Envisioning Spectrum Management in Virtualised C-RAN

Al-Samman, Imad; Artuso, Matteo; Christiansen, Henrik Lehrmann; Doufexi, Angela; Beach, Mark

Published in:
2017 IEEE Wireless Communications and Networking Conference (WCNC)

Link to article, DOI:
10.1109/WCNC.2017.7925555

Publication date:
2017

Document Version
Peer reviewed version

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
You may not further distribute the material or use it for any profit-making activity or commercial gain
You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Envisioning Spectrum Management In Virtualised C-RAN

Imad Al-Samman*, Matteo Artuso†, Henrik Christiansen†, Angela Doufexi* and Mark Beach*

*Communication Systems & Networks Group
University of Bristol
Bristol, United Kingdom
Email: {Imad.Al-samman, M.A.Beach, A.Doufexi }@bristol.ac.uk

†DTU Fotonik, Department of Photonics Engineering
Technical University of Denmark
2800 Kongens Lyngby, Denmark
Email: { matart, hlch }@fotonik.dtu.dk

Abstract—Cloud Radio Access Network (C-RAN) has attracted a worldwide attention in both academia and industry. This network architecture re-forming has been considered as a potential solution to meet the increasing capacity demands for future mobile data traffic. In addition, Network Virtualisation is a promising technique for efficient resource utilisation. This paper proposes a customisable resource virtualisation algorithm for multi user data scheduling in a Long Term Evolution (LTE) C-RAN deployment. The algorithm is based on the hypervisor specific assignment of air resources allocation between the virtual operators (VOs) dynamically, based on either joint scheduling or per cell schemes. The objective is to improve the resource allocation mechanism based on traffic conditions and a database of pre-defined services priorities. Two distinctive scenarios are considered and evaluated against standard Round Robin (RR) C-RAN scheduling technique. Simulation results show improvements in the overall traffic throughput and reduction in end-to-end delay for delay sensitive applications. In addition, an assessment of fairness guarantee is considered across all users.

Key Words—C-RAN; Wireless Virtualisation; LTE, Allocation algorithm.

I. INTRODUCTION

According to Cisco’s forecast traffic report [1] issued in 2016, it is expected that the data transmission volume will grow 10 folds by 2019. This will further increase the demands for the mobile network operators to handle the unprecedented rate of growth in network usage. Thus new approaches to reshape the network architecture are gradually evolving. C-RAN is foreseen to be the leading technology in future 5G mobile networks that can manage the growing capacity demands efficiently [2]. C-RAN and “traditional” Radio Access Networks (RANs) are significantly different. In LTE, the RAN is composed mainly of distributed base stations that are called eNodeBs. However the Base Stations (BSs) functionalities in C-RAN are split between two main entities known as the Remote Radio Head (RRH) that is designed to be technology-agnostic, and the Base Band Unit (BBU) that is responsible of the core baseband processing. The network that connects the BBUs with the RRHs is named fronthaul [2]. BBUs are congregated in a pool in order to centralise the signal processing whereas RRHs are located at the BSs sites. In this deployment, the RRHs can provide better coverage and capacity in very dense areas. In addition, the computational resources can be shared in the BBU pool for multiple sites [3]. This is useful when the BBU pool serves sites with diverse traffic profiles (e.g., residential or offices) and associated BSs loads. However, in C-RAN the communication between the BBUs and the RRHs has to be transported with Quadrature Signals I/Q data that leads to high capacity requirements in the fronthaul. Thus current researches consider the required capacity and latency based on the specification of the Common Public Radio Interface (CPRI) [4].

On the other hand, wireless network virtualisation has been a hot topic and a leading research direction. The aim of the virtualisation is to allow operators to share common physical infrastructure and to coexist on the same platforms. The related concept relies on decoupling the Mobile Network Operator (MNO) into two distinctive roles [5]. Firstly there are the Infrastructure Providers (InPs) who deploy the physical network, and secondly the Virtual Operator (VO) that is responsible of customised user services and delivers them by renting resources from InPs. The concept of virtualisation has been considered for servers and routers virtualisation such as in [6]. Resource assignment among VOs needs to take into account many factors such as scheduling fairness and end user Quality of Service (QoS). The authors in [7] introduced an algorithm based on C-RAN and network virtualisation to minimise the network latency. Their proposed scheme considers the cell reselection challenge in small cells environments. Moreover, relevant research has commenced to validate the potential deployment of this concept in LTE wireless technology such as in [8]. The authors in [8] concentrate on resource allocation across multiple VOs in a single cell and estimate the gain obtained by applying resource sharing (e.g., enhanced resource utilisation).

