Intra-ONU Bandwidth Scheduling in Ethernet Passive Optical Networks

N. Ghani, *Senior Member, IEEE*, A. Shami, *Member, IEEE*, C. Assi, *Member, IEEE*, and
M. Y. A. Raja, *Senior Member, IEEE*

***Abstract*—Quality-of-service (QoS) support in Ethernet passive optical networks (EPON) is a crucial concern. However, most studies have only focused on optical line terminal (OLT) capacity allocation amongst multiple optical network units (ONU), and the further issue of intra-ONU allocation remains open. In this work a novel decentralized intra-ONU solution is presented using virtual-time schedulers. Results confirm good performance for a wide range of input traffic classes and loads.**

***Index Terms*—Ethernet PON (EPON), optical access, quality of service (QoS), scheduling.**

I. Introduction

Ethernet Passive Optical Network (EPON) [1] is a very promising fiber-based ultra-broadband access technology. The architecture comprises of a centralized *optical line terminal* (OLT) connecting dispersed *optical network units* (ONU) over point-to-multipoint topologies, e.g., tree, ring, and bus. Here, downstream transmission is done in a broadcast manner, whereas upstream transmission is arbitrated by the OLT via *time-division multiple access* (TDMA) [1]. In particular, a new EPON *multi-point control protocol* (MPCP) is being defined in the IEEE 802.3ah working group.[1]

With increased application stringencies, EPON *quality of service* (QoS) has become a major focus, and many *dynamic bandwidth allocation* (DBA) schemes have been proposed [1]–[3]. Nevertheless, in practice a single ONU will host many end-users and this mandates added *intra-ONU* considerations [5] to ensure *end-to-end* service guarantees. This letter addresses these crucial concerns and is organized as follows. Section II presents a brief overview of DBA schemes and subsequently Section III details a novel intra-ONU virtual-time scheduler. Detailed simulation studies are then presented in Section IV along with final conclusions in Section V.

II. EPON Bandwidth Allocation Overview

Early OLT schemes used fixed *time division multiplexing* (TDM) to assign static capacity increments to ONU nodes [1]. However, these setups precluded idle capacity reuse and yielded low utilizations for bursty traffic. Hence, designers graduated to more advanced setups using OLT-ONU signaling to coordinate ONU allocations, i.e., MPCP *REQUEST, GRANT* messages [1]. A key proposal here was the frame-based *interleaved polling and adaptive cycle time* (IPACT) [1] scheme, which used time-overlapping of ONU transmission windows to improve OLT utilization. Nevertheless, since IPACT does not specify explicit multi-service QoS provisions, more advanced DBA schemes were evolved to tailor *aggregate* GRANT windows to ONU request/usage levels, e.g., *limited, gated, linear credit*, and *elastic* allocation [1].

Most OLT-ONU DBA schemes can essentially be classified as dynamic, distributed renditions of *weighted round-robin* (WRR) schedulers [3], as discussed in [6], (Fig. 1). Namely, allocations are made from a given frame size (fixed or variable) as per time-varying ONU usage reports. However, since the focus of the work here is strictly on *intra-ONU* allocation, a generic *weighted* inter-ONU DBA scheme from [2] is adopted in which ONU nodes are partitioned into two groups, *underloaded* and *overloaded*. Underloaded ONU nodes are those requesting below their minimum guaranteed bandwidth, $B_i^{\min}$, and hence their unused capacity is shared in a weighted manner amongst *overloaded* ONU nodes. Specifically, consider an OLT link of speed C bits/s serving N ONU nodes. The ith overloaded ONU allocation is

$$B_i = \frac{\left(T - \sum_{j\in\text{underload}} R_j\right)\omega_i}{\sum_{k\in\text{overload}}\omega_k} \tag{1}$$

where T is the frame size (minus overheads), R_j is the request size of the jth ONU (underloaded), and ω_i is the weight associated with the i-th ONU (overloaded). For example, results with $\omega_i = R_i$ (where R_i is received request size) show improved performance efficiency over the limited scheme [2]. Most EPON DBA schemes require all *REPORT* messages to be received in frame period, i.e., frame size greater than largest round-trip delay. Note that end-of-frame compute times can increase inter-frame idling and lower efficiency. Hence [2] sends underloaded ONU *GRANT* messages in the same cycle in which they were received, i.e., prior to (1) computation.

