Exploiting PONs for Mobile Backhaul

T. Orphanoudakis, E. Kosmatos, J. Angelopoulos, and A. Stavdas, University of Peloponnese

ABSTRACT

The growing popularity of mobile data services necessitates a rapid rise in network capacity not only on the air interface to the end user, but also in the backhaul network. The latter is quite important for the mobile operator business model, affecting capital investment, operational expenses, service deployment, and customer experience. Fiber infrastructure is inevitably the only long-term solution, and the deployment of passive optical networks presents an opportunity for a cost-effective, scalable, and future-proof solution. In this article we investigate the use of PONs for mobile backhaul and propose a resource allocation framework building on the efficiency of PONs to share resources, dynamically allocate bandwidth in real time, and enhance efficiency by improved statistical multiplexing. The main objective of this work is to exploit existing standardized technologies, and provide design and deployment guidelines regarding PON MAC operation, enabling a gradual and future-safe infrastructure upgrade of mobile backhaul systems.

INTRODUCTION

Mushrooming mobile traffic driven by third-/-fourth-generation (3/4G) systems and novel mobile data applications is saturating the current backhaul networks, which are based on dedicated links (mainly microwave or E1/T1). However, present solutions have limited upgrade potential, while the recent traffic trends will make the eventual need to connect the stations to some form of fiber inevitable sooner rather than later. Given that the situation is not static but rapidly changing, the introduction of fiber backhaul has to follow a careful migration path compatible with revenue generation since a large initial cash outlay does not seem a viable approach for even the more affluent operators. In this light, time-division multiple access (TDMA) passive optical networks (PONs) offer the significant advantage of cost-effective port and traffic consolidation [1]. Apart from the lower cost compared to dedicated fibers for the initial deployment, PONs retain the extra comfort of a secure, gradual, and future-proof evolution path. This can lead to any desired bandwidth in the form of upgrades to 10GPON and later wavelength-division multiplexing (WDM) PONs, which can even provide dedicated wavelengths without retrenching and cabling whenever this ever becomes necessary.

A mixed use of PONs for both fixed access/wireless backhaul offers obvious synergy, bringing forward the economic break-even point for both mobile and fixed line operators/providers while benefiting the end user at the same time. This serendipity provides a strong edge to the selection of PONs over competing technologies, particularly for the introductory phase, when costs will be critical and traffic not high enough to justify the full PON capacity for the residential market or mobile market alone.

In the mixed use case, a service level agreement (SLA) with the mobile operator based on over-provisioning may only be tolerable in the initial stage while the PON is still lightly loaded. However, as traffic picks up, the exploitation of a well-tuned TDMA PON medium access control (MAC) protocol will become indispensable for high system utilization and hence profitability. The purpose of this article is to investigate the critical worst-case delay and latency issues arising in this environment, assess the traffic handling capabilities of TDMA PONs under such a mixed initial traffic scenario, and provide design and deployment guidelines to both manufacturers and operators on the fine tuning of the PON MAC parameters.

To this end, typical initial deployment architectures are presented next; then guidelines for traffic handling are given, and computer simulation is used to assess the traffic performance under typical and worst-case service scenarios that such a TDMA PON can provide. In addition suggestions for PON MAC fine-tuning are presented.

ARCHITECTURAL SETUP

A typical setup under the presented introductory scenario of a PON used for mixed residential and mobile backhaul (MBH) is depicted in Fig. 1, where some optical network units (ONUs) support residential or professional users and small businesses, while one or two serve mobile base transceiver stations (commonly called BTS or eNodeB, depending on the technology; we use the broader term BTS hereafter) and are interconnected through the PON fiber tree to the optical line termination (OLT). The effect is that PONs allow deeper fiber penetration at lower cost, and simplify integration with optical metro and core networks.

No cost or technical advantage would make the PON a viable proposition for MBH as an initial or interim solution if it could not offer an obvious, easy, cost-effective, well defined, and safe evolution path to any desired bandwidth in

