

Cloud-based Queuing Model for Tactile Internet in Next Generation of RAN

Narges Gholipoor, Saeedeh Parsaeefard, Mohammad Reza Javan, Nader Mokari,

Hamid Saeedi

Abstract

Ultra low latency is the most important requirement of the Tactile Internet. In this paper, a new queuing model for the Tactile Internet is proposed for the cloud radio access network (C-RAN) architecture of the next generation wireless networks, e.g., fifth generation wireless networks (5G), assisted via power domain non-orthogonal multiple access (PD-NOMA) technology. This model includes both the radio remote head (RRH) and baseband processing unit (BBU) queuing delays for each end to end (E2E) connection between a pair of tactile users. For this setup, with the aim to minimize the transmit power of users subject to guaranteeing tolerable delay of users, and fronthaul and access limitations, we formulate a resource allocation problem. Since the proposed optimization problem is highly non-convex, to solve it in an efficient manner, we utilize diverse transformation techniques such as successive convex approximation (SCA) and difference of two convex functions (DC). In our proposed system model, we dynamically adjust the fronthaul and access links to minimize the transmit power. Simulation results reveal that by dynamic adjustment of the access and fronthaul delays, transmit power can be saved compared to the case of fixed approach per each transmission session. Also, energy efficiency of orthogonal frequency division multiple access (OFDM) and NOMA are compared for our setup.

Index Terms

C-RAN, Tactile Internet, NOMA.

I. INTRODUCTION

The Tactile Internet is a new service portfolio of the next generation of wireless networks, e.g., the fifth generation wireless networks (5G), where a novel communication paradigm is introduced. For instance, via the Tactile Internet touch sensation can be remotely transmitted. One of the most important requirements of Tactile Internet service is ultra low end-to-end (E2E) e.g., E2E delay should be less than one millisecond [1]–[4]. These requirements cannot be guaranteed via existing wireless networks such as fourth-generation wireless networks (4G) [4]. However, 5G platform via its own soft, virtualized, and cloud based architecture can be leveraged to implement the Tactile Internet services [1], [2].

For instance, via the concept of cloud radio access network (C-RAN) in 5G, spectral efficiency (SE) and energy efficiency (EE) along with cost can be efficiently optimized, where baseband processing is performed by the baseband units (BBUs) which is connected to remote radio heads (RRHs) via the fronthaul links [5], [6]. Specifically, C-RAN reduces energy consumption and cost, and improves throughput in dense environment [7], [8]. Therefore, this RAN architecture is a proper environment for the implementation of the Tactile Internet services in hotspot areas.

On the other hand, proper design and selection of multiple access technologies have a significant impact on the SE and EE. One of the promising approaches for the next generation wireless networks is the power domain non-orthogonal multiple access (PD-NOMA)¹ which can significantly improve the SE and reduce the transmission delay [9]. In NOMA, the available spectrum is shared among different users of one RRH. A NOMA transmitter requires implementing a method for superposition of different users signals. Consequently, to compensate interference between users, a complex method for decoding the signals are required at a NOMA receiver. Successive interference cancellation (SIC) technique is one of these methods which is implemented at the receiver to decode the desirable signal [10], [11].

¹We use PD-NOMA and NOMA interchangeably in the followings.

In the next generation wireless networks, due to the introduction of various services, each to be provided with a high quality of service (QoS) via the virtualization techniques, the concept of slice has been defined for each service in which each slice is a bundle of users with a specific set of QoS requirements [12], [13]. The slice concept adds flexibility to utilizing resources which leads to higher SE and EE. However, in this concept, the isolation between slices should be preserved such that the activities of users of one slice do not have harmful effects on QoS of the users of other slices. One of the major issues in the slicing is how to translate the isolation concept to the proper notation for the networks procedures. There exists a large body of work for this translation such as dynamic and static methods [12]–[14]. In this paper, we consider the minimum required rate of each slice as a means of preserving the isolation between slices [14], [15].

Obviously, for this setup due to the complexity of system architecture, diverse transmission parameters such as power, and different QoS requirements, resource allocation are highly essential which has drawn a lot of attention recently [3], [16]–[19]. For instance, in [16], a resource allocation problem for Tactile Internet in the Long-Term Evolution-Advanced (LTE-A) is investigated where the average queuing delay and queuing delay violation in one base station (BS) are optimized. Orthogonal frequency division multiple access (OFDM) and single carrier frequency division multiple access (SC-FDMA) are considered for downlink (DL) and uplink (UL), respectively. The cross-layer resource allocation problem for Tactile Internet is proessed in [3] for single BS where the packet error probability, maximum allowable queuing delay violation probability, and packet dropping probability are jointly optimized with the objective to minimize the total transmit power subject to maximum allowable queuing delays. In [16], queuing delay, packet loss induced by queuing delay violation, packet error, and packet drop caused by channel fading are considered for analyzing the E2E delay of RAN. In [17], the effect of frequency diversity and spatial diversity on the transmission reliability in UL is studied in the Tactile Internet service where the number of subcarriers, the bandwidth of each subcarrier,

and the threshold for each user are optimized for minimizing the total bandwidth to ensure the transmission reliability. In [18], a multi-cell network based on frequency division multiple access (FDMA) with a fixed delay for backhaul is studied in the Tactile Internet service. Moreover, queuing delay, delay violation probability, and decoding error probability are considered for analyzing the E2E delay of Tactile Internet service.

In the above-mentioned works, a network is considered in which for each user one queue at the BS is assumed. Therefore, by increasing the number of users, a lots of queues are needed at the BS. However, given that the Tactile Internet is assumed to be implemented in the 5G framework, it is necessary to consider C-RAN architecture. There exists a set of RRHs in highly dense network which are connected to BBU center via fronthaul links. Moreover, previous works consider orthogonal multiple access techniques instead of NOMA. In addition, the results in the above works generally ignore the fronthaul delay. However, due to the importance of delay in the Tactile Internet, it is crucial to consider queuing delay in fronthaul as well, otherwise, the resulting allocation of the resources may not practically fulfill the requirement of Tactile Internet.

To address the mentioned issues, we consider a C-RAN architecture serving a set of tactile users. The contributions of this paper are as follows, many of which have been considered for the first time in Tactile Internet:

- We propose a C-RAN scenario in ultra dense environment in 5G platform. This will impose new constraints to the system as far as the number of queues is concerned. For the considered C-RAN architecture, we propose a practical queuing model for sequential queues in Tactile Internet that can be implemented in realistic networks. We consider a PD-NOMA scheme for our system model while all earlier works are based on OFDMA and FDMA schemes. Moreover, we consider slicing for Tactile Internet service in our work.
- In contrast to [3], [16], [17] where fronthaul delay is ignored, we take this delay into consideration. Moreover, we consider dynamic adjustment of the access and fronthaul delays based on channel state information (CSI) for each pair of users instead of fixed maximum

delay values per each transmission part of our setup and show that it can significantly reduce the required total transmit power.

The rest of this paper is as follows. In Section II, the system model is described. In Section III, we formulate the optimization problem. Numerical results and simulation are presented in Section IV. Finally, Section V concludes the paper.

II. SYSTEM MODEL

We consider a C-RAN network where all RRHs are connected to the BBU via fronthaul links. In this region, there exist several pairs of tactile users where each user aims to send its information to its own tactile user pair via the closest RRH through the uplink transmission link. Then, RRH sends the received data to the BBU via the fronthaul link. The BBU processes all received data and then sends data to the corresponding RRH of its paired tactile user. Finally, this RRH transmits the relevant message to the paired tactile user via downlink (DL) transmission link. Assume each RRH has only one queue for UL transmission and all the data of tactile users is stored in this queue. RRHs send all data to the BBU to process. Then BBU sends data to the corresponding RRH. In DL, we consider each RRH has a queue for each user for sending data to the paired users such as connected car and telemedicine applications.

As shown in Fig. 1, in the considered system model, we have $\mathcal{J} = \{1, \dots, J\}$ RRHs, $\mathcal{S} = \{1, \dots, S\}$ slices and $\mathcal{I} = \{0, \dots, I\}$ pairs of tactile users. Slice s contains $\mathcal{I}_s = \{0, \dots, I_s\}$ tactile users and the total number of tactile users in our system model is equal to $\mathcal{I} = \bigcup_{s \in \mathcal{S}} \mathcal{I}_s$ pairs of users. The terms of access link and fronthaul link often are used to express the RRH-user connection and RRH-BBU connection, respectively. In order to reduce the cost of cabling, wireless fronthaul is used instead of fiber fronthaul [5], [20]. Considering that fronthaul links are provided via wireless channels in ultra dense environment, we assume there exist two sets of subcarriers $\mathcal{K}_1 = \{1, \dots, K_1\}$ and $\mathcal{K}_2 = \{1, \dots, K_2\}$ for access and fronthaul links, respectively. Moreover, we define $\mathcal{Q} = \{\text{UL}, \text{DL}\}$ for simplicity. We consider a two-phase transmission; in

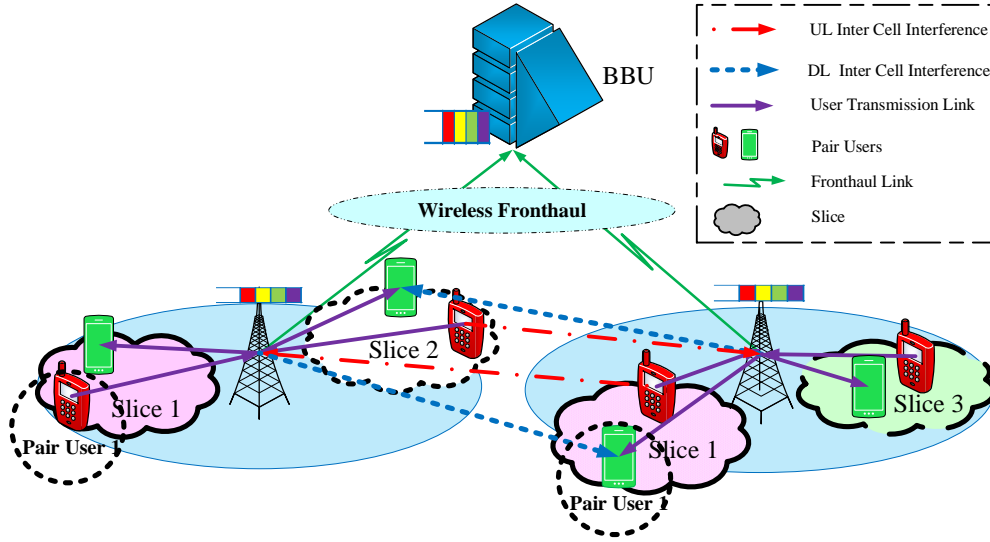


Fig. 1: The illustration of the considered network in which three slices with two RRHs are considered. Here as an example, a pair of tactile users is shown by the dotted circles.

the first phase, all tactile users send their data to the corresponding RRH and simultaneously all RRHs send their buffered data to one BBU via fronthaul links.

In the second phase, all RRH send data to the corresponding tactile users, and simultaneously, BBU sends buffered data to all RRH via fronthaul links. These two phases do not perform at the same frequency. Therefore, the proposed system model is based on the frequency division duplex (FDD) transmission mode in which each RRH can transmit and receive simultaneously in different frequencies. In order to isolate slices, a minimum required rate in each slice s must be reserved [14], [21], [22]. By considering the above definitions, we can now proceed to review the system parameters.

