A Low-Latency Traffic Estimation Based TDM-PON Mobile Front-Haul for Small Cell Cloud-RAN Employing Feed-Forward Artificial Neural Network

Ahmed Mohammed Mikaeil, Weisheng Hu, and Syed Baqar Hussain
State Key Laboratory of Advance Optical Communication System and Network
Shanghai Jiao Tong University, Shanghai 200240, China
Email: ahmed_mikaeil@yahoo.co.uk, wshu@sjtu.edu.cn

ABSTRACT
This introduces a novel method for low-latency time-division multiplexing passive optical networks (TDM-PONs) mobile front-haul for small cell cloud radio access network (C-RAN) based on traffic estimation. In this method, the amount of packets arrive to the optical network unit (ONU) buffer from the remote radio unit (RRU) link is predicted using a feedforward artificial neural network function integrated into the dynamic bandwidth allocation (DBA) module at the optical line terminal (OLT). This predictive method eliminates the latency required for reporting ONU buffer occupancy to the OLT. As the result, the delay is minimized as the required for mobile front-haul in C-RAN architecture. The performance of this new method is evaluated by means of simulation under XGS-PON standard. The simulation results confirmed the validity of the proposed method and proved its superiority to other XGS-PON standard complaint algorithms proposed for the same purpose.
Keywords: Artificial Neural Network, Bandwidth allocation, Cloud-RAN, Mobile Front-haul, Traffic Estimation, XGS-PON.

1. INTRODUCTION

TDM-PONs are a promising technology for realizing a cost-efficient mobile front-haul for small cell C-RAN as they allow the sharing of optical fibers and transmission equipment. However, the latency in TDM-PONs upstream transmission due to the DBA mechanism is much higher than the latency required for mobile front-haul in C-RAN architecture (i.e. 250~300µs in [1,2]). There is considerable research on different low-latency DBA methods to support mobile front-haul over TDM-PONs system for example Ref [3,4]. Also, in Ref [5] we able to achieve 300µs low-latency XG-PON based mobile front-haul in MAC-PHY split based C-RAN. In continuation of this work, we consider the architecture depicted in Fig. 1, a PHY layer split based small cell C-RAN [4]. In this architecture, the instantaneous actual wireless system load in form of antenna symbol is mapped into bit-stream then encapsulated into Ethernet packets and transmitted between BBU and RRU via XGS-PON front-haul. For efficient utilization of XGS-PON mobile front-haul and avoiding the collisions between the packets transmitting from different RRU at the same time a DBA mechanism is used at OLT side to govern the assignment of the transmission opportunity (i.e. the grant allocation) for each ONU (i.e. RRU). The DBA engine at OLT assigns the grant to the ONU based on its buffer occupancy reports or ONUs transmission containers also known as T-CONTs (Note: in this paper we assume that the ONU has only one T-CONT as in [6].

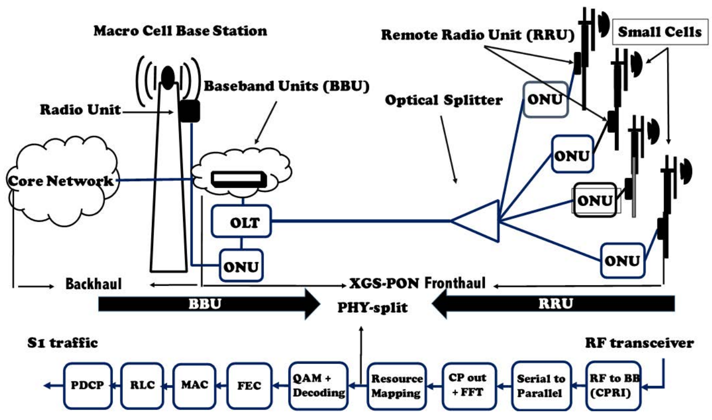


Figure 1. PHY- Split based small cell C-RAN.

In [5] we have shown that giving grants allocation to the ONUs based on their buffer occupancy reports, while adjusting the congested ONUs predetermined limit can effectively accommodate the burst-ness of wireless traffic at RRUs side and attains a considerable delay performance improvement. As an upgrade to this method, in this paper we propose to predict the amount of ONU's buffer occupancy in advance based on the historical ONUs reports collected during BBU processing time. Based on these predicted ONU reports the DBA engine at the OLT prepares in advance the grants for each ONU. We formulate the problem of predicting the future amount of ONU buffer occupancy (i.e. next ONU report) as machine learning function approximation problem. Then, we utilize the feedforward neural network (FANN) to solve this approximation problem. To the best of our knowledge, no prior studies have applied FANN approach in the context of mobile front-haul.

