Optical Switching and Networking 12 (2014) 1–13

Contents lists available at ScienceDirect

Optical Switching and Networking

journal homepage: www.elsevier.com/locate/osn

Impact of report message scheduling (RMS) in 1G/10G EPON and GPON

Anu Mercian [a], Michael P. McGarry [b,1], Martin Reisslein [a,*]

[a] School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287-5706, United States
[b] Department of Electrical and Computer Engineering, University of Texas at El Paso, 500 W University Ave, El Paso, TX 79968, United States

ARTICLE INFO

Article history:
Received 26 July 2013
Received in revised form
4 November 2013
Accepted 19 November 2013
Available online 4 December 2013

Keywords:
Dynamic bandwidth allocation (DBA)
Long-reach PON (LRPON)
Multi-thread polling (MTP)
Passive optical networks (PON)

ABSTRACT

A wide array of dynamic bandwidth allocation (DBA) mechanisms have recently been proposed for improving bandwidth utilization and reducing idle times and packet delays in passive optical networks (PONs). The DBA evaluation studies commonly assumed that the report message for communicating the bandwidth demands of the distributed optical network units (ONUs) to the central optical line terminal (OLT) is scheduled for the end of an ONU's upstream transmission, after the ONU's payload data transmissions. In this paper, we conduct a detailed investigation of the impact of the report message scheduling (RMS), either at the beginning (i.e., before the pay load data) or the end of an ONU upstream transmission on PON performance. We analytically characterize the reduction in channel idle time with reporting at the beginning of an upstream transmission compared to reporting at the end. Our extensive simulation experiments consider both the Ethernet Passive Optical Networking (EPON) standard and the Gigabit PON (GPON) standard. We find that for DBAs with offline sizing and scheduling of ONU upstream transmission grants at the end of a polling cycle, which processes requests from all ONUs, reporting at the beginning gives substantial reductions of mean packet delay at high loads. For high-performing DBAs with online grant sizing and scheduling, which immediately processes individual ONU requests, or interleaving of ONUs groups, both reporting at the beginning or end give essentially the same average packet delays.

1. Introduction

Passive optical networks (PONs) have emerged over the past decade as a highly promising access network technology for connecting individual distributed optical network units (ONUs) at distributed subscriber premises to a central optical line terminal (OLT), see Fig. 1, [1–9]. Recent advances in the underlying photonic and physical layer communications technologies and commensurate standardization efforts have paved the way for PONs operating at a channel bandwidth of 10 Gbps (compared to the 1 Gbps bandwidth considered in early PON development), cf. IEEE 802.3av [10] and G.987 [11]. Also, long-reach PONs operating up to distances of 100 km between the distributed ONUs and the central OLT have emerged [12–19]. Operating at high bandwidth over long distances, i.e., with a high bandwidth-delay product, poses significant challenges for coordinating the upstream transmissions of the distributed ONUs so as to avoid collisions on the shared upstream (from ONUs to OLT) channel. As reviewed in Section 2, a wide array of dynamic bandwidth allocation (DBA) mechanisms have been developed to solve this medium access control problem on the upstream channel for bursty ONU packet traffic.

The DBA mechanisms operate commonly within the context of the standardized signaling mechanisms for PONs which are based on a cyclical report-grant polling structure,

* Corresponding author. Tel.: +1 480 965 8593; fax: +1 480 965 8325.
E-mail addresses: amercian@asu.edu (A. Mercian), mpmcgarry@utep.edu (M.P. McGarry), reisslein@asu.edu (M. Reisslein).
[1] Tel.: +1 915 747 6955; fax: +1 915 747 7871.

1573-4277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.osn.2013.11.004

which is illustrated in Fig. 2. More specifically, ONUs signal their queue depths, i.e., current bandwidth demands, with a control (report) message to the OLT. The OLT then sets the sizes (lengths) of the upstream transmission windows (grants) for the individual ONUs and signals the length and starting time (schedule) of each transmission window to the individual ONUs through grant messages, which are represented by the downward arrows in Fig. 2. In particular, the Ethernet PON (EPON) employs the Report and Gate messages of the Multi-point Control Protocol (MPCP) according to the IEEE 802.3ah or 802.3av standards. The Gigabit PON (GPON) employs dynamic bandwidth reports upstream (DBRu) for signaling the queue depths and Bandwidth Maps (BWMaps) for signaling the upstream transmission windows following the G.984 or G.987 standards [20].

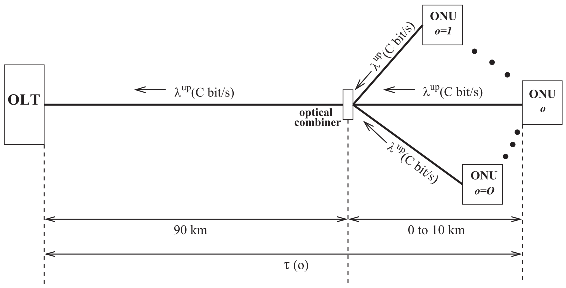

Fig. 1. Illustration of upstream transmission direction from distributed ONUs o, $o = 1, 2, \ldots, O$, to a central OLT in the PON structure. The O ONUs share a single wavelength channel with bit rate C bit/s for their upstream transmissions and have one-way propagation delay $\tau(o)$ to the OLT.

The report message from an ONU is typically lumped together with the upstream payload data transmission so as to avoid extra guard times for the short report message. While the EPON standard leaves the position of the report message within an ONU's upstream transmission open, the vast majority of EPON studies have assumed that the report message is positioned at the very end of an ONU's upstream transmission, after the ONU's payload data transmissions. This "reporting at the end" allows the ONU to signal the most up-to-date queue depth, at ideally the time instant of the end of the payload transmission, to the OLT. On the other hand, the GPON standard specifies that the report message be included at the beginning of the upstream transmission, i.e., to precede the payload data [11]. This "reporting at the beginning" has the advantage that the OLT receives the report message earlier (i.e., before the ONUs payload data) and can already size and schedule the transmission windows for the next cycle. On the downside, the report at the beginning does not contain the packets that were newly generated during the ONU's payload transmission.

To the best of our knowledge the impact of the report message scheduling at the beginning or end of an ONU's

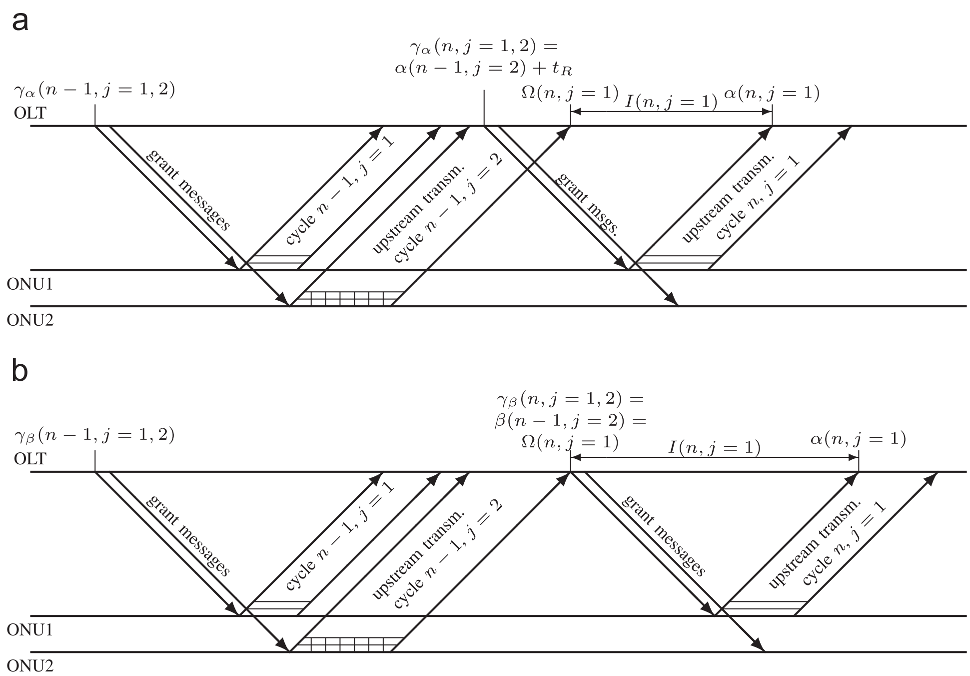

Fig. 2. Illustration of cyclical report-grant polling structure. Grant messages signal upstream transmission windows to the individual ONUs, which report their queue depths in report messages included in the upstream transmissions. The figure also illustrates scheduling at the beginning and at the end of an upstream transmission for $O=2$ ONUs with offline single-thread polling (STP). For illustration of the differences in scheduling for cycle n, scheduling decisions for cycle $n-1$ are assumed to be made at the same time instant, namely at $\gamma_\alpha(n-1, j=1,2) = \gamma_\beta(n-1, j=1,2)$. With scheduling at the beginning, the report is at the beginning of the upstream transmission, allowing the OLT to make the scheduling decision for cycle n at time instant $\gamma_\alpha(n, j=1,2) = \alpha(n-1, j=2) + t_R$, i.e., a report transmission time t_R after the upstream transmission of ONU $j=2$ of cycle $n-1$ begins to arrive at the OLT (at $\alpha(n-1, j=2)$). With scheduling at the end, the report is at the end of the upstream transmission, thus, the OLT can make the scheduling decision for cycle n at time instant $\gamma_\beta(n, j=1,2) = \beta(n-1, j=2)$, i.e., when the end of the upstream transmission of ONU $j=2$ arrives at the OLT. (a) Scheduling at the beginning: report message included at the beginning (left side) of an upstream transmission, (b) scheduling at the end: report message included at the end (right side) of an upstream transmission.

upstream transmission has not yet been investigated in detail. In this paper we examine this open research question in the context of both EPONs and GPONs operating at either 1 Gbps or 10 Gbps channel bandwidth for state-of-the-art DBA mechanisms. We analyze the channel idle time with reporting at the beginning or end for the different DBA mechanisms. We show that reporting at the beginning can reduce the channel idle time that precedes the arrival of an ONU upstream transmission at the OLT by up to the transmission time of an ONU's payload compared to reporting at the end. We conduct extensive simulations to evaluate the average packet delay. We find that reporting at the beginning significantly reduces the packet delay for DBA mechanisms that accumulate all reports from a cycle for offline transmission window sizing and scheduling. In contrast, DBA mechanisms that size and schedule transmission windows online or employ interleaving strategies for the cyclic polling processes perform equally well for both reporting at the beginning or end.

