Wireless Network Virtualization with Non-Orthogonal Multiple Access

Tai Manh Ho[1], Nguyen H. Tran[1], S.M Ahsan Kazmi[1], Zhu Han[1,2], and Choong Seon Hong[1]

[1]Department of Computer Science and Engineering, Kyung Hee University, 446-701, Republic of Korea
[2] Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204-4005, USA
[1]Email: {hmtai,nguyenth,ahsankazmi,cshong}@khu.ac.kr,
[2]Email: zhan2@mail.uh.edu

***Abstract*—We study the problem of joint user clustering and resource allocation for wireless network virtualization (WNV) using non-orthogonal multiple access (NOMA). We aim to maximize the weighted total sum-rate while taking into account the *isolation constraint* of the mobile virtual network operators (MVNOs). To solve the non-convex formulated problem, we decouple it into three subproblems, i.e., user clustering, resource block (RB) allocation and power assignment. We apply the framework of *matching game with externalities* to solve the user clustering problem while the solutions for RB and power allocation are derived by using the Lagrange dual approach and complementary Geometric programming, respectively. An alternative maximization algorithm is provided to achieve a suboptimal solution for the original problem. We propose to classify user equipments (UEs) into three classes, i.e., strong, normal and weak UEs and compare our proposed scheme with general NOMA scheme with two UEs per cluster. Simulation results revel a performance gain of 2.5% in terms of throughput. Moreover, the proposed scheme outperforms the traditional OFDMA scheme in terms of throughput and energy efficiency by up to 40% and 58%, respectively.**

Index Terms*—*Wireless network virtualization, non-orthogonal multiple access, user clustering, resource allocation.

I. Introduction

Wireless network virtualization (WNV) is considered among the key technologies to fulfill the stringent requirements of the forthcoming fifth generation (5G) cellular networks. WNV can enhance the data rates, spectrum/energy efficiency, lower the costs and end-to-end latency [1]-[3]. Moreover, abstraction and sharing of resources among different virtual parties can also be enabled. However, the deluge of data traffic in mobile Internet and the tsunami of mobile devices demand high spectrum efficiency and massive connectivity in 5G wireless communications. Furthermore, traditional orthogonal multiple access (OMA) schemes (i.e., OFDMA scheme) have a big disadvantage as it limits the number of served user equipments (UEs) by the number of spectrum resources, i.e., subchannels, resource blocks (RBs). This limitation can be addressed by using the Non-orthogonal multiple access (NOMA) scheme which is considered as a key enabling technique for 5G cellular systems [8], [9]. NOMA scheme alleviates the aforementioned challenge of OMA schemes and boosts WNV development by exploiting the spectrum sharing for guaranteeing isolation among multiple mobile virtual network operators (MVNOs). In NOMA, multiple UEs can be multiplexed into transmission power domain by exploiting the channel gain differences. These UEs can then be simultaneously scheduled over same RBs non-orthogonally. Furthermore, through NOMA scheme, we can achieve massive device connectivity when compared to traditional OMA schemes.

In this paper, we choose the NOMA technique in resource allocation for WNV to enhance the spectrum efficiency. The motivation behind applying NOMA comes from the spectrum sharing fact that can be applied for grouping multiple MVNOs users into same frequency band and time slots (RBs). Therefore, an optimization problem is formulated such that MVNOs' UEs are grouped into different NOMA clusters and each NOMA cluster will be allocated the appropriate resources (RBs and transmission power) based on their QoS requirements (i.e., isolation constraint in WNV). The formulated problem is a combinatorial mixed integer non-convex problem. Therefore, we decouple it into three subproblems, namely user clustering, RB allocation and power assignment.

The first subproblem of user clustering is solved via a one-to-many matching game with externalities [18], [19] as matching games provide practically efficient solutions for combinatorial problems. Then, for a given set of NOMA clusters, the second subproblem of RB allocation is solved via the Lagrange dual based approach. Here, we find the optimal RB allocation for each NOMA cluster. Finally, the complementary geometric programming (CGP) and arithmetic-geometric mean inequality (AGMA) are applied to solve the non-convex subproblem of power assignment. Finally, a joint algorithm is provided to solve all three subproblems iteratively that converges to a suboptimal solution of the proposed original problem. Moreover, extensive simulations are provided to validate the superiority of our proposal in terms of network throughput and energy efficiency when compared to the the traditional OFDMA scheme in WNV.

The rest of this paper is organized as follows. Section II presents the related works. Section III presents the system model and problem formulation. In Section IV, we apply the matching game to solve the user clustering subproblem. We derive solutions for RB allocation and power allocation subproblems in Sections V and VI, respectively. The joint algorithm is presented in Section VI.B, and the simulation results are provided in Section VII. Finally, we conclude the paper in Section VIII.

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2015-0-00567, Development of Access Technology Agnostic Next-Generation Networking Technology for Wired-Wireless Converged Networks) *Dr. CS Hong is the corresponding author

978-1-5386-3416-5/18/$31.00

II. Related Works

1) Wireless Network Virtualization: Resource allocation (RA) for WNV has been investigated in several works recently [4]-[7]. WNV in LTE is addressed in [5], [6]. In these works, resource blocks are allocated based on flexible service level agreements for each service provider (SP) expressed as a minimum bandwidth allocation. A hierarchical combinatorial auction mechanism is provided in [4], which is based on a truthful and sub-efficient RA framework. In [7], the authors propose a virtual RA scheme for OFDMA based WNV and show that the Pareto optimal allocation can be achieved based on the market equilibrium price theory. However, all these works consider resource allocation for WNV in OFDMA schemes, which is limited by number of spectrum resources. In the OFDMA scheme, UE connectivity drops significantly when there are large number of UEs in the network, i.e., the number of UEs exceeds the number of spectrum resources. Therefore, the performance of these schemes will dramatically decrease in dense network scenarios.

2) NOMA: The basic concept of NOMA with successive interference cancellation (SIC) receiver is introduced in [9]. The joint user clustering and power allocation for both downlink and uplink is discussed in [10]. In [11], a joint subchannel assignment and power allocation to maximize the weighted sum-rate is formulated. A many-to-many two-sided UE-subchannel matching game is proposed to achieve a stable matching. The work in [12] is the only recent study which addresses resource allocation in WNV using NOMA; however, the authors only consider the problem of power allocation *without considering the channel assignment problem, i.e., user clustering*. In all these works, except [12], either a minimum data rate constraint is strictly imposed so as to guarantee QoS of each individual UE [10] or this constraint is ignored [11]. The strict instantaneous QoS constraints (expressed in terms of a minimum data rate) for individual UEs cannot be guaranteed in case when a UE has a bad channel condition (e.g., cell edge UEs). Therefore, supporting instantaneous QoS for each UE can result in an infeasible solution, even with a high transmit power and/or a large bandwidth allocation (i.e., large number of RBs).

III. System Model and Problem Formulation

A. Wireless Network with NOMA

We consider the downlink of a single cell consisting of one macro base station (MBS). The MBS and spectrum are owned and managed by an infrastructure provider (InP) who provides its virtual network services to a set of slices $\mathcal{N}$, (i.e., MVNOs) by individual contracts. A slice $n \in \mathcal{N}$ provides service to a set $\mathcal{U}_n$ of subscribed UEs. Let $\mathcal{U} \triangleq \cup_{n=1}^{|\mathcal{N}|}\mathcal{U}_n$ denote the set of all UEs. Moreover, the InP owns a system bandwidth $\mathcal{C}$ which is divided into Ω RBs, each of bandwidth B.

