This document is published in:

Networks and Optical Communications (NOC) (2012) pp. 1-8

DOI: 10.1109/NOC.2012.6249913

Managing delay in the access

David Larrabeiti, Jose A. Hernandez, Isaac Seoane and Ricardo Romeral
Escuela Politecnica Superior
Universidad Carlos III de Madrid
Spain, Leganes 28911
Email: http://www.it.uc3m.es

***Abstract*—So far latency has been regarded as a minor issue in Passive Optical Networks (PONs). However it may become a key factor for the commercial success of PON and WDM-PON. In this work we review the relevance of very-low latency in the access as the enabler of new higly-interactive cloud services. Then, we propose an approach for delay-based differentiation in PON that causes the delay variance to become smaller than in regular IPACT, featuring a reduction of jitter for high-priority traffic. Then we analyse the impact of setting a maximum delay guarantee to high-priority traffic, on low-priority traffic.**

I. DELAY IN THE ACCESS

Operators and equipment vendors are facing a major challenge in the next years as the race to offer Gigabit rates to residential users has started under a tough global competition, where subscribers expect to pay not much more than they are currently paying for their ADSL access. A more radical investment in FTTH by network operators could be justified by both a radical increase of speed offered to subscribers, and the advent of new services for which operators are in an advantageous position to get actual additional profit from the investment. However, in our view, it is not just very-high speed what can give birth to killer services. It is also end-to-end latency, since it is very likely that the physical limits of the Internet Round-Trip-Times will make it indispensable to push a number of Future Cloud services closer to the user, i.e. not outside the metro-access area. These services that may guarantee a direct involvement of operators in service exploitation are envisioned to be in the short term : very-highly interactive immersive Cloud Gaming, low-level Cloud services such as virtualised storage devices (e.g. SAN disks) and hard real-time remote control of advanced electro-mechanical sytems such as home robots or e-health devices, to give just a few examples. Indeed, these services together with a proper business model can justify the investment, but it should be noted that they require extremely well controlled round-trip-time to the servers and, in some cases, physical proximity to the cloud, a key aspect for operators. Therefore, it is paramount to properly address latency in the design of the next generation of access and metropolitan area network technology.

The same fate follows WDM-PON to a large extent, with profound implications in the metro area. WDM-PON is the expected upgrade to provide guaranteed ultra-broadband rates far beyond the capacity of current GPON/EPON shared access. A massive deployment of this technology would require not just an enhancement of access technology, but also a strong investment in Metro and Core networks to supply the required capacity to deal with millions of individual Gigabit/s access rates in a scalable way. Consequently, the target market niche of technologies such as UDWDM-PON risks becoming constrained only to a) the corporative environment (Internet access and VPN) and b) inside the MAN network as traffic aggregators (usually in the optical backhaul of cellular wireless systems). Therefore, it seems that only the mainstreaming of immersive high-bandwidth low-latency services in the residential segment can justify the expense required to remove the new bottleneck in the Metro Area Network and make it scale as the access boosts its capacity.

Table I gives estimations of the processing and transmission delays involved in xDSL, PON and WDM-PON. Some of the data displayed are further explained in section III. The best case is given for 20in ADSL and $N = 32$ ONUs in TDM-PON. The worst case listed gives an estimate for the case of a long interleaving time configured in ADSL and 80More typical data can easily be computed from this table e.g. low user load and high downstream load, and also round-trip times (RTT). In the best-case scenario, the RTT for ADSL is around 20 ms, whereas TDM-PON takes 0.4 ms and TDM-PON 0.2ms. In the worst case, a 1500-byte packet takes 90 ms to go and return in ADSL, and 3 ms in TDM-PON or 0.3 ms in WDM-PON. From this data, it becomes clear that WDM-PON is in an advantageous position to justify its deployment if ultra-interactive services become widespread, since this technology gives an order of magnitude of latency improvement over TDM-PON in the worst case, and four times in the best case (low fan-out unloaded TDM-PONs). Nevertheless, before WDM-PON is widely available, it is possible to enhance TDM-PON with QoS mechanisms in order to provide latency awareness and delay differentiation between high priority interactive traffic and the rest. In the next section we review existing approaches that have worked on this direction.

