Energy-Efficient Resource Allocation in Heterogeneous Cloud Radio Access Networks via BBU-offloading

Nahid Amani, Hossein Pedram, Hassan Taheri, Saeedeh Parsaeefard

***Abstract*—Energy efficient resource allocation in heterogeneous cloud radio access networks (H-CRANs) is considered in this paper where a set of users within a specific region is served by femto-cell access points (FAPs) and remote radio heads (RRHs) connected to the base-band unit (BBU) pool through front-haul links. For orthogonal frequency division multiple access (OFDMA) based uplink H-CRANs, we formulate an energy efficient optimization problem with the novel utility function in order to jointly assign the access point (RRH/FAP), sub-carrier, transmit power, RRH, front-haul link and BBU. Due to highly complex relation between diverse optimization variables and their effects on each other, the formulated optimization problem is non-convex and NP-hard, suffering from high computational complexity. To tackle this issue, we develop an efficient two-step iterative algorithm with low computational complexity to solve the proposed problem based on the successive convex approximation (SCA) and complementary geometric programming (CGP). The simulation results reveal that our proposed approach can effectively offload the traffic from C-RAN to low-power femto-cells, and reduce the energy consumption cost by switching off the under-utilized BBUs.**

***Index Terms*—BBU offloading, fifth-generation (5G) of wireless networks, complementary geometric programming, successive convex approximation.**

I. Introduction

Increasing demand for wireless data traffic and mitigating energy consumption in different types of applications are inevitable for the fifth generation (5G) of wireless networks. Cloud radio access networks (C-RANs) are novel RAN architecture to reduce capital expenditure (CAPEX) and operating expense (OPEX) of 5G by separating remote radio head (RRH) from base band units (BBUs) based on network functions virtualization (NFV) concepts [1].

The use of different radio access points (APs) in heterogeneous networks (HetNets) will improve network capacity and energy efficiency (EE) of 5G networks [2]. To take advantage of both C-RANs and HetNets, heterogeneous cloud radio access networks (H-CRANs) are introduced as cost-efficient potential solutions to enhance spectrum efficiency (SE) and EE [3]. The H-CRAN architecture consists of two tiers: (1) a C-RAN tier consisting of a set of RRHs connected to a set of centralized BBUs through wire/wireless front-haul links where each RRH can be assumed to be served by one virtual network function (VNF) in the BBU; and, (2) a low power femto access points (FAPs) tier connected to the core network via back-haul links in this region. Deploying a large number of RRHs in H-CRAN can bring higher SE to 5G [4]. However, due to employing highly overlapped RRHs and interference among users from different coverage regions of APs, it is not sufficient to assign each user to specific AP based on the largest received signal-to-interference-plus-noise ratio (SINR). Moreover, to mitigate interference and enhance EE and SE, dynamic resource allocation is essential, e.g., [5], [6]. In this context, due to existence of inter-related diverse radio optimization variables, e.g., transmit power, sub-carrier and AP assignment to each user, and cloud optimization variables such as front-haul allocation to active BBUs, dynamic resource allocation problem is NP-hard and suffers from high computational complexity. Solving this type of problems efficiently and addressing their drivers aspects have been drawn a lot of attentions recently, e.g., [4], [5].

H-CRAN struggles on the trade-off between increasing the SE and enhancing EE [5]. To reach higher SE, large number of RRHs should be deployed which may lead to under-utilized BBUs and RRHs, and consequently, decreasing of EE. Therefore, to address this issue, proposing the EE resource allocation in H-CRAN is essential. Recently, a surge of research has been developed for proposing EE resource allocation for both C-RANs and traditional HetNets. A large body of the existing works focused either on decreasing the transmit power of users, e.g., [5], [7], [8], [9], or on decreasing the power consumption of APs by introducing the sleep mode strategies, e.g., [10], [11], [12]. For example, in [4], by virtualizing the resources into slices, joint power, sub-carrier, and antenna allocation problem in wireless small cell networks with large scale multiple antenna is proposed to attain EE. [5] formulates an EE optimization problem with user association to APs and power allocation where Lagrange dual decomposition method is applied to reduce power consumption of users in H-CRAN. [7] and [8] formulate joint sub-carrier and power allocation with the aim of maximizing EE as a noncooperative game problem in HetNet. [9] formulates joint physical resource block and power allocation without front-haul limitation with the aim of minimizing transmit power of users, and proposes a knapsack formulation to solve the RRH-BBU assignment

N. Amani and H. Pedram are with the Department of Computer Engineering, AmirKabir University of Technology Tehran, Iran, e-mails: nahid.amani@aut.ac.ir and pedram@aut.ac.ir

H. Taheri is with the Department of Electrical Engineering, AmirKabir University of Technology Tehran, Iran, e-mail: htaheri@aut.ac.ir

S. Parsaeefard is with the Department of Communication Technologies, ITRC, Tehran, Iran, e-mail: s.parsaeifard@itrc.ac.ir

TABLE I
SUMMARY OF RELATED WORKS

Ref.	System model	Objective function	Optimization variables	Solution
[5]	H-CRAN	Min sum power of users	AP assignment, sub-carrier and power	Lagrange dual decomposition/global optimal
[7]	HetNet	Max EE	Sub-carrier and transmit power	Noncooperative game/local optimal
[8]	HetNet	Max EE	Sub-carrier and transmit power	Noncooperative game/local optimal
[9]	C-RAN	Min sum power of users	AP, sub-carrier, power and RRH to BBUs assignment	Branch and Cut algorithm/local optimal
[10]	HetNet	Min sum power of APs	AP assignment and AP transmission power	Heuristic algorithms/local optimal
[11]	C-RAN	Min sum power of APs	AP transmit power	Reweighted ℓ_1 minimization and SCA/local optimal
[12]	HetNet	Min sum power of APs	AP transmit Power	Stochastic geometry models/local optimal
[13]	C-RAN	Min cost of installing front-haul links	Front-haul to BBUs assignment	Heuristic algorithm/local optimal
[14]	C-RAN	Min sum power of computation resources in BBUs	BBUs computation resources	Heuristic algorithm/local optimal
This work	H-CRAN	Max EE with novel utility function	AP assignment, sub-carrier and transmit power allocation, and RRH and front-haul allocation to active BBUs	CGP and SCA/local optimal

in C-RAN. [10] proposes heuristic algorithms to minimize energy consumption with user association where total energy consumption for each base station (BS) is computed based on mobile traffic statistics. Then, each user is associated with its nearest BS to obtain a minimum total energy consumption in HetNet. [11] formulates the optimization problem of minimizing the transmission power and activation power of BSs in C-RAN by applying reweighted ℓ_1 minimization technique and successive convex approximation (SCA). In [12], BS sleep mode strategies are proposed for HetNet where stochastic geometry models are used to derive the coverage probability of UEs. In [13], the cost of installing and maintaining front-haul links and BBUs is minimized by reducing the number of installed nodes. [14] and [15] reduce the power consumption of the BBU pool via dynamic allocation of computation resources to the virtual machines (VMs) in C-RAN. The above mentioned works are summarized in Table I. However, none of the mentioned works has jointly considered dynamic radio resource allocation, user association and C-RAN limitations, e.g., maximum front-haul capacity and BBU capacity in order to reduce energy consumption cost of BBUs by switching off under-utilized BBUs in H-CRAN and this paper aims to fill this gap.

To address this problem, we formulate the EE optimization problem with the novel utility function and new set of constraints in order to jointly assign APs (RRH/FAP), sub-carrier and transmit power to each user, and RRH and front-haul allocation to active BBUs. Hence, we consider energy consumption cost of BBUs and total transmit power of all users as operation cost in utility function. Considering this cost in utility function leads to migrate VNFs from under-utilized and high energy cost BBUs to low energy cost BBUs. Besides, associated user equipments (UEs) to the under-utilized BBUs and their related RRHs can be offloaded to neighboring FAPs. Therefore, traffic of this group of UEs will be offloaded from C-RAN to low cost FAPs which is referred to traffic offloading in this paper. Consequently, under-utilized BBUs can be switched off in order to reach higher EE. In this case, when BBUs are switched on/off according to the variations of traffic load, their corresponding front-haul links will be enabled/disabled subsequently.

Due to coupling between RRHs and BBUs and for the above mentioned problems, we introduce a new optimization variable to jointly assign RRH and BBU, which is referred to joint RRH-BBU association factor (JRBF). Since sub-carrier assignment and AP association are also inter-related, we introduce the user association factor (UAF) as a new optimization variable [16]. Due to existence of binary variables such as JRBF and UAF, and interference among users in our setup, the proposed optimization problem is non-convex and NP-hard with high computational complexity. We apply the frameworks of coplementary geometric programming (CGP) and the SCA [16], [17], [18], [19] to develop an efficient two-step iterative algorithm with low computational complexity to solve the proposed problem. We will show by applying various transformation techniques, such as DC-approximation and AGMA, a highly non-convex problem of each step can be transformed into its lower-bound geometric programming (GP) problem, which can be solved via efficient on-line available software packages, e.g., CVX [20].

We compare performance of the proposed approach with traditional approach where each user is assigned to the APs based on largest value of received SINR. The simulation results show that our proposed approach with novel utility function is more efficient than the traditional approach, in terms of increasing EE and reducing outage probability. Also, simulation results reveal that our proposed algorithm can effectively offload the traffic from C-RAN to low-power FAPs.

The rest of this paper is organized as follows. In Section II, the system model and problem formulation are introduced. Section III presents the proposed two-step iterative algorithm to solve the formulated problem. Section IV demonstrates the simulation results and their detailed computational complexity analysis, followed by concluding remarks in Section V.

II. System Model and Problem Formulation

We consider an uplink transmission of a two-tier orthogonal frequency division multiple access (OFDMA) based H-CRANs, as demonstrated in Fig. 1, where a set of $\mathcal{N} = \{1, ..., N\}$ UEs within a specific region is served by

- A set of $\mathcal{F} = \{1, ..., F\}$ FAPs,
- A set of $\mathcal{R} = \{1, ..., R\}$ RRHs,
- A BBU pool consisting of $\mathcal{B} = \{1, ..., B\}$ BBUs,

