Bandwidth Allocation for Multiple Functional Splitting Options over TWDM-EPON Networks with Multi-ONU Customers

Oscar J. Ciceri[†], Carlos A. Astudillo [†], Zuqing Zhu[‡] and Nelson L. S. da Fonseca[†]
[†]Institute of Computing, State University of Campinas, Campinas, Brazil
[‡]School of Information Science and Technology, University of Science and Technology of China, Hefei China
Email: oscar@lrc.ic.unicamp.br, castudillo@ic.unicamp.br, zqzhu@ieee.org, nfonseca@ic.unicamp.br

***Abstract*—The support of Mobile Fronthaul (MFH) over Passive Optical Networks (PONs) poses significant challenges due to the stringent latency and bandwidth requirements of Functional Splitting (FS). This paper addresses the problem of Quality of Service (QoS) provisioning in next-generation Ethernet PON (NG-EPON) for the transport of traffic generated by multiple different FS options. We propose a PON bandwidth allocation algorithm that distributes the resources for the Optical Network Units (ONUs) serving Functional Split (FS) options based on their bandwidth and latency requirements in networks with customers renting/owning more than one ONU (multi-ONU customers). Simulation results show that our proposal significantly improves network resource utilization for multi-ONU customers, meeting the latency requirements of the different FS options while reducing the required bandwidth.**

***Index Terms*—Functional splitting, O-RAN, Quality of service, Resource allocation, Passive optical networks.**

I. INTRODUCTION

The 3rd Generation Partnership Project (3GPP) introduced Functional Split (FS) in 5G networks for separating the Radio Access Network (RAN) functions [1], which allows the reduction of costs and network requirements for fronthauling compared to the conventional centralized architecture, also known as Cloud Radio Access Network (CRAN) [2]. However, such separation still imposes stringent bandwidth and latency requirements for the Mobile Fronthauling (MFH) network, that vary as a function of the FS option (Table I).

To maximize profitability, Infrastructure Service Providers (InPs) may accommodate customers with diverse requirements on the same infrastructure, including traditional residential and enterprise customers and Mobile Network Operator (MNO). MNOs employ, among other technologies, Passive Optical Network (PON), which enables a fine-granularity transport service suitable to Mobile Fronthauling (MFH). In an MFH scenario, MNOs typically rent/own multiple Optical Network Units (ONUs) distributed in different regions of a city, involving residential and commercial areas [3]. These customers are called *multi-ONU customers* and may have ONUs serving different FS options (Figure 1). Such a multi-ONU customer scenario increases the spatiotemporal variability of the PON traffic.

While efforts have been made to accommodate multiple FS options over the same PON [4]–[7], existing strategies fail to exploit the spatiotemporal variability of the MFH traffic that occurs in multi-ONU customer scenarios. Moreover, traditional PON business models guarantee bandwidth to individual ONUs, which implies that multi-ONU customers cannot take advantage of the load imbalance among their ONUs, and, consequently, bandwidth can be wasted and costs increased. To satisfy the resource demands of ONUs that support FS options with strict delay constraints, such as FS Opts 6, 7, and 8, multi-ONU customers end up over-provisioning resources, increasing operational costs [8].

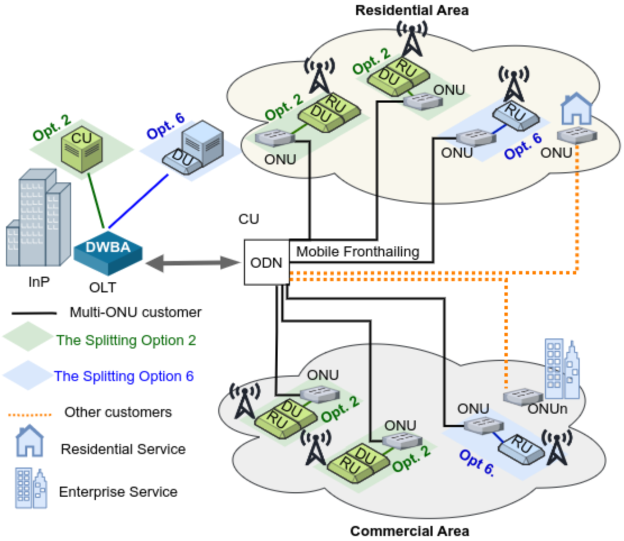


Fig. 1: An example of a PON deployment

In PONs, Dynamic Wavelength and Bandwidth Allocation (DWBA) algorithms are employed to provide bandwidth guarantees to the ONUs according to predefined Service Level Agreement (SLA) [9]. In the case of multi-ONU customers, the individual bandwidths of the ONUs to the same customer can be aggregated into a single SLA [10].