In our model, the UEs who are packed or scheduled over non-orthogonal set of the same RBs form a NOMA cluster. Each NOMA cluster operates on a set of RBs which is orthogonal to other sets of RBs allocated to other clusters (Fig. 1). Furthermore, the number of UEs per NOMA cluster ranges between 2 and $|\mathcal{U}|$. Finally, the number of RBs allocated to the k-th cluster is represented by δ^k where $1 \leq \delta^k \leq \Omega$.

Let $\mathcal{S}$ be the set of clusters and $\mathcal{S}_k$ be the set of active UEs grouped into the k-th cluster. The maximum MBS transmission

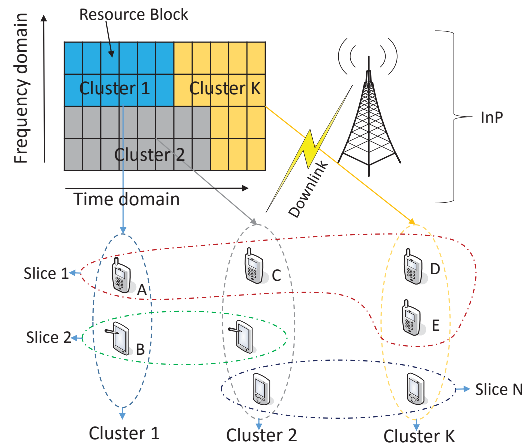

Fig. 1: System model: In this scenario, users A and B are set to the same NOMA cluster 1 due to their unbalanced channel gain. On the other hand, users A, C, D, and E in slice 1 must be guaranteed a minimum rate requirement (i.e., slice isolation), due to the MVNO's service level agreement.

power budget is P_T. The power allocated to UE $j_n \in \mathcal{U}_n$ is denoted by P_{j_n}. The complex coefficient of channel between UE j_n and the MBS is denoted by $h_{j_n} = \chi_{j_n}/\mathcal{D}(d_{j_n})$, where χ_{j_n} denotes the Rayleigh fading channel gain, $\mathcal{D}(\cdot)$ is the path loss function, and d_{j_n} is the geographical distance between UE j_n and the MBS. Let x_{j_n} be the transmitted symbol of UE j_n. The signal that UE j_n receives from the MBS in the k-th cluster is then given by

$$y_{j_n}^k = h_{j_n}\sqrt{P_{j_n}}x_{j_n}^k + \sum_{i \neq j | i \in \mathcal{S}_k} h_{j_n}\sqrt{P_{i_m}}x_{i_m}^k + z_{j_n}, \quad (1)$$

where z_{j_n} is the additive white Gaussian noise.

The optimal order of SIC decoding is in the order of the increasing channel gains normalized by the noise. To be specific, the receiver of UE $j_n \in \mathcal{S}_k$ can cancel the interference from any other UE $i_m \in \mathcal{S}_k$ with channel gain $|h_{i_m}|^2/z_{i_m} < |h_{j_n}|^2/z_{j_n}$, i.e., UE j_n first decodes the signal from UE i_m, then subtracts it and decodes its target signal $x_{j_n}^k$ correctly from the received signal $y_{j_n}^k$ in the k-th cluster. We now define a user clustering variable $\beta_{j_n}^k$ as follows:

$$\beta_{j_n}^k = \begin{cases} 1, & \text{if UE } j_n \text{ is grouped into cluster } k, \\ 0, & \text{otherwise.} \end{cases}$$

The achievable throughput for UE j_n in the downlink NOMA k-th cluster can be expressed as:

$$R_{j_n}^k = \delta^k B \log_2\left(1 + \frac{P_{j_n}|h_{j_n}|^2}{I_{j_n}^k + \delta^k B z_{j_n}}\right), \quad (2)$$

where $I_{j_n}^k$ is the interference that UE $j_n \in \mathcal{U}_n$ receives due to the other UEs in the k-th cluster.

$$I_{j_n}^k = \sum\nolimits_{i_m \in \mathcal{U} | \frac{|h_{i_m}|^2}{z_{i_m}} > \frac{|h_{j_n}|^2}{z_{j_n}}} \beta_{i_m}^k P_{i_m}|h_{j_n}|^2. \quad (3)$$

B. Wireless Virtualization Model

Isolation at the physical resource level can be implemented in different manners. The first is a static fixed sharing scheme in which each slice is preassigned a fixed subset of physical resources in different domains, and the access is restricted within this fixed subset. The second is a general dynamic sharing scheme that imposes no restriction on the resource access, while the isolation is achieved by guaranteeing certain pre-determined requirements or contract service agreement (e.g., minimum share of resource or minimum data rate) [4]. In this work, we adopt the second isolation scheme, i.e., the dynamic sharing scheme in which the InP guarantees service contract agreements with MVNOs by a minimum data rate constraint [4], [12]. We denote $R_n^{\min}$ as the minimum data rate requirement of the n-th slice. Then the isolation constraint for each slice n, i.e., the n-th slice is given as follows

$$\sum_{j_n \in \mathcal{U}_n} \sum_{k \in \mathcal{S}} \beta_{j_n}^k R_{j_n}^k \geq R_n^{\min},\ \forall n \in \mathcal{N}. \tag{4}$$

C. Problem Formulation

We consider the network sum-rate objective function and using a weight factor ω_{j_n} for UE j_n to adjust UEs' priority. The joint user clustering, RB allocation and power assignment problem for weighted sum-rate maximization in downlink NOMA can be formulated as:

$$\begin{aligned}
&\max_{\boldsymbol{\beta},\boldsymbol{\delta},\mathbf{P}} \sum_{n \in \mathcal{N}} \sum_{j_n \in \mathcal{U}_n} \omega_{j_n} \sum_{k \in \mathcal{S}} \beta_{j_n}^k \delta^k B \log_2 \left(1 + \frac{P_{j_n}|h_{j_n}|^2}{I_{j_n}^k + \delta^k B z_{j_n}}\right) \\
&\text{s.t.:}\ \ C_1: \sum_{n \in \mathcal{N}} \sum_{j_n \in \mathcal{U}_n} \sum_{k \in \mathcal{S}} \beta_{j_n}^k P_{j_n} \leq P_T, \\
&C_2: \sum_{j_n \in \mathcal{U}_n} \sum_{k \in \mathcal{S}} \beta_{j_n}^k \delta^k B \log_2 \left(1 + \frac{P_{j_n}|h_{j_n}|^2}{I_{j_n}^k + \delta^k B z_{j_n}}\right) \geq R_n^{\min},\ \forall n \\
&C_3: \sum_{k \in \mathcal{S}} \beta_{j_n}^k = 1,\ \forall j_n \in \mathcal{U}_n, n \in \mathcal{N}, \\
&C_4: 2 \leq \sum_{n \in \mathcal{N}} \sum_{j_n \in \mathcal{U}_n} \beta_{j_n}^k \leq |\mathcal{U}|,\ \forall k, \\
&C_5: \sum_{k \in \mathcal{S}} \beta_{j_n}^k \delta^k \leq \Omega,\ \forall j, n, \\
&C_6: \delta^k \in \{1, 2, ..., \Omega\},\ \beta_{j_n}^k \in \{0, 1\},\ \forall k, j, n.
\end{aligned} \tag{5}$$

where $\boldsymbol{\beta} \triangleq \{\beta_{j_n}^k\},\ \forall j_n \in \mathcal{U}, k \in \mathcal{S}$ is the user clustering matrix, $\boldsymbol{\delta} \triangleq \{\delta^k\},\ \forall k \in \mathcal{S}$ is the RB allocation vector, and $\mathbf{P} \triangleq \{P_{j_n}\},\ \forall j_n \in \mathcal{U}$ is the power allocation vector. Constraint C_1 denotes the total power constraint of the MBS and C_2 ensures the isolation requirement of the slices. User clustering constraints C_3 and C_4 ensures that one UE can be assigned to at most one cluster, while at least two UEs are grouped into each downlink NOMA cluster. Constraint C_5 provides the total downlink spectrum resource constraint.