This paper is structured as follows. Section 2 discusses the background and related work. In Section 3, we analyze the channel idle time for reporting at the beginning and identify the reduction in channel idle time compared to reporting at the end. In Section 4, we present extensive simulation results comparing reporting in the beginning and end for 1 Gbps and 10 Gbps EPON and GPON. In Section 5, we summarize our observations and outline future research directions.

2. Background and related work

2.1. Dynamic bandwidth allocation (DBA)

Efficient control of the access by the distributed ONUs to the shared upstream channel so as to serve bursty traffic with low delay while avoiding collisions is one of the key challenges in operating a PON [21]. Several dynamic bandwidth allocation (DBA) approaches have been developed for this channel access problem. A primary classification criterion for DBA mechanisms is the number of polling threads employed per ONU. Single-thread polling (STP) [22–29] employs one polling thread per ONU, while multi-thread polling (MTP) [14,30–33] employs multiple polling threads. The polling threads may operate in offline fashion, i.e., collect report messages from all ONUs before sizing and scheduling the upstream transmission windows, or make these decisions in online fashion after the receipt of each individual report message [34].

The vast majority of the existing studies on DBA in PONs has considered reporting at the end of the upstream transmission. Reporting at the beginning has only briefly been considered for STP with elementary gated grant sizing in [35] and for MTP in [36]. Also, the channel idle time has so far only been analyzed for reporting at the end of an upstream transmission in [32,37]. The present study provides the first analysis of the channel idle time for reporting at the beginning of an ONU upstream transmission as well as a detailed examination of the impact of the report message scheduling at beginning vs. end of an upstream transmission on the channel idle time and packet delay for a wide range of DBA mechanisms.

2.2. PON standards

PONs with 1 Gbps channel bandwidth were standardized a decade ago as EPON in IEEE 802.3ah [38] and as GPON in ITU-T G.984 [39]. On the other hand, corresponding standards for 10 Gbps channel bandwidth were established only recently as 10G-EPON in IEEE 802.3av [10] and as XG-PON in ITU-T G.987 [11]. As a result, DBA mechanisms for 10 Gbps EPON or XG-PON have yet to receive significant attention from the research community. Several comparisons of the physical layer and link layer overhead differences among the various 1 Gbps and 10 Gbps standards have appeared in the literature [40–43].

Some early investigations of DBA mechanisms for the 10 Gbps standards have been reported in [44–48]. The impact of the polling cycle time in single-thread polling with limited grant sizing on various performance measures, e.g., packet delay and jitter, for each of the 1 Gbps and 10 Gbps standardized PON variants was studied in [44]. Mechanisms to increase TCP throughput for 10 G-EPON were studied in [45]. A modification to an existing DBA algorithm to support a mixed network of both 1 Gbps and 10 Gbps EPON ONUs was proposed in [46]. Efficient utilization of unused bandwidth for XG-PON was investigated in [47,48]. The work presented in this paper augments this relatively small body of literature by investigating the impact of the reporting position on performance measures for each of the 1 Gbps and 10 Gbps standardized PON variants.

3. Analysis of channel idle time

3.1. PON model

In this section we analyze the idle time on the upstream channel of a PON before each upstream transmission. We consider a PON model with a total number of O ONUs, whereby ONU o, $o = 1, 2, \ldots, O$, has a one-way propagation delay of $\tau(o)$ [s] to the OLT. The O ONUs share the upstream wavelength channel with bit rate C [bit/s], as illustrated in Fig. 1. Polling-based medium access control with report-grant cycles is employed to avoid collisions on the shared upstream wavelength channel. We denote n for the polling cycle index and t_g for the guard time, i.e., the minimum required spacing between successive upstream transmissions from different ONUs. Moreover, we denote t_R and t_G for the transmission times of a report and grant message, respectively, as summarized in Table 1.

For DBA mechanisms employing multiple polling threads [14,30–33], we denote Θ for the total number of threads, with θ, $\theta = 1, 2, \ldots, \Theta$, denoting the thread index. The Θ threads operate in parallel, giving each ONU Θ opportunities to report the queue depth and transmit upstream payload data in a polling cycle. Note that $\Theta = 1$ corresponds to single-thread polling. We omit the thread index θ from the model notations for single-thread polling. We denote Z for the maximum cycle duration in terms of the sum (aggregation) of the upstream transmissions of all O ONUs (and all Θ threads) of a given cycle n. A particular grant scheduling policy may arrange the upstream transmission windows of the O ONUs of a given thread θ in

Table 1
PON modeling parameters.

Parameter	Meaning
Network and polling structure	
O	Total number of ONUs, indexed $o=1,2,\ldots,O$
$\tau(o)$	One-way propagation delay from OLT to ONU o (s)
C	Upstream bandwidth (bit/s)
Z	Maximum cycle duration, i.e., max. aggregate duration of upstream transmission windows of all O ONUs (and all Θ threads) in a cycle (s)
t_R	Transmission time of a report message (s)
t_g	Guard time (s)
t_G	Gate transmission time (s)
Cycle, thread, and upstream transmission indices	
n	Polling cycle index
Θ	Total number of threads
θ	Thread index, $\theta=1,\ldots,\Theta$
j	ONU index ordered by upstream transmission position for a given thread θ in a given cycle n, i.e., ONU j has jth upstream transmission grant of thread θ in cycle n
Upstream transmission window (grant) scheduling [all parameters are in units of seconds]	
$\gamma(n,\theta,j)$	Time instant when OLT makes scheduling decision for transmission window of jth ONU of thread θ in cycle n
$T(n,\theta,j)$	Gate signaling delay: time duration from instant of OLT scheduling decision to end of the GATE transmission for jth ONU of thread θ in cycle n plus round-trip propagation delay
$\alpha(n,\theta,j)$	Time instant when upstream transmission of jth ONU of thread θ in cycle n starts to arrive at OLT
$\beta(n,\theta,j)$	Time instant when end of upstream transmission of jth ONU of thread θ in cycle n arrives at OLT
$\Omega(n,\theta,j)$	Time instant of end of upstream transmission preceding arrival of upstream transmission of jth ONU of thread θ in cycle n
$I(n,\theta,j)$	Duration of channel idle time preceding the arrival of upstream transmission of jth ONU of thread θ in cycle n at OLT
$G_{\max}$	$=\frac{Z}{\Theta O}$, Maximum duration of granted upstream transmission window size for Limited grant sizing

a particular order. We use the index j, $j=1, 2, \ldots, O$ to denote the ordering of the ONU upstream transmissions (of a given thread θ) in a given cycle n.

3.2. Timing of reporting at the beginning and end of upstream transmission

We initially consider two report message scheduling approaches, namely reporting at the beginning and reporting at the end of an upstream transmission. With reporting at the beginning, the message indicating the queue depth at the ONU to the OLT is positioned at the beginning of the upstream transmission. Specifically, the report message that contains the queue depth for sizing the upstream transmission of ONU j of thread θ in cycle n begins to arrive at the OLT at time instant $\alpha(n-1,\theta,j)$ and is completely received by time instant $\alpha(n-1,\theta,j)+t_R$. Thus, neglecting processing delays, the OLT can make a scheduling decision based on this received report as early as time instant $\alpha(n-1,\theta,j)+t_R$, as illustrated for offline STP in Fig. 2(a). We denote $\gamma_\alpha(n,\theta,j)$ for the scheduling instant for the upstream transmission of ONU j of thread θ of cycle n with reporting at the beginning, and specify $\gamma_\alpha(n,\theta,j)$ for the different PON scheduling frameworks [34,37] in Section 3.3.

In contrast, with reporting at the end, the report message is positioned at the end of the upstream transmission, i.e., it begins to arrive at the OLT at instant $\beta(n-1,\theta,j)-t_R$ and is completely received by instant $\beta(n-1,\theta,j)$. Thus, the OLT can make grant sizing and scheduling decisions for the upstream transmission of ONU j of thread θ of cycle n as early as time instant $\beta(n-1,\theta,j)$, as illustrated for offline STP in Fig. 2(b). We denote $\gamma_\beta(n,\theta,j)$ for the scheduling instant for the upstream transmission of ONU j of thread θ of cycle n with reporting at the end.

3.3. Scheduling instants with reporting at beginning and end

We consider the following combinations of scheduling (polling) frameworks and grant sizing mechanisms:

- Offline single-thread polling with Gated grant sizing (S, offl., gat.) [23,49].
- Offline single-thread polling with Limited grant distribution (S, offl., lim.) [23,49].
- Offline single-thread polling with Excess grant distribution (S, offl., exc.) [50,51].
- Double Phase Polling with Excess grant distribution and share mechanism (D, exc. shr.) [37,52].
- Online single-thread polling with Limited grant distribution (S, onl. lim.) [23,49].
- Online single-thread polling with Excess grant distribution (S, onl., exc.) [32,50,51].
- Online Multi-thread polling with Excess grant distribution (M, onl. exc.) [32].

With the offline scheduling (polling) framework, reports from all O ONUs must be received before the OLT makes grant sizing and scheduling decisions. Thus, the scheduling instant with S, offl. polling is governed by the arrival of the report from the last ONU in a cycle, i.e., for reporting at the beginning the scheduling instant for the upstream transmission grants of a cycle n coincides with the arrival of the report message at the beginning of the upstream transmission of the last ONU in the preceding cycle $n-1$, $\gamma_\alpha(n,j)=\alpha(n-1,O)+t_R$, as illustrated for $O=2$ ONUs in Fig. 2(a). On the other hand, with reporting at the end, the OLT needs to wait until the end of the upstream transmission of the last ONU in cycle $n-1$ is received before sizing and

Table 2
Scheduling instants $\gamma(n,\theta,j)$ for upstream transmissions of ONU j (of thread θ in multi-thread polling) of cycle n.