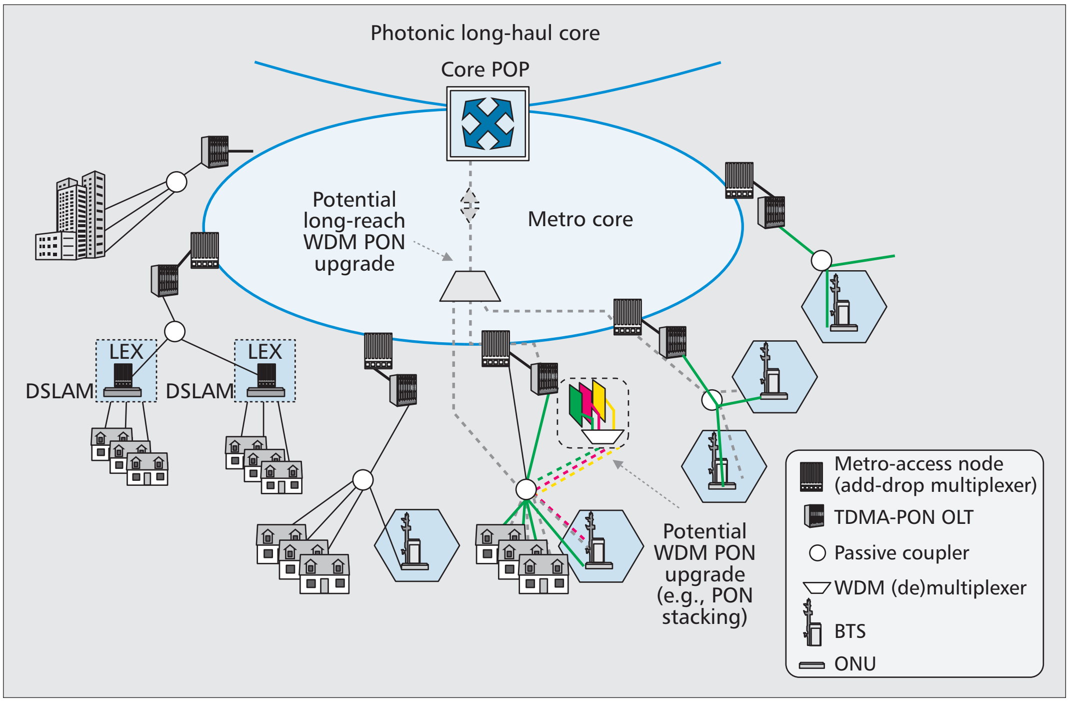

Figure 1. *Typical architecture and potential WDM evolution paths (dotted lines).*

A significant cost-advantage of the PON comes from the cheap and effective way that traffic multiplexing into a single port is realized by a low-cost passive combiner. Packets are marshaled one behind the other in perfect and gapless succession as a by-product of the MAC operation.

the future without tearing up the infrastructure and retrenching. Upgrades to faster data rates (e.g., 10 Gb/s) are the obvious first step, but also at a later time more technological alternatives may mature and find their way into standard systems exploiting subcarrier multiplexing (e.g., orthogonal frequency-division multiple access [OFDMA]-PONs), radio over fiber (RoF), and ultimately WDM [2]. In the latter case a dedicated PON wavelength may become feasible and justified for wireless systems, but a lot of income must have been generated before this can become a viable solution. The replacement of TDMA by WDM can be gradual, starting with the reduction of the splitting ratio by overlaying new PONs at different wavelengths, as shown in Fig. 1, reducing reliance on TDMA mode to eventual elimination. This would require the investment of deploying additional OLTs to handle the extra wavelengths and upgrade of ONUs of users connected to the WDM PON, exploiting the additional bandwidth through wavelength agility. The new OLTs could be collocated with legacy TDMA OLTs operating as a WDM overlay, guaranteeing backward compatibility (Fig. 1), or placed deeper in the network, exploiting long reach WDM PON technologies (shown in a feeder fiber-based topology in Fig. 1, while collector ring-based topologies are also possible, as discussed in [2, references therein]), leading to potential node consolidation, higher aggregation, and access core integration.

Traffic Handling in TDMA PONs

Mobile operators have been using static links (whether leased lines or microwave-based or even optical) to connect the traffic of BTSs. Compared to such links, TDMA PONs are quite a different affair. Their passive multiplexing, which underpins their cost effectiveness in a mixed residential-mobile backhauling deployment, requires much more complex traffic management. A thorough understanding and refinement of the dynamic bandwidth allocation aspects of the TDMA PON are essential to achieve efficient utilization of network resources while respecting quality of service (QoS) parameters and SLAs. A significant cost advantage of the PON comes from the cheap and effective way that traffic multiplexing into a single port is realized by a low-cost passive combiner. Packets are marshaled one behind the other in perfect and gapless succession as a by-product of the MAC operation.

When referring to the TDMA PON, there are two dominant standards that can be used for the mixed backhaul network: GPON [3, 4] and EPON [5]. Both foresee the support of different QoS levels embedded in the MAC (priorities) for a successful performance, but operators must be well aware of the idiosyncrasies in their packet multiplexing methods and their MAC mechanisms because the differences in each PON type (GPON, EPON) are not trivial, although their relevant functionality is based on the same general principles.

In this work the focus is on the GPON MAC, but where appropriate, comments will be provided on EPON differences and a performance comparison regarding the impact of such differences on their MBH support role. A crucial parameter when deploying a TDMA PON as an MBH is delay introduced by the shared nature of the fiber medium. This delay adds a significant burden against the requirements imposed by the evolving standards for wireless and mobile networking, which are quite stringent. For example, [6] requests lower latency bounds for the

