

Joint Power Allocation and Network Slicing In an End to End O-RAN System

1st Mojdeh Karbalaee Motalleb
Electrical and Computer Engineering
Tehran University
Tehran, Iran
mojdeh.karbalaee@ut.ac.ir

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

Abstract—Many major telecommunication companies confirmed the unification of the xRAN Group with the C-RAN Alliance to establish a more flexible and openness radio access network which is the Open-RAN (O-RAN) for the fifth generation of wireless technology.

To increase energy efficiency and optimize the allocation of resources, Network Slicing (NS) is considered as the best method for the fifth generation (5G) in order to virtualize the common physical network into several logical end-to-end networks. Every slice consists of a part of core network resources, network functions, and radio access network resources as a functional end-to-end network.

In this paper, we elaborate joint NS in RAN and Core of O-RAN system to investigate the power of each User Equipment (UE), connect slices to services and also connect physical Data Centers (DC) to slices to jointly maximize energy efficiency and minimize consumption power of RRHs and the cost of physical resources in a downlink channel. The problem is formulated as a mixed-integer optimization problem that can be decomposed into two independent sub-problems due to the fact that sub-problems are independent. Heuristic algorithms are proposed to each of sub-problems in order to connect slices to services, optimize power consumption and connect slices to physical resources to minimize the cost of total DCs simultaneously.

Index Terms—O-RAN, Network Slicing, Energy Efficiency, Data Center (DC)

I. INTRODUCTION

Recently, O-RAN, which is the integration and expansion of C-RAN and xRAN, is expected to be a key technology for 5G to enhance RAN performance and solve the part of challenges of mobile network operators (MNOs) in the best way. The Idea of O-RAN comes from two opinions. Firstly, according to real-time analytic used for artificial intelligence systems, the radio access networks must be evolved to be more intelligent and flexible than before. Furthermore, O-RAN can virtualize elements of the network with appropriate interfaces [1]. In an innovative O-RAN system, the programmable RAN software is decoupled from hardware, which can be run on any specific processing platforms, in order to be more flexible for MNOs especially for mobile virtual network operators (MVNOs) [2].

The core idea of C-RAN is to split the radio remote head (RRH) from baseband unit (BBU). Several BBUs operating on a cloud server will create a BBU-Pool, providing unified baseband signal processing with powerful computing capabilities. Moreover, in O-RAN technology, this separation is

implemented [3]. To communicate between BBU-Pool and RRHs, the fronthaul fiber link interface, is assumed with limited capacity. The compression of a message passed through these links is a consequence of limited fronthaul capacity [4]–[6].

xRAN technology, released in April 2018 as the next generation of RAN, has three fundamental features. The Control plane is decoupled from User plane. Besides, a modular eNB software stack is built to operate on common-off-the-shelf (COTS) hardware. Moreover, open north-bound and south-bound interfaces are introduced [7].

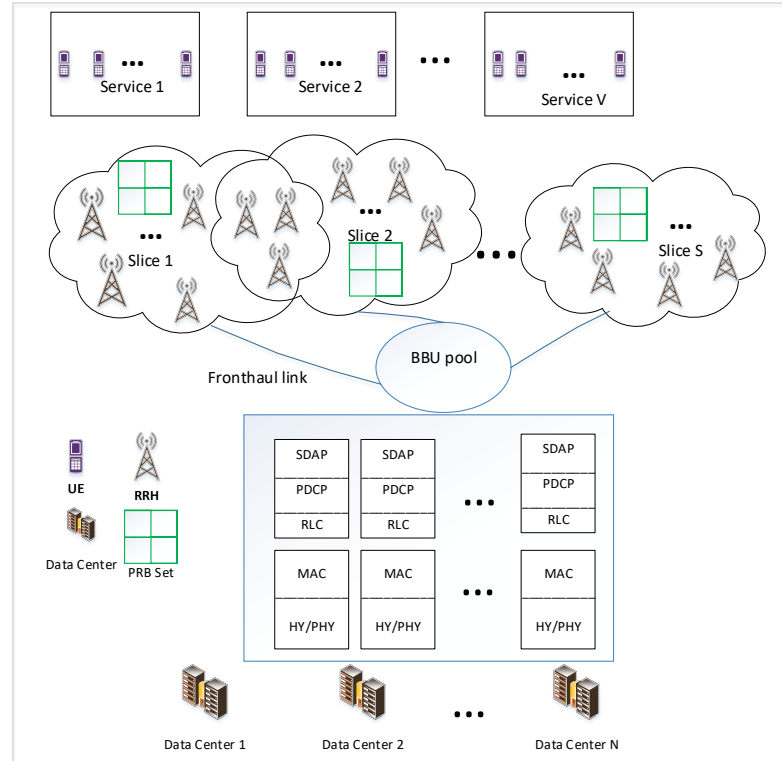


Fig. 1: Network sliced O-RAN system

To evolve servicing in 5G, separation of elements of software and hardware of network is deployed in network functions virtualization (NFV) technology. In this technology,

the functionality of networks is virtualized and divided into blocks of virtual network function (VNF). The responsibility of wireless systems of the fifth generation covers wide range types of services. In order to provide the requirements of these services, NS is implemented to virtualize the shared physical network into several logical end-to-end networks. Three different types of NS are introduced in [8] contains Core Slicing, RAN Slicing and Core-RAN Slicing. In Core-RAN Slicing, each slice of RAN is mapped to slices of Core. Also, UEs classified into a group of services according to their requirements. In addition, each service is connected to one or more Core-RAN slices based on the resource of slices. Using cloud-computing in BBU-Pool, the performance of the system is enhanced by virtualizing resources into virtual machines (VMs). Each VM has a computing processor in order to be run by the Virtual Network Functions (VNFs) which processed arrival data. Also, VNFs are mapped to physical resources through NS techniques [4], [9], [10].