Scheduling framework	ONU indices	Rep. at the beg. $\gamma_\alpha(n,\theta,j)=$	Rep. at the end $\gamma_\beta(n,\theta,j)=$
STP, offline	$1\le j\le O$	$\alpha(n-1,O)+t_R$	$\beta(n-1,O)$
DPP	$1\le j\le \frac{O}{2}$	$\alpha\left(n-1,\frac{O}{2}\right)+t_R$	$\beta\left(n-1,\frac{O}{2}\right)$
DPP	$\frac{O}{2}<j\le O$	$\alpha(n-1,O)+t_R$	$\beta(n-1,O)$
STP, online	$1\le j\le O$	$\alpha(n-1,j)+t_R$	$\beta(n-1,j)$
MTP, online	$1\le j\le O$	$\alpha(n-1,\theta,j)+t_R$	$\beta(n-1,\theta,j)$

scheduling the grants for cycle n, i.e., $\gamma_\beta(n,j)=\beta(n-1,O)$, see Fig. 2(b).

Similarly, the scheduling instants of the other scheduling frameworks depend on the arrival of the ONU report message triggering the OLT grant sizing and scheduling either at the beginning or end of the ONU upstream transmission, as summarized in Table 2. Double-phase polling (DPP) schedules the first ONU group with indices $j=1,2,\ldots,O/2$ when the report from ONU $O/2$ is received and the second ONU group when the report from the last ONU O is received. Online single-thread polling (STP) schedules each individual ONU j grant for a cycle n immediately after receipt of the report from ONU j in cycle $n-1$. Similarly, online multi-thread polling schedules each ONU j for a given polling thread θ in a cycle n immediately after receipt of the report of ONU j in thread θ of the preceding cycle $n-1$.

3.4. Summary of idle time analysis

In this section we summarize the analysis of the channel idle time $I(n,\theta,j)$ that precedes the arrival of the upstream transmission of ONU j of thread θ of cycle n at the OLT, which is detailed in the Appendix. The idle time $I(n,\theta,j)$ is the time span (period) from the instant $\Omega(n,\theta,j)$ of the arrival of the end of the preceding upstream transmission at the OLT to the arrival of the beginning of the upstream transmission of ONU j of thread θ in cycle n at time instant $\alpha(n,\theta,j)$ at the OLT, see Fig. 2. That is, the duration of the channel idle time is the difference

$$I(n,\theta,j)=\alpha(n,\theta,j)-\Omega(n,\theta,j). \tag{1}$$

The duration of this idle time span is governed by two constraints:

- Guard time constraint: There must be at least a guard time of duration t_g between the arrival of two successive upstream transmissions at the OLT.
- Signaling constraint: The upstream transmission of ONU j of thread θ of cycle n can arrive no earlier than the gate signaling delay $T(n,\theta,j)$ (transmission time of grant message t_G plus round-trip propagation delay 2τ) after the scheduling instant $\gamma(n,\theta,j)$.

As detailed in the Appendix, the earlier scheduling instant $\gamma_\alpha(n,\theta,j)$ with reporting at the beginning compared to $\gamma_\beta(n,\theta,j)$ for reporting at the end can reduce the channel idle time.

Depending on the combination of guard time and signaling constraints that govern the idle time for the reporting at the beginning and end, reporting at the beginning can reduce the idle time up to the difference between the two scheduling instants, i.e., up to $\gamma_\beta(n,\theta,j)-\gamma_\alpha(n,\theta,j)$.

3.5. Dynamic optimization of report message scheduling

The report message scheduling (RMS) can be dynamically selected for optimization. Reporting at the end (and thus including the packets that have been newly generated during an upstream transmission in the report) can be dynamically selected when reporting at the beginning would not reduce the channel idle time. Based on the detailed analysis in the Appendix, the idle time with offline polling hinges primarily on the reporting of the last ONU ($j=O$) in a cycle. Thus, all but the last ONU, i.e., ONUs $j=1,2,\ldots,O-1$, can report at the end, thus including the newly generated packets in the report, while the last ONU $j=O$ reports at the beginning.

For online scheduling, RMS dynamic selection is not possible. This is because the report schedule decision (beginning or end reporting) would need to be communicated by the OLT to the ONU before the parameters determining the channel idle time reduction Δ_I case in Table 5 are available at the OLT. In particular, to impact the idle time preceding the ONU j transmission arrival in cycle n, the ONU would need to be instructed to report at the beginning or end of the upstream transmission in cycle $n-1$ (when the report determining ONU j's upstream transmission window in cycle n arrives at the OLT). The OLT would need to send out these instructions for reporting at the beginning or end in the preceding cycle $n-2$. However, the queue depth of the preceding ONU $j-1$ used for sizing ONU $j-1$'s window in cycle n (which governs $\Omega(n,j)$) is not yet available at the OLT at that time (as it arrives only shortly before the report from ONU j in cycle $n-1$). Thus, the report scheduling cannot be optimized unless some traffic prediction [53,54] is employed.

4. Simulation results for packet delay

4.1. Simulation set-up

We employ a simulation model of the OLT and ONUs built on CSIM, a discrete event simulator using the C programming language, and validated in preceding studies [32,37]. We implement the LRPON in both EPON and GPON standards for $C=1$ Gbps (IEEE 802.3ah and G.984, respectively) and $C=10$ Gbps (IEEE 802.3av and G.987, respectively), with a total number of $O=32$ infinite-buffer ONUs (ONTs in GPON) placed around the OLT with a constant distance of 90 km from the OLT to the splitter and the ONUs placed randomly in the last 10 km range. The maximum round-trip delay is $2\tau=1$ ms.

We consider self-similar packet traffic with Hurst parameter 0.75 and four different packet sizes with distribution 60% 64 Byte, 4% 300 Byte, 11% 580 Byte, and 25% 1518 Byte packets. The traffic load is defined as long-run average of the payload bit rate.

Control messages for EPON and GPON follow the respective standards. In GPON, the control message is sent periodically every 125 μs. In the EPON, the ONUs report their queue depths with a REPORT message (64 Bytes), while a DBRu (4 Bytes) message is used in the GPON. We set the guard times for EPON $t_g = 1$ μs and for GPON $t_g = 30$ ns.

Simulations are performed for all DBAs noted in Section 3.3 for maximum cycle length $Z=2$, 4, and 8 ms. The maximum grant size for limited grant sizing [23,49], which is the initial basis for excess bandwidth allocation [50,51], is

$$G_{\max} = \frac{Z}{\Theta O}. \quad (2)$$

For MTP, we set the number of threads to $\Theta = 2$ (for consistent comparison with the two ONU groups in DPP [52]) and the threshold for thread tuning to $T_{\text{tune}} = 5$ [32,33].

We observe the average packet delay from the packet generation instant at an ONU to the delivery instant of the complete packet to the OLT. We also observe the average channel idle time $I(n, \theta, j)$.

In Figs. 3 and 7 we plot the average packet delay for all considered combinations of scheduling framework and grant sizing mechanism (see Section 3.3) for all three considered maximum cycle lengths Z for the 1G and 10G EPON respectively. The corresponding average channel idle times are plotted in Figs. 5 and 9. A few scheduling framework-grant sizing combinations were omitted from Figs. 5 and 9 to reduce clutter. Specifically, for reporting at the end, all offline STP approaches give essentially the same average idle times; we plot therefore only offline STP with gated grant sizing while omitting offline STP with limited and excess grant sizing. Moreover, for reporting at the beginning, offline STP with limited grant sizing gives very similar average idle times to offline STP with excess grant distribution (S. offl. exc.); therefore, we only plot S. offl. exc. We omitted online STP with excess grant distribution (S. onl. exc.) which gives very similar average idle times as online STP with limited grant sizing (S. onl. lim.). Due to space constraints, we include for the 1G and 10G GPON only the simulation results for the representative $Z=4$ ms maximum cycle length in Figs. 4, 6, 8, and 10.

4.2. General reporting at beginning vs. end trade-off

We observe across the set of plots in Figs. 3–10 that reporting at the beginning generally gives lower average packet delays and channel idle times than reporting at the end. That is, the effect of the OLT receiving reports earlier with reporting at the beginning and thus making earlier upstream transmission sizing and scheduling decisions generally outweighs the effect of reporting the newly generated packets (generated during an ONU upstream transmission) later (i.e., in the next cycle). The earlier reporting tends to reduce the channel idle time and thus increases the level of masking of idle times, resulting in overall shorter polling cycles and thus lower packet delays. The specific delay reduction effects for the various DBA mechanisms are discussed in detail in the following subsections.

