Energy Efficiency for MEC Offloading with NOMA through Coalitional Games

Michalis Eliodorou∗†, Constantinos Psomas†, Ioannis Krikidis‡, and Socratis Socratous∗

∗Department of Electrical and Computer Engineering, University of Cyprus, Cyprus
†Department of Network Planning, MTN Cyprus, Cyprus
e-mail: {eliodorou.michalis, psomas, krikidis}@ucy.ac.cy, socratis.socratous@mtn.com

Abstract—In this paper, we investigate the user association problem for mobile edge computation (MEC) offloading in non-orthogonal multiple access (NOMA) networks. According to NOMA, multiple users can access the MEC server to offload data simultaneously. However, resources are shared among the users which can potentially impact the required transmit power for offloading, thus increasing the total energy consumption. Aiming to minimize the overall energy consumption for all the users of the network, we formulate a problem where user association, optimal power allocation, data rate and offloaded data are jointly considered. More specifically, two coalition game algorithms are proposed and compared in an effort to efficiently reduce the total energy consumption. Simulation results show that both proposed algorithms can successfully reach a final state with low complexity, where the overall energy consumption is significantly reduced.

Index Terms—Coalition games, user association, mobile edge computing, non-orthogonal multiple access.

I. INTRODUCTION

The growing demand of Internet-based services emerging from the development of Internet of Things (IoT) has led to an increased number of devices with mobile access. The number of devices keeps increasing year after year and is expected to reach 12.3 billion by 2022 [1]. Despite the fact that mobile devices have become more powerful and sophisticated, finite battery life and limited computation capacity pose significant challenges. Due to these limitations, Internet-services in fifth generation (5G) and beyond 5G networks cannot be executed solely in isolated devices. Mobile edge computing (MEC) is an emerging key technology derived from the advancements of cloud-computing which is capable of addressing the aforementioned challenges [2]. More specifically, MEC leverages data offloading for execution at a server with superior computation resources deployed at the edge of the network. Another promising technology is non-orthogonal multiple access (NOMA) which allows multiple users to be served simultaneously [3]. The joint implementation of MEC with NOMA gives the advantage that the same resources (i.e., frequency, time, code) are utilized by multiple devices, allowing them to access the MEC server at the same time and execute tasks remotely.

In networks where MEC technology is integrated with NOMA, the energy required by the users to execute a task can be decreased. However, since the resources of the network are being shared among the users, the uplink transmission and the latency play a critical role for minimizing the energy consumption. Many recent papers have considered the application of MEC jointly with NOMA, proposing ways to solve such optimization problems. The authors in [4], improve the energy performance of a multiuser NOMA system with MEC using the Lagrange dual method to obtain the global optimum solution. In [5], the authors investigate a MEC system with one mobile user adopting NOMA to offload data to multiple base stations (BSs), aiming to minimize the total power consumption. The formulated problem is non-convex and is solved by decomposing it into two subproblems. Another optimization problem, aiming to minimize the maximum overall delay is investigated in [6]. In this work, the authors introduce the relation between resource allocation and uploading delay and propose an algorithm with greedy subcarrier assignment to reduce the overall delay of all users.

User association is a mechanism with significant importance that can, among other things, increase the benefits of NOMA [7], especially in ultra-dense heterogeneous networks using MEC. Game theoretical approaches have recently gained great interest for resource allocation problems due to their ability to reach a final state which is beneficial for all players. The authors in [8], propose a low complexity algorithm based on a Stackelberg game to maximize the sum rate of a NOMA system. It is shown that the proposed scheme is capable of achieving significant performance gain. In [9], a coalition game is considered for optimizing the bandwidth allocation, shared between the fronthaul and the backhaul links, in a mmWave environment. In this scenario, the angle of each antenna was optimized to maximize the directivity gain. The results show that when coalition is adopted the sum-rate is effectively maximized.

Motivated by the above, in this paper, we minimize the total power consumption for MEC offloading but, in contrast to [5], we formulate the problem for a cellular uplink network with multiple users. In addition, we consider the user-subcarrier association problem, introduced in [6], while also considering the optimization problem for both the rate and the transmit power of each user investigated in [4]. In order to increase the number of users being served we also exploit NOMA. The problem is formulated and solved via game theoretic tools. Specifically, we propose two coalition game algorithms with different initial allocation schemes and different user selection processes to minimize the total energy consumption. The two algorithms are compared showing the benefits of the game-theoretic approach for each case. We show that both algorithms can successfully reduce the energy consumption.
for both partial offloading and the conventional full offloading. The proposed algorithms exploit cooperation via game theory and are of great importance for future networks, as they are of low-complexity and can achieve significant performance gains.

