

Towards the Mobility Issues of 5G-NOMA Through User Dissociation and Re-association Control

Muhammad Kamran Naeem¹, Raouf Abozariba², Md Asaduzzaman³ and Mohammad Patwary²

¹Research Innovation and Enterprise, Solent University, UK

²School of Computing and Digital Technologies, Birmingham City University, UK

³School of Creative Arts and Engineering, Staffordshire University, UK

Abstract—Non-Orthogonal Multiple Access (NOMA) system is considered as core enabler for the Fifth-Generation (5G) of the wireless cellular networks, contributing to the improvement of spectral efficiency. NOMA groups users into clusters, based on the maximum channel gain-difference. However, user mobility continuously changes the channel gain and requires re-clustering, depending on the percentage of mobile users and the environment in which they operate. In this paper, we propose a set of re-clustering methods: arbitrary, one-by-one and simultaneous, that expedite link re-establishment and keep the clusters interference-free, taking into account the mobility of users. The methods are applied to dissociate identified users within clusters, when the gain-difference is lower than a given threshold, followed by re-association procedure, which integrates users into different clusters, maintaining an appropriate gain-difference. The proposed methods are based on mathematical formulation and algorithms and address many technical and computational challenges associated with the clustering techniques. Numerical and experimental investigation has been carried out to test their performance and the results show that the simultaneous method can provide lower number of clusters, making it more suitable in dense and highly mobile scenarios. Our findings also demonstrate that this method has the potential to minimize the number of re-clustering, improving resource utilization and lowering signaling loads.

Index Terms—5G networks, NOMA, user mobility, user clustering.

I. INTRODUCTION

Background: The frequency bands currently used in cellular communications are becoming ever more saturated due to high traffic generated by the overwhelming number of smartphones, IoT nodes and similar devices. Such devices are used to run applications which are bandwidth hungry and have stringent delay constraints. The demand will intensify further by smart city based applications such as intelligent transportation systems, smart healthcare services, smart home and environment, and public safety applications [1, 2]. The availability and usage of millimeter-wave frequencies and techniques such as beam-forming and Filter Bank Multi-carrier Modulation (FBMC) will alleviate the problem, but these are not the only solutions being explored by researchers. Non-Orthogonal Multiple Access (NOMA), a promising technology, aims at improving the spectral efficiency by using successive interference cancellation (SIC) [3]. NOMA clusters several users with sufficient gain-difference between their channels and assign the same frequency band and time slot [4]. The clustering is based on their Channel State Information (CSI) and distance from the base stations (BS).

Related work: NOMA technology has been addressed from many different standpoints in the literature. In [5], the authors investigated how poor user pairing may increase the probability of NOMA achieving a lower sum capacity than conventional Orthogonal Multiple Access (OMA) schemes. The authors concluded that the performance gain of fixed power NOMA (F-NOMA) over conventional multiple access can be further enlarged by selecting users, whose channel conditions are more distinct. It is also mentioned in [6] and [7] that NOMA's performance is determined by how different the users' channel conditions are. However, even if the pairing is performed optimally, the inherent high mobility in cellular networks, can rapidly change the channel gain difference, raising the possibility of users being paired incorrectly. There are a number of studies which address this scenario. For example, in [8], the gain difference is artificially created by degrading the channel gain of one user and enhance it for the other concurrently, through precoding and power allocation. Complex power allocation strategies such as cognitive radio power allocation were also proposed in [5]. The majority of these studies concentrate on static setup in cellular networks, without investigating the impact of high user mobility and it is unclear whether these solutions are effective in dense mobile environment, such as in modern cities.

In this paper we propose a mathematical formulation and a solution algorithm, designed to dissociate one or more users when the channel gain-difference is low. The dissociated users (UEs) will encounter outages as a result of losing the channel resources associated with their original cluster. Since the dissociation mechanism creates vacant positions in some of the clusters, our methods enable re-assigning other dissociated users to those positions in real-time, minimizing outages in a robust manner. We present three different re-association methods: arbitrary, one-by-one and simultaneous, and provide detailed comparisons, highlighting the advantageous of the simultaneous technique.

The rest of the paper is organized as follows. Section II presents the system model. In Section III the problem is described mathematically followed by the resource allocation procedures in Section IV. The simulation results are presented in Section V and we conclude with a few remarks in Section VI.

