IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 3, 2019, accepted October 30, 2019, date of publication November 5, 2019, date of current version November 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2951581

A Distributed Control Framework for TDM-PON Based 5G Mobile Fronthaul

ASM DELOWAR HOSSAIN [1], (Senior Member, IEEE), AND ABDULLAH RIDWAN HOSSAIN [2]

[1]Department of Electrical and Telecommunication Engineering Technology, The City University of New York, New York, NY 11201, USA
[2]Electrical and Computer Engineering Department, New Jersey Institute of Technology, Newark, NJ 07102, USA

Corresponding author: ASM Delowar Hossain (ahossain@citytech.cuny.edu)

ABSTRACT The 5G mobile fronthaul (MFH) is bound to be overwhelmed by the explosive growth and stringent requirements of future wireless traffic. The time division multiplexed (TDM) passive optical network (PON) is a candidate for supporting such an MFH. Typically, such a TDM-PON utilizes upstream dynamic bandwidth allocation (DBA) schemes that are centralized at a distant central office (CO) which results in a performance that is not best suited for the MFH. Therefore, this work proposes a novel control framework by shifting the centralized DBA tasks away from the CO towards the closely clustered Remote Radio Heads (RRHs). By exploiting the interconnectivity of the RRHs, the access nodes exchange signaling and control information among themselves to facilitate a fully distributed upstream DBA without the involvement of the CO. The proposed framework separates the control plane from the data plane which paves the way for an efficient exchange of control messages, increased available upstream bandwidth and improved cyclic bandwidth allocation. This translates into a TDM-PON based MFH with lower latency, higher utilization and reduced packet loss.

INDEX TERMS 5G, dynamic bandwidth allocation, mobile fronthaul, radio access network.

I. INTRODUCTION

The unprecedented interconnection among wireless devices sets the stage for Internet of Things (IoT), requiring virtually unlimited bandwidth and very low latency. The 4G wireless network, although revolutionized the networking paradigm, cannot support the above requirements. To address this inevitable limitation, a new networking concept which aims at much lower latency, higher bandwidth and reduced power consumption than 4G was envisioned, namely 5G. It is yet to be standardized but its expectations and stringent requirements necessitates advanced technologies, such as the exploitation of unused high frequency bands (mmWave), smaller cells, cell densification, massive Multiple Input Multiple Output (MIMO) etc. [1], [2].

A Radio Access Network (RAN) is a segment of the network, consisting of base stations (BSs), that wirelessly connects the user equipment (UE) to the core network. Traditional BSs are composed of mainly two key elements: the Base Band Unit (BBU) and the Remote Radio Head (RRH) which is connected to an antenna. In general, the BBU is responsible for digital signal processing and the RRH is responsible for transmitting and receiving radio frequency (RF). An RRH is a device that, while receiving, collects the wireless signal and forwards it to the BBU for processing; while transmitting, it performs the reverse function. The placement configuration of the BBU and RRH can result in mainly two types of architectures, namely the Distributed RAN (D-RAN) and the Centralized-RAN (C-RAN) [also referred to as *cloud or converged* RAN] [3]. In the D-RAN architecture, the BBU and the RRH are integrated and placed at each antenna site and function as a single unit. In contrast, the C-RAN places the two elements at separate locations; the BBUs are centralized at one location but are linked to the RRHs via the mobile fronthaul (MFH) network (Fig. 1).

The future wireless network of countless devices will generate large volumes of RAN traffic which raises concerns regarding bandwidth saturation, transmission latency, energy constraint, and data security. As such, there are a number of D-RAN schemes proposed in the recent literature to address these issues. One such evolving scheme is Mobile Edge Computing (MEC), which is an alternative to centralized mobile cloud computing, that can provide resources for bandwidth

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

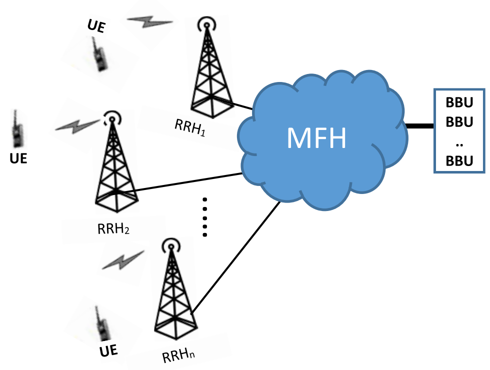

FIGURE 1. C-RAN MFH.

hungry and time sensitive future applications such as Real-time Traffic Management Systems, Road Safety, Smart Cities etc. [4]–[6].

The centralization of the BBUs in the C-RAN results in cost and energy savings due to the sharing of back-planes, power, computational and maintenance resources. It also facilitates advanced techniques for enhancing RAN performance via multi-cell processing, such as enhanced inter-cell interference coordination and coordinated multi-point (CoMP) transmission/reception; moreover, the antenna site designs are further simplified [7]–[9]. It is important to note that although the C-RAN simplifies the RRH and makes it economical, the system complexity and cost is shifted towards the MFH.

The unparalleled density of both wireless devices and BSs in 5G will result in massive volumes of traffic which will cause an MFH bottleneck. A fiber supported MFH is best suited for handling such massive traffic [10]–[15]. Among the various fiber supported networks, the Passive Optical Network (PON) [16], [17] is the economical choice for the 5G MFH. The MFH applications require low cost and power per bit; PON offers such by sharing cost effective and low maintenance optical infrastructure among countless mobile devices. Additionally, futuristic flexible-PON architecture paves the way for the enhancement of network throughput, data rates, latency, energy efficiency, and management through implementation of software-defined networking (SDN) and network function virtualization (NFV) [18]. Thus, a partnership between 5G and PON is reasonable.

There are two major PON options, Wave Division Multiplexing (WDM) and Time Division Multiplexing (TDM); the TDM-PON is still preferable for the future economical MFH because of its existing infrastructure, simplicity, ease of maintenance and sharing of fibers and communications equipment [11], [12]. A TDM-PON connects an Optical Line Terminal (OLT), located at a Central Office (CO), to several Optical Network Units (ONUs) at the customer end via fiber network (Fig. 2). This configuration requires some kind of arbitration scheme since a number of ONUs contend and share the fiber when transmitting to the OLT (upstream direction). Typical upstream scheme handles receiving

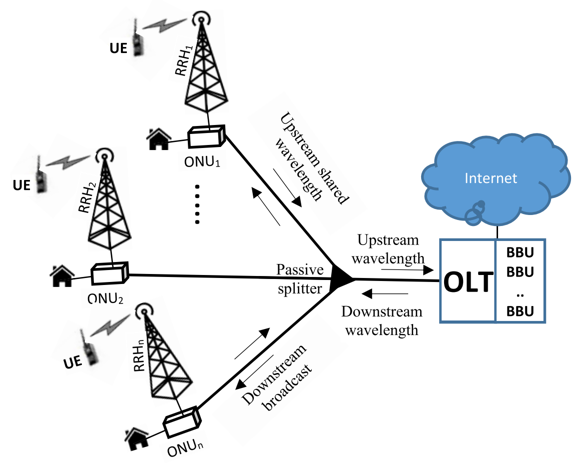

FIGURE 2. TDM-PON based MFH network.

transmission requests from the ONUs, allocating the available bandwidth based on those requests, and scheduling upstream transmissions as per the allocated bandwidth. Most of the proposed TDM-PON based dynamic bandwidth allocation (DBA) schemes are centralized at the OLT which is co-located with the BBUs; it is located far from the ONUs (connected to RRHs) requiring considerable round trip time (RTT). As the typical centralized scheme requires the exchange of control messages between the ONUs and the distant OLT, which usually results in unwanted delay and imposes other limitations as discussed in section III, is not best suited for the TDM-PON based MFH [19].

There are number of works presented in the current literature studying such centralized scheme of dynamic bandwidth allocation (DBA) for the TDM-PON based MFH. Among them, a DBA scheme for Constant Bit Rate (CBR) traffic utilizing a fixed bandwidth allocation (FBA) scheme was proposed [10], [13]. However, CBR is not suitable for wireless traffic characterization and the FBA scheme does not adequately address the need of bursty traffic [20]–[22]. Reference [14] has proposed the usage of a statistical function complementing the FBA scheme to accommodate non-CBR traffic. It utilizes periodical estimation of mobile traffic which is not very effective when trying to precisely capture minor variations in MFH traffic at every RRH. Reference [15] proposed integration of mobile scheduling with PON scheduling in advance. This approach shows an improvement for mobile traffic since it is scheduled to arrive at the ONU right before it transmits; *however, there remains the centralized PON DBA limitations suffered by ONUs (existing PON/wired traffic) which are not addressed here.* In other words, when the scheduling is observed in a holistic manner, including mobile and existing PON traffic, the limitation of centralization becomes apparent. Due to the above limitations, *it requires a viable alternative that can support 5G traffic along with PON traffic efficiently*. Therefore, this work proposes a *novel distributed control framework* for the TDM-PON based MFH, which essentially shifts the centralized DBA tasks from the OLT towards the ONU/RRH end. It is worthy to note that, in the proposed distributed framework, the TDM-PON is predominantly utilized for data transport as opposed to for both data and control in the centralized scheme.

