Evaluating Bandwidth Efficiency and Latency of Scheduling Schemes for 5G Fronthaul over TDM-PON

Sarvesh Bidkar*, Konstantinos Christodoulopoulos, Thomas Pfeiffer, Rene Bonk
*Nokia Bell Labs, 8 Magirusstrasse, Stuttgart, Germany, *sarvesh.bidkar@nokia-bell-labs.com*

Abstract *We propose scheduling schemes for Cooperative DBA in upstream TDM-PON to enable 5G fronthaul services and evaluate their bandwidth efficiency and latency performance against FBA and conventional DBA using a co-simulation of a 25GS-PON MAC and a 5G system level simulator.*

1. Introduction

Centralized Radio Access Network (C-RAN) architecture in 5G systems can split the physical layer signal processing functions between the Radio Unit (RU) and the Distributed Unit (DU). The functional split enables statistically varying traffic between the DU and RU that needs to be transported over a fronthaul interface with certain latency (few 100s of μs) and jitter (few 10s of μs) requirements [1]. In this context, Time Division Multiplex-Passive Optical Network (TDM-PON) offers a cost-efficient solution for statistical traffic multiplexing over a point-to-multipoint passive optical infrastructure as shown in Fig. 1. However, the main challenge in TDM-PON is that the typical upstream latency, from the optical network unit (ONU) to the optical line terminal (OLT), i.e. from RU to DU, is in the milliseconds range due to the dynamic bandwidth assignment (DBA) process for conventional upstream TDMA scheduling, which is unacceptable for fronthaul. In [2], a low-latency TDM-PON operation was demonstrated using multiple bursts per ONU per upstream frame and fixed bandwidth allocation (FBA) for constant bit rate Common Public Radio Interface (CPRI) based fronthaul traffic. Such solution results in low latency, but is bandwidth inefficient for statistically varying types of fronthaul traffic (e.g. 7.2x C-RAN split) and also increases the burst overhead which reduces the overall PON capacity.

The typical DBA process used in TDM-PON is based on the traffic monitoring (TM) and status reporting (SR) mechanism which is too slow (updating bandwidth assignments every few ms) to timely react to dynamic variations of mobile fronthaul traffic on a fine time granularity (e.g. per radio slot of 0.5 ms). Therefore, it is prone to providing incorrect instantaneous bandwidth assignments resulting in variable buffering of the fronthaul traffic at the ONU – finally leading to higher latency and increased jitter.

To address the slow reaction time and incorrect bandwidth estimation of DBA, the authors of [3] proposed a cooperative DBA (Co-DBA) mechanism for fronthaul traffic. In this concept, the mobile cell scheduler conveys the future uplink scheduling information to the Co-DBA process in advance of the actual uplink transmission from the cell. With accurate synchronization between the radio and the PON systems, the OLT can provision the exact amount of bandwidth at the exact time so that the fronthaul traffic is not buffered in the corresponding ONU while waiting to receive upstream bursts on the PON. The Co-DBA is described in ITU-T G.Sup.71 [4] and one of its use case is considered for the Cooperative Transport Interface (CTI) defined by ORAN for mobile fronthaul transport [5].

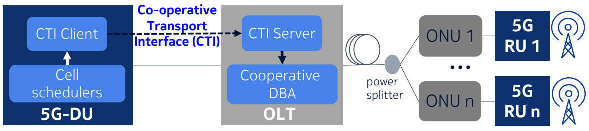

Fig. 1: C-RAN fronthaul architecture over TDM-PON

In this paper, we focus on the upstream scheduling of a Co-DBA enabled TDM-PON and propose two schemes specifically for fronthaul traffic. We evaluate and compare the bandwidth efficiency and latency performance of the proposed scheduling schemes with DBA and FBA through co-simulation of the TDM-PON Medium Access Control (MAC) and the 5G radio system level simulator.

