Semi-Centralized Optimization for Energy Efficiency in IoT Networks with NOMA

Abdullah Alajmi, *Graduate Student Member, IEEE,* Muhammad Fayaz, *Graduate Student Member, IEEE,* Waleed Ahsan, *Member, IEEE,* and Arumugam Nallanathan, *Fellow, IEEE*

***Abstract*—We propose a novel semi-centralized framework for Internet-of-Things (IoT) networks with non-orthogonal multiple access to maximize the energy efficiency (EE) of two types of clients, namely grant-based (GB) and grant-free (GF). We use a proximal policy optimization algorithm to maximize the EE of GB clients and a multi-agent deep Q-network to optimize resource allocation for GF clients aided by a gateway node. The proposed algorithm combines the advantages of fully centralized and fully distributed frameworks to compensate for their shortcomings (complexity and long learning time). The numerical results show that the proposed algorithm enhances the EE of GB clients by 6% and 11.5%, respectively, compared with the fixed power allocation and random power allocation strategies. Moreover, the results demonstrate a 47.4% increase in the EE of GF clients over the benchmark scheme. Additionally, we show that the increase in the number of GB clients has a significant impact on the EE of GB and GF clients.**

***Index Terms*—Non-orthogonal multiple access, resource allocation, Internet-of-Things, deep reinforcement learning.**

I. Introduction

Non-orthogonal multiple access (NOMA) is considered a promising solution to provide massive connectivity to the increasing number of Internet-of-Things (IoT) clients, one of the important use cases of massive machine-type communication (mMTC) [1]. IoT clients are characterized by a small data rate, long battery life, sporadic transmission and different quality of service (QoS) requirements [2]. Therefore, the energy efficiency (EE) for such devices requires more attention, and it is vital for networks with NOMA to allocate resources more efficiently and appropriately [3]. In conventional NOMA transmissions, the communication between the sender and receiver is based on four-way handshakes, which are known as grant-based (GB) transmission. GB transmission is unsuitable for IoT scenarios because of the long latency and signal overhead. Later, the grant-free (GF) scheme was introduced by removing the handshaking process from the GB scheme.

Recently, another multiple access scheme, known as the semi-grant-free (SGF) NOMA [4], was proposed; here, two types of clients (GF and GB) share the same time/frequency resource block. SGF-NOMA improves the connectivity through GF transmission and fulfils the QoS requirements of GB clients via GB communication. However, to secure the QoS of GB clients, SGF-NOMA schemes suffer from EE and complexity problems.

In existing work [3–8], the authors have optimized the transmit power for only GF or GB clients, with the main focus on successive interference cancellation (SIC) order. Moreover, they used either the conventional optimization methods or deep Q-network (DQN) algorithm, which can only handle discrete actions. According to authors in [3], DQN is not effective with continuous action spaces because discretization increases action space exponentially. Furthermore, the discretization of action space also throws away information that could be vital in finding a solution to the problem. In other words, it triggers a quantization error for those tasks with continuous actions (e.g., power allocation). Therefore, algorithms such as proximal policy optimization (PPO) are preferable with continuous action spaces and can be used to solve these issues. The authors in [4] proposed two SGF methods to limit the admitted GF clients to the same resource block (RB) reserved by GB clients. The proposed scheme ensures the QoS of GB clients while GF clients transmit with fixed power without considering the channel gain or location of the clients. Adaptive power allocation for GB clients was proposed in [5] using a conventional optimization approach to enhance the outage performance of GF clients without considering the impact of path loss. The work given in [6] assumed a homogeneous distribution of clients, and only the first two GF clients with the largest channel gain were admitted to ensure the GB clients' performance. The work given in [7] used stochastic geometry to analyze the ergodic rate and outage performance while considering a dynamic threshold for admitting GF clients. Recently, multi-agent deep reinforcement learning (MA-DRL) based SGF-NOMA was proposed in [8], where only the transmit power of GF clients is optimized and GB clients transmit with a fixed power. The SIC process used in the above-mentioned schemes has a significant impact on the performance of GB/GF clients in terms of EE. Because GB and GF clients share the same RB at the same time, this adds to the complexity and energy consumption of these clients. In addition, direct access (to the BS) is the focus of existing work due to its simplicity. However, path loss increases with increasing distance, resulting in low energy efficiency and reduced rates. To overcome the effect of distance-dependent path loss, in the existing work, the source node needs to transmit at higher power [9]. However, IoT users have small processing and limited transmit power capability, which makes it impractical to communicate over a