This paper investigates the virtualisation of the air interface at the base station in a C-RAN topology. However, as the BBU is the intelligent entity in the C-RAN, the resource allocation and the air interface virtualisation are presumed to occur at the BBU and performed based on two scheduling schemes. The first is termed as Traffic Aware Joint Scheduling (TAJS) that takes into account all cells under one particular BBU. The second scheme is named Traffic Aware per RRH Scheduling (TARS) and inspired by [8] (that only considers one cell). Nonetheless, the second scheme applies that technique on all RRHs to evaluate the performance from a system perspective unlike the one cell assumption in [8]. The goal of the proposed scheduling schemes is to apply collective scheduling and maximise the spectral efficiency. This study will propose different scenarios of interaction between VOs in each BBU based on the number of users associated with each traffic type per BBU. This can help in terms of exploring the contractual area of the relationship...
between the InPs and VOs, and the VOs themselves. This will enhance the market as new opportunities will rise for new players to provide new services to their clients using virtualised networks.

The rest of the paper is structured as follows. Section II introduces the C-RAN network system model. Section III illustrates the proposed algorithms and methodology. Scenario configuration and system parameters are presented in section IV. Simulation results and overall evaluation of the proposed schemes are covered in section V. Finally we conclude the paper in section VI.

II. C-RAN Network Model

In order to compare the proposed algorithms against the baseline, in this paper we make use of a computer model designed and developed with OPNET modeler and Matlab. The network model is comprised of different elements that model an Ethernet-based C-RAN end-to-end as shown in Fig 1, together with the relevant protocols employed in the data plane, extensively described in our previous work [3]. The network topology consists of 9 RRHs each modelling an urban macro cell with omnidirectional antennas and Inter-Site Distance (ISD) of 500m. The RRHs are connected to a pool of 3 BBUs through a model of an Ethernet fronthaul network. Lastly, the whole mobile core network and the internet are simplified using a model of a server that is used to generate realistic application-layer traffic. This simplification of the core network is considered a reasonable assumption, being the focus of this work on the performances of the LTE scheduler for the air interface, at the MAC layer of the BBU.

To complete the network topology, models of the Users (UEs) are uniformly distributed among the cells and are associated with the different traffic profiles and VOs detailed in the following. As a consequence the wireless channel is modelled as Extended Typical Urban model (ETU) in 3GPP TS 36.101 and implemented in Matlab, together with the remaining of the Physical layer (PHY) processing. Finally, to model the application behaviour closely to a realistic scenario, the LTE model implements fast retransmission techniques such as Automatic Retransmission reQuest (ARQ) and its hybrid counterpart (HARQ), as detailed in [3].

III. TAJS - TARS Algorithms & Proposed Dynamic VOs Allocation Scheme

The authors in [8] have proposed the virtualisation for the eNodeB in LTE. Consequently, the BBU has to be virtualised in an analogous approach to the eNodeB in [8] in order to apply the same concept in C-RAN. The “hypervisor” (the virtualisation enabler entity) is presumed to be embedded in the BBU pool. In order to virtualise the eNodeB into a number of virtual eNodeBs (each associated with VO), the hypervisor has to schedule the physical resource among different VOs. It collects the relevant data of the traffic load for the different user traffic types (mobile traffic video, mobile web/data, audio), user channel conditions and Quality of Service (QoS) on a pre-defined basis for each BBU’s individual cell, as well as contractual information between the InPs and the VOs. The a-priori knowledge of the daily experienced traffic status is employed to process the scheduling of the Physical Resource Blocks (PRBs).