III. Intra-ONU Bandwidth Allocation

Given the various EPON deployment scenarios, a single ONU will likely host many endusers with differing *service*

Manuscript received March 22, 2004. The associate editor coordinating the review of this letter and approving it for publication was Dr. M. Devetsikiotis. This work was supported in part by the Center for Manufacturing Research (CMR) at Tennessee Tech University, Cookeville, TN, USA.

N. Ghani is with the Department of Electrical and Computer Engineering, Tennessee Tech University, Cookeville, TN 38505 USA (e-mail: nghani@tntech.edu; nghani@ieee.org).

A. Shami is with the Department of Electrical and Computer Engineering University of Western Ontario, London, ON N6A 5C2, Canada.

C. Assi is with the Concordia Institute for Information Systems Engineering, Concordia University, W. Montreal, QC H3G 1M8, Canada

M. Y. A. Raja is with the Department of Physics and Optical Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223-0001 USA.

Digital Object Identifier 10.1109/LCOMM.2004.837664

[1] IEEE 802.3ah task force home page. [Online.] Available: http://www.ieee802.org/3/efm

1089-7798/04$20.00 © 2004 IEEE

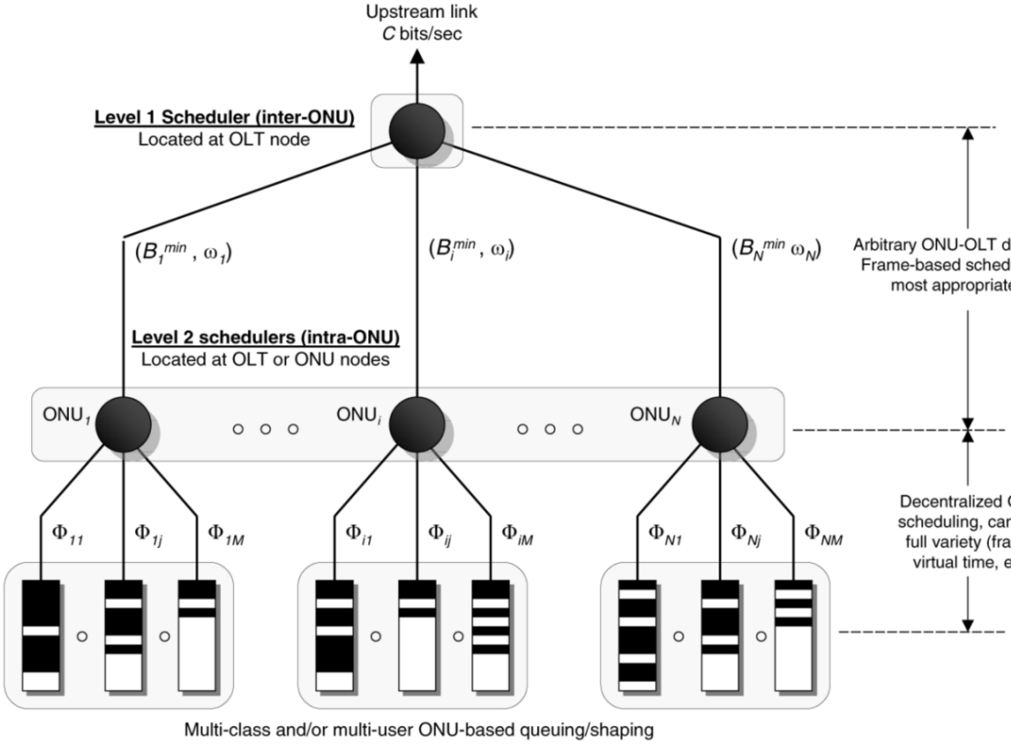
Upstream link
C bits/sec
Level 1 Scheduler (inter-ONU)
Located at OLT node
(B_1^{min}, ω_1)
(B_i^{min}, ω_i)
$(B_N^{min}\ \omega_N)$
Arbitrary ONU-OLT delays, Frame-based schedulers most appropriate
Level 2 schedulers (intra-ONU)
Located at OLT or ONU nodes
ONU_1
ONU_i
ONU_N
Φ_{11}
Φ_{1j}
Φ_{1M}
Φ_{i1}
Φ_{ij}
Φ_{iM}
Φ_{N1}
Φ_{Nj}
Φ_{NM}
Decentralized ONU scheduling, can use full variety (frame, virtual time, etc)
Multi-class and/or multi-user ONU-based queuing/shaping