Remark 1: Note that the complexity of SIC cancellation in NOMA is primarily a function of the number of code layers in superposition coding [23]. By assuming that two users is allocated to each subcarrier, the complexity of SIC receiver is constant and negligible due to the fact that only two layers superposition coding is required for this case [24], [25]. In addition, since strong and low latency codes such as low density parity check code (LDPC) can be utilized for

decoding in our setup, the decoding delay of the signals can be neglected [23]. In particular, for this type of detection codes, a decoding delay is around 100 ns which can be omitted in this setup [4], [26], [27]. Moreover, the LDPC can increase the reliability of decoding.

Remark 2: To estimate the CSI for DL transmission, pilot signals are transmitted via RRHs to all users. Then, each user sends the estimate channel sends to RRHs via feedback channels. To estimate the CSI for UL, pilot signals are transmitted via users to RRHs, and then, the estimate channels are sent to the users. For the CSI estimation, one of the proposed approaches in [28]–[30] can be applied.

A. Access Links Parameters

We introduce a binary variable $\tau_{i,k_1}^{s,j,q}$ which is set to 1 if subcarrier k_1 is assigned to user i in slice s at RRH j , i.e.,

$$\tau_{i,k_1}^{s,j,q} = \begin{cases} 1, & \text{if subcarrier } k_1 \text{ is assigned to user } i \text{ in slice } s \text{ at RRH } j \text{ and } q \in \mathcal{Q}, \\ 0, & \text{otherwise.} \end{cases}$$

Since we deploy PD-NOMA in this setup, each subcarrier can be allocated to L_1 users [9].

Therefore, we have the following constraint

$$\text{C1: } \sum_{s \in \mathcal{S}} \sum_{i \in \mathcal{I}_s} \tau_{i,k_1}^{s,j,q} \leq L_1, \forall j \in \mathcal{J}, k_1 \in \mathcal{K}_1, q \in \mathcal{Q}.$$

Here, for all $i \in \mathcal{I}$, $j \in \mathcal{J}$, $k_1 \in \mathcal{K}_1$, $s \in \mathcal{S}$ and $q \in \mathcal{Q}$, the achievable rate for user i on subcarrier k_1 at RRH j can be calculated as

$$r_{i,k_1}^{s,j,q} = \log_2(1 + \gamma_{i,k_1}^{s,j,q}), \quad (1)$$

where $\gamma_{i,k_1}^{s,j,q} = \frac{p_{i,k_1}^{s,j,q} h_{i,k_1}^{s,j,q}}{\sigma_{i,k_1}^{s,j,q} + I_{i,k_1}^{s,j,q} + \tilde{I}_{i,k_1}^{s,j,q}}$, in which $p_{i,k_1}^{s,j,q}$, $h_{i,k_1}^{s,j,q}$, and $\sigma_{i,k_1}^{s,j,q}$ represent the transmit power, channel power gain from RRH j to user i on subcarrier k_1 in slice s , and noise power, respectively.

Also, $I_{i,k_1}^{s,j,q}$ is intra-cell interference which is equal to $I_{i,k_1}^{s,j,q} = \sum_{v \in \mathcal{S}} \sum_{\substack{u \in \mathcal{I}_v, u \neq i \\ h_{u,k_1}^{s,j,q} > h_{i,k_1}^{v,j,q}}} \tau_{u,k_1}^{v,j,q} p_{u,k_1}^{v,j,q} h_{i,k_1}^{s,j,q}$,

and $\tilde{I}_{i,k_1}^{s,j,q}$ is inter-cell interference which is equal to $\tilde{I}_{i,k_1}^{s,j,q} = \sum_{f \in \mathcal{J}, f \neq j} \sum_{v \in \mathcal{S}} \sum_{u \in \mathcal{I}_v} \tau_{u,k_1}^{v,f,q} p_{u,k_1}^{v,f,q} h_{i,k_1}^{v,f,q}$.

Therefore, the total achievable rate in the access links at RRH j is as follows

$$R_{\text{RRH}_j}^q = \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} \tau_{u,k_1}^{s,j,q} r_{u,k_1}^{s,j,q} \forall j \in \mathcal{J}, q \in \mathcal{Q}. \quad (2)$$

Due to the power limitation of each RRH in DL transmission, we have the following constraint

$$\text{C2: } \sum_{i \in \mathcal{I}} \sum_{s \in \mathcal{S}} \sum_{k_1 \in \mathcal{K}_1} \tau_{i,k_1}^{s,j,\text{DL}} p_{i,k_1}^{s,j,\text{DL}} \leq P_{\text{RRH}_j}^{\text{DL}}, \forall j \in \mathcal{J}.$$

Moreover, due to the power limitation of each user, we have

$$\text{C3: } \sum_{j \in \mathcal{J}} \sum_{s \in \mathcal{S}} \sum_{k_1 \in \mathcal{K}_1} \tau_{i,k_1}^{s,j,\text{UL}} p_{i,k_1}^{s,j,\text{UL}} \leq P_{\text{USER}_i}^{\text{UL}}, \forall i \in \mathcal{I}.$$

Due to using NOMA in access, we have SIC constraint as follows

$$\text{C4: } \frac{\tau_{i,k_1}^{s,j,q} p_{m',k_1}^{s,j,q} h_{i,k_1}^{s,j,q}}{\tilde{I}_{i,k_1}^{s,j,q} + I_{i,k_1}^{s,j,q} + \sigma_{i,k_1}^{s,j,q}} \geq \frac{\tau_{i,k_1}^{s,j,q} p_{m',k_1}^{s,j,q} h_{m',k_1}^{s,j,q}}{\tilde{I}_{m',k_1}^{s,j,q} + I_{m',k_1}^{s,j,q} + \sigma_{m',k_1}^{s,j,q}}, \forall j, s, q, i, m' \in \mathcal{I}, m' \neq i, h_{i,k_1}^{s,j,q} > h_{m',k_1}^{s,j,q},$$

B. Fronthaul Links Parameters

We introduce a binary variable $x_{k_2}^{j,q}$ denoting that subcarrier k_2 is assigned to RRH j , and defined by

$$x_{k_2}^{j,q} = \begin{cases} 1, & \text{if subcarrier } k_2 \text{ is assigned to RRH}_j \text{ and } q \in \mathcal{Q}. \\ 0, & \text{Otherwise.} \end{cases}$$

Assuming that PD-NOMA is also deployed for the fronthaul links, again each subcarrier can be allocated to at most L_2 RRHs, and hence, we have the following constraint

$$\text{C5: } \sum_{j \in \mathcal{J}} x_{k_2}^{j,q} \leq L_2, \forall k_2 \in \mathcal{K}_2, q \in \mathcal{Q}.$$

The achievable rate for each RRH on subcarrier k_2 is calculated as follows

$$r_{k_2}^{j,q} = \log_2(1 + \gamma_{k_2}^{j,q}), \forall j \in \mathcal{J}, k_2 \in \mathcal{K}_2, q \in \mathcal{Q}, \quad (3)$$

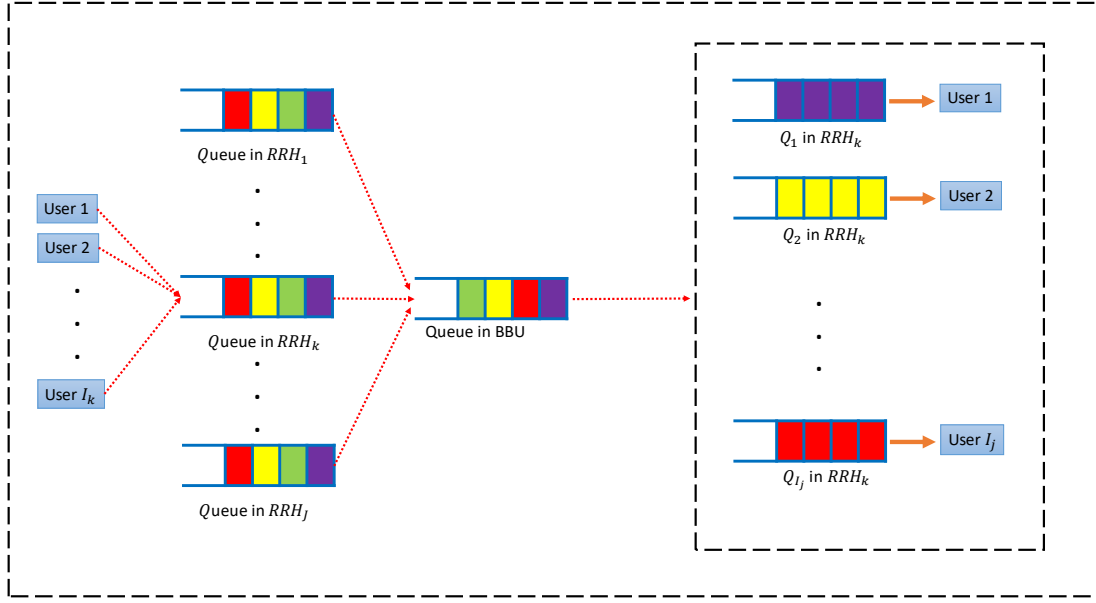


Fig. 2: Queuing model for our setup where each RRH has only one queue in UL transmission of its all users. RRHs send all data to the BBU Queue. In DL, each RRH has a specific queue for each user.

where $\gamma_{k_2}^{j,q}$ is defined as $\gamma_{k_2}^{j,q} = \frac{p_{k_2}^{j,q} h_{k_2}^{j,q}}{\sigma_{k_2}^{j,q} + I_{k_2}^{j,q}}, \forall j \in \mathcal{J}, k_2 \in \mathcal{K}_2, q \in \mathcal{Q}$, where $I_{k_2}^{j,q}$ is the interference among RRHs and is represented as $I_{k_2}^{j,q} = \sum_{\substack{f \in \mathcal{J}, \\ f \neq j, h_{k_2}^{f,q} > h_{k_2}^{j,q}}} x_{k_2}^{f,q} p_{k_2}^{f,q} h_{k_2}^{j,q}, \forall j \in \mathcal{J}, k_2 \in \mathcal{K}_2, q \in \mathcal{Q}$. Therefore, the total achievable rate in BBU is obtained as follows

$$R_{\text{BBU}}^q = \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} x_{k_2}^{j,q} r_{k_2}^{j,q}, \forall q \in \mathcal{Q}. \quad (4)$$

Due to the power limitation of each RRH in UL transmission, we have

$$\text{C6: } \sum_{k_2 \in \mathcal{K}_2} x_{k_2}^{j,\text{UL}} p_{k_2}^{j,\text{UL}} \leq P_{\text{RRH}_j}^{\text{UL}}, \forall j \in \mathcal{J}.$$

Moreover, due to the power limitation of BBU, we have

$$\text{C7: } \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} x_{k_2}^{j,\text{DL}} p_{k_2}^{j,\text{DL}} \leq P_{\text{BBU}}^{\text{DL}}.$$

Due to using NOMA in fronthaul, we have SIC constraint as follows

$$\text{C8: } \frac{x_{k_2}^{j,q} p_{k_2}^{j',q} h_{k_2}^{j,q}}{I_{k_2}^{j,q} + \sigma_{k_2}^{j,q}} \geq \frac{x_{k_2}^{j,q} p_{k_2}^{j',q} h_{k_2}^{j',q}}{I_{k_2}^{j',q} + \sigma_{k_2}^{j',q}}, \forall j, q \in \mathcal{Q} \forall j, j' \in \mathcal{J}, j' \neq j, h_{k_2}^{j,\mathcal{Q}} > h_{k_2}^{j',q}.$$

C. Queuing Delay Model

The total delay of this architecture consists of three components: delay resulting from UL queues at RRH, BBU queue, and DL queues at RRHs, as shown in Fig. 2. Due to delay constraint in Tactile Internet service, we have

$$\text{C9: } D_{\max}^j + D_{\max}^{i,j} + D_{\max}^{BBU} \leq D_{\max}^{\text{Total}}, \forall i \in \mathcal{I}, j \in \mathcal{J}, s \in \mathcal{S},$$

where D_{\max}^j , $D_{\max}^{i,j}$, D_{\max}^{BBU} , and D_{\max}^{Total} are delays of UL queues at RRH, BBU queue, DL queues at RRH, and total delay, respectively.

