REGISTER_REQUEST, REGISTER, and REGISTER_ACK). Such messages enable for instance distance equalization between the OLT and the various ONUs.

In its regular operation, the OLT generates downstream GATE messages to dynamically assign timeslots to the active ONUs. A GATE message contains the time when an ONU is allowed to start its transmission and the length of this transmission window (timeslot). It is the role of each ONU to coordinate access to the medium during its allocated timeslot among its different active traffic queues. In order to facilitate this coordination, MPCP stamps each GET message when it arrives at an ONU with a local clock. Upon receiving a message matching its MAC address, each ONU updates its local clock to that of the timestamp in the received GATE control message to avoid any potential clock drift. A transmission window may include multiple variable size Ethernet frames. Depending on the size of the allocated timeslot and the number of buffered packets (Ethernet frames) at the ONU, a REPORT message followed by upstream user data frames is sent in each allocated timeslot. Typically, a REPORT message contains the required size of the next timeslot based on the ONUs buffer occupancy. Upon receiving REPORT messages, the OLT passes the bandwidth requests of the various ONUs to the DBA module responsible for bandwidth allocation decision.

B. Interleaved Polling with Adaptive Cycle Time (IPACT)

Interleaved Polling with Adapting Cycle Time (IPACT), proposed in [4], provides a statistical multiplexing for ONUs and results in efficient upstream channel utilization. IPACT consists in a pipelining process using an interleaved polling strategy. For instance, ONU$_i$ is polled during ONU$_{i-1}$ is sending its user data to the OLT. In order to facilitate this pipelining, an ONU may generate its own control message which is piggybacked to user data transmission. Thus, a given ONU informs the OLT by means of a piggybacked control message how many bytes were in its buffer at the instant of its last user data transmission. In doing so, bandwidth is dynamically assigned to ONUs according to their buffer occupancy. An ONUs which is transitory not allowed to send data during a given transmission cycle is still polled by the OLT. This ONU is then able to report its queues occupancy for the next cycle. By convention, the transmission of a user frame that cannot fit in totality in a timeslot is postponed to the next timeslot. To prevent the upstream channel being monopolized by a single ONU with high data volume, a maximum transmission window size is assigned to each ONU.

We denote the specific maximum transmission window size of ONU$_i$ by $W_{MAX}^{[i]}$ (in bytes). The choice of the specific values of $W_{MAX}^{[i]}$, assigned to the various ONUs, determines the maximum polling cycle time T_{MAX} under heavy load conditions:

$$T_{MAX} = \sum_{i=1}^{N} \left(G + \frac{8 \times W_{MAX}^{[i]}}{R_U} \right) \tag{1}$$

where G is the guard interval (seconds), N is the number of ONUs, and R_U is the line rate (bps). The guard intervals provide protection for fluctuations of round-trip time (RTT) of different ONUs (distance ranging). Furthermore, the OLT receiver needs some time to proceed to two physical layer operations. First, the OLT needs to readjust its optical power detection thresholds according to its distance from the transmitting ONU (power ranging). Second, a clock resynchronization of the PLL (Phase Locked Loop) must be carried at the OLT at the beginning of each timeslot (clock recovery). The choice of the $W_{MAX}^{[i]}$ values will also determines the guaranteed bandwidth available for each ONU$_i$. We denote the minimum guaranteed bandwidth of ONU$_i$ by $\Lambda_{MIN}^{[i]}$ (bps). Obviously, the ONU is guaranteed to be able to send at least $W_{MAX}^{[i]}$ (bytes) in at most T_{MAX} (seconds):

$$\Lambda_{MIN}^{[i]} = \frac{8 \times W_{MAX}^{[i]}}{T_{MAX}} \tag{2}$$

An ONUs bandwidth is limited to its guaranteed bandwidth only if all other ONUs in the system also use all of their available bandwidth. If at least one ONU does not consume its guaranteed bandwidth, it is assigned a shorter transmission window, thus making the polling cycle time shorter. Therefore the available bandwidth to all other ONUs is in this case increased proportionally to their $W_{MAX}^{[i]}$. As a consequence, the polling cycle time is not static but is adapted to the instantaneous network load.

III. Analitical Model

A. Bandwidth management

Bandwidth management and fair scheduling of different traffic classes play very important role in supporting QoS in EPON-based access network. Diffserv [5], developed by Internet Engineering Task Force (IETF), provides method to classify the network traffic.