The rest of the paper is organised as follows. In Section II, the system model is described and the problem formulation is introduced. In Section III, the two coalition game algorithms are presented. The simulated results are shown in Section IV. Finally, in Section V conclusions are discussed.

II. SYSTEM MODEL

A. Network Model

We consider an uplink cellular network and we focus on a small base station (SBS) with \( N \) number of subcarriers located at the center of a circular area with radius \( R_D \) and \( K \) users which are randomly located, with \( K \geq N \). We denote by \( N = \{1, 2, \ldots, N\} \), the set of subcarriers and \( K = \{1, 2, \ldots, K\} \), the set of users. Each user must execute a task with \( L \) input bits and is able to offload all or part of the data by utilizing the MEC scheme. In addition, the users of each subcarrier can upload the offloaded data simultaneously using NOMA. Each subcarrier can receive data from up to \( N_{RF} \) users at the same time, where \( N_{RF} \) is the number of available radio frequency (RF) chains. The described system model is shown in Fig. 1.

B. Channel Model

The set of users associated with the \( j \)-th subcarrier is indicated with \( K_j \) and the cardinality of the set is denoted by \( K_j \), where \( K_j \leq N_{RF} \) and \( \sum_{j=1}^{N} K_j = K \). The channel vector of a subcarrier \( j \in N \) serving \( K_j \) users is

\[
h_j = [h_{1,j}, h_{2,j}, \ldots, h_{K_j,j}],
\]

where \( h_{k,j} \) is the channel gain of user \( k \in K_j \) associated with subcarrier \( j \). All channel coefficients are modeled as block Rayleigh fading with unit variance, i.e. \( h_{k,j} \sim \mathcal{CN}(0, 1) \), and the SBS is assumed to have full channel state information (CSI) for all \( N \) subcarriers. The path-loss model is considered to be \( d_{k,j}^{-\alpha} \), where \( d_{k,j} \) is the distance between user \( k \) and SBS \( j \) with \( \alpha > 2 \) being the path-loss exponent. All links exhibit additive white Gaussian noise with variance \( \sigma^2 \).

C. Mobile Edge Computing

We consider a MEC server integrated at the SBS to help users execute their computational tasks within a time slot of duration \( T \). We denote by \( L \) the overall bits of the task. The process of remote execution taking place during the distinct period \( T \) is partitioned in three phases: \( T_{UL} \), where the user uploads the data to the SBS, \( T_{EX} \), where the processing of the data at the MEC server takes place and \( T_{DL} \) for the downlink transmission of the final result back to the user. In this paper, partial offloading is implemented, meaning that the user can transmit part of the task’s bits, \( (L - l) \), while the remaining \( l \) bits are executed locally. The time block is shown in Fig. 2 and based on the above can be written as

\[
T = T_{UL} + T_{EX} + T_{DL}.
\]

Considering the advanced resources of the SBS and especially the MEC server’s high processing capacity, we assume that the time for the execution stage \( T_{EX} \) is negligible. In addition, since the data of the result, produced from the remote execution, are considerably small compared to the \( L \) number of bits of the task, we assume that the time needed for the final stage has minor impact to the entirety of the time. Based on the above, we assume that \( T_{EX} \approx 0 \) and \( T_{DL} \approx 0 \) [4], and we focus on the offloading process, which is the most time consuming part of the problem.