1) *UL Queuing Delay*: The aggregation of receiving bits from several nodes can be modeled as a Poisson process [3], [31]. The effective bandwidth for a Poisson arrival process in RRH j is defined as [3], [31], [32]:

$$E_B^j(\theta_j) = \lambda_j \frac{(e^{\theta_j} - 1)}{\theta_j}, \forall j \in \mathcal{J},$$

where θ_j is the statistical QoS exponent of the j^{th} RRH. A larger θ_j indicates a more stringent QoS and a smaller θ_j implies a looser QoS requirement. λ_j is the number of bits arrived at RRH j queue defined as $\lambda_j = \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} r_{u,k_1}^{s,j,\text{UL}}$, $\forall j \in \mathcal{J}$. The probability of queuing delay violation for RRH j can be approximated as

$$\epsilon_1^j = \Pr\{D_j > D_{\max}^j\} = \eta_1 \exp(-\theta_j E_B^j(\theta_j) D_{\max}^j), \quad (5)$$

for all $j \in \mathcal{J}$ where D_j is the j^{th} RRH delay, D_{\max}^j is the maximum delay, and η_1 is the non-empty buffer probability. Equation (5) can be simplified to

$$\exp(-\theta_j E_B^j(\theta_j) D_{\max}^j) = \exp(-\theta_j \lambda_j \frac{(e^{\theta_j} - 1)}{\theta_j} D_{\max}^j) = \exp(-\lambda_j (e^{\theta_j} - 1) D_{\max}^j) \leq \delta_1.$$

Therefore, we have

$$\text{C10: } \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} r_{u,k_1}^{s,j,\text{UL}} \geq \frac{\ln(1/\delta_1)}{(e^{\theta_j} - 1) D_{\max}^j}, \forall j \in \mathcal{J}.$$

2) *BBU Queuing Delay*: We consider a queue for all RRH at the BBU for processing data. Therefore, the formulas in the previous section can also be used for this section. The effective bandwidth for each queue in BBU is $E_B^{\text{BBU}}(\theta_{\text{BBU}}) = \Lambda_{\text{BBU}} \frac{(e^{\theta_{\text{BBU}}} - 1)}{\theta_{\text{BBU}}}$, where θ_{BBU} is the statistical QoS exponent in the BBU and Λ_{BBU} is the number of bits arrived at the queue in the BBU which is defined as $\Lambda_{\text{BBU}} = \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} r_{k_2}^{j, \text{UL}}$. The probability of queuing delay violation at the BBU can be approximated as

$$\epsilon_{\text{BBU}} = \Pr\{D_{\text{BBU}} > D_{\text{max}}^{\text{BBU}}\} = \eta_2 \exp(-\theta_{\text{BBU}}^* E_B^{\text{BBU}}(\theta_{\text{BBU}}) D_{\text{max}}^{\text{BBU}}), \quad (6)$$

where η_2 is the non-empty buffer probability. Equation (6) can be simplified to

$$\exp(-\theta_{\text{BBU}}^* E_B^{\text{BBU}}(\theta_{\text{BBU}}) D_{\text{max}}^{\text{BBU}}) = \exp(-\theta_{\text{BBU}}^* \Lambda_{\text{BBU}} \frac{(e^{\theta_{\text{BBU}}} - 1)}{\theta_{\text{BBU}}} D_{\text{max}}^{\text{BBU}}) \leq \delta_2.$$

Therefore, we have

$$\text{C11: } \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} r_{k_2}^{j, \text{UL}} \geq \frac{\ln(1/\delta_2)}{(e^{\theta_{\text{BBU}}} - 1) D_{\text{max}}^{\text{BBU}}}.$$

3) *DL Queuing Delay*: The effective bandwidth for each user in RRH j is defined as $E_B^{i,j}(\theta_i^j) = \lambda_i^j \frac{(e^{\theta_i^j} - 1)}{\theta_i^j}$, $\forall i \in \mathcal{I}, j \in \mathcal{J}$, where θ_i^j is the statistical QoS exponent of the i^{th} user in RRH j and λ_i^j is the number of bits arrived at user i queue in RRH j which is defined as $\lambda_i^j = \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} r_{u,k_1}^{s,j, \text{DL}}$, $\forall i \in \mathcal{I}, j \in \mathcal{J}$. The probability of queuing delay violation for user i can be approximated as

$$\epsilon_3^{i,j} = \Pr\{D_i^j > D_{\text{max}}^{i,j}\} = \eta_3 \exp(-\theta_i^j E_B^{i,j}(\theta_i^j) D_{\text{max}}^{i,j}), \forall i \in \mathcal{I}, j \in \mathcal{J}, \quad (7)$$

where D_i^j is the i^{th} user delay in RRH j and η_3 is the non-empty buffer probability. Equation (7) can be simplified to

$$\exp(-\theta_i^j E_B^{i,j}(\theta_i^j) D_{\text{max}}^{i,j}) = \exp(-\theta_i^j \lambda_i^j \frac{(e^{\theta_i^j} - 1)}{\theta_i^j} D_{\text{max}}^{i,j}) = \exp(-\lambda_i^j (e^{\theta_i^j} - 1) D_{\text{max}}^{i,j}) \leq \delta_3.$$

Therefore, we have

$$\text{C12: } \sum_{k_1 \in \mathcal{K}_1} r_{u,k_1}^{s,j, \text{DL}} \geq \frac{\ln(1/\delta_3)}{(e^{\theta_i^j} - 1) D_{\text{max}}^{i,j}}, \forall i \in \mathcal{I}, \forall j \in \mathcal{J}.$$

In order to avoid bit dropping, the output rate of queues must be greater than the input rate of queues. Therefore, we have two following constraints

$$\begin{aligned} \text{C13: } & \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{j \in \mathcal{J}} \sum_{k_1 \in \mathcal{K}_1} \tau_{u,k_1}^{s,j,\text{UL}} r_{u,k_1}^{s,j,\text{UL}} \leq \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} x_{k_2}^{j,\text{UL}} r_{k_2}^{j,\text{UL}}, \\ \text{C14: } & \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} x_{k_2}^{j,\text{DL}} r_{k_2}^{j,\text{DL}} \leq \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{j \in \mathcal{J}} \sum_{k_1 \in \mathcal{K}_1} \tau_{u,k_1}^{s,j,\text{DL}} r_{u,k_1}^{s,j,\text{DL}}. \end{aligned}$$

III. OPTIMIZATION PROBLEM FORMULATION

In this section, our aim is to allocate resources to minimize the overall power consumption in our setup by considering a bounded delay constraint to satisfy the E2E delay requirements. Based on the mentioned constraints C1-C14, the optimization problem can be written as

$$\min_{\mathbf{P}, \mathbf{T}, \mathbf{X}, \mathbf{D}} \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} x_{k_2}^{j,\text{UL}} p_{k_2}^{j,\text{UL}} + x_{k_2}^{j,\text{UL}} p_{k_2}^{j,\text{UL}} + \tau_{u,k_1}^{s,j,\text{DL}} p_{u,k_1}^{s,j,\text{DL}} + \tau_{u,k_1}^{s,j,\text{UL}} p_{u,k_1}^{s,j,\text{UL}} \quad (8)$$

s.t. : (C1)-(C14)

$$\text{C15: } \sum_{k_1 \in \mathcal{K}_1} \sum_{u \in \mathcal{I}_s} \sum_{j \in \mathcal{J}} \tau_{u,k_1}^{s,j,q} r_{u,k_1}^{s,j,q} \geq R_{\text{rsv}}^{s,q}, \forall s \in \mathcal{S}, q \in \mathcal{Q}.$$

The optimization variables in (8) are subcarrier and power allocation for different users in access and fronthaul as well as in both UL and DL where \mathbf{P} , \mathbf{T} , \mathbf{X} , and \mathbf{D} are the transmit power, the access subcarrier allocation, fronthaul subcarrier allocation, and delay vector for users, respectively. C15 is the rate constraint for isolation of slices. According to the delay and SIC constraints in the optimization problem, the one of the outputs of the optimization problem is the pair of NOMA users that satisfy SIC and delay constraints. In fact, the SIC constraint defines the conditions for executing SIC for two distinct users. If this condition is not met, each subcarrier is exclusively allocated to each user.

Constraints C1-C8 are linear functions of the optimization variables. In problem (8), the rate is a non-convex function, which leads to the non-convexity of the problem. In addition, this problem contains both discrete and continuous variables, which makes the problem more

challenging. Therefore, we resort to an alternate method to propose an efficient iterative algorithm [33], [34] with three sub-problems, namely, subcarrier allocation sub-problem, power allocation sub-problem, and delay adjustment sub-problem which will be explained in the followings.

IV. AN EFFICIENT ITERATIVE ALGORITHM

As mentioned earlier, to solve (8), we deploy an iterative algorithm that divides the problem into three sub-problems and solve them alternately [33], [34]. This procedure is presented in Algorithm.1. Let z be the iteration number and $\mathbf{P}^{(0)}$, $\mathbf{X}^{(0)}$, and $\mathbf{T}^{(0)}$ be the initial values. In each iteration, we solve each sub-problem with considering the optimization parameters of other sub-problems as fixed values derived in the previous steps. The iteration stops when the error in Step 5 is less than a predetermined threshold, i.e., ϵ_{TH} , or the number of iterations exceeds a predetermined value i.e., Z_{TH} . The solution of the last iteration is then declared as the solution of (8).

Proposition 1 : The presented iterative algorithm which is described in Algorithm.1 converges.

Proof: See Appendix A. ■

A. Subcarrier Allocation Sub-Problem

With assuming fixed value \mathbf{P} and \mathbf{D} , the subcarrier allocation sub-problem is written as follows

$$\min_{\mathbf{T}, \mathbf{X}} \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} x_{k_2}^{j, \text{UL}} p_{k_2}^{j, \text{UL}} + x_{k_2}^{j, \text{UL}} p_{k_2}^{j, \text{UL}} + \tau_{u, k_1}^{s, j, \text{DL}} p_{u, k_1}^{s, j, \text{DL}} + \tau_{u, k_1}^{s, j, \text{UL}} p_{u, k_1}^{s, j, \text{UL}} \quad (9)$$

$$\text{s.t. : (C1-C8), (C10-C15)}$$

While (9) has less computational complexity than (8), still it suffers from non-convexity due to the interference in rate functions. In addition, this problem contains discrete variables. We apply time sharing method and relax discrete variables as $x_{k_2}^{j, q} \in [0, 1], \forall k_2 \in \mathcal{K}_2, \forall j \in \mathcal{J}, \forall q \in \mathcal{Q}$ and

Algorithm 1 Three-Step Iterative Algorithm

Step 1: Initialization

$\mathcal{J} = \{1, \dots, J\}, \mathcal{K}_1 = \{1, \dots, K_1\}, \mathcal{K}_2 = \{1, \dots, K_2\}, \mathcal{I}_s = \{1, \dots, I_j\}, \mathcal{S} = \{1, \dots, S\}, \epsilon_{\text{TH}} = 10^{-4},$
 $Z_{\text{TH}} = 100$ and $z = 0$,

Calculate initial value $p^{(z)} = p^0, \tau^{(z)} = \tau^0$ and $x^{(z)} = x^0$ from Section (IV-D).