II. DELAY DIFFERENTIATION IN PON AND RELATED WORK

A Passive Optical Network (PON) works as a broadcast-and-select network in the downstream direction, since the data sourced at the Optical Line Terminal (OLT) is replicated by the passive splitter/combiner and delivered at all Optical Network Units (ONU). On the other hand, the upstream wavelength is shared by all ONUs, so a channel access arbitration mechanism must be defined to avoid collisions at the passive

TABLE I
ACCESS LATENCY ESTIMATE

	Legacy ADSL upstream	**Legacy ADSL downstream**	**TDM-PON upstream**	**TDM-PON downstream**	**WDM-PON**
Propagation delay					
1 Km	5.2 μs	5.2 μs	4.84 μs	4.84 μs	4.84 μs
20 Km	104 μs	104 μs	96.8 μs	96.8 μs	96.8 μs
Interleaving delay	2.5 − 12.5 ms	2.5 12.5 ms	–	–	–
Serialization delay (1500 byte packet)	12 ms (at 1Mb/s)	1.2 ms (at 10 Mbit/s)	12 μs (at 1 Gbit/s)	12 μs (at 1G bit/s)	12 μs (at 1G bit/s)
Queueing delay (Poisson traffic)					
$E(W_q) = \frac{\rho}{1-\rho}E(X)$					
At 20% load	3 ms	300 μs	3 μs	3 μs	3 μs
At 80% load	48 ms	4.8 ms	48 μs	48 μs	48 μs
TDMA upstream delay (20Km) [1]					
Absolute minimum, low load					
$T_{min} = RTT + T_{proc}$			235 μs		
Medium to high loads:					
$E[T] = \frac{NT_0}{(1-\rho_T)}$ to $2E[T]$)					
N=32 ONUs					
At 20% load			235 μs		
At 80% load	–	–	320 − 640 μs	–	–
N=128 ONUs					
At 20% load			320 − 640 μs		
At 80% load			1.28 − 2.56ms		
TOTAL					
Best case	17.6 ms	4.1 ms	0.35 ms	0.11 ms	0.11 ms
Worst case	72.6 ms	18.6 ms	2.7 ms	0.16 ms	0.16 ms

splitter/combiner. The MultiPoint Control Protocol for Ethernet PONs (MPCP described in the IEEE 802.3ah standard) specifies a mechanism where, the OLT firstly estimates the different RTTs (Round Trip Times) to the ONUs and secondly, it schedules transmission windows for the ONUs based on their bandwidth requirements.

A number of Dynamic Bandwidth Allocation (DBA) algorithms for timeslot scheduling have already been proposed in the literature [2], being the Interleaved Polling with Adaptive Cycle Time (IPACT) algorithm [3] the most popular one. Essentially, the OLT and the ONUs coordinate themselves via the exchange of two control messages: Report and Gate. The Reports are generated by the ONUs to inform the OLT about the bandwidth they need for the next window scheduling round. With this information, the OLT decides the window size and the starting transmission time for each ONU. Then, a Gate message with this information is sent to all ONUs.

Concerning Quality-of-Service (QoS) support in PONs, most previous studies have focused on service differentiation based on bandwidth, rather than on delay. A good summary can be found in [2]. For instance, the authors in [4] propose an algorithm that divides the total upstream bandwidth into fixed-sized bandwidth units (say 10 Mbps per bandwidth unit), such that, high-priority ONUs receive more bandwidth units than best-effort ones. Further refinements are proposed in [5], [6], [7] where the ONUs have several Virtual Output Queues (VOQs) for different traffic classes, such that the granted transmission window is shared by the ONU's traffic classes following some weighted algorithm that assigns more time to high-priority VOQs than to low priority ones. Essentially, in most cases, it is the OLT which guarantees a fair sharing of bandwidth between the ONUs, while each ONU partitions the granted transmission window to the traffic classes following some weighted algorithm [8].

In conclusion, most QoS DBA proposals for PONs have focused on service-differentiation based on bandwidth sharing between traffic classes. Only a few studies have taken into account the average delay and jitter of packets from a given ONU, see for instance the studies of the authors in [9], [10], [11], [12]. However, these studies do not provide a clear delay-based proportional differentiation algorithm, an important feature for a number of applications such as highly interactive networked gaming or IP telephony.

Thus, this work proposes a new delay-based service differentiation DBA algorithm for IPACT-based PONs and presents preliminary simulations to validate the approach. This new algorithm forces the transmission of high-priority traffic more often than low-priority one such that the expected average cycle time for high-priority packets is much smaller (and also proportional and configurable) than for low-priority ones. A complete explanation of this algorithm and the proposed analytical model can be found in [13]

We shall start by reviewing the formulae ruling delay in regular IPACT in the next section.

III. THE GATED PON TDM MODEL

As studied in [1], in the steady state and at medium-to-high loads, the average cycle time $E[T_i(k)]$ of IPACT is given by

$$E[T] = \frac{NT_0}{1-\sum_i \rho_i} = \frac{NT_0}{1-\rho_T} \tag{1}$$

thus, $E[T]$ only depends on the total load ρ_T, the number N of ONUs, and the guard and Report times $T_0 = T_{guard} + T_{report}$. T_{guard} is the guard time between windows ($1.5\mu s$ recommended at 1 Gbps) and $T_{report} = \frac{8\cdot 64}{10^9} \approx 0.5\mu s$ is the transmission time of a 64-byte Report message [1].

