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 global 'carrier-led' effort to bring greater transparency of Open-RAN (O-RAN), to the next-generation of radio access network.**
To increase energy efficiency and optimize allocation of resources, Network Slicing (NS) is considered as a best method in 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 elaborates joint NS in RAN and Core of O-RAN system and investigate optimal power of each User Equipment (UE) to jointly maximize Energy Efficiency and minimize consumption power of physical resources in a downlink channel. The problem is formulated as a mixed integer optimization problem that can be decompose into two independent sub-problems for RAN and Core since sub-problems are independent. Heuristic algorithms is proposed to each of sub-problems to solved the first sub-problem in order to simultaneously map slices to services and optimize power and the second sub-problem in order to map slices to physical resources and minimize number of active Data Centers (DC).

***Index Terms*—O-RAN, Network Slicing, Energy Efficiency, Data Center**

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 challenges in a best way. The Idea of O-RAN comes from two opinion. Firstly, according to real-time analytic used for artificial intelligence system, the radio access networks must be evolved to be more open and smarter than previous generations. Furthermore, O-RAN can virtualize elements of the network with appropriate interfaces [1].
The core idea driving C-RAN is to split radio remote head (RRH) and 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 [2]. To communicate between BBU-Pool and RRHs, fronthaul interface which is fiber link is assumed with limited capacity. Compression of message which is passed through these links, is a consequence of limited fronthaul capacity [3]–[5].
xRAN technology ,released in April 2018 as a next generation of RAN, has three fundamental features. Control plane is decoupled from User plane. In addition, a modular eNB software stack is built to operate on common-off-the-shelf (COTS) hardware. Moreover, open north-bound and south-bound interfaces is introduced [6].

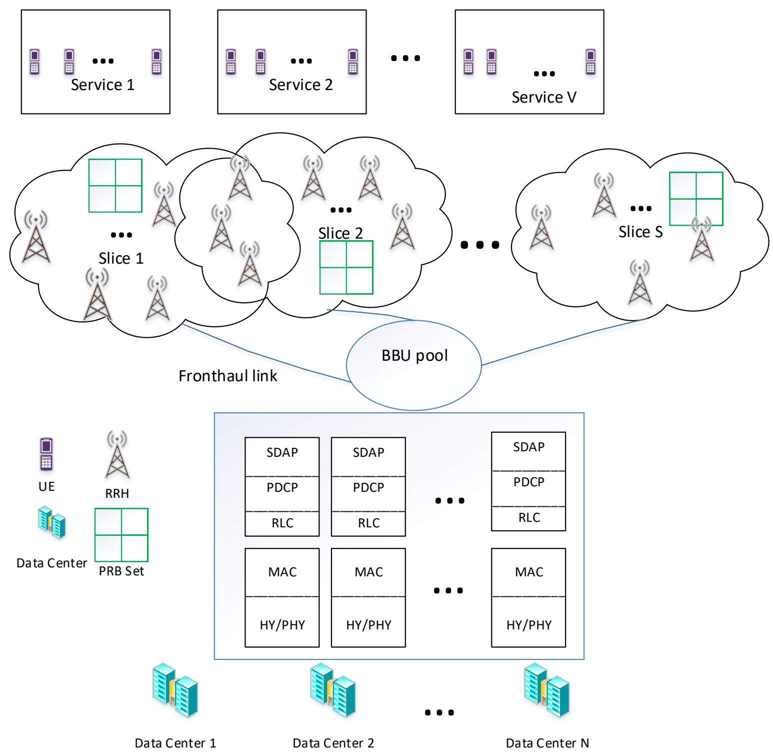

Fig. 1: Network sliced O-RAN system

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 to blocks of virtual network function (VNF). Responsibility of wireless systems of the fifth generation covers a wide range types of services. In order to provide the requirement of these services, NS is implemented to virtualize the common physical network into several logical end-to-end networks. Three different types of NS is introduced in [7] 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 to group of services according to their requirements. In addition, each service is connected to one or more Core-RAN slices base on the resource of slices. Using cloud-computing in BBU-Pool, enhance performance of system by virtualizing resources into virtual machines (VMs). Each VM has a computing processor mapped to Virtual Network Functions (VNFs) which processed arrival data. Also VNFs are mapped to physical resource through NS technique [3], [8], [9].

