A framework for resources allocation in virtualised C-RAN

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

Published in:
2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications

Link to article, DOI:
10.1109/PIMRC.2016.7794959

Publication date:
2016

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.

Peer reviewed version

Link to published version (if available):
10.1109/PIMRC.2016.7794959

Link to publication record in Explore Bristol Research

PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via IEEE at https://doi.org/10.1109/PIMRC.2016.7794959. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms
A Framework for Resources Allocation 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- The vast growth in mobile data traffic with its increasing capacity demands necessitates investigating future solutions to cope with these challenges. Network Virtualisation is considered one potential solution to simplify the current wireless networks. Recently efforts have been made to show the performance gains of different resource allocation schemes in legacy LTE air interface virtualization. Moreover Cloud – based mobile networks are anticipated to play a significant role for next generation mobile networks. This paper investigates the potential deployment benefits of a novel resource virtualization algorithm (Traffic Aware Joint Scheduling) in Cloud-RANs (C-RAN). Air interface resources are coordinated and allocated dynamically by a hypervisor among different virtual operators (VOs). Three distinctive schemes are proposed and evaluated against standard Round Robin (RR) C-RAN scheduling. Simulation results show improvements in the throughput of the mobile video traffic and reduction in end-to-end delay for delay sensitive applications. In addition, an assessment of fairness guarantee is considered across all users. Note that this paper considers the impact of the proposed schemes on the transmission/data plane.

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

I. INTRODUCTION

The exponential increase in number of devices connected in mobile networks will lead to a data tsunami in the coming years. According to [1], it is expected that the data transmission volume will grow 10 folds by 2019. This will require mobile networks to cope with an unprecedented rate of growth in network usage. Thus new approaches to reshape the network’s architecture are required to cope with an unprecedented rate of growth in network usage. This is advantageous particularly when the BBU pool serves sites with diverse traffic profiles (e.g., residential or offices). Nevertheless C-RAN has a disadvantage that the communication between BBUs and RRHs has to be done with I/Q data. The fronthaul network in this case requires high capacity. Thus current research considers 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 attracted attention since it aims to enhance the diversity, manageability, flexibility and energy efficiency of current networks [5]. The concept of mobile cellular virtualisation relies on decoupling the mobile network operator (MNO) into two distinctive roles [6]. Firstly there are the infrastructure providers (InPs) who deploy the physical network and secondly the Virtual Operator (VO) that handles the customised user services and delivers them by renting resources from InPs. Recently research has been carried out to investigate the potential of this concept in LTE wireless technology [7]. The authors in [7] concentrate on resource allocation across multiple VOs in a single cell and estimates the gain obtained by applying resource sharing (e.g., enhanced resource utilisation).

This paper studies the virtualisation of the air interface at the base stations in a C-RAN topology (RRHs). However, as the BBU is the intelligent entity in the C-RAN, the resource allocation and air interface virtualisation is presumed to occur jointly taking into account all cells under one particular BBU unlike [7] that only considers one cell. The goal is to implement an algorithm that manages the contract between VOs and InPs and between VOs themselves. In addition it applies collective scheduling and maximises the spectral efficiency.

The rest of this paper is structured as follows. Section II illustrates the motivation behind network virtualisation and describes relevant work in this context. The C-RAN network system model is introduced in section III. Section IV describes the proposed algorithm and methodology. Scenarios, configuration and system parameters are presented in section V. Simulation results and overall evaluation of the proposed scheme are covered in section VI. Finally we conclude the paper in section VII.

II. NETWORK VIRTUALISATION

The concept of virtualisation has been exploited in operating systems and personal computers memory. However, research is
III. C-RAN Network Model

This study employs a C-RAN network model that has been developed to provide dimensioning constraints for the fronthaul network [3]. The model implementation is based mainly on 5 node types as illustrated in Fig.1. The node types are Application server, BBU, gateway RRH and UE, the nodes are described in detail in [3] and therefore a brief description of each is given in this paper. The application-layer traffic is generated by an application server. Three BBUs as shown in Fig 1 that are utilised in this model.

The BBUs are grouped in one pool and logically parted but physically co-located. The BBU is implemented to play the role of eNodeB in LTE with the same protocol stack. The segmentation process of the data packets received on the BBU SI interface is executed at the Radio Link Control (RLC) to fit the Physical Resource Blocks (PRBs). The Medium Access Control layer (MAC) is responsible of passing the segmentation parameters to the RLC. These parameters are acquired via the channel Quality Indicator (CQI) feedback from the UEs side. The scheduling process of segmented packets occurs at the MAC layer in granularity of 1 ms in accordance with LTE standards. The communication between BBUs and the gateway is CPRI- based; the gateway is the interface between the BBU pool and the fronthaul network along with the RRHs. It plays the role of encapsulation of the CPRI packets from the BBU pool into Ethernet frames that are forwarded afterward to the associated RRH Ethernet interface.