The main contribution of this work is the proposed framework which splits the upstream control plane from the data plane (within the context of upstream DBA), where the control messages are exchanged among the RRHs while data is exchanged between the OLT and ONUs. The advantages of the proposed framework are summarized below:

(i) As the OLT is not participant in the DBA process, it does not exchange control messages with the ONUs resulting in minimized control overhead. This increases network utilization and PON resources for both upstream and downstream.

(ii) The RRHs wirelessly exchange DBA control messages among themselves rather than between the OLT and ONUs. As a result, the long round-trip delay between the OLT and ONUs is replaced by the shorter propagation delay between the RRHs; this lends more current network information for effective decision making resulting in an improved network delay performance.

(iii) This framework enables a holistic approach for cycle bandwidth allocation by considering the bandwidth requests of all ONUs and redistributing unallocated (surplus) bandwidth to the bandwidth hungry ONUs. This allows the network to serve more traffic within a given cycle and minimize the average upstream queueing delays.

(iv) Additionally, the following two capabilities are available:

  a. Local traffic can be transported wirelessly via the RRHs instead of through the OLT (with minimal network modifications); this lowers local traffic delays and increases PON resources for both upstream and downstream traffic.

  b. Advanced protection and restoration mechanisms are possible. In the event of a distribution fiber failure for an ONU, error messages can be sent from that ONU to a neighboring ONU wirelessly (via RRHs) so that the OLT can be notified to initiate recovery measures.

The rest of the paper is organized as follows: sections II and III provide brief overviews of MFH and TDM-PON respectively, section IV illustrates the proposed control framework, section V discusses the simulations and results, and section VI offers closing remarks.

II. OVERVIEW OF THE MOBILE FRONTHAUL

One of the key elements in the envisioned C-RAN architecture is the MFH, which transports the RRH's traffic originating at the UEs to the BBUs at a centralized location (Fig. 1). The exact location of the BBU placement is still under research [23]. This innovative C-RAN paradigm exploits the functional separation of conventional BSs, where the BBU interfaces with the backhaul network and performs physical layer digital processing of baseband signals and other upper layer functions. The RRH on the other hand interfaces with antennas and performs remaining physical layer functions such as frequency up/down-conversion, power amplification, digital-to-analog/analog-to-digital conversions etc. The popular approach in terms of interfacing between the RRHs and BBUs is via the Common Purpose Radio Interface (CPRI) [20] which is supported by digital radio-over-fiber (D-RoF) [21]. The CPRI utilizes CBR that can support different types of traffic with different requirements. While it is a popular industrial agreement, it imposes high overhead [21], [22] that is burdensome for the MFH. To ease the concern of the MFH traffic load, many alternative initiatives are underway, such as various Medium Access Control-Physical (MAC-PHY) layer functional splits [24]–[26], data compression methods [10], [27] etc. A detailed discussion can be found in the above references.

The massive MFH requires an optical network, preferably TDM-PON, to connect the standalone RRH cell sites to the BBUs. In the downstream direction, the BBU broadcasts the mobile traffic via the OLT to the ONUs over the fiber network; subsequently, the ONUs forward traffic to the connected RRHs to be transmitted to the UEs (Fig. 2). Conversely, in the upstream direction, the traffic from the UE is received by the RRH and forwarded (after minimal processing) to the connected ONU; the ONU then transmits the traffic over the shared fiber network to the OLT/BBU. Since the upstream fiber network (PON channel) is shared by several ONUs, some form of DBA is needed for the ONUs to have access to the channel. Typically, the DBA is centralized at the distant OLT which cyclically assigns upstream bandwidth to the ONUs as per their requests. Consequently, the *upstream MFH performance greatly depends on the upstream TDM-PON DBA* which is discussed next.

III. TDM-PON AND ITS LIMITATIONS

TDM-PON supports bidirectional transmission using two wavelengths; one for traffic from the OLT to the ONUs (downstream) and another from the ONUs to the OLT (upstream). For the downstream transmission, a passive splitter links the OLT to the ONUs through the distribution fiber network. The OLT broadcasts to all the ONUs; however, each ONU processes the traffic that matches its Medium Access Control (MAC) address only. For the upstream transmission, the transmission time slot and start time for each ONU are typically scheduled at the OLT in a centralized manner utilizing a DBA scheme. The DBA scheme takes advantage of the multi-point control protocol (MPCP) defined by the IEEE 802.3ah standard [28]. For the DBA tasks, the MPCP relies on two control and management messages to facilitate the exchange of information between the OLT and ONUs; they are known as REPORT (from ONUs to the OLT) and GRANT (from the OLT to ONUs). The ONUs send REPORTs to the OLT requesting transmission bandwidth and in response the OLT broadcasts GRANTs to the ONUs granting bandwidth as determined by the DBA scheme. Subsequently, each ONU initiates upstream transmission within the granted time window.

Many centralized DBA schemes (MPCP based) were proposed for TDM-PON in recent literature. Among them is the conventional DBA in [15] that offers a bare-bone picture of the centralized framework. Since this work intends to identify the core issues related to such a centralized framework (that limits MFH performance) to propose a viable distributed solution, it is appropriate to discuss the conventional DBA scheme in brief. It is an OLT-based centralized polling scheme based upon the exchange of REPORT and GRANT messages. It is implemented using an interleaved polling technique, where the succeeding ONU is polled prior to the completion of the upstream transmission from the preceding ONU. It is a cycle-based DBA, where a cycle (T_{CYC}) is the time it takes to serve all the ONUs. The cycle time is confined within a certain minimum (T_{CYC_MIN}) and maximum time (T_{CYC_MAX}). As the cycle time translates to cycle bandwidth, the cycle bandwidth is confined within a certain minimum and maximum as well. This cycle bandwidth is divided among the ONUs. An ONU is granted the reported bandwidth (bytes) as long it does not exceed the preset maximum (B_{MAX}) for an ONU. If ONU_i reports a bandwidth need of R_i, then the OLT grants bandwidth, $B_{Granted}$, as set below:

$$B_{Granted} = \begin{cases} R_i & \text{if } R_i \leq B_{MAX} \\ B_{MAX} & \text{if } R_i > B_{MAX} \end{cases}$$

B_{MAX} corresponds to T_{CYC_MAX}:
$B_{MAX} = \frac{1}{N}[R_{PON} * (T_{CYC\_MAX} - (N * T_G))]$, where
N = number of ONUs,
R_{PON} = PON transmission rate,
T_{CYC_MAX} = maximum cycle time, and
T_G = guard duration between successive ONU transmissions.

The typical centralized DBA algorithms like the conventional ones are OLT-based, where the OLT keeps track of the overall network information for bandwidth calculation purposes (especially for upstream transmission). Due to the centralization of the DBAs at the distant OLT, such DBA schemes suffer from certain limitations such as:

(i) *longer queueing delay:* the REPORTs and the GRANTs experience long RTT due to the distance of the OLT. In a high-speed network with bursty traffic, it may result in a GRANT that is outdated (not reflecting the actual need of the ONU due to the elapsed RTT time during which the ONUs' queues may have admitted more packets); thus such outdated bandwidth requests contribute to longer queueing delays.

(ii) *reduction of available bandwidth*: the frequent REPORTs and GRANTs needed for the DBA burden the OLT and consume upstream and downstream bandwidth; this results in a reduction of available network bandwidth for data transmission (as some bandwidth is consumed by control functions), which in turn increases delay and decreases network utilization *in both the upstream and downstream directions.*

(iii) *inefficient bandwidth allocation:* typically, the OLT calculates bandwidth for ONU_i solely based on the ONU_i REPORT received in the previous cycle. The bandwidth calculation does not consider the REPORTs from other ONUs; hence, this process is not efficient at the overall network level. *Note: offline centralized DBAs (at the OLT) address this issue by making bandwidth decisions upon receiving reports from all the ONUs. However, it cannot support interleave polling in the granting cycles which tends to result in an idle upstream channel time between granting cycles* [29].

(iv) *inflexible bandwidth allocation:* in a cycle, some ONUs may request bandwidth more than B_{max} (maximum limit) while others may request less than B_{max}. For instance, ONU_i requests bandwidth of $R_i > B_{max}$, while ONU_j requests bandwidth of $R_{j<}B_{max}$; this results in an excess amount of bandwidth (B_{max} - R_j) available from ONU_j that can be allocated to ONU_i. However, the OLT does not have the collective ONU REPORTs for the current cycle and therefore, allocates bandwidth solely based on the REPORT of a single ONU (see limitation # iii above); consequently, the maximum amount of bandwidth allocated to ONU_i cannot be more than B_{max}. *Note* that this issue can be addressed by an offline centralized DBA, however, it can cause *idle time between upstream granting cycles* (as mentioned in limitation # iii).

