Adaptive Fairness through intra-ONU Scheduling for Ethernet Passive Optical Networks

Ahmad R. Dhaini*, Chadi M. Assi*, Abdallah Shami†, and Nasir Ghani‡

*Concordia Institute for Information Systems Engineering
Concordia University, Montreal, Quebec, Canada
Email:{a_dhaini, assi}@ciise.concordia.ca
†Electrical and Computer Engineering Department
University of Western Ontario, Ontario, Canada
Email: ashami@eng.uwo.ca
‡Electrical and Computer Engineering Department
Tennessee Tech University, TN, USA
Email: nghani@tntech.edu

Abstract—Ethernet passive optical networks (EPONs) are being designed to deliver multiple services and applications, such as voice communications (VoIP), standard and high-definition video (STV and HDTV), video conferencing (interactive video) and data traffic access network. However, most of the current work focuses on inter-ONU dynamic bandwidth allocation (DBA) algorithms. In this paper, we concentrate on the intra-ONU bandwidth allocation for different classes of services. We present a new intra-ONU scheduling scheme based on the Deficient Weighted Round Robin (DWRR) scheduling to achieve adaptive fairness among different classes of services. We validate our reasoning by measuring both the end-to-end delay of different traffic along with the jitter performance of high priority traffic using extensive simulation experiments.

I. INTRODUCTION

Ethernet Passive Optical Network (EPON) [7] represents the convergence of inexpensive and ubiquitous Ethernet equipment with low-cost fiber infrastructure. It is viewed by many as an attractive solution for the broadband access network bottleneck and has been under intense research activities recently. EPON is a point-to-multipoint (P2MP) access network with no active elements in the signal’s path from source to destination. It has been standardized by the IEEE 802.3ah working group [6] and it comprises one Optical Line Terminal (OLT) and a set of associated Optical Network Units (ONUs). The OLT resides at the provider central office (CO) and connects the optical access network to the metropolitan area network (MAN) or wide area network (WAN). On the other hand, the ONU is usually located at either the curb (i.e., fiber-to-the-curb (FTTC) solution) or the end-user location (i.e., fiber-to-the-building (FTTB) and fiber-to-the-home (FTTH)), and provides broadband video, data, and voice services. As shown in Fig.1(a), a single fiber extends from an OLT to a 1 : N passive optical splitter. The splitter fans out to multiple single fiber drops, which are connected to different ONUs. EPON systems deploy only one channel for downstream traffic and another channel for upstream traffic. In the downstream, Ethernet frames are broadcast by the OLT and are selectively received by each ONU. Alternatively, in the upstream, multiple ONUs share the same transmission channel to transmit data and control packets to the OLT. Since ONUs are unable to detect collision and due to the difficulty to implement a carrier sense multiple access with collision detection (CSMA/CD), it is necessary to design a mechanism that arbitrates the access of ONUs to the shared medium. This is achieved by designing a Medium Access Control (MAC) protocol to prevent collision between packets of different ONUs transmitting simultaneously. Current MAC supports Time Division Multiplexing (TDM) [7], where each ONU is allocated a fixed or dynamic time slot to transmit data to the OLT and each ONU buffers packets received from different subscribers until they are transmitted in the assigned window. One distinguishing feature that broadband EPON is expected to support is the ability to deliver services to emerging IP-based multimedia traffic with diverse Quality-of-Service (QoS) requirements [2]. However, ensuring proper bandwidth allocation for every class of service (CoS) remains a challenging issue; a problem that is better known as fairness in EPON. Lately, Kramer et. al [7] addressed this problem, and proposed a new hierarchical scheduler that fairly divides the excess bandwidth (defined in a later section) among priority queues (PQs) from different ONUs in EPON.

In this paper we present a new intra-ONU bandwidth scheduling algorithm based on the Deficient Weighted Round Robin (DWRR) scheduling technique applied in ATM networks [9]. This scheme ensures that every class of traffic gets a fair share of the assigned bandwidth at the ONU by forcing the scheduler to visit every PQ for a specific period of time that is determined by the weight allocated to the corresponding PQ. The rest of the paper is organized as follows. Section II reviews related work in the field. In section III, we introduce DWRR and we present our Intra-ONU scheduling scheme. Section IV presents a simulation based comparative study of the proposed algorithms and finally, section V concludes our work.