D. Non-orthogonal Multiple Access

NOMA is a multiple access scheme which can be used for allowing more than one user to offload data simultaneously. In this paper, NOMA is applied at each subcarrier, hence \( K_n \) users can be assigned to the \( n \)-th subcarrier, \( n \in N \), utilizing the same resources simultaneously. For each set of users associated with a subcarrier, the users are ordered based on their channel conditions. For ease of use, the index is ordered in an ascending order, hence \( |h_1| \leq \cdots \leq |h_{K_n}| \), where \( k \in K_n \). The SBS applies successive interference cancellation (SIC) to decode the signal of each user. More specifically, the receiver decodes the information of the users in \( K_n \) stages. In the first stage, it decodes the data of the user with the highest index, \( K_n \), treating the signal from all other users as interference. Once the receiver decodes the data of user \( K_n \), it can reconstruct user \( K_n \)’s signal and subtract it from the aggregate received signal. The receiver can then decode the data of user \( K_n - 1 \), with interference from all the users with...
a lower index. The same process is repeated until the last user, i.e., user 1, which has only Gaussian noise [10]. Using a SIC receiver at the SBS, the signal-to-interference-plus-noise ratio (SINR) of user \( k \) associated with the \( n \)-th subcarrier is described as

\[
\text{SINR}_{k,n} = \frac{p_k|h_{k,n}|^2d_k^{-\alpha}}{\sum_{i=1}^{k-1}p_i|h_{i,n}|^2d_i^{-\alpha} + \sigma^2},
\]

and the rate (bits/sec/Hz) of each user can be expressed as

\[
\dot{r}_{k,n} = \log_2(1 + \text{SINR}_{k,n}).
\]

The capacity region \( C(p) \) of the uplink channel is characterized by the set of all rates \( \{r_1, \ldots, r_{K_n}\} \) [10], satisfying the conditions of the polymatroid, i.e.,

\[
C(p) = \left\{ \mathbf{r} \in \mathbb{R}^{K \times 1} : \sum_{k \in \mathcal{J}} r_k \leq \log_2(1 + \sum_{k \in \mathcal{J}} p_k|h_k|^2), \forall \mathcal{J} \subseteq K \right\}.
\]

where \( p \) is the power allocation vector \( \{p_1, \ldots, p_{K_n}\} \). The selection of \( p \) must consider the constraints set by (5).

E. Energy minimization

In this paper, the user association along with the partial offloading and the power allocation are jointly considered and the problem is formulated as a coalitional game, aiming to minimize the overall energy consumption of all \( K \) users of the network. The total energy consumption of a user depends on the transmit power needed for data offloading and the local computation of the remaining data which takes place at the device. The relation between the rate of user \( k \), shown in (4), and the number of bits that can be offloaded is

\[
r_k = \frac{L - l_k}{BT},
\]

where \( l_k \) is the remaining bits of the task which are executed locally, \( B \) is the available bandwidth and \( T \) is the time of the time block. The necessary energy to transmit the offloaded data is

\[
E_{k}^{\text{tx}} = p_l T.
\]

The local energy consumption is determined by the computation capability of the user’s device, which is related to the central processor unit (CPU) frequency \( f_k \) and indicates the number of CPU cycles/second. The energy consumption of the local computation is represented as

\[
E_{k}^{\text{loc}} = \zeta_k f_k^3 T = \frac{\zeta_k C_k^3 f_k^3}{T^2},
\]

where frequency \( f_k \) can be rewritten as \( C_k l_k/T \), where \( C_k \) is the number of instructions cycles per bit, and \( \zeta_k \) is the coefficient of the processor architecture that indicates the CPU capabilities of mobile user \( k \). The total energy consumption of the \( k \)-th user is expressed as

\[
E_{k}^{\text{tot}} = E_{k}^{\text{tx}} + E_{k}^{\text{loc}}.
\]

The association between user and subcarrier is critical since it affects the achievable data rate as shown in (5). Therefore, based on the above, the user association problem is formulated jointly with the transmit power allocation \( p \), the achievable data rate \( r \), and the offloaded data as

\[
\min_{r, p, \{x_1, x_2, \ldots, x_N\}} \sum_{n=1}^{N} x_{k,n}(E_{k}^{\text{loc}} + E_{k}^{\text{tx}}),
\]

subject to

\[
x_{k,n} \in \{0,1\}, \forall k \in \mathcal{K}, \forall n \in \mathcal{N},
\]

\[
\sum_{n=1}^{N} x_{k,n} = 1, \forall n \in \mathcal{N},
\]

\[
K \sum_{i=1}^{K} x_{i,n} \leq N_{RF}, \forall n \in \mathcal{N},
\]

\[
r \in C(p) \text{ with (6)},
\]

\[
0 \leq p_k \leq P_{\text{max}}, \forall k \in \mathcal{K},
\]