As the MFH is supported by the TDM-PON, such a centralized DBA will limit its performance. Therefore, this work proposes a framework that addresses the above issues to improve the MFH performance. It is important to note that while utilizing the same TDM-PON, the proposed scheme's inherent advantages unlock the full potential of TDM-PON.

IV. PROPOSED CONTROL FRAMEWORK

For the proposed distributed scheme to work and offer optimum performance (discussed later), it requires interconnectivity among the RRHs. As part of the improved network infrastructure for 5G's new paradigm in this work, it assumes that each RRH is equipped with a simple physical interface to communicate to each other; this interface is low powered, low latency and requires small bandwidth. It is analogous to the Long-Term Evolution (LTE) X2 logical interface [30]. A detailed discussion of the physical layer is beyond the scope of this work.

Each ONU is connected to an RRH (*collectively called a node*). Such an interconnectivity among the RRHs can allow the ONUs to communicate with each other (*via the connected RRHs)*. For instance, when ONU_1 intends to communicate with ONU_2, it needs to send a message via the connected RRH_1 to the neighboring RRH_2, which then passes the message to the connected ONU_2. Such an indirect interconnectivity among the ONUs can be utilized to exchange control

<table>
<tr><td rowspan="2">Source ID</td><td rowspan="2">Destination ID</td><td>Wired/PON Traffic Queue Size</td></tr>
<tr><td>5G traffic Queue Size</td></tr>
</table>

FIGURE 3. Format of report packet structure.

(REPORT) information (Fig. 3) regarding their upstream bandwidth needs (without involving the distant OLT); thus, facilitating a fully distributed control plane for TDM-PON via the RRH (wireless) network. The proposed scheme utilizes such *distributed control framework* to exchange ONU control information for upstream bandwidth allocation (see Fig. 4 - Fig. 7), while reserving the PON primarily for data traffic purposes (control and data plane separated). Once the ONUs exchange the appropriate control messages with each other regarding their upstream transmission needs, each ONU simultaneously and independently carries out a common DBA algorithm and generates identical results for the bandwidth allocation (due to identical network information received by each ONU). In summary, an RRH lends wireless access to its connected ONU; using that capability, ONUs exchange upstream control information among themselves so that each of them can run a simple bandwidth division algorithm that generates upstream transmission timeslots for all ONUs. Then the ONUs initiate upstream data transmission in a sequential manner, thereby relieving the OLT from its centralized upstream DBA tasks. While such a distributed scheme primarily improves upstream performance, *it benefits the downstream channel as well* [31]. This is because the centralized task of upstream DBA at the OLT (using REPORT/GRANT) costs downstream bandwidth and OLT resources; such downstream resources are freed up to support downstream tasks. This work limits the discussion to the upstream DBA framework; the downstream issues are beyond the scope this work. It should be noted that the proposed scheme primarily modifies the PON/MFH upstream control framework for the DBA only, the rest of the C-RAN functionality remains as is (centralized). The proposed system can work normally with a centralized scheme by simply disabling the distributed control frame exchange feature among the RRHs and restoring to the default centralized scheme (where the RRH receives and forwards information from the UE to the ONU to the OLT and vice-versa).

The proposed approach offers the following advantages:

(i) *Efficient resource allocation*: since the requests of all the ONUs are considered in the bandwidth calculation, this scheme can redistribute any surplus cycle bandwidth to the needy ONUs; thus, granting the needy ONUs sufficient bandwidth to serve their queues in a single cycle.

(ii) *Minimized queueing delay:* due to the optimized resource allocation above (#i), the upstream traffic does not have to wait for the next cycle to be transmitted; this minimizes the average delay for the upstream traffic.

(iii) *Efficient reporting:* the propagation delay among the RRHs is less than the RTT between the OLT and ONUs. For instance, the RTT required for a control message exchange between the RRH and OLT for a 20km distance is about 200μs, which is larger than the delay encountered in the proposed approach. The lower wireless RTT allows for more current network information to be used for effective decision making, which results in lower latency. Additionally, since the DBA control plane is separated from the data plane, the wireless reporting for the subsequent PON cycle (data) is concurrent with the current PON cycle. This typically facilitates the completion of reporting ahead of the next PON cycle; therefore, *no time from the PON channel is wasted* awaiting DBA outcomes (see Fig. 7).

(iv) *Improved network utilization:* since there is no exchange of DBA related messages between ONUs and the OLT, the PON bandwidth is *saved in both upstream and downstream directions*. This offers better network utilization and lesser delay in upstream direction. Additionally, the OLT resources and downstream channel are not burdened with the upstream PON DBA tasks and *such resources can be used to support other functionalities, including improving downstream performance*.

(v) *Improved inter-ONU/RRH performance:* when required, the inter-ONU/RRH traffic (local user traffic) can be transported wirelessly instead of via fiber to the OLT (local traffic travels upstream to the OLT to be broadcasted back downstream to the ONUs/RRHs for users in the same network); *this results in a lower delay for local traffic*. Furthermore, it *increases the available PON resources in both directions*. It is important to note this feature can burden the wireless network and may require some hardware and software modifications.

(vi) *Protection and restoration*: in case of a distribution fiber failure for an ONU, it can send an error message through its RRH to the neighboring RRH which can pass it on to the corresponding ONU; that ONU can then notify the OLT (to initiate failure recovery). Such details will be addressed in a future work.

As the distributed framework offers a number of advantages, it comes at a cost. Since this framework is not native to the typical TDM-PON based MFH, some minor modifications are required. In general, for the RRHs to communicate with each other, it requires the following additional elements: software, low powered wireless interface, nominal wireless spectrum, minimal signal processing capability and power. The added hardware and software require periodic maintenance. Such modifications are justified by the advantages of the proposed framework (see section V).

A. RRH REPORTING

Firstly, each of the RRHs starts the process of neighbor discovery to map the network topology. Once the RRHs (along with the connected ONU) are aware of the network mapping, each ONU cyclically exchanges its queue status (via the REPORT message) with other ONUs through the RRHs; this allows each ONU to obtain information from all the other

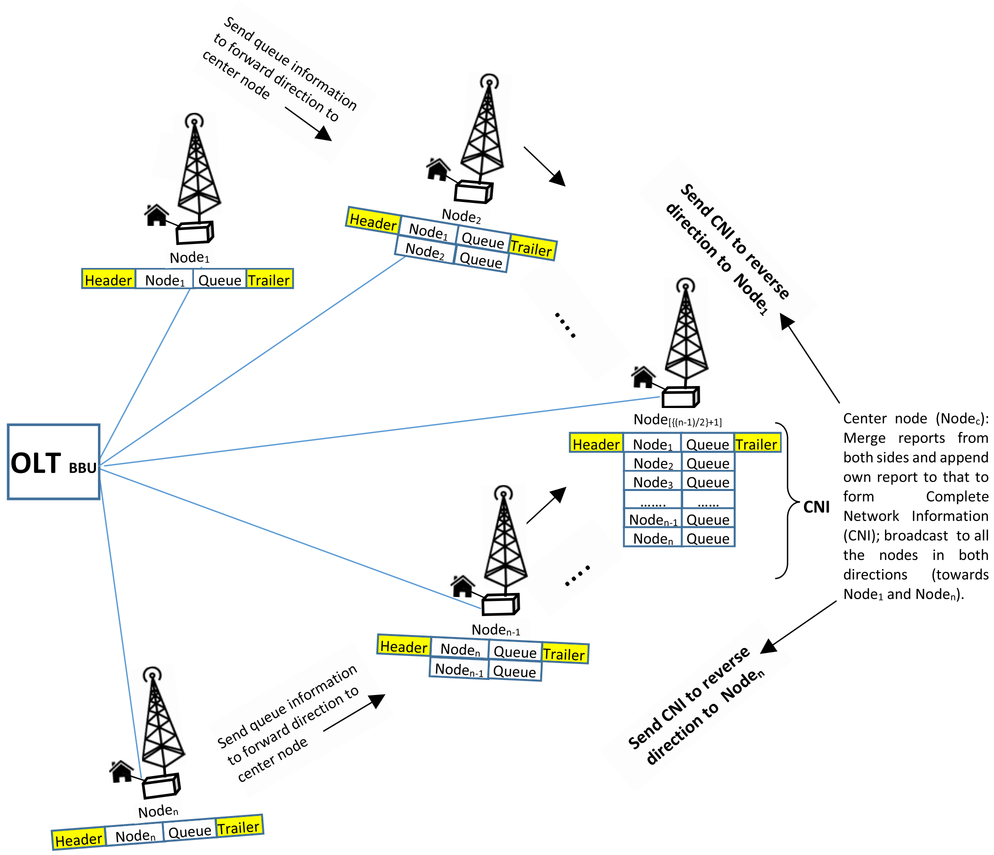

FIGURE 4. Overall reporting process.