II. BACKGROUND & RELATED WORK

Current EPONs support diverse applications, various traffic sessions are aggregated into a limited number of classes to
be serviced with differentiated services. These services are classified as follows: Best Effort (BE) "data” traffic, Assured Forwarding (AF) traffic such as variable-bit-rate (VBR) video stream and Expedited Forwarding (EF) traffic used to emulate point-to-point (P2P) connections or real time services, such as Voice over IP (VoIP). The high-priority class is EF, which is delay-sensitive and requires bandwidth guarantees. The medium-priority class is AF, which is not delay-sensitive but requires bandwidth guarantees. The low priority class is BE, which is neither delay-sensitive nor bandwidth guaranteed.

As shown in Fig.1(b), upon receiving traffic "flows” from the registered subscribers, the ONU performs three main operations before transmission in the upstream channel. First, it classifies every newly arriving packet using a "packet-based" classifier. Next, and before placing packets in the corresponding priority queues, the ONU decides whether a packet should be admitted depending on the adopted traffic policing (admission control) mechanism (e.g., Leaky Bucket). Finally, it selects packets from its queues, depending on the intra-/inter-ONU scheduling algorithm [11], and sends them to the OLT as "flows” in the assigned transmission window (TW). Moreover, there are two types of intra-ONU scheduling: strict and non-strict priority scheduling algorithms. In strict priority scheduling, a lower-priority queue is scheduled only if all queues with higher priority are empty. However, this may result in a starvation for low-priority traffic or as dubbed in [11], "light-load penalty".

Non-strict priority scheduling addresses this problem by allowing reported packets (regardless of their priority) to be transmitted first as long as they are transmitted in the allocated TW. In other words, here, the transmission order of different priority queues is based on their priorities. As a result, all traffic classes have access to the upstream channel while maintaining their priorities; which enables fairness in scheduling. Note that inter-ONU control messages for allocating bandwidth to different ONUs are transmitted via the MPCP (multi-point control protocol) access protocol. MPCP is a signaling access protocol which is being developed and standardized by the IEEE 802.3ah Task Force [6]. The OLT gathers information from different ONUs and dynamically allocates bandwidth to ONUs through the use of REPORT and GATE messages of MPCP. Within each cycle, ONUs use REPORT messages to report its bandwidth requirements (e.g. buffer occupancy) to the OLT. Upon receiving REPORT messages from the ONUs, the OLT performs the appropriate bandwidth allocation computation and broadcasts a GATE message to each ONU, containing the appropriate transmission grants (transmission start $T_{start}$ and $T_{end}$). Note that MPCP does not specify any particular bandwidth allocation algorithm. Instead, it is designed to facilitate the implementation of dynamic bandwidth allocation algorithm (DBA).

DBA uses the services offered by the MPCP protocol to communicate assigned transmission windows to their appropriate ONUs. In this section we will focus on related intra-ONU schedulers proposed to date. To the best of our knowledge, we are aware of only the work presented next on QoS intra-ONU scheduling. For details about inter-ONU scheduling we refer the reader to [10].

To cope with the light-load penalty caused by applying strict priority scheduling technique, the authors of [11] proposed two methods. The first method involves a two-stage queueing process. Here, the incoming packets after sending the REPORT message are placed in the second-stage queue. Consequently, when a new GATE is received, the second-stage queue is emptied first. This however results in an increased average delay for all types of traffic. In the second method, the "after-report” incoming traffic is estimated, and thus the grant window will be large enough to accommodate the newly arriving high priority packets.