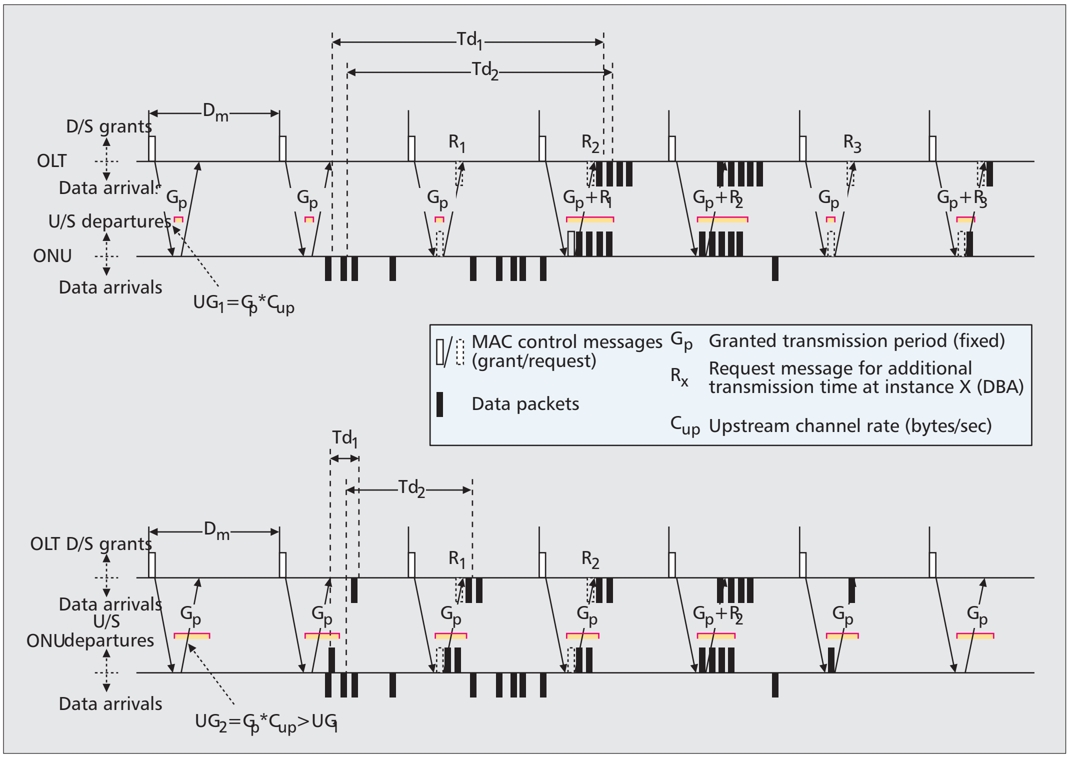

Figure 2. *Principle of operation of DBA in a TDMA PON.*

user plane (unload condition), control plane transitions, and real-time games of 5 ms, 50 ms, and 75 ms, respectively. A particularly demanding situation in terms of latency (when a packet arrives in a previously empty queue that therefore has no pending request) arises in the handling of the string of hard handover messages ([7, 8]) from a base station situated in one ONU to another station supported by a different ONU or a different PON. The performance of such a worst-case scenario is investigated in this article as it is of particular importance before feeling confident that a PON-added delay is no problem. It is also of interest to compare EPONs and GPONs in the handling of this, and investigate ways to alleviate the problem. However, before embarking in the quantitative performance assessment in the next section, the generic concept of operation will be presented based on GPON practice and terminology (the basic concept is the same also for the EPON MAC, although the details may vary), starting with the role of polling on latency. The main tool is the dynamic bandwidth allocation (DBA) mechanism, which has been well studied in GPONs [3, 9] and EPONs [10, 11], and, in fact, its status reporting option, which gives far more than the delay/efficiency performance the MBH environment requires, as shown in the performance section. As seen in Fig. 2, DBA relies on a continuous exchange of requests followed by grants a while later. The figure depicts the lifetime of packets in the PON. Arrivals are shown under the lines and departures above the lines. The time distance is marked to indicate the total transfer time (delay). Two instances (top and bottom) with different polling periods are shown to emphasize the importance of polling on the packet total delay. Polling is carried out by unsolicited grants (UGs), that is, grants sent without previous request. Normally, DBA works by first having the ONUs request service, indicating their queue length in a report field; then the OLT allocates enough upstream transmission grants to allow them to relieve the full content of their queues. Hence, at least a minimum portion of bandwidth should be reserved statically in any case in order to guarantee transmission opportunities for requests as well as traffic with low-latency requirements that cannot afford the delay of the request-grant cycles. The requests are piggybacked inside the transmissions departing from an ONU, and packets arriving into an already empty queue would never get a chance to declare their presence if it were not for the UGs arriving for the purpose of polling. Thus, polling involves granting a transmission interval to an ONU on the basis of time passed and not on known queued traffic. It is like a chain smoker who needs no fire to light one cigarette after the other, but will need a new light (polling) once he breaks the chain and extinguishes the last one. In the PON the new light comes from UGs. Frequent polling results in wasted bandwidth; large polling intervals, on the other hand, increase latency (i.e., the time waiting for the first grant when arriving in an empty queue, since non-empty queues can always transmit requests). The importance of UG rate (i.e., the resulting service rate allocated by UGs, expressing bytes over the D_m polling interval) is also illustrated in Fig. 2, where two scenarios with different UG rates are shown resulting in reduced packet delays (e.g., Td_1, Td_2) in the second scenario with the higher UG rate. This feature is exploited in our proposals below.

The situation with XG-PON and 10G-EPON is somehow improved for high priority traffic because the high rate allows faster polling while delay will not change significantly for DBA-based traffic since this is dominated by the round trip delay of the request/grant cycle.

EPONs support eight priority levels following the 802.1P approach and a somewhat restrictive native Ethernet support (i.e., Ethernet frames must be supported as a whole [5], while GPONs allow breaking up in smaller parts encapsulated in special frames).

The situation with XG-PON and 10G-EPON is somehow improved for high-priority traffic because the high rate allows faster polling, while delay will not change significantly for DBA-based traffic since this is dominated by the round-trip delay of the request/grant cycle.