ONUs. The wireless report cycle is initiated at both endpoints of the network (consider a symmetrical network with an odd number of nodes for simplicity as in Fig. 4). *As mentioned previously, each paired ONU and RRH is called a node, where the RRH has the wireless capability that is utilized by the ONU which has the capacity of simple bandwidth calculation.* The first node (Node$_1$) transmits to the second node (Node$_2$) and so on until it reaches the center node (Node$_c$); concurrently, the last node (Node$_n$) will transmit to the previous node (Node$_{n-1}$) and so on until Node$_c$. In other words, the transmission originates at both ends of the network and propagates to the center node (thus one half is transmitting in the forward direction while the other half is doing so in the backward direction). Although for simplicity, an odd number of nodes is considered, in the event a network has an even node count, there will be two center nodes, that is Node$_{n/2}$ and Node$_{(n/2)+1}$.

Fig. 4 depicts the overall reporting process; Fig. 5 and Fig. 6 illustrate the corresponding timing diagram and flow chart respectively. Fig. 7 demonstrates the timing of two concurrent processes, the wireless (control) and optical (data) cycles. The reporting stages are executed as detailed below.

(i) *Report collection*

a) *Node*$_1$: the first node (Node$_1$) sends its REPORT packet to the second node (Node$_2$) which appends its own REPORT to that of the first and forwards it to the third node (Node$_3$). This is repeated at each node until the aggregated REPORT reaches the center node (Node$_c$).

b) *Node*$_n$: simultaneously, the above process is carried out starting from the last node (Node$_n$) to Node$_c$ in the backward direction. Specifically, Node$_n$ sends its REPORT packet to Node$_{n-1}$ which appends its own REPORT to that of the previous node (in this case Node$_n$) and forwards it to the next node (Node$_{n-2}$) towards the direction of the center node (Node$_c$).

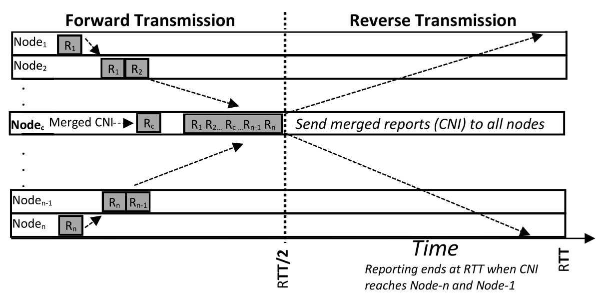

FIGURE 5. Wireless report timing diagram.

(ii) *Report merger:* upon receiving the aggregated REPORT from both segments of the network, Node$_c$

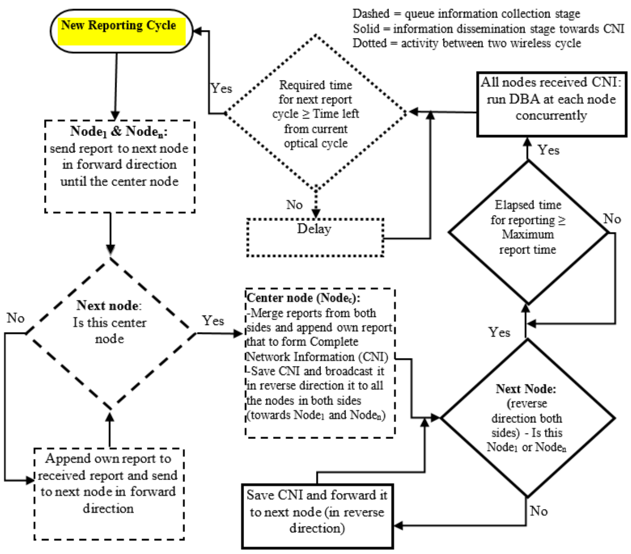

FIGURE 6. Reporting flow chart.

merges and appends its own REPORT to form what is called the Complete Network Information (CNI).

(iii) *CNI dissemination:* $Node_c$ sends the CNI towards both segments of the network by reversing the transmission directions described in stage (i). Thus, a CNI is sent in both directions of $Node_c$ to $Node_1$ and $Node_c$ to $Node_n$. Unlike stage (i), there is no appending carried out as this stage is solely for CNI distribution.

(iv) *Bandwidth division:* utilizing the CNI, each node concurrently carries out the same algorithm to generate timeslot assignments for the upcoming upstream PON (data) cycle.

Upon the completion of the above stages, the nodes transmit data as per their timeslot assignments avoiding collisions and more importantly, absolving the need for the OLT's intervention. As in any TDM scheme, a guard time is required between two successive ONU transmissions. Since the proposed scheme utilizes the RRH network for control tasks, it is important to observe its impact on the wireless network. The proposed scheme's CNI is equal to a *single report* packet with an aggregated payload (Fig. 4), carrying the node ID and queue size information; *it is not equal to* $\sum_{i=1}^{n}$ $REPORT_i$. Therefore, since it is a single packet, it results in a minimal burden on the RRH network bandwidth.

The timing specifications for the proposed scheme are discussed next as per the notations listed in Table 1. The wireless reporting cycle comprises of concurrent transmissions from both ends of the network towards the center node, which results in transmission among $\frac{(n-1)}{2}$ nodes. The network timing primarily depends on the signal propagation delay and the packet transmission time duration (proportional to the packet size). The signal propagation time duration (T_i^{fpt}) from the i^{th} node to the next node is defined as $T_i^{fpt} = \frac{D}{P}$, where D is the inter-nodal distance and P is the propagation speed. For simplicity, we assume that the network node layout is symmetrical with respect to the center node ($Node_c$) and the

TABLE 1. Proposed scheme's timing notations.

Notation	Description
F_{sz}	Packet size (bytes)
F_{OH}	Packet overhead size (bytes)
CNI	Aggregated packet size or complete network information (bytes)
N_{ID}	Node ID (bytes)
Q_{SZ}	Node queue size (bytes)
$Node_c$	Center node, c = $\frac{(n-1)}{2} + 1$, n≥ 3 and odd;
R	RRH transmission rate (Hz)
D	Inter-nodal distance (meter)
P	Propagation speed of radio signal (meter/second)
T_i^{fpt}	Inter-nodal propagation time for i^{th} node (seconds)
T_i^{ftt}	Forward transmit time duration for i^{th} node (seconds)
T_i^{rtt}	Reverse transmit time duration for i^{th} node (seconds)
T^{trc}	Total reporting cycle time duration (seconds)
T^{gce}	Grant cycle end time instant (seconds)
T^{gcs}	Grant cycle start time instant (seconds)
T_i^{gts}	Grant slot time duration for i^{th} node
T_i^{igt}	Inter-grant guard time duration for i^{th} node (seconds)
T_i^{nrs}	Report start time instant for i^{th} node (seconds)
T_i^{nrr}	Reverse report start time instant for i^{th} node (seconds)
T^{rcs}	Report cycle start time instant (seconds)
T^{rcr}	Reverse direction report start time instant (seconds)
T^{bct}	Bandwidth calculation time (seconds)

nodes are equally spaced; this results in a fixed propagation delay (T_i^{fpt}) between adjacent nodes in the network in any direction. The size of the report packet (F_{sz}) increases progressively in the forward direction at each subsequent node (due to appending) until it reaches the center node ($Node_c$); typically, it includes the queue size along with the node ID. For the sake of simplicity, we assume that in the forward direction, F_{sz} is fixed in both sides of the network and is equal to the forward accumulated packet size at the node prior to $Node_c$ (from either side- see Fig. 4): $F_{sz} = F_{OH} + \{\frac{(n-1)}{2} * (N_{ID} + Q_{SZ})\}$. Therefore, in the forward direction, the transmission times for all the nodes are equal to: $T_i^{ftt} = \frac{F_{sz}}{R}$.

Since it is a symmetrical network with concurrent transmissions from both ends towards the center (forward direction), the timing information is equal for both halves ($Node_1$ to $Node_c$ and $Node_n$ to $Node_c$). Therefore, demonstrating the network timing for one half should be sufficient; we choose to illustrate the timing information from $Node_1$ to $Node_c$. The total forward transmission duration until both sides reach $Node_c$ is $\sum_{i=1}^{\frac{(n-1)}{2}} (T_i^{fpt} + T_i^{ftt})$. The time instant of transmission of the i^{th} node equals to the total transmission duration of all the preceding nodes as denoted by $T_i^{nrs} = T^{rcs} + \sum_{j=1}^{i-1} (T_j^{fpt} + T_j^{ftt})$, where i≥2 and the cycle start time instant is T^{rcs}. Once the report packets from both sides reach $Node_c$, it subsequently appends its own packet; thus, completing the CNI from all the nodes for the current reporting cycle. The size of that packet (CNI) at $Node_c$ is: CNI $= F_{OH} + \{n * (N_{ID} + Q_{SZ})\}$.

