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—

Index Terms—component, formatting, style, styling, insert

I. Introduction

This document is a model and instructions for LaTeX. Please observe the conference page limits.

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 N_s PRBs. All the RRHs in a slice that is mapped to a service, transmit signals to all the UEs in specific service. Each RRH $r \in \{1, 2, ..., R\}$ is connected to BBU pool via an optical fiber link with limited fronthaul capacity. Also each RRH and PRB can serve more than one slice. It is considered that in BBU, the system has two processing layer consists of M_1 homogeneous VMs in first layer and M_2 homogeneous VMs in second layer.

B. Achievable Rate

In this subsection, the Achievable Rate is obtained as below. The achievable data rate for i^{th} UE in v^{th} service can be written as

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

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

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

Where, $P_{u(v,i)}$ represents the transmitted 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 equation (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 equation (3)

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

where $\boldsymbol{x}_{\mathcal{R}_s} = [x_{r_{(s,1)}}, ..., x_{r_{(s,\mathcal{R}_s)}}]^T \in \mathbb{C}^{R_s}$ depicts the transmitted symbol vector for the s-th set of Network slice, $\boldsymbol{z}_{U_v}$ 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 that map Physical Resource Blocks(PRB) to UE. Also as defined before, $\boldsymbol{H}_{\mathcal{R}_s,\mathcal{U}_v} = \left[\boldsymbol{h}_{\mathcal{R}_s,u_{(v,1)}}, \ldots, \boldsymbol{h}_{\mathcal{R}_s,v_{(v,\mathcal{U}_v)}}\right]^T \in \mathbb{C}^{R_s \times U_v}$ shows the channel matrix between RRH set $\mathcal{R}_s$ to UE set $\mathcal{U}_v$. The channel vector from the RRH of s^{th} slice to the i^{th} UE in the v^{th} service $\boldsymbol{h}_{\mathcal{R}_s,u_{(v,i)}} \in \mathbb{C}^{R_s}$ is modeled as below

$$\boldsymbol{h}_{\mathcal{R}_s,u_{(s,i)}} = \boldsymbol{\beta}^{\frac{1}{2}}_{\mathcal{R}_s,u_{(v,i)}} \boldsymbol{g}_{\mathcal{R}_s,u_{(v,i)}}, \tag{4}$$

where $\boldsymbol{g}_{\mathcal{R}_s,u_{(v,i)}} \backsim \mathcal{N}(0, N_0 \boldsymbol{I}_{\mathcal{U}_v})$ indicates the fast fading and flat fading channel vector and $\boldsymbol{\beta}_{\mathcal{R}_s,u_{(v,i)}} = \text{diag}(b_{r_{(s,1)},u_{(v,i)}}, \ldots, b_{r_{(s,\mathcal{R}_s)},u_{(v,i)}})$ represents the large scale fading matrix. Here, 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 follow

$$\boldsymbol{W}_{\mathcal{R}_s,\mathcal{U}_v} = \boldsymbol{H}_{\mathcal{R}_s,\mathcal{U}_v} (\boldsymbol{H}^H_{\mathcal{R}_s,\mathcal{U}_v} \boldsymbol{H}_{\mathcal{R}_s,\mathcal{U}_v})^{-1} \tag{5}$$

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

$$\begin{aligned} I_{u_{(v,i)}} &= \underbrace{\sum_{s=1}^{S}\sum_{n=1}^{N_s}\sum_{\substack{l=1\\l\neq i}}^{U_v}\gamma_1 p_{u_{(v,l)}} a_{vs}\zeta_{u_(v,i),n,s}\zeta_{u_(v,l),n,s}}_{\text{(intra-service interference)}} \\ &+ \underbrace{\sum_{\substack{y=1\\l\neq v}}^{V}\sum_{s=1}^{S}\sum_{n=1}^{N_s}\sum_{l=1}^{U_y}\gamma_2 p_{u_{(y,l)}} a_{ys}\zeta_{u_(v,i),n,s}\zeta_{u_(y,l),n,s}}_{\text{(inter-service interference)}} \\ &+ \underbrace{\sum_{s=1}^{S}\sum_{j=1}^{R_s}\sigma_{q_{r_{(s,j)}}}^{2}|\boldsymbol{h}_{r_{(s,j)},u_{(v,i)}}|^2 a_{vs}}_{\text{(quantization noise interference)}}. \end{aligned} \tag{6}$$