Step 2: Subcarrier Allocation: Allocate subcarrier by minimizing the transmit power and satisfying the problem constraints i.e, (9)

Step 3: Power Allocation: Allocate power to each user according to problem (11) and subcarrier allocated in Step 2.

Step 4: Delay Adjustment: Adjust delay of each user according to problem (13).

Step 5: Iteration $z = z + 1$, Repeat Step 2, 3 and 4 until $\|\mathbf{P}^{(z)} - \mathbf{P}^{(z-1)}\| \leq \epsilon_{\text{TH}}$ or $Z_{\text{TH}} < z$,

$\tau_{u,k_1}^{s,j,q} \in [0, 1], \forall u \in \mathcal{I}_s, \forall k_1 \in \mathcal{K}_1, \forall j \in \mathcal{J}, \forall q \in \mathcal{Q}$. To solve this problem, we use the difference of two convex (DC) functions to transform the problem into a convex form. The subcarrier

allocation sub-problem is transformed into the following form (See Appendix C):

$$\min_{\mathbf{T}, \mathbf{X}} \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} x_{k_2}^{j, \text{UL}} p_{k_2}^{j, \text{UL}} + x_{k_2}^{j, \text{UL}} p_{k_2}^{j, \text{UL}} + \tau_{u, k_1}^{s, j, \text{DL}} p_{u, k_1}^{s, j, \text{DL}} + \tau_{u, k_1}^{s, j, \text{UL}} p_{u, k_1}^{s, j, \text{UL}} \quad (10)$$

s.t. : (C1-C3), (C5-C7),

$$\begin{aligned} \bar{\text{C4:}} \quad & \tilde{I}_{i, k_1}^{s, j, q} h_{i, k_1}^{s, j, q^{-1}} + I_{i, k_1}^{s, j, q} h_{i, k_1}^{s, j, q^{-1}} + \sigma_{i, k_1}^{s, j, q} h_{i, k_1}^{s, j, q^{-1}} - \tilde{I}_{m', k_1}^{s, j, q} h_{m', k_1}^{s, j, q^{-1}} - \\ & I_{m', k_1}^{s, j, q} h_{m', k_1}^{s, j, q^{-1}} - \sigma_{m', k_1}^{s, j, q} h_{m', k_1}^{s, j, q^{-1}} \leq 0, \forall j, s, q, i, m' \in \mathcal{I}, m' \neq i, h_{i, k_1}^{s, j, q} > h_{m', k_1}^{s, j, q}, \\ \bar{\text{C8:}} \quad & \tilde{I}_{k_2}^{j, q} h_{k_2}^{j, q^{-1}} + \sigma_{k_2}^{j, q} h_{k_2}^{j, q^{-1}} - \tilde{I}_{k_2}^{j', q} h_{k_2}^{j', q^{-1}} - \sigma_{k_2}^{j', q} h_{k_2}^{j', q^{-1}} \leq 0, \forall q \in \mathcal{Q}, \forall j, j' \in \mathcal{J}, j' \neq j, h_{k_2}^{j, \mathcal{Q}} > h_{k_2}^{j', q}, \\ \bar{\text{C10:}} \quad & f_{\text{AC}}^{\text{UL}}(\mathbf{T}) - g_{\text{AC}}^{\text{UL}}(\mathbf{T}) \geq \frac{\ln(1/\delta_1)}{(e^{\theta_j} - 1)D_{\max}^j}, \forall j \in \mathcal{J}, \\ \bar{\text{C11:}} \quad & \hat{f}_{\text{AC}}^{\text{DL}}(\mathbf{T}) - \hat{g}_{\text{AC}}^{\text{DL}}(\mathbf{T}) \geq \frac{\ln(1/\delta_2)}{(e^{\theta_{\text{BBU}}} - 1)D_{\max}^{\text{BBU}}}, \\ \bar{\text{C12:}} \quad & \hat{f}_{\text{FH}}^{\text{UL}}(\mathbf{X}) - \hat{g}_{\text{FH}}^{\text{UL}}(\mathbf{X}) \geq \frac{\ln(1/\delta_3)}{(e^{\theta_i^j} - 1)D_{\max}^{i, j}}, \forall i \in \mathcal{I}, \forall j \in \mathcal{J}, \\ \bar{\text{C13:}} \quad & f_{\text{AC}}^{\text{UL}}(\mathbf{T}) + \tilde{f}_{\text{FH}}^{\text{UL}}(\mathbf{X}) - \tilde{g}_{\text{AC}}^{\text{UL}}(\mathbf{T}) - g_{\text{FH}}^{\text{UL}}(\mathbf{X}) \geq 0, \\ \bar{\text{C14:}} \quad & \tilde{f}_{\text{AC}}^{\text{DL}}(\mathbf{T}) + f_{\text{FH}}^{\text{DL}}(\mathbf{X}) - g_{\text{AC}}^{\text{DL}}(\mathbf{T}) - \tilde{g}_{\text{FH}}^{\text{DL}}(\mathbf{X}) \geq 0, \\ \bar{\text{C15:}} \quad & \tilde{f}_1(\mathbf{T}) - \tilde{g}_1(\mathbf{T}) \geq 0. \end{aligned}$$

The above problem is a convex problem and can be solved with the CVX toolbox in Matlab [35], [36].

Proposition 3: The proposed iterative algorithm based on the SCA method for subcarrier allocation step converges.

Proof: See Appendix B by considering fixed value for power (\mathbf{P}). ■

B. Power Allocation Sub-Problem

For the fixed value of \mathbf{T} , \mathbf{X} and \mathbf{D} , the power allocation sub-problem is obtained as follows

$$\min_{\mathbf{P}} \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} x_{k_2}^{j, \text{UL}} p_{k_2}^{j, \text{UL}} + x_{k_2}^{j, \text{UL}} p_{k_2}^{j, \text{UL}} + \tau_{u, k_1}^{s, j, \text{DL}} p_{u, k_1}^{s, j, \text{DL}} + \tau_{u, k_1}^{s, j, \text{UL}} p_{u, k_1}^{s, j, \text{UL}} \quad (11)$$

$$\text{s.t. : (C2)-(C4), (C6)-(C8), (C9)-(C15)}$$

Similar to the subcarrier allocation sub-problem, in problem (11), the rate is a non-convex function, which leads to the non-convexity of the problem. Therefore, it is necessary to approximate (11) with a convex problem. To solve this problem, we use the DC approximation to transform the problem into a convex form. Therefore, the power allocation sub-problem is transformed as follows (See Appendix C):

$$\min_{\mathbf{P}} \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} x_{k_2}^{j, \text{UL}} p_{k_2}^{j, \text{UL}} + x_{k_2}^{j, \text{UL}} p_{k_2}^{j, \text{UL}} + \tau_{u, k_1}^{s, j, \text{DL}} p_{u, k_1}^{s, j, \text{DL}} + \tau_{u, k_1}^{s, j, \text{UL}} p_{u, k_1}^{s, j, \text{UL}} \quad (12)$$

$$\text{s.t. : (C2-C3), (C6-C7),}$$

$$\tilde{\text{C4}}: \tilde{I}_{i, k_1}^{s, j, q} h_{i, k_1}^{s, j, q^{-1}} + I_{i, k_1}^{s, j, q} h_{i, k_1}^{s, j, q^{-1}} + \sigma_{i, k_1}^{s, j, q} h_{i, k_1}^{s, j, q^{-1}} - \tilde{I}_{m', k_1}^{s, j, q} h_{m', k_1}^{s, j, q^{-1}} - I_{m', k_1}^{s, j, q} h_{m', k_1}^{s, j, q^{-1}} - \sigma_{m', k_1}^{s, j, q} h_{m', k_1}^{s, j, q^{-1}} \leq 0,$$

$$\forall j, \forall s, q \in \mathcal{Q}, \forall i, m' \in \mathcal{I}, m' \neq i, h_{i, k_1}^{s, j, q} > h_{m', k_1}^{s, j, q},$$

$$\tilde{\text{C8}}: \tilde{I}_{k_2}^{j, q} h_{k_2}^{j, q^{-1}} + \sigma_{k_2}^{j, q} h_{k_2}^{j, q^{-1}} - \tilde{I}_{k_2}^{j', q} h_{k_2}^{j', q^{-1}} - \sigma_{k_2}^{j', q} h_{k_2}^{j', q^{-1}} \leq 0, \forall q, j, j' \in \mathcal{J}, j' \neq j, h_{k_2}^{j, q} > h_{k_2}^{j', q},$$

$$\tilde{\text{C10}}: f_{\text{AC}}^{\text{UL}}(\mathbf{P}) - g_{\text{AC}}^{\text{UL}}(\mathbf{P}) \geq \frac{\ln(1/\delta_1)}{(e^{\theta_j} - 1) D_{\max}^j}, \forall j \in \mathcal{J},$$

$$\tilde{\text{C11}}: \hat{f}_{\text{AC}}^{\text{DL}}(\mathbf{P}) - \hat{g}_{\text{AC}}^{\text{DL}}(\mathbf{P}) \geq \frac{\ln(1/\delta_2)}{(e^{\theta_{\text{BBU}}} - 1) D_{\max}^{\text{BBU}}},$$

$$\tilde{\text{C12}}: \hat{f}_{\text{FH}}^{\text{UL}}(\mathbf{P}) - \hat{g}_{\text{FH}}^{\text{UL}}(\mathbf{P}) \geq \frac{\ln(1/\delta_3)}{(e^{\theta_i} - 1) D_{\max}^{i, j}}, \forall i \in \mathcal{I}, \forall j \in \mathcal{J},$$

$$\tilde{\text{C13}}: f_{\text{AC}}^{\text{UL}}(\mathbf{P}) + \tilde{f}_{\text{FH}}^{\text{UL}}(\mathbf{P}) - \tilde{g}_{\text{AC}}^{\text{UL}}(\mathbf{P}) - g_{\text{FH}}^{\text{UL}}(\mathbf{P}) \geq 0,$$

$$\tilde{\text{C14}}: \tilde{f}_{\text{AC}}^{\text{DL}}(\mathbf{P}) + f_{\text{FH}}^{\text{DL}}(\mathbf{P}) - g_{\text{AC}}^{\text{DL}}(\mathbf{P}) - \tilde{g}_{\text{FH}}^{\text{DL}}(\mathbf{P}) \geq 0,$$

$$\tilde{\text{C15}}: \tilde{f}_1(\mathbf{P}) - \tilde{g}_1(\mathbf{P}) \geq 0.$$

Similar to the subcarrier allocation sub-problem, the above problem is a convex problem and can be solved with the CVX toolbox in Matlab [35], [36].

Proposition 3: The proposed iterative algorithm based on the SCA method for power allocation subproblem converges.