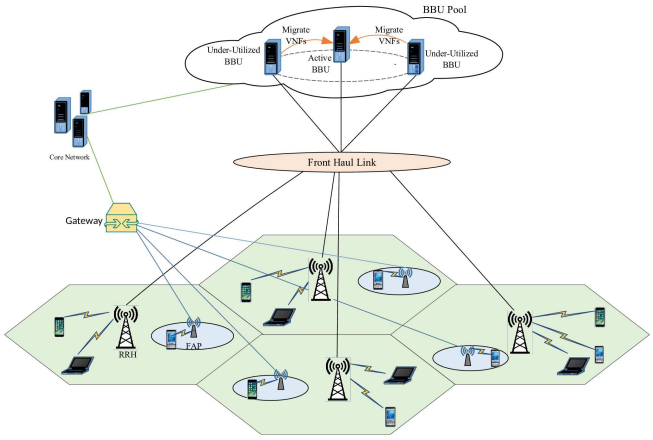

Fig. 1. Architecture of a H-CRAN

where each RRH is connected to the BBU pool by a limited capacity front-haul link and FAPs are connected to the core network via back-haul links. The BBU pool is responsible to process the baseband signals for all RRHs in this region. We use $\mathcal{K} = \mathcal{R} \bigcup \mathcal{F}$ to demonstrate the set of all APs in the network. Based on OFDMA, a bandwidth of W is divided into a set of $\mathcal{S} = \{1, ..., S\}$ sub-channels. A binary variable $x_{k,s,n}$ represents a joint sub-carrier and AP assignment for user $n \in \mathcal{N}$ on sub-carrier $s \in \mathcal{S}$ of AP $k \in \mathcal{K}$ as

$$x_{k,s,n} = \begin{cases} 1, & \text{if AP } k \text{ allocates } s^{th} \text{ sub-carrier to } n^{th} user, \\ 0, & \text{otherwise,} \end{cases}$$

where $x_{k,s,n}$ is defined as a UAF. Let $h_{k,s,n}$ and $p_{k,s,n}$ be the channel gain and the transmit power of user $n \in \mathcal{N}$ to AP $k \in \mathcal{K}$ on sub-channel $s \in \mathcal{S}$, respectively. Also consider $\mathbf{P} = [p_{k,s,n}]_{K \times S \times N}$ and $\mathbf{X} = [x_{k,s,n}]_{K \times S \times N}$ as matrices of transmit power and UAF of all users, respectively. For this setup, the throughput of user $n \in \mathcal{N}$ over sub-channel $s \in \mathcal{S}$ in AP $k \in \mathcal{K}$ is

$$R_{k,s,n}(\mathbf{P}) = \log_2(1 + \frac{p_{k,s,n} h_{k,s,n}}{\sigma^2 + I_{k,s,n}}), \tag{1}$$

in which $I_{k,s,n} = \sum_{\forall k' \in \mathcal{K}, k' \neq k} \sum_{\forall n' \neq n} p_{k',s,n'} h_{k,s,n'}$ is the interference to user $n \in \mathcal{N}$ in AP $k \in \mathcal{K}$ and sub-carrier $s \in \mathcal{S}$, and σ^2 is the noise power. Without loss of generality, σ^2 is assumed to be equal for all users in all sub-carriers and all APs. The QoS requirement of each user should be guaranteed which is represented by its minimum reserved rate, i.e., R_n^{rsv},

$$\text{C1}: \sum_{\forall k \in \mathcal{K}} \sum_{\forall s \in \mathcal{S}} x_{k,s,n} R_{k,s,n}(\mathbf{P}) \geq R_n^{\text{rsv}}, \quad \forall n \in \mathcal{N}.$$

TABLE II
Table of Notations

Notation	Description
$x_{k,s,n} \in \{0,1\}$	If $x_{k,s,n} = 1$, AP k allocates s^{th} sub-carrier to n^{th} user
$\mathcal{F} = \{1, ..., F\}$	A set of femto-cell access points
$\mathcal{R} = \{1, ..., R\}, r$	A set of RRHs, index of RRH
$\mathcal{B} = \{1, ..., B\}, b$	A set of BBU, index of BBU
$\mathcal{K} = \mathcal{R} \bigcup \mathcal{F}, k$	A set of all access points, index of AP
$\mathcal{S} = \{1, ..., S\}, s$	A set of sub-carriers, index of sub-carrier
$\mathcal{N} = \{1, ..., N\}, n$	A set of UEs, index of UE
$h_{k,s,n}$	Channel gain of user n to AP k on $s \in \mathcal{S}$
$p_{k,s,n}$	Transmit power of user n to AP k on $s \in \mathcal{S}$
σ^2	Noise power
$\mathbf{X} = [x_{k,s,n}]_{K \times S \times N}$	Matrix of user association factor
$\mathbf{P} = [p_{k,s,n}]_{K \times S \times N}$	Matrix of transmit powers
$I_{k,s,n}$	Interference to user n in AP k on $s \in \mathcal{S}$
I_s^{th}	Maximum interference for sub-carrier s
$R_{k,s,n}$	Throughput of user n to AP k on $s \in \mathcal{S}$
R_n^{rsv}	Minimum reserved rate of user n
$y_{r,b} \in \{0,1\}$	If $y_{r,b} = 1$, RRH r is assigned to BBU b
$\mathbf{Y} = [y_{r,b}]_{R \times B}$	Matrix of RRHs associated to BBUs
L_r	Load of UEs assigned to the RRH r
$bbu_b \in \{0,1\}$	If $bbu_b = 1$, BBU b is in state on
$\mathbf{bbu} = [bbu_b]_{1 \times B}$	Vector of all BBUs
$L_b^{\max}$	Maximum capacity of BBU b
$F_{r,b}^{\max}$	Maximum capacity of Front-haul $y_{r,b}$
$p_n^{\max}$	Maximum transmit power of user n.
C_b	Energy consumption cost of BBU b
C_p	Cost of the total transmit power
η	Network offloading ratio
$U(\mathbf{X}, \mathbf{P}, \mathbf{Y}, \mathbf{bbu})$	Network utility function

Due to OFDMA practical limitations, each user $n \in \mathcal{N}$ must be associated to only one AP and each sub-carrier should be assigned to only one UE in any AP. These two implementation limitations can be mathematically represented as [16]

$$\text{C2}: [\sum_{s \in \mathcal{S}} x_{k,s,n}][\sum_{\forall k' \neq k} \sum_{s \in \mathcal{S}} x_{k',s,n}] = 0, \quad \forall n \in \mathcal{N}, \forall k \in \mathcal{K},$$

and

$$\text{C3}: \sum_{\forall n \in \mathcal{N}} x_{k,s,n} \leq 1, \quad \forall k \in \mathcal{K}, \forall s \in \mathcal{S}.$$

Due to transmit power limitation of each user, we have

$$\text{C4}: \sum_{k \in \mathcal{K}} \sum_{s \in S} p_{k,s,n} \leq p_n^{\max}, \quad \forall n \in \mathcal{N},$$

where $p_n^{\max}$ is the maximum transmit power of user $n \in \mathcal{N}$. In the uplink, each RRH receives user's signals and transmits to the BBU through its front-haul link. Then, the BBU decodes them and forwards the processed signals to the core network via a back-haul link. Consider $\mathbf{Y} = [y_{r,b}]_{R \times B}$ as a matrix of RRHs associated to BBUs. Therefore, we define $y_{r,b}$ as a joint RRH to BBU association factor (JRBF) as

$$y_{r,b} = \begin{cases} 1, & \text{If the RRH } r \text{ is assigned to the BBU } b, \\ 0, & \text{Otherwise,} \end{cases}$$

representing front-haul link between RRH $r \in \mathcal{R}$ and BBU $b \in \mathcal{B}$. Assuming that each RRH $r \in \mathcal{R}$ can be associated to

only one BBU, we have

$$\text{C5}: \sum_{b\in\mathcal{B}} y_{r,b} = 1, \quad \forall r \in \mathcal{R}.$$

The front-haul link between each RRH $r \in \mathcal{R}$ and the BBU $b \in \mathcal{B}$ has a limited capacity represented as

$$\text{C6}: y_{r,b} L_r \leq F_{r,b}^{\max}, \ \forall r \in \mathcal{R}, \forall b \in \mathcal{B},$$

where L_r is a sum of traffic loads of UEs assigned to the RRH $r \in \mathcal{R}$, defined as $L_r = \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} x_{r,s,n} R_{r,s,n}(\mathbf{P})$. In C6, $F_{r,b}^{\max}$ is the front-haul capacity limitation between RRH $r \in \mathcal{R}$ and BBU $b \in \mathcal{B}$. Consider a vector of all BBUs as $\mathbf{bbu} = [bbu_b]_{1\times B}$ where bbu_b represents the on-off states of BBU $b \in \mathcal{B}$ as

$$bbu_b = \begin{cases} 1, & \text{if BBU } b \text{ is in state on,} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, the load allocation of each BBU $b \in \mathcal{B}$ is

$$\text{C7}: \sum_{r\in\mathcal{R}} y_{r,b} L_r \leq L_b^{\max} \times bbu_b, \quad \forall b \in \mathcal{B},$$

where $L_b^{\max}$ is the maximum load handled by each BBU $b \in \mathcal{B}$. The allocated load to each BBU is received from its own corresponding allocated RRHs. A front-haul link is created between RRH $r \in \mathcal{R}$ and BBU $b \in \mathcal{B}$ when BBU b is active. To represent this novel constraint, we define C8 as

$$\text{C8}: y_{r,b} - bbu_b \leq 0, \quad \forall r \in \mathcal{R}, \forall b \in \mathcal{B}.$$

In H-CRAN, a new concern of the operational cost is the energy consumption cost of BBUs and total transmit power of all users. Therefore, we define the total operation cost as $\text{C}_\text{p}[\sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} x_{k,s,n} p_{k,s,n}] + \sum_{b\in\mathcal{B}} \text{C}_b \times bbu_b$ where C_p is a cost of the total transmit power and can be considered as a dimension regulation factor and its unit is bps/Hz/Watt. C_b denotes an energy consumption cost of each BBU $b \in \mathcal{B}$ which is proportional to its maximum load capacity and its unit is bps/Hz. Hence, we define a novel network utility function as

$$U(\mathbf{X},\mathbf{P},\mathbf{Y},\mathbf{bbu}) = \sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} x_{k,s,n} R_{k,s,n}(\mathbf{P}) - \quad (2)$$
$$\text{C}_\text{p}[\sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} x_{k,s,n} p_{k,s,n}] - \sum_{b\in\mathcal{B}} \text{C}_b \times bbu_b,$$

which is the total throughput minus the total operation cost. Notably, the trade-off between the increasing total throughput and the decreasing total operation cost helps to find optimal solutions for maximizing EE of network [4]. Consequently, the joint UAF, JRBF and power allocation problem to maximize EE can be written as

$$\max_{\mathbf{X},\mathbf{P},\mathbf{Y},\mathbf{bbu}} U(\mathbf{X},\mathbf{P},\mathbf{Y},\mathbf{bbu}) \quad (3)$$
$$\text{subject to}: \text{C1} - \text{C8}.$$

The formulated problem (3) is non-convex and NP-hard with high computational complexity [16]. To tackle this issues, by applying the SCA and CGP, we will develop an efficient two-step iterative algorithm with low computational complexity to solve the proposed problem. We will demonstrate that operation cost considered in utility function leads to offload traffic from C-RAN to neighboring FAPs with low transmit power. Therefore, VNFs of under-utilized BBUs migrate to low cost BBUs. Due to traffic offloading, under-utilized and high cost BBUs can be switched off. Also, UEs can be connected to FAPs with low transmit power. Consequently, total utility and EE of network will be enhanced.