II. SYSTEM MODEL

Consider a base station (BS) supporting $|\mathcal{V}|$ user equipments (UEs), randomly placed across the cell. The UEs are considered to be mobile or stationary and their locations may change with respect to the BS over time. The magnitude and angle of mobility is determined by Random Waypoint mobility model, widely used in the simulation studies of cellular networks [9]. We assume that there is a set of users who are mobile and they move between different cells, capturing the real-world dynamics in cellular networks.

In this model, we assume that the BS has access to the entire set of data on UEs, including two dimensional positions and the channel coefficients of each UE. Consider \mathcal{V}' is a set of all users under coverage and mathematically this can be defined by $\mathcal{V}' = \{v_1, v_2, \dots, v_i, \dots, v_{|\mathcal{V}'|}\}$. To group UEs into clusters, we use the distance based criteria (Euclidean distance), for which a distance measure is specified between each pair of UEs, in respect to the BS. The BS forms the clusters \mathbf{c}_i based on criteria relevant to the UEs distances $\mathcal{D} = \{d_1, d_2, \dots, d_i, \dots, d_{|\mathcal{V}'|}\}$ from each other.

In the initial clustering, UEs are first arranged in ascending order based on \mathcal{D} (described in more details in Section IV). The set of available resources at each BS is \mathcal{R} , with $|\mathcal{R}| = N_{RB}$, where each resource block (RB) represents the minimum spatio-temporal scheduling unit. The number of RBs available bounds the number of clusters in the network. Let l be the number of UEs in each group, where

$$l = \lceil |\mathcal{V}'| / k_{max} \rceil,$$

and k_{max} is the maximum number of users allowed in each cluster. The groups of UEs can be arranged in \mathbf{C} , which is defined as

$$\mathbf{C} = \begin{bmatrix} v_1 & v_2 & \dots & v_l \\ v_{l+1} & v_{l+2} & \dots & v_{2*l} \\ \vdots & \vdots & \vdots & \vdots \\ v_{l(k-1)+1} & v_{l(k-1)+2} & \dots & v_{k*l} \end{bmatrix}, \quad (1)$$

where each column of \mathbf{C} represents the UEs in a cluster. Consider \mathbf{C} is the initial set of clusters in the network which is defined as

$$\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_i, \dots, \mathbf{c}_k\},$$

where k represents the total number of clusters. Moreover, the sum of all users within the BS is denoted as

$$\mathbf{c}_1 \cup \mathbf{c}_2 \cup \dots \cup \mathbf{c}_k \equiv \{1, 2, \dots, V'\}$$

where

$$\mathbf{c}_p \cap \mathbf{c}_q \equiv \emptyset, \quad p \neq q.$$

The size of the clusters k_{max} are defined by the range of the BS coverage and throughput requirements. The number of available resource blocks $|\mathcal{R}|$ is also taken into consideration when deciding the cluster size. For example the following relation can be used

$$k_{max} = \frac{|\mathcal{V}'|}{|\mathcal{R}|}.$$

Upon detection of low channel gain-difference, the UE, which is likely to cause the highest disruption to the overall cluster, is dissociated from their clusters. The process of UE dissociation and re-association is presented in the following section.

III. PROBLEM SETUP

NOMA arranges the users into clusters, and to keep the clusters with high gain-difference and to enable users to perform SIC successfully, our method discards UEs which cause interference. This process is defined as dissociation. The method also provides the identity of vacant positions, which allow new and dissociated UEs to join new clusters. This joining process is define as re-association. The dissociation and re-association processes are simplified in Figure 1.

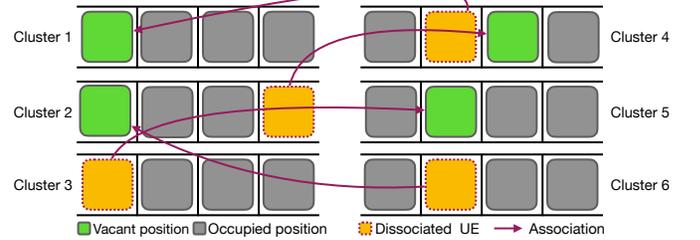


Fig. 1. An example of re-association/dissociation in shared clusters

A. Dissociation

In real world environments, wireless transmission suffers from various channel impairments, including path-loss, shadowing and fading. The channel of a wireless signal traversing a multiple N paths is usually represented as linear combinations of complex exponentials, given by [10]