On the RAN side, the model is composed of 9 omnidirectional antennas at the RRHs. Each RRH covers an urban macro cell with Inter-Site Distance (ISD) of 500m. The wireless channel is modelled as Extended Typical Urban model (ETU) provided by 3GPP TS 36.101[12].

In the uplink (UL), a wideband Channel Quality Indicator (CQI) is sent from the UEs periodically to the RRHs, while the CQI depends on measuring the Reference Signal (RS) embedded in the OFDM PRB. The Adaptive Modulation and Coding Scheme (AMC) can only work if the BBU is informed of the channel quality seen by the UE, and depending on the reporting mode (Wideband in this model), the CQI is used by the BBU as an input traffic scheduling process. The CQI reports are received by the RRH and encapsulated as packetized CPRI into the payload of the Ethernet frame [3], these frames are received by the gateway that decapsulates the CPRI payload in order to forward it to the corresponding BBU.

![Figure 1. C-RAN Network Model](Image)
IV. Tajs Algorithm & Proposed Dynamic Vos Allocation Scheme

The BBU in the C-RAN is the responsible entity of scheduling the air interface resources. The authors in [7] have proposed the virtualisation for the eNodeB in LTE. Therefore, in order to apply the same concept in C-RAN the BBU has to be virtualised similarly. Their research is based on a virtualisation enabler entity “Hypervisor”. 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. This paper assumes that the hypervisor is embedded in the BBU pool. It collects information about the traffic loads for different traffic types (mobile traffic video, mobile web/data, audio), user channel conditions, QoS on predefined basis for each BBU’s individual cell as well as contractual data between the InPs and the VOs. The a-priori knowledge of the traffic status is employed to process the scheduling of the interface resources (PRBs) between them. Fig 2 demonstrates the logic flow of the algorithm.

The proposed algorithm makes use of Joint Scheduling (JS) that has been a hot topic due to its encouraging results. Extensive CoMP scheduling algorithms have been investigated recently such as [13]. While, the authors in [14] proposed an interference-aware joint scheduling scheme. The algorithm proposed in this paper is based on the Traffic Aware JS (Tajs) technique that distributes (PRBs) among different users depending on their traffic profile. The algorithm considers several types of traffic that are mapped into three VOs. We have exploited the design of the BBU to apply the JS mechanism without the need for BSs coordination as suggested [14]. The lengthy monitoring of traffic loads across all BBUs can be added to the hypervisor database to be able to divide the spectrum between the VOs taking into account other criteria.

The embedded hypervisor in the BBU is designed to execute the Tajs algorithm in the BBUs based on users traffic load and cell’s type (residential or offices). Each cell type has a different traffic distribution during the times of the day, for instance, a residential cell is heavily loaded with traffic like video streaming or social media at evening/night time, while offices are more loaded with voice and heavy browsing during the working hours. The study scenarios (each with unique traffic distribution based on the time of the day) are introduced in the next section.

The Tajs algorithm is based on the collective scheduling for 3 cells that represent the number of cells served by single BBU. The C-RAN model is implemented on the assumption that each RRH has a channel BW of 20 MHz. In other words each RRH has a transmission BW of 100 RBs. Hence the Tajs algorithm has to split 300 PRBs between VOs which are assumed to be three in this paper and classified as:

First VO (premier): VO1 has the highest priority and corresponds to a premium class of service. It requests the greatest portion of the collective bandwidth at the BBU side with fixed guarantees as well as any unused PRBs from other VOs. Second VO: VO2 corresponds to an assured/controlled-load type of service and requests a guaranteed maximum BW with further more dynamic allocation based upon other VOs traffic allocation. Third VO (Best Effort) (BE): VO3 has minimum guarantees of collective BW at the BBU side with the chance of being allocated more PRBs that are rented out from other VOs if the load permits.