In [10], dynamic network slicing is considered in Heterogeneous CRAN (H-CRAN) to maximize the weighted sum rate. Moreover, in [3], [8], [9] minimization of cost obtained by power consumption and cloud processing is applied. Limited fronthaul capacity is considered. Also processing delay of each VM and wireless transmission delay is bounded.

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

II. System Model and Problem Formulation

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

A. System Model

Suppose that there are S slices Serving V services. Each Service $v \in \{1, 2, ..., V\}$, consists of U_v single antenna users that require certain service. Each slice $s \in \{1, 2, ..., S\}$ consists of R_s RRHs and K_s Physical Resource Blocks (PRBs) as it is defined in standard of LTE that the bandwidth (BW) of channel is divided to PRBs in time series [10]. All RRHs in a slice that is mapped to a service, transmit signals cooperatively to all the UEs in specific service [5], [11]. Each RRH $r \in \{1, 2, ..., R\}$ is connected to BBU pool via an optical fiber link with limited fronthaul capacity. Suppose we have two processing layer in 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 second layer. For simplicity, suppose each VM is connected to a VNF. Each VNF in both layers connects to one or more slices. So in s^{th} slice, there are M_{s_1} VNFs in first layer and M_{s_2} VNFs in second layer. All VNFs in first and second layer has 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

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

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

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

Where, $p_{u(v,i)}$ represents the transmition 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 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 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 Guassian additive noise and $I_{u(v,i)}$ is the power of interfering signals. 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)

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

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

Furthermore, $\zeta_{U_v,k,s} \triangleq \{\zeta_{u(v,1),k,s}, \zeta_{u(v,2),k,s}, ..., \zeta_{u(v,N_{U_v}),k,s}\}$, $\zeta_{u(v,i),k,s} \in \{0, 1\}$ is a binary parameter, 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. $\boldsymbol{H}_{\mathcal{R}_s,\mathcal{U}_v} = [\boldsymbol{h}_{\mathcal{R}_s,u_{(v,1)}}, \ldots, \boldsymbol{h}_{\mathcal{R}_s,v_{(v,\mathcal{U}_v)}}]^T \in \mathbb{C}^{R_s \times U_v}$ shows the channel matrix between RRH set $\mathcal{R}_s$ to UE set $\mathcal{U}_v$, besides. What's more, it is assumed we have perfect channel state information(CSI).

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

$$\boldsymbol{W}_{\mathcal{R}_s,\mathcal{U}_v} = \boldsymbol{H}_{\mathcal{R}_s,\mathcal{U}_v} (\boldsymbol{H}^H_{\mathcal{R}_s,\mathcal{U}_v} \boldsymbol{H}_{\mathcal{R}_s,\mathcal{U}_v})^{-1}. \qquad (4)$$

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

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

Where, $\gamma_1 = |\boldsymbol{h}^H_{\mathcal{R}_s,u_{(v,i)}} \boldsymbol{w}_{\mathcal{R}_s,u_{(v,l)}}|^2$ and $\gamma_2 = |\boldsymbol{h}^H_{\mathcal{R}_s,u_{(v,i)}} \boldsymbol{w}_{\mathcal{R}_s,u_{(y,l)}}|^2$. As it is clear, Interference signal

for each UE is comming 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,

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

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

$$C_{R_{(s,j)}} = \log\big(1 + \sum_{v=1}^{V} \frac{w_{r_{(s,j)},\mathcal{D}_s} \boldsymbol{P}_{\mathcal{U}_v}^{\frac{1}{2}} \boldsymbol{P}_{\mathcal{U}_v}^{H\frac{1}{2}} w_{r_{(s,j)},\mathcal{U}_v}^{H} a_{v,s}}{\sigma_{q_{r(s,j)}}^2}\big), \tag{7}$$