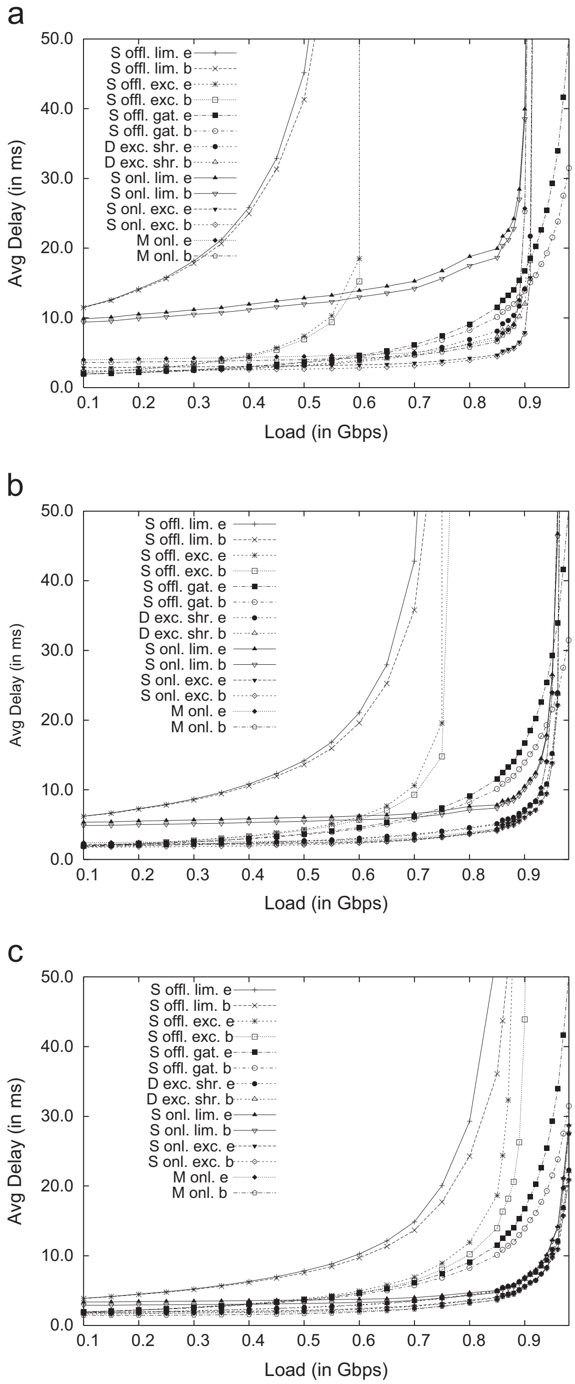

Fig. 3. Mean packet delay for EPON with upstream bandwidth $C=1$ Gbps. Abbreviations for DBA mechanisms (see Section 3.3): Threads: S, single-thread polling; D, double-phase polling; M, multi-thread polling; Scheduling framework: offl., offline; onl., online; Grant sizing: lim., limited; exc., excess distribution; gat., gated; exc. shr., excess share; Report scheduling: e, end; b, beginning. (a) Max. cycle length $Z=2$ ms, (b) Max. cycle length $Z=4$ ms, (c) Max. cycle length $Z=8$ ms.

Before examining the individual DBA mechanisms, we illustrate the effect of the number of ONUs O on the impact of report scheduling. In Table 3, we consider STP with offline gated DBA in a $C=1$ Gbps EPON at traffic load

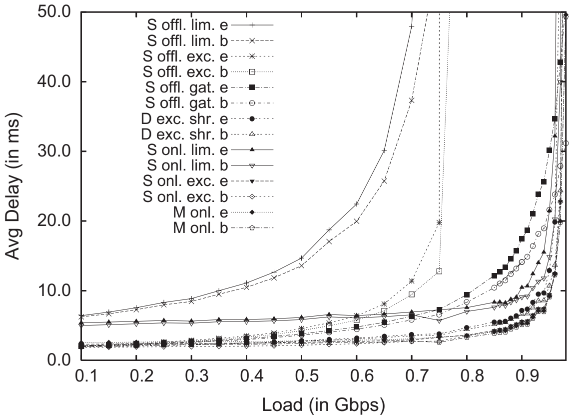

Fig. 4. Mean packet delay for XG-PON with $C = 1$ Gbps and maximum cycle length $Z = 4$ ms.

of 0.9 Gbps. We observe from the table that for a low number of $O = 8$ ONUs, reporting at the beginning reduces the average packet delay almost to half the delay for reporting at the end; whereas for the higher number of $O = 32$ ONUs, the delay reduction with reporting at the beginning is far less pronounced. For the smaller number of ONUs, each ONU upstream transmission window constitutes a relatively larger portion of the overall cycle duration, as illustrated by the average cycle length and average ONU transmission window length values $\overline{G}$ in Table 3.

For reporting at the end, the offline DBA considered in Table 3 has a 2τ channel idle period between successive cycles [34]. Thus, neglecting the guard times t_g and the small variations in the round-trip propagation delays, the average idle time is approximately $2\tau/O$.

Reporting at the beginning masks a portion of this propagation delay equal to the length of the last transmission window in a cycle. Thus, the average idle time is reduced to roughly $(2\tau - \overline{G})/O$. With each transmission window (including the last window in the cycle) constituting a relatively larger portion of the cycle for small O, this masking effect due to reporting at the beginning is significantly more pronounced for small O than for large O. The relatively stronger masking effect for small O leads to significantly more pronounced shortening of the average cycle duration and the average channel idle time, and, in turn, the average packet delay. To summarize, the performance improvements with reporting at the beginning generally are more pronounced in PONs with small numbers of ONUs. However, for the current trend of increasing numbers of ONUs served in a PON, the impact of report scheduling is reduced.

We observe from the results in the "opt." columns in Table 3 that the quantitative benefits from dynamic optimization of the report message scheduling (see Section 3.5) are generally small. Including the newly generated packets in the reporting at the end slightly increases the average transmission window length and cycle length as more packets are included in the ONU reports sent at the end of an upstream transmission. The overall longer cycle length increases also the window of the last ONU, thus increasing it in proportion to the round-trip propagation delay and, in turn, reducing average idle time compared to reporting at the beginning. The combined effects of including the newly generated packets in the end reports and the reduced idle time reduce the average packet delay. It is important to keep in mind though that these effects are relatively small, and this optimization through dynamic RMS selection is limited to offline scheduling.

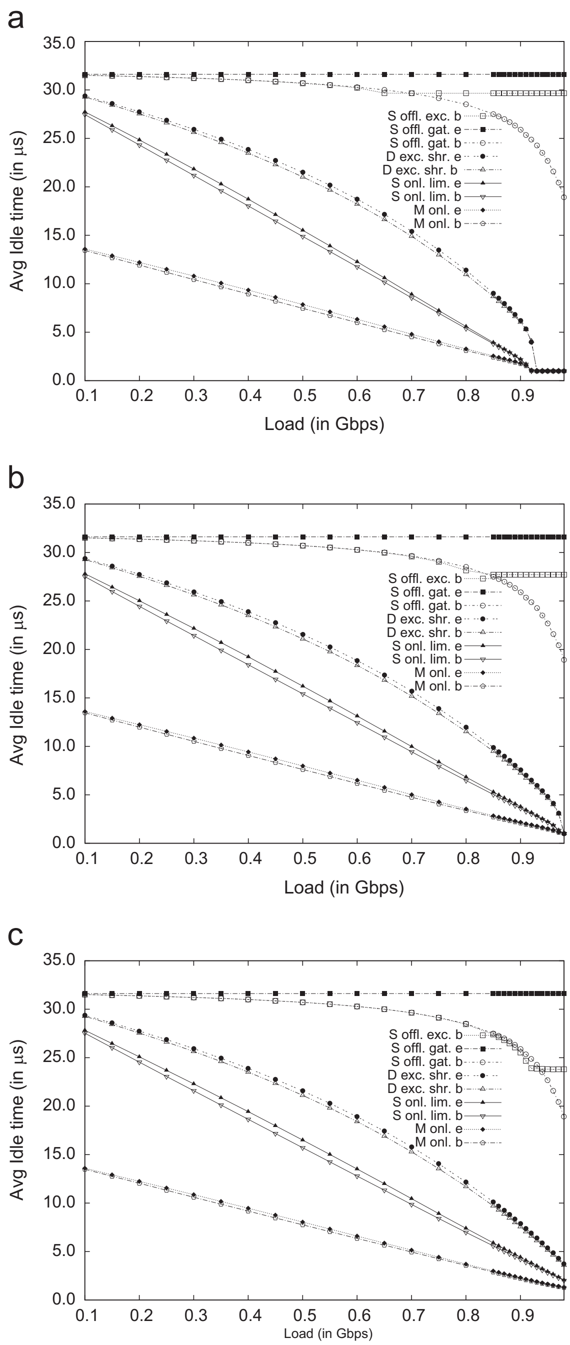

Fig. 5. Mean duration of channel idle time per ONU upstream transmission for EPON with upstream bandwidth $C = 1$ Gbps. (a) Max. cycle length $Z = 2$ ms, (b) Max. cycle length $Z = 4$ ms, (c) Max. cycle length $Z = 8$ ms.

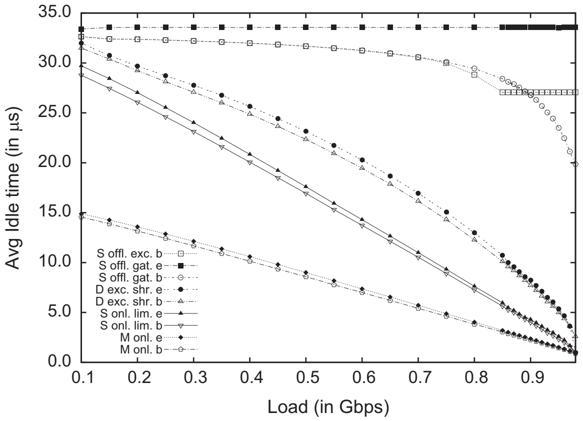

Fig. 6. Mean duration of channel idle time per ONU upstream transmission for XG-PON for $C=1$ Gbps and maximum cycle length $Z=4$ ms.

4.3. *Offline single-thread polling (STP) with limited grant sizing (S offl. lim.)*

We observe from the packet delay plots in Figs. 3, 4, 7 and 8 that offline STP with limited grant sizing gives the highest average packet delay among the compared DBA approaches. This is mainly due to the strict limit $G_{\max}$ on the transmission window length per ONU in a cycle, which results in inflexible bandwidth allocation to the individual ONUs. Offline STP with reporting at the end utilizes a maximum portion of $Z/(2\tau+Z)$ of a cycle for upstream transmissions since the upstream channel is idle during the upstream propagation of the last report of a cycle and downstream propagation of the first grant of the next cycle. That is, from the OLT perspective, the gate signalling delay T from the scheduling instant to the arrival of the corresponding upstream transmission at the OLT is roughly the round-trip propagation delay 2τ (when neglecting the small gate message transmission times).