2. REPORT ESTIMATION PROBLEM FORMULATION

In frequency division duplex (FDD) based LTE, the downlink and uplink subs-frames are time-aligned 4 ms. Within this 4 ms, BBU requires around 2.3 ~ 2.6 ms to finish preparing the downlink and sending the acknowledged (ACK) or non-acknowledged (NACK) to the UE, in a way that when it receives the uplink data from UE in sub-frame n it will send an ACK/NACK to the UE in sub-frame n+4 [7]. Our idea in this paper is to take advantage of the 2.3 ms BBU processing time in collecting data statistics (reports) from ONUs. Then utilize these reports as a training examples (samples) to train a prediction model to predict the next reports for the ONUs. After that these predicted reports are used to generate the scheduling grants that map the upcoming XGS-PON front-haul frames. The number of ONU reports that can be collected within one BBU processing time cycle is given by: T = BBU processing time / (one-way propagation delay to transmit one frame from ONU to OLT) + (125 µs: the time required for transmitting the previous frame). For example, for 10 km front-haul the number of the report will be equal to 2.3 ms / (50+125) µs = 13 ONU reports.

Assuming that we collect until t report for j-th ONU in one BBU processing time cycle, if we extend backward from t to the beginning of cycle then we will have a time series contains the following reports $\{R_t^j, R_{t-1}^j, R_{t-2}^j \ldots\ldots R_0^j\}$ where $j = 1,2,3 \ldots K$ is ONU index. From this time series, if we want to estimate R_t^j at some future time $t + s$, we can write this problem as follows:

$$\hat{R}_{t+s}^j = f(R_t^j, R_{t-1}^j, R_{t-2}^j \ldots\ldots R_0^j) \tag{1}$$

where s is the prediction horizon. Assuming s = 1, we can predict one time sample into the future $\hat{R}_{t+1}^j$, if we can find an optimal estimator that can accurately represents the function f .Finding such an estimator is a machine learning function approximation problem, which can be solved using machine learning time series prediction models such as neural network . In the next section we introduce adaptive deep learning approach to solve this problem.

3. REPORT ESTIMATION PROBLEM SOLUTION BASED ON FEEDFORWARD ARTIFICIAL NEURAL NETWORK

Artificial neural networks (ANN) are constructed using computational functions known as neurons. In Feedforward based artificial neural network (FANN) neurons are arranged in layers with non-cycled connections between the layers [8]. The first layer neurons and the last layer neurons in FANN represent the input and output variables of the function that we want to approximate. A number of hidden layers exist between the first layer and the last layer neurons (i.e. input and output layer) with weighted connections that determine how well the FANN performs. The output of the neuron in the hidden and last layer are often fed into an activation function (e.g. sigmoid function, linear and etc..) to determines the output value of the neuron. FANN employs function approximation by learning from the input and output variables or training examples that describe how the function works. Then it adjusts the network architecture and the internal weights to produce the same output as in the training examples. So that when it given a new examples or input variables, it will produce a correct output for them.

In order to solve the report estimation problem we mention in the previous section we first train FANN model to learn the function $f(R_t^j, R_{t-1}^j, R_{t-2}^j \ldots\ldots R_0^j)$.Then, we run the model to predict $\hat{R}_{t+1}^j$ based on the approximated function. And finally we use the predicted value of the report $\hat{R}_{t+1}^j$ to prepare in advance the grant allocation for the ONU.

3.1 FANN Training Phase

The training phase of FANN model is done by a function that take a set contains ONU reports collected during BBU processing time cycle as an input and return a trained FANN model for this ONU. We assume that during the first BBU processing time cycle the OLT gives fixed grants to each ONU in the network (i.e. $G_0^j = BW/K$, where BW is the total XGS-PON frame size in byte and K is the number of ONU. This to allow the OLT to collect the initial set of ONU reports. After collecting these reports in for every ONU in an individual training set $\{R_t^j, R_{t-1}^j, R_{t-2}^j \ldots\ldots R_0^j\}$ $j = 1, 2, 3 \ldots K$, we run FANN training algorithm on each ONU training set as follows

For every point in ONU training set, train the FANN model with R_t^j as inputs and what followed R_t^j as output until the model reaches the desired error we want (see also Fig. 2a). Then we save the FANN model for every ONU so as to be used later in the prediction phase.

3.2 FANN Prediction Phase

Upon arrival the of the last ONU report in the ONU training set (i.e. the newest report within the BBU processing time cycle). The first step in the pectin phase is to update the training set for the ONU by adding the new report to the training set and eliminating the earliest report from it. Then the saved FANN model for the ONU (i.e. from raining phase) is utilized to predict the upcoming report for that specific ONU (Fig. 2b). After predicting all ONU reports, the obtained estimated report set is provided as an input to the grants assignment phase algorithm in order to assign the final grant for every ONU.