Another important observation is that strict isolation between elastic and real-time traffic is required to provide performance guarantees, and this is achieved by strict prioritization into four classes of service (CoS) in GPONs. In contrast, EPONs support eight priority levels following the 802.1P approach and a somewhat restrictive native Ethernet support (i.e., Ethernet frames must be supported as a whole [5], while GPONs allow breaking up in smaller parts encapsulated in special frames [3, 4, 12]). This allows GPONs to introduce lower protocol overheads and the finest time granularity without wasting part of a granted upstream transmission time slot in case an integral packet does not fit into the remainder of this slot. This, in turn, leads to much lower levels of latency and delay than are possible in EPONs for the same level of efficiency, as becomes clear in the simulation results of the next section. In GPON terms, the traffic classes are five and are called traffic containers (TCONTs):

- TCONT1 traffic is intended for the emulation of leased line services (i.e., matching the need of real-time applications for guaranteed bandwidth and minimum latency) and is serviced only by unsolicited periodic grants (UGs).
- TCONT2 is intended for variable bit rate (VBR) traffic, and applications with demanding delay and throughput requirements. Bandwidth for this TCONT is ensured in the SLA, but assigned such that only a minimum is statically provisioned while additional bandwidth is made available on request to allow for multiplexing gain.
- TCONT3 is intended for better than best effort services and ensures service at a guaranteed minimum rate; any surplus bandwidth is assigned only on the basis of request and availability (best effort).
- TCONT4 is intended for purely best effort services and as such is serviced only on bandwidth availability.
- TCONT5 is a combined class of two or more of the other four TCONTs. The characteristic is that no target TCONT queue is specified (only ONU), and it is now left to the ONU to choose which queue to service (also called intra-ONU scheduling in [10, 11]). The use of this approach (sometimes referred to as using *colorless grants*) is left to the system designer, and its activation (when implemented) is left to the operator.

This last feature is of particular interest in this work, since it allows the PON to leave the local grant prioritization among classes to the ONU, with the MAC controller only allocating the ONU aggregate. This is made use of in the next section.

Obviously, the GPON or EPON MAC controller has to periodically visit all active ONU queues, and this leads to the concept of the mean scheduling period, D_m. The D_m parameter must be kept low enough in order to keep latency and delay variations low as well. In GPON the scheduling period can be quite small (integral multiples of 125 μs, enabling even the support of legacy time-division multiplexed [TDM] services), and, together with the low protocol overhead and the fragmentation of frames, easily achieves low latency and delay variation, while in the EPON this can only be achieved by selecting a low D_m at the expense of efficiency. This has been shown in several studies [9, 10]. The same methodology as in [9] is followed here, and the reader is referred to that article for a thorough presentation of the comparison. Consequently, the MAC protocol serves the top-priority class CoS1 (TCONT1 in GPON), periodically allocating an adequate number of unsolicited grants in every scheduling cycle, D_m, to cover any eventuality. This is the only way the operator can actually guarantee service to a contracted peak rate R_{p1} and a strict delay bound to CoS1, which can be derived as a function of D_m. The scheduling period D_m is used to calculate the bytes to be allocated to each queue to achieve the desired service rate. Hence, considering the case where unsolicited grants cover the sustainable rate R_{s2} of the second class, CoS2 (TCONT2), the total number of unsolicited grants for the ith ONU (UGi) in bytes is expressed as follows: UGi $=(R_{p1i} + R_{s2i}) * D_m$ (rates expressed in bytes per second). The remaining unallocated part of each scheduling period D_m is distributed dynamically in a weighted manner, and a service weight w_i can be used [11] to enforce proportional sharing of the upstream transmission window among ONUs to guarantee the portion reserved for CoS3 queues (TCONT3). Finally, CoS4 is served as best effort (i.e., whenever unallocated slots exist).

It follows from the above outline of the GPON MAC operation that in the simulation studies of the next section, when we assess the performance for the handover signaling exchange of messages, the relevant packets are assigned to the CoS1 (TCONT1) queue, and the operator must have foreseen an adequate number of unsolicited grants (resulting in a minimum guaranteed bandwidth) in every scheduling cycle D_m to satisfy the worst case latency. A typical hard handover scenario in GSM [7] and UMTS [8] consists of a sequence of four or five upstream single-packet messages, and the whole exchange must be completed well within the service interruption time allowed by the mobile standards. The specification for IMT [6] gives a maximum service interruption in a handover of 40 ms, while [8] specifies 50 ms. This includes all causes of delay (protocol processing, air interface, and propagation), but the new aspect of GPON or EPON MAC protocol delay, due to the queuing involved until a grant becomes available, is of course completely unaccounted for in the standard. It is reasonable to assume that only a small portion of this delay budget can be consumed by the new PON MBH. While some authors suggest a value of 2 ms, [13], in no case would it be acceptable to allow a value above 10 ms, and a safer margin might be warranted. (Obviously, soft handover presents no strict latency needs and is not considered here).

	CoS1 (TCONT1)		CoS2 (TCONT2)		CoS3 (TCONT3)	
	ONU load (%)	Profile	ONU load (%)	Profile	ONU load (%)	Profile
Residential ONUs	20	On-off, BF = 3	25	On-off, BF = 5	55	on-off, BF = 5
BTS ONU	10 10	CBR signaling	25	On-off, BF = 5	55	on-off, BF = 5

Table 1. *Simulated traffic load profiles per ONU.*

PERFORMANCE ASSESSMENT BY SIMULATION

In this section computer simulation results are presented to investigate the impact on delay of replacing a fixed backhaul link with a GPON and the increased probability of violating service-specific delay bounds, especially for the most critical case of the hard handover. Only the upstream is discussed in this study because this is where all the sensitive issues arise. As a next step, best practices for operators in allocating bandwidth to the ONU supporting the BTS and fine-tuning the UGs are presented and evaluated. The available margins and trade-offs between latency and utilization are also presented and evaluated.