The compression of messages passed through the fronthaul link due to the limited capacity of these links is considered in [5], [11]. In [12], dynamic network slicing is considered in Heterogeneous CRAN (H-CRAN) to maximize the weighted sum rate. Moreover, in [4], [9], [10] minimization of cost obtained by power consumption, also cloud processing and limited fronthaul capacity is considered. Furthermore, processing delay of each VM and wireless transmission delay is bounded.

In this paper, as depicted in figure 1, the downlink of the O-RAN system is assumed. UEs are divided to different groups according to their service requirements. Also RAN is decoupled to slices to provide requirements of services. Optimal power allocation and joint connecting slices to services are applied. In addition, connecting slices to physical resources is taken to account.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, first, we present the downlink (DL) of the O-RAN System. Then we obtain achievable rates and delays. Afterward, the main problem is expressed.

A. System Model

Suppose there are S slices Serving V services. Each Service $v \in \{1, 2, \dots, V\}$, consists of U_v single-antenna user equipments (UEs) that require certain service. Each slice $s \in \{1, 2, \dots, S\}$ consists of R_s RRHs and K_s Physical Resource Blocks (PRBs) as it is defined in the standard of LTE that the bandwidth (BW) of the channel is divided to PRBs in time series [12]. All RRHs in a slice that is connected to a service, transmit signals cooperatively to all the UEs in specific service [6], [13]. Each RRH $r \in \{1, 2, \dots, R\}$ is connected to BBU pool via an optical fiber link with limited fronthaul capacity. Suppose we have two processing layers in the BBU-Pool of O-RAN system. The lower layer is consist of high-PHY and MAC, and the upper layer is consist of RLC, PDCP and SDAP. Assume we have M_1 VMs in the first layer and M_2 VMs in the second layer for processing data. For simplicity,

suppose each VNF run on a VM and there is a one-to-one connection between VNFs and VMs. Each VNF in both layers connects to one or more slices. So in s^{th} slice, there are M_{s1} VNFs in the first layer and M_{s2} VNFs in the second layer. All VNFs in first and second layer has the computational capacity that is equal to μ_1 and μ_2 , respectively. Also, RRHs and PRBs can serve more than one slices.

B. Achievable Rate

The achievable data rate for i^{th} UE in v^{th} service can be written as

$$\mathcal{R}_{u(v,i)} = B \log_2(1 + \rho_{u(v,i)}); \quad (1)$$

Where B is the bandwidth of system and $\rho_{u(v,i)}$ is the SNR of i^{th} UE in v^{th} service which is obtained from

$$\rho_{u(v,i)} = \frac{p_{u(v,i)} \sum_{s=1}^S |\mathbf{h}_{R_s,u(v,i)}^H \mathbf{w}_{R_s,u(v,i)}|^2 a_{v,s}}{BN_0 + I_{u(v,i)}}; \quad (2)$$

Where $p_{u(v,i)}$ represents the transmission power allocated by RRHs to i^{th} UE in v^{th} service, also, $\mathbf{h}_{R_s,u(v,i)} \in \mathbb{C}^{R_s}$ is the vector of channel gain of a wireless link from RRHs in the s^{th} slice to the i^{th} UE in v^{th} service. In addition, $\mathbf{w}_{R_s,u(v,i)} \in \mathbb{C}^{R_s}$ depicts the transmit beamforming vector from RRHs in the s^{th} slice to the i^{th} UE in v^{th} service. Moreover, BN_0 denotes the power of Gaussian additive noise, and $I_{u(v,i)}$ is the power of interfering signals. Moreover, $a_{v,s} \in \{0, 1\}$ is a binary variable that illustrates whether slice s is connected to service v or not. If $a_{v,s} = 1$ then, v^{th} service is connected to s^{th} slice; otherwise, it is not connected.

To obtain SNR as formulated in (2), let $\mathbf{y}_{U_v} \in \mathbb{C}^{U_v}$ be the received signal's vector of all users in v^{th} service which is given by (3)

$$\mathbf{y}_{U_v} = \sum_{s=1}^S \sum_{k=1}^{K_s} \mathbf{H}_{R_s, \mathcal{U}_v}^H (\mathbf{W}_{R_s, \mathcal{U}_v} \mathbf{P}_{U_v}^{\frac{1}{2}} \mathbf{x}_{U_v} + \mathbf{q}_{R_s}) \zeta_{U_v, k, s} a_{v, s} + \mathbf{z}_{U_v}; \quad (3)$$

Where $\mathbf{x}_{U_v} = [x_{u(v,1)}, \dots, x_{u(v, U_v)}]^T \in \mathbb{C}^{R_s}$ depicts the transmitted symbol vector of UEs in v^{th} set of service, \mathbf{z}_{U_v} is the additive Gaussian noise $\mathbf{z}_{U_v} \sim \mathcal{N}(0, N_0 \mathbf{I}_{U_v})$ and N_0 is the noise power. In addition, $\mathbf{q}_{R_s} \in \mathbb{C}^{R_s}$ indicates the quantization noise which, is made from signal compression in BBU. Besides, $\mathbf{P}_{U_v} = \text{diag}(p_{u(v,1)}, \dots, p_{u(v, U_v)})$.