In our model, we classify network traffic into three priorities as defined in [5]: the expedited forwarding (EF), the assured forwarding (AF), and the best effort (BE). EF services gather the delay sensitive applications that require a bounded end-to-end delay and jitter specifications (such as voice over IP), whereas AF class is intended for services that are not delay sensitive but which require bandwidth guarantees. Finally, BE services are not delay sensitive and do not require neither jitter specifications nor minimum guaranteed bandwidth. As shown

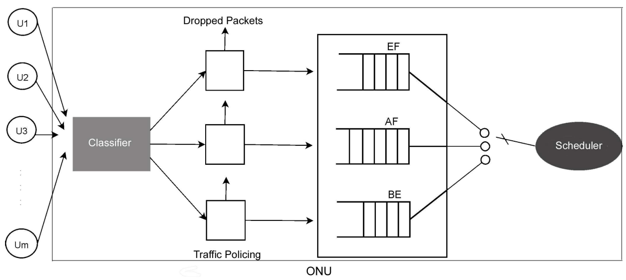

Fig. 1. ONU queue management

in Fig. 1, each ONU is provided with buffering space shared by three separate priority queues that are corresponding to service classes. Packets are first classified according to their type of service and then placed into their appropriate priority queues. Traffic policing is required at the ONU to ensure that packets conform to their service level agreement (SLA), non-conforming traffic being dropped. In the following, we assume that submitting traffic is conforming. As a result, we may neglect the traffic policing.

Packets transmission scheduling is performed by a priority-based scheduler. Strict priority scheduling algorithm, also known as Head-of-Line (HoL), schedules packets from the head of given queue only if all higher priority queues are empty. The HoL policy may then be very unfair for the lowest priority traffic. In order to prevent this drawback, we propose a priority-based scheduling mechanism which takes into account the traffic load in addition to its priority level.

B. ONU Model

Our ONU model is shown in Fig. 2. It consists of two separate stages. The first stage is composed of three separate priority queues as described previously, while the second stage contains only one common queue. Packets are scheduled in the second stage according to generalized processor sharing (GPS) strategy.

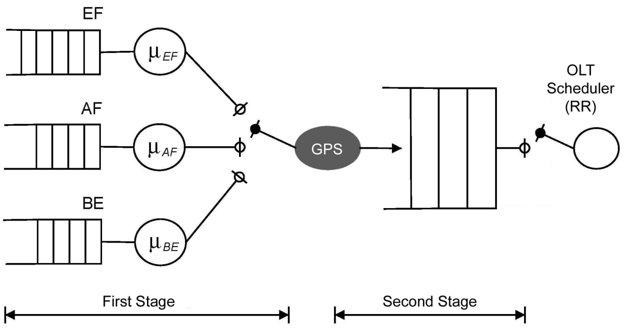

Fig. 2. ONU model

Each queue has a service share φ_c. In order to fairly distribute allocated bandwidth, we define the class service share considering its traffic load:

$$\varphi_c = \frac{\lambda_c}{\lambda} \times \delta_c \tag{3}$$

where λ_c is the class c traffic load, λ is the global traffic load (i.e. $\lambda = \sum_c \lambda_c$), and δ_c is the weight assigned to the class c based on its priority, with $\sum_c \delta_c = 1$. We notice that λ_c does not depend on index i, i.e. multi-priority traffic distribution being identical over the different ONUs. GPS strategy splits allocated bandwidth among all non-empty queues simultaneously, in proportion to their service shares as follows:

$$\mu_c^{[i]} = \frac{\varphi_c}{\sum_{c \in Q} \varphi_c} \Lambda_{MIN}^{[i]} \tag{4}$$

where Q denotes the set of non-empty queues.