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 $\alpha_{s_1} = \sum_{v=1}^{V} \sum_{u=2}^{U_v} a_{v,s} \lambda_{u(v,i)}$. Furthermore, the mean arrival rate of second layer is approximately equal to the mean arrival rate of first layer $\alpha_s = \alpha_{s_1} \approx \alpha_{s_2}$ since, by using Burkes Theorem, the arrival packets of second layer which is processed in 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 [3], [8], [9]. 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 first layer can be written as follow

$$d_{s_1} = \frac{1}{\mu_1 - \alpha_s / M_{s_1}}. \tag{8}$$

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

$$d_{s_2} = \frac{1}{\mu_2 - \alpha_s / M_{s_2}}. \tag{9}$$

In addition, $d_{s_{tr}}$ is another 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 for each slice that divide data rates and transmit divided rates 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. Therefore, the mean delay of transmission layer is

$$d_{s_{tr}} = \frac{1}{R_{tot_s} - \alpha_s}; \tag{10}$$

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

$$D_s = d_{s_1} + d_{s_2} + d_{s_{tr}} \forall s. \tag{11}$$

D. Physical Data Center Resource

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

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

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

$$\bar{\Omega}_s^{tot} = \sum_{f=1}^{M_{s_1}+M_{s_2}} \bar{\Omega}_{(f,s)}; \tag{13}$$

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 $\tau_{R_j}, \tau_{M_j}$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

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

Where, $\boldsymbol{w} = \{w_R, w_M, w_C\}$ 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

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

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 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 with the presence of constraints which is written as follow,

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

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

$$p_{u_{(v,k)}} \geq 0 \qquad \forall v, \forall k, \tag{16c}$$

$$\mathcal{R}_{u_{(v,k)}} \geq \mathcal{R}^{min}_{u_{(v,k)}} \qquad \forall v, \forall k, \tag{16d}$$

$$C_{r_{(s,i)}} \leq C^{max}_{r_{(s,i)}} \qquad \forall s, \forall i, \tag{16e}$$

$$D_s \leq D^{max}_s \qquad \forall s, \tag{16f}$$

$$\sum_{s=1}^{S} a_{v,s} \geq 1 \qquad \forall s, \tag{16g}$$

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

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

Where,$\boldsymbol{P} = [p_{u(v,k)}] \forall v, \forall k$, $\boldsymbol{A} = [a_{v,s}] \forall v, \forall s$ and $\boldsymbol{Y} = [y_{s,d}] \forall s, \forall d$. (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 optimizaiton 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

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

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

$$p_{u_{(v,k)}} \geq 0 \qquad \forall v, \forall k, \tag{17c}$$

$$\mathcal{R}_{u_{(v,k)}} \geq \mathcal{R}^{min}_{u_{(v,k)}} \qquad \forall v, \forall k, \tag{17d}$$

$$C_{r_{(s,i)}} \leq C^{max}_{r_{(s,i)}} \qquad \forall s, \forall i, \tag{17e}$$

$$D_s \leq D^{max}_s \qquad \forall s, \tag{17f}$$

$$\sum_{s=1}^{S} a_{v,s} \geq 1 \qquad \forall s \tag{17g}$$

and the problem B is

$$\min_{\boldsymbol{y}} \quad \phi_{tot}(\boldsymbol{Y}) \tag{18a}$$

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

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

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 optimum $\boldsymbol{A}$ and $\boldsymbol{P}$ 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 $\boldsymbol{A}$ by fixing $\boldsymbol{P}$ in the problem (17) and updating this parameter at the end of each iteration. Two different method is applied to acquire $\boldsymbol{A}$. The first method is using MOSEK and second method is a heuristic algorithm. The details of heuristic algorithm are represented in **Algorithm** (1).