Furthermore, $\zeta_{U_v, k, s} \triangleq \{\zeta_{u(v,1), k, s}, \zeta_{u(v,2), k, s}, \dots, \zeta_{u(v, U_v), k, s}\}$, $\zeta_{u(v,i), k, s} \in \{0, 1\}$ is a binary, demonstrates whether i^{th} UE in v^{th} service can transmit its signals through k^{th} PRB and also this PRB belongs to s^{th} slice or not. $\mathbf{H}_{R_s, \mathcal{U}_v} = [\mathbf{h}_{R_s, u(v,1)}, \dots, \mathbf{h}_{R_s, u(v, U_v)}]^T \in \mathbb{C}^{R_s \times U_v}$ shows the channel matrix between RRH set R_s to UE set \mathcal{U}_v , besides. What's more, it is assumed we have perfect channel state information (CSI).

Moreover, $\mathbf{W}_{R_s, \mathcal{U}_v} = [\mathbf{w}_{R_s, u(v,1)}, \dots, \mathbf{w}_{R_s, u(v, U_v)}] \in \mathbb{C}^{R_s \times U_v}$ is the zero forcing beamforming vector to minimize the interference which is indicated as below

$$\mathbf{W}_{R_s, \mathcal{U}_v} = \mathbf{H}_{R_s, \mathcal{U}_v} (\mathbf{H}_{R_s, \mathcal{U}_v}^H \mathbf{H}_{R_s, \mathcal{U}_v})^{-1}. \quad (4)$$

Hence, the interference power of i^{th} UE in v^{th} service can be represented as follow

$$\begin{aligned}
I_{u(v,i)} = & \underbrace{\sum_{s=1}^S \sum_{n=1}^S \sum_{\substack{l=1 \\ l \neq i}}^{U_v} \gamma_1 p_{u(v,l)} a_{v,s} \zeta_{u(v,i),n,s} \zeta_{u(v,l),n,s}}_{\text{(intra-service interference)}} \\
& + \underbrace{\sum_{\substack{y=1 \\ y \neq v}}^V \sum_{s=1}^S \sum_{n=1}^S \sum_{l=1}^{U_y} \gamma_2 p_{u(y,l)} a_{y,s} \zeta_{u(v,i),n,s} \zeta_{u(y,l),n,s}}_{\text{(inter-service interference)}} \quad (5) \\
& + \underbrace{\sum_{s=1}^S \sum_{j=1}^{R_s} \sigma_{q_{r(s,j)}}^2 |h_{r(s,j),u(v,i)}|^2 a_{v,s}}_{\text{(quantization noise interference)}}
\end{aligned}$$

Where, $\gamma_1 = |h_{\mathcal{R}_{s,u(v,i)}}^H \mathbf{w}_{\mathcal{R}_{s,u(v,i)}}|^2$ and $\gamma_2 = |h_{\mathcal{R}_{s,u(v,i)}}^H \mathbf{w}_{\mathcal{R}_{s,u(y,l)}}|^2$. Moreover, $\sigma_{q_{r(s,j)}}$ is the variance of quantization noise of j^{th} RRH in s^{th} slice. As it is clear, Interference signal for each UE is coming from UEs using the same PRB. If we replace $p_{u(v,l)}$ and $p_{u(y,l)}$ by P_{max} , an upper bound $\bar{I}_{u(v,i)}$ is obtained for $I_{u(v,i)}$. Therefore, $\bar{\mathcal{R}}_{u(v,i)} \forall v, \forall i$ is derived by using $\bar{I}_{u(v,i)}$ instead of $I_{u(v,i)}$ in (1) and (2).

let $\bar{p}_{r(s,j)}$ denote the power of transmitted signal from j^{th} RRH in s^{th} slice. from (3) we have,

$$\bar{p}_{r(s,j)} = \sum_{v=1}^V \mathbf{w}_{r(s,j),\mathcal{U}_v} \mathbf{P}_{\mathcal{U}_v}^{\frac{1}{2}} \mathbf{P}_{\mathcal{U}_v}^{H\frac{1}{2}} \mathbf{w}_{r(s,j),\mathcal{U}_v}^H a_{v,s} + \sigma_{q_{r(s,j)}}^2. \quad (6)$$

As a result the rate of users on the fronthaul link between BBU-Pool and the j^{th} RRH in s^{th} slice is formulated as

$$C_{R(s,j)} = \log \left(1 + \sum_{v=1}^V \frac{w_{r(s,j),\mathcal{D}_s} \mathbf{P}_{\mathcal{U}_v}^{\frac{1}{2}} \mathbf{P}_{\mathcal{U}_v}^{H\frac{1}{2}} w_{r(s,j),\mathcal{U}_v}^H a_{v,s}}{\sigma_{q_{r(s,j)}}^2} \right), \quad (7)$$

Where, $a_{v,s}$ is a binary variable denotes whether the slice s is connected to service v or not [5], [11].