The formula 1 is obtained under the following assumptions:

- All the N ONUs are d km distant from the OLT.
- The system is in the steady state.
- The ONUs offer a fixed traffic load over time ρ_i, $i = 1,\ldots,N$. Furthermore, the i-th ONU receives traffic from its users following a Poisson process with rate λ_i packets/sec. Also, each packet requires a fixed amount of service time $E[X] = \frac{1}{\mu}$ computed as:

  $$E[X] = \frac{1}{\mu} = \frac{8B}{C} \text{ secs} \tag{2}$$

  where B refers to the packet size and C denotes the line rate. For $B = 1518$ bytes and $C = 1$ Gbps, the service time required is $E[X] = 12.14\mu s$ per packet.

  Hence, the i-th ONU offers ρ_i, $i = 1,\ldots,N$ traffic load as:

  $$\rho_i = \frac{\lambda_i}{\mu} \tag{3}$$
- Finally, the total offered load ρ_T, that is, the sum of all individual traffic loads ρ_i must be smaller than unity:

  $$\rho_T = \sum_{i=1}^{N} \rho_i < 1 \tag{4}$$

The delay experienced by a random packet arrival lies between $E[T]$ and $2E[T]$. The former arises when the packet arrives exactly before the transmission of the Report message to the OLT, and assuming an empty queue, whereas the second occurs when the packet arrives just after the Report message has been sent to the OLT, as noted in [1]. Hence, the average delay experienced by a given packet is proportional to the average cycle time $E[T]$. We will use $E[T]$ as the performance metric to evaluate our DBA algorithm, bearing in mind that the average delay experienced by a packet is proportional to $E[T]$.

Eq. 1 is only valid at medium to high loads (see [1]) because the Gate message cannot depart before its associated Report packet has arrived at the OLT. To meet this condition, it is then required that the average cycle time $E[T]$ is greater than the Round-Trip Time (RTT) and processing delay of the Report packet T_{proc}. In other words, Eq. 1 is valid only when:

$$E[T] = \frac{NT_0}{1-\rho_T} > RTT + T_{proc} \tag{5}$$

which sets a minimum load value for Eq. 1 to be valid of:

$$\rho_T > 1 - \frac{NT_0}{RTT + T_{proc}} \tag{6}$$

Typically, for a 1 Gbps PON with $N = 32$ ONUs, $d = 10$ km distance between the OLT and ONUs and assuming a value of $T_{proc} = 35\mu s$, the above requirement is:

$$\rho_T > 0.24$$

for a $d = 10$ Km distance OLT-ONU, and

$$\rho_T > 0.73$$

for a $d = 20$ Km distance OLT-ONU.

In the cases where this condition cannot be met, the cycle time is imposed by RTT and the processing time of a report packet. For the examples above:

For 10 Km:

$$T_{min} = RTT + T_{proc} = 2\frac{10km}{2\cdot 10^5 km/s} + 35\mu s = 135\mu s$$

and for 20 Km:

$$T_{min} = 235\mu$$

Table I reflects these results and its relevance in the overall latency budget.

IV. ANALYSIS OF THE GATED PON TDM MODEL WITH DELAY DIFFERENTIATION

A. Scheduling algorithm with delay differentiation

Now, consider that each ONU employs two output queues, one per traffic class: High and Low Priority (HP and LP). The former traffic class is expected to contain packets from delay-sensitive applications, typically video streaming, voice over IP, or online gaming, whereas the second one is for best-effort applications.

The goal is to design a transmission scheduling algorithm that favors HP over LP traffic in terms of delay experienced at the ONU. Most previous studies have proposed to schedule the HP traffic ahead of the transmission window in order to save delay. Our algorithm differs from such studies since it proposes to separate HP and LP traffic Reports such that, HP traffic is sent to the OLT more regularly than LP traffic. This is expected to minimise the delay experienced by high-priority packet arrivals at the ONUs, however at the expense of a delay increase for the LP traffic. An example of operation for three ONUs is shown in Fig. 1.