Examining Figs. 3, 4, 7 and 8 closer for the impact of reporting at the end vs. reporting at the beginning, we observe relatively small delay differences for the short $Z=2$ ms maximum cycle length. For the longer $Z=4$ ms and 8 ms cycle lengths, we observe substantial delay reductions with reporting at the beginning at high traffic loads. These delay reductions can be explained with the average channel idle times plotted in Figs. 5, 6, 9 and 10, as discussed jointly with offline STP with excess bandwidth allocation in the next section.

4.4. *Offline STP with excess bandwidth allocation (S offl. exc.)*

Excess bandwidth allocation [50,51] makes the dynamic bandwidth allocation to the individual ONUs more flexible by redistributing the unused portions of the $G_{\max}$ limit from ONUs with presently low traffic to ONUs that presently have large traffic queues. As a result, the polling cycles become better utilized, which results in substantial delay reductions compared to limited grant sizing, as observed in Figs. 3, 4, 7 and 8.

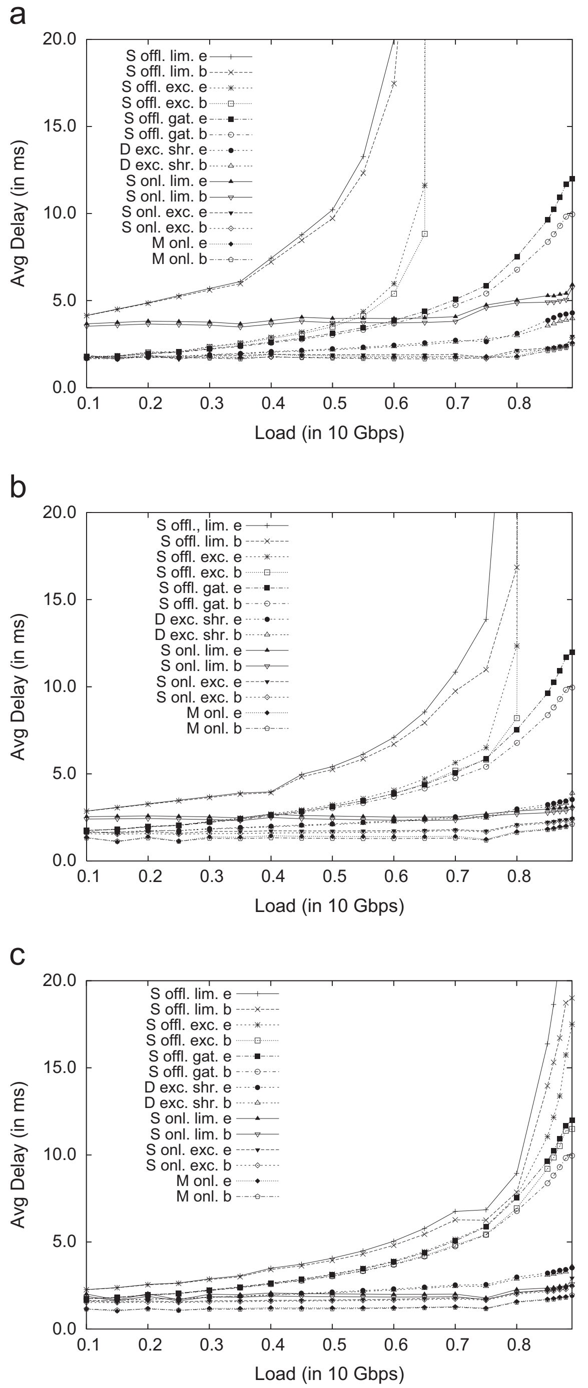

Fig. 7. Mean packet delay for EPON with upstream bandwidth $C=10$ Gbps. (a) Max. cycle length $Z=2$ ms, (b) Max. cycle length $Z=4$ ms, (c) Max. cycle length $Z=8$ ms.

As the traffic load increases, we observe from Figs. 5, 6, 9 and 10 reductions in the average idle time for offline STP with excess grant sizing and reporting at the beginning compared to offline STP with gated grant sizing with reporting at the end (which is plotted as a representative for all offline STP approaches with reporting at the end). As noted above and elaborated in more detail in the

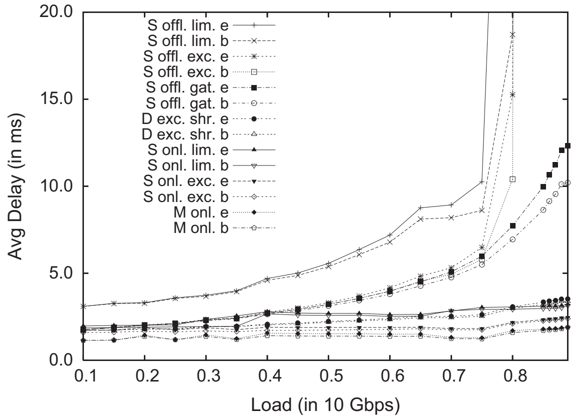

Fig. 8. Mean packet delay for XG-PON with $C=10$ Gbps and maximum cycle length $Z=4$ ms.

Appendix, with offline STP, there is a mandatory 2τ idle time between successive cycles. Thus, the arrival of the first ONU ($j=1$) transmission at the OLT is preceded by a 2τ idle time, while the subsequent ONU transmissions ($j=2,3,\ldots,32$) within the cycle are preceded by a guard time t_g (provided the traffic load and resulting grant lengths are sufficient to mask the propagation delay differences [55]). With reporting at the end, the average idle time per ONU transmission is thus approximately $2\tau/O \approx 31.25$ μs, where we neglect the t_g guard times and consider $2\tau=1$ ms. With reporting at the beginning of the upstream transmission, the idle time is reduced by the length of the upstream transmission of the last ONU in a cycle, which approaches $G_{\max}$ with high traffic load. Thus, the average idle time is reduced to $(2\tau-G_{\max})/O$, which is approximately 27.3 μs for $Z=4$ ms. This reduced average channel idle time per ONU upstream transmission reduces the average packet delay and increases the utilization of the upstream channel.

In additional idle time evaluations which we do not include in the plots to avoid clutter, we found that the differences between reporting at the beginning and reporting at the end of an upstream transmission are very similar for limited grant sizing and for excess grant sizing. The main difference between limited and excess grant sizing is that the average ONU transmission is longer with excess grant sizing, which results in the lower delays observed in Figs. 3, 4, 7 and 8. However, for very high traffic loads with delays beyond the plotted range, both limited and excess grant sizing exhibit the same respective utilization limits of $Z/(2\tau+Z)$ with reporting at the end and $Z/(2\tau-G_{\max}+Z)$ with reporting in the beginning.

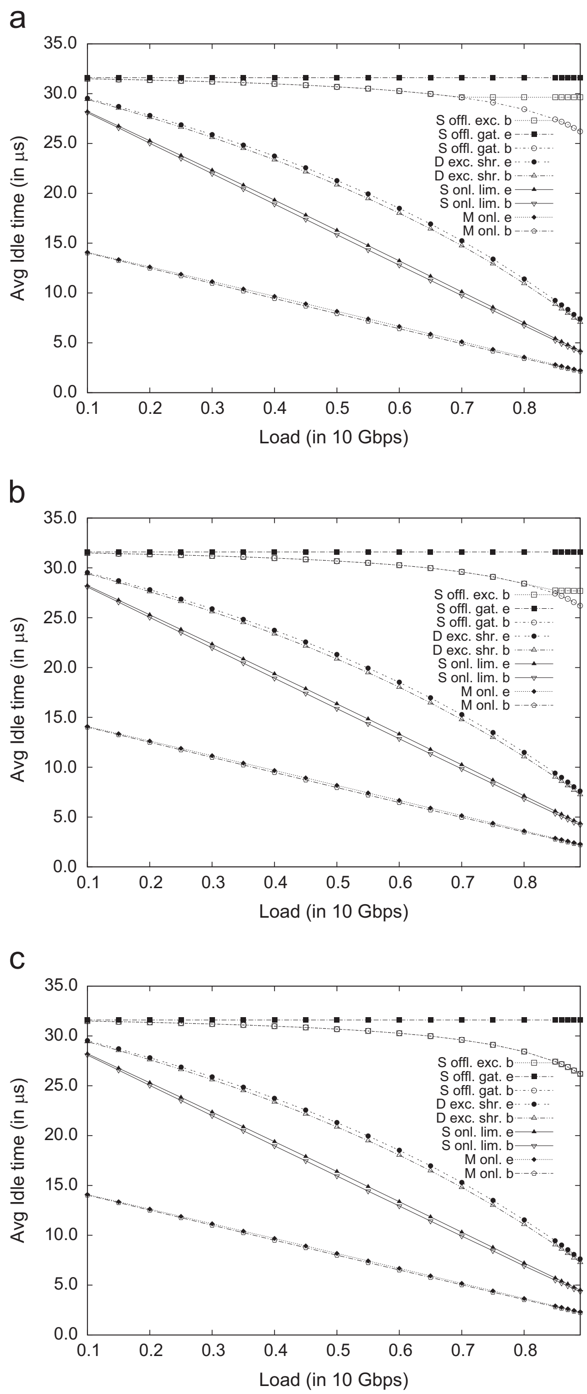

Fig. 9. Mean duration of channel idle time per ONU upstream transmission for EPON with bandwidth $C=10$ Gbps. (a) Max. cycle length $Z=2$ ms, (b) Max. cycle length $Z=4$ ms, (c) Max. cycle length $Z=8$ ms.

4.5. *Offline STP with gated grant sizing (S offl., gat.)*

Gated grant sizing does not limit the lengths of the ONU upstream transmission windows. Thus, for high traffic loads, the window lengths grow very large, substantially larger than $G_{\max}$. Accordingly, we observe in Figs. 5, 6, 9 and 10 a substantially more pronounced reduction of the average channel idle time per ONU transmission for reporting at the beginning with gated grant sizing than with excess grant sizing.

Correspondingly, we observe in Figs. 3, 4, 7 and 8 relatively large reductions of the average packet delay with reporting at the beginning compared to reporting at the end. The delay reduction reaches about 20 ms at the 0.98 Gbps load point in Figs. 3 and 4.