$$h = \sum_n^N a_n e^{-j2\pi \frac{d_n}{\lambda} + j\phi_n} \quad (2)$$

where λ is the wavelength, a is the path attenuation, d is the distance the path traverses, and ϕ is a frequency-independent phase that captures whether the path is direct or reflected. Now, consider any two channel gain measurements h_x^c , h_y^c , for the channel between the BS and any two UEs x and y in the cluster $\mathbf{c} \in \mathbf{C}$, respectively. The UEs in cluster \mathbf{c} share the same channel on NOMA basis. As mentioned in Section I, the performance gain of NOMA increases in channel gain, i.e., when difference in path-loss between any set of UEs in one cluster is large. We define the dissociation procedure employed by the BS as

$$D(h_x^c, h_y^c) = \begin{cases} 1, & \text{if } h_x^c \text{ and } h_y^c \text{ satisfy a certain condition,} \\ 0, & \text{otherwise.} \end{cases}$$

Decisions to disassociate UEs from clusters is based on binary decision rule $D(h_x^c, h_y^c)$, which is founded on a chosen distance between h_x^c and h_y^c and determined by the following model

$$D(h_x^c, h_y^c) = \mathbf{1} \{ \|h_x^c - h_y^c\| < \lambda_0 \},$$

for each $x, y \in \mathbf{c}$ and $x \neq y$, where $\mathbf{1}$ denotes the indicator function that takes the value 1 if its argument is true and 0 otherwise. To maintain the cluster optimality in the system, we choose one of the UEs in the cluster to be dissociated using the rule

$$(\mathcal{P}_1) \quad (x, y)^* = \arg \max_{x, y \in \mathbf{c}} \{h_x^c, h_y^c\} \quad (3)$$

subject to $D(h_x^c, h_y^c) = 1,$

for a chosen λ_0^c such that $0 < \lambda_0^c < \lambda'$, where $\{\lambda_0^c\}_{c \in \mathbf{C}}$ are the thresholds relative to the cluster size and λ' relative to the cell size, describing the desired degree of channel gain-difference where large λ_0^c gives less channel gain-difference. It is important to note that the BS needs to observe the values of h_x^c and h_y^c to make an informed decision.

- If $D(h_x^c, h_y^c) = 1$, this implies that observations of channel gain difference from users x and y , in the same cluster are too close, therefore one of the UEs (x or y) is dissociated.
- If $D(h_x^c, h_y^c) = 0$, then we do not consider the difference in gain to be close enough to impact the SIC process.

Dissociating UEs can simply be achieved by issuing a disassociate request to the UE.

B. Re-association

Dissociated UEs can switch to operate under OMA scheme, requiring additional resources. This is attainable if there are additional channels reserved for such circumstances and may provide low latency. However, this is undesirable practice, when channels are a scarce commodity and we only consider this option when a suitable cluster cannot be found. Due to the dynamic nature of cellular networks, dissociated UEs may join one of the clusters with empty positions which are given by

$$\sum_{i=1}^{|\mathbf{C}|} \sum_{j=1}^{|\mathbf{c}|} D_{ij}(h_x^c, h_y^c) = m, \quad (4)$$

where $D_{ij}(\cdot)$ is the decision value $\{0, 1\}$ for j th user in the i th cluster, m represents the total number of vacant positions in the network. In NOMA, it is desired to maximize the distance between UEs in a cluster, for example, in a cluster of two UEs the optimal clustering is in

$$\arg \max_{\{x_i, x_j\} \in \mathbf{C}; i \neq j} D(x_i, x_j) \quad (5)$$

subject to $\|h_{x_i} - h_{x_j}\| > \lambda_0 \quad \forall i, j \in \mathbf{C}.$