where, $\gamma_1 = |\boldsymbol{h}^H_{\mathcal{R}_s,u_{(v,i)}}\boldsymbol{w}_{\mathcal{R}_s,u_{(v,l)}}|^2$ and $\gamma_2 = |\boldsymbol{h}^H_{\mathcal{R}_s,u_{(v,i)}}\boldsymbol{w}_{\mathcal{R}_s,u_{(y,l)}}|^2$. It is assumed that interference signal applied from UEs with same PRB. let $\bar{p}_{r_{(s,j)}}$ denote the power of transmitted signal from j^{th} RRH in s^{th} slice. from equation (3) we have

$$\bar{p}_{r_{(s,j)}} = \sum_{v=1}^{V}\boldsymbol{w}_{r_{(s,j)},\mathcal{U}_v}\boldsymbol{P}^{\frac{1}{2}}_{\mathcal{U}_v}\boldsymbol{P}^{H\frac{1}{2}}_{\mathcal{U}_v}\boldsymbol{w}^H_{r_{(s,j)},\mathcal{U}_v}a_{vs} + \sigma^2_{q_{r(s,j)}}. \tag{7}$$

As a result the user data capacity on the fronthual link between BBU and the j^{th} RRH in s^{th} slice is formulated as below

$$C_{R_{(s,j)}} = \log\big(1 + \sum_{v=1}^{V}\frac{w_{r_{(s,j)},\mathcal{D}_s}\boldsymbol{P}^{\frac{1}{2}}_{\mathcal{U}_v}\boldsymbol{P}^{H\frac{1}{2}}_{\mathcal{U}_v}w^H_{r_{(s,j)},\mathcal{U}_v}a_{vs}}{\sigma^2_{q_{r(s,j)}}}\big), \tag{8}$$

C. Mean Delay

Suppose that we have two processing layer in BBU of O-RAN system. The lower layer is consist of PHY and MAC and the upper layer is consist of RLC, PDCP and SDAP.

As it is mentioned before, we have M_1 VMs in the first layer and M_2 VMs in second layer. Each VM in both layers map to one or more slices. So in s^{th} slice, there are M_{s_1} VMs in first layer and M_{s_2} VMs in second layer. Each VM in first and second layer has computational capacity that is equal to μ_1 and μ_2 respectively.

Let the packet arrival of UEs have a Poisson Process with arrival rate $\lambda_{u(v,i)}$ for i^{th} UE in v^{th} service. Therefore, the mean arrival data rate of UEs in s^{th} slice in the first layer is $\alpha_{s_1} = \sum_{v=1}^{V}\sum_{u=2}^{U_v}a_{vs}\lambda_{u(v,i)}$. 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 Burke's Theorem, the arrival packets of second layer which is processed in first layer is still Poisson with rate α_s It is assumed that there are dispatchers in each layer for each slice to divide the incoming traffic to VMs. Suppose the baseband processing of each VM is depicted as a M/M/1 processing queue. Each packet is routed by one of VMs of slices. So the mean delay of slice s which is related to incoming traffic rate routed to each VM in first layer can be written as follow

$$d_{s_1} = \frac{1}{\mu_1 - \alpha_s/M_{s_1}} \tag{9}$$

Also, the delay in s^{th} slice in second layer can be formulated as below

$$d_{s_2} = \frac{1}{\mu_2 - \alpha_s/M_{s_2}} \tag{10}$$

In addition, the arrival data rate to the queue of wireless transmission is equal to the arrival data rate of dispatcher. 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

$$d_{s_{tr}} = \frac{1}{R_{tot_s} - \alpha_s} \tag{11}$$

Where, $R_{tot_s} = \sum_{v=1}^{V}\sum_{u=2}^{U_v}a_{vs}R_{u(v,i)}$. We define a new parameter which indicates mean delay of each slice

$$D_s = d_{s_1} + d_{s_2} + d_{s_{tr}}\forall s \tag{12}$$

D. Physical Resource

Assume each VM is mapped to one virtual network function (VNF) for simplicity. Each VNF requires physical resources which contain RAM, Memory and CPU. Let, the required resources for VNF f in slice s is represented by

$$\Omega_{(f,s)} = \{\Omega_{R_{f,s}}, \Omega_{M_{f,s}}, \Omega_{C_{f,s}}\} \tag{13}$$