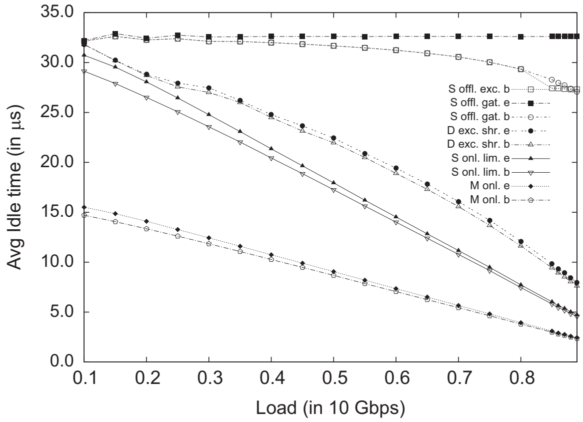

Fig. 10. Mean duration of channel idle time per ONU upstream transmission for XG-PON for $C=10$ Gbps and maximum cycle duration $Z=4$ ms.

4.6. *Online STP with limited and excess grant sizing (S onl. lim. and S onl. exc.)*

We observe from Figs. 3–10 that (a) online STP with excess grant sizing gives substantially smaller delays than online STP with limited grant sizing, and (b) reporting at the beginning gives only very minuscule reductions (on the order of 1–3 ms) in delay compared to reporting at the end for these two DBA approaches. The advantage of excess grant sizing is again due to the more flexible transmission window allocations to the individual ONUs, which more quickly serves their bursty traffic.

By closely examining the online STP with limited grant sizing delay performance across Figs. 3, 4, 7 and 8, we observe delay reductions with increasing maximum cycle length Z and upstream bandwidth C. For instance, we observe from Fig. 3 for $C=1$ Gbps that the average packet delay at traffic load 0.8 is close to 18 ms for $Z=2$ ms, but drops to around 7.5 ms for $Z=4$ ms and further to roughly 4.5 ms for $Z=8$ ms. Similarly, comparing Fig. 3(a) with Fig. 7(a) for $Z=2$ ms, we observe that the higher $C=10$ Gbps bandwidth reduces the average packet delay to less than half of the delays for $C=1$ Gbps. These observed delay reductions are due to the increased limit on the ONU upstream transmission G_{max} (2), which increases the flexibility of the dynamic bandwidth allocation of limited grant sizing. In particular, we observe from Fig. 8(c) that for the largest considered G_{max}, limited grant sizing attains essentially the same average delays as excess grant sizing. Also, the higher channel bit rate reduces the relative impact (in terms of time delay) due to the fixed-size (in terms of Byte count) overheads.

Online STP interleaves the polling processes to the individual ONUs (with a single polling process per ONU), eliminating the 2τ idle period between successive cycles in offline polling. Consequently, there are fewer and smaller opportunities for reducing unmasked idle time by shifting the report message from the end to the beginning of the upstream transmission, as validated by the idle time results in Figs. 5, 6, 9 and 10.

Considering online STP with excess grant sizing more closely, we observe from Figs. 3(a) and Fig. 7(a) that it achieves the smallest average packet delays for the short $Z=2$ ms maximum cycle duration. Whereby, both online STP with excess grant sizing with reporting at the end and with reporting at the beginning achieve similarly low average delays, with reporting at the beginning giving only very minuscule delay reductions for the mid-load range of the $C=1$ Gbps scenario in Fig. 3(a). Indeed, in additional evaluations that are not included in the plots to avoid clutter, we have observed that STP with excess grant sizing has similar average idle times as STP with limited grant sizing. We observe for STP with limited grant sizing from Fig. 5(a) that reporting at the beginning gives only very slight idle time reductions in the mid-load range, while both reporting approaches have essentially the same idle times for the $C=10$ Gbps scenario in Fig. 9(a).

Table 3
Impact of number of ONUs O: Average packet delay, idle time per ONU transmission I, cycle length, and ONU transmission window length $\overline{G}$ for $O=8$ and 32 ONUs; fixed parameters: $C=1$ Gbps EPON, STP offline gated DBA, traffic load $=0.9$ Gbps.

Perf. metric	$O=8$			$O=32$		
	End	Beg.	Opt.	End	Beg.	Opt.
Avg. pkt. del. (ms)	15.1	8.8	8.7	16.7	13.9	13.0
Avg. idl. tim. (μs)	123	67.5	61.9	31.6	25.9	22.6
Avg. cyc. len. (ms)	8.2	4.5	5.0	8.0	7.1	7.7
Avg win. len. $\overline{G}$ (ms)	0.90	0.49	0.55	0.22	0.19	0.20

4.7. *Double-phase polling (D exc. shr.)*

Double-phase polling (DPP) with excess sharing has slightly higher delays and noticeably longer idle times than online STP with excess grant sizing throughout the scenarios considered in Figs. 3–10. This is mainly because DPP employs offline scheduling based on two ONU groups. That is, the offline polling processes to the two ONU groups are interleaved, thus striving to mask the long 2τ idle period of offline scheduling. This strategy is quite effective, as illustrated by the dramatically lower packet delays and idle times compared to the offline polling approaches. In fact, the average delays of DPP approach quite closely those of online STP, but online STP achieves just a little bit lower average delays mainly due to its more extensive interleaving of the online polling processes to the individual ONUs.

The reporting strategy, reporting at the beginning or at the end of the ONU transmission, has essentially negligible impact on both the average packet delays and the idle times. This is mainly because the masking of idle times with the interleaving of the two ONU polling groups is quite effective. Further improving the interleaving by allowing an ONU group to proceed with the scheduling earlier, i.e., after receiving the last report message of the group at the beginning of the last ONU transmission of the group versus the end of the last ONU transmission has a very minor impact.

4.8. *Online MTP (M onl.)*

We observe from Fig. 3(a) that for the short $Z=2$ ms cycle length in the EPON, online MTP gives slightly higher delays than online STP with excess allocation. In all other plots, online MTP attains the smallest average packet delays. We observe from Figs. 7(b) and (c) and 8 that for the higher speed $C=10$ Gbps and longer $Z=4$ and 8 ms cycle lengths, online MTP achieves slightly lower delays than online STP with excess allocation. We also observe from these delay plots, as well as the idle time plots in Figs. 5, 6, 9 and 10 that reporting at the beginning gives very minor or no improvements compared to reporting at the end in online MTP.

Online MTP exploits the interleaving of the polling processes to the individual ONUs through the online scheduling framework as well as the interleaving of multiple polling threads for each ONU. Due to the multiple polling processes, i.e., more frequent polling, the average upstream transmission window lengths with MTP are typically smaller than with STP [32,33]. Shifting the reporting from the end to the beginning of an upstream transmission constitutes therefore a smaller shift of the report message compared to STP with its longer transmission windows. In addition, the multiple levels of interleaving in online MTP leave little unmasked idle times that could be shortened by shifting the report message to the beginning.

5. Conclusion

We have examined the effects of report message scheduling, specifically, scheduling the report message at the beginning or at the end of the upstream transmission of a optical network unit (ONU) in a passive optical network (PON). We have examined these two extreme positions of the report message (beginning or end of the upstream transmission) for a wide range of dynamic bandwidth allocation (DBA) mechanisms in an Ethernet PON (EPON) and Gigabit PON (GPON) for both 1 Gbps and 10 Gbps upstream channel bandwidth. Aside from providing insights into the effects of report message scheduling, this study provides insights into the performance of a wide range of DBA approaches at the 10 Gbps channel bandwidth for long-reach PONs (LRPONs). Most prior studies have only considered the 1 Gbps bandwidth.

We have found that report scheduling at the beginning achieves significant reductions of channel idle time and average packet delays for DBAs with the offline scheduling framework that requires reports from all ONUs before sizing and scheduling the upstream transmission windows for the next polling cycle. This is accomplished by reducing the unmasked idle time period, which is one round-trip propagation delay 2τ for reporting at the end, by the duration of the payload transmission time of one ONU through reporting at the beginning.

DBA approaches with short or few unmasked idle times provide little opportunity for increasing the masking of idle time through shifting the position of the report message. Thus, we observed that online single-thread polling (STP) that interleaves polling processes to the individual ONUs, double-phase polling (DPP) [52] that interleaves offline polling processes to two ONU groups, as well as online multi-thread polling (MTP) [32] are largely insensitive to the report scheduling.

There are several important direction for future research on effective dynamic bandwidth allocation for PON access networks. One direction is to integrated the PON DBA mechanisms with access networks involving other transmission media [7,56], such as wireless networks [57–62]. Another direction is to streamline the internetworking between access networks and metro/wide area networks, through specific network integration and internetworking mechanisms.

Table 4
Time instant $\Omega(n,\theta,j)$ of end of arrival of upstream transmission preceding the arrival of upstream transmission of ONU j of thread θ in cycle n at OLT.

Scheduling framework	Thread and ONU indices	$\Omega(n,\theta,j)=$
STP (both offl. and online)	$j=1$	$\beta(n-1,O)$
	$2\le j\le O$	$\beta(n,j-1)$
MTP	$\theta=1; j=1$	$\beta(n-1,\Theta,O)$
	$2\le\theta\le\Theta;\ j=1$	$\beta(n,\theta-1,O)$
	$1\le\theta\le\Theta;\ 2\le j\le O$	$\beta(n,\theta,j-1)$

Appendix A. Analysis of channel idle time

In this appendix, we build on the idle time analysis for reporting at the end for single-thread polling and DPP in [37] as well as the analysis for reporting at the end for multi-thread polling in [32] to analyze the idle time for both single- and multi-thread polling with reporting at the beginning. We then analyze the reduction of the idle achieved by reporting at the beginning.