At this point, the reverse transmission starts from $Node_c$ towards both sides ($Node_1$ and $Node_n$) to share the CNI with all the nodes. The elapsed time prior to the reverse transmission is T^{rcr} ($T_{cyc}/2$ in Fig. 5). Since the CNI is of a fixed size, the reverse wireless transmission time duration ($T_i^{rtt} = \frac{CNI}{R}$) at each node is same. Since the network is symmetrical, the reverse transmission timing from $Node_c$ to $Node_1$ is equal to the transmission timing from $Node_c$ to $Node_n$. Therefore, we demonstrate the reverse transmission timing for one of them ($Node_c$ to $Node_n$). The total duration of reverse transmission from ($Node_c$ to $Node_n$) is $\sum_{\frac{(n-1)}{2}+1}^{n-1} (T_i^{fpt} + T_i^{rtt})$ and the exact transmission start time of the i^{th} node equals to the total transmission time of all the preceding nodes $T_i^{nrr} = T^{rcr} + \sum_{j=1}^{i-1} (T_j^{fpt} + T_j^{rtt})$, where $i \geq \{\frac{(n-1)}{2}+2\}$ referring to ($Node_{c+1}$) and $T_i^{nrr} = T^{rcr}$ when $i = \left\{\frac{(n-1)}{2}+1\right\}$ referring to $Node_c$.

Therefore, the required time for a complete reporting cycle (forward and reverse transmission) is:

$$T^{trc} = \sum_{i=1}^{\frac{(n-1)}{2}} (T_i^{fpt} + T_i^{ftt}) + \sum_{\frac{(n-1)}{2}+1}^{n-1} (T_i^{fpt} + T_i^{rtt})$$

To ensure the most current queue size reporting in a cycle, it is desirable for a node to report as late as possible, meaning towards the end of the PON data cycle. The current PON grant cycle ends at

$$T^{gce} = T^{gcs} + \sum_{i=1}^{n} (T_i^{gts} + T_i^{igt}).$$

When the wireless reporting (plus bandwidth calculation time) for grant cycle k+1 is completed ahead of the of the ending of grant (data) cycle k, it results in optimal network performance (see Fig. 7). This means that the next cycle reporting start time is

$$T^{rcs(k+1)} = T^{gce(k)} - T^{trc(k+1)} - T^{bct(k+1)},$$

when

$$\sum_{i=1}^{n} (T_i^{gts(k)} + T_i^{igt(k)}) \geq (T^{bct(k+1)} + T^{trc(k+1)}).$$

Otherwise, the reporting for the $(k+1)^{th}$ cycle starts upon completion of the reporting of cycle as shown below:

$$T^{rcs(k+1)} = T^{trc(k)}.$$

B. INTEGRATED SCHEDULING AT ONU

Generally, there are two types of scheduling, the inter-ONU (DBA) and the intra-ONU (queue management). The inter-ONU scheduling at the OLT decides how much bandwidth is allocated to an ONU in a cycle so that it can transmit packets from its queue to the upstream direction. On the other hand, intra-ONU scheduling at the ONU controls the upstream traffic into and out of an ONU queue. Since both types of scheduling typically take place independently at two different locations, they are not optimized in terms of resource allocation. However, in the proposed scheme, *both types of scheduling are synchronized as they take place at the ONU, which can facilitate optimal bandwidth allocation at the queue level.*

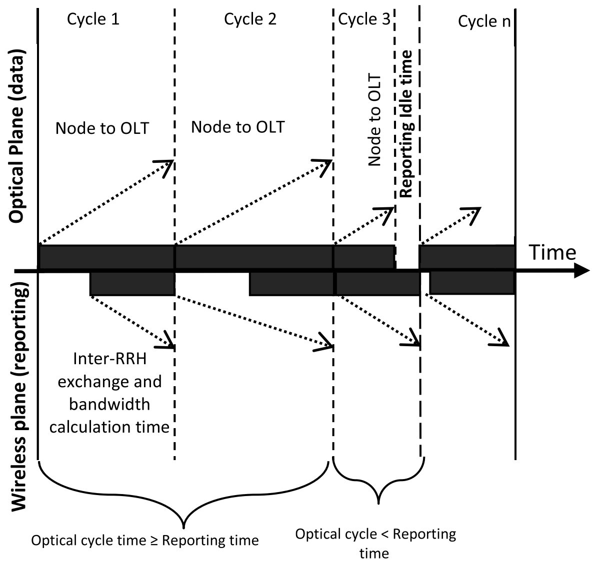

FIGURE 7. Overall upstream cycle timing diagram.

(i) Intra-ONU scheduling: Typically, the arriving traffic is admitted into a queue until the queue is full. Any traffic arriving after the queue is full results in packet drop. The queue is generally managed via first-in-first-out (FIFO) scheduling, where the packet that enters the queue first, exits the queue first. The queue management task is more involved when there are several classes of traffic arriving at the ONU queue. In this work, we utilize a relatively simple mechanism of *"priority queuing"* in a shared queue scenario; in practice it means that higher priority traffic has more right to be admitted into the queue than lower priority traffic. Furthermore, when the queue is full, a higher priority traffic can be admitted at the cost of a lower priority traffic from the queue. In our case, 5G traffic, compared to wired/PON traffic, is considered to be a higher priority traffic due to its stringent requirements. Therefore, under priority queueing, typical PON traffic from the queue can be dropped to make space for wireless traffic. In terms of transmitting traffic out of the ONU queue, a scheduler is needed as well. The function of this scheduler is to divide the ONU's allocated bandwidth among the classes of traffic in its queue. We employ a scheduler called *"strict priority"* for scheduling packet transmission. This mechanism transmits packets from the highest priority first, then the lower priority packets in the order of their importance. In our case, 5G traffic is transmitted first, then the wired traffic as per availability of bandwidth.

(ii) Inter-ONU scheduling: this mechanism is responsible for the distribution of cycle bandwidth among the ONUs. According to the ONU queue size in a cycle, ONUs can be divided into two groups: lightly loaded ONUs and heavily loaded ONUs. Lightly loaded ONUs in a cycle need bandwidth less than the maximum allowable bandwidth (B_{MAX}) while the heavily loaded ONUs need bandwidth greater

than B_{MAX}. During each PON cycle, the bandwidth calculation module keeps track of the cycle bandwidth unclaimed by the lightly loaded ONUs. This unclaimed excess cycle bandwidth is then divided among the ONUs that need bandwidth greater than B_{MAX} (heavily loaded ONU). This additional assignment is proportional to their excess need. Assume in a cycle, there are L number of lightly loaded ONUs ($R_i < B_{MAX}$) and H number of heavily loaded ONUs ($R_i > B_{MAX}$); the total excess cycle bandwidth due to the unclaimed bandwidth of the lightly loaded nodes is: $B_{Cycle\_Excess} = \sum_i^L (B_{MAX} - R_i)$. In contrast, the total excess need of the heavily loaded nodes is: $B_{Excess\_Need} = \sum_i^H (R_i - B_{MAX})$. Now the excess cycle bandwidth is proportionally distributed amongst the heavily loaded ONUs to increase their transmission bandwidth of that cycle as per [32]:

$$\Delta B_i^{extra} = B_{Cycle\_Excess} \left[\frac{R_i - B_{MAX}}{B_{Excess\_Need}} \right]$$

where ΔB_i^{extra} is the extra bandwidth assigned to ONU_i. The overall bandwidth assigned (B_{GH}) for a heavily loaded ONU_i is given by: $B_{GH} = \Delta B_i^{extra} + B_{MAX}$.

If R_i is the requested bandwidth (queue size) of ONU_i, then the bandwidth assigned using this DBA scheme is

$$B_{Granted} = \begin{cases} R_i & \text{if } R_i \leq B_{MAX} \\ R_i & \text{if } R_i > B_{MAX} \text{ and} \\ & B_{Cycle\_Excess} \geq B_{Excess\_Need} \\ B_{GH} & \text{if } R_i > B_{MAX} \text{ and} \\ & B_{Cycle\_Excess} < B_{Exces\_Need} \end{cases}$$

Note: DBAs that allow such distribution of excess bandwidth requires the requests from all the ONUs at the OLT, which typically demands extra time from the PON; however, the proposed scheme offers that inherently.

V. SIMULATIONS AND RESULTS

In this section, we present a comparative analysis of the upstream performance of the centralized scheme (referred to as "conventional" [15]) with the distributed scheme (referred to as "proposed"). The *objective is to compare the two control frameworks* of bandwidth allocation, one which is centralized at the OLT and the other which is distributed at the end nodes (ONU/RRH). Therefore, a *bare-bone centralized scheme* based on [15] is selected to compare with the proposed bare-bone distributed scheme. The centralized DBA enhancements such as Double Phase Polling (DPP) [33], Shortest Processing Time (SPT) first [34], Largest Number of Frames (LNF) first [35], Shortest Propagation Delay (SPD) first [36] etc. are not considered here since many of them are applicable to the distributed framework as well.