A. Alajmi. M. Fayaz, and A. Nallanathan are with Queen Mary University of London, London, UK (email:{a.alajmi, m.fayaz, a.nallanathan}@qmul.ac.uk).

A. Alajmi is also with Prince Sattam bin Abdulaziz University, Al-Kharj, SA (email: a.alajmi@psau.edu.sa).

M. Fayaz is also with the Department of computer Science and IT, University of Malakand, Pakistan.

W. Ahsan is with the School of Informatics, The University of Edinburgh, UK, (email: wahsan@ed.ac.uk).

long distance.

Therefore, the this paper aims to take advantage of GB clients as a gateway node to improve the EE of delay-tolerant GF clients and improve the EE of GB clients via the PPO algorithm. The main contributions are as follows:

- We propose a new optimization framework where the GF client transmits its signal to the serving GB client, which is known as a gateway node, via the NOMA protocol. Furthermore, we formulate the EE of both the GF and GB clients as an optimization problem.
- To jointly optimize the transmit power of GB and GF clients, we propose a semi-centralized framework that avoids the disadvantages of fully centralized and fully distributed RL algorithms. In particular, we use the PPO algorithm on the BS side (centralized part) to determine the optimal power levels for GB clients. However, considering the computational limitations of GF clients; a multi-agent deep Q-network (MA-DQN) algorithm (distributed part) is utilized on the GF client side.
- The experimental results show that our proposed scheme outperforms the random and fixed power allocation methods and the conventional GF transmission without a gateway node in terms of EE. Moreover, we show that the number of GB clients has a strong correlation with the EE of both types of clients.

II. System Model And Problem Formulation

We consider a NOMA IoT network with a single BS located at the center of a circle with a radius R. Two types of clients, namely GB[1] (represented by $\mathcal{X}=\{1,2,\ldots,N_X\}$) and GF[2] (listed as $\mathcal{F}=\{1,2,\ldots,N_F\}$) transmit their data in an uplink manner, which is given in Fig. 1(a). The GF clients send their data to the GB client acting as a cluster head (CH) [10] to reduce the impact of the path loss with the distance d, here given by $d^{-\alpha}$ on the energy constrained GF clients. The GB clients transmit their data to BS via W sub-channels. The locations of the GB and GF clients are modeled as two homogeneous Poisson point processes with densities λ^{GB} and λ^{GF}. Hence, the number of GB and GF clients follows a Poisson distribution [8].

A. Signal Model

The GB and GF clients transmit their data in a slotted manner. More specifically, the CH $j \in \mathcal{X}$ receives the combined signal from the N_F GF clients in time slot t, which can be expressed as follows:

$$y_j(t) = \sum_{w=1}^{W}\sum_{i=1}^{N_F}\sqrt{P_{w,i}(t)}h_{w,i}(t)s_{w,i}(t)+n_0, \tag{1}$$

where $s_{w,i}$, $h_{w,i}$, and $P_{w,i}$ denote the transmitted signal, channel gain, and transmit power of i-th GF client on sub-channel w, respectively. Here, n_0 represents the additive Gaussian noise with variance $(0,\sigma^2)$. The channel decoding order is, $P_{w,i}h_{w,i}(t) \geq \cdots \geq P_{w,N_F}h_{w,N_F}(t)$. The signal-to-interference-plus-noise ratio (SINR) for GF client $i \in \mathcal{F}$ can be given as follows:

$$\gamma_{w,i}(t)=\frac{P_{w,i}(t)|h_{w,i}|^2(t)}{\sum_{i=i+1}^{N_F}P_{w,i+1}(t)|h_{w,i+1}|^2(t)+\sigma^2}. \tag{2}$$

The data rate of each GF client is calculated as follows:

$$R_{w,i}(t) = B\log\left(1+\gamma_{w,i}(t)\right) \geq \varepsilon_{\mathcal{F}}, \tag{3}$$

where B is the bandwidth of sub-channel w and ε_F is the target data rate threshold for $\mathcal{F}$ clients.