Proof: See Appendix B by considering fixed values for subcarrier allocation parameters i.e, (T) and (X). ■

C. Delay Adjustment Sub-Problem

With assuming fixed values of P , T , and X , the delay adjustment sub-problem is obtained as

$$\text{find } D \quad (13)$$

$$\text{s.t. : C9-C12.}$$

The delay adjustment sub-problem can be solved by the linear programming (LP) of any optimization toolbox.

D. Calculate Initial Values

Due to complex nature of (8), and specially having C10, C11, C12, and C15, obtaining feasible initial values for Algorithm.1 is not trivial. Therefore, to find the initial values, we propose to solve the following optimization problem

$$\max_{\mathbf{P}, \mathbf{X}, \mathbf{T}} \sum_{j \in \mathcal{J}} \sum_{k_2 \in \mathcal{K}_2} \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{I}_s} \sum_{k_1 \in \mathcal{K}_1} (x_{k_2}^{j, \text{UL}} r_{k_2}^{j, \text{UL}} + \tau_{u, k_1}^{s, j, \text{UL}} r_{u, k_1}^{s, j, \text{UL}}) + (\tau_{u, k_1}^{s, j, \text{DL}} r_{u, k_1}^{s, j, \text{DL}} + x_{k_2}^{j, \text{UL}} r_{k_2}^{j, \text{UL}}) \quad (14)$$

$$\text{s.t. : (C1-C8), (C13-C14),}$$

which does not include C10, C11, C12, and C15, and the objective function is to maximize the total access and fronthaul rates. (14) is also non-convex function and we utilize DC approximation similar to Algorithm.1, except that we do not have C10, C11, C12, and C15. To solve (14), we set all initial values to zero which is a feasible point of (14). After deriving the solution of (14), we check out if the constraints C10, C11, C12, C15 and C9 hold or not. If these constraints

hold, the derived solution of (14) is an initial value of (8); otherwise, we consider that (8) is infeasible.

E. Computational Complexity

For subcarrier allocation and power allocation sub-problems, the number of required iterations for the DC approximation is $\frac{\log C}{\xi} \frac{t^0 \varrho}{\xi}$, where $0 \leq \varrho \leq \infty$ is the stopping criterion for the interior point method (IPM), ξ is used to update accuracy of the IPM, t^0 is initial point for approximating the accuracy of IPM, and C is the total number of constraints. For the subcarrier allocation sub-problem the total number of constraints is shown by C_{Sub} which is $C_{\text{Sub}} = 2JK_1 + 2K_2 + J + I + 2I^2JS + 2J^2 + JI + 2S + 4 - 2JSI$ [34]. Similarly, for the power allocation sub-problem the total number of constraints is shown by C_{Pow} which is $C_{\text{Pow}} = J + I + 2JSI^2 + 2J^2 + JI + 2S - 2JSI + 4$. For the delay adjustment sub-problem the total number of constraints is shown by C_{Delay} which is calculated as $C_{\text{Delay}} = J + 2JI + 1$. For instance, in subcarrier allocation subproblem and power allocation subproblem, the number of RRHs and the number of users have a significant impact on complexity. Moreover, the subcarrier allocation subproblem has more complexity than other subproblems. In contrast, the delay assignment subproblem has lower complexity than the others.

V. SIMULATION AND RESULTS

In this section, the simulation results are presented to evaluate the performance of the proposed system model. To simulate dense urban area, we consider a BBU is at the center of the coverage area whose distance is 1 Km from a set of RRHs. The coverage area is considered 10 square Kilometers. Moreover, we consider a Rayleigh fading wireless channel in which the subcarrier gains are independent. Channel power gains for the access links are set as $h_{i,k_1}^{s,j,q} = \Omega_{i,k_1}^{s,j,q} d_i^{j,q-\alpha}$ where $d_i^{j,q}$ is the distance between user i and RRH j , $\Omega_{i,k_1}^{s,j,q}$ is a random variable which is generated by Rayleigh distribution, and $\alpha = 3$ is the path-loss exponent. Channel power gains

for fronthaul links are set to $h_{k_2}^{j,q} = \Omega_{k_2}^{j,q} d_{j,q}^{-\beta}$ where similar to access links, the $d_{j,q}$ is the distance between RRH j and BBU, $\Omega_{k_2}^{j,q}$ is a random variable generated according to the Rayleigh distribution, and $\beta = 3$ is the path-loss exponent. The power spectral density (PSD) of the received Gaussian noise is set to -174 dBm/Hz. At each RRH, we set $P_{\text{RRH}_j}^{\text{DL}} = 43$ dBm $\forall j \in \mathcal{J}$ and $P_{\text{RRH}_j}^{\text{UL}} = 43$ dBm $\forall j \in \mathcal{J}$. For the BBU, we set $P_{\text{BBU}}^{\text{DL}} = 47$ dBm, and for each user, we set $P_{\text{USER}_i}^{\text{UL}} = 18$ dBm. The frequency bandwidth of wireless access and fronthaul links are $W_{\text{AC}} = 5$ MHz and $W_{\text{FH}} = 10$ MHz, respectively. Moreover, the bandwidth of each subcarrier is $W_S = 625 \times 10^3$ Hz. The QoS exponent is $\theta = 11$ [34]. Unless otherwise stated, we consider 2 RRHs, 2 users in each RRH and 2 slices. In this section, we use Monte-Carlo method for simulation where the optimization problem is solved for 1000 channel realizations and the total transmit power is the average value over all derived solutions.

A. PD-NOMA versus OFDMA for Tactile Internet in Our Setup

We initialize our study with comparing the performance of PD-NOMA and OFDMA for Tactile Internet in the proposed problem. For OFDMA-based system, problem (8) by considering $L1 = L2 = 1$ is solved and simulated which consists of both dynamic power allocation problem, subcarrier allocation subproblem, and delay adjustment problem. Therefore, for OFDMA-based system the set of variables is similar to that in NOMA-based system. Fig. 3 shows the total transmit power versus the number of users per slice in each RRH for these two cases. As expected, the total transmit power increases by increasing the number of users per slice in each RRH. This is due to the fact that each user has its own QoS determined by the corresponding delay requirement. Moreover, the total transmit power increases by increasing the value of reservation rate R_{rsv} .

As shown in this figure, the NOMA-based system outperforms the OFDM-based system. From Fig. 3, increasing the number of users can significantly increase the total transmit power in the OFDM-based system, e.g., the transmit power increases around 25% per each slice in each

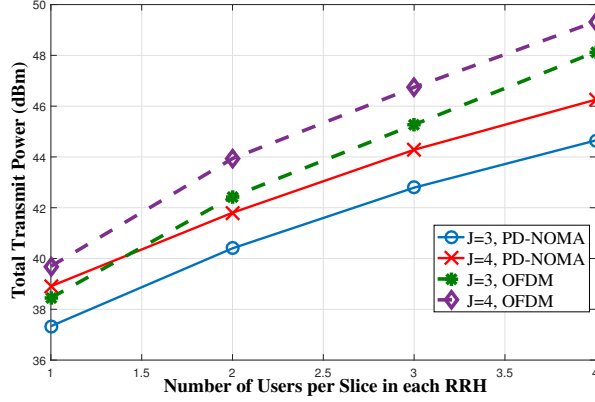


Fig. 3: Comparing PD-NOMA and OFDMA versus number of users per slice in each RRH

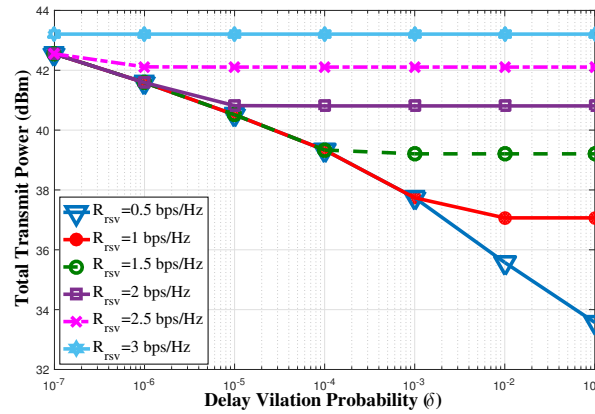


Fig. 4: Total transmit power versus delay violation probability

RRH. This simulation reveals the efficiency of NOMA compared to OFDMA for the tactile uses. Therefore, in the following figures, we only demonstrate NOMA-based simulation results.

B. Effects of Network Parameters

Now, here, we study the effects of other network parameters on the performance of the proposed algorithm based on PD-NOMA assisted C-RAN.

In Fig. 4, the total transmit power versus different delay violation probabilities is shown. As can be seen, decreasing the allowed delay violation probability can significantly increase the total transmit power. As an example, for reservation rate of 1bps/Hz, a second order of magnitude

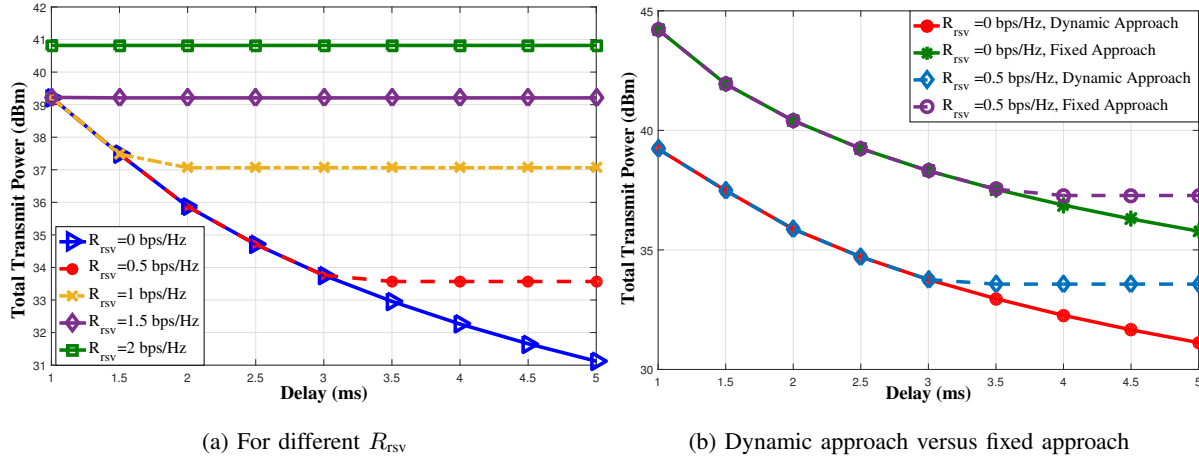


Fig. 5: Total transmit power versus delay

decrease in the violation probability requires almost twice more transmit power. However, at certain point, due to the reservation rate is dominant, the required transmit power becomes almost independent from the violation probability. Such an independence is more pronounced for larger values of reservation rate. By increasing the delay violation probability, the equivalent rate of the delay reduces and may become less than the reservation rate for slices.

To evaluate the performance of the proposed system model on a comparison basis, we consider a different scenario: in our proposed system model, we adjust the delay dynamically to minimize the transmit power which is called dynamic approach. In the new scenario which is called fixed approach, we assume the delay constraints are fixed and cannot be adjusted for access and fronthaul links. In this case, we have a new optimization problem, referred to as, the relaxed problem, in which we remove the delay variables from problem (8) and ignore constraint C9. Moreover, we set access and fronthaul delay manually in constraints C10, C11, and C12 as $D_{\text{max}}^j = D_{\text{max}}^{\text{BBU}} = D_{\text{max}}^{i,j} = D_{\text{max}}^{\text{Total}}/3$. The new problem can be solved with the DC approximation.