TABLE III
ALGORITHM1: JOINT JRBF, UAF ASSOCIATION AND POWER ALLOCATION ITERATIVE ALGORITHM

Initialization: Set $\tau = 0$, $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 10^{-3}$,
$\mathbf{P}(\tau = 0) = P^{\max}/S$.
Repeat: set $\tau = \tau + 1$
  Step 1.A Finding UAF and JRBF:
  Initialization for Step 1.A: set $\tau_1 = 0$,
  $\mathcal{V}(\tau_1) = \mathcal{V}(\tau)$, $\mathbf{P}(\tau_1) = \mathbf{P}(\tau)$ and set arbitrary initial for $q_{k,n}(\tau_1)$.
  Repeat: set $\tau_1 = \tau_1 + 1$.
    Step 1.A.1: Update CGP variables according to (7)-(17),
    Step 1.A.2: Solve $\mathcal{O}_1(\mathcal{V}(\tau_1))$ in (5) for $\mathbf{X}(\tau_1)$,
    $\mathbf{Y}(\tau_1)$ and $\mathbf{bbu}(\tau_1)$ using CVX [20],
  Until $\| \mathbf{X}^*(\tau_1) - \mathbf{X}^*(\tau_1 - 1) \| \leq \varepsilon_1$ and $\| \mathbf{Y}^*(\tau_1) - \mathbf{Y}^*(\tau_1 - 1) \| \leq \varepsilon_2$ and $\| \mathbf{bbu}^*(\tau_1) - \mathbf{bbu}^*(\tau_1 - 1) \| \leq \varepsilon_3$. set $\mathbf{X}(\tau) = \mathbf{X}^*(\tau_1)$, $\mathbf{Y}(\tau) = \mathbf{Y}^*(\tau_1)$, $\mathbf{bbu}^*(\tau) = \mathbf{bbu}^*(\tau_1)$.
  Step 1.B Power Allocation:
  Initialization for Step 1.B: set $\tau_2 = 0$,
    $\mathbf{X}(\tau_2) = \mathbf{X}(\tau)$, $\mathbf{Y}(\tau_2) = \mathbf{Y}(\tau)$, $\mathbf{bbu}(\tau_2) = \mathbf{bbu}(\tau)$.
  Repeat: set $\tau_2 = \tau_2 + 1$.
    Step 1.B.1: Update CGP variables according to (21)-(25),
    Step 1.B.2: Solve $\text{O}_2(\mathbf{P}(\tau_2))$ for $\mathbf{P}(\tau_2)$ using CVX [20],
  Until $\| \mathbf{P}^*(\tau_2) - \mathbf{P}^*(\tau_2 - 1) \| \leq \varepsilon_4$, set $\mathbf{P}^*(\tau) = \mathbf{P}^*(\tau_2)$.
Until $\| \mathbf{X}^*(\tau) - \mathbf{X}^*(\tau - 1) \| \leq \varepsilon_1$ and $\| \mathbf{Y}^*(\tau) - \mathbf{Y}^*(\tau - 1) \| \leq \varepsilon_2$ and $\| \mathbf{bbu}^*(\tau) - \mathbf{bbu}^*(\tau - 1) \| \leq \varepsilon_3$ and $\| \mathbf{P}^*(\tau) - \mathbf{P}^*(\tau - 1) \| \leq \varepsilon_4$.

III. TWO-STEP ITERATIVE ALGORITHM FOR DYNAMIC RESOURCE ALLOCATION

To solve (3), we apply an iterative algorithm to find the JRBF, UAF and power allocation for each user in two steps as summarized in Table III. In Step 1.A, with a given (fixed) power allocation vector, the optimal JRBF and UAF vectors are achieved by Algorithm 1.A (to be discussed in Section III.A). In Step 1.B, the power-allocation problem is solved based on obtained JRBF and UAF vectors from Step 1.A (to be discussed in Section III.B). These two steps are shown in Algorithm 1. The sequence of the JRBF, UAF and power allocation vector solutions can be expressed as

$$\underbrace{\mathbf{X}(0), \mathbf{Y}(0), \mathbf{bbu}(0) \rightarrow \mathbf{P}(0)}_{\text{Initialization}} \rightarrow \ldots$$
$$\rightarrow \underbrace{\mathbf{X}^*(\tau), \mathbf{Y}^*(\tau), \mathbf{bbu}^*(\tau) \rightarrow \mathbf{P}^*(\tau)}_{\text{Iteration } \tau} \rightarrow \quad \underbrace{\mathbf{X}^*, \mathbf{Y}^*, \mathbf{bbu}^* \rightarrow \mathbf{P}^*}_{\text{Optimal solution}},$$

where $\tau \geq 0$ is the iteration index. Also, $\mathbf{X}^*(\tau), \mathbf{Y}^*(\tau), \mathbf{bbu}^*(\tau)$ and $\mathbf{P}^*(\tau)$ are optimal values obtained at the iteration τ from convex transformation of related optimization problems in each step. The iterative procedure is stopped when convergence criteria are hold, as

$$\| \mathbf{X}^*(\tau) - \mathbf{X}^*(\tau - 1) \| \leq \varepsilon_1, \| \mathbf{Y}^*(\tau) - \mathbf{Y}^*(\tau - 1) \| \leq \varepsilon_2,$$
$$\| \mathbf{bbu}^*(\tau) - \mathbf{bbu}^*(\tau - 1) \| \leq \varepsilon_3, \| \mathbf{P}^*(\tau) - \mathbf{P}^*(\tau - 1) \| \leq \varepsilon_4,$$

where $0 < \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4 \ll 1$. Notably, optimization problems of these two steps are still non-convex and suffer from high computational complexity. To solve them efficiently, we apply CGP [19] for each step where via different transformations and convexification approaches, the sequence of lower bound GP approximation of relative optimization problems are solved. We refer interested reader about CGP to Section III.A in [16]. The sub-algorithms are described in the following subsections.

A. JRBF and UAF Allocation Algorithm

Assuming fixed value of $\mathbf{P}(\tau)$, the optimization variables for this step are $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{bbu}$, and (3) is transformed into

$$\max_{\mathbf{X},\mathbf{Y},\mathbf{bbu}} \widetilde{U}(\mathbf{X}, \mathbf{P}(\tau), \mathbf{Y}, \mathbf{bbu}) \tag{4}$$

$$\text{subject to: } \widetilde{\mathrm{C}}1, \widetilde{\mathrm{C}}2, \mathrm{C}3, \widetilde{\mathrm{C}}5, \mathrm{C}6, \mathrm{C}7, \widetilde{\mathrm{C}}8.$$

Utility function (2) with a given (fixed) power allocation is converted to $\widetilde{U}(\mathbf{X}, \mathbf{P}(\tau), \mathbf{Y}, \mathbf{bbu})$. In (4) optimization variables are $\mathbf{X}, \mathbf{Y}$ and $\mathbf{bbu}$, and (4) has less computational complexity than (3). However, due to binary variables $\mathbf{X}, \mathbf{Y}$ and $\mathbf{bbu}$, and non-linear constraints, the optimization problem (4) is still non-convex. To tackle these issues, we try to convert (4) into the standard form of GP. To reach this goal, we first relax the binary variables as $x_{k,s,n} \in [0,1]$, $y_{r,b} \in [0,1]$ and $bbu_b \in [0,1]$. Then, we apply iterative AGMA technique to get the monomial approximation $\widetilde{\mathrm{C}}1, \widetilde{\mathrm{C}}2, \widetilde{\mathrm{C}}5$ and $\widetilde{\mathrm{C}}8$ for C1, C2, C5 and C8, respectively, to reach standard GP format for optimization problem (4) as shown in proposition 1.

Proposition 1: Consider the positive auxiliary variables $\varpi_0(\tau_1) > 0$, $q_{k,n}(\tau_1)$, $z_{k,n}(\tau_1) = \sum_{s\in\mathcal{S}} x_{k,s,n}$, $z_n^{'}(\tau_1) = \sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} x_{k,s,n}$ and $\Xi_1 \gg 1$. Also, consider $\mathbf{q}$, $\mathbf{z}$ and $\mathbf{z}'$ as the vectors of all $q_{k,n}$, $z_{k,n}$ and $z_n^{'}$, respectively, for all $k \in \mathcal{K}$ and $n \in \mathcal{N}$ and $\mathcal{V}(\tau_1) = \{\mathbf{X}(\tau_1), \varpi_0(\tau_1), \mathbf{q}(\tau_1), \mathbf{z}(\tau_1), \mathbf{z}'(\tau_1), \mathbf{Y}(\tau_1), \mathbf{bbu}(\tau_1)\}$. The GP transformation of (4) at each iteration τ_1 is [16]

$$\mathcal{O}_1(\mathcal{V}(\tau_1)): \quad \min_{\mathcal{V}(t_1)} \varpi_0(\tau_1) \tag{5}$$
$$\text{subject to:} \quad \mathrm{C}3,$$

$$[\Xi_1 + \mathrm{C_p}(\sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} x_{k,s,n}(\tau_1) p_{k,s,n}(\tau)) + \sum_{b\in\mathcal{B}} C_b \times bbu_b(\tau_1)] \times (\frac{\varpi_0(\tau_1)}{\psi_0(\tau_1)})^{-\psi_0(\tau_1)} \times \prod_{k\in\mathcal{K},n\in\mathcal{N},s\in\mathcal{S}} (\frac{x_{k,s,n}(\tau_1) R_{k,s,n}(\mathbf{P}(\tau))}{w_{k,s,n}(\tau_1)})^{-w_{k,s,n}(\tau_1)} \leq 1, \tag{6}$$

$$\widetilde{\mathrm{C}}1: \ R_n^{\mathrm{rsv}} \times \prod_{k\in\mathcal{K},s\in\mathcal{S}} \left[\frac{x_{k,s,n}(\tau_1) R_{k,s,n}(\mathbf{P}(\tau))}{\varphi_{k,s,n}(\tau_1)}\right]^{-\varphi_{k,s,n}(\tau_1)} \leq 1, \forall n \in \mathcal{N},$$

$$\widetilde{\mathrm{C}}2.1: \ q_{k,n}^{-1}(\tau_1) + z_{k,n}(\tau_1) z_n^{'}(\tau_1) q_{k,n}^{-1}(\tau_1) \leq 1, \quad \forall n \in \mathcal{N}, \forall k \in \mathcal{K},$$

$$\widetilde{\mathrm{C}}2.2: \ q_{k,n}(\tau_1) \times \left[\frac{1}{\alpha_{k,n}(\tau_1)}\right]^{-\alpha_{k,n}(\tau_1)} \times \left[\frac{z_{k,n}^2(\tau_1)}{\beta_{k,n}(t_1)}\right]^{-\beta_{k,n}(\tau_1)} \leq 1, \forall n \in \mathcal{N}, \forall k \in \mathcal{K},$$

$$\widetilde{\mathrm{C}}2.3: \ z_{k,n}(\tau_1) \times \prod_{s\in\mathcal{S}} \left[\frac{x_{k,s,n}(\tau_1)}{\lambda_{k,s,n}(\tau_1)}\right]^{-\lambda_{k,s,n}(\tau_1)} = 1, \quad \forall n \in \mathcal{N}, \forall k \in \mathcal{K},$$

$$\widetilde{\mathrm{C}}2.4: \ z_{k,n}^{'}(\tau_1) \times \prod_{k\in\mathcal{K},\ s\in\mathcal{S}} \left[\frac{x_{k,s,n}(\tau_1)}{\mu_{k,s,n}(\tau_1)}\right]^{-\mu_{k,s,n}(\tau_1)} = 1, \quad \forall n \in \mathcal{N}, \forall k \in \mathcal{K},$$

$$\widetilde{\mathrm{C}}5: 2 \times (\frac{1}{\theta(\tau_1)})^{-\theta(\tau_1)} \times \prod_{b\in\mathcal{B}} \left[\frac{y_{r,b}(\tau_1)}{\S_{r,b}(\tau_1)}\right]^{-\S_{r,b}(\tau_1)} = 1, \forall r \in \mathcal{R},$$

$$\mathrm{C}6: \ \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} y_{r,b} x_{r,s,n}(\tau_1) R_{r,s,n}(\mathbf{P}(\tau)) \leq F_{r,b}^{\max}, \quad \forall r \in \mathcal{R}, \forall b \in \mathcal{B},$$

$$\mathrm{C}7: \ bbu_b(\tau_1)^{-1} \times \left[\sum_{r\in\mathcal{R}} \sum_{n\in\mathcal{N}} \sum_{s\in\mathcal{S}} x_{r,s,n}(\tau_1) y_{r,b}(\tau_1) R_{r,s,n}(\mathbf{P}(\tau))\right] \leq L_b^{\max}, \quad \forall b \in \mathcal{B},$$

$$\widetilde{\mathrm{C}}8: (y_{r,b}+1) \times (\frac{1}{\delta(\tau_1)})^{-\delta(\tau_1)} \times (\frac{bbu_b(\tau_1)}{\phi(\tau_1)})^{-\phi(\tau_1)} \leq 1, \quad \forall r \in \mathcal{R}, \forall b \in \mathcal{B},$$

where

$$\alpha_{k,n}(\tau_1) = \frac{1}{z_{k,n}^2(\tau_1 - 1) + 1}, \tag{7}$$

$$\beta_{k,n}(\tau_1) = \frac{z_{k,n}^2(\tau_1 - 1)}{z_{k,n}^2(\tau_1 - 1) + 1}, \tag{8}$$