The EE of the GF clients can be calculated as follows:

$$EE_{\mathcal{F}}(t) \triangleq \frac{\sum_{w=1}^{W}\sum_{i=1}^{N_F}R_{w,i}(t)}{\varsigma(t)+\vartheta_{\mathcal{F}}(t)}, \tag{4}$$

where $\varsigma(t) = \sum_{w=1}^{W}\sum_{i=1}^{N_F}p_{w,i}(t)$ and $\vartheta_{\mathcal{F}}(t)$ is the circuit power consumed by $\mathcal{F}$ clients similar to [11].

In the next time slot $(t+1)$, the BS receives the combined signal from the CHs and other GB clients as follows:

$$y_{BS}(t+1)=\sum_{w=1}^{W}\sum_{j=1}^{N_X}\sqrt{P_{w,j}(t+1)}g_{w,j}(t+1)s_{w,j}(t+1)+n_0,$$

where $s_{w,j}$, $g_{w,j}$, and $P_{w,j}$ represent the transmitted signal, channel gain, and transmit power of j-th GB client, respectively. We consider channel gain based decoding order at the BS; that is, the BS decode the GB client with strong channel gain in the first stage of SIC $\mathcal{G}=\{g_1 \geq g_2 \geq \cdots \geq g_{N_X}\}$.

Similarly, the SINR for GB clients can be shown as follows:

$$\gamma_{w,j}(t+1)=\frac{P_{w,j}(t+1)|g_{j,w}|^2(t+1)}{\sum_{j=j+1}^{N_X}P_{w,j+1}(t+1)|g_{w,j+1}|^2(t+1)+\sigma^2}. \tag{5}$$

The data rate of each GB client is calculated as follows:

$$R_{w,j}(t+1) = B\log\left(1+\gamma_{w,j}(t+1)\right) \geq \varepsilon_{\mathcal{X}}, \tag{6}$$

where $\varepsilon_{\mathcal{X}}$ is the target data rate threshold for $\mathcal{X}$ clients.

The EE of GB clients in time slot $(t+1)$, we have

$$EE_{\mathcal{X}}(t+1) \triangleq \frac{\sum_{w=1}^{W}\sum_{j=1}^{N_X}R_{w,j}(t+1)}{\varrho(t+1)+\vartheta_{\mathcal{X}}(t+1)}, \tag{7}$$

where $\varrho = \sum_{w=1}^{W}\sum_{j=1}^{N_X}p_{w,j}$ and $\vartheta_{\mathcal{X}}$ is the circuit power consumed by $\mathcal{X}$ clients. Based on equations (4) and (7), the EE of the system can be given as follows:

$$EE=EE_{\mathcal{F}}(t)+EE_{\mathcal{X}}(t+1). \tag{8}$$

B. Cluster Head and Sub-channel Selection (GF clients)

In time slot t, each GF client is allowed to select at most one sub-channel and one GB client as a CH. The following variables are used for CH and sub-channel selection, respectively:

$$c_i(t) = \begin{cases} 1, \text{if } i\text{-th GF client selects } j\text{-th client as CH} \\ 0, \text{otherwise,} \end{cases}$$

$$b_{w,i}(t) = \begin{cases} 1, \text{if } i\text{-th GF client selects sub-channel } w \\ 0, \text{otherwise.} \end{cases}$$