Note from (1) that the channel idle time is the difference between the instant $\alpha(n,\theta,j)$ when the beginning of the upstream transmission of ONU j of thread θ in cycle n starts to arrive at the OLT and the instant $\Omega(n,\theta,j)$ when the end of the preceding ONU transmission arrives at the OLT. We first determine $\Omega(n,\theta,j)$ for the various combinations of scheduling frameworks and ONU indices, as summarized in Table 4. For single-thread polling, both with online and offline scheduling, the transmission of the last ONU $j=O$ of the preceding cycle $n-1$ precedes the arrival of the transmission of the first ONU $j=1$ in cycle n, i.e., $\Omega(n,j=1)=\beta(n-1,O)$. In turn, the arrival of the transmission of ONU $j=1$ in cycle n precedes the arrival of the transmission of ONU $j=2$; in general for the ONUs "within" a single-thread polling cycle, the arrival of the transmission from ONU $j-1$ precedes the arrival of the transmission from ONU j, $j=2,3,\ldots,O$. For multi-thread polling, the transmission of the last ONU $j=O$ in the last thread $\theta=\Theta$ of a cycle $n-1$ precedes the arrival from the first ONU $j=1$ of the first thread $\theta=1$ of the subsequent cycle n. The first transmission of each subsequent thread $\theta=2,\ldots,\Theta$ within cycle n is preceded by the last transmission $j=O$ of the preceding thread $\theta-1$. The second and subsequent transmissions $j=2,3,\ldots,O$ within a thread are preceded by the preceding ONU transmission $j-1$.

As outlined in Section 3.4, the idle time constraint and the signaling constraint determine the arrival time instant $\alpha(n,\theta,j)$ of ONU transmission j of thread θ in cycle n at the OLT. Specifically, the idle time constraint requires that the arrival instant $\alpha(n,\theta,j)$ is no earlier than a guard time t_g after the end of the arrival of the preceding transmission at instant $\Omega(n,\theta,j)$. The signaling constraint imposes the gate signaling delay $T(n,\theta,j)$ between the scheduling instant

Table 5
Summary of cases for reduction Δ_I of channel idle time with report scheduling at the beginning compared to report scheduling at the end of an ONU upstream transmission.

Case	$\Delta_I =$
$\gamma_\beta(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)\le t_g$	0
$\gamma_\beta(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)> t_g$	$\gamma_\beta(n,\theta,j)+T(n,\theta,j)$
$\gamma_\alpha(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)< t_g$	$-\Omega(n,\theta,j)-t_g<\gamma_\beta-\gamma_\alpha$
$\gamma_\alpha(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)\ge t_g$	$\gamma_\beta-\gamma_\alpha$

$\gamma(n,\theta,j)$ of the ONU transmission and its arrival at the OLT. Thus,

$$\alpha(n,\theta,j)=\max\{\Omega(n,\theta,j)+t_g,\gamma(n,\theta,j)+T(n,\theta,j)\}. \quad (3)$$

Inserting the expression (3) for $\alpha(n,\theta,j)$ into Eq. (1) for evaluating the channel idle time gives

$$I(n,\theta,j)=\max\{t_g,\ \gamma(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)\}. \quad (4)$$

We evaluate the reduction of the channel idle time with report scheduling at the beginning compared to reporting at the end as

$$\Delta_I=I_\beta(n,\theta,j)-I_\alpha(n,\theta,j) \quad (5)$$

$$\Delta_I=\max\{t_g,\ \gamma_\beta(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)\}$$
$$-\max\{t_g,\ \gamma_\alpha(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)\}. \quad (6)$$

For the further analysis of (6), note that the scheduling instant with reporting at the end $\gamma_\beta(n,\theta,j)$ is always later (or at the same time) than the scheduling instant with reporting at the beginning $\gamma_\alpha(n,\theta,j)$. Specifically, these two time instants are the same, when the corresponding upstream transmission carries only the report message and no payload data. If the upstream transmission carries some payload data, then these two time instants are separated by the transmission time for the carried payload data. Thus,

$$\gamma_\beta(n,\theta,j)\ge\gamma_\alpha(n,\theta,j). \quad (7)$$

Considering (6), there are three cases for evaluating the channel idle time. First, in case

$$\gamma_\beta(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)\le t_g, \quad (8)$$

(7) implies that also

$$\gamma_\alpha(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)\le t_g. \quad (9)$$

Thus, both maxima in (6) are attained by t_g and the resulting reduction in channel idle time is zero, as summarized in Table 5. Intuitively, this first case occurs if the preceding ONU transmission, of which the end arrives to the OLT at $\Omega(n,\theta,n)$ is sufficiently long to mask the signaling time for the transmission of ONU j of thread θ in cycle n.

The other extreme case is that

$$\gamma_\alpha(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)\ge t_g, \quad (10)$$

which implies by (7) that also

$$\gamma_\beta(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)\ge t_g. \quad (11)$$

Hence, both maxima in (6) are attained by the terms involving the scheduling instants γ; specifically, $\Delta_I=\gamma_\beta-\gamma_\alpha$. That is, the reduction in channel idle time is equal to the duration of the transmission time of the payload of the ONU transmission. This case occurs if the signaling delay for ONU transmission j of thread θ in cycle n is not masked by the preceding ONU transmission. Such an unmasked idle time can occur if the preceding ONU transmission is too short to mask the gate signalling delay. Or the polling structure introduces a mandatory idle time that cannot be masked by a preceding transmission. For instance, offline scheduling requires the receipt of the report from the last ONU transmission ($j=O$) in cycle $n-1$ before sizing and scheduling the grants for cycle n. That is, the first ONU transmission ($j=1$) in a cycle n is preceded by the last ONU transmission of the preceding cycle and consequently, the time instant of the end of the preceding ONU transmission is $\Omega(n,1)=\beta(n-1,O)$, see Table 4, which coincides with the scheduling instant $\gamma(n,1)$, see Table 2 for reporting at the end. The resulting idle period (4) is the gate signalling delay $T(n,1)$ which equals one gate transmission time t_G (typ. negligible) and the round-trip time 2τ. This idle time can be reduced through shifting the report message to the beginning. Specifically, the unmasked idle time can be reduced by $\Delta_I=\gamma_\beta-\gamma_\alpha$, i.e., the transmission time for the payload in the last ONU ($j=O$) transmission in cycle $n-1$.

The intermediate case is that

$$\gamma_\beta(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)>t_g, \quad (12)$$

while

$$\gamma_\alpha(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)<t_g. \quad (13)$$

In this case, the maximum in the first line of (6) is attained by the term involving γ_β, while the maximum in the second line is attained by t_g. Thus,

$$\Delta_I=\gamma_\beta(n,\theta,j)+T(n,\theta,j)-\Omega(n,\theta,j)-t_g, \quad (14)$$

which by (13) is less than $\gamma_\beta-\gamma_\alpha$.

References

[1] C.-H. Chang, N. Alvarez, P. Kourtessis, R. Lorenzo, J. Senior, Full-service MAC protocol for metro-reach GPONs, IEEE/OSA J. Lightwave Technol. 28 (7) (2010) 1016–1022.
[2] J. Chen, L. Wosinska, M.N. Chughtai, M. Forzati, Scalable passive optical network architecture for reliable service delivery, IEEE/OSA J. Opt. Commun. Netw. 3 (9) (2011) 667–673.
[3] A. Dixit, B. Lannoo, G. Das, D. Colle, M. Pickavet, P. Demeester, Flexible TDMA/WDMA passive optical network: energy efficient next-generation optical access solution, Opt. Switch. Netw. 10 (4) (2013) 491–506.
[4] F. Effenberger, T. El-Bawab, Passive optical networks (PONs): past, present, and future, Opt. Switch. Netw. 6 (3) (2009) 143–150.
[5] F. Aurzada, M. Scheutzow, M. Reisslein, N. Ghazisaidi, M. Maier, Capacity and delay analysis of next-generation passive optical networks (NG-PONs), IEEE Trans. Commun. 59 (5) (2011) 1378–1388.
[6] K. Kerpez, Y. Luo, F. Effenberger, Bandwidth reduction via localized peer-to-peer (P2P) video, Int. J. Dig. Multim. Broadcast. (2010), http://dx.doi.org/10.1155/2010/562832.
[7] G. Kramer, M. De Andrade, R. Roy, P. Chowdhury, Evolution of optical access networks: architectures and capacity upgrades, Proc. IEEE 100 (5) (2012) 1188–1196.
[8] M. Mahloo, C. Machuca, J. Chen, L. Wosinska, Protection cost evaluation of WDM-based next generation optical access networks, Opt. Switch. Netw. 10 (1) (2013) 89–99.
[9] G. Sankaran, K. Sivalingam, ONU buffer reduction for power efficiency in passive optical networks, Opt. Switch. Netw. 10 (4) (2013) 416–429.
[10] K. Tanaka, A. Agata, Y. Horiuchi, IEEE 802.3av 10G-EPON standardization and its research and development status, IEEE/OSA J. Lightwave Technol. 28 (4) (2010) 651–661.