Fig. 1. Overview of OLT-ONU scheduling framework.

```
ONU_i packet receive at queue j (any time)
if (buffer j is empty)
        $S_{ij} = \max\{F_{ij}, v_i\}$
        $F_{ij} = S_{ij} + (L_{ij} / \Phi_{ij})$

ONU_i packet transmit (in GRANT window)
if (all buffers not empty)
        Find buffer j s.t. $S_{ij} < S_{ik}$ for all k
        if ( ($t+L_{ij}/C$) < grant window)
        Dequeue and send buffer j HOL packet
        $v_i = S_{ij}$
if (buffer j non-empty)
        $S_{ij} = \max\{F_{ij}, v_i\}$
        $F_{ij} = S_{ij} + (L_{ij} / \Phi_{ij})$
else
        $S_{ij} = \infty$
```

Fig. 2. Intra-ONU M-SFQ algorithm at ith ONU, $O\log(M)$.

level agreement (SLA) requirements, e.g., throughput, delay, delay variation. Although an initial solution in [2] incorporates per-queue statistics for intra-ONU allocation at the OLT, such *centralized* schemes pose scalability concerns and rely upon dated ONU reports. Hence, a more scalable, *decentralized* intra-ONU solution is developed here using robust packet scheduling for M input queues [3]. In particular, a *virtual time* approximation of the ideal *generalized processor sharing* (GPS) discipline [3], [4] is considered, as such schemes can achieve very fine bandwidth control. This solution fully decouples inter/intra-ONU bandwidth allocation and can interoperate with all OLT-ONU DBA algorithms.

Virtual time schedulers "time-stamp" incoming packets and maintain *virtual times* to track service levels. Hence, transmit orderings are simply determined by sorting the respective time-stamp values. Although a wide range of related schemes have been studied [3]—*weighted fair queuing* $(\mathrm{WFQ}/\mathrm{WFQ}^2)$, *self-clocked fair queuing* (SCFQ), *start-time fair queuing* (SFQ)—most have been applied for *intra-systems* roles in ATM or IP platforms. Clearly, the further application of such schedulers in intermittent EPON settings needs more consideration. In particular, since hardware cost/complexity is a huge concern for cost-sensitive ONU settings, a modified version of the *start-time fair queuing* (SFQ) [4], [5] algorithm is developed [6]. Unlike other virtual time schemes which time-stamp *all* packets, this scheme only maintains time-stamps for *head-of-line* (HOL) queue packets [6], yielding much lower complexity.