Algorithm 1 Mapping Slice to Service

Sort services according to their priority, the number of UEs in it and their requirements in descending order.
Sort slices according to the number of PRBs, RRHs and VNFs in two layers and the Capacity of their resources in descending order.
for $i \leftarrow 1$ to S **do**
    for $j \leftarrow 1$ to V **do**
        Set $a_{i,j} = 1$
        Obtain Parameters of System
        if conditions (16b), (16c), (16d) and (16e) is not applied **then**
            Set $a_{i,j} = 0$;
        else
            break from inner loop;
        end if
    end for
end for

B. Second Part of Sub-Problem A

In this part, by assuming that $\boldsymbol{A}$ 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*

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

Proof. See [12], Appendix A □

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

$$\begin{aligned}\mathcal{L}(\boldsymbol{P};\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\xi},\boldsymbol{\kappa}) = &\sum_{v=1}^{V}\sum_{k=1}^{U_v}\bar{\mathcal{R}}_{u_{(v,k)}} - \eta\sum_{v=1}^{V}\sum_{i=1}^{\mathcal{R}_s}\bar{p}_{r_{(s,i)}} \\ &+ \sum_{s=1}^{S}\sum_{k=1}^{U_v}\lambda_{u_{(v,k)}}(\bar{\mathcal{R}}_{d_{(s,k)}} - \mathcal{R}^{max}_{u_{(v,k)}}) \\ &- \sum_{s=1}^{S}\sum_{i=1}^{R_s}\mu_{r_{(s,i)}}(\bar{p}_{r_{(s,i)}} - P_{max}) \\ &- \sum_{s=1}^{S}\sum_{i=1}^{R_s}\xi_{r_{(s,i)}}(\bar{p}_{r_{(s,i)}} - \sigma^2_{q_{r(s,j)}}2^{C^{max}_{r_{(s,i)}}}). \\ &+ \sum_{v=1}^{V}\sum_{k=1}^{U_v}\kappa_{u_{(v,k)}}\sum_{s=1}^{S}(R_{u_{(v,k)}} - \mathfrak{D}_{\mathfrak{s}})a_{v,s}.\end{aligned} \tag{20}$$

Where,$\mathfrak{D}_{\mathfrak{s}} = \frac{1}{D^{max}_{s} - d_{s_1} - d_{s_2}} + \alpha_s$. Also, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, $\boldsymbol{\xi}$ and $\boldsymbol{\kappa}$ are the matrix of Lagrangian multipliers which have non-zero positive elements. Optimal power is obtained from equation (20) as follow

$$p^*_{u(v,i)} = \frac{\mathfrak{y}_{u(v,i)}\mathfrak{w}_{u(v,i)} - \mathfrak{x}_{u(v,i)}\mathfrak{z}_{u(v,i)}}{\mathfrak{x}_{u(v,i)}\mathfrak{w}_{u(v,i)}} \tag{21}$$

where, $\mathfrak{y}_{u(v,i)} = (-\mu_{u(v,i)} + \lambda_{u(v,i)} + \kappa_{u_{(v,k)}} + 1)\frac{B}{Ln_2}$ and $\mathfrak{w}_{u(v,i)} = \sum_{s=1}^{S}|\mathbf{h}^{H}_{R_s,u(v,i)}\mathbf{w}_{R_s,u(v,i)}|^2 a_{v,s}$. Also $\mathfrak{z}_{u(v,i)} = BN_0 + \bar{I}_{u(v,i)}$ and $\mathfrak{x}_{u(v,i)} = \sum\limits_{s=1}^{S}\sum\limits_{i=1}^{R_s}(\xi_{r_{(s,i)}} + \eta)||w_{r_{(s,j)},u_{(v,i)}}||^2$. By using sub-gradient method, the optimal power $\boldsymbol{P}$ 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, $\boldsymbol{A}$ 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

Set the maximum number of iterations I_{max}, convergence condition ϵ_η and the initial value $\eta^{(1)} = 0$
Set $\boldsymbol{P} = \boldsymbol{P}_{max}$
for $counter \leftarrow 1$ to I_{max} **do**
    Achieve $\boldsymbol{A}$ by applying Algorithm (1)
    Obtain $\boldsymbol{P}$ by using sub-gradient method which is mentioned in (III-B).
    if $\mathfrak{R}_{tot}(\boldsymbol{P}^{(i)}, \boldsymbol{A}^{(i)}) - \eta^{(i)}P_{r_{tot}}(\boldsymbol{P}^{(i)}, \boldsymbol{A}^{(i)}) < \epsilon_\eta$ **then**
        Set $\boldsymbol{P}^* = \boldsymbol{P}^{(i)}$, $\boldsymbol{A}^* = \boldsymbol{A}^{(i)}$ and $\eta^* = \eta^{(i)}$;
        break;
    else
        $i = i + 1$, Setting $\boldsymbol{P} = \boldsymbol{P}^{(i)}$;
    end if
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 $\boldsymbol{Y}$ 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.

