The Effect of Different Queuing Regimes on a Switched Ethernet Fronthaul

Mohamad Kenan Al-Hares[1], Philippos Assimakopoulos[1], Simon Hill[2] and Nathan J. Gomes[1]
[1]*Communications Research Group, University of Kent, Canterbury, UK*
[2]*Viavi Solutions UK Ltd, Newbury, Berkshire, UK*

ABSTRACT
This paper investigates the effects of different queuing regimes on the mean and standard deviation of the frame inter-arrival delay of a LTE traffic stream under the presence of background Ethernet traffic. The background traffic is used to represent traffic that would be generated by different functional subdivisions in the physical layer of traditional LTE base station. In this work, a Switched Ethernet architecture is used as the fronthaul section of a Cloud Radio Access Network (C-RAN). Contention in this fronthaul becomes an important issue since different traffic streams originating from different functional subdivisions with different quality of service specifications will be transmitted over the same physical links. Trunk ports then, will have to handle the queuing management and prioritization. Handling the traffic with different queuing regimes will reflect on the latency and latency variations of the LTE traffic.
Keywords: Fronthaul, C-RAN, LTE, Ethernet, Background traffic.

1. INTRODUCTION

The fronthaul of the centralised access network proposed for the next mobile generation must comprise a transport infrastructure for different types of traffic between base station baseband units (BBU) and remote radio heads (RRH) [1]. As a means of reducing the data rate requirements of the future fronthaul, the implementation of different physical layer functional subdivisions (or "splits") has been proposed [1, 2] Furthermore, a combination of functional splitting with the use of Ethernet technology in the fronthaul can lead to cost reductions (by leveraging the ubiquity of Ethernet equipment) and performance enhancements (mainly from the ability to obtain statistical multiplexing gains [3]), in addition to the reduction in data rate requirements. At Ethernet switch ports, queuing and scheduling algorithms, such as weighted round robin (WRR) and strict priority (SP) can be employed to prioritize and manage the traffic and different transported services [4]. It is, then, important to investigate the efficiency of different queuing regimes and measure the effect of each on the radio traffic, when this co-exists with other streams in the fronthaul. In this work, Long-Term Evolution (LTE) sampled radio waveform traffic is generated and transported over a switched Ethernet fronthaul together with background traffic, which emulates traffic that would be generated by different functional splits in the access network. Specifically we focus on the effect of using the WRR with different weights and SP queuing regimes on the frame inter-arrival delay of the LTE traffic, when it contends with background traffic.

2. STRICT PRIORITY AND WEIGHTED ROUND ROBIN QUEUING REGIMES

2.1 Strict Priority (SP)

With SP, the queue with highest priority transmits first and then the other queues with lesser priority will transmit one by one in order of their priority setting. Thus, a lower priority queue has to wait for all higher priority queues to finish transmission in order to have an opportunity to starts its own transmission. SP is used with very delay/jitter sensitive traffic (e.g. voice traffic in the mobile network) but can lead to starvation of lower priority queues [5].

2.2 Weighted Round Robin (WRR)

In this work we discuss a packet-count based WRR algorithm, where each queue is allowed to transmit a specific number of frames in every transmission round. This queuing regime is fairer than the SP regime. Each queue is assigned a weight which corresponds to the number of frames that it can transmit before it must cede transmission to other queues [6]. Increasing the number of transmitted streams will reduce the share of each stream in the link since each stream will be allocated a smaller percentage in the link resources. Note that SP can be considered as a special case of WRR with a weight equal to zero for the lower priority stream(s).