Now we shall discuss the criteria of joining users into a cluster. Let $\mathbf{X} = \{x_{i1}, x_{i2}, \dots, x_{ini}\}$ be a set of n_i UEs dissociated from the i th cluster \mathbf{c}_i , $\mathbf{c}_i \in \mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$. The set of remaining UEs in \mathbf{C} forms a new cluster \mathbf{C}^* with $n = \sum_{i=1}^k n_i$ vacant positions such that $\mathbf{C}^* = \{\mathbf{C} \setminus \mathbf{X}\}$. Let y_{ij} be a binary indicator variable, which takes value 1 if the i th element of j th cluster $x_{ij} \in \mathbf{c}_i$ is assigned to the new cluster \mathbf{c}_i^* i.e., $x_{ij} \in \mathbf{C}$ is assigned to $x_{ij} \in \mathbf{C}^*$. Therefore, the new set of cluster can be represented by $\mathbf{C}^* = \{\mathbf{c}_1^*, \mathbf{c}_2^*, \dots, \mathbf{c}_k^*\}$. The optimization

problem of UE association with the appropriate constraints can now be written as follows

$$(\mathcal{P}_2) \quad \arg \max_{\{\mathbf{c}_1^*, \mathbf{c}_2^*, \dots, \mathbf{c}_k^*\}} \sum_{i=1}^k \sum_{j \in \mathbf{c}_i, \mathbf{c}_i^*} \|h_{x_i} - h_{x_j}\| \cdot y_{ij}$$

subject to $\|h_{x_i} - h_{x_j}\| > \lambda_0 \quad \forall i, j \in \mathbf{c}_i \text{ and } \mathbf{c}_i^*$

$\mathbf{c}_i^* \cap \mathbf{c}_j^* = \phi \quad \forall i, j \in \mathbf{C} \text{ and } \mathbf{C}^*,$

$x_i, x_j \in \{0, 1\} \quad \forall \mathbf{c}_i \in \mathbf{C},$

and $y_{ij} \in \{0, 1\},$ (6)

where h_{x_i} is the channel gain between the BS and the dissociated user x_i and h_{x_j} is the channel gain between the BS and an element of the \mathbf{c}^* . The re-association problem (\mathcal{P}_2) is a binary optimization problem and can be viewed as a clustering technique. The optimization problem can be solved using three different approaches: (i) associating users arbitrarily, which does not guarantee the optimality (ii) associating users one-by-one, which would give near optimal solution but does not guarantee the global optimal and (iii) associating users simultaneously, which guarantees the global optimal solution.

The arbitrary algorithm solves the problem through assigning each dissociated user to a suitable cluster under a certain gain-difference threshold. The one-by-one algorithm assigns dissociated users to the best cluster available from among all possible cases via an exhaustive search, to reduce the interference, under a certain gain-difference threshold. In the simultaneous algorithm we deploy the Hungarian algorithm [11, 12] (also known as the Kuhn-Munkres algorithm) maximizing $D(h_x^c, h_y^c)$. For example, Figure 1 shows four vacant positions and there are four UEs required to be re-associated. The arbitrary algorithm will find the suitable cluster for first (second) UE and assign it to the first suitable cluster and so on. The one-by-one algorithm will check the first (second) UE with all the clusters with vacant spaces and assign it to the cluster with maximum channel gain-difference. The simultaneous algorithm will find the suitable cluster for all four UEs against all available vacant spaces and assign them to the clusters simultaneously while fulfilling the gain-difference requirement.

IV. NOMA RESOURCE ALLOCATION THROUGH USER DISSOCIATION/RE-ASSOCIATION PROCEDURE

Overview of Algorithm 1 — The algorithm that performs the initial clustering is composed of the following steps: **Step 1:** Construct the set \mathcal{D} , which consists of the distances between each UE and the BS and sort all the UEs, based on their distance from BS (Line 2–7). **Step 2:** Construct k groups containing UE from \mathcal{V}' , where each group consists of l number of UEs. We assign the first l UEs to form group 1, then the next set of l UEs to form group 2 and so on. The first (second) user in each group will form the first (second) cluster and so on (Line 8–13).

This form of clustering is the de facto method in NOMA allocation and used by several researchers [13]. However,