The dynamic allocation mechanism allows the VO to rent out unused PRBs to other VOs when it doesn’t experience any traffic running at that time slot. It ensures that all PRBs are being utilised regardless to which VO they belong. The performance of the Tajs algorithm will be measured against a standard Round Robin (RR) [15] scheduler that has been implemented in our C-RAN model. The users in RR are assigned the PRBs in turn (one after another) without considering their traffic profiles and QoS requirements, but it assures all users are equally treated. However, Tajs can be considered less fair than RR since it maps certain users based on their traffic profile to the associated VO. The VO’s chunk of bandwidth (in terms of PRBs) is determined at the starting stage of the algorithm. It relies on the VO’s number of the associated users and the VO class (premium or second …). As clarified in the next section, heavy traffic

Figure 2. Proposed Algorithm Principle
profile users are mapped to premier VO which are followed by other VO’s users based on their profile as well. To impose fairness, a re-sorting process occurs every scheduling turn as shown in Fig 2. This process depends on three factors ordered in accordance to their power as shown in Fig 2. The first factor is the user association (to which VO the user belongs) that premier VO has the greatest metric followed by other VOs in turn, the second is the user’s weight which is defined by how many times that user is scheduled until the current time and finally the last factor is the time of user last scheduling turn (LST). By considering the first factor, premier VO’s users are guaranteed to be scheduled first until VO limit is attained. The other two factors provide fairness within a particular VO scheduling process as users with less scheduling times move to the head of the list. Furthermore, if two users have the same weight, the user with oldest LST will have higher priority. The scheduling turn starts by allocating VO1’s users PRBs 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 used PRBs at the end of the scheduling turn. If the relevant number doesn’t exceed 300 PRBs then TAJS starts over the scheduling loop again to schedule the VO’s users who haven’t been scheduled at that turn. This occurs when there is a shortage of PRBs in their VOs at the time of the user’s scheduling turn. The VOs that experience shortage of resources will rent unutilised PRBs from other VOs to meet the need of their associated users. Each individual VO scheduling process runs as the right hand side of the flow chart (enclosed by the dashed rectangle) in figure 2. It commences in accordance with the users order after the re-sorting stage.

V. SIMULATION SCENARIOS & CONFIGURATIONS

This paper considers investigation of three proposed scenarios; each scenario combines two cases, one with the TAJS virtualisation algorithm and one with the standard RR. The RR case is based on per RRH scheduling while the hypervisor of the TAJS algorithm is on the BBU basis. The BBU collective scheduling 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. Table I summarises the simulation parameters.

The application server generates application-layer traffic; the traffic models are represented as traffic profiles and presented in table II. The configured traffic profiles take into account the traffic growth estimations done by Cisco [1] (e.g. mobile video will consume much of future mobile traffic). Vi-Str and Web profiles are the same for all scenarios, however SMV differs slightly for different times of the day.

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>DEPLOYED TRAFFIC PROFILES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Streaming Traffic Model</td>
<td></td>
</tr>
<tr>
<td>Incoming/Outgoing Stream inter-arrival time (seconds)</td>
<td>Const (0.01)</td>
</tr>
<tr>
<td>Incoming/Outgoing Stream Frame Size (Bytes)</td>
<td>Const (2000)</td>
</tr>
<tr>
<td>VoIP traffic Model</td>
<td></td>
</tr>
<tr>
<td>Encoder Scheme</td>
<td>GSM FR</td>
</tr>
<tr>
<td>Voice packets per frame</td>
<td>1</td>
</tr>
<tr>
<td>Compression Delay (Seconds)</td>
<td>0.02</td>
</tr>
<tr>
<td>Decompression Delay (Seconds)</td>
<td>0.02</td>
</tr>
<tr>
<td>Light Web browsing (HTTP v1.1)</td>
<td></td>
</tr>
<tr>
<td>Page Inter-arrival Time (seconds)</td>
<td>Exponential (Mean 120)</td>
</tr>
<tr>
<td>Page Size (Kbytes)</td>
<td>Uniform [2.5,10]</td>
</tr>
<tr>
<td>Social Media (Heavy Browsing)</td>
<td></td>
</tr>
<tr>
<td>Page Inter-arrival Time (seconds)</td>
<td>Constant (2)</td>
</tr>
<tr>
<td>Page Size (Kbytes)</td>
<td>(S1 &amp; S2) all pages include VoD videos</td>
</tr>
<tr>
<td></td>
<td>Uniform [80,400] plus 3 Short Videos (VoD)</td>
</tr>
<tr>
<td>Page Size (Kbytes) (S3), all pages include VoD videos</td>
<td>Uniform [160,800] plus 3 Short Videos (VoD)</td>
</tr>
</tbody>
</table>

Each user is mapped one traffic profile for the whole scenario simulation time. The traffic profile consists of one or more traffic types (e.g. SMV includes Social Media and VoIP). The related traffic profiles account for the aggregated traffic in the network in a certain time of the day. The proposed scenarios can be enlisted as follows:

Sc.1 has no association with time of the day, at BBU1 Video streaming (Vi-Str) is the predominant traffic (8 users) followed by the Social Media & VoIP (SMV) profile (6 users) and finally the Web profile (4 users) in the following distribution [3-2-1], [3-2-1], [3-2-1] for RRHs 1,2,3 respectively. The notation (3-2-1) is elaborated as VO1, VO2 and VO3 each is associated with 3, 2, 1 users respectively in a particular cell. BBU2 has no certain predominant traffic, BBU3 is SMV predominant with (9 users) then Vi-Str (7 users), and web (2 users) distributed as [3-3-0], [3-2-1], [3-2-1]. The virtualised case is termed as Dist.Virt. It assumes that three VOs are configured differently for different BBU’s, for instance in BBU1 (Vi-Str) users are mapped to VO1 (as (Vi-Str) is the predominant traffic) while SMV users are mapped to VO2. The algorithm assumes that when there is no dominant traffic in the BBU, SMV users will map to VO1 and Vi-Str are mapped to VO2 which is the case in BBU2.

Sc.2 is linked with evening time, where the Vi-Str is the dominant traffic across many cells, however some cells have
other prominent traffic profiles, but this will not change the criteria as Vi-Str users are always mapped to VO_1, SMV to VO_2 then Web to VO_3. BBU_1 has 8 Vi-Str users, 6 SMV and 4 Web [(3-2-1), (2-2-2), (3-2-1)]. BBU_2 has 7 Vi-Str, 7 SMV and 4 Web. BBU_3 has 7 Vi-Str users, 9 SMV users and 2 Web [(3-3-0), (2-3-1), (2-3-1)]. The virtualised case in this scenario is named as Unified2-Virt.

VI. SIMULATION RESULTS

The C-RAN model used in this paper has been implemented in the Discrete–Event Simulator (DES) tool OPNET modeler. The results section presents the potential performance gain that could be achieved by applying the TAJS algorithm in an actual deployment with comparisons with a standard RR. Fig 3 shows PRBs allocation per each VO at BBU_2. Sc.1 over 0.5 seconds.

![Figure 3. BBU per VO allocated bandwidth (PRBs)](image)

It can be observed that VO_{3,2} occupy almost all of the aggregated bandwidth and the number of allocated PRBs varies with time depending on the load of each traffic profile and the VO’s contract. The performance can be evaluated by different network metrics such as the average user, cell or BBU throughput (bps). Average cell throughput is defined by the sum of all users throughput across all cells divided by the number of cells. Fig 4 and Fig 5 demonstrate average cell throughput (Sc.1 & Sc.3). The results show that the virtualised cases in both scenarios outperform the standard RR by 8.13% and 21.23% for Sc.1 & Sc.3 respectively.

Mobile video content has much higher bit rates than other mobile content types. Thus it is intuitive that allocating more air interface resources for Video users than others with higher priority will result in higher average network throughput. In Sc.1 each BBU has its own priority-based algorithm (scheduling) as Vi-Str users are mapped to VO_1 in BBU_1. However, the SMV users are mapped to VO_1 in BBU_2/3. Sc.3 grants the priority to SMV users across all cells (SMV is predominant). The distinction between TAJS and RR can be clarified as following, assuming a cell composed of 6 users as in our model. The RR scheduler processes all users fairly with no priority. By assuming all users are running the same traffic profile all the time and being allocated the same packet sizes every turn that does the scheduling for one user only. Therefore all users have to wait the same time interval to be rescheduled again. However, in the priority-based algorithm, for instance VO_1 has a greater BW portion of 140 RBs (in case of 8 BBU users running its respective traffic). Thus its users have the priority needed to be rescheduled in less time interval than RR and other VOs.

![Figure 4. Average Cell Throughput Sc.1](image)

To draw a simple conclusion, VO_1 specific users have higher throughput than RR case for the same set of users in the same period slot that translates in higher average cell throughput. This conclusion is supported by the statistic results recorded in the simulation that show how many times VO_1 users have been scheduled against the same user in the RR case in scenario 1. Numbers show that an average VO_1 user has been scheduled 16.8% more than RR case at the end of the simulation time.