C. Mean Delay

Let the packet arrival of UEs have a Poisson Process with arrival rate $\lambda_{u(v,i)}$ for i^{th} UE of v^{th} service. Therefore, the mean arrival data rate of UEs connect to s^{th} slice in the first layer is $\alpha_{s1} = \sum_{v=1}^V \sum_{u=2}^{U_v} a_{v,s} \lambda_{u(v,i)}$. Furthermore, the mean arrival data rate of the second layer is approximately equal to the mean arrival data rate of first layer $\alpha_s = \alpha_{s1} \approx \alpha_{s2}$ since, by using Burkes Theorem, the mean arrival data rate of the second layer which is processed in the first layer is still Poisson with rate α_s . It is assumed there are dispatchers in each layer for each slice to divide the incoming traffic to VNFs [4], [9], [10]. Suppose the baseband processing of each VNF is depicted as a M/M/1 processing queue. Each packet is routed by one of VNFs of slices. So the mean delay of slice s which is related

to incoming traffic rate routed to each VNF in the first layer can be written as follow

$$d_{s1} = \frac{1}{\mu_1 - \alpha_s / M_{s1}}. \quad (8)$$

Also, the delay of s^{th} slice in the second layer is

$$d_{s2} = \frac{1}{\mu_2 - \alpha_s / M_{s2}}. \quad (9)$$

In addition, d_{str} is the transmission delay for s^{th} slice as a result of wireless transmission. The arrival data rate of wireless transmission is equal to the arrival data rate of dispatchers (used as a traffic divider) for each slice divide traffic data and transmit divided traffic to VNFs. Moreover, it is assumed that the service time of transmission queue for each slice s has an exponential distribution with mean $1/(R_{tot_s})$ and can be modeled as a M/M/1 queue [4], [9], [10]. Therefore, the mean delay of the transmission layer is

$$d_{str} = \frac{1}{R_{tot_s} - \alpha_s}; \quad (10)$$

Where, $R_{tot_s} = \sum_{v=1}^V \sum_{u=2}^{U_v} a_{v,s} R_{u(v,i)}$. Mean delay of each slice is

$$D_s = d_{s1} + d_{s2} + d_{str} \forall s. \quad (11)$$

D. Physical Data Center Resource

Each VNF requires physical resources that contain RAM, Memory, and CPU. Let, the required resources for VNF f in slice s is represented by a three-dimensional vector as follow

$$\bar{\Omega}_{(f,s)} = \{\Omega_{R_{f,s}}, \Omega_{M_{f,s}}, \Omega_{C_{f,s}}\}; \quad (12)$$

Where, $\bar{\Omega}_{(f,s)} \in \mathbb{C}^3$ and $\Omega_{R_{f,s}}, \Omega_{M_{f,s}}, \Omega_{C_{f,s}}$ indicate the amount of required RAM, Memory and CPU, respectively. Moreover total amount of required RAM, Memory and CPU of all VNFs of a slice is a three dimension vector which is defined as

$$\bar{\Omega}_s^{tot} = \sum_{f=1}^{M_{s1}+M_{s2}} \bar{\Omega}_{(f,s)}; \quad (13)$$

Also, there are D_c data centers (DC), which served VNFs. Each DC contains several servers that supply VNF's requirements. The amount of RAM, Memory and CPU is denoted by τ_{R_j}, τ_{M_j} and τ_{C_j} for j^{th} DC, respectively.

$$\tau_j = \{\tau_{R_j}, \tau_{M_j}, \tau_{C_j}\};$$

Also we define a weighted parameter of τ_j as follow

$$\begin{aligned}
\hat{\Omega}_s^{tot} &= w_R \bar{\Omega}_{R_s}^{tot} + w_M \bar{\Omega}_{M_s}^{tot} + w_C \bar{\Omega}_{C_s}^{tot} \\
\hat{\tau}_j &= w_R \tau_{R_j} + w_M \tau_{M_j} + w_C \tau_{C_j},
\end{aligned} \quad (14)$$

Where, $\mathbf{w} = \{w_R, w_M, w_C\}$ are the weight of RAM, Memory and CPU that is used for Algorithm 3. In this system model the placement of physical data center resources to VNFs is considered. $y_{s,d}$ is a binary variable indicates whether d^{th} data-center is connected to VNFs of s^{th} slice or not.