\[
0 \leq l_k \leq L,
\]

where \( x_n = \{x_{k,n}\} \), defines the set of users associated with the \( n \)-th subcarrier and \( x_{k,n} \) is a binary value denoting whether or not the \( k \)-th user is associated with the \( n \)-th subcarrier. Constraint (10c) ensures that each user is associated with only one subcarrier. Constraint (10d) guarantees that the number of users associated with a subcarrier does not exceed the number of available RF chains \( N_{RF} \). Constraint (10e) ensures that the uplink rates lie within the capacity region \( C(p) \) given by (5). Finally, constraints (10f) and (10g) ensure the values for \( p_k \) and \( l_k \) are non-negative and within a permitted maximum value which are denoted as \( P_{\text{max}} \) and \( L \), respectively. In case where (10e) cannot be satisfied, we assume that the entire task is executed locally.

The formulated problem is non-convex and difficult to transform into a convex problem. However, as described in [4], the sub-problem using only constraints (10e), (10f) and (10g) is convex and the optimal values for \( r^*, p^* \) and \( l^* \) can be obtained. The sub-problem can be solved with numerical tools such as CVX [11] or Gurobi [12]. Treating the formulated problem as a game we can jointly solve the user association problem of the network and the optimization problem of each subcarrier for the power allocation and the set of rates. The coalition algorithms, proposed in this paper, show that the general formulated problem can be solved as shown in Section III.

III. COALITIONAL GAME BASED ALGORITHMS

In this section, two algorithms based on coalition games are presented. The proposed algorithms provide a user association solution by exploiting the cooperation among the subcarriers where the CSI for all users are known, thus minimizing the total energy consumption. Game theory has been proven to be very efficient in multi-player scenarios [13]. The formulated problem described in Section II-E is defined as a game \((\mathcal{K}, \mathcal{X}, \mathcal{U})\) where \( \mathcal{K} \) is the player set consisting of the users, \( \mathcal{X} = \{x_1, x_2, \ldots, x_N\} \) are the sets of associated users for each subcarrier and \( \mathcal{U} \) is a non-transferable utility [13]. A partition of the users, among the available subcarriers, is denoted by \( S = \{S_1, S_2, \ldots, S_N\} \), where \( S_n \) is the coalition consisting
of the users associated with the \( n \)-th subcarrier. For each coalition \( S_n \in \mathcal{S} \), \( n \in \mathcal{N} \), the conditions \( S_n \cap S_p = \emptyset \), \( \forall n \neq q \) and \( \bigcup_{n=1}^{N} S_n = \mathcal{K} \) are satisfied. For both algorithms, the utility function of a coalition \( m \) is given by

\[
U(S_m) = \left\{ \sum_{k=0}^{K_m} v_k, \forall k \in S_m \right\}, \quad (11)
\]

where \( v_k \) is the payoff value of user \( k \) given the partition \( S \), which indicates that the payoff value of player \( k \) is a decreasing function of the energy consumption \( E_k^{\text{tot}} \) shown in (9). Before we present the proposed algorithms, the following three definitions are introduced.

**Definition 1:** (Preference) For any user \( k \in \mathcal{K} \), we use the symbol \( \succ_k \) to denote its preference between two coalitions \( S_m \) and \( S_{m'} \), \( m \neq m' \). The decision of a user \( k \) depends on whether or not the utility values of the two coalitions will increase i.e.,

\[
S_m \succ_k S_{m'} \Leftrightarrow U(S_m \setminus \{k\}) + U(S_{m'} \cup \{k\}) > U(S_m) + U(S_{m'}). \quad (12)
\]

**Definition 2:** (Split and merge operation) Given two different partitions \( S \) and \( S' \), where \( S' \) occurs from partition \( S \) if user \( k \) moves from coalition \( S_m \in S \) and joins \( S_{m'} \in S' \). User \( k \in \mathcal{K} \), decides to leave its current coalition if and only if its preference condition (Definition 1) is satisfied. The split and merge operation can be written as

\[
\{S_m, S_{m'}\} \rightarrow \{S_m \setminus \{k\}, S_{m'} \cup \{k\}\}. \quad (13)
\]

Note that for the above operation, the user \( k \) joins the other partition if \( |S_m'| < N_{RF} \). Otherwise, a user \( k' \) in coalition \( S_{m'} \) is selected at random and swapped with \( k \) based on the following definition.