In Fig. 5a, we investigate the effect of the actual value of the delay $D_{\text{max}}^{\text{Total}}$ on the total transmit power. As can be seen, for smaller values of reservation rate, every 1 ms decrease in delay

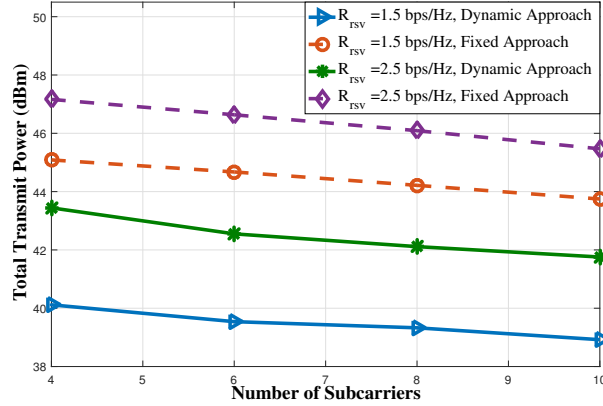


Fig. 6: Total transmit power versus number of subcarriers

requires 2 to 3 dB increase in the total transmit power. However, for larger values of reservation rate, the total transmit power is almost independent from the delay. In addition, it can be seen that by ignoring the reservation rate, the total transmit power decreases when delay increases. Moreover, from Fig. 5b, the proposed system model has a considerably better performance than the system model corresponding to the relaxed problem. As can be seen, by the dynamic adjustment of the delay, we can save around 5 dBm in transmit power.

In Fig. 6, the total transmit power versus the number of subcarrier for $D_{\text{max}}^{\text{Total}} = 1\text{ms}$. Fig. 6 is demonstrated where total transmit power decreases by increasing the number of subcarriers. This is due to the fact that increasing the number of subcarriers increases the diversity gain of the system. Moreover, the total transmit power increases by increasing the reservation rate for each slice. Moreover, from Fig. 6, up to 5 dBm transmit power can be saved.

C. Convergence Study of Algorithm.1

To study the proof the convergence of proposed system model in Propositions 1-3, in Fig. 7, the convergence of the alternate method for the proposed system model is demonstrated. It can be seen that the solution of the proposed algorithm converges to the fixed values after 15 iterations. For this simulation, we set $R_{\text{rsv}} = 1.5$ bps/Hz.

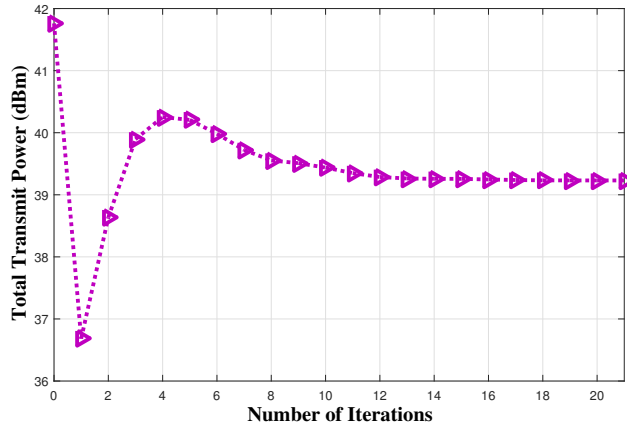


Fig. 7: Convergence of the algorithm

VI. CONCLUSION

In this paper, we propose a novel queuing model for the Tactile Internet services in PD-NOMA-based cloud radio access networks (C-RANs) serving several pairs of tactile users. For each pair of tactile users within C-RAN coverage area, our setup includes radio remote head (RRH) and baseband processing unit (BBU) queuing delays in one end to end (E2E) connection which is a more practical scenario in this context compared to previous works. We propose a resource allocation problem to minimize the transmit power by considering E2E delay of joint access and fronthaul links for each pair of tactile users where the delays of fronthaul and access links are dynamically adjusted to minimize the transmit power. To solve the highly non-convex proposed resource allocation problem, we applied successive convex approximation (SCA) method. Simulation results reveal that by dynamic adjustment of the access and fronthaul delays, transmit power can be considerably saved compared to the case of fixed approach per each transmission section.

APPENDIX A

In this algorithm, the convergence can be guaranteed if we can show that the objective function is a decreasing function with respect to the number of iterations. For the algorithm in Table.1,

in the first step of iteration $i + 1$, with a given power allocation at iteration i , $\mathbf{x} = \mathbf{x}^{(i+1)}$ and $\boldsymbol{\tau} = \boldsymbol{\tau}^{(i+1)}$ are derived. Based on DC approximation, we will have $f(\mathbf{x}^{(i)}, \mathbf{p}^{(i)}) \leq f(\mathbf{x}^{(i+1)}, \mathbf{p}^{(i)})$ and $f(\boldsymbol{\tau}^{(i)}, \mathbf{p}^{(i)}) \leq f(\boldsymbol{\tau}^{(i+1)}, \mathbf{p}^{(i)})$ [34]. In the second step, with a given subcarrier allocation at iteration $i+1$, the power allocation at iteration $(i+1)$ is obtained. Based on DC approximation, we will have $f(\mathbf{x}^{(i+1)}, \mathbf{p}^{(i)}) \leq f(\mathbf{x}^{(i+1)}, \mathbf{p}^{(i+1)})$ and $f(\boldsymbol{\tau}^{(i+1)}, \mathbf{p}^{(i)}) \leq f(\boldsymbol{\tau}^{(i+1)}, \mathbf{p}^{(i+1)})$. Finally, we have $\dots \leq f(\mathbf{x}^{(i)}, \mathbf{p}^{(i)}) \leq f(\mathbf{x}^{(i+1)}, \mathbf{p}^{(i)}) \leq f(\mathbf{x}^{(i+1)}, \mathbf{p}^{(i+1)}) \leq \dots \leq f(\mathbf{x}^*, \mathbf{p}^*)$, and $\dots \leq f(\boldsymbol{\tau}^{(i)}, \mathbf{p}^{(i)}) \leq f(\boldsymbol{\tau}^{(i+1)}, \mathbf{p}^{(i)}) \leq f(\boldsymbol{\tau}^{(i+1)}, \mathbf{p}^{(i+1)}) \leq \dots \leq f(\boldsymbol{\tau}^*, \mathbf{p}^*)$, where $\mathbf{x}^*, \boldsymbol{\tau}^*$ and \mathbf{p}^* are optimal solutions which are obtained in the previous iteration. After each iteration, we have $f(\boldsymbol{\tau}^{(i+1)}, \mathbf{p}^{(i+1)}) - f(\boldsymbol{\tau}^{(i)}, \mathbf{p}^{(i)})$ and $f(\mathbf{x}^{(i+1)}, \mathbf{p}^{(i+1)}) - f(\mathbf{x}^{(i)}, \mathbf{p}^{(i)})$ which is a decreasing function and consequently the proposed algorithm converges.

APPENDIX B

We approximate the rate function with the DC approximation as $r(\mathbf{P}, \mathbf{x}) = f(\mathbf{P}, \mathbf{x}) - g(\mathbf{P}, \mathbf{x})$. In each iteration for each sub-problem, $f(\mathbf{P}, \mathbf{x})$ and $g(\mathbf{P}, \boldsymbol{\tau})$ are single variable functions. With this approximation the rate function becomes a concave function. Thus, by using this method, the non-convex problem can be converted into a convex problem [37]. Given the fact that the functions in each iteration for each sub-problem are single-variable, we show the functions $f(\mathbf{P}, \mathbf{x})$ and $g(\mathbf{P}, \boldsymbol{\tau})$ as a function $\nu(\boldsymbol{\rho})$, where according to the subproblem $\boldsymbol{\rho}$ can be \mathbf{P} or \mathbf{x} or $\boldsymbol{\tau}$ for simplicity. Therefore, we have $\nu(\boldsymbol{\rho}^{(i)}) \leq \nu(\boldsymbol{\rho}^{(i-1)}) + \nabla \nu(\boldsymbol{\rho}^{(i-1)})(\boldsymbol{\rho}^{(i)} - \boldsymbol{\rho}^{(i-1)})$. Consequently, from [34], for iteration i , we have $f(\boldsymbol{\rho}^{(i)}) - \{g(\boldsymbol{\rho}^{(i-1)}) + \nabla g(\boldsymbol{\rho}^{(i-1)})(\boldsymbol{\rho}^{(i)} - \boldsymbol{\rho}^{(i-1)})\} \geq R_0$. Moreover, we have $f(\boldsymbol{\rho}^{(i+1)}) - g(\boldsymbol{\rho}^{(i+1)}) \geq f(\boldsymbol{\rho}^{(i)}) - g(\boldsymbol{\rho}^{(i)}) - \nabla g(\boldsymbol{\rho}^{(i)})(\boldsymbol{\rho}^{(i+1)} - \boldsymbol{\rho}^{(i)}) \geq f(\boldsymbol{\rho}^{(i)}) - g(\boldsymbol{\rho}^{(i)})$ [37]. In other words, after each iteration, the new solution is always closer to optimum than solution of previous iteration. Therefore, the SCA with the DC approximation converges to a sub-optimal solution.

APPENDIX C

First, we transform the access rate into a convex function by using the DC approximation as

$$\begin{aligned}
\sum_{s \in S} \sum_{u \in I_s} \sum_{j \in J} \sum_{k_1 \in \mathcal{K}_1} \tau_{u,k_1}^{s,j,q} r_{u,k_1}^{s,j,q} &= \sum_{j \in J} \sum_{s \in S} \sum_{u \in I_s} \sum_{k_1 \in \mathcal{K}_1} \log_2(1 + \tau_{u,k_1}^{s,j,q} \gamma_{u,k_1}^{s,j,q}) = \\
&\sum_{j \in J} \sum_{s \in S} \sum_{u \in I_s} \sum_{k_1 \in \mathcal{K}_1} \log_2\left(1 + \frac{\tau_{u,k_1}^{s,j,q} p_{u,k_1}^{s,j,q} h_{i,k_1}^{s,j,q}}{\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}}\right) = \\
&\sum_{j \in J} \sum_{s \in S} \sum_{u \in I_s} \sum_{k_1 \in \mathcal{K}_1} \log_2\left(\frac{\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q} + \tau_{u,k_1}^{s,j,q} p_{u,k_1}^{s,j,q} h_{i,k_1}^{s,j,q}}{\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}}\right) = \\
&\sum_{j \in J} \sum_{s \in S} \sum_{u \in I_s} \sum_{k_1 \in \mathcal{K}_1} \log_2(\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q} + \tau_{u,k_1}^{s,j,q} p_{u,k_1}^{s,j,q} h_{i,k_1}^{s,j,q}) - \log_2(\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}).
\end{aligned}$$