E. Problem Statement

One of the most important parameters to estimate the optimality of the system is energy efficiency which is represented as a sum-rate to sum-power

$$\eta(\mathbf{P}, \mathbf{A}) := \frac{\sum_{v=1}^V \sum_{k=1}^{U_v} \mathcal{R}_{u(v,k)}}{\sum_{s=1}^S \sum_{i=1}^{R_s} \bar{p}_{r(s,i)}} = \frac{\mathfrak{R}_{tot}(\mathbf{P}, \mathbf{A})}{P_{r_{tot}}(\mathbf{P}, \mathbf{A})}, \quad (15)$$

Assume the power consumption of baseband processing at each data center that is connected to VNFs of a slice is depicted as ϕ . So the total power of system for all active DCs that are connected to slices can be represented as

$$\phi_{tot} = \sum_{s=1}^S \sum_{d=1}^{D_c} y_{s,d} \phi.$$

In this paper, the main goal is to simultaneously maximize sum-rate and minimize sum-power which is the total power of all RRHs and the total power consumption of baseband processing at all DCs with the presence of constraints which is written as follow,

$$\max_{\mathbf{P}, \mathbf{A}, \mathbf{Y}} \quad \eta(\mathbf{P}, \mathbf{A}) + \frac{1}{\phi_{tot}(\mathbf{Y})} \quad (16a)$$

$$\text{subject to} \quad \bar{p}_{r(s,i)} \leq P_{max} \quad \forall s, \forall i, \quad (16b)$$

$$p_{u(v,k)} \geq 0 \quad \forall v, \forall k, \quad (16c)$$

$$\mathcal{R}_{u(v,k)} \geq \mathcal{R}_{u(v,k)}^{min} \quad \forall v, \forall k, \quad (16d)$$

$$C_{r(s,i)} \leq C_{r(s,i)}^{max} \quad \forall s, \forall i, \quad (16e)$$

$$D_s \leq D_s^{max} \quad \forall s, \quad (16f)$$

$$\sum_{s=1}^S a_{v,s} \geq 1 \quad \forall s, \quad (16g)$$

$$\sum_{d=1}^{D_c} \sum_{v=1}^V y_{s,d} a_{v,s} \geq 1 \times \sum_{v=1}^V a_{v,s} \quad \forall s, \quad (16h)$$

$$\bar{\Omega}_{\mathfrak{z}(s)}^{tot} = \sum_{f=1}^{F_s} \bar{\Omega}_{\mathfrak{z}(f,s)} \leq \sum_{d=1}^{D_c} y_{s,d} \tau_{\mathfrak{z}d} \quad \forall \mathfrak{z}, \forall s \quad (16i)$$

Where, $\mathbf{P} = [p_{u(v,k)}] \forall v, \forall k$, is the matrix of power for UEs, $\mathbf{A} = [a_{v,s}] \forall v, \forall s$ denotes the binary variable for connecting slices to services and $\mathbf{Y} = [y_{s,d}] \forall s, \forall d$ is a binary variable shown whether the a physical DC is connected to a VNFs of a slice or not. (16b) and (16c) indicate that the power of each RRH do not exceed the maximum power and power of each UE is a positive integer value, respectively. Also (16d) shows that the rate of each UE is more than a threshold. (16e) and (16f) depicts that the capacity of fronthaul link is limited and the delay of receiving signal should be less than a threshold, respectively. Furthermore, (16g) ensures that each service is connected to one or more slices. Also, (16h)

guarantees that each slice (VNFs in two layers of slices) has been placed to one or more physical resources of DCs. Moreover, (16i) $\mathfrak{z} \in \{M, R, C\}$, supports that we have enough physical resource for VNFs of each slice.

The main optimization problem which is formulated as (16) can be decomposed into two independent optimization problem A and B since the variables can be obtained independently. The problem A is defined as

$$\max_{\mathbf{P}, \mathbf{A}} \quad \eta(\mathbf{P}, \mathbf{A}) \quad (17a)$$

$$\text{subject to} \quad \bar{p}_{r(s,i)} \leq P_{max} \quad \forall s, \forall i, \quad (17b)$$

$$p_{u(v,k)} \geq 0 \quad \forall v, \forall k, \quad (17c)$$

$$\mathcal{R}_{u(v,k)} \geq \mathcal{R}_{u(v,k)}^{min} \quad \forall v, \forall k, \quad (17d)$$

$$C_{r(s,i)} \leq C_{r(s,i)}^{max} \quad \forall s, \forall i, \quad (17e)$$

$$D_s \leq D_s^{max} \quad \forall s, \quad (17f)$$

$$\sum_{s=1}^S a_{v,s} \geq 1 \quad \forall s \quad (17g)$$

and the problem B is

$$\min_{\mathbf{Y}} \quad \phi_{tot}(\mathbf{Y}) \quad (18a)$$

$$\text{subject to} \quad \sum_{d=1}^{D_c} \sum_{v=1}^V y_{s,d} a_{v,s} \geq 1 \times \sum_{v=1}^V a_{v,s} \quad \forall s, \quad (18b)$$

$$\bar{\Omega}_{\mathfrak{z}(s)}^{tot} = \sum_{f=1}^{F_s} \bar{\Omega}_{\mathfrak{z}(f,s)} \leq \sum_{d=1}^{D_c} y_{s,d} \tau_{\mathfrak{z}d} \quad \forall \mathfrak{z}, \forall s \quad (18c)$$

III. PROPOSED METHOD FOR PROBLEM (17)

In this subsection, the proposed method is applied to solve the optimization problem. We want to solve (17). Since the problem is non-convex and NP-Hard, iterative algorithm is applied. To solve the problem and obtain \mathbf{A} and \mathbf{P} we divide problem (17) into two different part that can be solved iteratively.