A. NETWORK MODEL

A packet-based network model was developed for the purpose of simulation. Each point in the simulation result corresponds to a sample of about 100 million packets averaged over three different runs *(wired and wireless traffic)*. The simulation focuses on a number of important network performance metrics, such as packet queuing delay, upstream channel utilization, packet loss ratio, queue size etc. In this model, both schemes have similar network parameters.

There are 16 ONUs in the network, which are connected by a fiber network to an OLT 20 km away. Each ONU is connected to an RRH via a full duplex fiber link of 1 Gbps and OLT via a shared fiber link of 10 Gbps. Note that the transmission speeds are selected arbitrarily; they do not have any specific relevance in terms of demonstrating improved performance of the proposed framework. The upstream traffic to each ONU (via RRH) is assumed to be bursty [37]; furthermore, the incoming traffic load at each ONU is assumed to be uneven. In other words, for a given average ONU load, half of the ONUs have a higher network load while the other half have a lower network load. The 5G traffic arriving at the ONU is assumed to be 1470 bytes as per [37]. The upper limit of an optical cycle duration (T_{CYC}) is 230μs, the guard time (T_G) that separates two consecutive ONU transmissions is 1 μs, and the queue size at each ONU is 50 MB. For the proposed scheme, the concurrent bandwidth/timeslot calculation time at each ONU (excluding reporting time) is ~20 μs and the average inter-RRH propagation time is 1.5 μs (inter-RRH distance is ~0.5 km).

At the presence of both 5G traffic and wired traffic in the network, few parameters are modified accordingly. For instance, the 50 MB queue at the ONU is shared by both the 5G traffic and the wired traffic. The incoming arrival rate for the ONU remains the same (1Gbps), however, it is divided evenly between the 5G traffic and wired traffic (500 Mbps each). The packet size of wired PON traffic varies from 64 bytes to 1518 bytes as per Ethernet standards.

B. RESULTS AND ANALYSIS

(i)5G Traffic Scenario: Fig. 8 shows the packet queuing delay versus the Aggregated Network Load (ANL) for both the conventional and proposed schemes. The proposed scheme demonstrates better results than the conventional scheme for all ANL. The proposed scheme's queueing delay remains below 127 μs until the ANL is 55% (0.55), which is well below the MFH delay requirement of about 300 μs [38]. However, for the same scenario, the conventional scheme displays up to 10 ms delay, which is two orders of magnitude higher than the proposed scheme. At low loads, the dominant cause for the conventional scheme's poor performance is the RTT (200 μs) which is inevitable due to the required exchange of report and grant messages between the ONUs and the OLT. At moderate loads, the dominant cause is the inefficient distribution of bandwidth which deprives the needy ONUs from receiving a grant more than the maximum limit set for a cycle (even though there may be surplus bandwidth from the lowly loaded ONUs which can be redistributed among the heavily loaded ONUs). Conversely, the improved performance of the proposed framework is due to the implementation of the DBA locally, which eliminates RTT and allows the redistribution of surplus cycle bandwidth

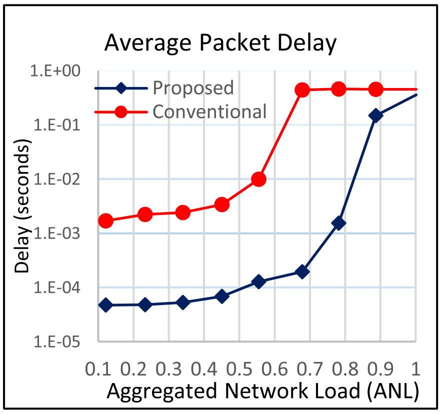

FIGURE 8. Average packet delay vs ANL.

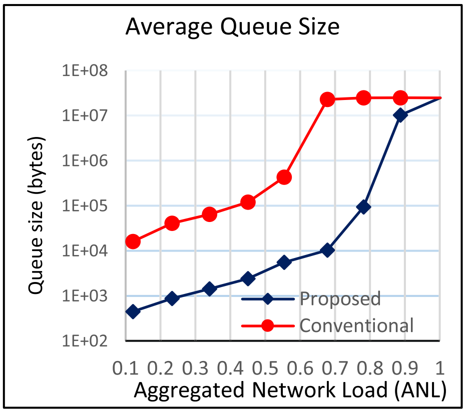

FIGURE 9. Average queue size vs ANL.

to the needy ONUs. Consequently, more packets are transmitted in a cycle (rather than waiting for the next cycle), which minimizes the packet queuing delay. Although the overall delay for both schemes increase as the ANL increases (until ∼85%), the proposed scheme continues to demonstrate better performance than the conventional scheme. As the ANL increases beyond 85%, the delay difference between the two schemes narrows. At 100% ANL, both schemes perform about the same; because, at a higher load, most ONUs request the maximum allowed bandwidth and there is no excess cycle bandwidth available to be redistributed among the needy ONUs. It is important to note that at ANL 70% the conventional scheme reaches network saturation, which results in maximum delay of 450 ms. However, the proposed scheme reached network saturation at 100% ANL.

Fig. 9 depicts the queue size as a function of ANL. The proposed scheme's queue size is smaller than the conventional scheme by an order of magnitude until the ANL is 55%. This implies that more packets are served in a given time in the proposed scheme (shorter delay) relative to the conventional scheme, which is consistent with the queueing delay figure (Fig. 8). The queue size of the proposed scheme continues to show improved performance than the conventional scheme until the ANL is 80%. As the ANL reaches 100%,

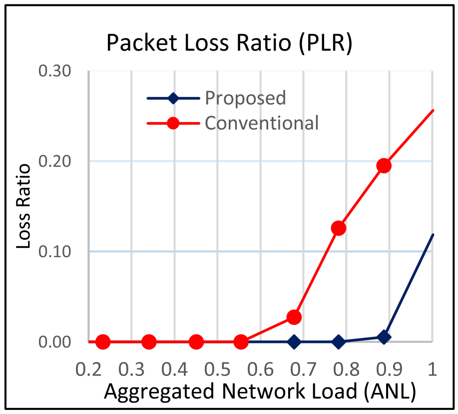

FIGURE 10. Average PLR vs ANL.

both schemes reflect maximum queue sizes indicating maximum network delays as shown in Fig. 8. It is important to note that the conventional scheme reaches network saturation at 70% ANL, which is much earlier than the proposed scheme. The underlying reasons were discussed above.

As the queue size impacts packet loss, we can see the correlation in Fig. 10. It shows the packet loss ratio (PLR) for both schemes against the ANL. The PLR is almost zero for both schemes until the ANL is 55%. This is because at lower loads the ONUs are adequately served and queues are not fully saturated to result in packet drops. Beyond that load, the conventional scheme shows growth in PLR reaching 25% at 100% ANL. On the contrary, the proposed scheme shows zero PLR until 90% ANL and reaches about 12% PLR at maximum network load (100% ANL). This figure aligns with network delay and queue size figures. The proposed scheme performs better because it facilitates the redistribution of excess cycle bandwidth to needy ONUs, which helps to free up queue space for the arriving packets (by serving more packets in a given time); the available additional queue space allows the admission of more packets into the queue which minimizes PLR. It is important to explain the cause of packet loss when the average queue size is still below the maximum. Occasionally the instantaneous queue size can reach the maximum, especially due to bursty traffic, while the average queue size remains below it. Any such instant of maximum queue size results in packet loss.

Fig. 11 depicts the upstream channel utilization. Until the ANL is 55%, there is no significant performance difference between the two schemes. This is because, at a lower ANL, most ONUs are served adequately for both schemes; this results in a similar ratio of cycle overhead versus packets served (for both schemes). It is important to note that each packet transmitted from an ONU requires a preamble (8 bytes) and an inter packet gap (12 bytes) for a total of 20 bytes of overhead per packet. Likewise, another overhead of 1 μs of guard time is required between two consecutive ONU transmissions. These overheads, regardless of the DBA scheme, prevents maximum network utilization. As the ANL grows to 100%, the proposed scheme's