For the hard handover delay, we investigated by computer simulation the time it takes to complete the signaling chain of messages under different loading conditions and polling distance programmed by the PON operator to check safety margins and by how much utilization must be sacrificed to ascertain safe service. The simulation setup employs 16 ONUs, one of which serves exclusively a mobile BTS, while the others carry residential traffic. Three CoS are simulated, of which the highest priority, COS1, is served in GPON by TCONT1 and by the top priority in EPON, while the other two by TCONT2 and 3, respectively. The fourth class (best effort) is not represented here as its study offers no useful conclusions. The traffic mix characteristics per ONU type are shown in Table 1. For all ONUs CoS1 traffic is considered to account for 20 percent of its total offered load and is modeled as either constant bit rate (CBR) voice traffic or control message traffic in the case of the wireless BTS, or data traffic modeled following an on-off model with a low burst factor $BF = (T_{\text{on}} + T_{\text{off}})/T_{\text{on}}$ in the case of residential users. All other traffic sources (CoS2 and CoS3 traffic) are considered as highly bursty data sources following an on-off model. The traffic mix was selected to demonstrate a scenario where a substantial portion of the traffic is generated from services with demanding SLAs (CoS1, CoS2) — also usually associated with a higher tariff [9–11] — while at the same time a significant signaling load is generated at the BTS ONU.

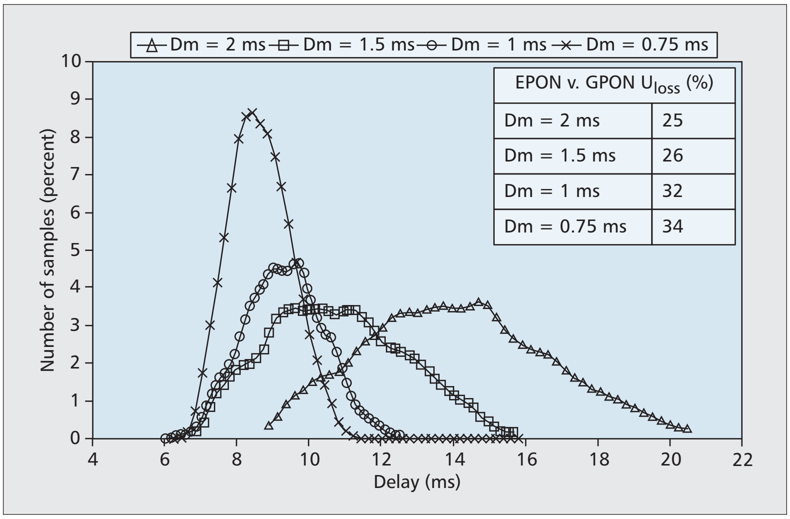

Figure 3. *PDF of signaling delay (load 40%) and inset showing throughput loss (EPON v. GPON).*

First, the total delay for the typical signaling exchange of a handover scenario was measured, and the probability density function (PDF) of this delay is depicted in Fig. 3 for a total PON load of 40 percent of its upstream channel capacity (no significant difference is observed at higher loads because of the highest priority, as is clear in Fig. 4). The signaling exchange consisted of five upstream single packets modeling the two-way handover protocol message exchange, which had to endure the access delay of the GPON backhaul and an additional processing delay before a response message was generated (random processing delay following a Poisson distribution was assumed with a mean of 1 ms). Values near or above 10 ms would risk unacceptable service interruption. As can be seen in Fig. 3, the impact of the D_m parameter is dominant, as expected from previous studies of TDMA GPON/EPON delay. In reality, only the value of $D_m = 0.75$ ms provides a small enough tail to give confidence in the mixed architecture studied in this article. This D_m value is six times the frame size of GPON and can also easily be programmed in the EPON, but at the penalty of some inefficiency. This is due to the way EPON is designed to carry whole Ethernet frames, leaving an unused space remainder (USR) at the end of each upstream allocation. Lowering D_m decreases the mean upstream transmission length, thus increasing this waste. There is no need to repeat the interesting investigation of this EPON idiosyncrasy, which has been extensively studied (e.g., [9, 12]); however, to give a quantitative indication of this effect here, we provide for comparison in the inset of the same figure the values of upstream PON utilization lost U_{loss} (i.e., the throughput reduction in EPON as a percentage of that of a GPON) for each of the same D_m value and the same load-