$$\lambda_{k,s,n}(\tau_1) = \frac{x_{k,s,n}(\tau_1 - 1)}{\sum_{s\in\mathcal{S}} x_{k,s,n}(\tau_1 - 1)}, \tag{9}$$

$$\mu_{k,s,n}(\tau_1) = \frac{x_{k,s,n}(\tau_1 - 1)}{\sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} x_{k,s,n}(\tau_1 - 1)}, \tag{10}$$

$$\theta(\tau_1) = \frac{1}{1 + \sum_{b\in\mathcal{B}} y_{r,b}(\tau_1 - 1)}, \tag{11}$$

$$\S_{r,b}(\tau_1) = \frac{y_{r,b}(\tau_1 - 1)}{\sum_{b\in\mathcal{B}} y_{r,b}(\tau_1 - 1)}, \tag{12}$$

$$\varphi_{k,s,n}(\tau_1) = \frac{x_{k,s,n}(\tau_1 - 1) R_{k,s,n}(\mathbf{P}(\tau))}{\sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} x_{k,s,n}(\tau_1 - 1) R_{k,s,n}(\mathbf{P}(\tau))}, \tag{13}$$

$$\psi_0(\tau_1) = \frac{\varpi_0(\tau_1-1)}{\varpi_0(\tau_1-1)+\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} X_{k,s,n}(\tau_1-1)R_{k,s,n}(\mathbf{P}(\tau))}, \tag{14}$$

$$w_{k,s,n}(\tau_1) = \frac{x_{k,s,n}(\tau_1-1)R_{k,s,n}(\mathbf{P}(\tau))}{\varpi_0(\tau_1-1)+\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}(\tau_1-1)R_{k,s,n}(\mathbf{P}(\tau))}, \tag{15}$$

$$\delta(\tau_1) = \frac{1}{1+bbu_b(\tau_1-1)}, \tag{16}$$

and

$$\phi(\tau_1) = \frac{bbu_b(\tau_1-1)}{1+bbu_b(\tau_1-1)}. \tag{17}$$

Proof. See Appendix A. □

$\mathcal{O}_1(\mathcal{V}(\tau_1))$ includes $\widetilde{\mathrm{C}}2.1-\widetilde{\mathrm{C}}2.4$ which are utilized for converting C2. Therefore C2 is demonstrated by the approximated monomial equalities and posynomial inequalities. Now, from Step 1.A of Algorithm 1, instead of (4), optimization problem $\mathcal{O}_1(\mathcal{V}(\tau_1))$ at each iteration is solved until the optimal value of JRBF and UAF vectors are derived based on CVX software [20].

Proposition 2: By applying AGMA transformation technique, Step 1.A of Algorithm 1 converges to a local optimal solution that satisfies the KKT conditions of the original problem.

Proof. [18] proves that the solutions of the series of approximations by AGMA converge to a point satisfying the Karush-Kuhn-Tucker (KKT) conditions for the convergence of the SCA method of (4). Based on [21] and [22], Algorithm 1 is a type of block SCA. Therefore, when (4) is feasible, the outer loop of Algorithm 1 converges to a local optimal solution of the KKT optimality conditions. □

B. Power Allocation Algorithm

For a given fixed set of JRBF and UAFs obtained from Step 1.A, we propose Step 1.B of Algorithm 1 to solve the power allocation problem. Consequently, the optimization problem of this step is

$$\max_{\mathbf{P}}[\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}(\tau)R_{k,s,n}(\mathbf{P}(\tau_2))- \mathrm{C_p}[\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}p_{k,s,n}(\tau_2)] - \sum_{b\in\mathcal{B}}\mathrm{C}_b\times bbu_b(\tau)] \tag{18}$$

$$\text{subject to: } \widetilde{\mathrm{C}}1.1, \mathrm{C}4, \widetilde{\mathrm{C}}6, \widetilde{\mathrm{C}}7.$$

In (18), the only optimization variable is $\mathbf{P}$, and therefore, (18) has lower computational complexity than (3). Since throughput is a logarithmic function and non-linear, (18) is non-convex optimization problem. Therefore, again by applying DC-approximation and iterative AGMA technique, we will reach GP approximation of (18). $\widetilde{\mathrm{C}}1.1, \widetilde{\mathrm{C}}6, \widetilde{\mathrm{C}}7$ are the monomial approximation for C1, C6, C7. First, we apply DC approximation of $R_{k,s,n}(\mathbf{P})$ at iteration τ_2. Therefore, (18) is converted into (see Appendix B)

$$\min_{\mathbf{P}(\tau_2)}(\sum_{b\in\mathcal{B}}\mathrm{C}_b\times bbu_b(\tau)+\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}}(x_{k,s,n}[\mathrm{C_p}\times p_{k,s,n}(\tau_2) + \frac{h_{k,s,n}}{\sigma^2+\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}} p_{k,s,n}(\tau_2-1)h_{k,s,n}}\times\frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)} - \frac{h_{k,s,n}}{\sigma^2+\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}} p_{k,s,n}(\tau_2-1)h_{k,s,n}}\times\frac{p_{k,s,n}(\tau_2)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2)} - \log_2(1+\frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)})])) \tag{19}$$

$$\text{subject to: } \widetilde{\mathrm{C}}1.1, \mathrm{C}4, \widetilde{\mathrm{C}}6, \widetilde{\mathrm{C}}7.$$

Now, based on (19) and Proposition 3, we derive the GP transformation of (18) for each iteration of Algorithm 1.B proposed in Table III.

Proposition 3: Consider the positive auxiliary variable $\varpi_1(\tau_2) > 0$ and $\Xi_2 \gg 1$. The objective function of (19) at each iteration τ_2 can be transformed into the following standard GP problem

$$\mathcal{O}_2(\mathbf{P}(\tau_2)) \quad : \min_{\mathbf{P}} \varpi_1(\tau_2) \tag{20}$$

$$\text{subject to :} \quad \mathrm{C}4,$$

$$\widetilde{\mathrm{C}}0: \ (\Xi_2+\sum_{b\in\mathcal{B}}\mathrm{C}_b\times bbu_b(\tau)+\mathrm{C_p}[\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} p_{k,s,n}(\tau_2)]+ \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}}\frac{h_{k,s,n}}{\sigma^2+\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}} p_{k,s,n}(\tau_2-1)h_{k,s,n}}\times\frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)}) (\frac{\varpi_1(\tau_2)}{a_0(\tau_2)})^{-a_0(\tau_2)}\prod_{k\in\mathcal{K},n\in\mathcal{N},s\in\mathcal{S}}\left[\frac{\log_2(1+\frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)})}{b_0(\tau_2)}\right]^{-b_0(\tau_2)} \prod_{\substack{k\in\mathcal{K},\\ n\in\mathcal{N},\\ s\in\mathcal{S}}}(\frac{\frac{h_{k,s,n}}{\sigma^2+\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}} p_{k,s,n}(\tau_2-1)h_{k,s,n}}\times\frac{p_{k,s,n}(\tau_2)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2)}}{d_0(\tau_2)})^{-d_0(\tau_2)}\le 1,$$

$$\widetilde{\mathrm{C}}1.1: \prod_{k\in\mathcal{K},s\in\mathcal{S}}(\sigma^2+I_{k,s,n}(\tau_2))\times\left[\frac{\sigma^2}{\lambda_0(\tau_2)}\right]^{-\lambda_0(\tau_2)}\times \prod_{k\in\mathcal{K},s\in\mathcal{S}}\left[\frac{p_{k,s,n}(\tau_2)h_{k,s,n}}{\lambda_{k,s}(\tau_2)}\right]^{-\lambda_{k,s}(\tau_2)}\le 2^{-R_n^{\mathrm{rsv}}},\quad \forall n\in\mathcal{N},$$

$$\widetilde{\mathrm{C}}6: \prod_{n\in\mathcal{N},s\in\mathcal{S}} y_{r,b}(\tau)x_{r,s,n}(\tau)(\frac{\sigma^2+I_s^{th}+p_{r,s,n}(\tau_2)h_{r,s,n}}{\sigma^2+I_s^{th}}) \le 2^{F_{r,b}^{\max}},\quad \forall r\in\mathcal{R},\forall b\in\mathcal{B},$$

$$\widetilde{\mathrm{C}}7: \prod_{r\in\mathcal{R},n\in\mathcal{N},s\in\mathcal{S}} y_{r,b}(\tau)x_{r,s,n}(\tau)(\frac{\sigma^2+I_s^{th}+p_{r,s,n}(\tau_2)h_{r,s,n}}{\sigma^2+I_s^{th}}) \le 2^{L_b^{\max}\times bbu_b(\tau)},\forall b\in\mathcal{B},$$

where

$$a_0(\tau_2) = \frac{\varpi_1(\tau_2-1)}{T}, \tag{21}$$

$$b_0(\tau_2) = \frac{\log_2(1+\frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)})}{T}, \tag{22}$$

$$d_0(\tau_2) = \frac{\frac{h_{k,s,n}}{\sigma^2+\sum_{k\in K}\sum_{n\in N} p_{k,s,n}(\tau_2-1)h_{k,s,n}} \times \frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)}}{T}, \quad (23)$$

$$\lambda_0(\tau_2) = \frac{\sigma^2}{\sigma^2 + \sum_{k\in\mathcal{K},s\in\mathcal{S}} p_{n,k,s}(\tau_2-1)h_{k,s,n}}, \quad (24)$$

$$\lambda_{k,s}(\tau_2) = \frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2 + \sum_{k\in\mathcal{K},s\in\mathcal{S}} p_{k,s,n}(\tau_2-1)h_{k,s,n}}, \quad (25)$$

and

$$T = \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} \log_2(1+\frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)}) + \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} (\frac{h_{k,s,n}}{\sigma^2+\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}} p_{k,s,n}(\tau_2-1)h_{k,s,n}} \times \frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)}) + \varpi_1(\tau_2-1).$$

Proof. See Appendix C. □

To simplify C6 and C7, we assume that interference for each sub-carrier $s \in \mathcal{S}$ is limited to the maximum aggregated value of I_s^{th}. By this assumption, the interference is considered in the worst-case condition [23] and [24]. While this assumption is not optimal, it can significantly reduce computational complexity. We apply iterative AGMA technique to get the monomial approximation for C1. The optimization problem $\mathcal{O}_2(\mathbf{P})$ is iteratively solved until the convergence criteria $\| \mathbf{P}^*(\tau_2) - \mathbf{P}^*(\tau_2-1) \| \leq \varepsilon_4$ are met. Similar to Proposition 2, we can conclude that Step 1.B of Algorithm 1 converges to a local optimal solution.