Algorithm 1: Initial clustering described in pseudo-code.

```

1 Phase 1: Choosing initial clusters
2 Input: Number of UEs  $|\mathcal{V}'|$ , cluster size  $k_{\max} \triangleright$  e.g.,
   (  $1 \leq |K| \leq k_{\max}$  )
3 Output:  $C$ ;
4 for  $i \in \{1, 2, \dots, |\mathcal{V}'|\}$  do
5   calculate the distance between UE  $v_i$  and BS and
   save it in set  $\mathcal{D}$ 
6 sort  $\{\mathcal{D}\}$ 
7 sort  $\{\mathcal{V}'\}$  according to  $\{\mathcal{D}\}$ 
8 for  $j \in \{1, 2, \dots, k_{\max}\}$  do
9    $l = j \times k$ 
10  if  $l \leq |\mathcal{V}'|$  then
11     $\mathbf{c}_j = \{(k \times (j - 1)) + 1, (k \times (j - 1)) + 2, \dots, (k \times$ 
       $(j - 1)) + k_{\max}\}$  assigning UEs to the clusters
       $\{1, 2, \dots, k\}$ .
12    else
13     $\mathbf{c}_j = \{(k \times (j - 1)) + 1, (k \times (j - 1)) + 2, \dots, |\mathcal{V}'|\}$ 
      assigning UEs to the clusters
       $\{1, 2, \dots, |\mathcal{V}'| - (k \times (j - 1))\}$ .

```

this user clustering technique would inevitably be invalid in networks where users are mobile. Our next algorithm deals with this problem. The detailed pseudo-code is presented in Algorithm 1.

Overview of Algorithm 2 — In Algorithm 2 we applied a backtracking procedure to continuously monitor and update the clusters. The pseudo-code consists of the following steps: **Step 1:** Monitor the gain-difference between each pair of UEs within a cluster in real-time (Line 4–9). **Step 2:** If the gain-difference between any two UEs within a cluster falls below λ_1 but remain greater than λ_0 , then the UEs in that cluster are prompted to increase the frequency of obtaining and reporting channel gain measurements. This minimizes channel feedback and communications overhead, when user mobility in the network is low (Line 10–12). **Step 3:** If the gain-difference between UEs within a cluster is less than the given threshold λ_0 , the UE which is introducing interference to other UEs within the cluster is dissociated (Line 13–21). **Step 4:** The dissociated UEs from the set \mathcal{M} are then assigned to clusters in set \mathcal{N} , using arbitrary, one-by-one or simultaneous algorithms, following the corresponding techniques therein. The dissociated users are assigned to new suitable positions in other clusters, within multiple position candidates (Line 22).

It is reasonable for the UEs to be placed to the position where the gain-difference between the users in the new cluster is higher than λ_1 . The logic behind this algorithm is to keep track of the gain-difference changes between users which occurs as a consequence of mobility and to maintain the clusters in working order, minimizing outages. The detailed pseudo-code of the algorithm is presented in Algorithm 2.

Algorithm 2: Dissociation and re-association described in pseudo-code

```

1 Phase 2: Managing clusters
2 Input: Let  $\lambda_0$  and  $\lambda_1$  gain-difference threshold values,
    $\mathcal{N}$  is a set of available positions and  $\mathcal{M}$  is a set of
   dissociated and new UEs.
3 Output: Updated set of clusters  $C$ 
4 for  $k = 1 \rightarrow |C|$  do
5    $\mathbf{s} \leftarrow \mathbf{s}_{emp} \triangleright$  an empty array.
6    $\mathbf{W} \leftarrow ((1 : (|\mathbf{c}_k| - 1)) \times (1 : |\mathbf{c}_k|)) \triangleright$  is a matrix of
   zeros, where each  $v$  represents the status of UE
   interference.
7   for  $i = 1 \rightarrow (|\mathbf{c}_k| - 1)$  do
8     for  $j = i + 1 \rightarrow |\mathbf{c}_k|$  do
9        $d_{ij} = |(h_{x_i} - h_{y_j})| \triangleright d_{ij}$  is a function to
       find the gain-difference between all the
       UEs in a cluster.
10      if  $\lambda_0 \leq d_{ij} \leq \lambda_1$  (where  $\lambda_0 < \lambda_1$ ) then
11         $\mathbf{c}_k \leftarrow \mathbf{c}_k$ 
12        increase the rate of CSI measurements
13      else if  $d_{ij} \leq \lambda_0$  then
14         $\mathbf{s} \leftarrow [\mathbf{s}, v_{(i,j)}] \triangleright$  save index of
        interfering UEs in  $\mathbf{s}$ .
15         $\mathbf{W}(i, j) \leftarrow 1 \triangleright$  interference status of
        UEs becomes 1 when their
        gain-difference is less than  $\lambda_0$ .
16    for  $l = 1 \rightarrow (1 : |\mathbf{c}_k|)$  do
17       $\psi \leftarrow \mathbf{W}(:, l) \triangleright$  identify the UEs which are
      causing interference by analyzing each
      column of  $\mathbf{W}$ .
18     $\mathbf{c}_k \leftarrow (\mathbf{c}_k \setminus \psi) \triangleright$  disassociate UE with desired
    criteria
19    check  $\mathbf{s}$  if all the interfering UEs are dissociated
    from  $k^{th}$  cluster otherwise go to 16.
20     $\mathcal{M} \leftarrow \psi \triangleright$  set of all the dissociated UEs
21     $\mathcal{N} \leftarrow$  save cluster number  $k$ 
22 Solve the re-association problem ( $\mathcal{P}2$ ) using arbitrary,
    one-by-one or simultaneous optimization method
    from set  $\mathcal{M}$  into clusters in  $\mathcal{N}$ 