Indeed, in [11], a Dynamic Bandwidth Allocation (DBA) algorithm was proposed to support multi-ONU SLA models in TDM-EPON networks, which increases the statistical multiplexing gain for multi-ONU customers. In [1], we introduced an algorithm to take advantage of the spatiotemporal imbalance of MFH traffic in PONs with multi-ONU customers for the support of FS options with latency requirements as low as 250 μs. These algorithms distribute the unused bandwidth employing the compensation method [12], which either distributes the unused bandwidth on a per cycle basis [11] or retains

TABLE I: The functional split options and their requirements. HLS: The Higher-Layer Split; LLS: The Lower-Layer Split; D: Dependent; I: Independent; V: Variable; C: Constant. The requirements are tailored by following 3GPP TR 38.801 V14.0.0.

Option	**Split Point**	**Category**	**Layer Functions**	**User Traffic**	**Bit Rate**	**UL/DL Load** $[Gbps]$	**Delay Requirement**	**Max Delay** $[\mu s]$	**Traffic Type**	**Compensation Cycles**
O1	RRC-PCCP	HLS	3	D	V	$3/4$	Low Strict	10000	$T1$	$n1$
O2	PCDP-HRLC	HLS	3	D	V	$3.024/4.016$	Low Strict	10000	$T1$	$n2$
O3	HRLC-Low RLC	HLS	2	D	V	$< 3.024/< 4.016$	Low Strict	10000	$T1$	$n3$
O4	Low RLC-High MAC	HLS	2	D	V	$3/4$	Strict	1000	$T1$	$n4$
O5	High MAC-Low MAC	HLS	2	D	V	$3/4$	Strict	1000	$T1$	$n5$
O6	Low MAC-High PHY	HLS	2	D	V	$5.640/4.133$	Very strict	250	$T2$	$n6$
O7	High PHY-Low PHY	LLS	1	I	C	$10.1 \sim 86.1/16.6 \sim 86.1$	Very strict	250	$T3$	$n7$
O8	Low PHY-RF	LLS	1	I	C	$157.3/157.3$	Very strict	250	$T3$	$n8$

it for use in the next scheduling cycle [1]. However, unused bandwidth is allocated to ONUs based solely on the bandwidth request, without taking into account the delay requirements.

Due to the limitation of employing a few compensation cycles and the inability to allocate bandwidth considering delay requirements, previous algorithms cannot deal with the unbalanced nature of MFH traffic, leading to inefficient resource utilization and delay requirement violations. Consequently, ONUs supporting FS options with less stringent delay requirements may occasionally be prioritized over ONUs with FS options with more stringent delay constraints, and the latter may not have their latency requirements supported.

Thus, it is crucial to devise algorithms for allocating resources to different FS options in PONs with multi-ONU customers. Indeed, the efficient bandwidth allocation to distinct FS options in MFH scenarios with multi-ONU customers is still an open problem.

To address such a challenge, this paper introduces a DWBA algorithm for TWDM-EPON networks that allocates resources based on the requirements of the FS options. Depending on the FS option, each ONU can receive excess bandwidth from a predefined number of compensation cycles. Such an approach allows MFH ONUs with strict Quality of Service (QoS) demands to utilize the unused resources from other MFH ONUs of the same customer without compromising the guaranteed bandwidth of other customers in the PON.

Simulation results demonstrate that customers owing ONUs with distinct FS options can reduce the bandwidth required to meet the delay requirements compared to the baseline algorithms. This capability is useful in PON scenarios with MNOs renting/owning part of the PON to support multiple FS options (*e.g.* Opt. 7.2, Opt. 6, and Opt. 2).

The rest of the paper is organized as follows. Section II discusses related work. Section III describes the proposed DWBA algorithm. Section IV details the simulation model and the scenarios used and analyzes the results derived via simulations. Finally, Section V concludes the paper.

II. Resource Allocation Schemes for Supporting Mobile Traffic over EPON Networks

In TWDM-PON, Optical Line Terminals (OLTs) employ DWBA algorithms to dynamically distribute the PON bandwidth among the ONUs by assigning time and wavelengths for data transmission. The work in [9] proposes a solution for low-latency services in which the OLT allocates resources based on the reports of the ONU buffer occupancy (*i.e.* Scheduling Request (SR)). In [13], the OLT dynamically infers the traffic state from the buffer reports sent by the ONUs (*i.e.* Traffic-Monitoring). In [14], the OLT receives information on upcoming mobile traffic from the Distributed Units (DUs) in advance through dedicated signaling (*i.e.* Cooperative Interface (CO)). Moreover, in [11], the OLT redistributes the unused bandwidth among the ONUs of the same multi-ONU customer, which increases the statistical multiplexing gain.