2. Scheduling schemes

The typical DBA process in TDM-PON can be split into 2 steps: (1) *bandwidth assignment* that decides the #bits to be assigned over a time period (e.g. multiple PON frame duration) based on service configuration input and bandwidth demand estimation, (2) *bandwidth allocation* that decides how the assigned bandwidth is realized with bursts over that period (e.g. by specifying burst duration and frequency). The latter is realized with bandwidth maps (BWmap) that define the exact burst timing and duration per upstream frame. Depending on the accuracy of the bandwidth assignment w.r.t. the traffic demand and specific bandwidth allocation parameters, there is a trade-off between PON bandwidth utilization and latency. For example, assigning lower than demanded bandwidth leads to lower bandwidth utilization but results in higher latency and vice versa. Similarly, higher burst frequency leads to higher burst overheads and higher bandwidth utilization but achieves lower latency and vice versa.

In this paper, we consider four scheduling schemes for mobile fronthaul traffic distinguished by the parameters of the bandwidth assignment input and bandwidth allocation output as shown in Fig. 2 and described in the following subsections.

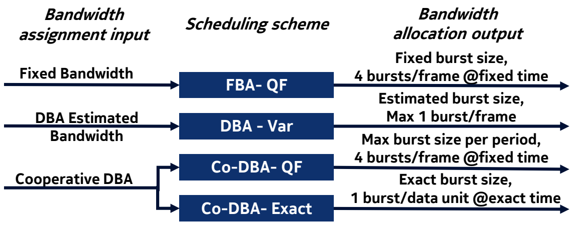

Fig. 2: Evaluated scheduling schemes for fronthaul over TDM-PON

2.1 FBA – Quarter Frame (FBA-QF)

This scheduling scheme uses the fixed bandwidth input calculated using the theoretical peak throughput of the specific RU configuration during a radio slot (based on the spectral bandwidth, subcarrier spacing and #antenna streams/ #spatial streams) and the details of fronthaul interface implementation. Each fronthaul traffic flow is allocated one PON burst of fixed duration every 31.25 µs (i.e. quarter of a PON frame) at a fixed time within each frame.

2.2 DBA – Varying (DBA – Var)

This scheduling scheme uses a periodic DBA process which estimates the bandwidth demands of fronthaul traffic over the next period based on traffic monitoring and status reporting [6] The traffic estimates are used as bandwidth assignment input and the PON bursts of certain duration are scheduled at varying periodicity (≥ 1 PON frame) is the bandwidth allocation output.

2.3 Co-DBA - Quarter Frame (Co-DBA-QF)

In this scheduling scheme, a Co-DBA process periodically receives an accurate bandwidth demand for fronthaul traffic over a future time period using the CTI between the DU and the OLT. The Co-DBA adapts the bandwidth assignments for fronthaul traffic flows at the time granularity as reported in the CTI message (≥ 1 radio slot). As an example, for a CTI period of 1 ms and radio slot of 0.5 ms, the Co-DBA will assign bandwidth for 2 radio slots. The bandwidth assigned within the reported time duration is allocated as one burst every 31.25 µs at a fixed time within each frame. The PON burst duration is assigned to carry the maximum fronthaul traffic burst during the CTI reported time period.

2.4 Co-DBA - Exact (Co-DBA-Exact)

This scheduling scheme uses the same bandwidth assignment mechanism as that of 2.3. The CTI message additionally has information about the fronthaul traffic pattern over the reported time period as a series of fronthaul traffic bursts and #bytes per burst. The bandwidth allocation over the CTI reported time duration is done by matching the PON burst allocation pattern (both size and timing) with the traffic pattern of the fronthaul traffic.

3. Evaluation methodology

To evaluate the bandwidth utilization and latency of the proposed scheduling schemes, we used a co-simulation of our 5G radio system level simulator and our TDM-PON transmission convergence (TC) layer simulator. The main co-simulation components are shown in Fig. 3.

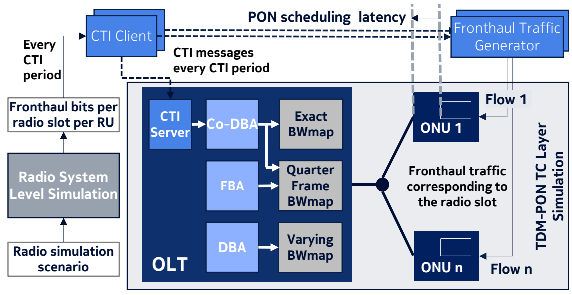

Fig. 3: Simulation setup

3.1 TDM-PON TC layer simulator

The TDM-PON TC layer simulator particularly focusses on scheduling in upstream direction. The simulator implements the bandwidth assignment process with per frame BWmap generation in the OLT and multiple ONUs with transmission container (TCONT) queues that follow the BWmaps. The TC layer model is implemented as per the 25GS-PON specification [7] and consists of downstream frames every 125 µs with BWmap structures and upstream frames containing ONU bursts for scheduled TCONTs.