Alternatively, in [2], the intra-ONU scheduler employs priority scheduling only on the packets that arrive before sending the REPORT message. This scheme eliminates the "light-load penalty” and allows all services to access the shared medium. On the other hand, the authors of [1] proposed a new intra-ONU scheduling scheme named "Modified Start-Time fair queueing” (M-SFQ) that muses the performance of VBR traffic. Here, the scheduler selects for transmission the queue with the minimal start time, derived from the head-of-line (HOL) packet in each queue, and synchronized with a Global Virtual Time.
III. INTRA-OUN SCHEDULING

To date, a wide range of scheduling schemes have been studied [1], (e.g., weighted fair queuing ($WFQ/WFQ^2$), self-clocked fair queuing (SCFQ), start-time fair queuing (SFQ), Weighted Round Robin (WRR) and Stratified Round Robin (SRR)). Most of these methods have been applied for intrasystems roles in ATM or IP platforms. One distinguished scheme for achieving fairness with low complexity is the Deficient Weighted Round Robin (DWRR) [9]. DWRR is well suited for EPON settings. In this paper we propose a modified algorithm (M-DWRR) to enforce fairness among the various classes of service.

A. DWRR scheduling discipline

DWRR as proposed, defines the following three main parameters for each CoS queue $i$:

1) A "weight" $\alpha_i$ that defines the percentage of the output port bandwidth allocated to the queue.

2) A "Deficit Counter" $DC(i)$ that specifies the total number of bytes that the queue is permitted to transmit in each scheduler’s visit. The DC saves "credits" remaining from previous scheduling visit and adds them to the DC of the next visit until the queue is empty and hence $DC(i) = 0$.

3) A "quantum" $Q(i)$ that is proportional to $\alpha_i$ and is expressed in bytes.

First, a Round Robin (RR) scheduler initializes the deficient counters, $DC(i) = 0, i = 0, ... , x$, then visits each non-empty queue and determines the size (in bytes) of the Head Of Line (HOL) packet. $Q(i)$ is computed from the available port bandwidth as follows:

$$Q(i) = [\alpha_i \times B_{port}] \tag{1}$$

Where $B_{port}$ is the bandwidth available on the transmission port (in bytes). Next, the scheduler computes:

$$DC(i) = DC(i) + Q(i) \tag{2}$$

At this time, it checks if the HOL packet is greater than $DC(i)$; if yes, it moves to the next queue and "saves" the remaining credits in $DC(i)$, otherwise will select the packet for transmission and updates its deficient counter:

$$DC(i) = DC(i) - S^{HOL}_{i} \tag{3}$$

Where $S^{HOL}_{i}$ is the size of the HOL packet in queue $i$. When queue $i$ is empty, $DC(i)$ is reset to 0, and the pointer of the RR scheduler moves to the lower priority queue.

Fig. 2 shows a DWRR numerical example with three queues [9]. Here, the first queue is considered the high priority queue with $\alpha_1 = 50\%$, the other two, medium and low priority, queues are allocated an equal weight rate $\alpha_2 = \alpha_3 = 25\%$; $B_{port} = 2000$ Bytes.

Consequently, $Q(1) = 1000$ bytes and $Q(2) = Q(3) = 500$ bytes. As Figure 2 shows, when the scheduler starts, it looks at the HOL packet in the HP queue and sets $DC(1) = Q(1) = 1000$ bytes. Here, $S^{HOL}_{1} < DC(1)$ and thus will be selected for transmission. Hence, DC(1) is updated and is now equal to 400 bytes. The HOL packet is now of size 300 and is going to be selected for transmission (following equation (2) in DWRR rules). DC(1) becomes equal to 100 bytes. Next, the HOL is of size 400 bytes which is greater than DC(1). For that reason 100 bytes are then saved as credits for the next scheduling round on queue 1, and the scheduler moves to next non-empty queue. In other words, when the scheduler re-visits queue 1, DC(1) will be equal 1100 bytes instead of 1000 bytes.

The advantages of DWRR over other schemes are listed below:

- DWRR accurately supports weighted fair bandwidth distribution for CoS queues of variable-length packets.
- DWRR combines both the class-based queuing approach along with the Weighted round robin scheduling scheme.
- DWRR has lower complexity than WFQ and can be implemented in hardware.