The authors in [9] have presented adaptive and distributed Coordinated Multi-Point (CoMP) and coordinated Scheduling (CS). Their proposed algorithm selects Joint Transmission JT or (CS/CB) transmission mode in accordance with UEs SINR threshold that is determined as a system parameter. Other studies have considered spectrum efficiency when evaluating JS such as [10]. Furthermore, the authors in [11] proposed an interference-aware JS scheme.

Fig 2 displays the logic flow of the (TAJS) algorithm that will be referred along with TARS algorithm as TAJS and TARS respectively in the rest of the paper. TAJS employs a Joint Scheduling (JS) technique. The design of the BBU in C-RAN has inspired us to exploit the related implementation to apply the JS mechanism without the need for BSs coordination as suggested in [9] and [11]. The regular observations of traffic loads across all BBUs can be formulated in a database built-in the hypervisor to be able to divide the spectrum between the VOs taking into account different performance criteria.

The embedded hypervisor in the BBU pool is designed to execute the TAJS algorithm in the BBUs based on user traffic profile and cell type (residential or offices). Each cell type has a different traffic distribution during the times of the day, for instance, a residential cell is more loaded with traffic like video streaming or social media at evening/night time, while offices are loaded with voice and heavy browsing during the working hours.
The next section introduces the scenarios where the times of the day affect the traffic distribution. The TAJSA collective scheduling aims to allocate resources for 3 cells that represent the number of cells served by a single BBU. The dynamic allocation mechanism allows the VO to rent out unused PRBs to other VOs when it does not experience any traffic running at that time slot. It ensures that all PRBs are utilised regardless to which VO they belong. The C-RAN model is implemented with the assumption that each RRH has a channel Bandwidth (BW) of 20 MHz. In other words each RRH can utilize 100 PRBs. Hence in TAJSA, the collective BW from all cells served by single BBU (300 PRBs) has to be divided among the VOs which are assumed to be three in this paper and classified as:

**First VO (premier):** VO1 requests the greatest portion of the collective bandwidth at the BBU side with fixed guarantees as well as any unused PRBs from other VOs traffic allocation. **Second VO:** VO2 requests a guaranteed maximum BW with dynamic allocation based upon other VOs traffic allocation. **Third VO (Best Effort) (BE):** VO3 has minimum guarantees of BW at the BBU side. It has the chance of being allocated more PRBs that are rented out from other VOs if they do not experience heavy traffic load at that time slot. As clarified in the next section, heavy traffic profile users are the next are mapped to premier VO which are followed by other VO’s users based on their profile as well. To impose fairness, a re-ordering process occurs every scheduling turn as shown in Fig 2. This process depends on three factors sorted according to their weight as illustrated in Fig 2. The first metric is the user association (to which VO the user belongs). The premier VO has the greatest value followed by the second and the BE. The algorithm by this metric can guarantee that VO1’s users will be granted the highest priority over other users in the same cell. The second metric is the user’s weight which is defined by the number of times a particular user is scheduled up to the current time slot, and last factor to be considered is the time of user last scheduling turn (LST). The last two factors provide a fairness within the VO’s users as the users with least number of scheduling times will be prioritised. At every scheduling turn, VO1’s users PRBs are allocated collectively at the BBU side (VO1’s users could be from all cells), after that VO2’s users are scheduled followed by the BE VO’s users. As the hypervisor has 300 PRBs for each BBU, it calculates the number of utilised PRBs at the end of the scheduling turn. If the relevant number does not exceed 300 PRBs, then TAJSA starts over the scheduling loop again to schedule the VO’s users who have not been allocated any resources at that turn. This occurs when there is a shortage in PRBs in their VOs at the time of the user’s scheduling turn.

On the other hand, TARSA is based on per RRH scheduling. The logic flow of TARSA is slightly deviated from TAJSA as scheduling occurs per RRH individually and not collectively. The VOs are allocated the associated portion of BW at the start and that will not change until the end of the simulation. The slight change in TARSA scheme is briefed as First VO (premier) is allocated only a fixed guaranteed number of RBs as agreed in its contract with the InPs. Second VO is configured with a dynamic guaranteed allocation up to certain threshold (50 RBs if there are 50% associated users or more out of total cell users, 45 RBs for 40% associated users, and else is 38 RBs). Third VO (Best Effort) (BE) is allocated in BE manner with no guaranteed PRBs depending on its associated users. Unlike TAJSA, in this algorithm each VO utilises only its initially allocated PRBs, thus they will not be able to rent out any unused resources. This might cause a waste of resources especially when the owner VO is idle while other VOs need resources. The same re-ordering fairness procedure is imposed in TARSA.