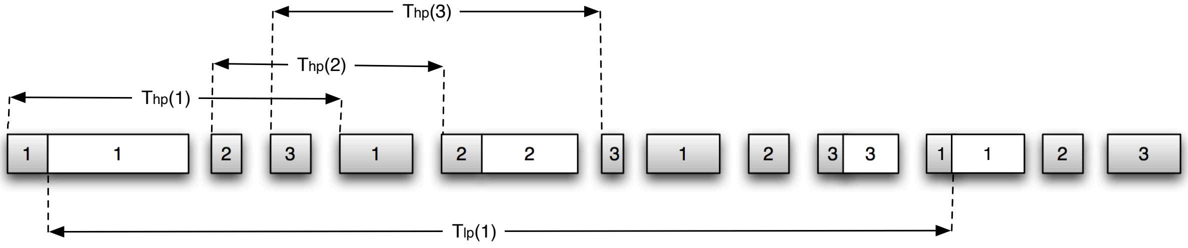


Fig. 1. Algorithm with delay differentiation. Case $M = 1$.

In this example, HP traffic (shadowed boxes) from the three ONUs are sent to the OLT more regularly than LP traffic (white boxes). We observe two types of cycle times, one for HP traffic $T^{(HP)}$ and another one for LP traffic $T^{(LP)}$, where $T^{(HP)} \leq T^{(LP)}$. Basically, the OLT polls the HP packets for the three ONUs and the LP traffic for a single ONU (ONU 1 in the example) in one HP cycle. On the second round, the OLT polls the HP packets for the three ONUs again, and the LP traffic from the second ONU. Finally, on the third round, the OLT polls the HP traffic from the three ONUs and the LP traffic for the third ONU, thus completing an LP cycle. In the fourth round, the OLT polls the HP traffic from the three ONUs and the LP traffic from the first ONU, just like in the first round. Hence, in this case, the LP cycle is as large as three HP cycles, as shown in Fig. 1.

Typically, the HP transmission windows are smaller than the LP ones for two reasons: First, delay-sensitive applications generate less traffic than best-effort applications. Secondly, the HP VOQ is polled more regularly than the LP VOQ, thus the latter aggregates more packets on every LP cycle.

As a generalisation, let M refer to the number of LP ONUs polled within the HP polling cycle. Fig. IV-A shows the cases for $M = 1, 2, 3$. As shown, case $M = 3$ produces the same cycle time for HP and LP traffic, therefore no delay-based service differentiation is performed.

B. Analysis of scheduling algorithm with delay differentiation

Now, let $E[T^{(HP)}]$ and $E[T^{(LP)}]$ refer to the average HP and LP cycle times respectively. Remark from Fig. IV-A that the average HP cycle time accounts for the N HP transmission windows plus another M LP windows plus the N guard and Report times, denoted by T_0. Hence it can be deduced that:

$$E[T^{(HP)}] = N\rho_i^{HP} E[T^{(HP)}] + M\rho_i^{LP} E[T^{(LP)}] + NT_0 \quad (7)$$

where $\rho_i^{(HP)} = p_h\rho_i$ and $\rho_i^{(LP)} = (1 - p_h)\rho_i$ for some $p_h \in (0, 1)$. Here p_h refers to the percentage of HP traffic over the total. Again, the total traffic must be smaller than unity:

$$\rho_T = \sum_{i=1}^{N} \left(\rho_i^{(HP)} + \rho_i^{(LP)}\right) < 1 \quad (8)$$

Similarly, for $E[T^{(LP)}]$:

$$E[T^{(LP)}] = \frac{N}{M}\sum_{i=1}^{N} E[V_i^{(HP)}] + \frac{N}{M}\sum_{i=1}^{M} E[V_i^{(LP)}] + \frac{N^2}{M}T_0 \quad (9)$$

From eqs. 7 and 9, it can be derived [13]:

$$E[T^{(HP)}] = \frac{NT_0}{1-\rho} \quad (10)$$

$$E[T^{(LP)}] = \frac{N}{M}\frac{NT_0}{1-\rho} \quad (11)$$

Concludingly, it can be observed that there is an M/N relationship between the HP and LP average cycle times, as shown from:

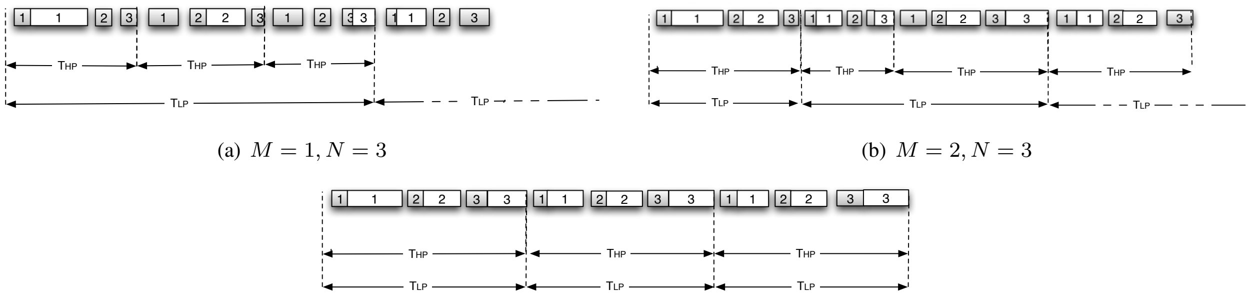
$$\frac{E[T^{(HP)}]}{E[T^{(LP)}]} = \frac{M}{N} \quad (12)$$

This relationship allows the scheduler to define the level of delay-based differentiation between HP and LP traffic, just by adjusting the appropriate value of M in the scheduler. The case $M = 1$ refers to maximum delay difference between HP and LP traffic, whereas the $M = N$ case refers to no difference between HP and LP traffic.