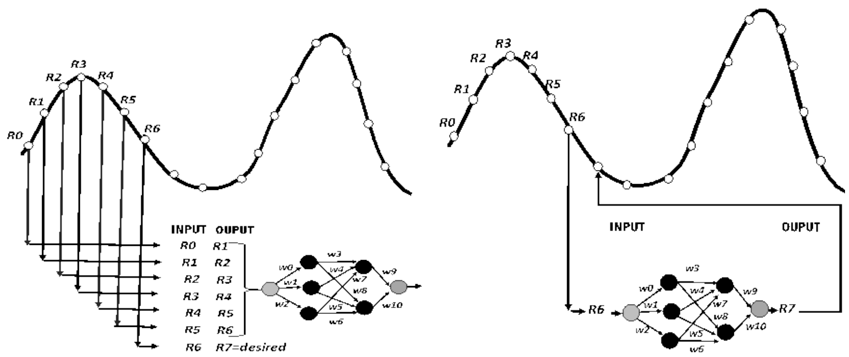


Figure 2: (a) FANN network model training phase; (b) FANN network model prediction phase.

3.3 Grants Assignment Phase

After receiving the set contains the predicted ONU reports from FANN prediction phase, we use these to assign the grants for the ONUs. For efficient utilization of XGS-PON upstream frames during upstream allocation cycles we utilize these predicted ONU reports as in input to our previous algorithm which known as Optimized-RR DBA [5] to assign the grants for every ONU.

4. SIMULATION RESULTS

In this section we evaluate the performance of our proposed method in terms of delay (the end to end time required for the packet to travel from ONU buffer ingress to the OLT buffer egress), packet drop ratio and upstream link utilization against three other XGS-PON compliant DBA: 1) Round Robin DBA (RR-DBA) [9], 2) Group Assured Giga-PON Access Network DBA (gGIANT) [6], 3) Optimized-RR DBA [5].

In order to evaluate the performance our algorithm which we refer as "Deep-Learning DBA" against the above-mentioned algorithms we conduct simulation experiments using XG-PON module of the network simulator NS-3. In our experiments, we consider a front-haul network with 8 LTE small cell RRUs (i.e. 20MHz bandwidth channel single carrier and single user per transmission time interval) connected to 8 ONUs, each ONU has a buffer size equal 1 Mbyte. We set 120 μs as a round trip propagation delay to represent 10 km distance front-haul (note: the round trip propagation delay of 10 km is 100μs, the 20μs is additional delay margin to tolerate the latency results from ONUs and other front-haul elements processing). We generate the mobile front-haul uplink traffic by injecting each ONU with Poisson Pareto Burst Process (PPBP) traffic with Hurst parameter 0.8 and shape parameter 1.4 for a period of 10 seconds as in [5]. To assure that the generated traffic is similar to the practical LTE network traffic we allow 10% of RRUs to generate 60% of the total aggregated front-haul uplink load. For the neural network module, we used fast artificial neural network library [10] and the value of the parameters used to train fast artificial neural network during the simulation are summarized in Table 1.

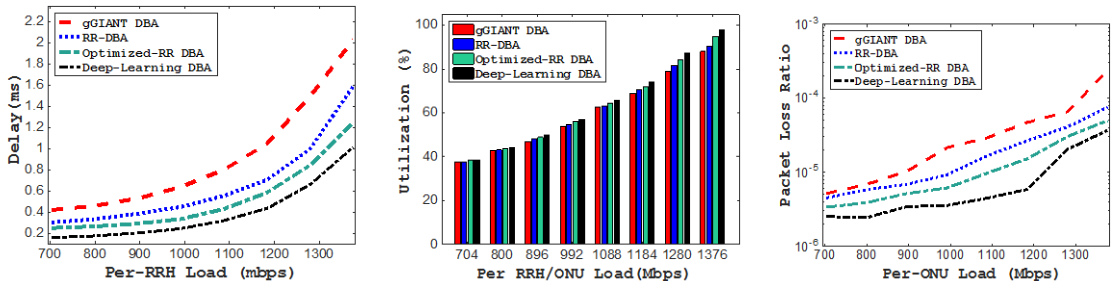
TABLE 1

Parameter	Details
Total number of layers	4
Number Inputs layers	1
Number Outputs layers	1
Neurons in first hidden layer	3
Neurons in second hidden layer	2
Desired error	10^{-7}
Activation function	Linear
Initial weights	Random[-0.1~ 0.1]
Learning rate	0.7
Max number of epochs	10000