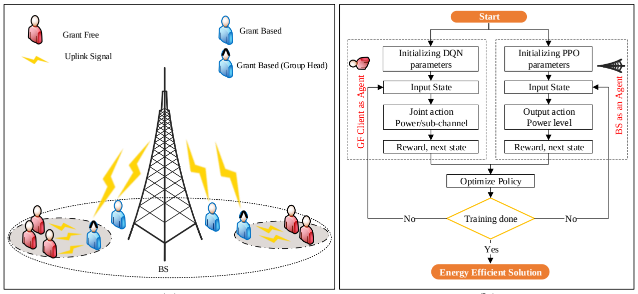
C. Sub-channel selection for GB clients

We use the following binary variable for GB clients to select sub-channel as follows:

$$m_{w,j}(t+1) = \begin{cases} 1, \text{if } j\text{-th client selects sub-channel } w \\ 0, \text{otherwise.} \end{cases}$$

[1] We assume that GB clients are delay sensitive and have enough processing capability to act as cluster head.

[2] The GF clients are delay tolerant, e.g, a sensor for temperature monitoring.



(a) (b)

Fig. 1: Sub-figure (a) shows the system model. Sub-figure (b) represents the flow chart of the proposed algorithm.

D. Problem Formulation

Our aim is to maximize the EE by optimizing m, b, c, and P. Therefore, the optimization problem can be formulated as,

$$\underset{m,b,c,P}{\text{maximize}} \sum_{t=1}^{T}\sum_{w=1}^{W}\sum_{j=1}^{N_X}\sum_{i=1}^{N_F} EE \tag{9}$$

$$\text{s.t.} \quad P_{w,i}(t), P_{w,j}(t{+}1) \le P_{max}, \forall w, i, j, t, \tag{9a}$$

$$\sum_{j=1}^{N_X} c_i(t) \in \{1,0\}, \quad \forall i, t, \tag{9b}$$

$$\sum_{w=1}^{W} b_{w,i}(t) \in \{1,0\}, \quad \forall i, t, \tag{9c}$$

$$\sum_{w=1}^{W} m_{w,j}(t{+}1) \in \{1,0\}, \quad \forall j, t, \tag{9d}$$

$$\sum_{w=1}^{W} R_{w,j}(t{+}1) \ge \varepsilon_{\mathcal{X}}, \quad \forall j, t, \tag{9e}$$

$$\sum_{w=1}^{W} R_{w,i}(t) \ge \varepsilon_{\mathcal{F}}, \quad \forall i, t, \tag{9f}$$

where (9a) is the maximum transmit power limit of clients. Constraint (9b) shows that GF clients can select only one cluster head at time slot t. Constraints (9c) and (9d) show that GF and GB clients are allowed to select at most one sub-channel in a given time slot t. Constraints (9e) and (9f) represent the minimum required data rate of GB and GF clients to ensure successful SIC, respectively.

III. Semi-centralized ML framework for EE

Machine learning (ML) algorithms for resource management are based on a centralized or distributed framework. In particular, in a centralized framework, a central entity (e.g., BS) is responsible for resource allocation, whereas, in the decentralized framework, resource allocation is handled by multiple agents (e.g., IoT clients). The downside of the former is increased computational complexity (CC) arising from the overwhelming demand, and the downside of the former is the lengthy learning/training time required to converge to optimality as a result of non-stationarity. To alleviate these challenges, we have designed a semi-centralized framework that minimizes the CC and reduces the learning time. The work flow of the proposed algorithm is given in Fig.1(b). Next, we formulate the EE problem as a Markov decision process (MDP) problem with the semi-centralized framework.

Remark 1. *This model can be used for orthogonal multiple access-NOMA (OMA-NOMA) scenarios for some applications, for example, the GB clients (enhanced Mobile Broadband (eMBB) client) can transmit using OMA, whereas GF clients can transmit using NOMA (mMTC client).*

A. MDP Elements with a Semi-Centralized Framework

An MDP consists of a tuple of $\big(\mathcal{N}, \mathcal{S}, \mathcal{A}$, and $\mathcal{R}(.)\big)$, where $\mathcal{N}$ is the number of agent(s) (BS, GF clients), $\mathcal{S}$ is the set of states, actions are denoted by $\mathcal{A}$, and $\mathcal{R}(.)$ is the reward function. To start the learning process, RL agents interact with the environment to maximize the long-term reward following some policy π.