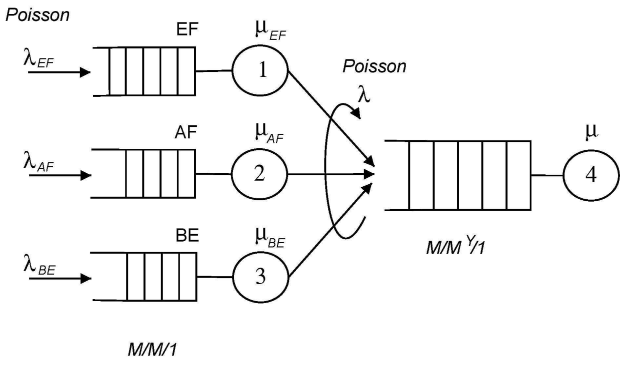

Fig. 3. Multi-class Open queueing network

We assume that new packets arrive at the system according to a Poisson process with rate λ_c and have an exponential service time distribution with mean $1/\mu_c^{[i]}$, where c corresponds to traffic class. Since we are mainly interested in the average access delay and not in packet loss, we can suppose that ONU buffer is infinite. Thus, each queue of the first stage can be modeled as a $M/M/1$ queue which has been analyzed extensively in the literature [6], [7]. For $M/M/1$ queue at equilibrium, packets departure process will also be Poisson, as well as combining independent Poisson processes leads to a process which will also be Poisson in nature [6]. Furthermore, the average flow rate leaving the $M/M/1$ queue at equilibrium is the same as the average flow rate entering this queue. Therefore, packets arrival at the second stage is a Poisson process with rate λ. Assume that service time is exponentially distributed with parameter $\mu^{[i]} = \Lambda_{MIN}^{[i]}$ and that up to $W_{MAX}^{[i]}$ packets are served simultaneously in considering packets transmission delay is negligible compared to queueing delay. Hence, the second stage can be modeled as a $M/M^Y/1$ queue, which has been discussed in [7]. Fig. 3 shows the obtained queueing system which is referred in queueing theory to multi-class open queueing network.

C. Performance evaluation

In our study, we consider a multi-class open queueing network which assumes the following hypothesis: (1) it contains only mono-server stations that operate under FIFO discipline; (2) service time is exponentially distributed; (3) packets arrival is a Poisson process. Thus, this network is equivalent to a mono-class open queueing network, also called open Jackson's network. As a result, the state probability can be expressed as follows:

$$p(n_1, n_2, n_3, n_4) = \prod_{i=1}^{4} p_i(n_i) \tag{5}$$

where $p_i(n_i)$ are the marginal probabilities associated to the EF, AF, BE, and $M/M^Y/1$ queues respectively. The Markov chain associated to the three queues of first stage ($i = 1$, $i = 2$, and $i = 3$) is illustrated by figure 4. Similarly, the Markov chain associated to the second stage is illustrated by figure 5. As mentioned previously, we are interested in the average access delay $\mathrm{E}[T]$. In order to determine the analytical expression of $\mathrm{E}[T]$, we need other performance parameters, namely the total network throughput γ and the expected number of packets $\mathrm{E}[N]$, where N stands for the random variable associated to the number of packets in the system.

All performance parameters are computed at equilibrium with regard to queueing network stability. System stability condition is defined as following:

$$\rho_i < 1 \tag{6}$$

where $\rho_i = \lambda_i/\mu_i$ and i refers the queue. Given that queues buffer are infinite and all queues are in stable state, then the total queueing system throughput is:

$$\gamma = \lambda = \sum_c \lambda_c \tag{7}$$

The expected total number of packets in the queueing system is given by:

$$\mathrm{E}[N] = \sum_{i=1}^{4} \mathrm{E}[N_i] \tag{8}$$

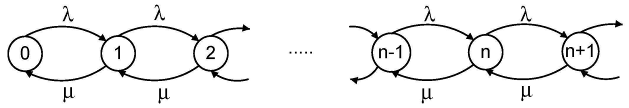

Fig. 4. Markov chain associated to $M/M/1$ queue

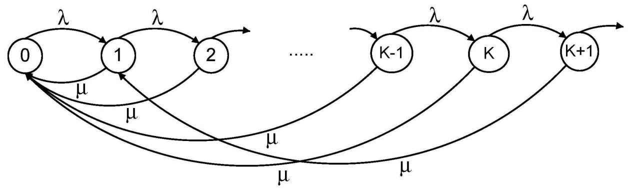

Fig. 5. Markov chain associated to $M/M^Y/1$ queue

where $\mathrm{E}[N_i]$ is the expected number of packets at queue i. We admitted that all first stage queues are $M/M/1$ queues then we can compute their expected number of packets as following:

$$\mathrm{E}[N_i] = \frac{\rho_i}{1-\rho_i}, \quad i \in \{1,2,3\} \tag{9}$$

Moreover, the second stage which has been modeled with a batch services queue ($M/M^Y/1$) contains a mean number of packets given by:

$$\mathrm{E}[N_4] = \frac{r_0}{1-r_0} \tag{10}$$

where r_0 is the single root of Eq. 11 which satisfies $|r_0| < 1$, obtained by means of Rouche's theorem [7].

$$[\mu p^{K+1}(n+1) - (\lambda+\mu)p(n+1) + \lambda p(n)] = 0 \tag{11}$$

Finally, we can compute the average access delay using Little's formula:

$$\mathrm{E}[T] = \frac{\mathrm{E}[N]}{\gamma} = \frac{r_0}{\lambda(1-r_0)} + \frac{1}{\lambda}\sum_c \frac{\rho_c}{(1-\rho_c)} \tag{12}$$

