Secure Energy Aware Power Control in Consumer Internet of Things with Semi Grant Free NOMA

Sohail Abbas, Muhammad Fayaz, Abdulrahman Ghandoura, Muhammad Zahid Khan, and Ateeq Ur Rehman

***Abstract*—The Consumer Internet of Things (CIoT), a key aspect of the IoT, aims to integrate smart technologies into everyday life. In order to improve the spectral efficiency and provide massive connectivity to IoT networks, non-orthogonal multiple access (NOMA) variants like semi-grant-free (SGF) NOMA are employed. This paper aims to maximize secrecy energy efficiency (EE) for SGF-NOMA enabled CIoT in the presence of untrusted users (eavesdroppers) by utilizing a single-agent multi-agent deep reinforcement learning (SAMA-DRL) algorithm to overcome scalability and expensive learning issues. Given the limited long-distance transmission capabilities of CIoT devices, which typically have low transmit power, relay nodes are used to decode and forward data from grant-free (GF) users to the base station. Moreover, to enhance the coverage for GF users, the K-nearest neighbors (KNN) algorithm is utilized to place the relay nodes at an optimal positions. Furthermore, we design a collaborative contribution reward system to discourage user (agent) laziness. Simulation results show that the proposed SAMA-DRL-based SGF-NOMA algorithm for CIoT networks is more effective than baseline algorithms, achieving a 20% increase in secrecy EE compared to DRL-based SGF-NOMA without KNN. Moreover, the proposed scheme outperforms benchmark schemes in terms of EE across different radii. Additionally, we show that the proposed algorithm with quality of service based successive interference cancellation (SIC) is more power efficient as compared to conventional SIC decoding order.**

***Index Terms*—Non-orthogonal multiple access, grant-free, Internet of things, grant-based, deep reinforcement learning.**

I. Introduction

INTERNET of Things (IoT) is a concept that involves connecting numerous physical objects or things that are equipped with sensors, actuators, and communication abilities, enabling them to gather and share data via the Internet [1]. IoT is gaining significant interest in the context of fifth generation and upcoming sixth generation communication systems. In particular, IoT-driven communication technology has potential applications in different domains like smart home, smart grid, and smart transportation, where extensive connectivity is needed and can significantly enhance current services [2]. It is predicted that there will be approximately 80 billion connected devices by 2030, which is roughly 21 devices per person [3]. While IoT spans a multitude of industries, consumer IoT (CIoT) focuses on bringing smart technologies into people's everyday life improving it across different facets. Applications of CIoT includes home automation, wearable devices, connected personal spaces etc. [4].

In the context of massive CIoT, important traffic features include predominantly uplink traffic, where numerous devices send small data transmissions. This scenario highlights the significance of energy efficiency (EE) due to the need to conserve battery life in potentially large-scale deployments. Moreover, there is a requirement for partially or fully autonomous communication and most importantly, the handling of sporadic transmissions, where devices communicate infrequently or at irregular intervals. Moreover, CIoT devices are typically lightweight, and a significant challenge is the limited energy capacity and providing massive connectivity to these devices. To accommodate the vast number of devices in cellular-enabled CIoT networks, non-orthogonal multiple access (NOMA) technique is a promising solution that enhances spectrum efficiency by allowing multiple users to transmit data using different power levels, thus enabling multiplexing of the same spectrum resource [5]. According to the CIoT features mentioned above, NOMA offers various access methods to meet the diverse requirements of CIoT networks.

- Grant-Based (GB) NOMA: In the GB NOMA access method, IoT users are required to accomplish a series of handshakes with the base station (BS) before they can actually transmit data [6]. This approach is specifically designed for IoT applications that exhibit structured and predictable communication patterns, enabling efficient management of network resources and maintaining organized communication between devices. However, this method results in a significant increase in signal overhead due to the handshake process.
- Grant-Free (GF) NOMA: In this approach, the need for a grant process is removed, allowing users to transmit data directly to the BS [7]. This is ideal for applications that need instant data transfer without the delay caused by the handshake process. GF communication reduces signal overhead and latency. However, GF communication is susceptible to reliability collision problems.
- Semi-Grant-Free (SGF) NOMA: In SGF access, both GB and GF users utilize the same resource block for transmitting data. SGF-NOMA represents a balanced approach, integrating the advantages of both GB and GF methods to serve the diverse CIoT requirements [6]. Additionally, it provides adaptive Quality of Service (QoS), which

Sohail Abbas is with the Department of Computer Science, College of Computing and Informatics, University of Sharjah, Sharjah, UAE (email: sabbas@sharjah.ac.ae).