Another statistic figure has been added to this evaluation in order to compare the average delay between each consecutive scheduling turns. This figure has been recorded for three users in the first scenario for both cases. Each user belongs to a distinctive VO in order to compare against the RR case as shown in table III.
User 0-2 (RRH1, VO1 user) has less delay when the TAJS algorithm is applied than the standard RR, this conclusion aligns with above mentioned result. User1-4 (RRH2, VO2 user) has longer delay than RR. This is logical as VO1 users are being scheduled more frequently, therefore other users will be scheduled less often. However VO2 and VO1 user's traffic data is less intensive traffic profile than VO1. The rest of the section investigates the potential impact on their throughput and the end to end delay. Although the algorithm objective is achieving higher bit rates for mobile video content services that can be assessed by computing the average video user throughput (this metric is calculated by averaging the throughput of all video users across all RRHs in our C-RAN), other traffic profiles should be monitored to assure fairness. 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 delay. Other factors to consider are packet delay variation that can be defined as variance among end to end delays for voice packets and jitter. Jitter is a significant parameter used in packet switch networks, it is defined as the variation in the delay of received packets. The packets at the sender are sent continuously with packet spaced evenly apart, however the delay between the packets can vary instead of being constant [16]. The average jitter should be less than 60 ms according to International Telecommunications Union (ITU) [17]. Jitter in Opnet can take negative values as it is computed as the difference between the delays of two consecutive packets at the receiver and transmitter side respectively. All previous voice parameters are averaged for all voice users’ packets across the simulation time. Table IV highlights the above mentioned metrics for the proposed schemes and standard C-RAN RR. The granularity of computing the aforesaid values is taken every 1 ms, however the average value is computed for each.

The voice users in both scenario 1 &3 are allocated with higher priority than others, this will introduce less buffering delay. Scenario 2 shows improvement as well although VoIP users are mapped to the second VO.

In order to evaluate the performance gain of the Vi-Str and SMV profiles users, the average Vi-Str and SMV user throughput is taken across all cells for three scenarios. The average Vi-Str user throughput variance around the mean is limited, therefore a mean is considered for average Vi-Str throughput in each scenario as illustrated in Fig 6. It can be noticed that Dist-Virt and Unified2-Virt show significant improvement as compared to RR in terms of throughput by 38.9% and 25.74% respectively. However, Unified1-Virt shows slight improvement over standard RR. In this scenario the number of SMV exceeds the number of Vi-Str users. It can be concluded that mapping video streaming traffic users to VO1 when they don’t form the majority in the BBU has minor improvement. Figure 7 & 8 demonstrates the VoIP Users average air throughput in cumulative density function form for both Sc. 2 & 3.

![Figure 6. Average Video User Throughput (Mbps)](image)

![Figure 7. Average SMV User Throughput (Sc.2) CDF](image)

It can be noted that Unified1-Virt outperforms RR in Sc.2; the probability that an average SMV user experiences throughput less than 3.324 Mbps is 50% against 95% for the standard RR for this rate. In the same manner scenario 3 has similar performance since the proposed scheme Unified2-Virt shows improvement against the standard RR. According to Fig 8, we observed that in Unified2-Virt, around 15% of SMV users achieved throughput of less than 3.66 Mbps, while 100% of RR users are within this range. This implies that more resources are allocated for SMV users in Unified2-Virt than RR. With respect to the light web users, their throughput is monitored as other VO’s users’ throughput to assure fairness. The algorithm
implementation guarantees a limited number of PRBs for VO₃ in every scheduling turn. At the same time VO₃ can make use of the available PRBs left by VO₁ & VO₂ when they don’t have traffic to run which allows more efficient utilisation of spectrum as compared to static allocation as discussed in section IV.

![Figure 8. Average SM Profile User Throughput (bps)](image)

**Fig 8** depicts the average web profile throughput for each case. The average throughput is measured for all associated users. The results show a throughput gain for the proposed schemes in all scenarios by 18% and 11% for Sc2 and Sc3 respectively. Although the algorithm has the least priority for web profile users, the bursty nature of that profile has low level of traffic running most of the time with few sudden increases (sudden traffic peak) according to its distribution profile when the user starts the relevant light browsing.

![Figure 9. Average Web User Throughput (bps)](image)

**VII. CONCLUSION**

In this paper we have proposed a Traffic aware Joint Scheduling (TAJS) mechanism for network air interface virtualisation in a C-RAN architecture. TAJS has been designed to adapt resource virtualisation dynamically according to VOs traffic load balance. Greater throughput from a network point of view and an improved performance at the end user side are observed. The results showed a higher cell, $Vi-Str$ and $SMW$ user throughput against the RR scheduler in C-RAN. In addition, end-to-end packet delay and jitter are improved. This work can be developed further by defining more complex hypervisor scheduling techniques that addresses other points such as VO’s interference, RRH’s coordination, joint transmission and the impact of fronthaul delay on them.

**REFERENCES**