3. MEASUREMENT SET-UP

Fig. 1 shows the testbed used for the measurement procedure. A workstation runs an emulated LTE base station (Amari LTE-100) that produces I/Q samples corresponding to a 5 MHz channel bandwidth (sampling rate of 6.25 MHz). The samples are then inserted into the payload section of a UDP/IP packet and transmitted over a layer 2 Ethernet network. The transmission is bursty, as the LTE source will buffer a number of OFDM symbols before transmitting the corresponding I/Q samples over the Ethernet link. The network comprises of two 3COM-5500G

Ethernet switches, operating in store-and-forward mode with standard 1000BASE-LX small form-factor pluggable (SFP) transceivers with LC connectors and single mode fiber (SMF) patchcords. The packet stream containing the I/Q samples is received by an Ettus N210 remote radio head (RRH) where it is digitally processed prior to up-conversion and transmission over the wireless channel. A Viavi hardware-based traffic generator is used to generate bursty background traffic of different payload sizes (500 Bytes, 1500 Bytes, and 4000 Bytes) and at variable data rates (45Mbps, 105 Mbps, 215 Mbps, and 450 Mbps). The two streams of traffic are assigned to different Virtual Local Area Network (VLAN) IDs and transmitted to the destination through a trunk link between the switches. A Viavi in-line Ethernet probe ("smart" probe) system is used to capture the Ethernet frame header and time stamp each packet header. These time stamps are sent back to a Packet Routing Engine (PRE) where they are routed to a management station for further processing. For a more detailed description of the probing system set-up, see [7]. The time stamps are used to estimate the frame inter-arrival delay of the LTE traffic when this contends with the background traffic over the trunk port.

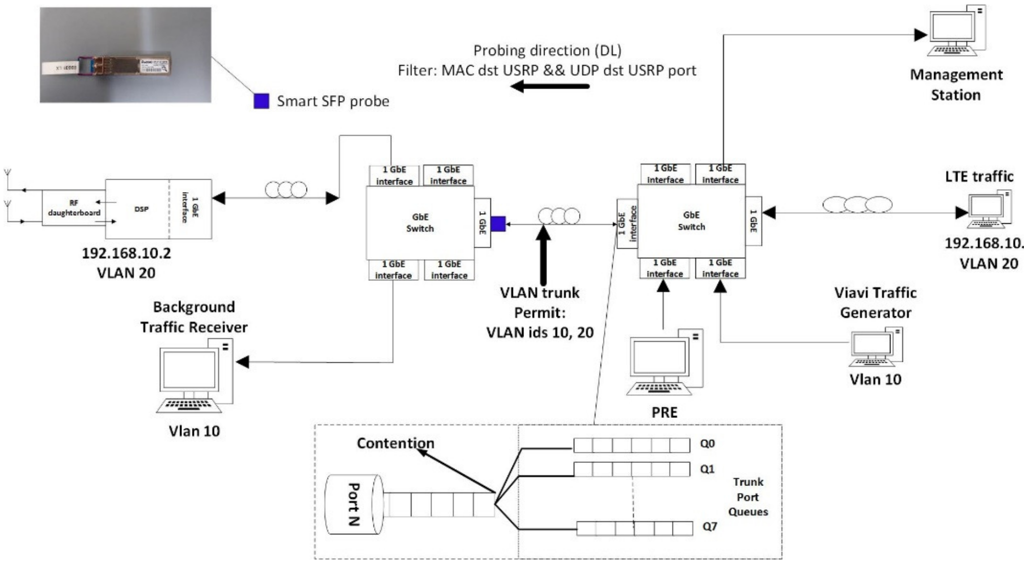

Fig.1. Testbed used for the measurement procedure. PRE=Packet Routing Engine, GbE=Gigabit Ethernet, SFP=small form-factor pluggable.

4. MEASUREMENT RESULTS

Here, the results from different queuing regimes, with a range of priority and weight combinations, and different background traffic rates (45, 105, 215 and 450Mbps) and frame sizes (500, 1500 and 4000 Bytes) are presented. In the results that will be presented, different packet-count based WRR weight combinations are used: (WRR1: LTE 8, Background 1), (WRR2: LTE 8, Background 4), (WRR3: LTE 8, Background 6) and (WRR4: LTE 8, Background 8). Standard deviation (std) values are superimposed on the mean values in the form of error bars. The base line case which corresponds to transmitting just the LTE traffic (i.e. no background traffic) is indicated in all the following plots.