V. ANALYSIS WITH GUARANTEED HP CYCLE TIMES

The previous section proposed a mechanism to set a proportional delay differentiation between HP and LP traffic. However, this mechanism does not guarantee an upper delay bound for HP traffic, that is, HP delay may be ten times smaller than LP delay, but still too high for real-time application performance. This section extends the mechanism proposed previously by defining a hard bound on the HP cycle times, even at the expense of LP traffic loss. Let $T_{Limit}^{(HP)}$ refer to such an upper delay bound on the HP cycle times.

Essentially, when the OLT collects the Reports of all ONUs for the next HP cycle (this includes the N HP windows and the M LP windows), it must check whether the sum of all the transmission windows requested by the ONUs is below such threshold $T_{Limit}^{(HP)}$. If it does, then the OLT can assign all the transmission windows requested by the ONUs. However, if the total transmission window requested exceeds the $T_{Limit}^{(HP)}$ value, then the OLT must assign smaller LP transmission windows. Such excess in LP traffic would be either delayed to the next



(a) $M = 1, N = 3$ (b) $M = 2, N = 3$

(c) $M = 3, N = 3$

Fig. 2. Algorithm with delay differentiation. Cases $M = 1$, 2 and 3. Timing indicated for ONU 1

LP cycle or loss. Fig. 3 shows an example of this situation for $N = 3$, $M = 1$.

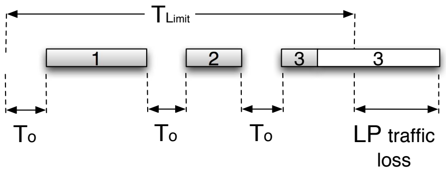


Fig. 3. Algorithm with delay differentiation and bounded HP cycle. Case $M = 1$.

Under these premises it can be found that the percentage of LP packet loss equals:

$$P_{Loss} = \frac{T^{HP}_{Limit}(\rho - 1) + NT_0}{\rho T^{HP}_{Limit}} \tag{13}$$

VI. SIMULATION RESULTS

This section tries to validate the analysis outlined in the previous sections in a number of realistic situations.

A. Validation of the average cycle times $E[T]$, $E[T^{(HP)}]$ and $E[T^{(LP)}]$

In this first experiment, the goal is to validate the equations obtained previously for $E[T^{(HP)}]$ and $E[T^{(LP)}]$ (eqs. 7 and 9) in the delay-based service differentiation algorithm for IPACT-based PONs.

The experiments in both cases consider the following system parameters:

- Line rate: $C = 1$ Gbps.
- Guard time: $T_{guard} = 1.5\mu s$
- Processing time of a Report packet: $T_{report} = 0.512\mu s$
- Percentage of HP traffic over the total: $p_h = 0.4$

Figs. 4(a)-4(c) show the $E[T^{(HP)}]$ and $E[T^{(LP)}]$ for 32 ONUs and different values of M in the scheduler. As shown, for $M = 1$, the delay differentiation is maximum (1/32 ratio), and for $M = 32$ there is no delay differentiation. Intermediate values of $M = 4$ and $M = 16$ are also shown. Additionally, it can be shown that the equations accurately match with the simulation experiments, thus validating the analytical sections.

Figs. 5(a)-5(b) show the average cycle time plus/minus twice its standard deviation for $E[T]$ and $E[T^{(HP)}]$. As shown, although the average cycle time are the same in both cases, the standard deviation is smaller in our algorithm, thus producing less spread HP cycle times, which favors for a small jitter for delay-sensitive applications, especially at high loads and for low values of M. The average cycle times and their standard deviation have been estimated from simulation.