Most DWBA algorithms support a single FS option [15]. Only a few handle traffic from diverse FS options over PON networks [2], [4]–[6]. The work in [4] employs the SR technique to support Fiber To The Home (FTTH) and High-Level Splitting (HLS) options, while the CO is used to support Low-Level Splitting (LLS) options. The work in [5] proposed a hybrid DBA in which bandwidth is allocated to MFH traffic employing the CO. In contrast, bandwidth is allocated to the ONUs carrying non-MFH traffic based on the Immediate Allocation with Colorless Grant (IACG) SR approach [16]. The work in [2] proposes a DBA algorithm that can support MFH and Best Effort (BE) data services. The algorithm utilizes three reporting variants: one based on SR for BE services and two alternative methods for MFH services, which take into account the variable and constant bit rate traffic involved in MFH services.

In previous work [1], we proposed a DWBA algorithm that leverages traffic imbalance to allocate resources among MFH ONUs of multi-ONU customers. It employs a two-cycle compensation method to distribute excess bandwidth. The CO and SR techniques were used to gather information about MFH and traditional traffic. However, to our knowledge, no solution has been proposed considering diverse FS options, delay constraints, and bandwidth demands for resource allocation in multi-ONU customer scenarios.

III. DWBA Scheme for Supporting Multiple FS Options

This section introduces the proposed DWBA scheme that allocates resources to the ONUs based on the FS options, called Resource Allocation with FS Options Support (*RAF-SOS*). It employs a compensation mechanism to allocate unused resources from recent scheduling cycles to MFH ONUs of a multi-ONU customer according to the requirements of their FS options. This feature significantly benefits MNOs leasing/owning part of the PON infrastructure from InP to support cell sites with different FS options.

Current Ethernet PON (EPON) DWBA algorithms for MFH do not allow the distribution of excess bandwidth among

ONUs of a multi-ONU customer based on their FS options, although Multi-ONU customers may have an *SLA* that specifies bandwidth guarantees for individual ONUs as well as for their groups of ONUs [11]. In these algorithms, the OLT distributes the unused bandwidth to the overloaded MFH ONUs based on the requested bandwidth, which is sent on the Report messages. Thus, it is not possible to prioritize the unused resources for MFH ONUs with strict low latency requirements. Such inability increases the bandwidth required by multi-ONU customers with diverse FS options, and, consequently, increases network costs. Therefore, improving the EPON DWBA algorithms is essential to reduce cost while providing a better bandwidth distribution, ensuring efficient utilization of resources for diverse MFH traffic.

RAFSOS differs from existing DWBAs in the way that the OLT allocates the unused bandwidth of the compensation cycles based on the FS options. *RAFSOS* defines a maximum number of compensation cycles for each FS option, as shown in Table I. The main idea is that the ONUs supporting FS with lower delay requirements have a higher number of compensation cycles compared to ONUs supporting FS with more flexible delay requirements. This allows *RAFSOS* to control the allocation of resources for each FS option based on latency requirements. The compensation cycles assigned to each FS option create a prioritization effect in the excess bandwidth distribution, where FS options with more strict requirements obtain more resources due to a higher number of compensation cycles than FS options with less strict latency constraints. In this way, multi-ONU customers can distribute the excess resources of recent scheduling cycles considering the delay requirements of their ONUs, allowing a better resource distribution and reducing the total bandwidth required to support various FS options.

The OLT maintains a *record* of the unused guaranteed bandwidth of the current and previous m scheduling cycles for each multi-ONU customer. This *record* is employed for later distribution of the excess bandwidth among the overloaded ONUs of the same customer, according to the requirements of their MFH traffic. The ONUs include the information of the served FS option on the standard Report message via the Class of Services (CoS) option. The OLT can determine the QoS requirements of the MFH traffic based on the reported CoS. Let l be the FS option between 1 and 8. The OLT associates a value $n_l \in [0, m+1], \forall l \in \{1, 2, ..., 8\}$ to each FS option. The n_l value defines the number of compensation cycles (*i.e.* excess bandwidth) in the *record* (*i.e.* bandwidth) assigned to FS option l. If the excess bandwidth distribution is disabled for a given FS option l, $n_l = 0$. For example, the ONUs serving FS option 6 and FS option 2, a suitable configuration for a multi-ONU customer, $n_2 < n_6$.

When a report message is received from an underloaded ONU, the OLT grants the requested bandwidth, and the unused guaranteed bandwidth (*i.e.* , excess bandwidth) of the current scheduling cycle is added to the *record*. When a report message arrives from an overloaded ONU, the OLT grants the maximum guaranteed bandwidth for that ONU plus an excess bandwidth, which is the minimum between the bandwidth needed and the excess bandwidth available of the last n_l scheduling cycles.