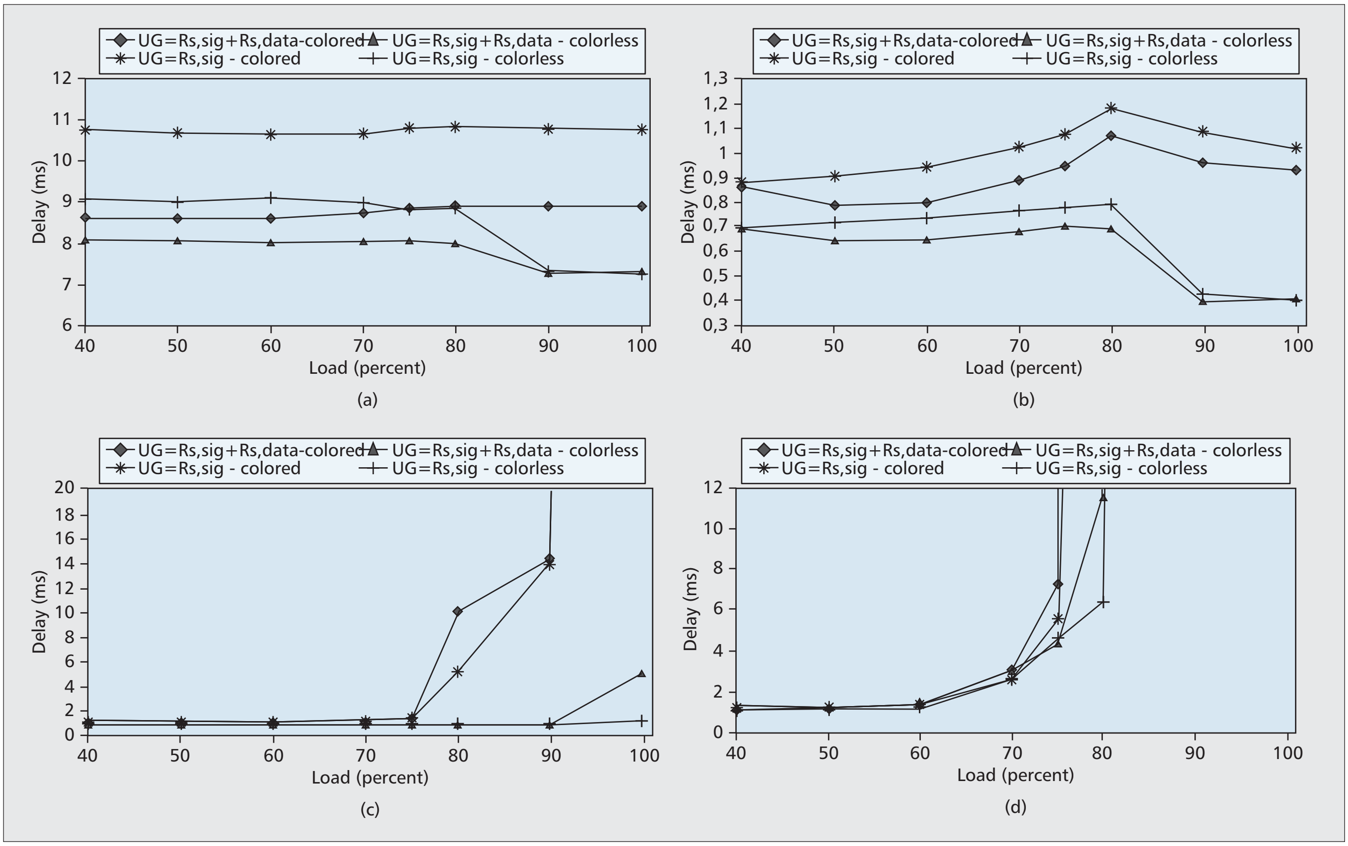

Figure 4. *Average packet delay vs. load, service policy and traffic type: a) handover message exchange; b) CoS1 (BTS voice and residential high priority); c) CoS2; d) CoS3.*

ing. It is worth noting that the GPON can still improve on the latency by using an even lower $D_m = 0.5$ ms without noticeable inefficiency.

Next, attention in this simulation study was directed to fine-tuning strategies on bandwidth allocation, investigating alternative policies more suited to MBH traffic and the consequent performance trade-offs. The core idea is to demonstrate the performance benefits arising from allowing more leeway in queue management to the ONU attached to the BTS than is the practice with the other ONUs. The reason is that this ONU carries traffic from an operator and differs in two major ways from the rest: first, it already has multiplexed traffic from many users, and second, the mobile operator can understand more complex SLAs, enabling better local handling of the queuing process than is possible with the other users. For this reason we propose to aggregate all traffic as one MAC entity belonging to TCONT5 (this allows a mix of priority levels in grant allocations). The characteristic is that no target TCONT queue is specified (only ONU), and it is now left to the ONU to choose which queue to service (also called intra-ONU scheduling in [10, 11]). The use of this approach is left by the standards [4] to the system designer, and its activation (when implemented) is left to the operator. Once such tools are available, it is possible to use the arriving grants locally for higher CoS traffic (whenever such traffic is queued) under the control of the local ONU. In this case requests refer to the aggregate sustainable rate, and the ONU decides which specific queue to serve using local queuing information, which is more current and more responsive to sensitive traffic. This approach, often called a "colorless" grant policy, is in contrast to the alternative where grants target specific queues (and specific CoS classes) already decided by the far away MAC controller at the OLT. The latter (which is by far the most common practice in today's PONs) is called "colored" since the grants are intended for specific target queues (colors) in the ONU and is quite appropriate whenever all customers are plain residential or small businesses with no complex SLA needs. The advantage of the colorless policy in the MBH case is that by leaving the allocation to be decided locally in the ONU, lower latency and better utilization can be achieved (e.g., it and can serve packets that were not even present when the grant was sent).

In addition to color or none, two polling policies were also investigated. In the first one, the polling rate (by means of UG) was set to just that required for the expected signaling rate of the first priority (while the rest used requests made possible by these UGs). This policy is indicated as $R_{s,sign}$ in the result figures. In the other, UGs are issued at a rate equal to the sum of signaling plus the sustainable rate contracted by the SLA and is indicated as $R_{s,sign} + R_{s,data}$. This, of course, refers to the top priority incurring no waste by providing UGs, since they will be used anyway by a packet of any class and will also provide the opportunity to send requests for the lower classes, provoking corresponding grants in a second round since TCONT5 works as an aggregate of many classes of traffic.

In combination, these policies create four alternatives, which were investigated in the simulations. For each one, the queuing delay per individual packet (in addition to the signaling scenario) is measured against increasing total PON load (expressed as percent of its upstream channel capacity). The results are shown in Figs. 4a–d, the first for the signaling exchange and then one for each CoS. A mean D_m of 0.75 ms was used. As expected, the first two classes have an almost steady delay across all loads since they do not feel any competition from the lower priorities and therefore enjoy an always lightly loaded medium. The temporal bursts, when the total offered load temporarily exceeds the available bandwidth, are borne by the lower classes, which, as expected, become unstable before reaching 100 percent total offered load, but at what load strongly depends on the specific MAC policy.