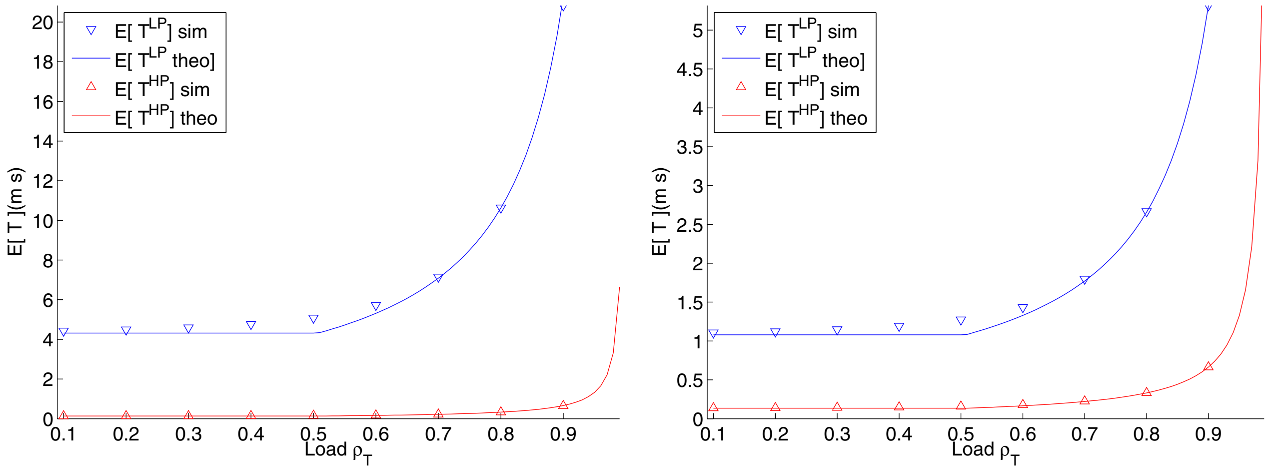
B. Working under heavy traffic conditions

Figures 6(a)-6(d) show the behaviour of the delay-based service differentiation DBA algorithm proposed in this article under heavy-traffic conditions, that means, for total loads in the range $1 \leq \rho_T \leq 2$. Two delay limits are considered in the experiments: $T^{(HP)}_{Limit} = 3.2ms$ (top) and $T^{(HP)}_{Limit} = 8ms$ (bottom). As shown, the values of the HP traffic cycles are kept below the $T^{(HP)}_{Limit}$ bound, in both cases while the LP traffic cycles are much higher. In addition, to guarantee such bounded delays, LP priority traffic is lost, as explained in Section V.

VII. CONCLUSION

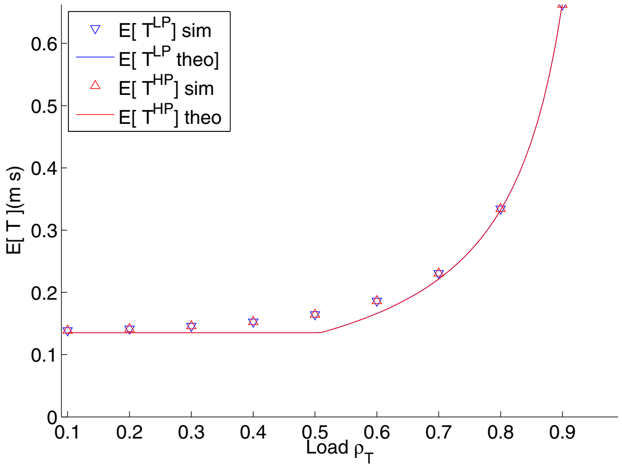
In our view, in order to foster the creation of new services and applications that a) are not feasible today with xDSL or GPON, and hence b) can guarantee a short-term return of investment (ROI) and profits to operators straight out of the Capital Expenditure (CAPEX) investment, the network must also provide the ability to transport a large amount of 1Gb/s circuits PON-to-Cloud across the MAN featuring sub-ms latency. Those 1Gb/s low-latency-demanding services can be run by the operator or a service provider located within the metropolitan area. Therefore the development of an ultra-dense capacity technology that is designed to flash-transfer packets PON-to-Cloud is a fundamental capability of a new generation metro-access network technology to be investigated.

On the way to WDM-PON, which provides the best scenario to low latency, we believe that some research should be devoted to latency in regular PON infrastructure. In this work we outlined a new Dynamic Bandwidth Allocation (DBA) algorithm to provide delay-based service differentiation of IPACT-based PONs. Essentially, the algorithm defines two types of polling cycles for the OLT: the high- and low-priority (HP and LP) cycles, being the former more regularly