We present the access rate in a DC approximation as $\tau_{u,k_1}^{s,j,q} r_{u,k_1}^{s,j,q} = f_{AC}^q(\mathbf{P}, \mathbf{T}) - g_{AC}^q(\mathbf{P}, \mathbf{T})$ where $f_{AC}^q(\mathbf{P}, \mathbf{T},) = \log_2(\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q} + \tau_{u,k_1}^{s,j,q} p_{u,k_1}^{s,j,q} h_{i,k_1}^{s,j,q})$, and $g_{AC}^q(\mathbf{P}, \mathbf{T},) = \log_2(\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q})$ are concave functions. Then, we use $g_{AC}^q(\mathbf{P}, \mathbf{T}) \approx g_{AC}^q(\mathbf{P}, \mathbf{T})^{(t_I-1)} + \nabla g_{AC}^q(\mathbf{P}, \mathbf{T})^{(t_I-1)}((\mathbf{P}, \mathbf{T}) - (\mathbf{P}, \mathbf{T})^{(t_I-1)})$ where

$$\nabla g_{AC}^q(P_{u,k_1}^{s,j,q}, \tau_{u,k_1}^{s,j,q}) = \begin{cases} 0, & \text{if } h_{i,k_1}^{s,j,q} < h_{u,k_1}^{s,j,q} \text{ and } m = j, \\ \frac{h_{u,k_1}^{s,m,q} p_{u,k_1}^{s,j,q}}{\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}}, & \text{if } m \neq j, \\ \frac{h_{i,k_1}^{s,j,q} p_{u,k_1}^{s,j,q}}{\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}}, & \text{if } h_{i,k_1}^{s,j,q} > h_{u,k_1}^{s,j,q} \text{ and } m = j \text{ and } u \neq i. \end{cases}$$

Now, we transform the fronthaul rate into a convex function by using the DC approximation as

$$\begin{aligned}
\sum_{j \in J} \sum_{k_2 \in \mathcal{K}_2} x_{k_2}^{j,q} r_{k_2}^{j,q} &= \sum_{j \in J} \sum_{k_2 \in \mathcal{K}_2} \log_2(1 + x_{k_2}^{j,q} \gamma_{k_2}^{j,q}) = \sum_{j \in J} \sum_{k_2 \in \mathcal{K}_2} \log_2\left(1 + \frac{x_{k_2}^{j,q} p_{k_2}^{j,q} h_{i,k_2}^{j,q}}{\sigma_{k_2}^{j,q} + I_{k_2}^{j,q}}\right) \\
&= \sum_{j \in J} \sum_{k_2 \in \mathcal{K}_2} \log_2\left(\frac{\sigma_{k_2}^{j,q} + I_{k_2}^{j,q} + \tilde{I}_{k_2}^{j,q} + x_{k_2}^{j,q} p_{k_2}^{j,q} h_{i,k_2}^{j,q}}{\sigma_{k_2}^{j,q} + I_{k_2}^{j,q}}\right) \\
&= \sum_{j \in J} \sum_{k_2 \in \mathcal{K}_2} \log_2(\sigma_{k_2}^{j,q} + I_{k_2}^{j,q} + \tilde{I}_{k_2}^{j,q} + x_{k_2}^{j,q} p_{k_2}^{j,q} h_{i,k_2}^{j,q}) - \log_2(\sigma_{k_2}^{j,q} + I_{k_2}^{j,q}).
\end{aligned}$$

We present the access rate in a DC form as $x_{k_2}^{j,q} r_{k_2}^{j,q} = f_{FH}^q(\mathbf{P}, \mathbf{X}) - g_{FH}^q(\mathbf{P}, \mathbf{X})$, where $f_{FH}^q(\mathbf{P}, \mathbf{X})$ and $g_{FH}^q(\mathbf{P}, \mathbf{X})$ are concave functions as $f_{FH}^q(\mathbf{P}, \mathbf{X}) = \log_2(\sigma_{k_2}^{j,q} + I_{k_2}^{j,q} + \tilde{I}_{k_2}^{j,q} + x_{k_2}^{j,q} p_{k_2}^{j,q} h_{i,k_2}^{j,q})$ and $g_{FH}^q(\mathbf{P}, \mathbf{X}) = \log_2(\sigma_{k_2}^{j,q} + I_{k_2}^{j,q} + \tilde{I}_{k_2}^{j,q})$.

Then, we deploy $g_{\text{FH}}^q(\mathbf{P}\mathbf{X}) \approx g_{\text{FH}}^q(\mathbf{P}, \mathbf{X})^{(t_l-1)} + \nabla g_{\text{FH}}^q(\mathbf{P}, \mathbf{X})^{(t_l-1)}((\mathbf{P}, \mathbf{X}) - (\mathbf{P}, \mathbf{X})^{(t_l-1)})$,

where

$$\nabla g_{\text{FH}}^q(P_{k_2}^{j,q}, x_{k_2}^{j,q}) = \begin{cases} 0, & \text{If } h_{k_2}^{j,q} < h_{k_2}^{m,q} \text{ and } m = j, \\ \frac{h_{k_2}^{s,m,q} p_{k_2}^{j,q}}{\sigma_{k_2}^{j,q} + I_{k_2}^{j,q}}, & \text{Otherwise.} \end{cases}$$

C4 and C8 are SIC constraints and are transformed into the linear functions. For C4, we have

$$\begin{aligned} & \tilde{I}_{i,k_1}^{s,j,q} (h_{i,k_1}^{s,j,q})^{-1} + I_{i,k_1}^{s,j,q} (h_{i,k_1}^{s,j,q})^{-1} + \sigma_{i,k_1}^{s,j,q} (h_{i,k_1}^{s,j,q})^{-1} - \\ & \tilde{I}_{m',k_1}^{s,j,q} (h_{m',k_1}^{s,j,q})^{-1} - I_{m',k_1}^{s,j,q} (h_{m',k_1}^{s,j,q})^{-1} - \sigma_{m',k_1}^{s,j,q} (h_{m',k_1}^{s,j,q})^{-1} \leq 0. \end{aligned}$$

Hence, C4 becomes a linear function. Similarly, C8 can be rewritten as

$$I_{k_2}^{j,q} (h_{k_2}^{j,q})^{-1} + \sigma_{k_2}^{j,q} (h_{k_2}^{j,q})^{-1} - I_{k_2}^{j',q} (h_{k_2}^{j',q})^{-1} - \sigma_{k_2}^{j',q} (h_{k_2}^{j',q})^{-1} \leq 0.$$

C13 and C14 are functions of the rate???. Therefore, by the DC approximation, C13 is transformed into convex function as

$$\begin{aligned} & \sum_{j \in J} \sum_{s \in S_j} \sum_{u \in I_j} \sum_{k_1 \in \mathcal{K}_1} \sum_{k_2 \in \mathcal{K}_2} \log_2(\sigma_{u,k_1}^{s,j,\text{UL}} + I_{u,k_1}^{s,j,\text{UL}} + \tilde{I}_{u,k_1}^{s,j,q}) + \log_2(\sigma_{k_2}^{j,\text{UL}} + I_{k_2}^{j,\text{UL}} + x_{k_2}^{j,\text{UL}} p_{k_2}^{j,\text{UL}} h_{k_2}^{j,\text{UL}}) \\ & - \log_2(\sigma_{u,k_1}^{s,j,\text{UL}} + I_{u,k_1}^{s,j,\text{UL}} + \tilde{I}_{u,k_1}^{s,j,\text{UL}} + \tau_{u,k_1}^{s,j,\text{UL}} p_{u,k_1}^{s,j,\text{UL}} h_{i,k_1}^{s,j,\text{UL}}) - \log_2(\sigma_{k_2}^{j,\text{UL}} + I_{k_2}^{j,\text{UL}}) = \\ & f_{\text{AC}}^{\text{UL}}(\mathbf{P}, \mathbf{T}) + \tilde{f}_{\text{FH}}^{\text{UL}}(\mathbf{P}, \mathbf{X}) - \tilde{g}_{\text{AC}}^{\text{UL}}(\mathbf{P}, \mathbf{T}) - g_{\text{FH}}^{\text{UL}}(\mathbf{P}, \mathbf{X}), \end{aligned}$$

where $\tilde{f}_{\text{FH}}^{\text{UL}}(\mathbf{P}, \mathbf{X})$ and $\tilde{g}_{\text{AC}}^{\text{UL}}(\mathbf{P}, \mathbf{T})$ are

$$\tilde{f}_{\text{FH}}^{\text{UL}}(\mathbf{P}, \mathbf{X}) = \sum_{j \in J} \sum_{k_2 \in \mathcal{K}_2} \log_2(\sigma_{k_2}^{j,\text{UL}} + I_{k_2}^{s,j,\text{UL}} + x_{k_2}^{j,\text{UL}} p_{k_2}^{j,\text{UL}} h_{k_2}^{j,\text{UL}}),$$

and

$$\tilde{g}_{\text{AC}}^{\text{UL}}(\mathbf{P}, \mathbf{T}) = \sum_{j \in J} \sum_{s \in S_j} \sum_{u \in I_j} \sum_{k_1 \in \mathcal{K}_1} \log_2(\sigma_{u,k_1}^{s,j,\text{UL}} + I_{u,k_1}^{s,j,\text{UL}} + \tilde{I}_{u,k_1}^{s,j,\text{UL}} + \tau_{u,k_1}^{s,j,\text{UL}} p_{u,k_1}^{s,j,\text{UL}} h_{i,k_1}^{s,j,\text{UL}}),$$

in which

$$\nabla \tilde{g}_{\text{AC}}^{\text{UL}}(P_{u,k_1}^{s,j,\text{UL}}, \tau_{u,k_1}^{s,j,\text{UL}}) = \begin{cases} 0 & \text{if } h_{i,k_1}^{s,j,\text{UL}} < h_{u,k_1}^{s,j,\text{UL}} \text{ and } m = j, \\ \frac{h_{u,k_1}^{s,m,\text{UL}} p_{u,k_1}^{s,m,\text{UL}}}{\sigma_{u,k_1}^{s,j,\text{UL}} + I_{u,k_1}^{s,j,\text{UL}} + \tilde{I}_{u,k_1}^{s,j,\text{UL}}} & \text{if } m \neq j, \\ \frac{h_{i,k_1}^{s,j,\text{UL}} p_{u,k_1}^{s,m,\text{UL}}}{\sigma_{u,k_1}^{s,j,\text{UL}} + I_{u,k_1}^{s,j,\text{UL}} + \tilde{I}_{u,k_1}^{s,j,\text{UL}}} & \text{if } h_{i,k_1}^{s,j,\text{UL}} > h_{u,k_1}^{s,j,\text{UL}} \text{ and } m = j. \end{cases}$$