[11] ITU-T G.987.3, 10-Gigabit-capable passive optical networks (XG-PON): transmission convergence (TC) specifications, ⟨http://www.itu.int/rec/T-REC-G.987.3/en⟩.
[12] M. De Andrade, A. Buttaboni, M. Tornatore, P. Boffi, P. Martelli, A. Pattavina, G. Gavioli, Design of long-reach TDM/WDM passive optical access networks, in: Proceedings of Telecommunications Network Strategy and Planning Symposium (NETWORKS), 2012, pp. 1–6.
[13] A. Helmy, H. Fathallah, H. Mouftah, Interleaved polling versus multi-thread polling for bandwidth allocation in long-reach PONs, IEEE/OSA J. Opt. Commun. Netw. 4 (3) (2012) 210–218.
[14] B. Kantarci, H. Mouftah, Bandwidth distribution solutions for performance enhancement in long-reach passive optical networks, IEEE Commun. Surv. Tutor. 14 (3) (2012) 714–733.
[15] Y. Liu, L. Guo, C. Yu, Y. Yu, X. Wang, Planning of survivable long-reach passive optical network (LR-PON) against single shared-risk link group (SRLG) failure, Opt. Switch. Netw., 11 (1, Part B) (2014), 167–176.
[16] B. Schrenk, F. Bonada, J.A. Lazaro, J. Prat, Remotely pumped long-reach hybrid PON with wavelength reuse in RSOA-based ONUs, IEEE/OSA J. Lightwave Technol. 29 (5) (2011) 635–641.
[17] L. Shi, A. Nag, D. Datta, B. Mukherjee, New concept in long-reach PON planning: BER-aware wavelength allocation, Opt. Switch. Netw. 10 (4) (2013) 475–480.
[18] A. Sivakumar, G.C. Sankaran, K.M. Sivalingam, A comparative study of dynamic bandwidth allocation algorithms for long reach passive optical networks, IETE Techn. Rev. 29 (5) (2012) 405–413.
[19] H. Song, B.W. Kim, B. Mukherjee, Long-reach optical access networks: a survey of research challenges, demonstrations, and bandwidth assignment mechanisms, IEEE Commun. Surv. Tutor. 12 (1) (2010) 112–123.
[20] D. Hood, E. Trojer, Gigabit-capable PONs, Hoboken, NJ: Wiley, 2012.
[21] A. Gumaste, K. Pulverer, A. Teixeira, J.S. Wey, A. Nouroozifar, C. Badstieber, H. Schink, Medium access control for the next-generation passive optical networks: the OLIMAC approach, IEEE Netw. 26 (2) (2012) 49–56.
[22] S. De, V. Singh, H. Gupta, N. Saxena, A. Roy, A new predictive dynamic priority scheduling in ethernet passive optical networks (EPONs), Opt. Switch. Netw. 7 (4) (2010) 215–223.
[23] G. Kramer, B. Mukherjee, G. Pesavento, IPACT: a dynamic protocol for an ethernet PON (EPON), IEEE Commun. Mag. 40 (2) (2002) 74–80.
[24] X. Liu, G. Rouskas, MPCP-ℓ: look-ahead enhanced MPCP for EPON, in: Proceedings of IEEE ICC, 2013.
[25] Y. Luo, N. Ansari, Bandwidth allocation for multiservice access on EPONs, IEEE Commun. Mag. 43 (2) (2005) S16–S21.
[26] A. Razmkhah, A.G. Rahbar, OSLG: a new granting scheme in WDM ethernet passive optical networks, Opt. Fiber Technol. 17 (6) (2011) 586–593.
[27] T. Jimenez, N. Merayo, P. Fernandez, R. Duran, I. de Miguel, R. Lorenzo, E. Abril, Implementation of a PID controller for the bandwidth assignment in long-reach PONs, IEEE/OSA J. Opt. Commun. Netw. 4 (5) (2012) 392–401.
[28] K. Kanonakis, I. Tomkos, Offset-based scheduling with flexible intervals for evolving GPON networks, IEEE/OSA J. Lightwave Technol. 27 (15) (2009) 3259–3268.
[29] Y. Qin, D. Xue, L. Zhao, C.K. Siew, H. He, A novel approach for supporting deterministic quality-of-service in WDM EPON networks, Opt. Switch. Netw. 10 (4) (2013) 378–392.
[30] A. Buttaboni, M. De Andrade, M. Tornatore, A multi-threaded dynamic bandwidth and wavelength allocation scheme with void filling for long reach WDM/TDM PONs, IEEE/OSA J. Lightwave Technol. 31 (8) (2013) 1149–1157.
[31] B. Kantarci, H. Mouftah, Periodic GATE optimization (PGO): a new service scheme for long-reach passive optical networks, IEEE Syst. J. 4 (4) (2010) 440–448.
[32] A. Mercian, M. McGarry, M. Reisslein, Offline and online multi-thread polling in long-reach PONs: a critical evaluation, IEEE/OSA J. Lightwave Technol. 31 (12) (2013) 2018–2228.
[33] H. Song, B.-W. Kim, B. Mukherjee, Multi-thread polling: a dynamic bandwidth distribution scheme in long-reach PON, IEEE J. Sel. Areas Commun. 27 (2) (2009) 134–142.
[34] J. Zheng, H. Mouftah, A survey of dynamic bandwidth allocation algorithms for ethernet passive optical networks, Opt. Switch. Netw. 6 (3) (2009) 151–162.
[35] F. Aurzada, M. Scheutzow, M. Herzog, M. Maier, M. Reisslein, Delay analysis of ethernet passive optical networks with gated service, OSA J. Opt. Netw. 7 (1) (2008) 25–41.
[36] A. Dixit, G. Das, B. Lannoo, D. Colle, M. Pickavet, P. Demeester, Adaptive multi-gate polling with void filling for long-reach passive optical networks, in: Proceedings of ICTON, 2011, pp. 1–4.
[37] M. McGarry, M. Reisslein, Investigation of the DBA algorithm design space for EPONs, IEEE/OSA J. Lightwave Technol. 30 (14) (2012) 2271–2280.
[38] G. Kramer, Ethernet Passive Optical Networks, New York: McGraw-Hill, 2005.
[39] ITU-T G.984.3, Gigabit-capable passive optical networks (G-PON): transmission convergence layer specification, ⟨http://www.itu.int/rec/T-REC-G.984.3/en⟩.
[40] R. Roy, G. Kramer, M. Hajduczenia, H. Silva, Performance of 10G-EPON, IEEE Commun. Mag. 49 (11) (2011) 78–85.
[41] P. Begovic, N. Hadziahmetovic, D. Raca, 10G EPON vs. XG-PON1 efficiency, in: Proceedings of ICUMT, 2011, pp. 1–9.
[42] M. Hajduczenia, H.J.A. da Silva, P. Monteiro, 10G EPON development process, in: Proceedings of ICTON, vol. 1, 2007, pp. 276–282.
[43] G. Talli, C.W. Chow, E.K. MacHale, P.D. Townsend, High split ratio 116 km reach hybrid DWDM-TDM 10 Gb/s PON employing R-ONUs, in: Proceedings of ECOC, 2006, pp. 1–3.
[44] B. Skubic, J. Chen, J. Ahmed, L. Wosinska, B. Mukherjee, A comparison of dynamic bandwidth allocation for EPON, GPON, and next-generation TDM PON, IEEE Commun. Mag. 47 (3) (2009) S40–S48.
[45] H. Ikeda, K. Kitayama, Dynamic bandwidth allocation with adaptive polling cycle for maximized TCP throughput in 10G-EPON, IEEE/OSA J. Lightwave Technol. 27 (23) (2009) 5508–5516.
[46] L. Gutierrez, P. Garfias, S. Sallent, Flexible joint scheduling DBA to promote the fair coexistence in 1G and 10G EPONs, in: Proceedings of ONDM, 2011, pp. 1–6.
[47] M.-S. Han, Simple and feasible dynamic bandwidth allocation for XGPON, in: Proceedings of ICACT, 2013, pp. 341–344.
[48] M. Han, H. Yoo, D. Kee, Development of efficient dynamic bandwidth allocation algorithm for XGPON, ETRI J. 35 (1) (2013) 18–26.
[49] G. Kramer, B. Mukherjee, G. Pesavento, Interleaved polling with adaptive cycle time (IPACT): a dynamic bandwidth distribution scheme in an optical access network, Photon. Netw. Commun. 4 (1) (2002) 89–107.
[50] C. Assi, Y. Ye, S. Dixit, M. Ali, Dynamic bandwidth allocation for quality-of-service over ethernet PONs, IEEE J. Sel. Areas Commun. 21 (9) (2003) 1467–1477.
[51] X. Bai, C. Assi, A. Shami, On the fairness of dynamic bandwidth allocation schemes in ethernet passive optical networks, Comput. Commun. 29 (11) (2006) 2123–2135.
[52] S. Choi, S. Lee, T.-J. Lee, M. Chung, H. Choo, Double-phase polling algorithm based on partitioned ONU subgroups for high utilization in EPONs, IEEE/OSA J. Opt. Commun. Netw. 1 (5) (2009) 484–497.
[53] Y. Luo, N. Ansari, Limited sharing with traffic prediction for dynamic bandwidth allocation and QoS provisioning over EPONs, OSA J. Opt. Netw. 4 (9) (2005) 561–572.
[54] Y. Zhu, M. Ma, IPACT with grant estimation (IPACT-GE) scheme for ethernet passive optical networks, IEEE/OSA J. Lightwave Technol. 26 (14) (2008) 2055–2063.
[55] M. McGarry, M. Reisslein, F. Aurzada, M. Scheutzow, Shortest propagation delay (SPD) first scheduling for EPONs with heterogeneous propagation delays, IEEE J. Sel. Areas Commun. 28 (6) (2010) 849–862.
[56] Y. Luo, Activities, drivers, and benefits of extending PON over other media, in: Proceedings of National Fiber Optic Engineers Conference, 2013.
[57] B. Kantarci, H. Mouftah, Energy efficiency in the extended-reach fiber-wireless access networks, IEEE Netw. 26 (2) (2012) 28–35.
[58] J. Coimbra, G. Schutz, N. Correia, A game-based algorithm for fair bandwidth allocation in fibre-wireless access networks, Opt. Switch. Netw. 10 (2) (2013) 149–162.
[59] K. Yang, S. Ou, K. Guild, H.-H. Chen, Convergence of ethernet PON and IEEE 802.16 broadband access networks and its QoS-aware dynamic bandwidth allocation scheme, IEEE J. Sel. Areas Commun. 27 (2) (2009) 101–116.
[60] M. Milosavljevic, M. Thakur, P. Kourtessis, J. Mitchell, J. Senior, Demonstration of wireless backhauling over long-reach PONs, IEEE/OSA J. Lightwave Technol. 30 (5) (2012) 811–817.
[61] C. Ranaweera, E. Wong, C. Lim, A. Nirmalathas, Next generation optical-wireless converged network architectures, IEEE Netw. 26 (2) (2012) 22–27.
[62] H.T. Win, A.-S.K. Pathan, On the issues and challenges of fiber-wireless (Fi-Wi) networks, J. Eng., http://dx.doi.org/10.1155/2013/645745, in press.