```

V. SIMULATION RESULTS

The cluster size has impact on the number of dissociation and re-association. The density of users in the network increases or decreases the number of dissociations and re-associations, which in turn, impact the latency perceived by the users. This latency is made up largely of the transition duration between dissociation and re-association of a user. The increase in search time for a new cluster will certainly impact the outages. In addition, the percentage of mobile users in the network will also influence the dissociation and re-association rate in the system. Our goals in this section is to explore the effectiveness of the proposed clustering algorithms in minimizing the number of clusters and the number of re-association by

implementing the algorithms and the optimizations in our simulation.

Experiment setup: In the simulation, we consider an area of 500×500 meters squared with a total of 800 users, randomly located with uniform distribution. The area is covered by one BS, located at the center, with coverage radius of 200 meters. For these experiments, we assume that 80% of users are mobile, using Random Waypoint mobility model. The mobile users can randomly move in or out of the BS coverage, but remain inside the defined area. This is to test our solutions in handling inter and intra-cell mobility.

Channel gain: There are many path-loss models available in the literature obtained through real-world measurements in urban, suburban and rural areas, based on probabilistic approaches, designed for different frequencies and environments [14–16]. To predict the channel gain between the BS and users, we used the ABG model for Urban Micro (UMi) under Non-Line-of-Sight (NLOS) that describes large-scale propagation path loss at sub-6GHz and millimeter-wave frequencies [17, 18].

$$\begin{aligned} \text{PL}[\text{dB}] = & 10\alpha \log_{10} \left(\frac{d_i}{1\text{m}} \right) + \beta \\ & + 10\gamma \log_{10} \left(\frac{f}{1\text{GHz}} \right) + \chi_{\sigma}^{\text{ABG}}, \end{aligned} \quad (1)$$

where α and γ are coefficients showing the dependence of path loss on distance and frequency, respectively, β is an optimized offset value for path loss in dB, f is the carrier frequency in GHz, and $\chi_{\sigma}^{\text{ABG}}$ is the shadow fading (SF) standard deviation, describing large-scale signal fluctuations about the mean path loss over distance. The following parameters are used in the simulation: $\alpha = 3.5$, $\beta = 24.4$ dB, $\gamma = 1.9$, $\sigma = 8$ dB and $f = 3.5$ GHz. We have built a simulation flexible enough to evaluate our solutions against network variations, enabling the configuration of cluster size, number of users in the cell, percentage of the mobile users and size of the cell.

A. Impact of mobility on the number of clusters

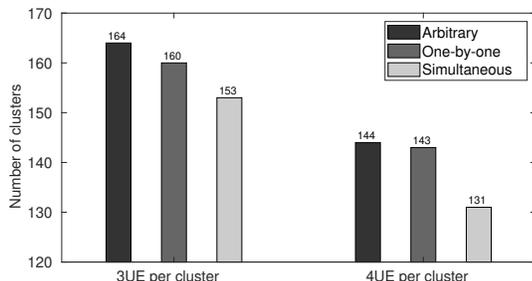


Fig. 2. Number of clusters with maximum cluster size of 3 and 4.

Due to the fact that the proposed solutions manage clusters differently, the number of clusters produced by each algorithm varies. The number of clusters is directly proportional to the number of required resource blocks. In Figure 2 we provide analysis for each method based on 1000 mobility instances, given two cluster sizes: 3 and 4. As can be seen in the

figure, with this metric, and with 443 average number of users under the coverage area, the number of clusters formed by the simultaneous algorithm is considerably lower. This can be explained by the fact that simultaneous algorithm maintains optimal gain-difference across all clusters simultaneously, while the other two algorithms are less optimal in this respect.