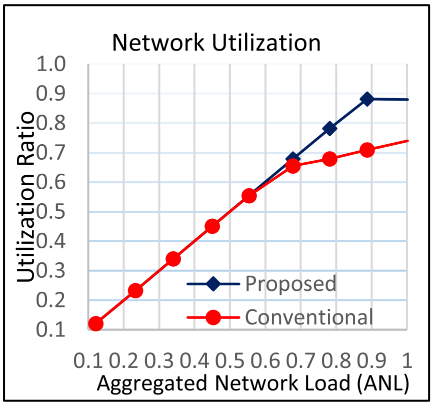

FIGURE 11. **Average utilization vs ANL.**

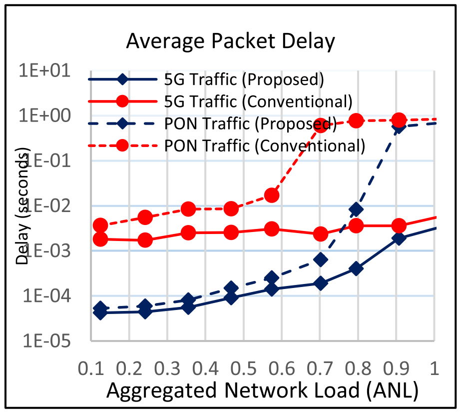

FIGURE 13. **Packet delay vs ANL.**

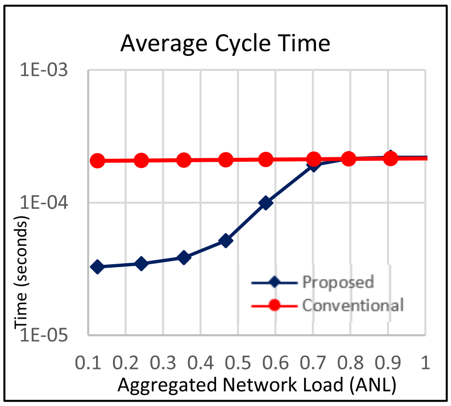

FIGURE 12. **Average cycle time vs ANL.**

utilization (∼88%) shows improved performance than the conventional scheme (∼ 74%) because the proposed scheme allows the redistribution of excess cycle bandwidth, which facilitates higher data transmission in a cycle. It means that lesser cycles are needed to transmit a certain number of packets which reduces cycle overhead allowing higher network utilization. The queue size, delay and PLR all are consistent with this figure.

Fig. 12 shows the cycle times for both schemes. In the conventional scheme, the cycle time is a little above 200 μs for low load, because the minimum RTT between the ONUs and the OLT is about 200 μs; and a cycle cannot be shorter than that. However, at higher loads, cycles may reach the maximum cycle length set by the scheme. In this simulation, the maximum cycle length was set to 230 μs considering the MFH delay requirement; it can however be up to 300 μs without significantly affecting the performance. In the proposed scheme, the cycle time is proportional to the network load. At higher loads, the cycles require larger bandwidth per node to transport more traffic per cycle which translates to longer cycle durations up to the maximum duration allowed.

(ii) 5G and Wired Traffic Scenario: Fig. 13 shows the packet queuing delay versus the ANL for the 5G traffic as well as the wired PON traffic. The proposed scheme demonstrates better results than the conventional one for both types of traffic. Since the proposed scheme's overall performance is better than that of the conventional scheme [as explained in section V.B.(i)], this advantage benefits both types of traffic for the proposed scheme. Between the two types of traffic, the 5G traffic is designated as the higher priority traffic (*"strict priority" transmission*) due to its stringent requirements. The 5G traffic's delay for the proposed scheme remains below 100μs (starting at 42 μs) until the ANL reaches 50%, while the conventional scheme's delay reaches 2.5 ms at the same ANL. The proposed scheme not only demonstrates superior performance, but also meets the 5G delay requirement. For an ANL of above 50%, the proposed scheme's queueing delay grows progressively and reaches about 3.3 ms for an ANL of 100% (network saturation). On the other hand, the conventional scheme's delay reaches 5.7 ms at network saturation. The underlying cause of performance improvement of the proposed scheme is same as that discussed previously. Additionally, since the 5G traffic gains from priority queuing when entering the queue and strict priority while exiting the queue, all the benefits of the proposed scheme is enjoyed by the 5G traffic (before the wired traffic).

Regarding the wired PON traffic, the proposed scheme performs better than the conventional scheme for all ANL in Fig. 13. The proposed scheme reaches about 160 μs (starting at 52 μs) for 50% ANL and reaches about 640 μs at 70% ANL. It is well within the range of PON traffic delay requirements [39]. In contrast, the conventional scheme's delay reaches about 9 ms (starting at 3.6 ms) at 50% of ANL and reaches about 600 ms at 70% of ANL. The performance difference is quite visible, especially at higher loads when the delay is of significant concern. As the ANL increases and reaches 100%, the delay saturates; the proposed scheme reaches a maximum delay of about 690 ms and the conventional scheme reaches about 830 ms. Therefore, even at the maximum network load, the proposed scheme not only performs better than the conventional one but also reaches maximum delay at a much higher load (90% ANL) than the conventional scheme (70% ANL).

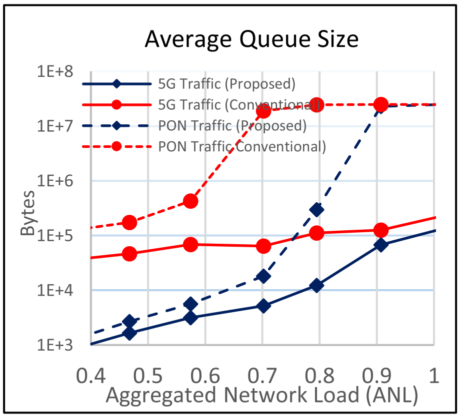

FIGURE 14. **Queue size vs ANL.**

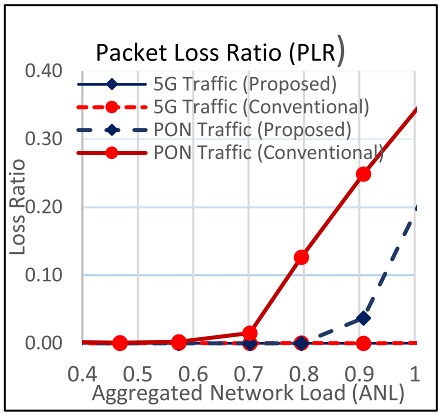

FIGURE 15. **PLR vs ANL.**

Fig. 14 depicts the queue size as a function of ANL for both types of traffic under two different schemes. As the queuing delay (Fig. 13) is proportional to the queue size, we observe similar advantages of the proposed scheme in Fig. 14. The proposed scheme exhibits smaller queue sizes than that of the conventional scheme for all ANL. Regarding the 5G traffic, at moderate load (∼50%), there is a noticeable difference in queue size between the two schemes; the queue size of the proposed and conventional schemes are 2 KB and 50 KB respectively. Although the queue sizes for both schemes increase at an ANL of 70%, the proposed scheme demonstrates better results (5 KB vs 64 KB). As the 5G traffic load increases to 100% ANL, the proposed scheme's queue size is about 130 KB while that of the conventional scheme is about 225 KB. This implies that more packets are served in a given time in the proposed scheme (smaller queue) than are served in the conventional scheme resulting in a smaller queueing delay for the proposed scheme; this agrees with the queueing delay figure.

In relation to the wired PON traffic, the proposed scheme displays a smaller queue size than that of the conventional scheme for all ANL; this aligns with the wired traffic's queuing delay in Fig. 13. At a moderate ANL (∼50%), the queue size of the proposed and conventional schemes are about 4 KB and 200 KB respectively. As the wired traffic grows above 60% ANL, the queue size grows rapidly for both schemes. At 70% ANL, the proposed and conventional schemes reach about 18 KB and 19 MB respectively. The queues reach maximum at 100% and 80% ANL for the proposed and conventional schemes correspondingly. The proposed scheme out performs the conventional scheme.

Fig. 15 depicts the PLR for the 5G and wired traffic for both schemes. In terms of the 5G traffic, both schemes perform similarly, because there is no packet loss for either scheme due to the higher priority of the 5G traffic. In priority queuing, there cannot be any 5G traffic loss if there is wired traffic in the queue that can be forced out to make space for 5G traffic. Therefore, both scheme's 5G traffic analysis demonstrates optimum performance at the expense of lower priority traffic (wired) which has to bear the loss. For the proposed scheme, the wired traffic packet loss is about 3% at 90% ANL and about 21% at maximum load. This agrees with Fig. 14, where it shows that the wired traffic queue size reaches maximum around 100% ANL causing significant packet loss. For the conventional scheme, the wired traffic packet loss is about 13% at 80% ANL and about 35% at full load. It is also consistent with Fig. 14, where the queue size approaches maximum at 80% ANL, which results in the beginning of significant packet loss. Note that the packet loss for the conventional scheme starts at a much lower load than that of the proposed scheme; the performance improvement is clear.

VI. CONCLUSION

This work demonstrated a novel control framework for the DBA for the TDM-PON based MFH to address the limitations of centralization at the OLT. The proposed framework shifted the DBA tasks from the OLT towards the ONU/RRH end thereby improving upstream network performance in terms of latency, packet loss and network utilization. Furthermore, *this framework can offer downstream performance improvements*. While this scheme displayed promising performance, it resulted in a nominal burden on the wireless network for the control tasks, specifically, by adding signal processing tasks at the RRHs; however, the improvements of the proposed scheme outweigh its limitations.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] I. A. Alimi, A. L. Teixeira, and P. P. Monteiro, "Toward an efficient C-RAN optical fronthaul for the future networks: A tutorial on technologies, requirements, challenges, and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 708–769, 1st Quart., 2018.