4.1 WRR Queuing Regime Results

First, the frame inter-arrival statistics for the LTE traffic in the switched Ethernet fronthaul when using a WRR regime in the switched Ethernet fronthaul is examined. The results in Fig. 2A show that increasing the weight of the background traffic causes the mean delay of the LTE stream to increase by approximately 2.6% on average for each weight increase and the std increases by approximately 13.7% on average. The increase in both the mean and std for higher data rates is a result of the corresponding increase in the frame transmission rate of the background traffic which will cause the background traffic queue to be filled more often (note that the traffic source is bursty). The increase among the different weight combinations for each data rate is simply a result of allocating more resources to the background traffic. Figure 2.B shows a comparison between different frame sizes with background traffic data rate 215Mbps. The results show that using larger frame sizes in the background traffic will lead to an increase in the mean and std of the inter-arrival delay (compared to using smaller frame sizes) for all WRR weight combinations and all background traffic data rates (although not shown in this Figure). This can be explained since larger frames require a longer time to be serialized out of the trunk port and as a result occupy the channel for longer time.

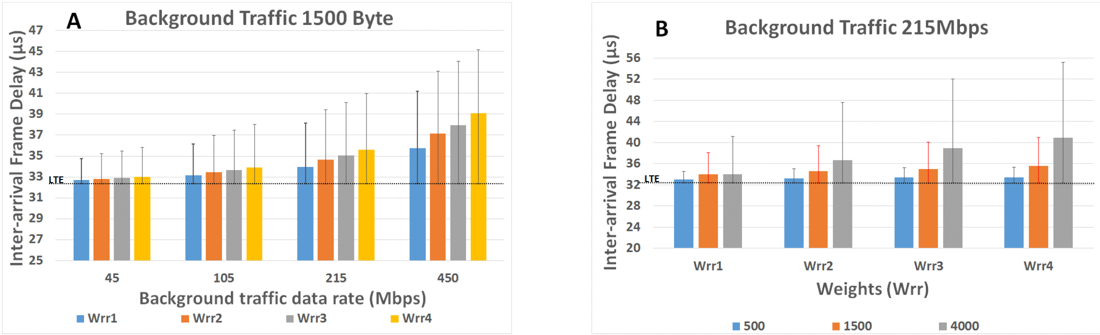

Fig.2. Comparison for mean and std of frame inter-arrival delays of the LTE traffic for different WRR weights for A) different background traffic data rates for a frame size of 1500 bytes and B) different background traffic frame sizes at a data rate 215 Mbps. The "LTE" trace corresponds to transmitting only the LTE traffic (i.e. no background traffic).

4.2 Strict Priority Queuing Regime Results

Fig. 3A is a plot of the mean and std using SP for different frame sizes and two different data rates for each frame size, while Fig. 3B shows the corresponding complementary cumulative distribution functions (CCDFs). The results show that with SP, using larger frame sizes will cause a small reduction in the mean frame inter-arrival delay but will lead to a higher std. This bahavior can be explained using Fig. 3B as follows: when a background traffic frame is being serialized out of the switch port while a new LTE frame arrives in the queue (which until that point was unoccupied), the time that the LTE frame will have to wait, until the serialisation of the other frame is complete, will be higher for larger background frame sizes (and bounded by one background frame serialisation delay), resulting in an increase in the std. The mean value on the other hand reduces, as with larger background frame sizes the occurrence of such an event is less likely (as the packet transmission rate is reduced). These results clearly show the effects of lack of pre-emption (that is the interruption of a lower priority frame by higher priority traffic) in this set-up.