**Definition 3:** (Swap operation) Two users are said to be swapped, if and only if, the preference condition (Definition 1) is satisfied for both of them. Then, the partitions are updated accordingly as

\[
\{S_m, S_{m'}\} \rightarrow \{S_m \setminus \{k\} \cup \{k'\}, S_{m'} \setminus \{k'\} \cup \{k\}\}. \quad (14)
\]

### A. Coalition game algorithm with random swapping

In the first coalition game algorithm all users are initially allocated randomly to the available subcarriers. At each iteration, a user associated with subcarrier, say \( n \), is again randomly selected. By selecting a different subcarrier \( n' \), \( n \neq n' \), thus selecting another coalition, we check if the above definitions are satisfied. In the case where this is true, (13) or (14) are applied accordingly. The pseudocode of the proposed algorithm is provided in Algorithm 1. In what follows, proof is provided that Algorithm 1 converges.

**Convergence:** Starting at any initial combination, the user association game of Algorithm 1 is guaranteed to converge at a final state.

**Proof.** In order to increase the game utility \( U \), the users perform either the *split and merge* or the *swap* operation, thus constantly modifying the partition set. Consider two successive iterations \( i \) and \( i + 1 \), and assume that partition \( S_{i+1} \) was formed from \( S_i \), after an operation is applied. Both operations, take place if and only if the game utility \( U \) is strictly increased. This can be written as

\[
S_i \rightarrow S_{i+1} \Leftrightarrow U(S_i) < U(S_{i+1}). \quad (15)
\]

Therefore, the game utility value is always increasing, that is, \( S_{ini} \rightarrow S_1 \rightarrow S_2 \rightarrow \cdots \rightarrow S_{fin} \). (16)

where \( S_{ini} \) and \( S_{fin} \) are the initial and final partition sets of the game, respectively. Hence, the overall total energy consumption of all \( \mathcal{K} \) users is guaranteed to decrease at each new partition set. Since the number of users is finite, the number of partition sets is also finite and is based on the Bell number [14]. Therefore, the sequence in (16) is guaranteed to converge to the final state \( S_{fin} \).

We can show that Algorithm 1 is \( D_p \) stable. A partition \( S \) is \( D_p \) stable, if for any partition set \( S' \) that occurs from \( S \), when a user moves or a pair of users is swapped, \( U(S) \geq U(S') \).

\( D_p \) stability: The final partition set \( S_{fin} \) is \( D_p \) stable.

**Proof.** Suppose the final partition \( S_{fin} \) of Algorithm 1 is not \( D_p \) stable. Then, there must exist a user \( k \in \mathcal{K} \) that prefers to leave its current coalition and join another. This will form a new partition \( S_{tmp} \), where \( S_{tmp} \succ_k S_{fin} \) which contradicts the fact that \( S_{fin} \) is the final partition. Therefore, the final partition of Algorithm 1 is \( D_p \) stable.

**Complexity:** Each iteration executes \( K \) number of computational operations, to calculate the energy consumption of each user assuming optimal values for power allocation, offloaded data and rate have been found. When Algorithm 1 is performed for \( L \) number of iterations, then the complexity of the algorithm is \( O(LK) \), which is much smaller compared to the complexity of the exhaustive search which is \( O(L^K) \) [9].

---

**Algorithm 1 Random Coalition Game Algorithm**

1. Initializing users with a random partition \( S_{ini} \)
2. Denote current partition \( S_c \leftarrow S_{ini} \)
3. repeat
4. Randomly select a user \( k \) of coalition \( S_n \in S_c \)
5. Randomly select a user \( k' \) of coalition \( S_n' \in S_c \)
6. if \( |S_n'| = N_{RF} \) then
7. Assume \( S_{tmp} \leftarrow \{ S_k \} \) with user \( k' \)
8. if \( S_{tmp} \succ_k S_c \) then
9. User \( k \) leaves \( S_n \) and joins \( S_n' \)
10. User \( k' \) leaves \( S_n' \) and joins \( S_n \)
11. Update current partition:
12. \( S_c \leftarrow \{ S_c \setminus \{S_n, S_n'\}\} \cup \{ S_n \setminus \{k\} \cup \{k'\}, S_n' \setminus \{k'\} \cup \{k\}\} \)
13. else
14. Assume \( S_{tmp} \leftarrow \) user \( k \) joins \( S_n \)
15. if \( S_{tmp} \succ_k S_c \) then
16. User \( k \) leaves \( S_n \) and joins \( S_n' \)
17. Update current partition:
18. \( S_c \leftarrow \{ S_c \setminus \{S_n, S_n'\}\} \cup \{ S_n \setminus \{k\}, S_n' \cup \{k\}\} \)
19. until
Algorithm 2 Sequential Game Algorithm