- **Agent(s):** $\underbrace{\text{The BS}}_{\text{Centralized Part}}$ and $\underbrace{\text{GF clients}}_{\text{Decentralized part}}$
- **State:** $\underbrace{\text{Channel gain (GB)}}_{\text{Centralized part}}$ and $\underbrace{\text{data rate (GF)}}_{\text{Decentralized part}}$
- **Action (BS):** $\underbrace{\text{Transmit power}}_{\text{Centralized part}}$
- **Action (GF client):** $\underbrace{\text{Sub-channel, transmit power, and CH}}_{\text{Decentralized part}}$
- **Reward:** The BS as an agent receives the EE of the GB clients as a reward, whereas, the i-th GF client receives the EE of the GF clients as a reward signal, as given below.

$$r_i(t) = \begin{cases} EE_{\mathcal{F}}(t), & \text{if } EE_{\mathcal{F}}(t) \ge EE_{\mathcal{F}}(t{-}1) \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

$$r_{BS}(t{+}1) = \begin{cases} EE_{\mathcal{X}}(t{+}1), & \text{if } EE_{\mathcal{X}}(t{+}1) \ge EE_{\mathcal{X}}(t) \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Remark 2. *The PPO uses two deep neural networks (DNN) and handles a continuous action space, which increases the complexity; therefore, the PPO is used on the BS side. In contrast, the IoT clients are resource and computation constrained and can handle discrete actions; hence, such algorithms cannot be applied to IoT clients.*

Algorithm 1 Semi-Centralized Framework for EE NOMA Systems

```
Initialize hyperparameter                                          ▷ MA-DQN
for Episode = 1: N_e  do
    for iteration at time step (t) = 1: T_e  do
        for agent = 1: I do
            Input s_i(t), take a_i(t), receive r_i(t) using (10) and s_i(t+1)
            Store s_i(t), a_i(t), r_i(t), s_i(t+1) to replay memory
        end for
    end for
    Sample mini-batches from memory and minimize the loss using (13)
end for
Initialize policy parameters                                       ▷ PPO
for Episode = 1: N_e  do
    for actor = 1, 2, . . . , N  do
        Run policy π_θold for T_e time steps
        Calculate advantage estimates Â_1, . . . , Â_T
    end for
    Optimize L (14) w.r.t θ with N_e and mini-batch size M ≤ Te θ_old ← θ
end for
```

For the decentralized part (GF clients act as agents), we define a Q-function as the expected cumulative discounted reward to find the optimal policy π^*, which can be given as follows:

$$Q_i^{\pi}(s_i, a_i) = \mathbb{E}^{\pi}\big[Re(t)\big|s_i(t) = s, a_i(t) = a\big], \tag{12}$$

where Re is the discounted reward $Re = \sum_{n=0}^{N_e} \beta^n r^{(t+n+1)}$.

The conventional Q-learning algorithm maintains a Q-table for each agent to store the corresponding Q-values of the Q-function, which is computationally expensive because of large state and action spaces. Therefore, DQN with weights θ is used for the Q-function approximation $Q_i^{\pi}(s_i, a_i; \theta)$. The DQN consists of a primary network used to evaluate the primary Q-value and a target network for target Q-value estimation. The weights of the primary network are updated in every learning step, while the weights of the target network are updated after a fixed number of training iterations. To train the Q-network, stochastic gradient descent (SGD) is used to update the weights and minimize the loss between the primary and target Q-network.

$$L(\theta) = (y_i(t) - Q_i(t)(s_i(t), a_i(t)))^2, \tag{13}$$

where $y_i(t) = r_i(t) + \max_{a_i} Q(s_i(t+1), a_i; \theta)$.