Algorithm 1: *RAFSOS* DWBA Algorithm

```
: $r_k = [0, 0, ..., 0], |r_k| = m, \forall k \in \mathcal{G}$
for each received report R from ONU i in cycle j do
    if ONU $i \in O_T$ then
        $W_i^{limited} = min(W_i^{max}, W_i^{reported})$
        GateGenerator($W_i^{limited}$)
    else
        BandwithAllocationSplit()
        if # of Report messages $R \in O_k$ received is $|O_k|$ then
            Remove the first element in $r_k$
            Add an element with 0 value in the last position of the
              list $r_k$
GateGenerator($W_i^{grant}$)
    Calculate $t_{txStart}$
    Calculate $f_i$ based on a wavelength allocation policy
    $Gate_i^j \leftarrow (W_i^{grant}, t_{txStart}, f_i)$
    Send $Gate_i^j$
BandwithAllocationSplit()
    $W_i^{required} = W_i^{reported} + W_i^{forecasted}$
    $W_i^{limited} = min(W_i^{max}, W_i^{required})$
    if $n_l = 0$ then
        GateGenerator($W_i^{limited}$)
    else if $W_i^{max} > W_i^{required}$ then
        $W_i^{excess} = W_i^{max} - W_i^{required}$
        $r_k[m] = W_i^{excess} + r_k[m]$
        GateGenerator($W_i^{Limited}$)
    else
        ExcessDistribution()
ExcessDistribution()
    Obtain the $n_l$ based on the CoS option of the ONU $i$
    $W_i^{Excess} = 0$
    for $n \leftarrow 0$ to $n_l$ do
        $W_i^{Excess} + = r_k[n]$
        if $W_i^{Limited} + W_i^{Excess} \geq W_i^{Required}$ then
            $W_i^{Excess^{-1}} = W_i^{Excess} - r_k[n]$
            $W_i^{aux} = W_i^{Required} - (W_i^{Limited} + W_i^{Excess^{-1}})$
            $r_k[n] = r_k[n] - W_i^{aux}$
            $W_i^{Excess} = W_i^{Excess^{-1}} + W_i^{aux}$
            Stop the iteration
        else
            $r_k[n] = 0$
    GateGenerator($W_i^{Limited} + W_i^{Excess}$)
```

RAFSOS is designed to employ multiple wavelengths and implement an adaptive polling cycle for dynamic resource allocation [9], meeting both the requirements of the IEEE 50 Gb/s EPON standard and the SLA demands of PON customers. Moreover, *RAFSOS* utilizes the widely adopted Cooperative Interface, proposed by Tashiro *et. al* in [14], which allows latency reduction while improving statistical multiplexing. Thus, the OLT can obtain in-advance information of the MFH traffic from the mobile network based on the Cooperative Transport Interface (CTI) that allows the exchange of control messages between the mobile and the PON systems [17].

A. *RAFSOS Algorithm*

Algorithm 1 summarizes *RAFSOS*. Let O the set of ONUs in the EPON in a given EPON $\mathcal{G}$ be the set of *multi-ONUs* customers, O_T the set of ONUs belonging to single-ONU customers, and O_k the set of ONUs belonging to the k-th multi-ONU customer. It employs $|\mathcal{G}|$ *records* to save the excess bandwidth values of the multi-ONU customers. The *record* for the k-th multi-ONU customer is $r_k = [E_m, E_{m-1}, ..., E_0]$, which maintains the values of the excess bandwidth of the current and previous m scheduling cycles.

For each Report message R received by the OLT, it is checked whether this message comes from a traditional customer O_T (Line 2) or from a multi-ONU customer (Line 5). If the Report message comes from a traditional customer, the granted transmission window is calculated using the limited policy (Line 3), which is the minimum between the requested bandwidth $W_i^{reported}$ and the maximum guaranteed bandwidth W_i^{max}. Afterward, the OLT issues and sends a Gate message to the ONU employing the $GateGenerator$ function (Line 10).

However, if the Report message comes from an ONU belonging to a multi-ONU customer, the $BandwithAllocationSplit$ procedure is applied (Line 6). The OLT calculates the required transmission window $W_i^{required}$ (Line 16) using the requested window $W_i^{reported}$ received in the Report message and the forecasted window $W_i^{forecasted}$ required for the packets that will arrive before starting the next transmission. The forecasted window is calculated based on the information provided by the Cooperative Transport Interface (CTI) Report message for 5G front hauling [18].

If the distribution of excess bandwidth among the ONUs of the k-th multi-ONU customer is disabled ($n_l = 0$, Line 18), a Gate message is issued and sent using the limited policy (Line 19). If the ONU is underloaded ($W_i^{required} \leq W_i^{max}$), it is fully served with its guaranteed bandwidth (Lines 20 and 23). The unused bandwidth ($W_i^{max} - W^{required}$) is added to the excess bandwidth *record* of the current scheduling cycle (Line 22). When the ONU is overloaded ($W_i^{required} > W_i^{max}$), the OLT applies the $ExcessDistribution$ procedure to distribute the excess bandwidth of the *record* (Line 25). If the sum of the maximum guaranteed bandwidth and the excess bandwidth allocated to the ONU i surpasses the required bandwidth ($W_i^{max} + W_i^{excess} > W_i^{Required}$), the ONU is fully served (Lines 35). Otherwise, the OLT completely allocates the excess bandwidth of the n_l cycles (Line 30).