3.2 Fronthaul traffic generator

The radio system level simulator uses a certain C-RAN deployment and user traffic profile as a simulation scenario input and provides the radio resource utilization of each RU at the radio slot time granularity. The CTI client obtains the radio resource utilization information every CTI period (≥ 1 radio slot) and translates it into #bits on the fronthaul interface over the reported time duration based on fronthaul interface implementation [8]. The CTI client then creates a CTI message with the bandwidth requirement and the traffic pattern per RU. The CTI message is sent to the CTI server in the OLT as well as to the fronthaul traffic generator. The CTI server provides the CTI message information to the Co-DBA process which calculates the BWmaps for the time duration specified in the CTI message. The traffic generator creates the corresponding fronthaul traffic bursts as a series of Ethernet packets.

Since a radio symbol is the smallest granularity of data unit at the fronthaul interface, the PON scheduling latency of the radio symbol is considered for evaluation. The cumulative bandwidth assigned in the upstream within a DBA period is considered for evaluating bandwidth efficiency of each scheduling scheme.

4. Simulation parameters and results

The radio simulation scenarios from [8] are used where each RU is considered as a radio cell operating in a frequency division duplex mode with 100 MHz bandwidth at 30 KHz subcarrier spacing and supports up to four antenna streams (or MIMO layers). This RU configuration results in the radio slot duration of 0.5 ms consisting of 14 radio symbols where the maximum packet size on the 7.2x split fronthaul interface is 30704 bytes per radio symbol, resulting in maximum fronthaul bandwidth of 6.89 Gbps. The selected user traffic scenario generates average user traffic request per RU corresponding to 30% (0.6 Gbps) of the theoretical air interface capacity of the RU.

The TDM-PON TC layer simulator is configured to operate at 25Gbps and includes XGEM framing and burst overheads. All ONUs have a user network interface at 25Gbps, are at 5 km distance from the OLT and hence the forward error correction is disabled.

The DBA period and the CTI messaging period is set to 1 ms for all scheduling schemes. The CTI message provides fronthaul traffic information over two consecutive radio slots of 0.5 ms each. Since all RUs operate in FDD mode, the fronthaul traffic pattern consists of a fronthaul traffic burst every radio symbol duration.

The RU configuration allows up to three RUs on a 25GS-PON for the *FBA-QF* scheme considering all PON overheads. Note that with the other scheduling schemes described in section 2, a few more RUs can be aggregated on the 25GS-PON. However, for a fair comparison of all schemes, three RUs were used. The three RUs are randomly selected from the radio simulation scenario and their radio resource utilization data over 10,000 radio slots is used to generate CTI messages and corresponding fronthaul traffic as described in section 3.2. The TDM-PON TC layer simulator performs the bandwidth assignment and bandwidth allocation per fronthaul traffic flow and reports the cumulative PON bandwidth utilized in each DBA period and the scheduling latency of every radio symbol. This process is repeated over 10 random combinations of three RUs and the descriptive statistics for bandwidth utilization and scheduling latency for the entire set are reported next.

Fig. 4 shows box-plots (i.e. min, max and interquartile ranges) of cumulative bandwidth utilized on the PON upstream (left Y axis) as well as scheduling latency per radio symbol (right Y axis) for the scheduling schemes described in section 2. The results in Fig. 4 show that the FBA-QF scheme results in the highest bandwidth utilization whereas DBA-Var and Co-DBA-QF utilize on average 9.5% and 21% higher bandwidth, respectively, w.r.t. Co-DBA-Exact. The higher bandwidth utilization of the DBA-Var is due to the bandwidth estimation errors whereas for the *Co-DBA-QF* it is due to the allocation of higher number of PON bursts than fronthaul traffic bursts (i.e. 32 PON bursts/ms for 28 fronthaul traffic bursts/ms) and allocating maximum fronthaul traffic burst duration over two radio slots reported in the CTI message. The *Co-DBA-Exact* results in lowest bandwidth utilization as it allocates the exact PON burst duration for the exact number of fronthaul traffic bursts.