The first observation is that increasing the polling rate with UG equal to both the expected maximum signaling traffic plus the sustainable rate improves delay performance, but this comes at the expense of utilization. This is to be expected as the resulting denser polling reduces latency, but a lot of these UGs go unused, thus wasting bandwidth. Clearly, a trade-off is needed, but there is no straightforward solution, so it is worth elaborating further on the UG rate choice.

To understand the incentive for an elaborate UG policy, one must approach the issue from the operator's perspective. What is really needed is a way for the operator to predict the volume of signaling traffic in order to pre-allocate enough bandwidth via polling to guarantee a lower than maximum tolerable latency. Since, in real life, the generation of signaling traffic is unpredictable, it is natural to consider overprovisioning thus allocating the UG rate equal to the expected worst-case peak rate of the signaling traffic. However, in that case, the available bandwidth is underutilized during periods of low signaling load. It then follows naturally that an improvement can be reached by multiplexing signaling and no signaling high-priority traffic into the same queue. In that case, the unused bandwidth of signaling traffic is allocated to data traffic, resulting in better efficiency, without at all compromising the critical latency and delay of the signaling traffic, or that of first-priority traffic, since the weak point of the PON lies in initiating transmissions from an initially empty queue and not for continuing service on a queue (which takes place by the chain of requests). Exploiting this idiosyncrasy of the TDMA PON, we propose this strategy (i.e., $R_{s,sign} + R_{s,data}$) by pre-allocating bandwidth equal to the sustainable rate of signaling traffic plus the sustainable rate of high-priority traffic. As can be seen from Figs. 4a–d, this policy outperforms the mode ($R_{s,sign}$) in terms of signaling delay under both low and high loads. On the other hand, as seen from Fig. 4, using ($R_{s,sign} + R_{s,data}$) in case of high loads drives CoS2 and CoS3 into high delay values at high load. The solution to this problem is the use of *colorless mode* instead of *colored mode*.

As seen from the same Fig. 4a–d, the colorless policy gives consistently better results in all cases and its adoption is recommended. This is to be expected since the OLT has limited knowledge of the local situation in comparison with a centralized multiplexer, which instantly knows all queue lengths. Unable to have this knowledge, one should at least delegate the remote multiplexing enacted by the PON MAC protocol controller to the local ONU (unfortunately with the limited scope of the local queues), thus improving performance. This is particularly useful among the different priority queues of the ONU, resulting in the obviously useful effect of high-priority queues "stealing" grants directed to lower priority, forcing the latter to report the same packet again in their request, suffering no real harm since they are delay-tolerant.

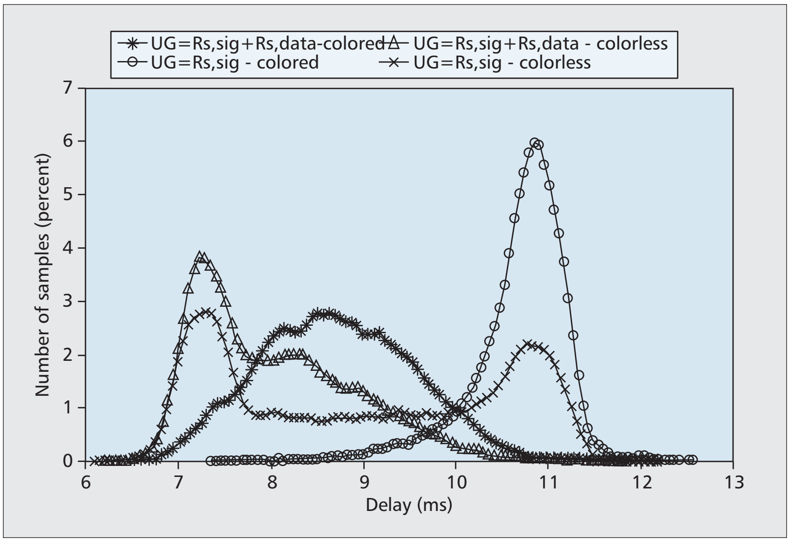

Figure 5. *Comparison of colored vs. colorless mode.*

This warrants a more careful look into the colorless mode, and this is provided in Fig. 5, which depicts the PDF of signaling delay at a high total offered load of 90 percent and a D_m of 0.75 ms.

The superior delay performance of the colorless policy is also evident in Fig. 5, which shows the frequency of delay values. The more values to the left of the figure, the lower the delay. So the best performance is achieved by colorless with UG = $R_{s,sign} + R_{s,data}$ (little triangles) with a peak below 7.5 ms, while the same UG rate but with colored policy (asterisks) gives a more even distribution but clearly moved to the right (higher delay values). When the color and colorless policies are compared for UG = $R_{s,sign}$, again the colored one (little circles) gives higher delay, peaking near 11 ms, while the colorless (x marked) has two distinct peaks, one again near 11 ms but much lower and another just above 7 ms. As expected, on the other hand, the higher UG rate gives lower delay for both the colored and colorless cases. The appearance of two peaks shows the importance of the polling rate (which in this case coincides with the D_m) in creating service opportunities. On average each packet waits 1/2 of D_m for the grant and an additional 1 ms for processing, (thus totaling about 7 ms for 5 packets). This is, of course, not fixed, but a distribution around this value that corre-

The widespread deployment of PON systems for fixed communications covering first mile access for residential and small business customers provides a serendipity for MBH that cannot be missed as it offers a smooth migration path both in technical as well as financial terms.

sponds to the first peak. Now some packets miss the first round and need another D_m to get the grant on the next round, giving a second concentration of values around 11 ms.