IV. Numerical Results

In this section, we study the impact of priority queueing on the overall performance of the network and we compare our analytical results with simulation results presented in [8]. We consider a PON access system with 16 ONUs and all of them have the same SLA. The upstream channel capacity R_u is equal to 1 Gbps. The maximum cycle time T_{MAX} is set to 2 ms and the guard time G separating two consecutive transmission windows is set to 5 μs. The Ethernet frame size is fixed at 1500 bytes. Using both of Eq. 1, and Eq. 2 and taking into account the fact that all the ONUs have the same SLA, we obtain the maximum size of transmission windows and the minimum guaranteed bandwidth: $W_{MAX} = 15000$ bytes, $\Lambda_{MIN} = 60$ Mbps.

The following results include the average packet delay and the average queue length. Each of these parameters has been evaluated numerically from Eq. 9, Eq. 10, and Eq. 12 by means of the formal mathematics Maple VIII software. Fig. 6 shows the relationship between average packet delay (Eq. 12) and network traffic load (Eq. 7). The average packet delay is defined as the average time elapsed between the instant of generation of a user packet and the instant of transmission of the last bit of this packet.

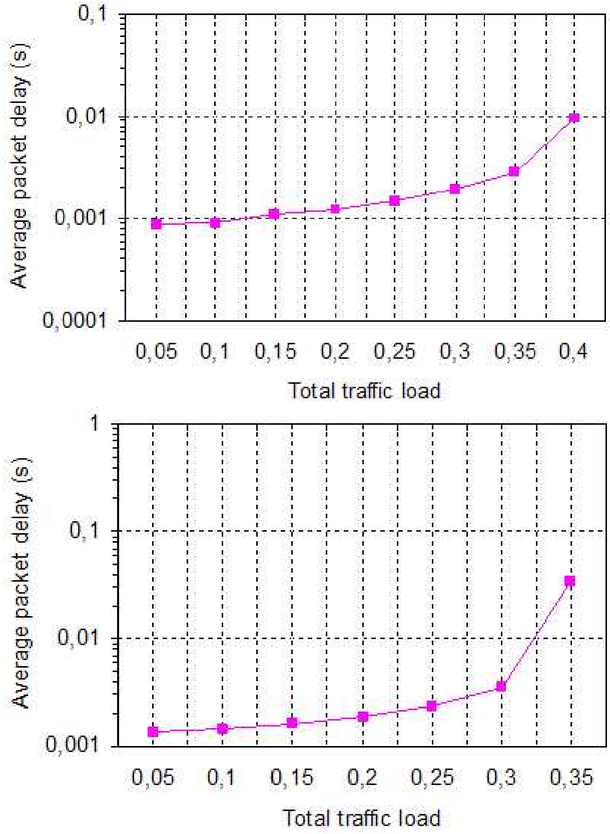

Fig. 6. Average packet delay of all packets

In Fig. 6.a (upper figure), 20 percent, 30 percent, and 50 percent of the global offered load are EF, AF, and BE traffic, respectively. To prevent BE queue instability, we have observed numerically that the total traffic load must remain under 0.4. Hence, the offered load fluctuates from 0.05 to 0.4. Under light traffic load, the results shown in Fig. 6.a can be compared to those obtained in [8]. With an offered load of 0.4, average packet delay increases to almost 10 ms. The reason is that the arriving packets rate of BE traffic reached the service rate (i.e., $\rho_{BE} \simeq 1$).

Unlike Fig. 6.a, one assumes in figure 6.b (lower figure) that the total offered load is uniformly distributed between EF, AF, and BE traffic classes (33 percent, 33 percent, and 33 percent, respectively). The longest average delay, observed at a load of 0.35, is due to the BE queue which reaches its stability limit. Figure 7 adopts the same weight distribution between AF, EF and BE traffic classes as in Fig. 6.b. Unlike Fig. 6.b, we have plotted in Fig. 7 the average packet delay proper to each class of traffic versus the global offered load. Finally, we have plotted in Fig. 8 the mean number of packets in each of the 4 queues of our queueing model versus the offered load, again assuming a fair distribution between the three traffic classes. We have the confirmation of this figure that BE traffic reaches its stability limit around a 0.35 offered load whereas the AF and EF traffic classes remain in their