Algorithm 3 Plecement of Physical resources into Virtual resources

Sort Slices according to $\hat{\Omega}^{tot}_s, \forall s$ in descending order.
Sort DCs according to $\hat{\tau}_j, \forall j$ in descending order.
$\boldsymbol{Y} = \boldsymbol{0}$
for $d \leftarrow 1$ to D_c **do**
    for $s \leftarrow 1$ to S **do**
        if $\sum_{d=1}^{D_c} y_{s,d} == 0$ and $\bar{\Omega}^{tot}_{\mathfrak{z}(s)} \leq \tau_{\mathfrak{z}j}\forall\mathfrak{z}, \forall s$ **then**
            Set $y_{s,d} = 1$;
            $\tau_j \leftarrow \tau_j - \bar{\Omega}^{tot}_s$
        end if
    end for
end for
$\{ind_{residual} = s|(\sum_{d=1}^{D_c} y_{s,d} == 0)\}$
Sort residual amount of DCs same as before in descending order.
Sort residual slices same as before in descending order.
for $r \leftarrow 1$ to $S_{residual}$ **do**
    for $n \leftarrow 1$ to D_c **do**
        Set $y_{s,d} = 1$;
        $\bar{\Omega}^{tot}_s \leftarrow \bar{\Omega}^{tot}_s - \tau_j$
        if $\bar{\Omega}^{tot}_s == 0$ **then**
            Set $y_{s,d} = 1$;
            $\tau_j \leftarrow \tau_j - \bar{\Omega}^{tot}_s$
            break inner loop
        end if
    end for
end for
Remapping DCs must be done to prevent wasting Energy

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

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

TABLE II: Simulation Parameter

Parameter	Value
w_C	320
w_R	64
w_M	1

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 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.

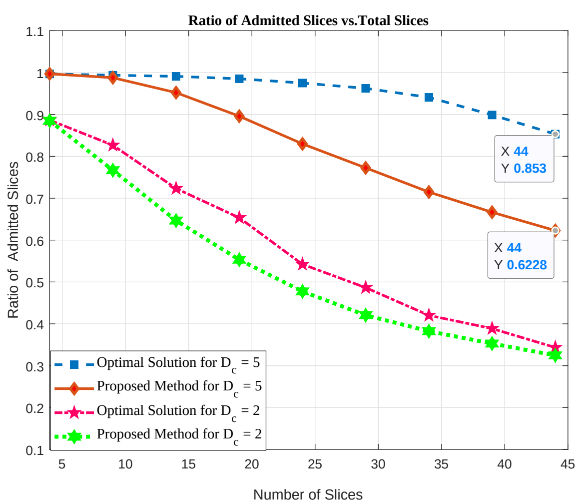

Fig. 2: Ratio of Admitted Slices connected to just one DC vs. Total slices

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.

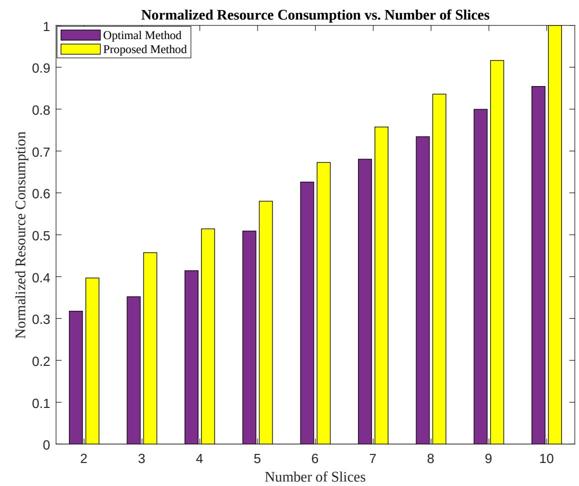

Fig. 3: Normalized Resource Consumption vs. Number of Slices

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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] (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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.