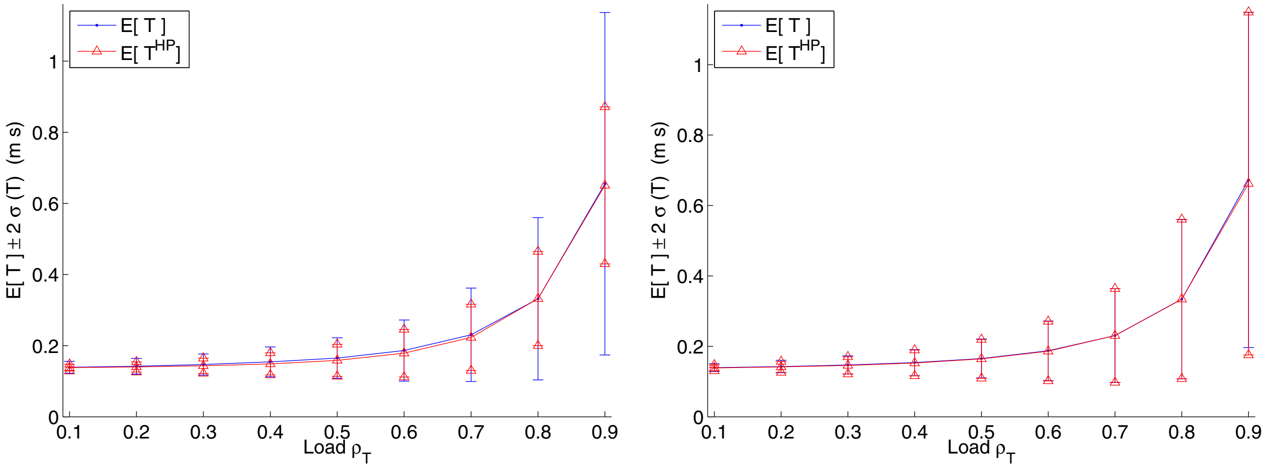
(a) $M/N = 1/32$

(b) $M/N = 4/32$



(c) $M/N = 32/32$

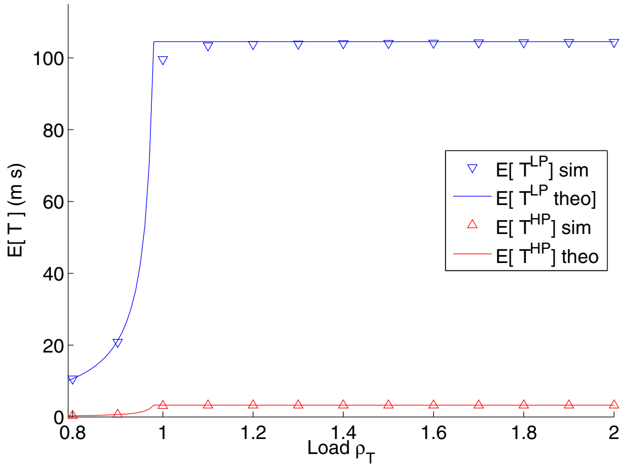
Fig. 4. Average cycle times $E[T^{(HP)}]$ and $E[T^{(LP)}]$ for M/N ratios: (a) $\frac{M}{N} = \frac{1}{32}$, (b) $\frac{M}{N} = \frac{4}{32}$, (c) $\frac{M}{N} = \frac{32}{32}$ (no differentiation)



(a) $M/N = 1/32$

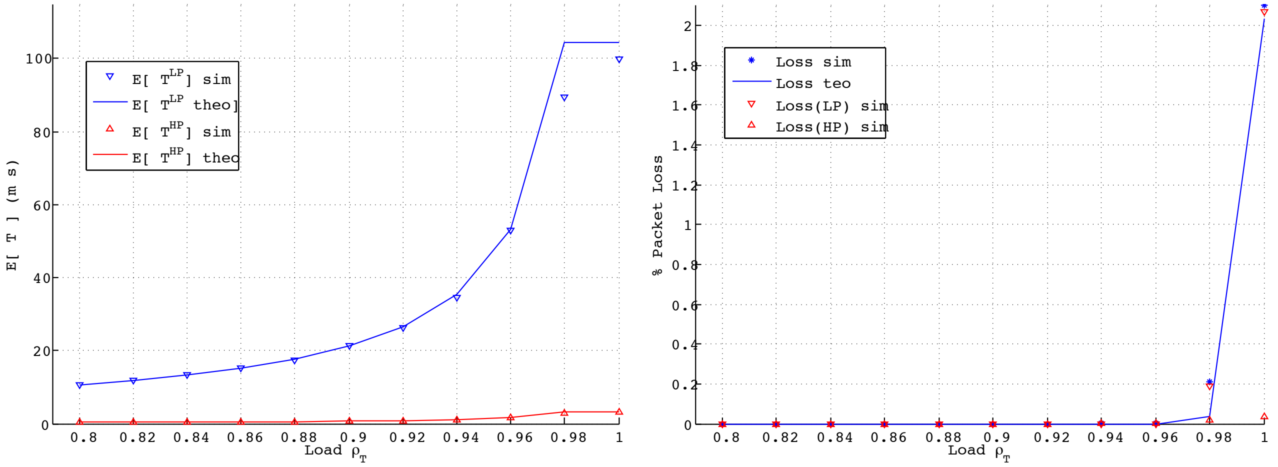
(b) $M/N = 32/32$

Fig. 5. Average cycle times $E[T]$, $E[T^{(HP)}]$ and their standard deviations ($E[T] \pm 2Std[T]$) for M/N ratios: (a) $\frac{M}{N} = \frac{1}{32}$,(b) $\frac{M}{N} = \frac{32}{32}$ (no delay differentiation)