The problem in (5) is a mix-integer non-convex optimization problem due to the integer constraint in C_6 and the existence of the interference term in the objective function. To tackle the above problem, we decouple the user clustering and resource allocation problems, and propose a joint solution in which the user clustering and resource allocation problems are solved iteratively.

IV. Matching Game for User Clustering (NOMA Clustering)

In NOMA, UEs with significantly different channel gains over a RB are grouped together. However, in our case, we find a cluster including a number of RBs shared by a group of UEs. Therefore, our aim is to find a set of UEs that can be grouped into the same cluster. Note that the number of clusters in a network depends on the network UEs' channel conditions, i.e., a large number of good channel UEs (or bad channel UEs) cannot be grouped together as they would experience strong interference. We classify the network UEs into three[1] classes of UEs: A, B and C corresponding to strong, normal, and weak UEs depending on their respective channel gains. Our choice of classifying UEs into three classes comes from the intuition that a large number of UEs per cluster will increase the SIC receiver complexity, whereas a small number of UEs per cluster (i.e., two classes) increases the number of required clusters, thus increasing the convergence time.

A. User Classification

In this subsection, first, we classify UEs and then find the number of available clusters in the network. The intuition comes from the fact that in NOMA, we try to group UEs into one cluster that have significantly different channel gains. Next, we describe our novel approach for user classification and cluster formation.

Firstly, UEs are sorted in descending order of channel gains. Then, we divide UEs into three classes denoted by A, B and C corresponding to strong, normal, and weak UEs depending on their respective channel gains. Note that, we can use a pre-defined threshold to classify the UEs into three classes, i.e., UEs greater than a certain threshold fall into the same class. UEs in one class are similar to each other in terms of channel gain and dissimilar to the UEs belonging to other classes. The similarity between UEs is based on a measure of the channel gains between themselves and the MBS. Secondly, we determine the number of clusters required based on the output of the previous step, i.e., classification of UEs. Here, we calculate the cardinality of each class and choose the number of required NOMA clusters $\mathcal{S}$ equal to the maximum cardinality among all classes.

Once we obtain the classes and number of clusters in the network, we aim to group UEs into these clusters, i.e., subproblem of user clustering. To solve this subproblem, we assume that the number of RBs[2] allocated to each cluster and power level assigned to each UE are given, then, the problem of user clustering is as follows

$$\begin{aligned}
\textbf{UC}:\ &\max_{\boldsymbol{\beta}} \sum_{n \in \mathcal{N}} \sum_{j_n \in \mathcal{U}_n} \omega_{j_n} \sum_{k \in \mathcal{S}} \beta_{j_n}^k R_{j_n}^k \\
&\text{s.t.:}\ \ C_1: \sum_{n \in \mathcal{N}} \sum_{j_n \in \mathcal{U}_n} \sum_{k \in \mathcal{S}} \beta_{j_n}^k P_{j_n} \leq P_T, \\
&C_2: \sum_{j_n \in \mathcal{U}_n} \sum_{k \in \mathcal{S}} \beta_{j_n}^k R_{j_n}^k \geq R_n^{\min},\ \forall n \\
&C_3: \sum_{k \in \mathcal{S}} \beta_{j_n}^k = 1,\ \forall j_n \in \mathcal{U}_n, n \in \mathcal{N}, \\
&C_4: 2 \leq \sum_{n \in \mathcal{N}} \sum_{j_n \in \mathcal{U}_n} \beta_{j_n}^k \leq |\mathcal{U}|,\ \forall k, \\
&C_5: \sum_{k \in \mathcal{S}} \beta_{j_n}^k \delta^k \leq \Omega,\ \forall j, n, \\
&C_6: \beta_{j_n}^k \in \{0, 1\},\ \forall k, j, n.
\end{aligned} \tag{6}$$

[1] Any number of users' classification can be defined. In simulation results, we have also considered two network classes, i.e., strong and weak.

[2] Initially, for the first iteration we can assume equal RBs for all clusters.

The user clustering problem in (6) is still a combinatorial problem, and finding the solution becomes NP-hard for a large set of UEs and clusters in a practical amount of time [19]. Note that problem (6) is desired to be solved in a distributed manner. Therefore, we use matching theory to map the problem (6) into a matching game and then discuss the details of the solution in the following subsections.

For solving the user clustering problem in (6) via matching game, we do not take into account the ${C_1}^3$ constraint. Similarly, we relax C_4 by also allowing a single user in a cluster. Note that, we can consider such a cluster as an OMA cluster. This case generally occurs when the network consists of more UEs that have almost indistinguishable channel gains.

B. Matching Game with Externalities for User Clustering

Once user classification and cluster formation is executed, we get the required number of clusters and UE classes. The next goal here is to perform user clustering into clusters for the given number of clusters. Therefore, in this section, we apply matching theory for solving the problem of user clustering as this problem is a combinatorial problem. The motivation to apply matching theory for our problem comes from matching theory's ability to tackle combinatorial problems and achieve a distributed solution [17], [18], [19].

1) Matching Game Formulation: In our game there are two disjoint sets of agents, the set of clusters, $\mathcal{S}$, and the set of MVNO UEs, $\mathcal{U}$. Each cluster k has a strict, transitive, and complete preference profile $\mathcal{P}_k$ defined over UEs. Note that in this game, from constraint C_3 in (6), it is given that each UE can be assigned to a single cluster. However, different UEs can exist in same cluster, i.e., property of NOMA. Therefore, the preference profile $\mathcal{P}_u$ of UEs is defined over the clusters, i.e., $\mathcal{S}$. Note that, other UEs j' operating in the same cluster, implicitly affect the preference ranking of UE j. Therefore, our design corresponds to the *one-to-many matching* given by the tuple $(\mathcal{U}, \mathcal{S}, \succ_{\mathcal{U}}, \succ_{\mathcal{S}})$. Here, $\succ_{\mathcal{U}} \triangleq \{\succ_j\}_{j \in \mathcal{U}}$ and $\succ_{\mathcal{S}} \triangleq \{\succ_k\}_{k \in \mathcal{S}}$ represent the set of the preference relations of the UEs and clusters, respectively. Formally, we define the matching as follows:

Definition 1. *A* matching β *is defined on the set* $\mathcal{U} \cup \mathcal{S}$ *which satisfies for all* $k \in \mathcal{S}$ *and* $j \in \mathcal{U}$*:*

1) $|\beta(j)| \leq 1$ *and* $\beta(j) \in \mathcal{S} \cup \phi$,
2) $|\beta(k)| \leq q_k$ *and* $\beta(k) \in 2^{\mathcal{U}} \cup \phi$,
3) *If* $j \in \beta(k)$ *then* $\beta(j) = k$,
4) *If* $\beta(j) \in k$ *for cluster* k *then* $\beta(k) = j$,

where q_k denotes the quota of cluster k, $|\beta(.)|$ denotes the cardinality of matching outcome $\beta(.)$. The first two conditions here represent constraints C_3 and C_4 in (6), respectively; where $q_k \leq |\mathcal{U}|$ represents the total quota of cluster k. Here, $\beta(j) = \phi$ means that j is not matched to any cluster. Similarly, if $\beta(k) = \phi$ then there are no UE matched to cluster k.