Figure 3a we can see that Optimized-RR outperforms the three others algorithms in term of upstream delay performance. From the same figure we can notice that Optimized-RR DBA attains a delay performance in the range of 290 to 300μs when the per-ONU/RRU load is ranging from 903 to 922 Mbps (i.e. the required front-haul uplink capacity when considering PHY split based small cell [11]), while gGIANT and RR-DBA show a delay performance higher 300μs for the same range of traffic load. When it comes to the deep learning based DBA, we can see that is attains a delay performance in the range of 200 to 205 μs for the same range traffic load range. This delay performance is much better than the other DBA algorithms and lower than the latency required for mobile front-haul stated by the next generation mobile networks alliance [1] (250μs). The key idea behind the low latency performance achieved by Deep-Learning DBA is because of eliminating one-way propagation delay for receiving the buffer occupancy report from ONU s as well as the waiting time for DBA processing at OLT.

Figure 3b shows the utilization performance comparison between the four algorithms. From this figure we can see that the highest utilization performance is achieved by Deep-Learning DBA. This is because of two reasons the first one is because of faster grants assignment process by Deep-Learning DBA which allow ONUs to effectively utilize XGS-PON upstream bandwidth comparing to the three other DBAs. The second reason is because Deep-Learning DBA also inherits the property of dynamical adjusting of ONU pre-determined limit same as Optimized-RR DBA [5]. This allows ONUs to efficiently exploit any unallocated reminder of the upstream frames improving the overall XGS-PON upstream link utilization.

Figure 3c shows the packet loss ratio performance of the fours algorithms. As we can see that the deep learning DBA performs better than the three others algorithm Optimized-RR and attains the lowest packet loss rate. The reason behind this is because the deep learning based DBA increases the speed of serving ONU. In other words, it clears ONU buffer from data faster than the other three algorithms resulting in more reduction in network congestion and waiting time for the packet to get served.



(a) (b) (c)

Figure 3. The performance comparison between gGIANT, RR-DBA, Optimized-RR and Deep learning DBA: (a) The delay performance, (b) The utilization performance, (c) The packet loss ratio performance.

5. CONCLUSIONS

In this paper we proposed a novel method for predicting the future amount of ONU buffer occupancy based on the historical set of ONU reports collected during BBU processing time. Based on predicted ONU report by the aforementioned method, the OLT allocates the scheduling grant in advance to the ONU eliminating the delay required for reporting the ONU buffer occupancy to the OLT and the additional DBA processing delay. The simulation results for evaluating our new method against three others XGS-PON compliant DBAs have shown a significant improvement in the upstream delay, packet loss and utilization performance.

ACKNOWLEDGEMENTS

This work was supported by China NSFC (61431009, 61371082, and 61521062) and the National Science and Technology Major Project of the Ministry of Science and Technology of China (2015ZX03001021).

REFERENCES

[1] Alliance NGMN. Further study on critical C-RAN technologies. Next Generation Mobile Networks. 2015.
[2] N. Yoshimoto: Operator perspective on next-generation optical access for future radio access, in *Proc. 2014 IEEE International Conference on Communications Workshops (ICC)*, pp. 376-381.
[3] H. Nomura *et al.*: First demonstration of optical-mobile cooperation interface for mobile front-haul with TDM-PON, *J. IEICE Communications Express*, vol. 6, pp. 375-80, 2017.
[4] S. Zhou *et al.*: Low-latency high-efficiency mobile front-haul with TDM-PON (Mobile-PON), *Journal of Optical Communications and Networking*, vol. 1, pp. A20-A26, Jan. 2018.
[5] A. M. Mikaeil *et al.*: Performance evaluation of XG-PON based mobile front-haul transport in cloud-RAN architecture, *Journal of Optical Communications and Networking*, vol. 9, pp. 984-994, Nov. 2017.
[6] P. Alvarez *et al.*: Backhauling mobile systems with XG-PON using grouped assured bandwidth, in *Proc. 2014 19th European Conference on Networks and Optical Communications (NOC)*, pp. 91-96.
[7] N. Nikaein: Processing radio access network functions in the cloud: Critical issues and modeling, in *Proc. 6th Int. Workshop on Mobile Cloud Computing and Services*, pp. 36-43. ACM, 2015.
[8] A K. Jain *et al.*: Artificial neural networks: A tutorial, *Computer* 29, pp. 31-44, Mar. 1996.
[9] J. A . Arokkiam *et al.*: Experimental evaluation of TCP performance over 10Gb/s passive optical networks (XG-PON), in *Proc. IEEE GLOBECOM*, pp. 2223-2228, 2014.
[10] Fast artificial neural network library, "Library" [Online]. Available: http://leenissen.dk/fann/.
[11] Virtualization, small cell, functional splits and use cases, in *Proc. Small Cell Forum Release*, vol. 6, 2016.