M. Fayaz, and M. Zahid Khan are with the Department of Computer Science and IT, University of Malakand, Pakistan (email:{m.fayaz, mzahid-khan}@uom.edu.pk).

Abdulrahman Ghandoura is with the Department of Engineering and Applied Sciences, Applied College, Umm Al-Qura University Makkah, 24382, Saudi Arabia (email:amghandoura@uqu.edu.sa).

Ateeq Ur Rehman Department of computing, Staffordshire University, UK (email: ateequr.rehman@staffs.ac.uk).

dynamically modifies the access method to maintain QoS for different types of CIoT traffic.

The above NOMA access techniques are designed to meet the specific demands and needs of CIoT networks, including the support for a large number of devices, EE, low latency, and consistent and reliable communication. However, the effectiveness of these access methods depends on users' power control and clustering approach. Moreover, the inherent broadcast nature of radio and the implementation of successive interference cancellation (SIC) at the receivers render NOMA vulnerable to potential security breaches from both external and internal eavesdroppers. Therefore, an effective power control method is required not only for optimizing EE, which aims to optimize the use of power for data transmission, but also for ensuring secrecy EE, which aims to optimize energy consumption for data transmission as well as ensures that this power usage safeguards the data from unauthorized access.

A. Related Work

To enhance EE, researchers have explored joint optimization of resources and communication radio in the literature. To increase the overall EE of the GB system, the authors in [13] optimized the transmit beamforming at the BS and the reflecting beamforming at the intelligent reflecting surface (IRS). Motivated by the limitation of user power, the study given in [14] focuses on maximizing EE of GB-NOMA system. The problem of maximizing EE is transformed into a series of sub problems aimed at maximizing sum rate thereby using fractional programming. Every sub-problem is then addressed using the proposed iterative water-filling solution. In order to maximize EE, the authors of [15] investigate the issue of jointly allocating power and subchannels in an uplink multi-user NOMA system. In order to tackle this non-convex optimization problem, the study suggests three deep reinforcement learning (DRL) based frameworks, in contrast to conventional model-based resource allocation techniques. The goal of many research works in literature is to maximize the sum rate for GF and SGF-NOMA based IoT networks. However, there are limited studies that specifically address the issue of EE. For example, GF schemes utilizing traditional optimization methods are examined in [16] and [17]. The authors of [16] and [17] partitioned the cell area and users and sub-channels into equal segments in order to avoid collisions among IoT users, employing orthogonal resources in various layers. The allocation of resources for GF transmission using DRL is discussed in several studies, including [18], [19], [20], and [21]. In [18], the authors have suggested a technique for minimizing collisions by arranging sub-channel clusters and users within a designated area, where users compete for access to available sub-channels using a GF approach. They addressed the long-term cluster throughput issue by employing a DRL algorithm for optimal power and sub-channel allocation in GF-NOMA. In [19], users were represented as cluster heads to optimize capacity and fulfill time constraints through the application of a multi-agent learning algorithm. The study in [20] explored the issue of maximizing data rate and the number of successful long-term transmissions using a Q-learning algorithm. The authors of [21] designed a transmit power pool for GF-NOMA to maximize system throughput. Only a single work given in [22] maximizes the network EE using multi-agent (MA) DRL.

TABLE I: Primary outcomes and limitations of recent works

Reference	Primary outcomes	Limitations
[8]	EE	Optimized power for GB users only
[9]	Ergodic rate	Perfect SIC
[10]	Outage performance	Perfect SIC
[11]	Minimize waiting delay	Restrict number of GF users
[12]	Secrecy rate	Perfect SIC