2) Preference Profiles of Players: In our formulated game, both sides need to rank each other using the preference profile. However, the preference profiles of UEs here depend on the resources offered to a cluster as well as other UEs assigned to that cluster. Such interdependence relations are known in matching theory as *externalities*, and have important implications in the design of the proposed solution. Due to these externalities, an agent may continuously change its preference order, in response to the formation of other agents and never reach a final assignment, unless externalities are well-handled.

In order to build the preference profile of UEs ($\mathcal{P}_j$), each UE calculates the achievable data rate for each cluster and then ranks them in descending order. Moreover, for enabling distributed matching game, the minimum data rate for each slice (constraint C_2) is relaxed to be minimum data rate for each UE, i.e., $R_j^{\min} = R_n^{\min} \frac{|h_{j_n}|^2}{\sum_{j_n \in \mathcal{U}_n} |h_{j_n}|^2}$, $\forall j \in \mathcal{U}_n$, minimum data rate demand is proportional to the UE's channel gain. Therefore, the utility of each UE can be defined as follows:

$$U_j(k, \beta) = \max(R_j^k, R_j^{\min}), \quad \forall k. \tag{7}$$

Thus, for any UE j_n, a preference relation $\succ_{j_n}$ is defined over the set of clusters $\mathcal{S}$ such that, for any two clusters $k, k' \in \mathcal{S}, k \neq k'$, and two matchings β and $\beta' \in \mathcal{U} \times \mathcal{S}, k = \beta(j), k' = \beta'(j)$

$$(k, \beta) \succ_j (k', \beta') \Leftrightarrow U_j(k, \beta) > U_j(k', \beta'). \tag{8}$$

Similarly, each cluster k creates its preference profile ($\mathcal{P}_k$) by using the following preference function:

$$U_k(j, \beta) = \max(R_j^s : D(R_j^{\min}, P_j) \geq \delta^k, 0), \quad \forall j, \tag{9}$$

where $D(R_j^{min}, P_j)$ represents the demand function in terms of the number of RBs for a UE at a given power and required minimum rate. According to (9), each cluster k wants to select the UEs whose minimum demand D can be fulfilled and maximize the achievable rate of a cluster. If the demand cannot be met, the UEs is not preferred and is ranked at the end which is represented by a zero preference, i.e., constraint C_5. Moreover, for any cluster k a preference relation $\succ_k$ is defined as follows, for any two UEs $j, j' \in \mathcal{U}$, where $j \neq j'$, and $j = \beta(k), j' = \beta'(k)$:

$$(j, \beta) \succ_s (j', \beta') \Leftrightarrow U_k(j, \beta) > U_k(j', \beta'). \tag{10}$$

Once the matching game and preference profile of both the agent sides have been defined, we now aim at finding a stable clustering scheme for the proposed game.

However, it is evident from (7), and (9) that our preferences are a function of the existing matching β and from (3), it is clear that UEs affect each other's performance through interference produced by high SINR UEs. Therefore, in the next subsection, we present a novel approach adopted to handle such externalities.

3) Preferences and Externalities: Next, we develop a novel approach to handle externalities in the proposed game and analyze its solution.

In the proposed game if UE j is assigned to a cluster k, it will interfere with other UEs using the same cluster k if its gain is higher than those UEs of clusters. Consequently, an agent may change its preference order with regards to a given cluster k in response to the action of other agents, i.e., UE j' which have been assigned to the same cluster k. This may lead to a case in which agents never reach a final clustering solution.

Therefore, for building the UE preference which can also handle the externalities, we propose that the initial network

[3] The constraint C_1 will be handled in power allocation section.

information (i.e., CSI of all MVNO's UEs) is broadcasted to the MVNO's UEs by the InPs after collecting it from each individual MVNO's UEs. Through this information, each UE can find the set of UEs that have a higher gain with MBS. Note that, here we only care about UEs that fall in the same class only. Thus, each UE would have a different set. We name this set as an externality set for UEs j that has a set of conflicting UEs and represent it by $\mathcal{C}_j$ as follows:

$$\mathcal{C}_j = \left\{ j' \in \mathcal{U} : \frac{|h_{j'}|^2}{z_{j'}} > \frac{|h_j|^2}{z_j}, j, j' \in \mathcal{A} \right\}, \tag{11}$$

where $\mathcal{A}$ represents UEs in the same class. From (11), we select the UEs that belong to the same class of user j and have a higher gain compared to user j. The main idea is restrict the UEs that belong to the same class to be grouped into same cluster.

4) Clustering Algorithm: In order to find a stable clustering scheme, we need to first define the blocking pair for our game. Note that, in the formulated game there is an additional challenge of externalities. Thus, traditional solution designed for one to many games based on Gale-Shapley do not apply to our game. Therefore, first, we design the blocking pair for the formulated game with externalities followed by a stable algorithm. The blocking pair for the formulated game is defined as follows:

Definition 2. *A matching β is said to be stable if there exists no blocking pair (j, k) such that, $j \succ_k \beta(k)$, $k \succ_j \beta(j)$, and $\beta(k) \notin \mathcal{C}_j$.*

Definition 2 is based on the following intuition. Whenever an user j prefers an cluster s over its assigned cluster $\beta(j)$ that does not contain a conflicting user (i.e., $\beta(k) \notin \mathcal{C}_j$), and cluster k is also willing to admit j (i.e., $k \succ_k \beta(k)$) by rejecting some accepted UEs in $\beta(k)$ which are ranked lower than j, then j and k can deviate from their assigned matching to form a blocking pair. A matching is stable only if there exists no blocking pair. Moreover, to achieve stability, a sufficient condition is that formation of any new agent pair does not undermine the stability of existing matched pairs. By employing such a condition, the preference profile of currently matched UEs on an cluster will remain unaltered even after this new pair formation. Stability in our solution ensures that after clustering, no matched pair (user-cluster) in the network would benefit from replacing their assigned RB with a new better cluster, and vice versa. This property is important to ensure the stable matching for one to many matching problems with externalities.