B. Integrating DWRR with QoS EPON

In EPON, every ONU maintains a number of priority queues where incoming packets are classified and queued based on their priorities. Unlike the system discussed in section III-A, in EPON, the ONU accesses the channel during the assigned TW that is specified by $T_{start}$ and $T_{end}$. Hence, ONU $j$ will compute the quantum for each queue $i$ based on the weight assigned to the queue and the transmission window allocated by the OLT.

Therefore, DWRR will have to set its three defined parameters (i.e., $\alpha_{i,j}$, $DC(i,j)$ and $Q(i,j)$) for each queue $i$. Suppose that the allocated TW is of size $S_j$ (bytes) and is computed as follows:

$$S_j = \min(B_{min} + B^j_{excess}, \sum_{i=1}^x R_{i,j}) \tag{4}$$

Where $R_{i,j}, i = 1...x$ is defined as the requested size of each queue $i$, $B^j_{excess}$ is the excess bandwidth allocated to ONU$_j$ and $B_{min}$ is the minimum bandwidth guaranteed (see next section). Then the quantum is computed:

$$Q(i,j) = [\alpha_{i,j} \times S_j] \tag{5}$$

The update of the deficient counter is computed as in (2). Note that $Q(i,j)$ can be set by the OLT and incorporated in the GRANT message.

C. Modified DWRR (M-DWRR)

As mentioned before, DWRR scheduling discipline visits each PQ in a round robin fashion. Moreover, after each visit
made by the scheduler to all PQs, the deficient counter is updated according to the rules explained in section A. On the other hand, in M-DWRR, once the scheduler has finished visiting all the queues, the remaining bandwidth from the assigned TW of the current cycle is distributed to all the PQs based on the corresponding weights:

\[ DC(i, j) = DC(i, j) + [\alpha_{i,j} \times B^j_{\text{remain}}] \]  

(6)

Where \( B^j_{\text{remain}} \) is the remaining bandwidth (in bytes) from the assigned TW of the same cycle. This remaining bandwidth is found from the unutilized bandwidth after the first scheduling visit to all PQs. In other words, since the TW is divided among priority queues depending on their weights (and not their needs), some queues might not utilize all their corresponding assigned bandwidth. Thus, in order to eliminate the waste of bandwidth, we re-allocate this portion to the PQs based on the same weight assignment.

Alternatively, the ONU might follow a different "update scheme" and hence re-validates the deficient counters based on a different weight assignment scheme, that might be derived/concluded from the different traffic requirements and queues occupancies rather than the original weight agreement. Furthermore, another "update discipline" might be implemented, where \( DC(i, j) \) is computed as in (6), but yet if the allocated bandwidth of higher-priority is not needed (i.e., queue is empty), it will be distributed to the lower priority queues. However, since high priority traffic (EF) are delay-sensitive and since incoming packets might arrive after the described distribution, the scheduler must permit transmission of these packets by setting a flag that triggers its pointer, upon the arrival of these packets, to the appropriate queue. In this way, high priority traffic delay is preserved and its jitter is protected.

On the other hand, the scheduler might allocate the remaining bandwidth in a traditional round robin fashion while assigning bandwidth for each non-empty queue such that this allocated bandwidth is "just" equal to the HOL packet of each queue. Algorithm 1 illustrates this scheme (where \( R_{i,j} \) is the variable remaining bandwidth after each allocation done by the scheduler for each PQ \( i \)).

The advantage of such a scheme and of DWRR in general, is that each ONU can adaptively set (depending on the traffic demand and the Service Level Agreement (SLA)) its own weights in both phases (i.e., initially and/or after computing \( B^j_{\text{remain}} \)). Another major advantage over the previously proposed intra-ONU schemes [1], [2], is that this scheme ensures guaranteed bandwidth for all types of traffic. Namely, the algorithm guarantees for every class of traffic the "agreed upon" bandwidth and quality of service by enforcing the weight policy. However, if the traffic of one priority queue is light, then the allocated resources can be utilized by other traffic classes.