For the centralized part (agent BS), we apply the PPO algorithm to find the optimal transmit power for GB clients. The PPO is a policy gradient method that utilizes the actor-critic method and can be used in environments with continuous action space. In the stochastic policy, the actor maps an observation to an action, and the critic calculates the reward for the given observation. A stochastic gradient ascent optimizer is used to update the policy, and an SGD technique is used to fit the value function. The loss function can be calculated as follows:

$$L(\theta) = \hat{\mathbb{E}}(t)[min\big(r(t)(\theta)\hat{A}(t), clip\big(r(t)(\theta)1-\epsilon, 1+\epsilon\big)\hat{A}(t)\big)], \tag{14}$$

where $\hat{\mathbb{E}}(t)$ represents the empirical expectation over time steps and θ represents the policy parameter. The $\hat{A}(t)$ is the estimated advantage at time step (t), $r(t)$ denotes the ratio of the probability under the new and old policies. This equation has two parts; first it minimizes the loss of conservative policy iteration $(min(r(t)(\theta)\hat{A}(t))$, and in the second part, we have $(\text{clip}(r(t)(\theta)1-\epsilon, 1+\epsilon)\hat{A}(t))$, where we clip the policy ratio between $1+\epsilon$ and $1-\epsilon$.

B. Proposed Semi-Centralized Algorithm

To maximize the EE of GF clients, each GF client and BS acts as an agent to find the optimal policy. The details of the algorithm are given in **Algorithm-1**. In the distributed part, we initialize the network and training parameters before the start of agents training. All agents (GF clients) jointly explore the environment using a ϵ-greedy policy. The agents receive states from the environment, and they take a joint action. Based on the action taken, agents obtain a reward and next state from the environment. All agents save experiences to their replay memories (line 6). To train the primary network, all the agents randomly sample mini-batches from the memory and compute the loss (line 9). For the centralized part (line 11), first, we initialize the policy parameters. We run the policy $\pi_{\theta_{old}}$ for T_e time steps to calculate the advantage estimates. Finally, we calculate the loss with respect to θ using a mini-batch of size M and update θ_{old} with θ (line 17).

C. Complexity

The complexity results from a number of GB N_X and GF N_F clients connecting to BS via sub-channel W. The total number of all clients (GB and GF) is denoted as $\hat{\mathcal{N}}$. At each time step t, the CC of the proposed algorithm is given by, $\mathcal{O}\big[N_e \times T_e\big((N_X^W) + N_F\big)\big]$. In contrast, the CC of the centralized framework can be given by, $\mathcal{O}\big[N_e \times T_e(\hat{\mathcal{N}}^W)\big]$. For example, if we have five GB clients and five GF clients in a centralized framework, the complexity is increased exponentially. On the other hand, in our proposed algorithm, the complexity is distributed among the BS and GF clients.

IV. Simulation Results

In this section, we evaluate the performance of both the GB and GF clients. The parameters given in Table I are used to obtain the simulation results.

TABLE I: Simulation Parameters

GB clients	$(3-15)$	GF clients	$(3-15)$
Power levels	$[0.1, ..., 0.9]$ W [12]	Sub-channels	3 [12]
Sub-channel bandwidth	10 KHz [12]	α	2.8
Min rate (GB clients)	10 bps/Hz [12]	Episodes	300
Min rate (GF clients)	4 bps/Hz	Learning rate	0.001 [12]
DNN activation	ReLU	Optimizer	Adam

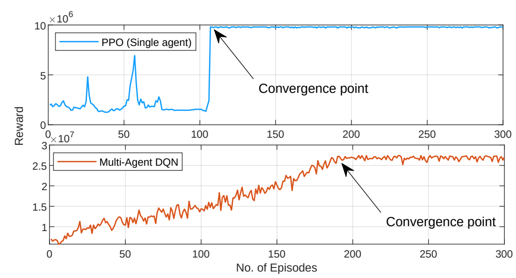


Fig. 2: Shows the convergence of the PPO and MA-DQN.