Next, we present a novel and stable user clustering algorithm. In this algorithm, the InP first decides the proposing order based on the set of available classes. The intuition behind this assumption comes from the fact that in NOMA we would like to restrict UEs from the same class to be in the same cluster. Thus, by allowing a sequential proposing manner in terms of classes will allow same class UEs to compete with each other. Therefore, in our algorithm, we assume proposal staring by the strongest class to the weakest class, i.e., staring from class A to class C. Note that by allowing this proposed order, we can guarantee that no matched user from a higher class can be affected by lower class UEs. The algorithm starts by using the local information to build the preference profiles (lines 1-3). At each iteration t, each user j that belong to a

Algorithm 1 User Clustering Algorithm

1: **input**: $\mathcal{P}_j^{(t)}$, $\mathcal{P}_k^{(t)}$, $\mathcal{C}_j$, $\forall k, j$.
2: **initialize**: $t = 0$, $\beta^{(1)} \triangleq \{\beta(j)^{(1)}, \beta(k)^{(1)}\}_{j \in \mathcal{U}, k \in \mathcal{S}} = \emptyset$, $\mathcal{J}_k{}^{(1)} = \emptyset$, $C_s^{(1)} = \emptyset$, $q_k^{(1)} = |\mathcal{U}|$, $\forall k, j$.
3: **repeat**
4: $t \leftarrow t + 1$.
5: Update $\forall j$, $\mathcal{P}_j{}^{(t)}$ for given $\beta(s)^{(t-1)}$.
6: $\forall j \in$ class $\mathcal{K}$ with k as its most preferred in $\mathcal{P}_j^{(t)}$.
7: **while** $j \notin \beta(k)^{(t)}$ and $\mathcal{P}_j^{(t)} \neq \emptyset$ **do**
8: **if** $C_k^{(t)} = \{j' \in \beta(k)^{(t)} \cup C_j\} \neq \emptyset$ **then**
9: $\mathcal{X}'^{(t)}_k = \{j' \in \beta(k)^{(t)}, k' \in \mathcal{C}_j | j \succ_k j'\}$.
10: $j_{lp} \leftarrow$ the least preferred $j' \in \mathcal{X}'^{(t)}_k$.
11: **for** $j_{lp} \in \mathcal{X}'^{(t)}_k$ **do**
12: $\beta(k)^{(t)} \leftarrow \beta(k)^{(t)} \setminus j_{lp}$, $q_k{}^{(t)} \leftarrow q_k{}^{(t)} + 1$.
13: **if** $C_k^{(t)} = \{j' \in \beta(k)^{(t)} \cup C_j\} \neq \emptyset$ **then**
14: $j_{lp} \leftarrow j$.
15: **else**
16: $\beta(k)^{(t)} \leftarrow \beta(s)^{(t)} \cup j$, $q_k{}^{(t)} \leftarrow q_k{}^{(t)} - 1$.
17: **else**
18: **if** Check $q_k{}^{(t)} > 0$ **then**
19: $\beta(k)^{(t)} \leftarrow \beta(k)^{(t)} \cup j$, $q_k{}^{(t)} \leftarrow q_k{}^{(t)} - 1$.
20: **else**
21: $j_{lp} \leftarrow j$.
22: $\mathcal{J}_k{}^{(t)} = \{j \in \mathcal{X}'_k{}^{(t)} | j_{lp} \succ_k j\} \cup \{k_{lp}\}$.
23: **for** $j \in \mathcal{J}_r{}^{(t)}$ **do**
24: $\mathcal{P}_j{}^{(t)} \leftarrow \mathcal{P}_j{}^{(t)} \setminus k$ $\mathcal{P}_k{}^{(t)} \leftarrow \mathcal{P}_k{}^{(t)} \setminus j$.
25: **Check**: $\beta^{(t-1)} = \beta^{(t)}$.
26: **until** $\forall$ classes, i.e., A, B, C.

specific class, first calculates its utility and re-ranks all the clusters based on the previous matching $\beta(k)^{(t-1)}$ (line 4).

Then, each user j proposes to the most preferred k (line 6). On receiving the proposals, each cluster k first investigates if there exists a conflicting user j' in $\beta(k)$ that can result in either of the two cases. The first case is there exists conflicting UEs, i.e., $\mathcal{C}_k$ is non-empty (line 8). In this case, k removes all lower ranked UEs j' compared to j from its current matching (lines 9-12) and rechecks the conflict set (line 13). If conflict set is still non-empty, j is also rejected along with other removed UEs and is considered as the least preferred user j_{lp} (line 14) otherwise it is accepted (lines 15-16). The second case is the conflict set is empty. In this case, the quota of cluster k (q_k [4]) is first checked, if enough quota exists to accommodate j, and then, user j is accepted by the cluster k otherwise user j is rejected by the cluster k and considered as least preferred (lines 17-21). Finally, the least preferred user, i.e., j_{lp} and all UEs ranked lower than j_{lp} are removed from $\mathcal{P}_k{}^{(t)}$, and similarly these UEs also remove k from their respective $\mathcal{P}_j{}^{(t)}$ (lines 22-24). Once all UEs of a class have either been accepted or rejected by all the clusters, the next class starts the proposal process. Note that here the matching for a specific class terminates when the results of two consecutive iterations t remains unchanged (line 25). *With this process, we guarantee*

[4]The quota q_k is set to three UEs per clusters as we assume three classes. The intuition comes from the fact that only the best user of each class is selected by a cluster k.

that any less preferred user will not be accepted by that cluster even if it has sufficient quota to do so, which is crucial for the matching stability of our design. This process is repeated until the matching converges.

Theorem 1. *Alg. 1 converges to a stable allocation.*

Proof: We prove this theorem by contradiction. Assume that Alg. 1 produces a matching β with a blocking pair (j,k) by Definition 2. Since $k \succ_j \beta(j)$, j must have proposed to k and has been rejected due to a more preferred conflicting user j' on cluster k (lines 13-14). Thus, in this case (j,k) cannot form a blocking pair as $j' \succ_k j$, a contradiction. Moreover, when j was rejected, then any lower ranked user j' was rejected either before j (lines 9-12), or was made unable to propose because k is removed from j' preference list (lines 22-24). Thus, any lower ranked j' cannot be matched by k, i.e., $j' \notin \beta(r)$, a contradiction. ■

V. Resource Allocation

Given the user clustering $\boldsymbol{\beta}$ (i.e., each NOMA cluster $\mathcal{S}_j$ is given), power level assignment $\mathbf{P}$, the problem of resource allocation, i.e., the number of RBs allocated to each cluster, is formulated as follows

$$\textbf{RBA}: \max_{\boldsymbol{\delta}} \sum_{k\in\mathcal{S}}\sum_{j_n\in\mathcal{S}_k}\omega_{j_n}\delta^k B\log_2\left(1+\frac{P_{j_n}|h_{j_n}|^2}{I^k_{j_n}+\delta^k B z_{j_n}}\right)$$

s.t.:

$$\begin{aligned}
&C_2: \sum_{j_n\in\mathcal{U}_n}\sum_{k\in\mathcal{S}}\beta^k_{j_n}\delta^k B\log_2\left(1+\frac{P_{j_n}|h_{j_n}|^2}{I^k_{j_n}+\delta^k B z_{j_n}}\right)\geq R^{\min}_n,\ \forall n\\
&C_5: \sum_{k\in\mathcal{S}}\delta^k\leq\Omega,\\
&C_6: \delta^k\in\{1,2,...,\Omega\},\ \forall k.
\end{aligned} \tag{12}$$

where $I^k_{j_n}=\sum_{i_m\in\mathcal{S}_k|\frac{|h_{i_m}|^2}{z_{i_m}}>\frac{|h_{j_n}|^2}{z_{j_n}}}P_{i_m}|h_{j_n}|^2$.

We see that problem (12) is an integer programming problem, as the parameters to be optimized, i.e., δ^k is an integer variable. To address the integer difficulties, we relax the discrete variables δ^k to be real numbers within the interval $[1,\Omega]$. Thus the relax problem is a concave problem of real variable δ^k since the objective and constraints C_2 are in form of $x\log(1+\frac{1}{x})$.