Hence, C13 becomes a convex constraint via the DC approximation. Similarly, for C14, we have

$$\begin{aligned} & \sum_{j \in J} \sum_{s \in S_j} \sum_{u \in I_j} \sum_{k_1 \in \mathcal{K}_1} \sum_{k_2 \in \mathcal{K}_2} \log_2(\sigma_{u,k_1}^{s,j,\text{DL}} + I_{u,k_1}^{s,j,\text{DL}} + \tilde{I}_{u,k_1}^{s,j,\text{DL}} + \tau_{u,k_1}^{s,j,\text{DL}} p_{u,k_1}^{s,j,\text{DL}} h_{i,k_1}^{s,j,\text{DL}}) + \log_2(\sigma_{k_2}^{j,\text{DL}} + I_{k_2}^{j,\text{DL}}) \\ & - \log_2(\sigma_{u,k_1}^{s,j,\text{DL}} + I_{u,k_1}^{s,j,\text{DL}} + \tilde{I}_{u,k_1}^{s,j,\text{DL}}) - \log_2(\sigma_{k_2}^{j,\text{DL}} + I_{k_2}^{j,\text{DL}} + x_{k_2}^{j,\text{DL}} p_{k_2}^{j,\text{DL}} h_{k_2}^{j,\text{DL}}) \\ & = \tilde{f}_{\text{AC}}^{\text{DL}}(\mathbf{P}, \mathbf{T}) + f_{\text{FH}}^{\text{DL}}(\mathbf{P}, \mathbf{X}) - g_{\text{AC}}^{\text{DL}}(\mathbf{P}, \mathbf{T}) - \tilde{g}_{\text{FH}}^{\text{DL}}(\mathbf{P}, \mathbf{X}), \end{aligned}$$

where $\tilde{f}_{\text{AC}}^{\text{DL}}(\mathbf{P}, \mathbf{T})$ and $\tilde{g}_{\text{FH}}^{\text{DL}}(\mathbf{P}, \mathbf{X})$ can be rewritten as

$$\begin{aligned} \tilde{f}_{\text{AC}}^{\text{DL}}(\mathbf{P}, \mathbf{T}) &= \sum_{j \in J} \sum_{s \in S_j} \sum_{u \in I_j} \sum_{k_1 \in \mathcal{K}_1} \log_2(\sigma_{u,k_1}^{s,j,\text{DL}} + I_{u,k_1}^{s,j,\text{DL}} + \tilde{I}_{u,k_1}^{s,j,\text{DL}} + \tau_{u,k_1}^{s,j,\text{DL}} p_{u,k_1}^{s,j,\text{DL}} h_{i,k_1}^{s,j,\text{DL}}) + \log_2(\sigma_{k_2}^{j,\text{DL}} + I_{k_2}^{j,\text{DL}}), \\ \text{and } \tilde{g}_{\text{FH}}^{\text{DL}}(\mathbf{P}, \mathbf{X}) &= \sum_{j \in J} \sum_{s \in S_j} \sum_{u \in I_j} \sum_{k_1 \in \mathcal{K}_1} \sum_{k_2 \in \mathcal{K}_2} \log_2(\sigma_{k_2}^{j,\text{DL}} + I_{k_2}^{j,\text{DL}} + x_{k_2}^{j,\text{DL}} p_{k_2}^{j,\text{DL}} h_{k_2}^{j,\text{DL}}). \end{aligned}$$

Also, we have

$$\nabla \tilde{g}_{\text{FH}}^{\text{DL}}(p_{k_2}^{j,\text{DL}}, x_{k_2}^{j,\text{DL}}) = \begin{cases} 0, & \text{If } h_{k_2}^{s,m,\text{DL}} < h_{k_2}^{j,\text{DL}}, \\ \frac{p_{k_2}^{j,\text{DL}} h_{k_2}^{m,\text{DL}}}{\sigma_{k_2}^{j,\text{DL}} + I_{k_2}^{j,\text{DL}} + p_{k_2}^{j,\text{DL}} h_{k_2}^{j,\text{DL}} x_{k_2}^{j,\text{DL}}}, & \text{Otherwise.} \end{cases}$$

For C10, we have [31], [32]

$$\exp(-\theta_j E_B^j(\theta_j) D_{\max}) = \exp(-\theta_j \lambda_j \frac{(e^{\theta_j} - 1)}{\theta_j} D_{\max}) = \exp(-\lambda_j (e^{\theta_j} - 1) D_{\max}) \leq \delta_1.$$

Therefore, we have $\sum_{s \in S_j} \sum_{u \in I_j} \sum_{k_1 \in \mathcal{K}_1} r_{u,k_1}^{s,j,q} \geq \frac{\ln(1/\delta_1)}{(e^{\theta_j} - 1) D_{\max}}$ which can be rewritten to

$$\begin{aligned} & \sum_{s \in S_j} \sum_{u \in I_j} \sum_{k_1 \in \mathcal{K}_1} \log_2(p_{u,k_1}^{s,j,q} \tau_{u,k_1}^{s,j,q} + \sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}) - \log_2(\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}) - \frac{\ln 1/\delta_1}{(e^{\theta_j} - 1) D_{\max}} \\ & = f_{\text{AC}}^q(\mathbf{P}, \mathbf{T}) - g_{\text{AC}}^q(\mathbf{P}, \mathbf{T}) \geq 0. \end{aligned}$$

Similarly, for C12, we have

$$\exp(-\theta_i^j E_B^{i,j}(\theta_i^j) D_{\max}) = \exp(-\theta_i^j \lambda_i^j \frac{(e^{\theta_i^j} - 1)}{\theta_i^j} D_{\max}) = \exp(-\lambda_i^j (e^{\theta_i^j} - 1) D_{\max}) \leq \delta,$$

Therefore, we have $\sum_{k_1 \in \mathcal{K}_1} r_{u,k_1}^{s,j,q} \geq \frac{\ln(1/\delta)}{(e^{\theta_i^j} - 1) D_{\max}}$ which can be simplified to

$$\begin{aligned} & \sum_{k_1 \in \mathcal{K}_1} \log_2(p_{u,k_1}^{s,j,q} \tau_{u,k_1}^{s,j,q} + \sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}) - \log_2(\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}) - \frac{\ln 1/\delta}{(e^{\theta_i^j} - 1) D_{\max}} = \\ & \hat{f}_{\text{AC}}^q(\mathbf{P}, \mathbf{T}) - \hat{g}_{\text{AC}}^q(\mathbf{P}, \mathbf{T}) \geq 0. \end{aligned}$$

Again, via the DC approximation, C15 is transformed into the convex functions as

$$R_{\text{rsv}}^s = \sum_{k_1 \in \mathcal{K}_1} \sum_{u \in \mathcal{I}} \sum_{j \in \mathcal{J}} \log_2(\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q}) \\ - \log_2(\sigma_{u,k_1}^{s,j,q} + I_{u,k_1}^{s,j,q} + \tilde{I}_{u,k_1}^{s,j,q} + \tau_{u,k_1}^{s,j,q} p_{u,k_1}^{s,j,q} h_{i,k_1}^{s,j,q}) = \tilde{f}_1(\mathbf{P}, \mathbf{T}) - \tilde{g}_1(\mathbf{P}, \mathbf{T}) \geq 0.$$

REFERENCES

- [1] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, “Realizing the tactile internet: Haptic communications over next generation 5G cellular networks,” *IEEE Wireless Communications*, vol. 24, no. 2, pp. 82–89, April 2017.
- [2] G. P. Fettweis, “The tactile internet: Applications and challenges,” *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, March 2014.
- [3] C. She and C. Yang, “Ensuring the quality-of-service of tactile internet,” in *Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–5.
- [4] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, “5G-Enabled tactile internet,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, March 2016.
- [5] S. H. Park, K. J. Lee, C. Song, and I. Lee, “Joint design of fronthaul and access links for C-RAN with wireless fronthauling,” *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1657–1661, November 2016.
- [6] M. Peng, Y. Li, Z. Zhao, and C. Wang, “System architecture and key technologies for 5G heterogeneous cloud radio access networks,” *IEEE Network*, vol. 29, no. 2, pp. 6–14, March 2015.
- [7] —, “System architecture and key technologies for 5G heterogeneous cloud radio access networks,” *IEEE Network*, vol. 29, no. 2, pp. 6–14, March 2015.
- [8] H. Zhang, Y. Dong, J. Cheng, M. J. Hossain, and V. C. M. Leung, “Fronthauling for 5G LTE-U ultra dense cloud small cell networks,” *IEEE Wireless Communications*, vol. 23, no. 6, pp. 48–53, December 2016.
- [9] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, “Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges,” *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 721–742, Secondquarter 2017.
- [10] L. Song, Y. Li, Z. Ding, and H. V. Poor, “Resource management in non-orthogonal multiple access systems: State of the art and research challenges,” *arXiv preprint arXiv:1610.09465*, 2016.
- [11] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-orthogonal multiple access (NOMA) for cellular future radio access,” pp. 1–5, June 2013.
- [12] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, “Network slicing in 5G: Survey and challenges,” *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, May 2017.
- [13] H. Zhang, N. Liu, X. Chu, K. Long, A. Aghvami, and V. C. M. Leung, “Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges,” *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, April 2017.

- [14] S. Parsaefard, R. Dawadi, M. Derakhshani, and T. Le-Ngoc, "Joint user-association and resource-allocation in virtualized wireless networks," *IEEE Access*, vol. 4, pp. 2738–2750, June 2016.
- [15] M. I. Kamel, L. B. Le, and A. Girard, "LTE wireless network virtualization: Dynamic slicing via flexible scheduling," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, September 2014, pp. 1–5.
- [16] A. Aijaz, "Towards 5G-enabled tactile internet: Radio resource allocation for haptic communications," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2016 IEEE*. IEEE, April 2016, pp. 145–150.
- [17] C. She, C. Yang, and T. Q. S. Quek, "Uplink transmission design with massive machine type devices in tactile internet," in *2016 IEEE Globecom Workshops (GC Wkshps)*, December 2016, pp. 1–6.
- [18] —, "Joint uplink and downlink resource configuration for ultra-reliable and low-latency communications," *IEEE Transactions on Communications*, vol. 66, no. 5, pp. 2266–2280, May 2018.
- [19] —, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 72–78, 2017.
- [20] R. G. Stephen and R. Zhang, "Joint millimeter-wave fronthaul and OFDMA resource allocation in ultra-dense CRAN," *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1411–1423, March 2017.
- [21] S. Parsaefard, V. Jumba, M. Derakhshani, and T. Le-Ngoc, "Joint resource provisioning and admission control in wireless virtualized networks," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 2020–2025.
- [22] V. Jumba, S. Parsaefard, M. Derakhshani, and T. Le-Ngoc, "Energy-efficient robust resource provisioning in virtualized wireless networks," in *2015 IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB)*, October 2015, pp. 1–5.
- [23] B. Ling, C. Dong, J. Dai, and J. Lin, "Multiple decision aided successive interference cancellation receiver for NOMA systems," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 498–501, Aug 2017.
- [24] A. Benjebbovu, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, "System-level performance of downlink NOMA for future LTE enhancements," in *2013 IEEE Globecom Workshops (GC Wkshps)*, December 2013, pp. 66–70.
- [25] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, November 2013, pp. 770–774.
- [26] D. J. Costello Jr, A. E. Pusane, S. Bates, and K. S. Zigangirov, "A comparison between LDPC block and convolutional codes," in *Proc. Information Theory and Applications Workshop*, 2006, pp. 6–10.
- [27] M. Ismail, I. Ahmed, J. Coon, S. Armour, T. Kocak, and J. McGeehan, "Low latency low power bit flipping algorithms for LDPC decoding," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2010, pp. 278–282.
- [28] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Transactions on Broadcasting*, vol. 48, no. 3, pp. 223–229, September 2002.

- [29] O. Edfors, M. Sandell, J. . van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," vol. 46, no. 7, July 1998, pp. 931–939.
- [30] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," vol. 49, no. 4, July 2000, pp. 1207–1215.
- [31] C. Li, A. Burchard, and J. Liebeherr, "A network calculus with effective bandwidth," *IEEE/ACM Transactions on Networking*, vol. 15, no. 6, pp. 1442–1453, December 2007.
- [32] C.-S. Chang and J. A. Thomas, "Effective bandwidth in high-speed digital networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1091–1100, August 1995.
- [33] L. Venturino, N. Prasad, and X. Wang, "Coordinated scheduling and power allocation in downlink multicell OFDMA networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2835–2848, July 2009.
- [34] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, "Joint subchannel assignment and power allocation for OFDMA femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 342–355, January 2014.
- [35] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2008.
- [36] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [37] H. H. Kha, H. D. Tuan, and H. H. Nguyen, "Fast global optimal power allocation in wireless networks by local D.C. programming," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 510–515, February 2012.