Fig. 2 shows the convergence of the PPO algorithm at the BS side to allocate the power to GB clients and the MA-DQN algorithm for the GF clients to find the optimal power level. The centralized agent, i.e., the BS, finds the optimal power level for each GB client after 100 episodes, as seen in the top sub-figure of Fig. 2. Compared with the decentralized MA-DQN, the PPO converges quickly. However, for a large number of GB clients, the PPO may require more training time because of the continuous action space. The bottom sub-figure shows the convergence of MA-DQN. There is a fluctuation in the reward because the actions of one agent affect other agents in the environment. Therefore, MA-DQN requires more episodes for convergence.

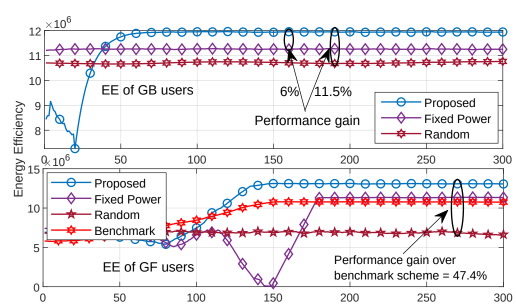


Fig. 3: Performance comparison of the proposed scheme with fixed power, random power, and the benchmark scheme.

Fig. 3 provides a comparison of the EE of GB clients and GF clients with other methods. The top sub-figure shows the EE of GB clients, and it can be observed that the proposed scheme outperforms the fixed power allocation and random power allocation methods. Because the BS identifies the accurate power levels according to the client channel gain, maintaining the QoS requirements of those GB clients with a minimum power consumption. In contrast, in the other two methods, clients transmit power without considering the channel gain, which increases intra-cluster interference, hence recording a low EE. The EE of GF clients is depicted in the bottom sub-figure of Fig. 3. For comparison, we use the conventional GF method as a benchmark, where the clients directly transmit their data to the central BS. It is observed that our proposed scheme performs well when compared with the benchmark scheme, fixed power selection,and random power selection methods. Unlike the benchmark scheme, the GF clients in our proposed scheme transmit their data to the nearest cluster head, which requires a minimum transmit power and enhances the EE.

Fig. 4 shows the trade-off between the number of GB clients and EE of both types of clients. It can be observed that the EE of GF clients increases with the number of GB clients. Because the GF clients have an increased choice in cluster head selection, they transmit to their nearest cluster head. On the other hand, as the number of GB clients increases, it decreases the EE of GB clients because this increases the number of clients in each cluster, which increases the intra-cluster interference. To achieve the required data rate threshold with increased interference, GB clients are required to transmit with high transmit powers.

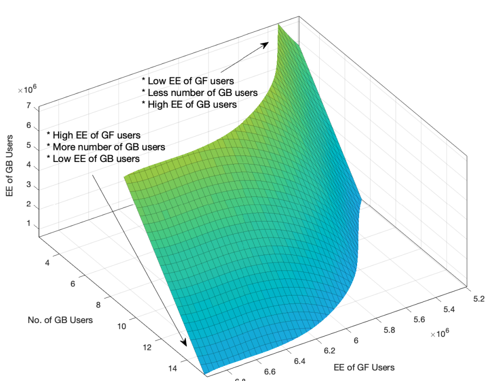
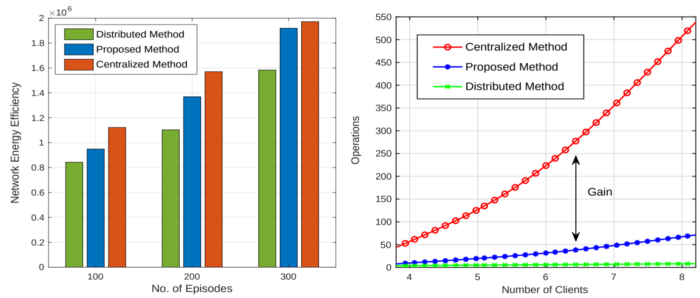


Fig. 4: The EE of GF and GB clients w.r.t increasing GB clients.