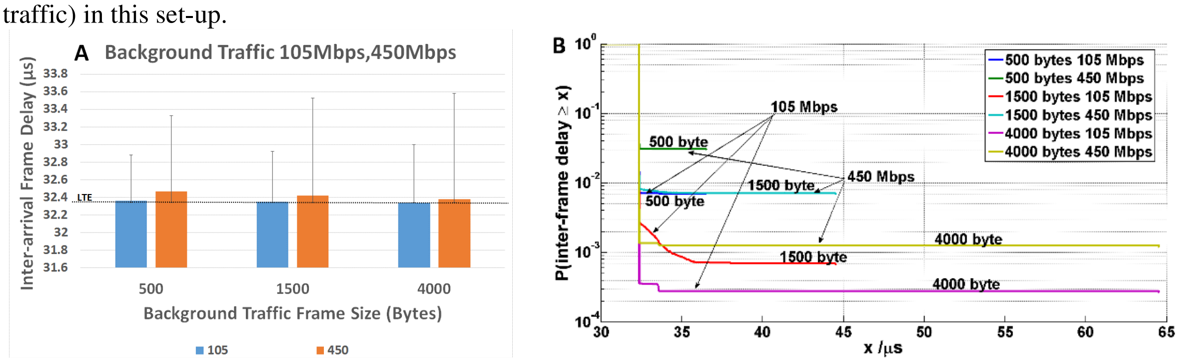

Fig.3. Mean and std of frame inter-arrival delays of the LTE traffic under SP regime for A) different Frame sizes and background traffic data rates (105 Mbps, 450Mbps). B) CCDFs of the results in A).

4.3 Comparison with Single-Queue Case

The difference between using no-priority (i.e. a single queue) and WRR with equal-weights (i.e. two queues with equal priority) can also be investigated. These two special cases are important for two reasons: With many different types of traffic streams potentially being transported through the fronthaul, there may be cases where two streams have equal (or approximately equal) weight definitions. Additionally, there is only a limited number of priority definitions at layer 2 which means that different streams may need to be accommodated by the same queue. The results in Fig. 4A show that using no priority in the network will cause higher mean delays than using two equal-weight queues for smaller frame sizes (not jumbo frame regimes) for both measured LTE traffic rates. On the other hand, the no-priority case will result in smaller mean delay than the equal-weights WRR case when using jumbo frames in the background traffic, for both LTE traffic rates (see Fig. 4B). Note that for the no-priority case the delay does not change considerably between the two frame sizes in Figures 5A and 5B. This is expected since a smaller frame size simply means that the traffic source will be transmitting a larger number of frames (in this case eight 500 byte frames instead of a single 4000 byte frame) over the same time interval. The delay of the equal-weight case in Fig. 4A remains very low in value even at the higher data rates. Note that the delay here is bounded by the sum of the serialisation delays of the two traffic sources (approx. 36.4 µs) but the mean delay will also depend on how frequently frames from the two sources interact in the port (i.e. how often both queues are filled due to the bursty nature of the sources). Obviously, there is a clear increase in the delay for higher frame sizes for the equal weight case as now the serialisation time of each background frame will be higher (by a factor of eight in this case).

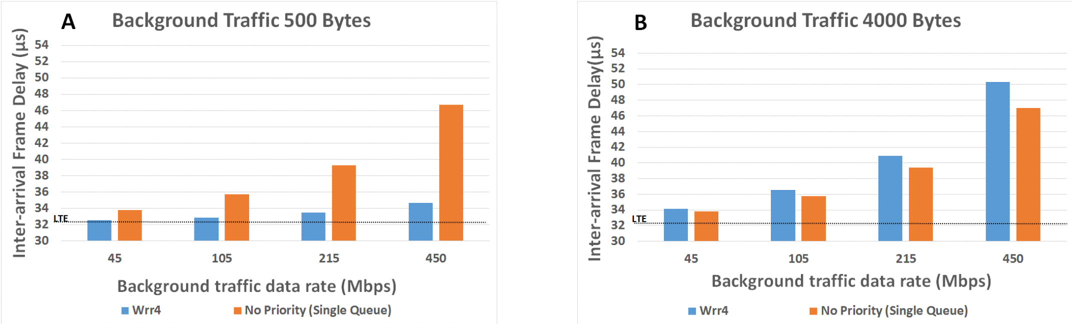

Fig.4. Comparison for the mean of the frame inter-arrival delay of the LTE traffic for equal-weight (Wrr4) and No-priority cases with different LTE traffic bandwidths and different background traffic data rates with frame sizes A) 500 Bytes and B) 4000 Bytes. The "LTE" trace corresponds to transmitting only the LTE traffic (i.e. no background traffic).