1: Initialization of $S_{ini}$:
   All CHs are associated with different subcarriers
   All CMs are randomly associated with the subcarriers
2: Denote current partition $S_c \leftarrow S_{ini}$
3: repeat
4:   For any CH $i$ of cluster $S_i$, $i \in \{1, 2, \ldots, N\}$, user $i$ makes a new proposal $\sigma_i$ to all CMs sequentially
5:   For any CM $k$, $k \in S_j, j \neq i$:
6:      if $|S_i| = N_{\text{RF}}$ then
7:         Select a CM $k'$ of $S_i \in S_c$ to investigate swap
8:         if utilities of $S_i$ and $S_j$ are increased then
9:            The proposal is accepted by the CMs $k$ and $k'$
       else
10:            The proposal is declined (no operation)
       end
11: else
12:    if utilities of $S_i$ and $S_j$ are increased then
13:       The proposal is accepted by the CM $k$
14:     else
15:        The proposal is declined (no operation)
16: until

B. Coalition game algorithm with sequential swapping

The second proposed algorithm has two distinguished characteristics. The first one is that the initial association is not random. Instead, for each subcarrier $n$, $n \in N$, one user is selected first based on the best channel conditions and distance. The selected dominant user is called coalition head (CH) and is guaranteed a significantly low value for the transmit power to offload data. All remaining users are called coalition members (CMs). The second difference is that at each iteration the investigated player is not selected randomly. In contrast, the CHs make proposals and invite the CMs sequentially to join their coalitions. The strategy of any CH in coalition $S_i$, towards all CMs of different clusters is expressed as

$$\sigma_i = \{S_i|k \in S_j, \forall j \neq i\},$$

and takes place sequentially for all coalitions $n \in N$. In each iteration, the corresponding answer of any CM is to either accept the proposal and join the coalition of the CH or reject it and remain at the current coalition. Similar to Algorithm 1, the response is evaluated based on the sum of the two coalition utility values. The response of a CM $k$ in the $j$-th coalition, $S_j$ to a CH in the $i$-th coalition, $S_i$, where $i \neq j$ is represented as

$$\sigma_{i,k} = \begin{cases} 
\text{Yes, if } U(S_i) + U(S_j) < U(S_i \setminus \{k\}) + U(S_j \cup \{k\}), \\
\text{No, if } U(S_i) + U(S_j) \geq U(S_i \setminus \{k\}) + U(S_j \cup \{k\}). 
\end{cases}$$

The second coalition game algorithm is shown in Algorithm 2. Similar to the previous algorithm, convergence and stability are satisfied, therefore a final partition will be reached within a limited number of iterations converging to a final solution. In this algorithm, only $K - N$ operations are executed in each iteration, since the association of $N$ selected users (CHs) is pre-determined according to the initial association scheme. Therefore, the complexity of Algorithm 2 is $O(L(K - N))$ for $L$ number of iterations which is lower than Algorithm 2.

IV. Numerical Results

In this section, numerical results are presented to demonstrate the performance gains from the proposed algorithms on the total energy of all users. The following parameters were used: $N = 4$, $N_{\text{RF}} = 4$, $K = 12$, $\sigma^2 = -90$ dBm and the path-loss exponents is set as $\alpha = 3$. The users are randomly distributed within a cell of radius $R_D = 60$ m hence the distance between a user and the MEC server can be anywhere between 0-60 m. The random location of the $K$ users is modelled for 100 cases and the average is provided in our results. The bandwidth for each subcarrier is considered to be 200 KHz and the time slot $T$ is 1 sec. The coefficient of the processor architecture is the same for all users with $\xi_k = 10^{-14}$ and $C_k = 10^3$. The overall number of bits of the task is $L = 200$ bits.