A. First Part of Sub-Problem A

Firstly, we need to obtain \mathbf{A} by fixing \mathbf{P} in the problem (17) and updating this parameter at the end of each iteration. Two different method is applied to acquire \mathbf{A} . The details of heuristic algorithm are represented in **Algorithm** (1).

B. Second Part of Sub-Problem A

In this part, by assuming that \mathbf{A} is fixed, the optimal power of UEs in each service is achieved.

Theorem 1. η^* which is the optimum energy efficiency can be achieved if

$$\begin{aligned} \max_{\mathbf{P}} (\mathfrak{R}_{tot}(\mathbf{P}) - \eta^* P_{r_{tot}}(\mathbf{P})) = \\ \mathfrak{R}_{tot}(\mathbf{P}^*) - \eta^* P_{r_{tot}}(\mathbf{P}^*) = 0. \end{aligned} \quad (19)$$

Proof. See [14], Appendix A □

Algorithm 1 Mapping Slice to Service

- 1: Sort services according to the number of UEs in it and their requirements in descending order.
 - 2: Sort slices according to the weighted linear combination of number of PRBs, RRHs and VNFs in two layers and the Capacity of their resources in descending order.
 - 3: **for** $i \leftarrow 1$ to S **do**
 - 4: **for** $j \leftarrow 1$ to V **do**
 - 5: Set $a_{i,j} = 1$
 - 6: Obtain Parameters of Systems (power and rate of UEs, rate of fronthaul links, power of RRHs)
 - 7: **if** conditions (16b), (16c), (16d) and (16e) is not applied **then**
 - 8: Set $a_{i,j} = 0$;
 - 9: **else**
 - 10: break from inner loop;
 - 11: **end if**
 - 12: **end for**
 - 13: **end for**
-

The second subproblem can be solved using Lagrangian function and iterative algorithm. Since, Interference is a function of power of UEs, for simplicity, we assume an upper bound $\bar{I}_{u(v,i)}$ for interference (the worst-case). In order to make (17) as a standard form of convex optimization problem, it is required to change the variable of equations (17e) and (17f). Lagrangian function is written as follow

$$\begin{aligned} \mathcal{L}(P; \lambda, \mu, \xi, \kappa) = & \sum_{v=1}^V \sum_{k=1}^{U_v} \bar{\mathcal{R}}_{u(v,k)} - \eta \sum_{v=1}^V \sum_{i=1}^{\mathcal{R}_s} \bar{p}_{r(s,i)} \\ & + \sum_{s=1}^S \sum_{k=1}^{U_v} \lambda_{u(v,k)} (\bar{\mathcal{R}}_{d(s,k)} - \mathcal{R}_{u(v,k)}^{max}) \\ & - \sum_{s=1}^S \sum_{i=1}^{\mathcal{R}_s} \mu_{r(s,i)} (\bar{p}_{r(s,i)} - P_{max}) \\ & - \sum_{s=1}^S \sum_{i=1}^{\mathcal{R}_s} \xi_{r(s,i)} (\bar{p}_{r(s,i)} - \sigma_{q_{r(s,i)}}^2 2^{C_{r(s,i)}^{max}}). \\ & + \sum_{v=1}^V \sum_{k=1}^{U_v} \kappa_{u(v,k)} \sum_{s=1}^S (R_{u(v,k)} - \mathcal{D}_s) a_{v,s}. \end{aligned} \quad (20)$$

Where, $\mathcal{D}_s = \frac{1}{D_{s1}^{max} - d_{s1} - d_{s2}} + \alpha_s$. Also, λ, μ, ξ and κ are the matrix of Lagrangian multipliers which have non-zero positive elements. Optimal power is obtained from equation (20) as follow

$$p_{u(v,i)}^* = \frac{\eta_{u(v,i)} \mathfrak{w}_{u(v,i)} - \mathfrak{r}_{u(v,i)} \mathfrak{z}_{u(v,i)}}{\mathfrak{r}_{u(v,i)} \mathfrak{w}_{u(v,i)}} \quad (21)$$

where, $\eta_{u(v,i)} = (\lambda_{u(v,i)} + \kappa_{u(v,k)} + 1) \frac{B}{Ln_2}$ and $\mathfrak{w}_{u(v,i)} = \sum_{s=1}^S |\mathbf{h}_{R_s, u(v,i)}^H \mathbf{w}_{R_s, u(v,i)}|^2 a_{v,s}$. Also $\mathfrak{z}_{u(v,i)} = BN_0 + \bar{I}_{u(v,i)}$ and $\mathfrak{r}_{u(v,i)} = \sum_{s=1}^S \sum_{i=1}^{\mathcal{R}_s} (\mu_{r(s,i)} + \xi_{r(s,i)} + \eta) \|w_{r(s,i), u(v,i)}\|^2$. By using sub-gradient method, the optimal power \mathbf{P} is obtained.