Proposition 1: The objective function of problem (12) is concave for positive δ^k. Thus, the optimal RBs allocated for each cluster k satisfies:

$$\begin{aligned}
\frac{\partial L(\delta^k,\lambda_{j_n})}{\partial\delta^k} &= \sum_{j_n\in\mathcal{S}_k}\omega_{j_n}\left(\mathcal{R}^k_{j_n}-\frac{\delta^{k*}\mathcal{H}^k_{j_n}}{\mathcal{I}^k_{j_n}\mathcal{Q}^k_{j_n}}\right)\\
&+\sum_{n\in\mathcal{M}}\lambda_n\sum_{j_n\in\mathcal{U}_n}\beta^k_{j_n}\left(\mathcal{R}^k_{j_n}-\frac{\delta^{k*}\mathcal{H}^k_{j_n}}{\mathcal{I}^k_{j_n}\mathcal{Q}^k_{j_n}}\right)=0,
\end{aligned} \tag{13}$$

where $L(\delta^k,\lambda_{j_n})$ is the Lagrangian and λ_n is the multiplier correspond to the isolation constraints C_2 and $\mathcal{I}^k_{j_n}=I^k_{j_n}+\delta^{k*}Bz_{j_n}$, $\mathcal{Q}^k_{j_n}=I^k_{j_n}+P_{j_n}|h_{j_n}|^2+\delta^{k*}Bz_{j_n}$, $\mathcal{H}^k_{j_n}=P_{j_n}|h_{j_n}|^2B^2z_{j_n}$, $\mathcal{R}^k_{j_n}=B\log_2\left(1+\frac{P_{j_n}|h_{j_n}|^2}{I^k_{j_n}+\delta^{k*}Bz_{j_n}}\right)$.

The optimal solution obtained from (13) can be mapped to the largest previous integer, i.e., $\overline{\delta}^*=\lfloor\delta^*\rfloor$, where $\lfloor x\rfloor$ denotes the largest integer no more than x.

VI. Power Assignment

Given the user clustering $\boldsymbol{\beta}$ and resource allocation $\boldsymbol{\delta}$, the problem of power assignment is formulated as follows:

$$\textbf{PA1}: \max_{\mathbf{P}}\sum_{k\in\mathcal{S}}\sum_{j_n\in\mathcal{S}_k}\omega_{j_n}\beta^k_{j_n}\delta^k B\log_2\left(1+\frac{P_{j_n}|h_{j_n}|^2}{I^k_{j_n}+\delta^k Bz_{j_n}}\right)$$

$$\begin{aligned}
\text{s.t.:}\ &C_1: \sum_{k\in\mathcal{S}}\sum_{j_n\in\mathcal{S}_k}P_{j_n}\leq P_T,\\
&C_2: \sum_{j_n\in\mathcal{U}_n}\sum_{k\in\mathcal{S}}\beta^k_{j_n}\delta^k B\log_2\left(1+\frac{P_{j_n}|h_{j_n}|^2}{I^k_{j_n}+\delta^k Bz_{j_n}}\right)\geq R^{\min}_n,\ \forall n.
\end{aligned} \tag{14}$$

The problem in (14) is not convex because the rate function in (2) is non-concave. To overcome such a major difficulty, we adopt the following successive convex approximation (SCA) approach [15], [16] and find the optimal power allocation.

A. Arithmetic-Geometric Mean Approximation and Centralized SCA-based Power Assignment

Define subset $\bar{\mathcal{S}}_k(j_n)\triangleq\left\{i_m\in\mathcal{S}_k|\frac{|h_{i_m}|^2}{z_{i_m}}>\frac{|h_{j_n}|^2}{z_{j_n}}\right\}$, $\tilde{\beta}^k_{j_n}\triangleq\beta^k_{j_n}\delta^k B$, and $\tilde{z}_{j_n}\triangleq z_{j_n}\delta^k B$, then the achievable throughput of UE $j_n\in\mathcal{S}_k$ can be rewritten as follows

$$R^k_{j_n}=\log_2\left(\frac{\sum_{i_m\in\bar{\mathcal{S}}_k(j_n)}P_{i_m}|h_{j_n}|^2+\tilde{z}_{j_n}+P_{j_n}|h_{j_n}|^2}{\sum_{i_m\in\bar{\mathcal{S}}_k(j_n)}P_{i_m}|h_{j_n}|^2+\tilde{z}_{j_n}}\right)^{\delta^k B}. \tag{15}$$

Define $u^k_{j_n}(\mathbf{P})=\sum_{i_m\in\bar{\mathcal{S}}_k(j_n)}P_{i_m}|h_{j_n}|^2+\tilde{z}_{j_n}+P_{j_n}|h_{j_n}|^2$, the arithmetic-geometric mean (AGM) inequality states that

$$\begin{aligned}
&u^k_{j_n}(\mathbf{P})\geq\underline{u}^k_{j_n}(\mathbf{P})\\
&=\prod_{i_m\in\bar{\mathcal{S}}_k(j_n)}\left(\frac{P_{i_m}|h_{j_n}|^2}{\kappa_{i_m}}\right)^{\kappa_{i_m}}\left(\frac{\tilde{z}_{j_n}}{\lambda_{j_n}}\right)^{\lambda_{j_n}}\left(\frac{P_{j_n}|h_{j_n}|^2}{\gamma_{j_n}}\right)^{\gamma_{j_n}},
\end{aligned} \tag{16}$$

where for all $j_n\in\mathcal{S}_k$, $i_m\in\bar{\mathcal{S}}_k(j_n)$, $\kappa_{i_m}=P_{i_m}|h_{j_n}|^2/u^k_{j_n}(\mathbf{P})$, $\lambda_{j_n}=\tilde{z}_{j_n}/u^k_{j_n}(\mathbf{P})$, and $\gamma_{j_n}=P_{j_n}|h_{j_n}|^2/u^k_{j_n}(\mathbf{P})$. The following approximate problem belongs to the class of geometric programs:

$$\textbf{PA2}: \min_{\mathbf{P}}\prod_{k\in\mathcal{S}}\prod_{j_n\in\mathcal{S}_k}\left(\frac{\sum_{i_m\in\bar{\mathcal{S}}_k(j_n)}P_{i_m}|h_{j_n}|^2+\tilde{z}_{j_n}}{\underline{u}^k_{j_n}(\mathbf{P})}\right)^{\tilde{\beta}^k_{j_n}\omega_{j_n}}$$

$$\begin{aligned}
\text{s.t.:}\ &C_1: \sum_{j_n\in\mathcal{U}}P_{j_n}\leq P_T,\\
&C_2: \prod_{j_n\in\mathcal{U}_n}\prod_{k\in\mathcal{S}}\left(\frac{\sum_{i_m\in\bar{\mathcal{S}}_k(j_n)}P_{i_m}|h_{j_n}|^2+\tilde{z}_{j_n}}{\underline{u}^k_{j_n}(\mathbf{P})}\right)^{\tilde{\beta}^k_{j_n}}\leq 2^{-R^{\min}_n},\ \forall n.
\end{aligned} \tag{17}$$

From the above discussion, we present in Alg. 2 a power allocation scheme based on the SCA approach with the AGM approximation to solve (14).