Fig. 3 shows the total energy consumption achieved by the proposed schemes, using partial offloading over the number of iterations. We also present the two algorithms when applied with full offloading. For the case of full offloading the formulated shown in (10a) is adjusted. Specifically, constraint (10f) for the value of $p_k$ does not include a maximum limit $P_{\text{max}}$ and (10g) is removed since the offloaded bits are set to $L$. Along with the algorithms applied for partial and full offloading, the total energy consumption with local computation is included by using (8) with $l_k = L$. As it can be observed, partial offloading provides significant improvement over full offloading and local computation. As the number of iterations increases, the user association changes reaching a final state where both algorithms decrease the total energy consumption. Algorithm 1 can provide a better final outcome but requires more iterations than Algorithm 2 which makes the sequential algorithm better for time-critical communication.
higher performance gain, reducing the energy consumption even further. The results verify that coalition games are ideal for such optimization problems and the two algorithms can provide solutions with low complexity, efficiently reducing the total energy consumption. A potential extension of this work could also take into account the execution time at the MEC server where a meaningful delay might be present when many users access the server. Future work could also examine the behaviour of such algorithms for different data load sizes to see the impact of the size towards the execution time.

ACKNOWLEDGMENT
This work has been partially supported by the H2020-MSCA-ITN-2016 framework under grant agreement number 722788 (SPOTLIGHT) and partially co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research Promotion Foundation, under the project INFRASTRUCTURES/1216/0017 (IRIDA).

REFERENCES
[1] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update,” Cisco, White Paper, Mar. 2017.
[2] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, “A survey on mobile edge computing: The communication perspective,” IEEE Commun. Surveys Tuts., vol. 19, no. 4, pp. 2232–2358, 4th Quart., 2017.
[3] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, “A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends,” IEEE J. Sel. Areas Commun., vol. 35, pp. 2181–2195, October 2017.
[4] F. Wang, J. Xu, and Z. Ding, “Optimized Multiuser Computation Offloading with Multi-antenna NOMA,” in IEEE Global Commun. Conf. Workshops, Singapore, China, Dec. 2017.
[5] G. Gu, G. Wang, J. Liu, R. Fan, D. Fan and Z. Zhong, “Optimal offloading with non-orthogonal multiple access in mobile edge computing,” in Proc. IEEE Global Commun. Conf., Abu Dhabi, United Arab Emirates, Dec. 2018.
[6] Y. Dai, M. Sheng, J. Liu, N. Cheng and X. Shen, “Resource allocation for low-latency mobile edge computing offloading in NOMA networks,” in Proc. IEEE Global Commun. Conf., Abu Dhabi, United Arab Emirates, Dec. 2018.
[7] S. Ali, E. Hossain, and D. L. Kim, “Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation,” IEEE Access, vol. 5, pp. 565–577, Nov. 2016.
[8] K. Wang, J. Cui, Z. Ding and P. Fan, “A Stackelberg game approach for NOMA in mmWave systems,” in Proc. IEEE Global Commun. Conf., Abu Dhabi, United Arab Emirates, Dec. 2018.
[9] Y. Liu, X. Fang, P. Zhou, and K. Cheng, “Coalition game for user association and bandwidth allocation in ultra-dense mmWave networks,” in Proc. IEEE/CIC Int. Conf. Commun., China, Qingdao, China, Oct. 2017.
[10] D. Tse and P. Viswanath, “Fundamentals of Wireless Communication.” Cambridge University Press, New York, NY, USA, 2005.
[11] M. Grant, S. Boyd, and Y. Ye, “CVX: Matlab software for disciplined convex programming,” 2009. [Online]. Available: http://cvxr.com/cvx/
[12] “Gurobi Optimizer: State of the Art Mathematical Programming Solver.” [Online]. Available: http://www.gurobi.com/products/gurobi-optimizer
[13] Z. Han, D. Niyato, W. Saad, T. BA Sar, and A. Hjorungnes, Game Theory in Wireless and Communication Networks - Theory, Models, and Applications, Cambridge University Press, 2012.
[14] G.-C. Rota, “The number of partitions of a set,” The American Mathematical Monthly, vol. 71, pp. 498–504, 1964.