The performance of the TAJSA will be compared against TARSA and the standard RR scheduler [12] that has been implemented in our C-RAN model. The users in RR are assigned the RBs in turn (one after another) without considering their traffic profiles and QoS requirements, but it assures all users are equally scheduled. However, TAJSA and TARSA can be considered less fair than RR since they map certain users based on their traffic profile to the associated VO.

**IV. SIMULATION SCENARIOS & CONFIGURATIONS**

The paper considers two scenarios; each scenario combines three cases, firstly with the TAJSA virtualisation, secondly with the TARSA, and finally with standard RR. The application server generates application–layer traffic; the traffic models are represented as traffic profiles and presented in Table I. Video streaming denoted as V-Str and Web profiles are the same for all scenarios, however social media & VoIP (SMV) differs for different times of the day. The configured traffic profiles take into account the traffic growth estimations by Cisco [1] that states mobile video will consume much of the mobile traffic.
The proposed scenarios can be enlisted as follows: Sc.1 has association with night time where (Vi-Str) is the predominant traffic across all BBU1 cells (8 users) followed by the (SMV) profile (6 users) and finally the Web profile (4 users) in the following distribution [(3-2-1), (2-2-2), (3-2-1)] for RRHs 1, 2, 3 respectively. The notation (3-2-1) is elaborated as VO 1, VO2 and VO3 have 3, 2, 1 users respectively, in a particular cell. BBU1 has 8 (Vi-Str) users, 7 SMV and 3 web users as [(2-2-2), (3-3-0), (3-2-1)], BBU3 has 7 (Vi-Str) users, 6 SMV users and 5 Web users as [(2-2-2), (3-2-1), (2-2-2)]. The TAJSA virtualised case is termed as UnifVirt1TARS whereas the TARSA case is named as UnifVirt2TARS. Both schemes assume that Vi-Str users in all cells are mapped to VO1, while SMV users are mapped to VO2 and Web users to VO3. The portion of BW (in terms of PRBs) is allocated to each VO in accordance with the number of the users running the corresponding traffic profiles. Sc.2 is linked with daytime, where the SMV is the dominant traffic across majority of cells, however some cells have other prominent traffic profiles, and this will not change the criteria as SMV users are always mapped to VO1, Vi-Str to VO2 then Web to VO3. BBU1 has 8 SMV users, 6 Vi-Str and 4 Web [(3-2-1), (2-2-2), (3-2-1)], BBU3 has 7 Vi-Str, 7 SMV and 4 Web. BBU1 has 9 SMV users, 7 Vi-Str users and 2 Web [(3-3-0), (3-2-1), (3-2-1)]. The TAJSA virtualised case in this scenario is named as UnifVirt2TARS and TARSA case is called UnifVirt1TARS.

The BBU collective scheduling in this case means users that belong to BBU VO1 will be scheduled first taking advantage of the highest priority than others across all cells in that BBU. VO2 & VO3 users are scheduled after in turn. Table II summarizes the simulation parameters.
divided by the number of cells. Fig 4 displays the average cell throughput (Sc.1 & Sc.2). The results show that the virtualised cases in scenario 1 (UnifVirt1-TARS, UnifVirt1-TARS) have the same performance with RR with no additional gain. Conversely, TAJSA and TARS outperform RR in Sc.2. When applying the TARS proposed in literature over nine cells (as in our model), the overall cell average shows better performance with 4% gain (similar with [8]). However, the proposed TAJSA (UnifVirt2-TARS) outperforms TARS & RR by 5.2% & 9.5% respectively.