(a) Average cycle time for $T_{Limit} = 3.2ms$

(b) Packet loss for $T_{Limit} = 3.2ms$



(c) Average cycle time for $T_{Limit} = 3.2ms$ (zoom)

(d) Packet loss for $T_{Limit} = 3.2ms$ (zoom)

Fig. 6. Average cycle times and packet loss ratios for the delay differentiation DBA algorithm featuring cycle limitation under heavy traffic conditions: $T_{Limit} = 3.2ms$, avg. cycle time (a) and packet loss ratio (b); and zooms (a) (b).

performed than the latter ones. This strategy favors the creation of short HP cycles and long LP cycles, with a predefined ratio between them. In this paper we showed simulation results of this idea, fully developed in [13]. Then we introduced un upper bound on the HP cycle times in order to set a hard limit to HP traffic delay, and analysed its impact on LP traffic loss. Simulation results validated our approach and served to quantify the impact of the different design parameters of the algorithm on the whole system performance.

ACKNOWLEDGMENT

The authors would like to thank the MEDIANET project and the FIERRO thematic network.

REFERENCES

[1] B. Lannoo, L. Verslegers, D. Colle, M. Pickavet, M. Gagnaire, and P. Demeester, "Analytical model for the IPACT dynamic bandwidth allocation algorithm for EPONs," *OSA J. Optical Networking*, vol. 6, no. 6, pp. 677–688, 2007.

[2] M. McGarry, M. Maier, and M. Reisslein, "Ethernet PONs: A survey of dynamic bandwidth allocation (DBA) algorithms," *IEEE Communications Magazine*, vol. 42, no. 8, pp. 8 – 15, Aug. 2004.

[3] 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.

[4] M. Ma, Y. Zhu, and T. H. Chen, "A bandwidth guaranteed polling MAC protocol for Ethernet Passive Optical Networks," in *Proc. of IEEE INFOCOM*, vol. 1, San Francisco, CA, USA, 2003, pp. 22–31.

[5] L. Zhang, E.-S. An, C.-H. Youn, H.-G. Yeo, and S. Yang, "Dual DEB-GPS scheduler for delay-constraint applications in Ethernet Passive Optical Networks," *IEICE Trans. Communications*, vol. E86-B, no. 5, pp. 1575–1584, May 2003.

[6] C. M. Assi, Y. Ye, S. Dixit, and M. A. Ali, "Dynamic bandwidth allocation for Quality of Service over Ethernet PONs," *IEEE J. Selected Areas in Communications*, vol. 21, no. 9, pp. 1467–1477, Nov. 2003.

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

[8] B. Moon and H. Song, "Hierarchical Ethernet-PON design with QoS supporting by eligible reporting," in *Proc. of Asia-Pacific Conf. in Communications*, Busan, Aug. 2006, pp. 1–5.

[9] C. Park, D. Han, and B. Kim, "Packet delay analysis of dynamic bandwidth allocation scheme in an Ethernet PON," *Lecture Notes in Computer Science*, vol. 3420, pp. 161–168, 2005.

[10] C. G. Park, D. H. Han, and K. W. Rim, "Packet delay analysis of symmetric gated polling system for DBA scheme in an EPON," *Telecommunication Systems*, vol. 30, pp. 13–34, 2005.

[11] T. Berisa, Z. Ilic, and A. Bazant, “Absolute delay variation guarantees in passive optical networks,” *IEEE/OSA J. Lightwave Technology*, no. 99, 2011.

[12] F. An, Y. Hsueh, K. Kim, I. White, and L. Kazovsky, “A new dynamic bandwidth allocation protocol with quality of service in ethernet-based passive optical networks,” in *Proc. of IASTED Int. Conf. Wireless and Optical Communications (WOC 2003)*, vol. 3, 2003, pp. 165–169.

[13] I. Seoane, J. A. Hernández, R. Romeral, and D. Larrabeiti, “Analysis and simulation of a delay-based service differentiation algorithm for IPACT-based PONs,” *Photonic Network Communications*, 2012.