Once all reports from the k-th multi-ONU customer are received in the current cycle (Line 7), the OLT updates the *record* r_k. The excess bandwidth of the cycle m is removed (Line 8), and an empty bandwidth is added to *record* (Line 9). Thus, the OLT can record the unused bandwidth of ONUs of the k-th multi-ONU customer for the next scheduling cycle.

B. Complexity Analysis

The complexity of the proposed algorithm is analyzed as follows. There is a loop in which each Report message is processed per cycle. Thus, the allocation is performed in $O(n)$, n is the number of ONUs in the PON. In this loop, there is an if-else condition. If ONU i is in O_T, it performs a constant number, O(1). If ONU i is NOT in O_T, it calls the $BandwithAllocationSplit$ process, which, if the ONU is underloaded, the OLT immediately calculates the bandwidth granted, incurring in $O(1)$. If the ONU is overloaded, the $ExcessDistribution$ process is employed, which performs a loop that iterates at maximum m times for each multi-ONU customer. Thus, the time complexity in the worst case is $O(n + m \times g)$, where g is the total number of overloaded ONUs in the multi-ONU customers.

C. Specific Variations of RAFSOS

Our proposed algorithm allows a differentiated allocation of excess bandwidth for ONUs supporting distinct FS options, and particular cases of *RAFSOS* are equivalent to well-known algorithms. When the excess bandwidth distribution is not enabled, the resulting bandwidth allocation is based on the Limited policy, which behaves as the First-Fit (FF) scheme [9]. If the excess bandwidth distribution is only from the current scheduling cycle for a given FS option l (*i.e.* $n_l = 1$), the algorithm performs as the MOS-IPACT algorithm [11]. If the distribution of excess bandwidth is from the current and past scheduling cycle (*i.e.* $n_l = 2$), the algorithm performs as the RALM algorithm [1]. In these cases, the distribution of network resources is based on the bandwidth demand without considering the FS requirement, negatively impacting the provisioning of FS that requires strict delay values.

IV. PERFORMANCE EVALUATION

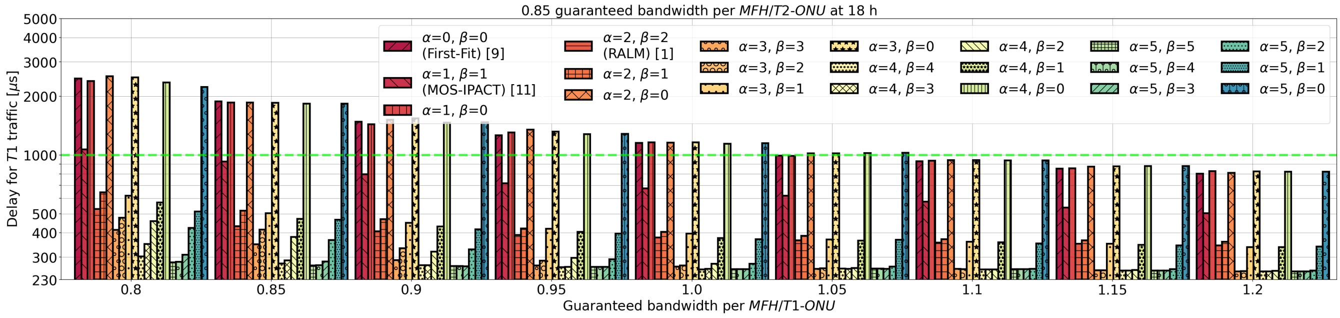
In this section, we assess the performance of *RAFSOS* by using the EPON simulator EPON-Sim, developed in Java and previously validated in [1]. EPON-Sim implements the FF DWBA algorithm together with the limited discipline [9]. This simulator also implements the Cooperative Transport Interface recently standardized in ITU-T Rec. Series G Supplement 71 (G.Sup.coDBA) and O-RAN Cooperative Transport Interface Transport Control Plane Specification (O-RAN.WG4.CTI-TCP.0-v02.00). *RAFSOS* was introduced in EPON-Sim, and the new version of the simulator was also validated extensively.