Both TAJSA and TARS algorithms differ from RR in terms of the scheduling intervals. The users are scheduled fairly with no priority in RR, where we assume that each cell contains 6 users as illustrated in our model with all users having the same traffic profile all the time and are allocated the same packet size every turn (one user to schedule at each turn). In this case, all users have to wait the same time interval to be rescheduled. While in priority-based algorithms (TAJSA & TARS), VO1 users have to wait the same time interval to be rescheduled. In both cases, all traffic profiles are the same and all users are allocated the same packet size.

To draw a simple conclusion, TAJSA VO1 specific users have higher throughput than the standard TARS and RR cases for the same set of users in the same period slot that translates to higher average cell throughput (taking into account the traffic profile and number of users supported by VO1). This conclusion is also supported by the average delay results between consecutive scheduling turns. Note, this metric has been recorded for the same three users that belong to three distinctive VOs in all scenarios. Figure 7 shows the mean value of the delay in BBU1. User 0-2 (RRH1, VO1 user) has less delay with TAJSA than TARS and RR in both scenarios. For User1-4 (RRH2, VO2 user) and User2-6 (RRH2, VO3 user), the delay depends on the scenario considered. However, having longer delay will not have a major impact on VO2 users as they have lighter traffic load.

Although the algorithm objective is to achieve higher data rates for VO1 users services that can be assessed by the average Vi-Str or SMV user throughput (this metric is calculated by averaging the throughput of all the related users across all RRHs in our C-RAN topology). Other metrics should be monitored to assure better QoS. VoIP traffic requires different QoS criteria than video. The relevant QoS includes latency which is measured in Opnet as voice packet end-to-end delay that is defined as the time for packet to be transmitted from the source to the destination including encoding/decoding, transmission, propagation processing and queue delays. Other factors to consider are packet delay variation that can be defined as the variance among end to end delays for voice packets, and Jitter that is a significant parameter and it is defined as the variation in the delay of received packets [13]. Jitter in Opnet is computed as the difference between the delays of two consecutive packets at the receiver and transmitter side respectively. The previous voice parameters are averaged for all users’ voice packets during the simulation time. Table III illustrates the above mentioned metrics for the proposed cases (UnifVirt1-TARS, UnifVirt1-TARS and RR) in Sc.1.

The purpose of the selection of Sc.1 for voice evaluation is that voice users in the proposed cases (UnifVirt1-TARS, UnifVirt1-TARS) are mapped to second VO with less priority than VO2 users, therefore there is a potential impact on their services metrics. It can be observed that users in (UnifVirt1-TARS) experience less delay for all measurements than UnifVirt1-TARS and less than RR in terms of Jitter. The dynamic JS proposed in TAJSA allows even better performance in VoIP metrics than TARS. The authors in [8] presumed that VO1 has a fixed guarantees of the cell bandwidth that is allocated all the time regardless of being utilised or not.
This implies that 27% of resources are being wasted. In the same manner, in BBU1, the mean number of utilised PRBs is 122 and the percentage of wasted resources is 18%. The static mechanism of not renting out resources to other VOs when they are not utilised results in lower system spectral efficiency. TAJS addresses that point to enhance the utilisation techniques by allowing VOs to use other VOs resources when needed. For instance, BBU1 has utilised in UnifVirt2_TJS an average number of 125 RBs per scheduling turn efficiently. BBU3 has used in UnifVirt2_TJS 251 RBs efficiently, that is greater than 122 in the same BBU for TARSA.

VI. CONCLUSION

In this paper we have proposed Traffic aware Joint & per RRH Scheduling (TAJS) and (TARS) mechanisms for network air interface virtualisation in a C-RAN deployment and LTE platform. TAJS has been designed to adapt dynamically to the variation in VO’s traffic load in different times of the day. TARS only depends on the time of the day with static predefined allocated resources only. Results indicate that TAJS outperforms TARSA and standard RR in terms of average cell end user throughput. From a system point of view, TAJS has also improved the spectral efficiency. In addition, no negative impact on the end-to-end VoIP packets delay and jitter at the end user side is observed. This work can be developed further by defining more complex scheduling algorithms that address other areas such as RRH’s coordination, joint transmission and the impact of fronthaul delay on them.

REFERENCES