C. Solving two part of Sub-problem A iteratively

In (III-A) and (III-B) the details of solving each part of subproblem is depicted. Firstly, by fixing power of each UE, \mathbf{A} is obtained using Algorithm (1). Then by using sub-gradient method, (III-B) is solved and iteratively solve these problems until variables converge. Here, the algorithm of solving sub-problem A is shown in **Algorithm (2)**

Algorithm 2 Joint Network Slicing and Power Allocation

- 1: Set the maximum number of iterations I_{max} , convergence condition ϵ_η and the initial value $\eta^{(1)} = 0$
 - 2: Set $\mathbf{P} = \mathbf{P}_{max}$
 - 3: **for** $counter \leftarrow 1$ to I_{max} **do**
 - 4: Achieve \mathbf{A} by applying Algorithm (1)
 - 5: Obtain \mathbf{P} by using sub-gradient method which is mentioned in (III-B).
 - 6: **if** $\Re_{tot}(\mathbf{P}^{(i)}, \mathbf{A}^{(i)}) - \eta^{(i)} P_{r_{tot}}(\mathbf{P}^{(i)}, \mathbf{A}^{(i)}) < \epsilon_\eta$ **then**
 - 7: Set $\mathbf{P}^* = \mathbf{P}^{(i)}$, $\mathbf{A}^* = \mathbf{A}^{(i)}$ and $\eta^* = \eta^{(i)}$;
 - 8: break;
 - 9: **else**
 - 10: $i = i + 1$, Setting $\mathbf{P} = \mathbf{P}^{(i)}$;
 - 11: **end if**
 - 12: **end for**
-

D. Sub-Problem B

In this subsection, we want to solve (18) which is the placement of virtual resources to physical resource in order to minimize the number of using DCs. To achieve optimum \mathbf{Y} heuristic algorithm and Mosek is applied. The details of heuristic algorithm is written in **Algorithm (3)**. In this algorithm, firstly we sort slices and DCs according to their sum-weighted of their requirements. Secondly, we start mapping from the most needed slices to the DC with the most physical resources. After mapping DCs to slices, if some slices are not admitted, we start admitting residual slices to more than one DCs. At the end, if DC with the lowest physical resources is free and can served instead of DC with highest physical resource, the slices remapped to new DC with lowest physical resource since it has lowest power consumption. if a slice does not admit to a specific DC, it is remain for next placement. In next placement the residual slices, map to more than one DC according to their requirements.

IV. SIMULATION

In this section, Simulation and numerical results for the main problem are depicted. In Fig. 2, the ratio of admitted slices is demonstrated for two different number of DCs with different number of slices (the parameters for simulation listed in table I and II). In this simulation, it is assumed that each slice can be served by just one DC and it is not admitted by more than one DC. Proposed method is based on Algorithm 3 and optimal method is done by MOSEK toolbox. When we have two DCs, proposed method and optimal method have approximately same ratio of admitted slices. But by increasing

Algorithm 3 Plcement of Physical resources into Virtual resources

```

1: Sort Slices according to  $\hat{\Omega}_s^{tot}, \forall s$  in descending order.
2: Sort DCs according to  $\hat{\tau}_j, \forall j$  in descending order.
3:  $\mathbf{Y} = \mathbf{0}$ 
4: for  $d \leftarrow 1$  to  $D_c$  do
5:   for  $s \leftarrow 1$  to  $S$  do
6:     if  $\sum_{d=1}^{D_c} y_{s,d} == 0$  and  $\bar{\Omega}_{\hat{s}(s)}^{tot} \leq \tau_{\hat{s}(s)} \forall \hat{s}, \forall s$  then
7:       Set  $y_{s,d} = 1$ ;
8:        $\tau_j \leftarrow \tau_j - \bar{\Omega}_s^{tot}$ 
9:     end if
10:   end for
11: end for
12:  $\{ind_{residual} = s | (\sum_{d=1}^{D_c} y_{s,d} == 0)\}$ 
13: Sort residual amount of DCs same as before in descending order.
14: Sort residual slices same as before in descending order.
15: for  $r \leftarrow 1$  to  $S_{residual}$  do
16:   for  $n \leftarrow 1$  to  $D_c$  do
17:     Set  $y_{s,d} = 1$ ;
18:      $\bar{\Omega}_s^{tot} \leftarrow \bar{\Omega}_s^{tot} - \tau_j$ 
19:     if  $\bar{\Omega}_s^{tot} == 0$  then
20:       Set  $y_{s,d} = 1$ ;
21:        $\tau_j \leftarrow \tau_j - \bar{\Omega}_s^{tot}$ 
22:       break inner loop
23:     end if
24:   end for
25: end for
26: Remapping DCs must be done to prevent wasting Energy

```

TABLE I: Simulation Parameter

Parameter	Value
Mean of CPU for DCs	25.6GHz
Mean of RAM for DCs	128G
Mean of Memory for DCs	10T
Mean of CPU for Slices	3.2GHz
Mean of RAM for Slices	16G
Mean of Memory for Slices	1T

the number of DCs to five, performance of proposed method reduced. Using five DCs, the difference between proposed method and optimal method in worst case (44 slices) is about 23 percentage.