**D. Dynamic Bandwidth Allocation (DBA) Scheme**

In our study, we focus on the performance of the intra-ONU scheduling algorithm. We choose the limited DBA scheme [11] with excess consideration [2] for our inter-ONU scheduling. Here, the OLT is unaware of the quality of service requirements of the ONU. Hence, the ONU reports its bandwidth requirement using a single four-byte field specified in the MPCP REPORT message. Meanwhile, the OLT waits until all REPORTs from all ONUs are received, and allocates the appropriate bandwidth based on the requested bandwidth of each ONU. In other words, if ONU \( j \) is requesting less than a minimum bandwidth guaranteed \( B_{MIN} \), a bandwidth equivalent to the requested one is allocated. In contrast, if the requested bandwidth is larger than \( B_{MIN} \), the OLT allocates \( B_{MIN} \) plus a share of excess bandwidth (if available). The minimum bandwidth guaranteed \( B_{MIN} \) is defined in [2] and is dependant on the weight assigned to each ONU based on the SLA between the service provider (SP) and users.

There are two ways to assign transmission windows via the excess bandwidth, namely Controlled Excess (CE) and Un-Controlled Excess (UE) allocation schemes [4]. In UE scheme, the OLT collects from the received REPORTs all the surplus bandwidth available for the next cycle and assigns this total excess uniformly to all highly loaded ONUs regardless of their requested bandwidth. The advantage of this uncontrolled scheme is that highly loaded ONUs are assigned enough bandwidth to satisfy their high demands (assuming the excess is enough). However, if some ONUs are only "slightly" highly loaded, they are being assigned an unfair share of the excess bandwidth that could ultimately be not utilized. Hence, the assignment of the excess bandwidth must be controlled (i.e., CE) by the OLT in order to guarantee a fair bandwidth allocation for all highly loaded ONUs. In [4], we showed that CE bandwidth allocation improves the overall performance and significantly minimizes the wasted bandwidth caused by the UE allocation. Hence, we also use the CE allocation scheme in our DBA scheme.

**IV. Simulation Results**

In this section, we compare the performance of the intra-ONU bandwidth scheduling schemes presented in the previous sections and their impact on the overall performance of the network. For this reason, an event-driven packet-based simulation model is developed using C++. The total number of ONUs \( N = 16 \), and the PON speed = 1Gbps. The guard time is equal to 1\( \mu \)s, the cycle time to 2ms and the ONU buffering queue size to 10Mbytes. For the traffic model considered

**Algorithm 1 M-DWRR Deficient Counter Update**

1: \( R_{i,j} = B^j_{\text{remain}} \)
2: while \( R_{i,j} > 0 \) do
3:   for all \( i \in ONUs \) do
4:     if queue \( i \) empty() \&& \( S^H_{i,j} \leq R_{i,j} \) then
5:       \( DC(i, j) \leftarrow S^H_{i,j} \)
6:       \( R_{i,j} \leftarrow R_{i,j} - S^H_{i,j} \)
7:     end if
8:   end for
9: end while
the network behavior, we consider two sets of PQ weights: 

- Set 1: $\alpha_1 = 20\%$, $\alpha_2 = 70\%$ and $\alpha_3 = 10\%$
- Set 2: $\alpha_1 = 50\%$, $\alpha_2 = 40\%$ and $\alpha_3 = 10\%$

The reason behind this choice is to show how the selection of the priority queue weights can affect the overall network performance. At the same time, this selection will show that our proposed scheme can adaptively ensure fairness among all traffic classes even if the weights selection was not fair to lower priority queues.

Our simulator takes into account the queuing delay, transmission delay and the packet processing delay. The metrics of comparison are: average packet delay, network throughput, packet drop rate and the jitter performance of EF traffic [5].

A. Average Packet Delay

Fig. 3 compares the average packet delay of EF, AF, and BE services for M-SFQ and both our proposed DWRR and M-DWRR scheduling schemes under both sets of weights. Note that the traffic load of a high loaded ONU is varied between 0.1 and 1 (i.e., 10Mbps and 100Mbps). As shown, all schemes behaved similarly under light and heavy loads for EF traffic. On the other hand, the impact of the weight selection appears on both AF and BE traffic behaviors. Here, DWRR and M-DWRR schemes exhibit better results under set 2 (best suiting the AF traffic) than M-SFQ, given that M-SFQ is designed to satisfy HP and MP traffic and especially VBR traffic; (e.g., $\approx 90$ ms difference at load 0.8).