Algorithm 2 SCA-based Power Allocation with Arithmetic-Geometric Mean Approximation

Initialize: $t = 1$;
repeat
  Compute each coefficient

$$\kappa_{i_m}[t] = \frac{P_{i_m}[t-1]|h_{j_n}|^2}{u^k_{j_n}(\mathbf{P}[t-1])}, \; \lambda_{j_n}[t] = \frac{\tilde{z}_{j_n}}{u^k_{j_n}(\mathbf{P}[t-1])},$$

$$\gamma_{j_n}[t] = \frac{P_{j_n}[t-1]|h_{j_n}|^2}{u^k_{j_n}(\mathbf{P}[t-1])}.$$

  Compute monomial

$$\underline{u}^k_{j_n}(\mathbf{P})[t] = \prod_{i_m \in \bar{\mathcal{S}}_k(j_n)} \left(\frac{P_{i_m}[t-1]|h_{j_n}|^2}{\kappa_{i_m}[t]}\right)^{\kappa_{i_m}[t]} \times \left(\frac{\tilde{z}_{j_n}}{\lambda_{j_n}[t]}\right)^{\lambda_{j_n}[t]} \left(\frac{P_{j_n}[t-1]|h_{j_n}|^2}{\gamma_{j_n}[t]}\right)^{\gamma_{j_n}[t]}. \tag{18}$$

  With $\underline{u}^k_{j_n}(\mathbf{P})[t]$, solve geometric program (17), e.g., by an interior-point method, for an optimal power $\mathbf{P}[t]$.
  Set $t := t + 1$;
until $\mathbf{P}$ converge;

Algorithm 3 Joint User Clustering and Resource Allocation in NOMA (JUCRAN)

Step 1: Initialization
The MBS obtains CSI of all the UEs.
Classify UEs and calculate number of cluster
The MBS allocates the transmitted power equally to each UE.
The MBS allocates number of RBs equally to each cluster.
Step 2: Joint User Clustering and Resource Allocation
repeat
  Update the user clustering $\boldsymbol{\beta}$ using Alg. 1.
  Update RBs allocation using Lagrange dual method.
  Update power allocation $\mathbf{P}$ using Alg. 2.
until convergence;

B. Joint User Clustering and Resource Allocation in NOMA - JUCRAN

With the above user clustering and resource allocation algorithms, we now present the overall joint algorithms JUCRAN for our proposed problem as shown in Alg. 3.

Theorem 2. *The proposed algorithm JUCRAN converges to the suboptimal solution of the original problem in* (5).

VII. Simulation Results

A. Simulation Setting

In this section, we evaluate the performance of the proposed joint algorithm JUCRAN for user clustering and resource allocation in downlink NOMA WNV. In the simulations, we consider one base station located in the cell center and the MVNOs' UEs are uniformly distributed in a circular range with radius of 500 m. The system bandwidth is 20 MHz, bandwidth of a RB is $B = 180$ KHz, and number of available RBs is $\Omega = 100$. The maximum transmit power of the MBS is set to be 40 W in some simulations, however, we vary this parameter to investigate trade-off between system throughput and energy efficiency. The number of MVNOs is 3, and the number of UEs of each MVNO and its QoS isolation constraint are uniformly distributed, and the total number of all MVNOs' UEs are varied to investigate the performance of our proposed scheme. The noise power is assumed to be 10^{-13} W for all RBs. The small-scale fading coefficients of the MBS-to-user links are generated as i.i.d. Rayleigh random variables with unit variance. We compare the performance of our proposed scheme with the traditional OFDMA scheme. In the OFDMA scheme, UEs are assigned with an orthogonal number of RBs, i.e., no interference between UEs. Moreover, we also draw a comparison between our proposed scheme, i.e., JUCRAN with 3 classes of UEs (i.e., strong, normal and weak UEs) denoted JUCRAN-3, with the a similar scheme that allows only 2 classes of UEs (i.e., strong and weak UEs) denoted by JUCRAN-2. This benchmark scheme of two classes is similar to the proposal in [10]. Moreover, all the results are averaged over 100 simulation runs with random topology for each number of UEs.

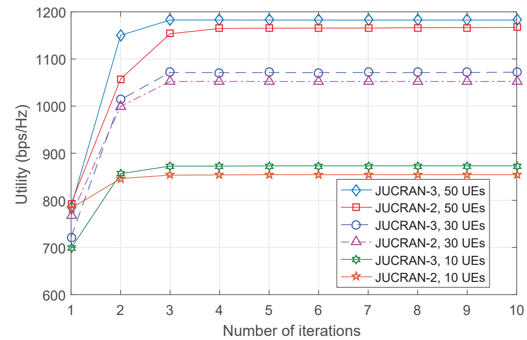

Fig. 2: Convergence of JUCRAN-2 and JUCRAN-3

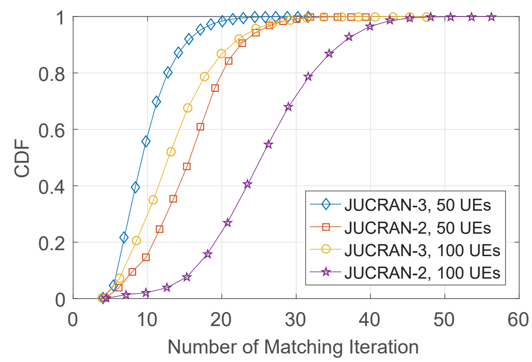

Fig. 3: Number of matching iteration

B. Numerical Results

In Fig. 2, we illustrate the convergence of our joint scheme, i.e., JUCRAN. We investigate the performance for three different instances, i.e., with 10, 30, and 50 UEs. From the results, it can be seen that our design JUCRAN can converge quickly with a limited number of iterations (i.e., less than 5 iterations) under all cases for both schemes JUCRAN-2 and JUCRAN-3. Moreover, JUCRAN-3 scheme has a relatively higher utility (i.e., the objective function of problem (5)) compared to the JUCRAN-2 scheme. The main reason is that using JUCRAN-3,

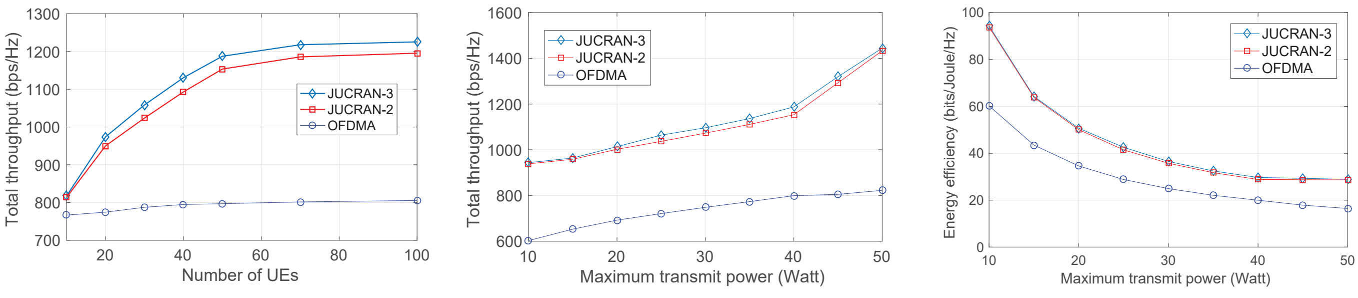

Fig. 4: (a) Throughput vs number of UEs, (b) Throughput vs maximum transmit power, (c) Energy efficiency vs maximum transmit power.

each cluster can accommodate more number of UEs compared to JUCRAN-2 scheme, thus, allowing to meet the minimum rate requirements of more UEs in the network.