The *modified SFQ* (M-SFQ) algorithm for the i-th ONU is shown in Fig. 2. The scheme assigns a weighting to each queue, ϕ_{ij} $(\sum_j \phi_{ij} \leq 1)$ and tracks *aggregate* ONU service via a global virtual time, v_i. Variables are also maintained to track local per-queue (i.e., HOL) start and finish times, namely S_{ij} and F_{ij}, respectively. Upon arrival at an empty queue j, the HOL start and finish times are updated as:

$$S_{ij} = \max(F_{ij}, v_i) \tag{2}$$

$$F_{ij} = S_{ij} + \frac{L_{ij}}{\phi_{ij}} \tag{3}$$

respectively, where L_{ij} is the packet length. Meanwhile, packets arriving at nonempty buffers are simply queued. Conversely for transmission, the queue with the minimum HOL start time is selected. If this HOL packet can be transmitted in the transmit window (no fragmentation), it is de-queued and the global virtual time updated to the *start* time [5], i.e., S_{ij}. After transmission, the related local start and finish times are also updated. Now if the queue is nonempty, these updates are identical to the arrival case, otherwise, the start time is set to large value to remove it from contention, i.e., $S_{ij} = \infty$. Overall, M-SFQ is of order $\mathrm{O}(\log M)$, as it sorts up to M values per transmission [5]. Clearly, this is acceptable for Ethernet QoS frameworks which

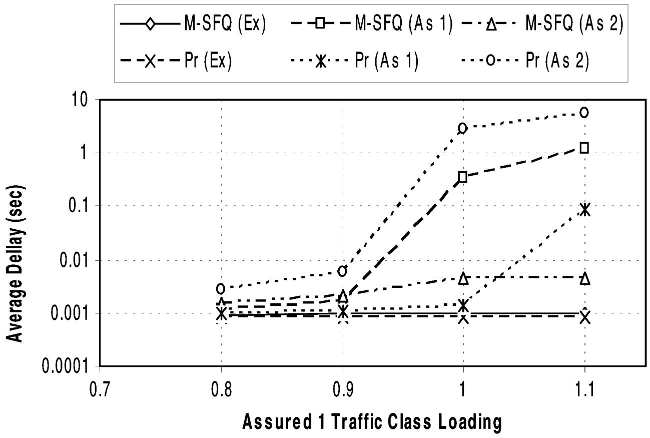

Fig. 3. Average ONU-to-OLT queueing and scheduling delay.

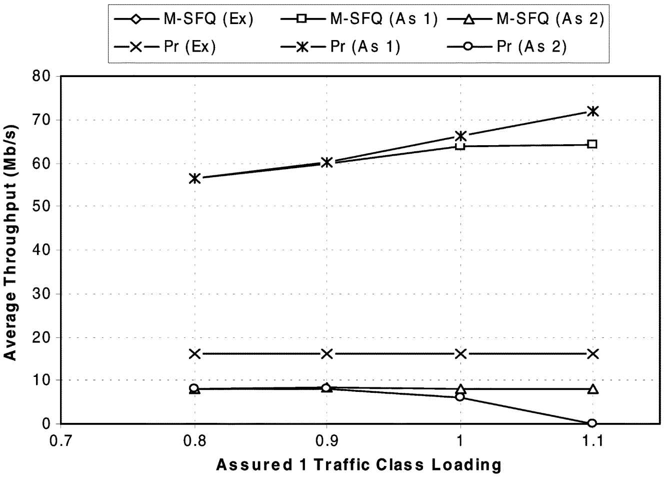

Fig. 4. Average throughputs (80% Expedited, 80% Assured 2).

usually comprise up to eight classes. Note that this scheme is not strictly work-conserving due to intermittent *GRANT* arrivals and ONU fragmentation effects. Nevertheless, global virtual time (v_i) and all finishing times (F_{ij}) still satisfy the required monotonic-increase property [3].

IV. Simulation Results

The M-SFQ scheme is studied using the *OPNET Modeler 10.0* simulation tool. In particular, several *real-time* classes are defined based upon the IETF *Differentiated Services* (DiffServ) model, including an Expedited class (delay-sensitive voice/leased-line), a high-grade Assured 1 class (high-speed video), and a lower-grade Assured 2 class (compressed slow-speed play-back voice/video). Delay-insensitive best-effort traffic is not modeled as it can easily be buffered and serviced via idle ONU capacity. Performance is compared against an intra-ONU priority scheduler, which transmits packets using a simple, strict ordering, e.g., Expedited, Assured 1, Assured 2. The only exception is during ONU fragmentation, in which case lower priority classes are considered to boost utilization, i.e., $\mathrm{O}(M)$ [6].