A. Simulation Model and Setup

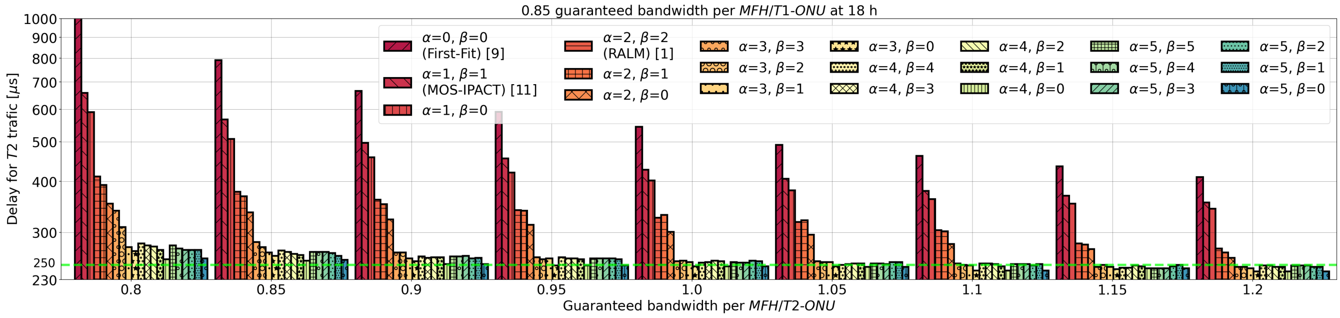
We simulated an InP employing a 50 Gbps EPON network and a 5 km radius covering. The OLT serves a set of 32 ONUs organized in a tree topology. Each ONU transmits on a single 25G wavelength that is allocated dynamically. The guard time between bursts of data from different ONUs was set to $0.624\,\mu s$ to avoid collisions. The maximum cycle length was set to $250\,\mu$s and the propagation delay to $5\,\mu$s/km. Each simulation scenario lasted 60 s and was replicated ten times.

We assumed that an MNO that rented parts of the PON from the InP to support MFH network traffic. The group of ONUs $O_M \subset O$; $|O_M| = 6$ belonging to this MNO served six BSs. Four ONUs served traffic T_1 produced by any of FS options from 1 to 5. The traffic loads generated from FS options 1 to 5 are identical (see Table I); the difference is the delay requirement, which is under 10 ms and 1 ms for FS options $1-3$ and $4-5$, respectively. The other ONUs supported traffic $T2$ produced by the BSs configured with FS option 6, the most demanding one with the largest bandwidth and lowest latency requirements among the upstream variable-rate split options. Hereafter, we refer to the MFH ONUs supporting the traffic T1 and T2 as $ONU\text{-}T1$ and $ONU\text{-}T2$, respectively.

The BS geographical distribution and their average traffic were obtained from processing a large dataset comprising data from a MNO in Dublin [19]. Three commercial and three residential BSs were selected within the coverage region of



(a) Delay for $T1$ traffic varying Guaranteed Bandwidth for $ONU\text{-}T1s$.



(b) Delay for $T2$ traffic varying Guaranteed Bandwidth for $ONU\text{-}T2s$.

Fig. 2: Delay performance of MFH ONUs for different resource allocation schemes with RAN functional split option 6 and any option from 1-5; Each bar shows 99.99th percentile delay values.

the InP. The $ONU\text{-}T1$ and $ONU\text{-}T2$ were equally allocated to these regions. There were 2 $ONU\text{-}T1$ and 1 $ONU\text{-}T2$ located in residential areas and 2 $ONU\text{-}T1$ and 1 $ONU\text{-}T2$ located in commercial areas (see Figure 1).

The peak load values of the kth MFH ONUs (P_k) were those in [1], and the same assumptions suggested in the 3GPP TR 38.801v14 document was used. The traffic flows were generated by using a Poisson distribution. The obtained P_k values for the $ONU\text{-}T2$ were 4836 Mbps and 4681 Mbps in the residential and commercial areas, respectively. For the $ONU\text{-}T1$ were (2220 and 1893) Mbps and (2065, and 2431) Mbps, in the residential and commercial areas, respectively. We simulated the scenario at 18:00 due to its high traffic variability. The guaranteed bandwidth B_k of the MFH ONUs was varied for the $ONU\text{-}T1$ and $ONU\text{-}T2$ from $0.8 \cdot P_k$ to $1.2 \cdot P_k$. Hereinafter, P_k is omitted from the B_k values.

To properly assess the performance of the proposed algorithm, the other 26 ONUs in the PON were conventional ones supporting traditional services. For those ONUs, the guaranteed bandwidth is equal to the remaining available bandwidth in the PON, divided by the total number of conventional ONUs. The loads produced by those ONUs comprised three different traffic classes: Expedited Forwarding (EF), Assured Forwarding (AF), and Best Effort (BE); the implementation details of these three traffic types can be found in [11]. The offered load on the conventional ONUs was set to 85% of their guaranteed bandwidth to replicate a network scenario with heavy traffic.

We compared the performance of the *RAFSOS* algorithm by varying the n_6 value from 0 to 5, and the n_h $\forall h \in \{1, 2, .., 5\}$ varied from 0 to n_6. Hereinafter, α and β are, respectively, n_6 and n_h. Then, the OLT can allocate for the $ONU\text{-}T1$ the excess bandwidth of a number of compensation cycles less than or equal to the used by the $ONU\text{-}T2$, which supports splitting options with strict QoS demands.

B. Results

This section assesses the proposed algorithm performance in supporting two different MFH traffic, $T1$ and $T2$, of a multi-ONU customer. The figures presented in this section show the 99.99th percentile delay values derived from 10 independent replications.