CONCLUSIONS

The widespread deployment of PON systems for fixed communications covering first mile access for residential and small business customers provides a serendipity for MBH that cannot be missed as it offers a smooth migration path in both technical as well as financial terms. Although the TDMA technique will exhaust itself at some point, PONs still constitute a future-proof solution because of their ability to accommodate WDM extensions without further fiber laying or other costly operations. However, the TDMA aspect presents certain peculiarities, and a careful traffic management by the operator is needed. As demonstrated in this article, the added access delay jeopardizes specified limits for sensitive services. Also, quantitative assessment showed that this can be improved by delegating more multiplexing decisions to the local ONU of the mobile BTS, while aggregating traffic for several flows relying on TCONT5. This policy carries distinct advantages in terms of latency and delay bounding for sensitive traffic without sacrificing efficiency under high load. This is particularly useful in the EPON case since it does not possess the better frame fill level afforded by GPON because of its tighter encapsulation thanks to frame fragmentation.

ACKNOWLEDGMENTS

This work was supported by the 09S-71-698 CONFES project, partially funded by the Greek General Secretariat of Research and Technology (GSRT).

The authors would like to thank the anonymous reviewers, who greatly helped to improve the accuracy of the manuscript.

REFERENCES

[1] Ed. Mallete, “Mobile Backhaul for PON: A Case Study,” *Proc. OFC*, NTh4E.4, 2012.
[2] K. Grobe and J.-P. Elbers,“PON in Adolescence: From TDMA to WDM-PON,” *IEEE Commun. Mag.*, vol. 46, no. 1, Jan. 2008, pp. 26–34.
[3] J. Angelopoulos *et al.*, “Efficient Transport of Packets with QoS in an FSAN-Aligned GPON,” *IEEE Commun. Mag.*, vol. 42, 2004, pp. 92–98.
[4] ITU-T, Rec. G.984.3, “Gigabit-capable Passive Optical Networks (G-PON): Transmission Convergence Layer Specification,” Mar. 2008.
[5] IEEE Std 802.3ah-2004, 2004.
[6] ITU-R M.2134, “Requirements Related to Technical Performance for IMT-Advanced Radio interface(s),” 2008.
[7] GSM 04.18, “Digital Cell Telecommunications System (Phase 2+); Mobile Radio Interface Layer 3 Specification, Radio Resource Control Protocol.”
[8] 3GPP TS 23.009 V5.6.0, “Handover Procedures,” tech. spec., TS Group Core Network.
[9] T. Orphanoudakis *et al.*, “Performance Evaluation of GPON vs EPON for Multi-Service Access,” *Int’l. J. Commun. Sys. 2009*, vol. 22, pp. 187–202.
[10] A. Shami *et al.*, “Jitter Performance in Ethernet Passive Optical Networks,” *IEEE J. Lightwave Tech.*, vol. 23, no. 4, Apr. 2005, pp. 1745–53.
[11] C. M. Assi *et al.*, “Dynamic Bandwidth Allocation for Quality-of-Service Over Ethernet PONs,” *IEEE JSAC*, vol. 21, no. 9, Nov. 2003.
[12] S. Lallukka and P. Raatikainen , “Link Utilization and Comparison of EPON and GPON Access Network Cost,” *Proc. IEEE GLOBECOM 2005*, St. Louis, MO, 28 Nov.–2 Dec. 2005.
[13] J. Chung *et al.*, “Application of 2W-PON to the Backhaul of A Wireless Network,” *J. Opt. Net.*, vol. 7, no. 11, Nov. 2008 , pp. 905–13.

BIOGRAPHIES

THEOFANIS G. ORPHANOUDAKIS [S’93, M’98] (fanis@uop.gr) received his Dipl-Ing. degree in electrical and computer engineering in 1995 and his Ph.D. in telecommunications in 1998, both from the National Technical University of Athens (NTUA). Since 2010 has been elected as a lecturer at the University of Peloponnese. Prior to joining the university he was with Ellemedia Technologies as a technical manager. His research interests include design and implementation of high-speed networks and systems, performance evaluation of MAC protocols, traffic engineering, and network processor architectures. He has led several research and development projects in the above areas, and has published over 60 papers in international journals and conferences.

EVAGGELOS KOSMATOS received Dipl.Ing. and Ph.D. degrees, both in electrical and computer engineering, from NTUA, Athens, Greece, in 2002 and 2007, respectively. His research interests lie in the area of network control plane design, resource allocation in broadband networks, and intelligent multimedia content delivery. He has participated in several EU and national research projects in the above areas.

JOHN D. ANGELOPOULOS [S’92, M’95] received his electrical engineering degree from NTUA in 1973, his M.Sc. degree in automatic control from Nottingham University in 1977, and his Ph.D. degree in telecommunications from NTUA in 1992. He is currently a professor in the Automation Engineering Department of the Technological Institute of Piraeus. His research interests cover most aspects of high-speed communications with special focus on shared medium access protocols and implementations.

ALEXANDROS STAVDAS [M’00] received his B.Sc. degree in physics from the University of Athens, Greece, his M.Sc. degree in optoelectronics and laser devices from Heriot-Watt University/St Andrews University, Edinburgh, United Kingdom, and his Ph.D. degree in wavelength-routed wavelength-division multplexed networks. He is currently a professor of optical networking with the Department of Telecommunications Science and Technology, University of Peloponnese, Tripolis, Greece. His current interests include physical-layer modeling of optical networks, ultra-high-capacity end-to-end optical networks, optical crossconnect architectures, and optical packet switching.