B. Impact of the proposed algorithms on clustering efficiency

Figure (3-left) compares the ratio of number of users per cluster of the three proposed algorithms, given a maximum cluster size of 3. The arbitrary algorithms is the simplest of the three algorithms, but results in a degraded performance, where more than 20% of the clusters are underutilized, compared to under 10% for the simultaneous. When the maximum number of users per cluster is set to 4, the performance of the arbitrary is even worse with more than half the clusters not reaching full capacity, as shown in Figure (3-right). On the other hand, the one-by-one and the simultaneous are comparable, with a slight advantage in favor of the latter. The simultaneous clustering algorithm represent the most efficient solution in handling mobility, achieving an improved clustering efficiency, over its counterparts. The clustering efficiency translates to increased number of served users, given a fixed set number of frequency blocks.

C. Impact of the proposed algorithms on the number of re-associations

In Figure 4, we analyze the number of re-associations, given two maximum cluster sizes: 3 and 4, by plotting the cumulative probability against the number of re-associations. The re-association rate affects different system performance metrics such as signaling load and user perceived quality of experience (QoE). When the cluster size is 3 (Figure 4-left) the simultaneous and the one-by-one algorithms are competitive and both methods yield better results than the arbitrary. By comparison, the simultaneous algorithm is confirmed as the best variant of the three algorithms when we set the maximum cluster size to 4. The low number of re-associations implies shorter transition duration time between dissociation and re-association of a user.

Implications: From our analysis we can conclude that the number of clusters can be lowered by using the simultaneous algorithm, especially in the case where the cluster sizes is greater than 3. Nevertheless, our analysis can also provide indications on selecting the appropriate maximum cluster size, given the size of the network and other parameters. Through these results, we have shown that these techniques can be used in NOMA clusters to manage mobility instead of complex power allocation techniques, such as cognitive radio power allocation, with high computational costs [19, 20].

The clusters arrangement vary according to the frequency band used for transmissions, since the channel to and from the user change with the frequency. In cellular networks, the uplink and the downlink operate on different frequencies (over 20 MHz apart). The clustering techniques presented in this paper is not restricted to downlink NOMA only, the same methods could also be applied to uplink NOMA transmissions, with minor changes to the algorithms.

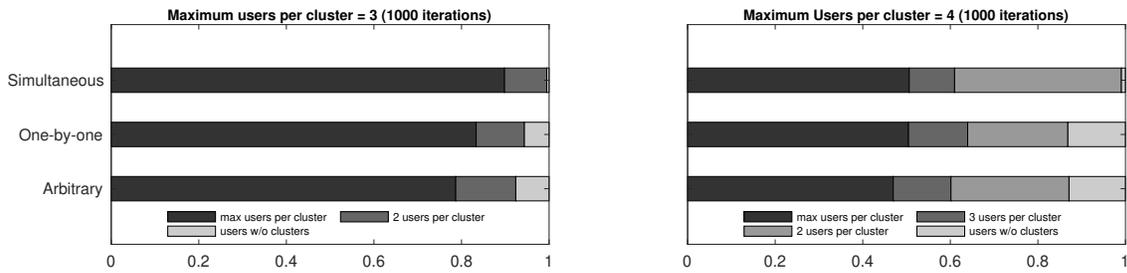


Fig. 3. Percentage of number of users in clusters using arbitrary, one-by-one and simultaneous algorithm for (left) Maximum users per cluster = 3 and (right) Maximum users per cluster = 4.

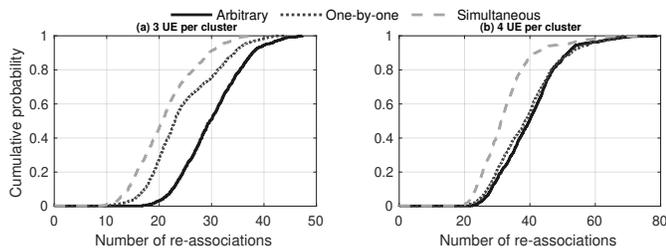


Fig. 4. Simulation results for comparing simultaneous algorithm with arbitrary and one-by-one in terms of number of re-associations. (left) 3 UEs per cluster and (right) 4 UEs per cluster. Simultaneous algorithm reduces the number of re-associations when cluster size is higher than 3.