[3] T. Pfeiffer, *Next Generation Mobile Fronthaul Architectures*. Chennai, India: OFC 2015.

[4] X. Wang, Z. Ning, and L. Wang, "Offloading in Internet of vehicles: A fog-enabled real-time traffic management system," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4568–4578, Oct. 2018.

[5] Z. Ning, X. Kong, F. Xia, W. Hou, and X. Wang, "Green and sustainable cloud of things: Enabling collaborative edge computing," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 72–78, Jan. 2019.

[6] Z. Ning, X. Wang, J. J. Rodrigues, and F. Xia, "Joint computation offloading, power allocation, and channel assignment for 5G-enabled traffic management systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 3058–3067, May 2019.
[7] P. Rost, C. J. Bernardos, A. De Domenico, M. Di Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, "Cloud technologies for flexible 5G radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–76, May 2014.
[8] B. Haberland, F. Derakhshan, H. Grob-Lipski, R. Klotsche, W. Rehm, P. Schefczik, and M. Soellner, "Radio base stations in the cloud," *Bell Labs Tech. J.*, vol. 18, no. 1, pp. 129–152, 2013.
[9] K. Sundaresan, M. Y. Arslan, S. Singh, S. Rangarajan, and S. V. Krishnamurthy, "FluidNet: A flexible cloud-based radio access network for small cells," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 915–928, Apr. 2015.
[10] N. Shibata, T. Tashiro, S. Kuwano, N. Yuki, Y. Fukada, J. Terada, and A. Otaka, "Performance evaluation of mobile front-haul employing Ethernet-based TDM-PON with IQ data compression," *J. Opt. Commun. Netw.*, vol. 7, no. 11, pp. B16–B22, Nov. 2015.
[11] S. Zhou, X. Liu, F. Effenberger, and J. Chao, "Low-latency high-efficiency mobile fronthaul with TDM-PON (mobile-PON)," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 10, no. 1, pp. A20–A26, Jan. 2018.
[12] N. P. Anthapadmanabhan, A. Walid, and T. Pfeiffer, "Mobile fronthaul over latency-optimized time division multiplexed passive optical networks," in *Proc. IEEE ICC*, Jun. 2015, pp. 62–67.
[13] N. Shibata, T. Tashiro, S. Kuwano, N. Yuki, J. Terada, and A. Otaka, "Mobile front-haul employing Ethernet-based TDM-PON system for small cells," in *Proc. Opt. Fiber Commun. Conf. Exhibit. (OFC)*, 2015, pp. 1–3.
[14] T. Kobayashi, D. Hisano, T. Shimada, J. Terada, and A. Otaka, "Bandwidth allocation scheme based on simple statistical traffic analysis for TDM-PON based mobile fronthaul," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2016, pp. 1–3, Paper W3C-7.
[15] 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 *Proc. Opt. Fiber Commun. Conf. Exhibit. (OFC)*, Mar. 2014, pp. 1–3.
[16] *Gigabit-Capable Passive Optical Networks (G-PON)*, document ITU-T recommendations G.984, 2008.
[17] *Ten Gigabit-Capable Passive Optical Networks (XG-PON)*, document ITU-T Recommendations G.987, 2012.
[18] E. Kosmetos, C. Matrakidis, A. Stevdas, and T. Orfanoudakis, "An SDN architecture for PON networks enabling unified management using abstractions," in *Proc. ECOC*, Sep. 2018, pp. 1–3.
[19] B. Skubic, J. Chen, J. Ahmed, L. Wosinska, and B. Mukherjee, "A comparison of dynamic bandwidth allocation for EPON, GPON, and Next-Generation TDM PON," *IEEE Commun. Mag.*, vol. 47, no. 3, pp. S40–S48, Mar. 2009.
[20] *CPRI Specification V5.0*, CPRI, Bengaluru, Karnataka, 2011.
[21] A. Pizzinat, P. Chanclou, T. Diallo, and F. Saliou, "Things you should know about fronthaul," *J. Lightw. Technol.*, vol. 33, no. 5, pp. 1077–1083, Mar. 1, 2015.
[22] M. Fiorani, "On the design of 5G transport networks," *Photon. Netw. Commun.*, vol. 30, no. 3, pp. 403–415, Dec. 2015.
[23] F. Musumeci, C. Bellanzon, N. Carapellese, A. Pattavina, and S. Gosselin, "Optimal BBU placement for 5G C-RAN deployment over WDM aggregation networks," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1963–1970, Apr. 15, 2016.
[24] G. Mountaser, M. L. Rosas, T. Mahmoodi, and M. Dohler, "On the feasibility of MAC and PHY split in cloud RAN," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.
[25] *Project RAN Evolution: Further Study on Critical C-RAN Technologies*, Next Gener. Mobile Netw. (NGMN) Alliance, Frankfurt, Germany, Mar. 2015.
[26] N. Shibata, K. Miyamoto, S. Kuwano, J. Terada, and A. Otaka, "System level performance of uplink transmission in split-PHY processing architecture with joint reception for future radio access," in *Proc. 25th PIMRC*, Aug./Sep. 2015, pp. 1375–1379.
[27] K. Tanaka and A. Agata, "Next-generation optical access networks for C-RAN," in *Opt. Fiber Commun. Conf. (OSA) Tech. Dig.* Washington, DC, USA: OSA, 2015, paper Tu2E.1.
[28] *IEEE Standard for Information technology. Telecommunications and Information Exchange Between Systems. Local and Metropolitan Area Networks. Specific Requirements*, IEEE Standard 802.3ah, 2004. [Online]. Available: http://www.ieee802.org/21/doctree/2006_Meeting_Docs/2006-11_meeting_docs/802.3ah-2004.pdf
[29] M. P. McGarry and M. Reisslein, "Investigation of the DBA algorithm design space for EPONs," *J. Lightw. Technol.*, vol. 30, no. 14, pp. 2271–2280, Jul. 2012.
[30] A. Mohamed, A. Babiker, and N. Mustafa, "Functions of X2 interface," *Int. J. Sci. Res.*, vol. 5, no. 2, pp. 1048–1051, Feb. 2016.
[31] A. D. Hossain, M. Ummy, and A. R. Hossain, "Improving fiwi access network downstream performance: A distributed approach," *Int. J. Modern Eng.*, vol. 17, no. 2, pp. 20–25, 2017.
[32] C. M. Assi, Y. Ye, S. Dixit, and M. A. Ali, "Dynamic bandwidth allocation for quality-of-service over Ethernet PONs," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 9, pp. 1467–1477, Nov. 2003.
[33] S. Choi, S. Lee, T.-J. Lee, M. Chung, and H. Choo, "Double-phase polling algorithm based on partitioned ONU subgroups for high utilization in EPONs," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 1, no. 5, pp. 484–497, Oct. 2009.
[34] S. Bhatia and R. Bartos, "IPACT with smallest available report first: A new DBA algorithm for EPON," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2007, pp. 2168–2173.
[35] M. P. McGarry, M. Reisslein, C. J. Colbourn, M. Maier, F. Aurzada, and M. Scheutzow, "Just-in-time scheduling for multichannel EPONs," *J. Lightw. Technol.*, vol. 26, no. 10, pp. 1204–1216, May 2008.
[36] M. McGarry, M. Reisslein, F. Aurzada, and M. Scheutzow, "Shortest propagation delay (SPD) first scheduling for EPONs with heterogeneous propagation delays," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 6, pp. 849–862, Aug. 2010.
[37] A. Mikaeil, W. Hu, T. Ye, and S. B. Hussain, "Performance evaluation of XG-PON based mobile front-haul transport in cloud-RAN architecture," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 9, no. 11, pp. 984–994, Nov. 2017.
[38] S. Kuwano, J. Terada, and N. Yoshimoto, "Operator perspective on next-generation optical access for future radio access," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Jun. 2014, pp. 376–381.
[39] G. Kramer, B. Mukherjee, S. Dixit, Y. Ye, and R. Hirth, "Supporting differentiated classes of service in Ethernet passive optical network," *J. Opt. Netw.*, vol. 1, no. 8, pp. 280–298, Aug./Sep. 2002.

ASM DELOWAR HOSSAIN (M'95–SM'10) received the B.S., M.S., and Ph.D. degrees in electrical engineering from The City University of New York, in 1997, 1999, and 2007, respectively. He was a Research and Development Engineer in transportation industry, from 1997 to 2011. He joined the Electrical and Telecommunications Engineering Technology Department, New York City College of Technology, The City University of New York, in fall 2011. He contributed to many journal and conference publications. His research interest includes the wireless and optical access networks.

ABDULLAH RIDWAN HOSSAIN received the B.S. and M.S. degrees in electrical engineering from The City University of New York, in 2017 and 2019 respectively. He is currently pursuing the Ph.D. degree with the New Jersey Institute of Technology (NJIT). He contributed to eight publications in journals and conferences. His research interest includes the fiber optics communications.