Where, $\Omega_{R_{f,s}}, \Omega_{M_{f,s}}, \Omega_{C_{f,s}}$ indicate the amount of required RAM, Memory and CPU. Also, in the Core Network(CN), there are N_D data centers(DC), which served VNFs. Each DC contains several servers that supply VNF's needs. The amount of RAM, Memory and CPU is denoted respectively by τ_{R_j}, τ_{M_j} and τ_{C_j} for j^{th} DC.

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

E. Problem Statement

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

$$\eta(\boldsymbol{P},\boldsymbol{A}) := \frac{\sum\limits_{v=1}^{V}\sum\limits_{k=1}^{U_v}\mathcal{R}_{u_{(v,k)}}}{\sum\limits_{s=1}^{S}\sum\limits_{i=1}^{R_s}\bar{p}_{r_{(s,i)}}} = \frac{R_{tot}(\boldsymbol{P},\boldsymbol{A})}{P_{r_{tot}}(\boldsymbol{P},\boldsymbol{A})}, \tag{14}$$

Assume the power consumption of baseband processing at each data center that is mapped to VMs of a slice is depicted as ϕ. So the total power can be represented as

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

Where, $y_{s,d}$ is a binary variable which indicates whether d^{th} data-center is mapped to VNFs of s^{th} slice or not.
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,

$$
\begin{aligned}
\max_{\boldsymbol{P},\boldsymbol{A},\boldsymbol{y}} \quad & \eta(\boldsymbol{P},\boldsymbol{A}) + \frac{1}{\phi_{tot}(\boldsymbol{Y})} && && (15a)\\
\text{subject to} \quad & \bar{p}_{r_{(s,i)}} \le P_{max} && \forall s, \forall i, && (15b)\\
& p_{u_{(v,k)}} \ge 0 && \forall v, \forall k, && (15c)\\
& \mathcal{R}_{u_{(v,k)}} \ge \mathcal{R}^{min}_{u_{(v,k)}} && \forall v, \forall k, && (15d)\\
& C_{r_{(s,i)}} \le C^{max}_{r_{(s,i)}} && \forall s, \forall i, && (15e)\\
& D_s \le D^{th}_s && \forall s, && (15f)\\
& \sum_{s=1}^{S} a_{vs} \ge 1 && \forall s, && (15g)\\
& \sum_{d=1}^{D_c}\sum_{v=1}^{V} y_{s,d} a_{vs} \ge 1 \times \sum_{v=1}^{V} a_{vs} && \forall s, && (15h)\\
& \sum_{f=1}^{F_s} \Omega_{\mathfrak{z}(f,s)} \le \sum_{d=1}^{D_c} y_{s,d}\tau_d && \forall f, \forall s && (15i)
\end{aligned}
$$

Where, $\boldsymbol{P} = [p_{u(v,k)}]\forall v, \forall k$, $\boldsymbol{A} = [a_{vs}]\forall v, \forall s$ and $\boldsymbol{Y} = [y_{s,d}]\forall s, \forall d$. Equation (15b), (15c) indicates respectively that the power of each RRH do not exceed the maximum power and power of each UE is a positive integer value. Also (15d) shows that the rate of each UE is more than a threshold. (15e) and (15f) depicts respectively that the capacity of fronthaul link is limited and the delay of receiving signal should be less than a threshold. Furthermore, (15g) ensure that each service is mapped to one or more slice. Also, (15h) guarantee that each slice (VNFs in two layers of slice) has been placed to one or more physical resources (DC). Moreover, in (15i) $\mathfrak{z} \in \{M, R, C\}$, which supports that we have enough physical resource for VNFs of each slice.
The main optimizaiton problem which is formulated as (15) can be decomposed into two independent optimization problem. The First and main problem is