The EPON comprises ten ONU nodes at 0.125-ms delays, chosen to stress OLT-ONU bandwidth-delay effects. Link speeds are 1.0 Gbps and the OLT frame size is set to 2 ms. Here, the individual traffic categories are assigned *fixed* proportions of the aggregate load, e.g., 20% Expedited, 70% Assured 1, 10% Assured 2. Expedited traffic emulates voice services and is generated using fixed 70-byte packets with exponential interarrival times. Meanwhile, both assured traffic categories emulate packet video and are generated using uniform packet sizes between 64–1518 bytes with exponential inter-arrivals. Note that interarrival times will depend upon the desired loading. Furthermore, intra-ONU weights (M-SFQ) are assigned per the proportional loadings, i.e., $\phi_{i1} = 0.2$ (Expedited), $\phi_{i2} = 0.7$ (Assured 1), $\phi_{i3} = 0.1$ (Assured 2). Since the main focus is on intra-ONU performance, OLT DBA essentially allocates equal capacity to each ONU, yielding approximately 100 Mbps/ONU at full loading (i.e., $\omega_\mathrm{i} = 0.10$, Section II).

The M-SFQ scheme is tested against the priority scheduler for aggressive traffic scenarios. Namely, the average Expedited and Assured 2 traffic class loads are fixed at 80% of their maximum values (16 and 8 Mbps, respectively), whereas the heavier Assured 1 traffic class loads are varied from 80%–110% of their maximum values (56–77 Mbps). The total delay is computed as the end-to-end delay minus fixed propagation and transmission times, and infinite ONU buffering is assumed to ascertain maximum latencies. The related delay and throughput performances are shown in Figs. 3 and 4, and confirm the improved performance (inter-class isolation) of the M-SFQ scheme. Here, Expedited class delays are very stable and lower-volume Assured 2 traffic throughput is also maintained, albeit with a slight increase in average delay (2.5 ms, 110% Assured 1 load). Conversely near link saturation, the priority scheduler yields unacceptably high delay and throughput degradation for Assured 2 traffic, e.g., almost zero throughput at 110% load. Clearly, the M-SFQ scheme achieves a very fine degree of bandwidth resolution (over 90% of fair share after signaling overheads).

V. Conclusion

Intra-ONU bandwidth allocation is a key issue in EPON. In this letter, a novel M-SFQ virtual time scheduler is presented for decentralized ONU bandwidth operation. The scheme features low implementation complexity and can interoperate with any OLT-ONU (inter-ONU) DBA scheme. Simulation results confirm that the M-SFQ scheme achieves a very fine degree of bandwidth allocation and good delay performance.

References

[1] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT: A dynamic protocol for an ethernet PON (EPON)," *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 74–80, Feb. 2002.

[2] C. Assi *et al.*, "Dynamic bandwidth allocation for quality of service over ethernet PON," *IEEE J. Select. Areas Commun.*, Nov. 2003.

[3] H. Zhang, "Service disciplines for guaranteed packet performance service in packet-switching networks," *Proc. IEEE*, Oct. 1995.

[4] P. Goyal *et al.*, "Start-time fair queuing: A scheduling algorithm for integrated services packet switching networks," in *ACM Sigcomm 1996*, Stanford, CA, Aug. 1996.

[5] N. Ghani and J. W. Mark, "Hierarchical scheduling for integrated ABR/VBR services in ATM networks," in *Proc. IEEE GLOBECOM 1997*, Phoenix, AZ, Oct. 1997.

[6] N. Ghani *et al.*, "Quality of service in Ethernet passive optical networks," in *IEEE Sarnoff Symp.*, Princeton, NJ, Apr. 2004.