Meanwhile, our schemes demonstrate better performance for BE traffic. For example in fig. 3(a), at load 0.8, BE traffic, in both DWRR and M-DWRR, shows better performance of almost 75% improvement (i.e., 1 s and 4 s) over M-SFQ where a dramatic delay degradation starts, because almost all BE packets are dropped. Moreover, as discussed, the weight selection does affect the delay performance for both DWRR and M-DWRR. Accordingly, with Set 1, the behavior of all traffic classes in all schemes (in terms of delay) is similar, while with Set 2, the performance varies. This shows the advantage of our schemes, that is the capability of assigning weights in a way to poise the performance of each traffic aside. At the same time, by enforcing the weight policy, our schedulers make sure to provide each PQ a share of the assigned bandwidth. In this way, fairness is ensured among different classes of traffic, and the traffic priority (delay sensitivity) is respected.

B. Network Throughput & Packet Loss Rate

Our proposed schemes goal is to ensure fairness among all classes of traffic, especially for BE traffic that, in most of the “to-date” advocated schemes, pays the performance penalty to gratify HP traffic. For more illustration, we measured the packet loss rate for a new set of weights where $\alpha_1 = \alpha_2 = 70\%$ and $\alpha_3 = 10\%$.
20%, α2 = 40% and α3 = 40%. Fig. 4 shows the packet loss rate. Here, regardless of the weight combination, DWRR and M-DWRR improve the buffer occupancy of all PQs, and thus significantly reduce the packet drop rate (e.g., 50% better than M-SFQ at load 0.8). On the other hand, the packet loss rate in M-DWRR with Set 3 is ≈ 80% improved over M-SFQ and DWRR. These noticeable results are achieved because the weight profile selection meets the pre-selected traffic profile. This minimized packet loss rate eventually maximizes the overall network throughput. Fig. 5 shows that DWRR and M-DWRR schedulers offer the same level of throughput as the M-SFQ and DWRR. These noticeable results are achieved because the adaptive weight policy allows a unique ONU-gripping to QoS traffic. Extensive simulation results confirm that the proposed schemes maintain the desired delay and jitter performance of high priority traffic and at the same time achieve a very fine degree of bandwidth allocation, and improved packet loss rate and network throughput.

C. Jitter Performance

Another positive contribution of our proposed schedulers is the ability to preserve the packet delay variation for EF services discussed in details in [5]. The jitter is represented by the packet delay variation of two consecutively departed EF packets from the same ONU in the same TW [5]:

\[ J_i = D_i - D_{i-1} \]  

(7)

Where \( J_i \) is the ith delay jitter within the window and \( D_i \) is the ith packet delay within the window.

Fig. 6 shows the probability density function (pdf) of EF service packet delay at full loading scenario in Set 1 (proved to provide best EF delay). It is shown that the EF delay sequence presents a dispersion with enough number of data points in a tail until 3.75 ms for DWRR scheduler and a centralization with all data points condensed before 2.5 ms for both M-SFQ and M-DWRR schedulers. The reason behind this difference between DWRR and M-DWRR is that M-DWRR efficiently re-allocates the remaining bandwidth to all CoS traffic. Thus, M-DWRR prevents the MP traffic (here, 40% weight) from monopolizing the bandwidth after the second scheduling visit; a case that occur using the DWRR scheduler.

V. CONCLUSION

Fair intra-ONU bandwidth allocation is a key issue in QoS EPON. In this paper, a novel decentralized bandwidth scheduler is presented, based on DWRR scheduling scheme. The scheme features low implementation complexity, hardware implementation ability, fairness assurance and possible inter-operation with any inter-OUN (DBA) scheme. Moreover, the adaptive weight policy allows a unique ONU-gripping to QoS traffic. Extensive simulation results confirm that the proposed schemes maintain the desired delay and jitter performance of high priority traffic and at the same time achieve a very fine degree of bandwidth allocation, and improved packet loss rate and network throughput.

REFERENCES