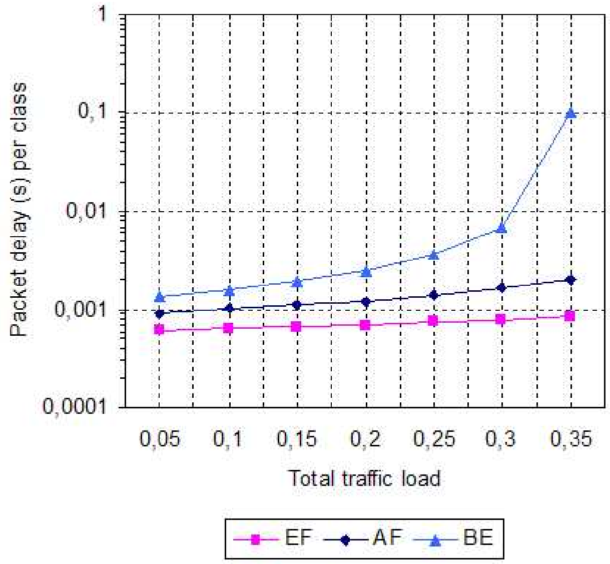

Fig. 7. Average packet delay per class

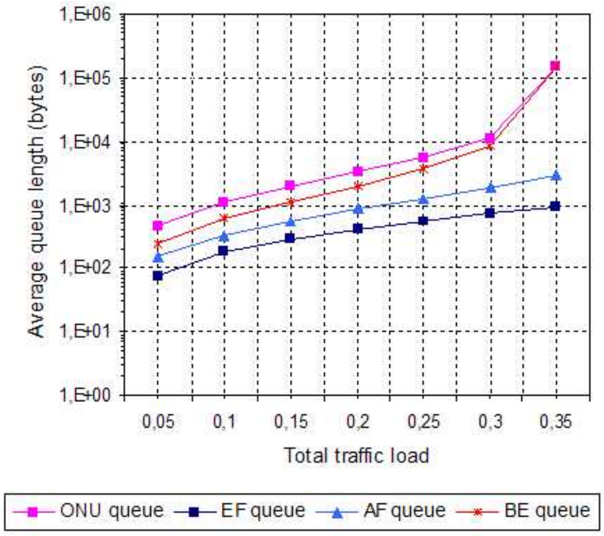

Fig. 8. Average queue length in bytes

stable state. It is possible to check the stability of the global system for an offered load of 0.3 by means of the Little's formula. Indeed, we see from Fig. 8 that an average queue length of 10e4 bytes is obtained for a load equal to 0.3 of the channel capacity. This gives an average packet delay of 2.66 ms. Such a delay is confirmed by Fig. 6.b.

V. Conclusion

EPONs are today considered as an economically mature alternative to xDSL access systems. The MPCP protocol has been adopted as the IEEE 802.3ah standard MAC protocol for EPONs. Multiple dynamic bandwidth allocation schemes such as IPACT have been recently proposed as complement of MPCP. At the best of our knowledge, we propose in this paper the first analytical model of the MPCP+IPACT protocol. Our model is based on a two stages open Jackson's queueing network. Our numerical results applied to the three Diffserv classes of traffic are confirmed by discrete event computer simulation already published in the recent literature.

References

[1] "IEEE 802.3ah Task Force home page [Online]. Available: http://www.ieee802.org/3/emf."

[2] G. Kramer, A. Banerjee, N.K.Singhal, B. Mukherjee, S. Dixit and Y. Ye , "Fair queueing with service envelopes (FQSE): A cousin-fair. hierarchical scheduler for subscriber access networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 8, pp. 1497–1513, october 2004.

[3] M. P. McGarry, M. Maier, M. Reisslein, "Ethernet PONs: a survey of Dynamic Bandwidth Allocation (DBA) algorithms," *IEEE Communications Magazine*, vol. 1, pp. S8–S15, august 2004.

[4] G. Kramer, B. Mukherjee and G. Pesavento, "IPACT: A Dynamic Protocol for an Ethernet PON (EPON)," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 74–80, february 2002.

[5] S. Blake, D. Blak, M. Carlson, E. Davies, Z. Wang and W. Weiss, "An architecture for differentiated services," IETF, RCF 2475, Tech. Rep., december 1998.

[6] B. Baynat, *Théorie des files d'attente – des chaînes de Markov aux réseaux à forme produit*. Paris, France: Hermes Science Publications, 2000.

[7] Guy Pujolle, Serge Fdida, *Modèles de systèmes et de réseaux – Tome 2 : Files d'attente*. Paris, France: Eyrolles, 1989.

[8] Y. Luo and N. Ansari, "Bandwidth allocation for multiservice access on epons," *IEEE Optical Communications*, vol. 43, no. 2, pp. S16–S21, february 2005.