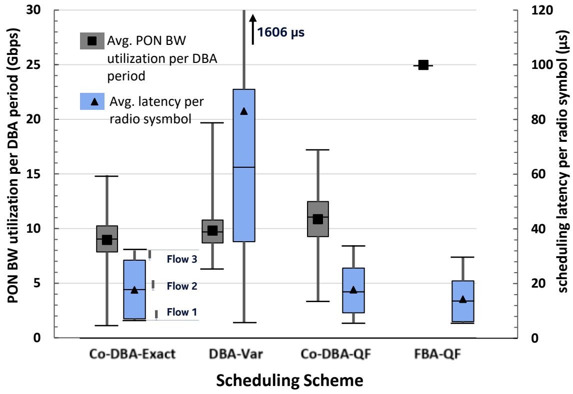

Fig. 4: Comparison of PON bandwidth utilization and scheduling latency per radio symbol

The PON scheduling latency per radio symbol is worst for *DBA-Var* as it has a higher average latency and a large latency jitter. This is due to the high burst periodicity of ≥ 1 PON frame coupled with the incorrect fronthaul bandwidth assignment - especially in case of a big low-to-high traffic transition over a short time duration. The *QF* allocation schemes perform similarly in terms of latency. The average and maximum latency of *Co-DBA-QF* is only 3.5 µs and 4 µs higher than the *FBA-QF*, respectively, due to the higher inter-burst gap of *Co-DBA-QF*. The latency and jitter in both *QF* allocation schemes is less than 31.25 µs as this is their maximum inter-burst gap. The *Co-DBA-Exact* scheme achieves low-jitter of < 4 µs for individual fronthaul traffic flows as shown in the inset of Fig. 4. However, it can align the PON bursts exactly in time with one fronthaul flow traffic bursts only (i.e. flow 1). The remaining fronthaul traffic flows have a fixed delay equal to the sum of max. symbol durations of the previously scheduled flows, to minimize the latency jitter.

6. Conclusions

We propose and compare bandwidth efficiency and latency performance of two scheduling schemes for Co-DBA enabled TDM-PON for C-RAN traffic w.r.t. FBA and conventional DBA. Although the *Co-DBA-Exact* provides optimal performance, the scheduling complexity of burst alignment for multiple fronthaul traffic flows is challenging. The *Co-DBA-QF* provides acceptable performance with a scheduling scheme suitable for a practical implementation.

Acknowledgements

This work has been supported by the German BMBF funded project KIGLIS (16KIS1227K) as well as by the German BMBF within the EU-Celtic project AI-NET ANTILLAS (FKZ 16KIS1305).

References

[1] O-RAN-WG4.CUS.0-v08.00 “Control, User and Synchronization Plane Specification”, O-RAN Alliance (March 2022)

[2] S. Bidkar, J. Galaro, Th. Pfeiffer, "First Demonstration of an Ultra-Low-Latency Fronthaul Transport Over a Commercial TDM-PON Platform", OFC 2018, paper Tu2K.3

[3] T. Tashiro *et al.*, "A novel DBA scheme for TDM-PON based mobile fronthaul," *OFC 2014*, San Francisco, CA, 2014, Tu3F.3.

[4] G.Sup71, “Optical line termination capabilities for supporting cooperative dynamic bandwidth assignment”, ITU-T (04/2021)

[5] ORAN-WG4.CTI-TCP.0-v02.00 “Cooperative transport interface, transport control plane specification”, O-RAN Alliance (2021)

[6] ITU-T Recommendation G.9807.1 Amd2 “10-Gigabit-capable symmetric passive optical network (XGS-PON)”, ITU-T (10/2020)

[7] Multi-source Agreement “25GS-PON Specification v1.0” (10/2020), https://www.25gspon-msa.org/wp-content/uploads/2020/10/25GS-PON-Specification-V1.0-public.pdf

[8] S. Bidkar, P. Dom, R. Bonk, Th. Pfeiffer, "Mobile Xhaul Traffic Modelling for High-Speed TDM-PON", ECOC 2020, Brussels (virtual), paper We2J-5