C. Traditional Algorithm

In this paper, Algorithm 2 is suggested for comparing performance of proposed Algorithm 1 which is summarized in Table IV. In traditional wireless networks such as forth generation of wireless networks (4G), each user is assigned to the AP (RRH/FAP) with the largest average received SINR [25] which is calculated based on reference signal received power (RSRP) broadcasted by APs. Therefore, the resource allocation problem can be formulated as

$$\max_{\mathbf{X}',\mathbf{Y},\mathbf{P},\mathbf{bbu}} [\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x'_{k,s,n} R_{k,s,n}(\mathbf{P}) - \mathrm{C_p} \times [\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x'_{k,s,n} p_{k,s,n}] - \sum_{b\in\mathcal{B}} \mathrm{C}_b \times bbu_b] \quad (26)$$
$$\text{subject to :} \quad \mathrm{C1}, \mathrm{C3}-\mathrm{C8},$$

where $x'_{k,s,n}$ shows the sub-carrier allocation of user $n \in \mathcal{N}$ on sub-carrier $s \in \mathcal{S}$ where user $n \in \mathcal{N}$ is allocated to the AP $k \in \mathcal{K}$. Since AP-user association is based on the largest average received SINR, C2 is removed. Therefore, (26) has four sets of variables, i.e., $\mathbf{X}'$, $\mathbf{Y}$, $\mathbf{P}$, **bbu**, and it has less computational complexity compared to (3). Clearly, (26) is non-convex. Similar to (3), we decompose it into two sub-problems and apply CGP to solve it. The iterative algorithm to solve (26) is summarized in Table IV. Algorithm 2 is based on CGP and similar to Algorithm 1 which develops the efficient iterative algorithm with two-step to solve the proposed problem. For a given fixed power allocation, Step 2.A derives the optimal UAF and C-RAN JRBF that consists of sub-carrier allocation, BBU selection and RRH assignment to active BBU. In Step 2.B, the power allocation problem for this case can be solved similar to $\mathcal{O}_2(\mathbf{P}(\tau_2))$. When the convergence criteria in Algorithm 2 are hold, this iterative algorithm is terminated.

TABLE IV
ALGORITHM2: TRADITIONAL ALGORITHM

Initialization: Set $\tau_3 = 0$, $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 10^{-3}$, $\mathbf{P}(\tau=0) = P_k^{\max}/S$, AP assignment: where user $n \in \mathcal{N}$ is assigned to AP $k \in \mathcal{K}$ based on largest average received SINR.

Repeat: $\tau_3 = \tau_3 + 1$

Step 2.A: Compute, $\mathbf{X}'^*(\tau_3)$, $\mathbf{Y}^*(\tau_3)$, $\mathbf{bbu}^*(\tau_3)$ by using Step 1.A, except that users assigned to AP based on the signal strength.

Step 2.B: For a given fixed set of $\mathbf{X}'^*(\tau_3)$, $\mathbf{Y}^*(\tau_3)$, and $\mathbf{bbu}^*(\tau_3)$ that obtained from Step 1.A, we find the optimal power allocation $\mathbf{P}^*(\tau_3)$ by using Step 1.B.

Until $\| \mathbf{X}'^*(\tau_3) - \mathbf{X}'^*(\tau_3-1) \leq \varepsilon_1 \|$ and $\| \mathbf{Y}^*(\tau_3) - \mathbf{Y}^*(\tau_3-1) \leq \varepsilon_2 \|$ and $\| \mathbf{bbu}^*(\tau_3) - \mathbf{bbu}^*(\tau_3-1) \leq \varepsilon_3 \|$ and $\| \mathbf{P}^*(\tau_3) - \mathbf{P}^*(\tau_3-1) \leq \varepsilon_4 \|$.

TABLE V
SIMULATION PARAMETERS

Parameter	Value
Cell radius	500 m
$\mathcal{K} = \mathcal{R} \bigcup \mathcal{F}$	10
Path loss exponent ($\jmath$)	3 dbm [26]
σ^2	[1,2] dbm
I_s^{th}	[1,2] dbm
$p_n^{\max}$	1Watt (normalized)
$F_{r,b}^{\max}$	[2,7] bps/Hz (normalized)
$L_b^{\max}$	[6,15] bps/Hz (normalized)
C_b	[2,24] bps/Hz (normalized)

IV. SIMULATION RESULTS

In this section, we evaluate the performance of Algorithm 1 and Algorithm 2 by simulation results. Consider $N = 30$ users located at random positions inside a region of 4×4 Km^2 and served by $B = 3$ BBUs, $R = 4$ RRHs and $F = 6$ FAPs. For all of the simulations, we set $\mathrm{C}_P = 1$ bps/Hz/Watt, $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 10^{-3}$. Also we consider the ratio of maximum transmit power of FAPs to maximum transmit power of RRHs is 0.5. The values of maximum BBU load, transmission front-haul link limitation and operation cost of BBUs are randomly chosen in the rang of $L_b^{\max} \in [6, 14]$ bps/Hz, $F_{r,b}^{\max} \in [2, 7]$ bps/Hz and $\mathrm{C}_b \in [2, 24]$ bps/Hz, respectively. Channel gain between user $n \in \mathcal{N}$ and AP $k \in \mathcal{K}$ is modeled as $h_{k,s,n} = \varsigma_{k,s,n} d_{k,n}^{-\jmath}$ where $\jmath = 3$ is the path loss exponent, $d_{k,n} > 0$ is the normalized distance between user $n \in \mathcal{N}$ and AP $k \in \mathcal{K}$ and $\varsigma_{k,s,n}$ is the exponential random variable with mean of 1 [26]. Also I_s^{th} for each sub-carrier $s \in \mathcal{S}$ is randomly chosen for each simulation in the rang of $[1, 2]$ dBm. The simulation parameters are summarized in Table V. For our proposed approach all simulation results are averaged over 50

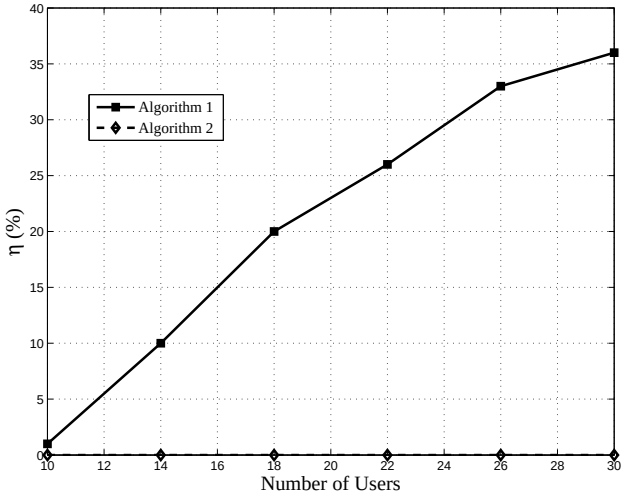

Fig. 2. Traffic offloading from RRHs to FAPs versus number of users.

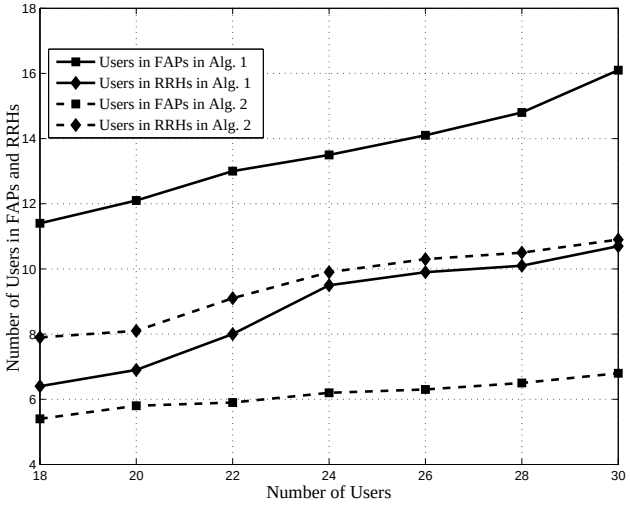

Fig. 3. Users association to RRHs and FAPs versus number of users.

random realization of channel gains, where at each realization, locations of UEs are randomly generated based on the uniform distribution.

In the followings, we evaluate and compare the total throughput, outage probability, traffic offloading and energy efficiency achieved by Algorithm 1 and Algorithm 2 for different system parameters. In Figs. 2 and 3, traffic offloading of BBUs by Algorithm 1 and Algorithm 2 are compared. Based on Algorithm 1, RRHs'users offload their traffic to neighbouring FAPs according to the minimum operation cost of BBUs. Therefore, under-utilized BBUs are switched off. Hence, the number of active BBUs and operation cost (energy consumption cost) of BBUs will be minimized. To study traffic offloading, we define network offloading ratio η, as

$$\eta = [\text{Associated UEs to FAPs by Algorithm1 or Algorithm 2 - Associated UEs to FAPs by Largest Received SINR}]/\text{Total Number of UEs},$$

where η denotes the ratio of total number of UEs that are moved from RRHs to FAPs and offload their traffic. Fig. 2 demonstrates that η is increased with increasing the number of users in Algorithm 1, while it is equal to zero in Algorithm 2. This is because in Algorithm 2, the AP assignment is predetermined. For instance, in Fig. 2, when $N = 26$, more than 30% of traffic is offloaded from RRHs to FAPs in Algorithm 1, while η is always equal to zero in Algorithm 2. Based on Algorithm 2, UEs are associated to RRHs/FAPs according to the largest average received SINR [25] while ignoring the maximum load limitation of BBUs, front-haul link limitations and minimizing operation cost of BBUs. Therefore, Algorithm 2 may exceed the BBUs and front-haul limitations during user association, resulting in reducing the performance of the network. However, UEs, especially for the cell-edge users in the dense RRH deployed H-CRAN, can be moved to the neighboring FAPs via Algorithm 1, and consequently, Algorithm 1 can efficiently manage interference and resources of network compared to Algorithm 2. Hence, EE and network coverage will be enhanced. This fact is illustrated in Fig. 3 where the number of UEs assigned to low-power

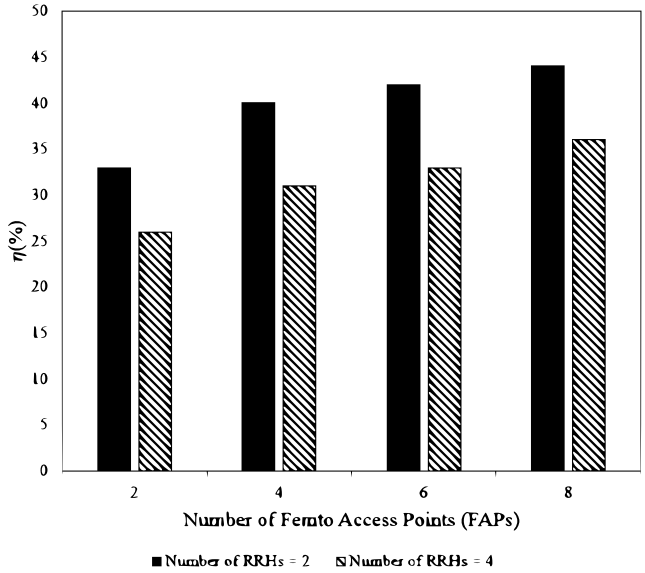

Fig. 4. Traffic offloading versus total number of access points (RRH/FAP).

FAPs are more than the number of UEs allocated to RRHs for Algorithm 1. However, the number of UEs assigned to RRHs are more than the number of UEs allocated to low power FAPs for Algorithm 2. This is because in Algorithm 1, UEs are associated to RRHs/FAPs based on reducing operation cost of BBUs; hence we have traffic offloading.

We investigate effect of increasing the number of APs (RRHs/FAPs) on the traffic offloading in our proposed Algorithm 1. In Fig. 4, we demonstrate traffic offloading from C-RAN to low cost FAPs versus the total number of APs (RRHs/FAPs). To show clear presentation of our approach, Fig. 4 is plotted for $R = 2$ and $R = 4$. We assume that in both cases RRHs are located in a 4×4 Km^2 square area. Fig. 4 shows that traffic offloading is increased with increasing the number of FAPs for both cases.