VI. CONCLUSION

In this work, we highlighted the problem of mobility associated with NOMA clustering, in which the gain-difference between users may decrease to a level, where the successive interference cancellation (SIC) fails. This in turn diminishes the NOMA performance gain over its OMA counterpart. Here, we presented a new approach, fully-automatic, to manage and update users clusters and addressed this problem through user dissociation, followed by re-association procedure to link the dissociated users to new clusters. The dissociation is a combinatorial linear optimization problem in polynomial time, whereas the re-association, whenever this is at all possible, is a more complex NP -hard problem. We proposed three solutions in this domain: (i) arbitrarily (ii) one-by-one and (iii) simultaneous. To test this, we conducted experiments through simulations to compare the performances of each solution. We showed that the developed algorithms jointly able to maintain the clusters with the desired gain difference between users, while the rate of dissociation and re-association are minimized. We believe that this clustering solution is the first in this field. It is also generic and can be extended to cover several other kinds of networks underpinned by NOMA technology, where handling the mobility is essential, e.g., connectivity of autonomous and semi-autonomous vehicles.

REFERENCES

- [1] N.-S. Vo, T. Q. Duong, M. Guizani, and A. Kortun, "5G optimized caching and downlink resource sharing for smart cities," *IEEE Access*, vol. 6, pp. 31 457–31 468, 2018.
- [2] M. Erol-Kantarci and S. Sukhmani, "Caching and computing at the edge for mobile augmented reality and virtual reality (AR/VR) in 5G," in *Ad Hoc Networks*. Springer, 2018, pp. 169–177.
- [3] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Vehicular Technology Conference (VTC Spring)*, 77th. IEEE, 2013, pp. 1–5.
- [4] J. Choi, "On the power allocation for a practical multiuser superposition scheme in NOMA systems," *IEEE Communications Letters*, vol. 20, no. 3, pp. 438–441, 2016.
- [5] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G non-orthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, 2015.
- [6] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1462–1465, 2015.
- [7] B. Kimy, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *IEEE Military Communications Conference (MILCOM)*. IEEE, 2013, pp. 1278–1283.
- [8] Z. Ding, L. Dai, and H. V. Poor, "Mimo-noma design for small packet transmission in the internet of things," *IEEE access*, vol. 4, pp. 1393–1405, 2016.
- [9] E. Hytiä and J. Virtamo, "Random waypoint mobility model in cellular networks," *Wireless Networks*, vol. 13, no. 2, pp. 177–188, 2007.
- [10] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [11] R. Matushita and T. Tanaka, "Low-rank matrix reconstruction and clustering via approximate message passing," in *Advances in Neural Information Processing Systems*, 2013, pp. 917–925.
- [12] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.
- [13] A. Kiani and N. Ansari, "Edge computing aware NOMA for 5G networks," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1299–1306, 2018.
- [14] G. R. MacCartney and T. S. Rappaport, "Rural macrocell path loss models for millimeter wave wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1663–1677, 2017.
- [15] S. Y. Seidel and T. S. Rappaport, "914 MHz path loss prediction models for indoor wireless communications in multifloored buildings," *IEEE transactions on Antennas and Propagation*, vol. 40, no. 2, pp. 207–217, 1992.
- [16] J. Wu, S. Rangan, and H. Zhang, *Green communications: theoretical fundamentals, algorithms, and applications*. CRC press, 2016.
- [17] K. Haneda, J. Zhang, L. Tan, G. Liu, Y. Zheng, H. Asplund, J. Li, Y. Wang, D. Steer, C. Li *et al.*, "5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments," in *IEEE 83rd Vehicular Technology Conference (VTC Spring)*. IEEE, 2016, pp. 1–7.
- [18] S. Sun, T. S. Rappaport, S. Rangan, T. A. Thomas, A. Ghosh, I. Z. Kovacs, I. Rodriguez, O. Koymen, A. Partyka, and J. Jarvelainen, "Propagation path loss models for 5G urban micro-and macro-cellular scenarios," in *IEEE 83rd Vehicular Technology Conference (VTC Spring)*. IEEE, 2016, pp. 1–6.
- [19] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 721–742, 2016.
- [20] T. Takeda and K. Higuchi, "Enhanced user fairness using non-orthogonal access with SIC in cellular uplink," in *2011 IEEE vehicular technology conference (VTC Fall)*. IEEE, 2011, pp. 1–5.