The SGF-NOMA transmission scheme was initially presented in [6] to enhance connectivity and reduce collisions. The scheme involves a single GB user sharing the channel with multiple GF users through NOMA, and proposes two contention control mechanisms to minimize interference to the GB user from the GF users. The researchers derived closed-form expressions for the outage probability of the GF users and explored the impact of different SIC decoding orders. The study in [23] enhances transmission resilience and successfully lowers error floors in outage probability without requiring users to precisely control their power. A method for adaptive power allocation was introduced in [8] to control the transmission power of GB users according to their channel conditions and desired data rate, thereby ensuring reliable decoding of their signals in the second stage of SIC. In a study conducted in [9], the operation of an uplink SGF-NOMA system was analyzed, which included multiple uniformly distributed GF and GB users. The proposed scheme paired the GF user with the received power below that of the GB user. The study derived closed-form expressions for the precise and approximate ergodic rates of both the GB and GF users. The researchers in [10] investigated the impact of GF users' random locations on the effectiveness of SGF-NOMA systems through the application of stochastic geometry. To enhance the throughput and minimize the waiting delay of GF users, the work in [11] introduces a NOMA-assisted SGF scheme with a hybrid SIC technique, enabling a specific number of GF users to share the GB user's channel. The authors in [24] introduce a SGF-NOMA scheme, facilitating multiple multi-antenna mobile terminals and a single earth station to utilize the satellite network concurrently within a shared resource block. A recent study in [12] examined the effectiveness of SGF-NOMA in enhancing the secrecy performance. The research derives analytical expressions for both exact and asymptotic secrecy outage probability.

B. Motivation and Contributions

The aforementioned studies primarily focus on maximizing the sum rate and assume perfect SIC at the receiver which is impractical. Moreover, these methods mainly concentrate on optimizing power allocation, but they frequently neglect the numerous QoS requirements security issues that are common in CIoT networks. Additionally, in these approaches, IoT users transmit directly to the base station. However, in IoT scenarios, IoT devices are typically lightweight with lower transmit power capability, limiting them to short-distance transmissions. Furthermore, in the SGF-NOMA schemes mentioned above, only the transmit power for GB or GF users is opti-

TABLE II: List of symbols

Symbol	Description
R	Radius of the circular cell area
$\mathcal{N}$	Set of GB users
$\mathcal{M}$	Set of GF users
$\mathcal{G}$	Set of CIoT relays
$\mathcal{S}$	Set of sub-channels
E	Eavesdropper
λ	Density of GF users
d_n	Distance from n-th GB user to BS
d_m	Distance from m-th GF user to BS
α	Path loss exponent
$h_{n,s}$	Channel gain from n-th GB user to BS on sub-channel s
$h_{m,s}$	Channel gain from m-th GF user to BS on sub-channel s
$p_{n,s}$	Transmit power of n-th GB user on sub-channel s
$p_{m,s}$	Transmit power of m-th GF user on sub-channel s
$x_{n,s}$	Transmitted signal of n-th GB user on sub-channel s
$x_{m,s}$	Transmitted signal of m-th GF user on sub-channel s
w_0	Additive white Gaussian noise (AWGN)
σ^2	Variance of the AWGN
η_m	Battery level of m-th GF user
I_{SIC}	Residual interference from SIC imperfection
$\gamma_{n,s}^{\mathrm{GB}}$	SINR of n-th GB user on sub-channel s
$\gamma_{m,s}^{\mathrm{GF}}$	SINR of m-th GF user at CIoT relay node
$\gamma_{n,E}$	SINR of eavesdropper intercepting n-th GB user
$\gamma_{m,E}$	SINR of eavesdropper intercepting m-th GF user
$R_{n,s}$	Secrecy rate of n-th GB user on sub-channel s
$R_{m,s}$	Secrecy rate of m-th GF user on sub-channel s
$E(t)$	SEE of the network
$\delta_{\mathcal{N}}$	Total transmit power of GB users
$\phi_{\mathcal{N}}$	Circuit power of GB users
$\delta_{\mathcal{M}}$	Total transmit power of GF users
$\phi_{\mathcal{M}}$	Circuit power of GF users
$c_{n,s}$	Sub-channel selection indicator for n-th GB user
$b_{m,g}$	Relay node selection indicator for m-th GF user
$c_{m,s}$	Sub-channel selection indicator for m-th GF user
τ	Required data rate for GB users
$\bar{\tau}$	Required data rate for GF users
P_{max}	Maximum transmit power
Z	Maximum number of GF users connected to one CIoT relay

mized. However, it is essential to jointly optimize the transmit power for both GF and GB users in order to take full advantage of SGF-NOMA's benefits for IoT networks. Therefore, this article focuses on the EE maximization with the help of a relay node and simultaneously optimizing the transmit power of both GB and GF users. The primary contributions of this study are outlined as follows:

- *Problem Formulation:* We formulate the secrecy energy efficiency of GF and GB users in an SGF-NOMA based CIoT network as an optimization problem, which seeks to jointly optimize the transmit power, sub-channel selection, and relay node selection. Secrecy energy efficiency ensures the power used for data transmission also ensure secure communication in the presence of eavesdroppers. To enhance the network lifetime, we have proposed a QoS-based SIC decoding order. This methodology not only prioritizes users based on their QoS requirements, but also considers their battery levels, leading to a more balanced SIC decoding order. Additionally, the GB users are cooperating as a decode and forward relay to extend the coverage area and enhance signal quality, particularly for GF users situated at the network's edge or in areas with poor signal strength.
- *SAMA-DRL Framework with Collaborative Contribution Reward Function:* We design single agent multi-agent (SAMA) DRL framework to address the defined optimization problem. In this framework, the base station (BS) acts as a single agent and interacts with multiple agents (GF users) to determine the best actions, such as transmit power, sub-channel, and relay selection. Moreover, we employ the K-Nearest Neighbor (KNN) algorithm to position the relay node in a suitable location in order to maximize coverage for GF users. Additionally, a reward function has been developed for the multi-agent system to assess each agent's contribution to the objective function and discourage the lazy agents.
- *Proposed Scheme Performance Evaluation:* Initially, we investigate the performance of our designed reward function by examining both overall reward and individual reward to find out the effectiveness. Following this, we evaluate our suggested SAMA-DRL algorithm against various baseline algorithms. The simulation findings confirm the effectiveness of the proposed algorithm, demonstrating superior performance compared to the baseline methods in terms of EE and user lifespan across various system parameter setups.

The rest of this article is structured as follows. Section II introduces the system model and the optimization problem. Section III examines the proposed SAMA-DRL framework. Section IV outlines the results of our simulations. Section V concludes and summarizes the article.

II. System Model

We consider an CIoT network with a single BS using SGF-NOMA technology. The BS is situated in the middle of a circular cell area with a radius of R. We consider a set of GB users represented by $\mathcal{N} = \{1, \cdots, N\}$ share the RBs via NOMA principles with a set of GF users denoted as $\mathcal{M} = \{1, \cdots, M\}$. Moreover, we consider a set of CIoT relays $\mathcal{G} = \{1, \cdots, G\}$ distributed inside the cell area. These CIoT users transmit their uplink data via sub-channels $\mathcal{S} = \{1, \cdots, S\}$. Additionally, we assume that there is an evesdropper E that trying to intercept the signal of both GF and GB users. We assume that in a given time slot t, the GB users transmit their data directly to the central BS and GF users to their nearest CIoT relay node to save energy. The distribution of GF and GB users are modelled using Homogeneous Poisson point processes (PPPs) with densities λ_{GF} and λ_{GB}, respectively. We express the channel gain from n-th GB CIoT user to BS with distance d_n and path loss exponent α in the s-th sub-channel as $h_{n,s} = |h_n|^2 (d_{n,s})^{-\alpha}$. Similarly, the channel gain between the m-th GF user and BS is defined as $h_{m,s} = |h_m|^2 (d_{m,s})^{-\alpha}$, where d_m is the distance from m-th GF user to the BS. The channel gains of both types of users are determined by small-scale Rayleigh fading and path loss. List of notations used in this paper is given in Table II.

A. Transmission with SGF-NOMA

In SGF-NOMA, GB and GF users utilize a common RB to improve the connectivity that creates SGF-NOMA. It is important to mention that the traditional GB transmission offers a greater capacity than what is typically needed by CIoT users in many cases. This surplus capacity can be used

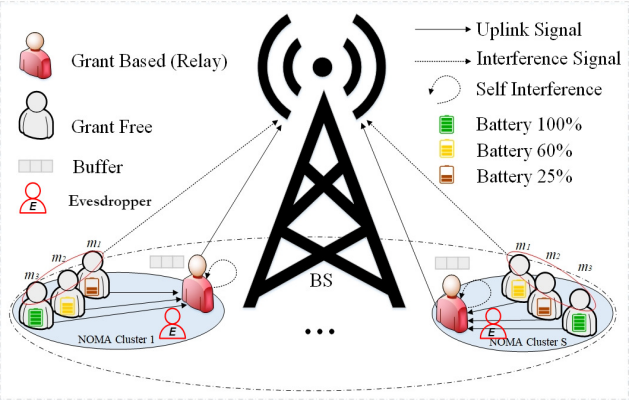


Fig. 1: An illustrative layout of the proposed SGF-NOMA based CIoT network. GF users send their data to the closest relay node, while GB users send their data to the BS.