Besides, Fig. 4 illustrates when $R = 4$, traffic offloading is less than that of $R = 2$. This is because in the case of $R = 4$, users are closer to the RRHs, therefore, with less transmit power, they can achieve larger SINR and causing less

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2018.2882466, IEEE Transactions on Vehicular Technology

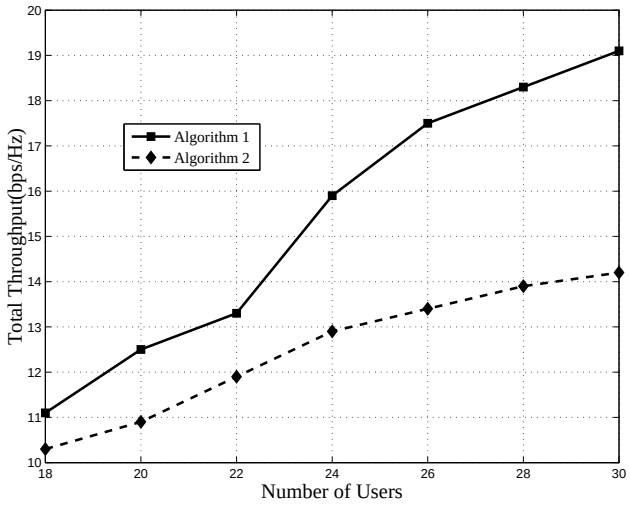

Fig. 5. Total Throughput versus number of users for $R_n^{\text{rsv}} = 0.3$ bps/Hz.

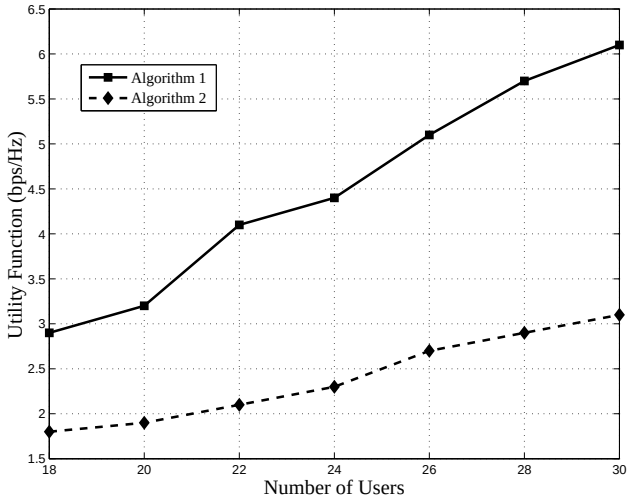

Fig. 6. Utility function versus number of users for $R_n^{\text{rsv}} = 0.3$ bps/Hz.

interference at the RRHs which leads to decrement of traffic offloading. In the simulation results in both cases, only one BBU with the lowest operation cost is in state on. This result along with Fig. 4 shows that $R = 4$ is over planned and $R = 2$ is enough for this case. To cover this point, and reduce energy consumption, we need to consider RRH switching off which is remained as a future work of this paper.

Figs. 5 and 6 illustrate total throughput and utility function versus the total number of UEs for Algorithm 1 and 2, respectively. Fig. 5 shows that the total throughput is increased with increasing the number of users for both algorithms, as expected from multiuser diversity gain [26]. Fig. 5 demonstrates that Algorithm 1 outperforms Algorithm 2 in terms of total throughput, because it can efficiently control interference between different APs. Additionally, femto-cells can reuse spectrum and offload traffic from RRHs to FAPs. Hence, Fig. 6 demonstrates that the utility function in Algorithm 1 is improved compared to that of Algorithm 2. Consider the outage probability of C1 as

$$\Pr(\text{outage}) = \Pr\{\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}R_{k,s,n}(\mathbf{P},\mathbf{X}) \leq R_n^{\text{rsv}}\}, \forall n \in \mathcal{N}.$$

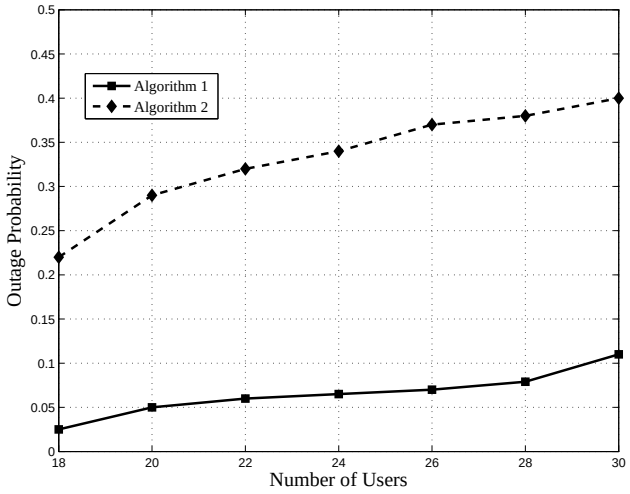

Fig. 7. Outage probability versus number of users.

In Fig. 7, we demonstrate outage probability versus the total number of UEs for Algorithm 1 and 2. In Fig. 7, with increasing the number of users, the outage probabilities of both Algorithm 1 and Algorithm 2 increase. However, the outage probability in Algorithm 2 is more than that of Algorithm 1. Due to traffic offloading and interference management between RRHs and FAPs, Algorithm 1 has more chance to choose AP and sub-carriers, and allocate transmit power among users while the AP assignment is predetermined in Algorithm 2. Consequently, the feasibility region of resource allocation in Algorithm 1 is larger than that of Algorithm 2, which can provide more achieved total rate and better coverage of users for the dense deployed H-CRAN. Fig. 7 also highlights the outage probability of users in Algorithm 1 is much less than that of Algorithm 2.

Table VI illustrates numerical results when N = 26 and $R_n^{\text{rsv}} = 0.3$ bps/Hz. Here, we consider two scenarios: heterogeneous energy consumption cost, i.e., $\text{C}_1 = 2$, $\text{C}_2 = 5$, $\text{C}_3 = 4$ and homogeneous energy consumption cost, i.e., $\text{C}_1 = 2$, $\text{C}_2 = 2$, $\text{C}_3 = 2$ in Table VI. In both scenarios, the simulation results illustrate that via Algorithm 1, only one BBU with the lowest operation cost is in state on. Furthermore, from the results it can be seen that via Algorithm 1, more than 30% of traffic is offloaded from the RRHs to FAPs and the outage probability is less than that of Algorithm 2. Also, Table VI demonstrates that via Algorithm 2, when the values of C_bs are different, only one BBU with the largest operation cost is in state on. However, when the values of C_b=2 for all $b \in \mathcal{B}$, all three BBUs are in state on for Algorithm 2. From the results of Table VI, it can be concluded that Algorithm 1 outperforms Algorithm 2 in terms of traffic offloading, outage probability, total throughput and operation cost of BBUs in both cases.

The simulation results show that in Algorithm 1, some UEs with lower transmit power can achieve a high data rate that is significantly higher than their required data rate, e.g., R_n^{rsv}, compared to that of Algorithm 2. For instance, via Algorithm 1, user 2 can achieve a rate of 1.6843 bps/Hz when its total transmit power is 0.5 Watt. However, in Algorithm 2, when its total transmit power is 0.7 Watt, its total throughput is

TABLE VI
COMPARING ALGORITHM 1 AND ALGORITHM 2 FOR $R_n^{\text{RSV}} = 0.3$ AND $N = 26$.

Algorithm	BBUs Cost			Active BBUs	$\eta(\%)$	Pr(outage)	Throughput	Total cost of BBUs
	BBU1	BBU2	BBU3					
Alg. 1	$C_1 = 2$	$C_2 = 5$	$C_3 = 4$	C_1	32	0.07	17.35	2
Alg. 2	$C_1 = 2$	$C_2 = 5$	$C_3 = 4$	C_2	0	0.3	11.8	5
Alg. 1	$C_1 = 2$	$C_2 = 2$	$C_3 = 2$	C_1	32	0.07	17.35	2
Alg. 2	$C_1 = 2$	$C_2 = 2$	$C_3 = 2$	C_1, C_2, C_3	0	0.3	11.8	6

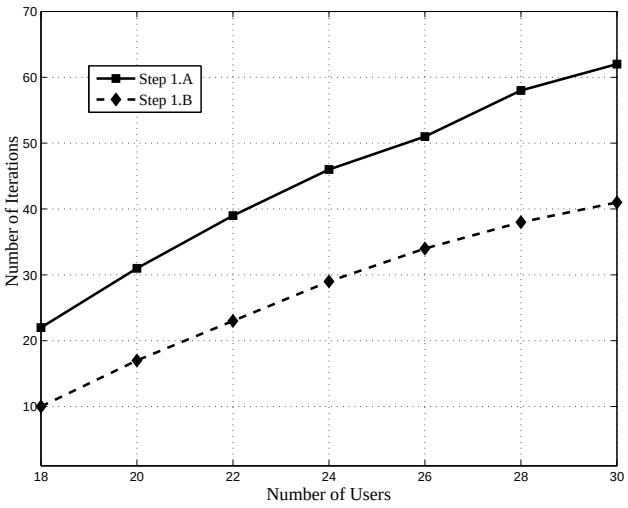

Fig. 8. Number of required iterations for Algorithm 1 versus number of users.

0.32 bps/Hz. It is because Algorithm 1 can offload traffic from C-RAN to FAPs with low transmit power, and manage interference between different APs.

Finally, from the results of this section, Algorithm 1 can considerably improve network performance in terms of EE, total throughput, operation cost of H-CRAN and outage probability as compared to those in Algorithm 2. This is because joint AP assignment and sub-carrier allocation manage the interference between different APs in Algorithm 1.

A. Computational Complexity

Here, we analyze the computational complexity of JRBF, UAF and power allocation sub-problems in Algorithm 1 via CVX [20]. Since CVX is used to solve GP sub-problems with the interior point method in Steps 1.A and 1.B, the number of required iterations is $\frac{\log(m/(\xi t^0))}{\log \jmath}$ where m is the total number of constraints, t^0 is the initial point for approximating the accuracy of interior point method, $0 < \xi < 1$ is the stopping criterion for interior point method, and $\jmath$ is used to update the accuracy of interior point method [27]. The number of constraints in (5) and (20) are $m_{\text{JRBF-UAF}} = 4KN + N + 2RB + B + R + KS + 1$ for Step 1.A and $m_{\text{PA}} = 2N + RB + B + 1$ for Step 1.B, respectively. Moreover, in Steps 1.A and 1.B, for each iteration, the number of computations required to convert the non-convex problems using AGMA into (5) and (20) are $i_{\text{JRBF-UAF}} = NK^2S + 3NKS + NK + 2RB$ and $i_{\text{PA}} = 4NKS + 2NRBS$, respectively. Therefore, the total number of computations for Algorithm 1 is

$$\begin{cases} i_{\text{JRBF-UAF}} \times \frac{\log(m_{\text{JRBF-UAF}}/(\xi_{\text{JRBF-UAF}} {t_{\text{JRBF-UAF}}}^0))}{\log \jmath_{\text{JRBF-UAF}}}, & \text{in Step 1.A,} \\ i_{\text{PA}} \times \frac{\log(m_{\text{PA}}/(\xi_{\text{PA}} {t_{\text{PA}}}^0))}{\log \jmath_{\text{PA}}}, & \text{in Step 1.B.} \end{cases} \quad (27)$$

Based on this analysis, the computational complexity of Step 1.A is significantly higher than that of Step 1.B. Moreover, Step 1.A is more sensitive to N compared to Step 1.B because the total number of constraints for Step 1.A is higher than that for Step 1.B. For further investigation by simulation, in Fig. 8, the number of iterations required for convergence for Steps 1.A and 1.B versus the total number of users is plotted for $R_n^{\text{rsv}} = 0.3$ bps/Hz, $C_1 = 2$, $C_2 = 5$, $C_3 = 4$ and $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 10^{-3}$. Fig. 8 illustrates that with increasing N, the number of iterations required for convergence of both Step 1.A and Step 1.B increases.