Simulation results show that the delay values of T1 and T2 traffic depend on the assured bandwidth per MFH ONU. Increased guaranteed bandwidth per MFH ONU reduces these delays. Thus, we examined how the guaranteed bandwidth of the $ONU\text{-}T1$ and $ONU\text{-}T2$ affected the delay of T1 and T2 traffic independently, under peak traffic conditions at 18:00h (Figure 2). There was no packet loss since the aggregated guaranteed bandwidth was higher than the aggregated offered load, adequately serving these traffics.

The proposed scheme attained the two possible delay requirements of the T1 traffic, *i.e.* < 1ms and < 10ms, with a guaranteed bandwidth per $ONU\text{-}T1$ greater than or equal to 85% of its peak hour average load value and $\beta \geq 1$ (Figure 2a). Additionally, our proposal met the delay boundary for the $T2$ traffic ($< 250\mu$s), with a guaranteed bandwidth per $ONU\text{-}T2$ greater than or equal to 95% and with ($\alpha \geq 3$ and $\beta = 0$) configuration (Figure 2b). On the other hand, the other configurations failed to produce satisfactory delays for the $ONU\text{-}T2$ with the same guaranteed bandwidth. These configurations (*i.e.* $\alpha \geq 3$ and $\beta \geq 1$) required guaranteed bandwidth per $ONU\text{-}T2$ greater than or equal to 105% to satisfy delays for the $ONU\text{-}T2$.

Our proposal enables the support of both traffic due to an increase in the utilization of unused resources at the multi-ONU customer level. Our proposal prioritizes those with stringent delay requirements by limiting the distributed unused bandwidth for $ONU\text{-}T1$ that requires a low strict delay level. Moreover, accounting for the excess bandwidth from previous cycles allows the OLT to fully serve the bandwidth demands of $ONU\text{-}T2$ in each scheduling cycle.

The utilization of the limited policy ($\beta = 0$) for the $T1$ traffic, which restricts the distribution of excess bandwidth in such traffic, resulted in $ONU\text{-}T1$ requiring 25% additional guaranteed bandwidth (*i.e.* , 110% of BW) to satisfy a delay lower than < 1ms compared to the *RAFSOS* configured with $\beta \geq 1$ (*i.e.* , 85% of BW). However, the bandwidth restriction on the $T1$ traffic had a positive effect on the delay of traffic $T2$, as it reduced the required bandwidth for $ONU\text{-}T2$ to 95% to achieve a delay of less than $250\,\mu$s when $\alpha = 3$.

Our proposal successfully supported the requirements of the two FS options; $\alpha \geq 3$ and $\beta \leq \alpha$ allow the support of delay requirements of $< 250\mu$s and < 1ms for the $ONU\text{-}T1$ and $ONU\text{-}T2$, respectively, with an assured bandwidth of 110% for both split options. On the other hand, none of the baseline algorithms met the delay requirements for both FS options. These algorithms met only the delay requirement of traffic T1 (*i.e.* ≤ 1ms). Specifically, ONUs required a guaranteed bandwidth of 105%, 85%, and 80% when First-Fit, MOS-IPACT, and RALM algorithms were used, respectively, to support $T1$ traffic.

The results show that *RAFSOS* not only satisfies the delay requirements but also significantly reduces the required guaranteed bandwidth per MFH ONU compared to the other algorithms. As a result, MNOs can benefit from increased bandwidth utilization and reduced costs. Moreover, it promotes better control of the unused bandwidth allocated to each splitting option, increasing resource allocation flexibility in resource allocation when the network condition varies.

The results described above illustrate the effect of resource sharing among functional splitting options. As expected, these gains do not compromise the guaranteed bandwidth of PON customers. This implies that a multi-ONU customer's MFH ONUs can achieve the delay budget by employing less assured bandwidth than those defined in the SLAs, even under high traffic variability. Moreover, multiple types of splittings can coexist in PONs by setting appropriate values of n_l, for a fair distribution of excess bandwidth.

V. Conclusion

This paper introduced a novel DWBA algorithm for the provisioning of QoS requirements of MFH ONUs supporting different FS options over EPONs. *RAFSOS* employs a bandwidth compensation scheme to account for the excess bandwidth from underloaded ONUs and distributes it based on the FS option of the overloaded ONUs. This capability supports delay requirements while using less guaranteed bandwidth than the baseline algorithms. Results demonstrate that our proposal provides lower delay values than existing algorithms, still reducing the guaranteed bandwidth. RAFSOS efficiently utilizes the network capacity while maintaining the desired delay requirements of the 5G mobile services.

In future work, we plan to employ techniques such as reinforcement learning to optimize the n_l values on the fly. Furthermore, we plan to evaluate additional scenarios to understand the potential limitations of our proposal.

References

[1] O. J. Ciceri, C. A. Astudillo, G. B. Figueiredo, Z. Zhu, and N. L. da Fonseca, "Passive optical networking for 5g and beyond 5g low-latency mobile fronthauling services," *arXiv preprint:2109.08989*, 2021.