$$
\begin{aligned}
\max_{\boldsymbol{P},\boldsymbol{A}} \quad & \eta(\boldsymbol{P},\boldsymbol{A}) && && (16a)\\
\text{subject to} \quad & \bar{p}_{r_{(s,i)}} \le P_{max} && \forall s, \forall i, && (16b)\\
& p_{u_{(v,k)}} \ge 0 && \forall v, \forall k, && (16c)\\
& \mathcal{R}_{u_{(v,k)}} \ge \mathcal{R}^{min}_{u_{(v,k)}} && \forall v, \forall k, && (16d)\\
& C_{r_{(s,i)}} \le C^{max}_{r_{(s,i)}} && \forall s, \forall i, && (16e)\\
& D_s \le D^{th}_s && \forall s && (16f)
\end{aligned}
$$

and the second problem is

$$
\begin{aligned}
\min_{\boldsymbol{y}} \quad & \phi_{tot}(\boldsymbol{Y}) && && (17a)\\
\text{subject to} \quad & \bar{p}_{r_{(s,i)}} \le P_{max} && \forall s, \forall i, && (17b)\\
& \sum_{s=1}^{S} a_{vs} \ge 1 && \forall s, && (17c)\\
& \sum_{d=1}^{D_c}\sum_{v=1}^{V} y_{s,d} a_{vs} \ge 1 \times \sum_{v=1}^{V} a_{vs} && \forall s, && (17d)\\
& \sum_{f=1}^{F_s} \Omega_{(f,s)} \le \sum_{d=1}^{D_c} y_{s,d}\tau_d && \forall f, \forall s && (17e)
\end{aligned}
$$

F. Proposed Method For Problem (16)

In this subsection, the proposed method is applied to solve the optimization problem. We want to solve (16). 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 (16) to two different part that can be solved iteratively.

1) First Sub-Problem: Firstly, we need to obtain $\boldsymbol{A}$ by fixing $\boldsymbol{P}$ in the problem (16) 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 1 ≤ i ≤ S do
    for 1 ≤ j ≤ V do
        Set a_ij = 1
        Obtain Parameters of System
        if conditions (15b), (15c), (15d) and (15e) is not applied then
            Set a_ij = 0;
        else
            break from inner loop;
        end if
    end for
end for
```

2) Second Sub-Problem: 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

$$
\begin{aligned}
&\max_{\boldsymbol{P}} (R_{tot}(\boldsymbol{P}) - \eta^* P_{r_{tot}}(\boldsymbol{P})) = \\
&R_{total}(\boldsymbol{P}^*) - \eta^* P_{r_{tot}}(\boldsymbol{P}^*) = 0.
\end{aligned} \tag{18}
$$

This 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 for interference (the worst-case) as follow

$$
\begin{aligned}
\bar{I}_{u_{(v,i)}} = & \underbrace{\sum_{s=1}^{S} \sum_{n=1}^{N_s} \sum_{\substack{l=1 \\ l \neq i}}^{U_v} \gamma_1 p_{max} a_{vs} \zeta_{u_(v,i),n,s} \zeta_{u_(v,l),n,s}}_{\text{(intra-service interference)}} \\
& + \underbrace{\sum_{\substack{y=1 \\ l \neq v}}^{V} \sum_{s=1}^{S} \sum_{n=1}^{N_s} \sum_{l=1}^{U_y} \gamma_2 p_{max} a_{ys} \zeta_{u_(v,i),n,s} \zeta_{u_(y,l),n,s}}_{\text{(inter-service interference)}} \\
& + \underbrace{\sum_{s=1}^{S} \sum_{j=1}^{R_s} {\sigma_{q}}^2_{r_{(s,j)}} |\boldsymbol{h}_{r_{(s,j)},u_{(v,i)}}|^2 a_{vs}}_{\text{(quantization noise interference)}} .
\end{aligned} \tag{19}
$$

where, $\gamma_1 = |\boldsymbol{h}^H_{\mathcal{R}_s,u_{(v,i)}} \boldsymbol{w}_{\mathcal{R}_s,u_{(v,l)}}|^2$ and $\gamma_2 = |\boldsymbol{h}^H_{\mathcal{R}_s,u_{(v,i)}} \boldsymbol{w}_{\mathcal{R}_s,u_{(y,l)}}|^2$.

Lagrangian function is written as follow

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