Additionally, Based on (27), it can be concluded that the computational complexity of JRBF-UAF and PA become $O((NK^2S + RB)\log(K(N + S) + RB))$ and $O(NS(K + RB)\log(N + RB))$, respectively, which are logarithm functions. As compared to exhaustive search algorithm, Algorithm 1 has less computational complexity. It is mainly because in the exhaustive search algorithm, with the aim of maximizing EE, for each possible combination of active BBUs, the RRHs are allocated to active BBUs and APs (RRH/FAP), sub-carrier and transmit power are jointly assigned to each user. The maximum network EE is achieved by searching all the combinations of active BBUs. Hence, the exhaustive search algorithm has an exponential complexity of $O(2^{K\times S\times N} + 2^{R\times B})$. However, the complexity of Algorithm 1 is much less than the exponential complexity of exhaustive search algorithm. Besides, we compare the utility function achieved by Algorithm 1 and exhaustive search versus $p_n^{\max}$ for $B = 2$, $R = 2$, $F = 2$, $S = 4$ and $N = 5$ users. From Fig. 9, it can be observed that the utility function of Algorithm 1 is approaching the exhaustive search solution by increasing $p_n^{\max}$, because the optimal solution obtained by AGMA approach is the best fit approximation for the large SINR scenario.

B. Summary of Simulation Results

The main simulation results of this paper are summarized as follows:

- We demonstrated that operation cost existing in our proposed utility function leads to offload traffic from under-utilized and high cost BBUs to neighboring FAPs with low transmit power, and migrates associated VNFs to low cost BBUs.

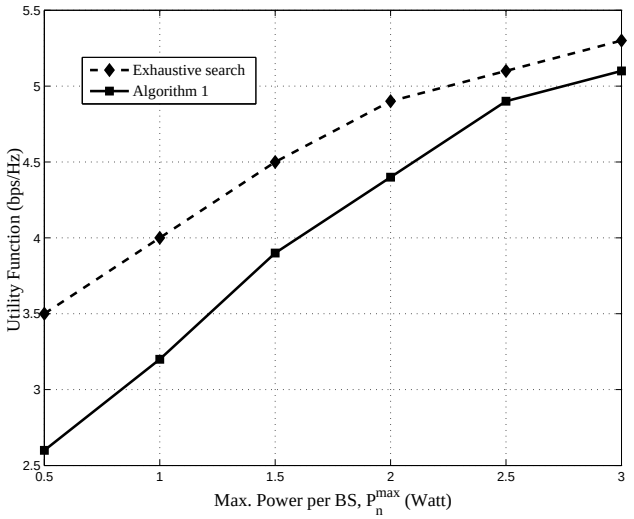

Fig. 9. Utility function versus $p_n^{\max}$ for exhaustive search and Algorithm 1.

- Due to traffic offloading from C-RAN to low cost FAPs, under-utilized BBUs can be switched off. Consequently the number of active BBUs and total operation cost will be reduced. Also, due to frequency reuse by femto-cells, total network throughput will be increased. On the other hand, new utility function leads to increase EE.
- Joint AP assignment and sub-carrier allocation manage the inter-tier interference between different APs. Therefore, the outage probability and coverage of UEs will be enhanced via Algorithm 1 compared to Algorithm 2.

V. Conclusion

In this paper, we investigate energy efficient optimization problem for H-CRANs. In the proposed approach, to reduce energy consumption cost, we define a novel utility function to minimize energy consumption cost of BBUs. We demonstrate via this utility function, associated VNFs of under-utilized BBUs migrate to low operation cost BBUs; also, associated UEs of under-utilized BBUs offload their traffic to neighboring FAPs with low transmit power. Consequently under-utilized BBUs and corresponding front-haul links will be switched off. We formulate this problem as a joint sub-carrier, RRH-BBU, front-haul and AP assignment and power allocation optimization problem in the OFDMA based H-CRANs while satisfying the minimum required rate of each user. The formulated optimization problem is non-convex and NP-hard with high computational complexity. By applying CGP and SCA Algorithm, we develop an efficient two-step iterative approach to dynamically assign resources. The simulation results show that our proposed approach is more efficient compared to the traditional scenario, in terms of increasing total throughput, reducing H-CRAN energy consumption cost, and decreasing outage probability of users. Also, simulation results show that our proposed algorithm can effectively offload the traffic from C-RAN to low-power FAPs, specifically for dense RRH deployed H-CRAN. Besides, in low traffic time, a significant number of RRHs will be under-utilized and served only low or even no traffic, but consume a considerable amount of energy based on our simulation results. To cover this point, and reduce energy consumption, we need to consider RRH switching off which is remained as a future work of this paper.

Appendix A
Proof of Proposition 1

Proof of Proposition 1 has four steps: 1) C2 converts to $\widetilde{\text{C}}2.1-\widetilde{\text{C}}2.4$, 2) C5 converts to $\widetilde{\text{C}}5$, 3) C8 converts to $\widetilde{\text{C}}8$, and 4) objective function (4) is transformed into (5).

1) At iteration τ_1, we define $z_{k,n}(\tau_1) = \sum_{s\in\mathcal{S}} x_{k,s,n}$ and $z_n^{'}(\tau_1) = \sum_{k\in\mathcal{K}}\sum_{s\in S} x_{k,s,n}$. Then we rewrite C2 as

$$z_{k,n}(\tau_1)[z_n^{'}(\tau_1) - z_{k,n}(\tau_1)] = 0,\ \forall n\in\mathcal{N},\ k\in\mathcal{K}, \quad (28)$$

which is not a monomial function. Now we use an auxiliary variable $q_{k,n}\geq 0$ to relax and convert (28) into the posynomial inequalities as

$$z_{k,n}(\tau_1)z_n^{'}(\tau_1)+1 \leq q_{k,n}(\tau_1) \leq z_{k,n}^2(\tau_1)+1,\ \forall n\in\mathcal{N},\ k\in\mathcal{K}. \quad (29)$$

(29) can be rewritten as $\frac{z_{k,n}(\tau_1)z_n^{'}(\tau_1)+1}{q_{k,n}(\tau_1)}\leq 1$ and $\frac{q_{k,n}(\tau_1)}{z_{k,n}^2(\tau_1)+1}\leq 1$. Since, above constraints are non-posynomial functions, we transform them to posynomial functions by AGMA approximation as follows

$$\widetilde{\text{C}}2.1:\ q_{k,n}^{-1}(\tau_1) + z_{k,n}(\tau_1)z_n^{'}(\tau_1)q_{k,n}^{-1}(\tau_1)\leq 1,$$

$$\widetilde{\text{C}}2.2:\quad q_{k,n}(\tau_1)\times\left[\frac{1}{\alpha_{k,n}(\tau_1)}\right]^{-\alpha_{k,n}(\tau_1)}\times\left[\frac{z_{k,n}^2(\tau_1)}{\beta_{k,n}(\tau_1)}\right]^{-\beta_{k,n}(\tau_1)} \leq 1, \forall n\in\mathcal{N},\forall k\in\mathcal{K},$$

where $\alpha_{k,n}(\tau_1)$ and $\beta_{k,n}(\tau_1)$ are defined in (7) and (8), respectively. Also, we have

$$\text{C}2.3:\quad z_{k,n}(\tau_1) = \sum_{s\in\mathcal{S}} x_{k,s,n},$$

$$\text{C}2.4: z_n^{'}(\tau_1) = \sum_{k\in\mathcal{K}}\sum_{s\in\mathcal{S}} x_{k,s,n}.$$

Note that, the equality constraints in C2.3 and C2.4 are not monomial since we have $z_{k,n}(\tau_1) - \sum_{s\in\mathcal{S}} x_{k,s,n} = 0$ and $z_n^{'}(\tau_1) - \sum_{k\in K}\sum_{s\in S} x_{k,s,n} = 0$, and, they have negative terms. To convert C2.3 and C2.4 to the monomial functions, we apply AGMA approximation as

$$\widetilde{\text{C}}2.3:\quad z_{k,n}(\tau_1)\times\prod_{s\in\mathcal{S}}\left[\frac{x_{k,s,n}(\tau_1)}{\lambda_{k,s,n}(\tau_1)}\right]^{-\lambda_{k,s,n}(\tau_1)} = 1, \quad \forall n\in\mathcal{N},\forall k\in\mathcal{K},$$

$$\widetilde{\text{C}}2.4:\quad z_{k,n}^{'}(\tau_1)\times\prod_{k\in\mathcal{K},\ s\in\mathcal{S}}\left[\frac{x_{k,s,n}(\tau_1)}{\mu_{k,s,n}(\tau_1)}\right]^{-\mu_{k,s,n}(\tau_1)} = 1, \quad \forall n\in\mathcal{N},\forall k\in\mathcal{K},$$

where $\lambda_{k,s,n}(\tau_1)$ and $\mu_{k,s,n}(\tau_1)$ are defined in (9) and (10), respectively.

2) To have a standard GP formulation, the equality constraint in C5 should only involve monomial functions. In the followings, we relax the binary variable as $y_{r,b} \in [0,1]$, and then, apply iterative AGMA algorithm. By adding 1 to both the left and right hand sides of C5, e.g., $\sum_{b\in B} y_{r,b} + 1 = 2$, we have $\frac{2}{1+\sum_{b\in B} y_{r,b}} = 1$. Now, by applying iterative AGMA algorithm, we get the monomial approximation for C5 as

$$\widetilde{\text{C}}5: \quad 2 \quad \times (\frac{1}{\theta(\tau_1)})^{-\theta(\tau_1)} \times \prod_{b\in\mathcal{B}} \left[\frac{y_{r,b}(\tau_1)}{\S_{r,b}(\tau_1)}\right]^{-\S_{r,b}(\tau_1)} = 1,$$

where $\theta(\tau_1)$ and $\S_{r,b}(\tau_1)$ are defined in (11) and (12), respectively.

3) The equality constraint in C8 is not standard GP formulation, because it should be involved positive and only monomial functions. Similar to C5, by adding 1 to both the left and right hand sides of C8, e.g., $y_{r,b} + 1 \leq bbu_b \times L_b^{\max} + 1$, we can apply iterative AGMA algorithm as

$$\widetilde{\text{C}}8: (y_{r,b}+1) \times (\frac{1}{\delta(\tau_1)})^{-\delta(\tau_1)} \times (\frac{bbu_b(\tau_1)}{\phi(\tau_1)})^{-\phi(\tau_1)} \leq 1, \quad \forall r \in \mathcal{R}, \forall b \in \mathcal{B},$$

where $\delta(\tau_1)$ and $\phi_{r,b}(\tau_1)$ are defined in (16) and (17), respectively.

4) In standard GP formulation, the aim of optimization problem is to minimize the positive objective function [27]. To transform (4) to GP form, we rewrite it as

$$\min_{\mathbf{X}(\tau_1),\mathbf{Y}(\tau_1),\mathbf{bbu}(\tau_1)} [\sum_{b\in\mathcal{B}} \text{C}_b \times bbu_b(\tau_1) + \text{C}_\text{p}[\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}p_{k,s,n}(\tau)] - \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}R_{k,s,n}(\mathbf{P}(\tau))]. \tag{30}$$

Since there are negative terms in (30), to meet the positive conditions of objective function in GP, we consider the positive auxiliary variable $\varpi_0(\tau_1)$ and $\Xi_1 \gg 1$, and rewrite the objective function as

$$\frac{\Xi_1 + \text{C}_\text{p}[\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}p_{k,s,n}(\tau)] + \sum_{b\in\mathcal{B}} C_b \times bbu_b(\tau_1)}{\varpi_0(\tau_1) + \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}(\tau_1)R_{k,s,n}(\mathbf{P}(\tau))} \leq 1, \tag{31}$$

which is always positive. Now, by applying AGMA approximation, (31) is transformed into

$$[\Xi_1 + \text{C}_\text{p}(\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}p_{k,s,n}(\tau)) + \sum_{b\in\mathcal{B}} C_b \times bbu_b(\tau_1)] \times (\frac{\varpi_0(\tau_1)}{\psi_0(\tau_1)})^{-\psi_0(\tau_1)} \times \prod_{k\in\mathcal{K},n\in\mathcal{N},s\in\mathcal{S}} (\frac{x_{k,s,n}(\tau_1)R_{k,s,n}(\mathbf{P}(\tau))}{w_{k,s,n}(\tau_1)})^{-w_{k,s,n}(\tau_1)} \leq 1, \tag{32}$$

where in each iteration, $[\psi_0(\tau_1)]$ and $[w_{k,s,n}(\tau_1)]$ are updated from (14) and (15), respectively. Thus, the equivalent optimization problem can be transformed into (5).