5. CONCLUSIONS

Different queuing regimes in the switched fronthaul network of future cloud radio access network (C-RAN) are presented and examined. The results show the importance of using a suitable queuing regime in the fronthaul network based on the requirements of each individual stream. SP can be used with sensitive delay/jitter services where the mean and std of frame inter-arrival delay is slightly increased. However this regime does not guarantee that the time-sensitive traffic will not encounter higher delays (due to lack of a pre-emptive mechanism). The WRR regime provides more capability to balance and distribute the available trunk capacity between different streams in the fronthaul network. The results show that the background traffic rate and Ethernet frame size effect the mean and std values of frame inter-arrival delay. Using bigger frame sizes (jumbo frame regimes) will increase the mean and std of the inter-arrival delay. Two interesting special cases, which are the equal-weight WRR and the single-queue (i.e. no-priority) regime are also compared. The equal-weight regime results is smaller mean frame inter-arrival delays than the no-priority case but only when using small frame sizes. When using Jumbo frames the opposite behaviour is observed. The main factors for choosing a queuing regime in the mobile fronthaul network is the requirements of each stream to meet given end-to-end latency and latency variation requirements.

Acknowledgment

This work was carried out within the framework of the European Union's Horizon 2020 research and innovation programme under grant agreement No 644526 (iCIRRUS project) and EPSRCs "Towards an Intelligent Information Infrastructure (TI3)" programme (NIRVANA project). Mohamad Kenan Al-Hares acknowledges the funding through an EPSRC Doctoral Training Partnership (DTP) and Philippos Assimakopoulos by the NIRVANA project. Data used in this work is stored in Kent Academic Repository (https://kar.kent.ac.uk/).

REFERENCES

[1] NGMN (MAR. 2015), *A deliverable by the NGMN alliance :Further study on critical C-RAN technologies* [Online]. Available: https://www.ngmn.org/publications/technical.html

[2] iCIRRUS (July 2015), *iCIRRUS intelligent C-RAN architecture* [Online]. Available: http://www.icirrus-5gnet.eu/wp-content/uploads/2016/02/D2.1-Intelligent-C-RAN-architecture.pdf

[3] iCIRRUS (Jan. 2016), "D3.1: Verification of Ethernet as transport protocol for fronthaul / midhaul," [Online]. Available: http://www.icirrus-5gnet.eu/category/deliverables/

[4] Cisco.(30 May 2006), *QoS Scheduling and Queueing on Catalyst 3550 Switches* [Online]. Available: http://www.cisco.com/c/en/us/support/docs/lan-switching/lan-quality-of-service/24057-187.html .

[5] J. Segarra, V. Sales and J. Prat, "Queue Management and Priority Scheduling Disciplines for QoS Control in Wavelength Routed OBS (WROBS) Access Networks," in *IEEE Int. Conf. on Transp .Optic.Net*., Nottingham,United Kingdom (ICTON) , pp. 207-214, 2006.

[6] I. Saidu , S. Subramaniam and A. Jaafar. "A load-aware weighted round-robin algorithm for IEEE 802.16 networks," *EURASIP J. on Wireless Commun. and Networking,* vol. 13, no. 18, pp. 3,4,Sep 2014.

[7] P. Assimakopoulos, M. K. Al-Hares, S. Hill, A. Abu-Amara, and N. J. Gomes, "Statistical distribution of packet inter-arrival rates in an Ethernet fronthaul," in *IEEE Int. Conf. on Commun. Workshops (ICC)*, Kuala Lumpur, Malaysia, 2016.