[2] A. Zaouga, A. F. de Sousa, M. Najjar, and P. P. Monteiro, "Self-adjusting dba algorithm for next generation pons (ng-pons) to support 5g fronthaul and data services," *J. Light. Technol.*, vol. 39, no. 7, pp. 1913–1924, 2020.

[3] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud ran for mobile networks—a technology overview," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 1, pp. 405–426, 2015.

[4] H. Nomura, H. Ujikawa, H. Uzawa, H. Nakamura, and J. Terada, "Novel dba scheme integrated with sr-and co-dba for multi-service accommodation toward 5g beyond," in *Euro. Conf. on Opt. Commun. (ECOC)*. IET, 2019, pp. 1–4.

[5] S. O. Edeagu, R. A. Butt, S. M. Idrus, and N. J. Gomes, "A hybrid dynamic bandwidth allocation scheme operating with iacg and cooperative dba for converged fronthaul networks," *Opt. Fiber Technol.*, vol. 75, p. 103199, 2023.

[6] M. Waqar and A. Kim, "Qos assurance for pon-based fronthaul and backhaul systems of 5g cloud radio access networks," *Int. J. Internet Technol. Secur. Trans.*, vol. 11, no. 4, pp. 416–430, 2021.

[7] D. Harutyunyan and R. Riggio, "Flex5g: Flexible functional split in 5g networks," *IEEE Transactions on Network and Service Management*, vol. 15, no. 3, pp. 961–975, 2018.

[8] B. Ojaghi, F. Adelantado, A. Antonopoulos, and C. Verikoukis, "Impact of network densification on joint slicing and functional splitting in 5g," *IEEE Communications Magazine*, vol. 60, no. 7, pp. 30–35, 2022.

[9] S. B. Hussain, W. Hu, H. Xin, and A. M. Mikaeil, "Low-latency dynamic wavelength and bandwidth allocation algorithm for ng-epon," *J. Opt. Commun. Netw.*, vol. 9, no. 12, pp. 1108–1115, 2017.

[10] P. Alvarez, N. Marchetti, and M. Ruffini, "Evaluating dynamic bandwidth allocation of virtualized passive optical networks over mobile traffic traces," *Journal of Optical Communications and Networking*, vol. 8, no. 3, pp. 129–136, 2016.

[11] O. J. Ciceri, C. A. Astudillo, and N. L. d. Fonseca, "Dynamic bandwidth allocation with multi-onu customer support for ethernet passive optical networks," in *2018 IEEE Symposium on Computers and Communications (ISCC)*, 2018, pp. 00 230–00 235.

[12] I. Hwang, J. Lee, K. Lai, and A. Liem, "Generic qos-aware interleaved dynamic bandwidth allocation in scalable epons," *Journal of Optical Communications and Networking*, vol. 4, no. 2, pp. 99–107, 2012.

[13] D. Hisano, T. Kobayashi, H. Ou, T. Shimada, H. Uzawa, J. Terada, and A. Otaka, "Tdm-pon for accommodating tdd-based fronthaul and secondary services," *J. Light. Technol.*, vol. 35, pp. 2788–2796, 2017.

[14] T. Tashiro, S. Kuwano, J. Terada, T. Kawamura, N. Tanaka, S. Shigematsu, and N. Yoshimoto, "A novel dba scheme for tdm-pon based mobile fronthaul," in *OFC 2014*, 2014, pp. 1–3.

[15] H. Uzawa, K. Honda, H. Nakamura, Y. Hirano, K.-i. Nakura, S. Kozaki, and J. Terada, "Dynamic bandwidth allocation scheme for network-slicing-based tdm-pon toward the beyond-5g era," *J. Opt. Commun. Netw.*, vol. 12, no. 2, pp. A135–A143, 2020.

[16] M.-S. Han, H. Yoo, B.-Y. Yoon, B. Kim, and J.-S. Koh, "Efficient dynamic bandwidth allocation for fsan-compliant gpon," *J. Opt. Netw.*, vol. 7, no. 8, pp. 783–795, 2008.

[17] J. Maes, S. Bidkar, M. Straub, T. Pfeiffer, and R. Bonk, "Efficient transport of ecpri fronthaul over pon," in *2023 Optical Fiber Communications Conference and Exhibition (OFC)*. IEEE, 2023, pp. 1–4.

[18] S. Bidkar, R. Bonk, and T. Pfeiffer, "Low-latency tdm-pon for 5g xhaul," in *2020 22nd International Conference on Transparent Optical Networks (ICTON)*. IEEE, 2020, pp. 1–4.

[19] P. Francesco, F. Malandrino, L. A. Effenberger, DaSilva, and J. Chao, "Assembling and using a cellular dataset for mobile network analysis and planning," *Trans. Big Data*, vol. 4, pp. 614–620, 2017.