Fig. 5 compared the proposed method with the centralized and distributed framework in terms of EE and CC. Fig. 5 (a) represents the total network EE w.r.t a different number of episodes. It is concluded that the fully centralized method provides the highest EE. However, as the number of episodes increases the EE of the proposed method approaches the EE of centralized method. Because distributed methods need a long learning time to fully explore the environment. From Fig. 5 (b), it can be seen that the complexity of the centralized model is increasing exponentially as the number of clients increases. In our proposed framework, the complexity is distributed between the BS and GF clients. In a fully distributed model, all the



(a) (b)

Fig. 5: Performance Comparison: Sub-figure (a) Network EE. Sub-figure (b) CC.

clients are independently searching for resources without any centralized entity (BS), which reduces the CC. However, a fully distributed framework requires a long learning time to reach the Nash Equilibrium. On the other hand, the centralized method can easily find the optimal resources for users but at the cost of a high CC.

V. Conclusion

In this letter, we have proposed a low-complexity semi-centralized framework for NOMA networks to avoid the disadvantages of fully centralized and fully distributed systems. The proposed scheme improves the EE of GB and GF clients and outperforms the fixed and random power allocation methods. The EE of GF clients surpasses the EE of the conventional GF scheme where no group head exists.

References

[1] W. Ahsan, W. Yi, Z. Qin, Y. Liu, and A. Nallanathan, "Resource allocation in uplink NOMA-IoT networks: A reinforcement-learning approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5083–5098, 2021.

[2] Z. Ding, "Harvesting devices' heterogeneous energy profiles and QoS requirements in IoT: WPT-NOMA vs BAC-NOMA," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 2837–2850, 2021.

[3] Y. Zhang, X. Wang, and Y. Xu, "Energy-efficient resource allocation in uplink NOMA systems with deep reinforcement learning," in *Proc. 11th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2019.

[4] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4464–4478, 2019.

[5] Z. Yang, P. Xu, J. A. Hussein, Y. Wu, Z. Ding, and P. Fan, "Adaptive power allocation for uplink non-orthogonal multiple access with semi-grant-free transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1725–1729, 2020.

[6] N. Jayanth, P. Chakraborty, M. Gupta, and S. Prakriya, "Performance of semi-grant free uplink with non-orthogonal multiple access," in *IEEE 31st Annual Int. Symp. Personal, Indoor and Mobile Radio Commun.*, 2020, pp. 1–6.

[7] C. Zhang, Y. Liu, W. Yi, Z. Qin, and Z. Ding, "Semi-grant-free NOMA: Ergodic rates analysis with random deployed users," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 692–695, 2020.

[8] M. Fayaz, W. Yi, Y. Liu, and A. Nallanathan, "Competitive MA-DRL for transmit power pool design in semi-grant-free NOMA systems," *arXiv preprint arXiv:2106.11190*, 2021.

[9] A. Mondal, A. M. A. Junaedi, K. Singh, and S. Biswas, "Spectrum and energy-efficiency maximization in RIS-aided IoT networks," *IEEE Access*, vol. 10, pp. 103 538–103 551, 2022.

[10] S. Han, X. Xu, L. Zhao, and X. Tao, "Joint time and power allocation for uplink cooperative non-orthogonal multiple access based massive machine-type communication network," *Int. J. of Dist. Sensor Net.*, vol. 14, no. 5, 2018.

[11] F. Fang, Z. Ding, W. Liang, and H. Zhang, "Optimal energy efficient power allocation with user fairness for uplink MC-NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1133–1136, 2019.

[12] M. Fayaz, W. Yi, Y. Liu, and A. Nallanathan, "Transmit power pool design for grant-free NOMA-IoT networks via deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7626–7641, 2021.