APPENDIX B
PROOF OF (19)

Similar to (4), the objective function (18) is transformed into

$$\min_{\mathbf{P}(\tau_2)} [\sum_{b\in\mathcal{B}} \text{C}_b \times bbu_b(\tau) + \text{C}_\text{p}[\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}p_{k,s,n}(\tau_2)] - \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} x_{k,s,n}R_{k,s,n}(\mathbf{P}(\tau_2))]. \tag{33}$$

Objective function (33) is not in a GP standard form because throughput is logarithm function which is non-linear function. We apply DC-approximation to solve this issue. Consider linear approximation of $\log_2(\upsilon_j(\tau_2))$ as

$$\log_2(\upsilon_{k,s,n}(\tau_2)) \approx \log_2(\upsilon_{k,s,n}(\tau_2-1)) + \nabla\log_2(\upsilon_{k,s,n}(\tau_2-1))(\upsilon_{k,s,n}(\tau_2) - \upsilon_{k,s,n}(\tau_2-1)),$$

where $\upsilon_{k,s,n}(\tau_2) = 1 + \frac{p_{k,s,n}(\tau_2)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2)}$. Then, by some mathematical manipulation, we obtain

$$\log_2(1+\frac{p_{k,s,n}(\tau_2)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2)}) \approx \log_2(1+\frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)}) + \frac{h_{k,s,n}}{\sigma^2+\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}} p_{k,s,n}(\tau_2-1)h_{k,s,n}}(\frac{p_{k,s,n}(\tau_2)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2)} - \frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)}). \tag{34}$$

With substituting (34) into (33), we have (19).

APPENDIX C
PROOF OF PROPOSITION 3

Proof of Proposition 2 has two steps: 1) C1 converts to $\widetilde{\text{C}}1.1$, and 2) objective function (19) is transformed into (20).

1) We can rewrite C1 as: $\log_2(\prod_{\substack{k\in\mathcal{K}\\ s\in\mathcal{S}}} (1+\frac{p_{k,s,n}(\tau_2)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2)})^{-1}) \leq -R_n^{\text{rsv}}$, which can be mathematically represented as

$$\prod_{\substack{k\in\mathcal{K}\\ s\in\mathcal{S}}} (\frac{\sigma^2+I_{k,s,n}(\tau_2)}{\sigma^2+I_{k,s,n}(\tau_2)+p_{k,s,n}(\tau_2)h_{k,s,n}}) \leq 2^{-R_n^{\text{rsv}}}.$$

By applying AGMA approximation, we reach to $\widetilde{\text{C}}1.1$ in $\mathcal{O}_2(\mathbf{P})$.

2) Since there are negative terms in (19), to meet the positive conditions of objective function in GP, we consider the positive auxiliary variable $\varpi_1(\tau_2)$ and $\Xi_2 \gg 1$, and rewrite the objective function as

C0 :

$$\Xi_2 + \sum_{b\in\mathcal{B}} \text{C}_b \times bbu_b(\tau) + \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} [\text{C}_\text{p} \times p_{k,s,n}(\tau_2) + \frac{h_{k,s,n}}{\sigma^2+\sum_{k\in\mathcal{K}}\sum_{n\in N} p_{k,s,n}(\tau_2-1)h_{k,s,n}} \times \frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)} -$$

$$\frac{h_{k,s,n}}{\sigma^2+\sum\limits_{k\in\mathcal{K}}\sum\limits_{n\in\mathcal{N}}p_{k,s,n}(\tau_2-1)h_{k,s,n}}\times\frac{p_{k,s,n}(\tau_2)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2)}-$$
$$\log_2(1+\frac{p_{k,s,n}(\tau_2-1)h_{k,s,n}}{\sigma^2+I_{k,s,n}(\tau_2-1)})]\leq\varpi_1(\tau_2), \quad (35)$$

which is always positive constraint. Finally, we reach to the following equivalent optimization problem

$$\min_{\mathbf{P}}\ \varpi_1(\tau_2) \quad (36)$$
$$\text{subject to :} \qquad \mathrm{C0},\mathrm{C1},\mathrm{C4},\mathrm{C6},\mathrm{C7}.$$

Since constraints $\mathrm{C0},\mathrm{C1},\mathrm{C6},\mathrm{C7}$ cannot satisfy the properties of monomials and posynomials in GP formulations, we apply AGMA approximations for these constraints to attain $\mathcal{O}_2(\mathbf{P})$.

REFERENCES

[1] C. Berger, M. Stubert, K. Georgios, D. Lars, and C. H. Lehrmann, *Cloud Radio Access Network architecture Towards 5G mobile networks*. Technical University of Denmark, PhD Thesis, 2016.

[2] E. Hossain and M. Hasan, "5G cellular: Key enabling technologies and research challenges," *IEEE Instrumentation and Measurement Magazine*, vol. 18, no. 3, pp. 11–21, May 2015.

[3] 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-April 2015.

[4] Z. Chang, Z. Han, and T. Ristaniemi, "Energy efficient optimization for wireless virtualized small cell networks with large scale multiple antenna," *IEEE Transactions on Communications*, vol. 65, no. 4, pp. 1696 – 1707, February 2017.

[5] M. Peng, K. Zhang, J. Jiang, J. Wang, and W. Wang, "Energy efficient resource assignment and power allocation in heterogeneous cloud radio access networks," *IEEE Transactions on Vehicular Technolgy*, vol. 64, no. 11, pp. 5275–5287, November 2015.

[6] M. Fallgren, "An optimization approach to joint cell, channel and power allocation in multicell relay networks," *IEEE Transaction Wireless Communication*, vol. 11, no. 8, pp. 2868–2875, June 2012.

[7] T. Z. Oo, N. H. Tran, W. Saad, D. Niyato, Z. Han, and C. S. Hong, "Offloading in HetNet: A coordination of interference mitigation, user association and resource allocation," *IEEE Transactions on Mobile Computing*, vol. 16, no. 8, pp. 2276–2291, August 2017.

[8] Z. Zhou, M. Dong, K. Ota, G. Wang, and L. T. Yang, "Energy-efficient resource allocation for D2D communications underlaying cloud-RAN-based LTE-A networks," *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 223–234, June 2016.

[9] M. Y. Lyazidi, N. Aitsaadi, and R. Langar, "Dynamic resource allocation for cloud-RAN in LTE with real-time BBU/RRH assignment," *in Proceeding 2016 IEEE International Conference on Communication (ICC)*, pp. 1–6, May 2016.

[10] B. Wang, Q. Yang, L. T. Yang, and C. Zhu, "On minimizing energy consumption cost in green heterogeneous wireless networks," *Computer Networks*, pp. 1–14, March 2017.

[11] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE Journal on Selected Area in Communication*, vol. 34, no. 4, pp. 1037 – 1050, June 2016.

[12] Y. S. Soh, T. Q. S. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *IEEE Transactions on Communications*, vol. 31, no. 5, pp. 840–850, May 2013.

[13] R. Mijumbi, J. Serrat, J. Gorricho, J. Rubio-Loyola, and S. Davy, "Server placement and assignment in virtualized radio access networks," *in Proceeding 11th International Conference On Network and Service Management (CNSM)*, pp. 398–401, November 2015.

[14] M. Qian, W. Hardjawana, J. Shi, and B. Vucetic, "Baseband processing units virtualization for cloud radio access networks," *IEEE Wireless Communucations Letters*, vol. 4, no. 2, pp. 189–192, April 2015.

[15] X. Wang, S. Thota, M. Tornatore, H. S. Chung, H. H. Lee, S. Park, and B. Mukherjee, "Energy-efficient virtual base station formation in optical-access-enabled cloud-RAN," *IEEE Journal of Selected Area in Communucations*, vol. 34, no. 5, pp. 1130–1139, May 2016.

[16] S. Parsaeefard, 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, April 2016.

[17] M. Chiang, "Geometric programming for communication systems," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 1-2, pp. 1–154, 2005.

[18] M. Chiang, C. W. Tan, D. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Transactions Wireless Communication*, vol. 6, no. 7, pp. 2640–2651, July 2007.

[19] G. Xu, "Global optimization of signomial geometric programming problems," *European Journal of Operational Research*, vol. 233, no. 4, pp. 500–510, July 2014.

[20] I. CVX-Research, "CVX: Matlab software for disciplined convex programming, version 2.0," *http://cvxr.com/cvx*, August.

[21] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, "Joint subchannel assignment and power allocation for OFDMA femtocell networks," *IEEE Transactions on Wireless Communication*, vol. 13, no. 1, pp. 342–355, January 2014.

[22] M. Razaviyayn, "Successive convex approximation: Analysis and applications," *Ph.D. dissertation, Dept. Elect. Eng., Univ. Minnesota. Minneapolis, MN, USA*, p. [Online]. Available: (http://hdl.handle.net/11299/163884), May 2014.

[23] A. Abdelnasser, E. Hossain, and D. I. Kim, "Tier-aware resource allocation in OFDMA macrocell-small cell networks," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 695–710, March 2015.

[24] M. Hasan and E. Hossain, "Distributed resource allocation in D2D-enabled multi-tier cellular networks: An auction approach," *in Proceeding 2015 IEEE International Conference on Communication (ICC)*, pp. 2949–2954, September 2015.

[25] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communication*, vol. 32, no. 6, pp. 1100–1113, June 2014.

[26] A. Goldsmith, "Wireless comunications. cambridge," *U.K.: Cambridge Univ. Press*, 2004.

[27] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, 2009.

Nahid Amani received her B.S. degree from Amirkabir University of Technology; the M.S. degree from Iran University of Science and Technology, Tehran, Iran, in 2000 and 2003, respectively, all in computer engineering. She is currently pursuing the Ph.D. degree in Computer Engineering in Amirkabir University of Technology and a faculty member in the Iran Telecommunication Research Center. Her current research area includes resource allocation in wireless networks.

Hossein Pedram received his BS degree in Electrical Engineering from Sharif University in 1979, and MS and PhD degrees from Ohio State University and Washington State University respectively in Computer Engineering. He served as a faculty member in Amirkabir University of Technology from 1992 to 2018. His research interests include data communications, nano-scale digital circuits and distributed systems.

Hassan Taheri received his B.S. degree from Amirkabir University of Technology, Tehran, Iran; the M.S. and Ph.D. degrees from University of Manchester Institute of Science and Technology, Manchester, U.K., in 1978 and 1988, respectively, all in electrical engineering. He is an associate professor in department of electrical engineering, Amirkabir University of Technology, Tehran, Iran. His research interests include different aspects of fixed and mobile communication networks.

Saeedeh Parsaeefard (S'09–M'14) received the Ph.D. degree in electrical and computer engineering from Tarbiat Modares University, in 2012. She was a Post-Doctoral Research Fellow in the Department of Electrical and Computer Engineering at the McGill University, Canada. From November 2010 to October 2011, she was a Visiting Ph.D. Student in UCLA, USA. She is currently a faculty member in the Iran Telecommunication Research Center. She received the IEEE Iran Section Women in Engineering (WIE) awards in 2018.