to enhance connectivity for GF users through GF access. The combined received signal from GB and GF users at the BS on the s-th sub-channel in time slot t can be given as

$$y_{\mathrm{BS},s}(t) = \underbrace{\sum_{n=1}^{N_s} \sqrt{p_{n,s}(t)} h_{n,s}(t) x_{n,s}(t) +}_{\text{Desired signal}} \underbrace{\sum_{m=1}^{M_s} \sqrt{p_{m,s}(t)} h_{m,s}(t) x_{m,s}(t) + w_0}_{\text{Interference from GF users}} \quad (1)$$

where $p_{n,s}$, $x_{n,s}$ is the transmit power and transmitted signal of GB CIoT user n on sub-channel s, respectively. The $p_{m,s}$, $x_{m,s}$ represents the transmit power and transmitted signal from GF m-th GF user, respectively. The w_0 is the additive white Gaussian noise (AWGN) with zero mean and variance σ^2. Likewise, the combined signal received at g-th IoT gateway[1] node can be expressed as

$$y_{g,s}(t) = \underbrace{\sum_{m=1}^{M_s} \sqrt{p_{m,s}(t)} h_{m,s}(t) x_{m,s}(t) +}_{\text{Desired signal}} \underbrace{\sum_{n=1}^{N_s} \sqrt{p_{n,s}(t)} h_{n,s}(t) x_{n,s}(t) + w_0}_{\text{Interference from GB users}} \quad (2)$$

B. Signal Model and QoS Based SIC Decoding

In NOMA communication, the receiver uses SIC to decode and separate the signal for different CIoT users from the combined received signal. The SIC will be used at the GB (relay node) user and at the BS. The GF users send their signals to the nearest relay node which decodes and saves the data of GF users for a single time slot in a buffer and transmits it in the next time slot with their own data to the central BS.

[1]A gateway node acts as relay node that receives data from GF users and forwards it to the BS, thereby enhancing connectivity and extending the communication range.

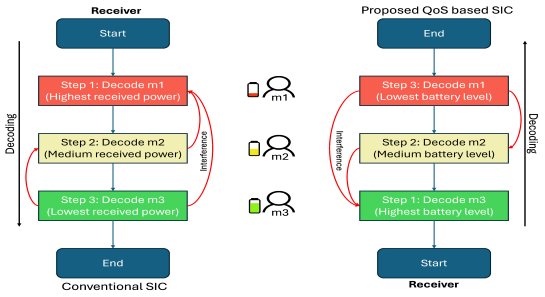


Fig. 2: Illustrates conventional and proposed SIC order.

The SIC at the GB (relay node) is based on the battery level of each GF user. Unlike conventional SIC process (where the user with highest received power level at the receiver is decoded first), as shown in Fig. 2, the user with highest battery power level will be decoded first in our proposed scheme[2]. The user with lowest battery power level will be decoded in the last stage of SIC in order to extend his battery life by achieving his QoS requirements with lowest transmit power. Let the battery level of each GF user m is represented by η_m, then the SIC decoding order in terms of battery power level is given as

$$\eta_{1,s}(t) \geq \eta_{2,s}(t) \geq \cdots \geq \eta_{m,s}(t) \geq \cdots, \geq \eta_{M,s}(t). \quad (3)$$

We assume that the SIC at the BS is determined by the order of received power strength, as stated in [25], we have

$$p_{1,s}h_{1,s}(t) \geq p_{2,s}h_{2,s}(t) \geq \cdots p_{n,s}h_{n,s}(t) \cdots \geq p_{N,s}h_{N,s}(t) \quad (4)$$

For practical implementation, we assume imperfect SIC, where some residual interference I_{SIC} arises from SIC imperfection.