In Fig. 3, we compare the number of iterations required for the matching game in JUCRAN-3 and JUCRAN-2 schemes with 50 UEs and 100 UEs. It can be seen that the average number of iterations of the matching game with JUCRAN-2 is higher than for JUCRAN-3, and the number of iterations increases with the increase in the number of network users. This can be explained by the number of users being higher in the network when more matching proposals are sent, thus increasing the convergence time. Moreover, in the JUCRAN-3 scheme, less iterations are required for convergence compared to JUCRAN-2. One of the main reasons for this is the quota of each cluster is set to two in JUCRAN-2, whereas in JUCRAN-3 it is set to three, thus accepting more users in its proposal phase.

In Fig. 4a, we illustrate the total network throughput vs. the number of UEs under three schemes, i.e., JUCRAN-2, JUCRAN3, and OFDMA. We evaluate the total throughput by obtaining the average total sum-rate over different numbers of UEs. It can be seen that the total throughput of JUCRAN-2 and JUCRAN-3 increase when the number of UEs increases until it saturates when the number of UEs in the network increases significantly. It can be observed that when the number of UEs are larger than 50, the total throughput continues to increase due to the multiuser diversity gain but grows at a slower speed and saturates when the number of UEs is sufficient large. Moreover, we can see that both JUCRAN-2 and JUCRAN-3 performance is almost comparable (i.e., JUCRAN-3 is slightly higher than JUCRAN-2), however, both schemes outperform the traditional OFDMA scheme significantly. Furthermore, through Fig. 4a it can be implied that the an InP cannot not take full advantage of resource allocation using traditional OFDMA since the number of RBs is limited and each user is only assigned an orthogonal dedicated number of RBs. This signifies the importance of NOMA in bringing WNV into fruition.

In Fig. 4b, we compare the impact of varying the total power in the InP over the network throughput with 50 UEs. It is evident that the total throughput increases as the MBS maximum budget power grows. Both scheme JUCRAN-2 and JUCRAN-3 outperform the OFDMA scheme significantly. While the total throughputs of JUCRAN-2 and JUCRAN-3 increase dramatically as the MBS transmit power budget increases, the throughput of OFDMA scheme increases gradually. This can be explained by the logarithms function of transmit power from the Shannon's formula of the OFDMA scheme.

Finally, Fig. 4c illustrates the total energy efficiency vs. MBS transmit power budget with 50 UEs. It is observed that the energy efficiency decreases when the MBS transmit power budget increases. This is because there is a tradeoff between transmission capacity and power consumption. From Fig. 4c, it is seen that both JUCRAN-2 and JUCRAN-3 can achieve better performance than OFDMA scheme in terms of energy efficiency. Moreover, both JUCRAN-2 and JUCRAN-3 schemes have indistinguishable performance.

VIII. Conclusion

In this paper, we proposed an efficient user clustering and resource allocation scheme for a downlink NOMA wireless network virtualization. In particular, we formulated an optimization problem with the objective to maximize the weight sum-rate while supporting the minimum reserved data rate per each slice to ensure effective isolation among UEs in the different slices. The proposed scheme includes user clustering, RB allocation and power allocation. To solve the proposed problem, we applied the one to many matching game with externalities to solve the user clustering subproblem and propose a stable matching algorithm. Next, for the power allocation, we applied complementary geometric programming and arithmetic-geometric mean inequality, since the objective function is non-convex and finally propose a computationally tractable iterative algorithm. Through extensive simulation, we observed that the use of NOMA scheme in WNV can significantly increase the performance in terms of network throughput and energy efficiency compared to the traditional OFDMA scheme. Moreover, we also observed that by classifying network UEs into three classes, it enhances the network performance compared to the general two user classification in terms of throughput performance and convergence time.

References

[1] C. Liang and F. R. Yu, "Wireless virtualization for next generation mobile cellular networks," *IEEE Wireless Comm.*, vol. 22, no. 1, pp. 61-69, Feb. 2015.

[2] C. Liang and F. R. Yu, "Wireless Network Virtualization: A Survey, Some Research Issues and Challenges," *IEEE Comm. Surveys & Tutorials*, vol. 17, no. 1, pp. 358-380, First quarter 2015.

[3] H. Wen, P. K. Tiwary, and T. Le-Ngoc, "Current trends and perspectives in wireless virtualization" *IEEE International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT)*, Montreal, Canada, Aug. 2013.

[4] K. Zhu and E. Hossain, "Virtualization of 5G Cellular Networks as a Hierarchical Combinatorial Auction," *IEEE Trans. Mobile Computing,* vol. 15, no. 10, pp. 2640-2654, Oct. 1 2016.

[5] M. I. Kamel, L. B. Le, and A. Girard, "LTE multi-cell dynamic resource allocation for wireless network virtualization" *IEEE Wireless Communications and Networking Conference (WCNC)*, New Orleans, LA, Mar. 2015.

[6] M. I. Kamel, L. B. Le, and A. Girard, "LTE Wireless Network Virtualization: Dynamic Slicing via Flexible Scheduling" *IEEE Vehicular Technology Conference (VTC Fall)*, Vancouver, Canada, Sept. 2014.

[7] G. Zhang , K. Yang, J. Wei, K. Xu, and P. Liu, "Virtual resource allocation for wireless virtualization networks using market equilibrium theory" *Computer Communications Workshops (INFOCOM WKSHPS), IEEE Conference,* Hong Kong, China, Apr. 2015.

[8] L. Song, Y. Li, Z. Ding and H. V. Poor, "Resource Management in Non-Orthogonal Multiple Access Networks for 5G and Beyond," *IEEE Network*, vol. 31, no. 4, pp. 8-14, July-August 2017.

[9] Y. Saito, A. Kishiyama, T. Benjebbour, A. Nakamura, Li, and H. K., "Non-orthogonal multiple access (NOMA) for cellular future radio access," *Proc. IEEE Veh. Tech. Conf. (VTC)*, Dresden, Germany, Jun 2013, pp. 1–5.

[10] M. S. Ali, H. Tabassum and E. Hossain, "Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems", *IEEE Access*, vol. 4, no. , pp. 6325-6343, 2016.

[11] B. Di, L. Song and Y. Li, "Sub-Channel Assignment, Power Allocation, and User Scheduling for Non-Orthogonal Multiple Access Networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7686-7698, Nov. 2016.

[12] R. Dawadi, S. Parsaeefard, M. Derakhshani and T. Le-Ngoc, "Power-Efficient Resource Allocation in NOMA Virtualized Wireless Networks,“ *IEEE Global Communications Conference (GLOBECOM)*, Washington, DC, Dec. 2016.

[13] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, "Joint Subchannel Assignment and Power Allocation for OFDMA Femtocell Networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 342-355, Jan. 2014.

[14] S. Boyd and L. Vandenberhe, *Convex Optimization*, Cambridge University Press, 2004.

[15] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for non-convex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[16] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, Jul 2007.

[17] A. Roth and M. Sotomayor, *"Two Sided Matching: A Study in Game Theoretic Modeling and Analysis"*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 1989.

[18] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, "Matching theory for future wireless networks: fundamentals and applications," *IEEE Comm. Magazine*, vol. 53, no. 5, pp. 52-59, May 2015.

[19] S. M. A. Kazmi, N. H. Tran, W. Saad, Z. Han, T. M. Ho, T. Z. Oo, C. S. Hong, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Trans. Mobile Comput.,* 2017.