TABLE II: Simulation Parameter

Parameter	Value
w_C	320
w_R	64
w_M	1

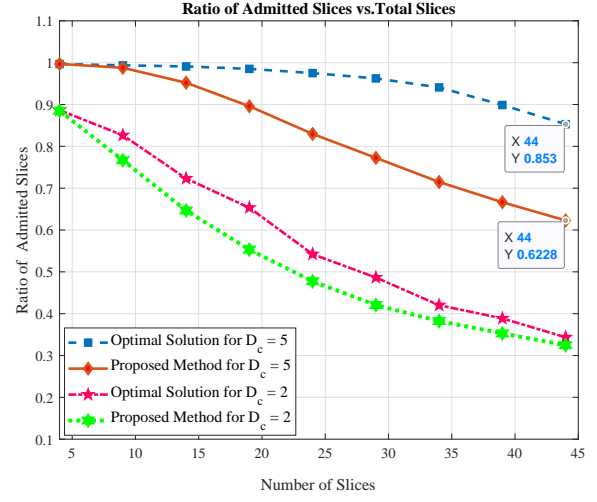


Fig. 2: Ratio of Admitted Slices connected to just one DC vs. Total slices

In Fig. 3, the normalized resource consumption is depicted due to the number of slices (the parameters for simulation listed in table I and II). In this simulation, it is assumed that number of DCs is completely enough to cover all slices. The optimality of placement of DCs to slices is measured. It is shown that how much resources of active DCs are not used. For ten slices, the difference between optimal solution and proposed solution is about 15 percent.

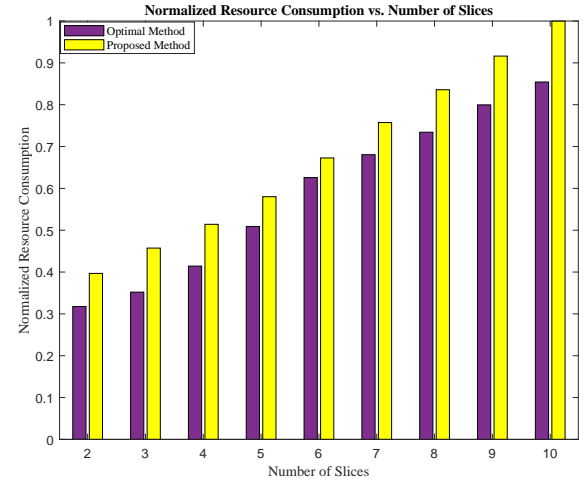


Fig. 3: Normalized Resource Consumption vs. Number of Slices

REFERENCES

- [1] I. Scales. (2018) The open ran (oran) alliance formed to lever open 5g for other technologies? and much more... [Online]. Available: <https://www.telecomtv.com/content/mwc/the-open-ran-oran-alliance-formed-to-lever-open-5g-for-other-technologies-and-much-more-16429>
- [2] B. Fletcher. (2019) Vodafone initiates first open ran trials in the u.k., challenging traditional vendors. [Online]. Available: <https://www.fiercewireless.com/tech/vodafone-initiates-first-open-ran-trials-uk-challenging-traditional-vendors>

- [3] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (c-ran): a primer," *IEEE Network*, vol. 29, no. 1, pp. 35–41, 2015.
- [4] J. Tang, W. P. Tay, T. Q. Quek, and B. Liang, "System cost minimization in cloud ran with limited fronthaul capacity," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3371–3384, 2017.
- [5] O. Simeone, J. Kang, J. Kang, and S. Shamai, "Cloud radio access networks: Uplink channel estimation and downlink precoding," *arXiv preprint arXiv:1608.07358*, 2016.
- [6] M. K. Motalleb, A. Kabiri, and M. J. Emadi, "Optimal power allocation for distributed mimo c-ran system with limited fronthaul capacity," in *2017 Iranian Conference on Electrical Engineering (ICEE)*. IEEE, 2017, pp. 1978–1982.
- [7] (2018) xran forum merges with c-ran alliance to form oran alliance. [Online]. Available: <https://www.businesswire.com/news/home/20180227005673/en/xRAN-Forum-Merges-C-RAN-Alliance-Form-ORAN>
- [8] X. Zhou, R. Li, T. Chen, and H. Zhang, "Network slicing as a service: enabling enterprises' own software-defined cellular networks," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 146–153, 2016.
- [9] P. Luong, C. Despins, F. Gagnon, and L.-N. Tran, "A novel energy-efficient resource allocation approach in limited fronthaul virtualized c-rans," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, 2018, pp. 1–6.
- [10] P. Luong, F. Gagnon, C. Despins, and L.-N. Tran, "Joint virtual computing and radio resource allocation in limited fronthaul green c-rans," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2602–2617, 2018.
- [11] S.-H. Park, O. Simeone, O. Sahin, and S. S. Shitz, "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 69–79, 2014.
- [12] Y. L. Lee, J. Loo, T. C. Chuah, and L.-C. Wang, "Dynamic network slicing for multitenant heterogeneous cloud radio access networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2146–2161, 2018.
- [13] P.-R. Li, T.-S. Chang, and K.-T. Feng, "Energy-efficient power allocation for distributed large-scale mimo cloud radio access networks," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2014, pp. 1856–1861.
- [14] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation for secure ofdma systems," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 6, pp. 2572–2585, 2012.