The signal-to-interference-plus-noise ratio (SINR) at the BS for n-th GB user on sub-channel s in time slot t is given by

$$\gamma_{n,s}^{\mathrm{GB}}(t) = \frac{p_{n,s}h_{n,s}(t)}{\sum_{\bar{n}=n+1}^{N_s} p_{\bar{n},s}h_{\bar{n},s}(t) + \sum_{m=1}^{M_s} p_{m,s}h_{m,s}(t) + I_{SIC} + \sigma^2},$$

whereas the SINR of the m-th GF user at CIoT relay node can be expressed as

$$\gamma_{m,s}^{\mathrm{GF}}(t) = \frac{p_{m,s}h_{m,s}(t)}{\sum_{\bar{m}=m+1}^{M_s} p_{\bar{m},s}h_{\bar{m},s}(t) + \sum_{n=1}^{N_s} p_{n,s}h_{n,s}(t) + I_{SIC} + \sigma^2},$$

where I_{SIC} is a random variable follows Gaussian distribution with zero mean and variance σ_{SIC}^2. The eavesdropper E trying to intercept the signal of a GB user n or a GF user m at time t, the SINR expressions could be defined as:

$$\gamma_{n,E}(t) = \frac{p_{n,s}(t)|h_{n,E}(t)|^2}{\sum_{i\neq n}^{N_s} p_{i,s}(t)|h_{i,E}(t)|^2 + \sum_{j=1}^{M_s} p_{j,g}(t)|h_{j,E}(t)|^2 + \sigma^2},$$

$$\gamma_{m,E}(t) = \frac{p_{m,s}(t)|h_{m,E}(t)|^2}{\sum_{i=1}^{N_s} p_{i,s}(t)|h_{i,E}(t)|^2 + \sum_{j\neq m}^{M_s} p_{j,g}(t)|h_{j,E}(t)|^2 + \sigma^2},$$

[2]We assume that GF users periodically report their battery status to the BS or gateway node. Moreover, to prevent excessive signaling overhead, updates can be triggered by events, such as when the battery level drops below predefined critical thresholds.

where $h_{n,E}(t)$ and $h_{m,E}(t)$ are the channel gains from the users to the eavesdropper. The secrecy rate for the GB and GF users at time t can be calculated as:

$$R_{n,s}(t) = [R_{n,BS}(t) - R_{n,E}(t)]^+$$
$$R_{m,s}(t) = [R_{m,R}(t) - R_{m,E}(t)]^+,$$

where $R_{n,BS}(t)$ and $R_{m,R}(t)$ are the rates of the GB user n and GF user m at the BS and relay node respectively, $R_{n,E}(t)$ and $R_{m,E}(t)$ are the rates at which the eavesdropper can potentially decode the signals of n and m, and $[x]^+$ denotes the positive part of x, i.e., $\max(x, 0)$. The SEE of the network can be calculated as follows

$$E(t) \triangleq \sum_{t=1}^{T}\sum_{s=1}^{S}\left(\frac{\sum_{n=1}^{N} R_{n,s}(t)}{\delta_{\mathcal{N}}(t)+\phi_{\mathcal{N}}(t)} + \frac{\sum_{m=1}^{M} R_{m,s}(t)}{\delta_{\mathcal{M}}(t)+\phi_{\mathcal{M}}(t)}\right), \quad (5)$$

where $\delta_{\mathcal{N}} = \sum_{s=1}^{S}\sum_{n=1}^{N} p_{n,s}(t)$ and $\phi_{\mathcal{N}}(t)$ is the circuit power used by GB users. The $\delta_{\mathcal{M}} = \sum_{s=1}^{S}\sum_{m=1}^{M} p_{m,s}(t)$ and $\phi_{\mathcal{M}}(t)$ is the amount of circuit power for GF users.

C. Relay Node and Sub-channel Selection

For the given system model, sub-channel selection for GB users and two selections needed to be optimized for GF users, i.e., CIoT relay selection and sub-channel selection. We define three variables $c_{n,s}$, $b_{m,g}$ and $c_{m,s}$ for sub-channel of GB users, gateway node and sub-channel selection for GF users, respectively. We have

$$c_{n,s}(t) = \begin{cases} 1, & \text{if } n \in \mathcal{N} \text{ select sub-channel } s, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

$$b_{m,g}(t) = \begin{cases} 1, & \text{if } m \in \mathcal{M} \text{ select relay node } g, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

$$c_{m,s}(t) = \begin{cases} 1, & \text{if } m \in \mathcal{M} \text{ select sub-channel } s \\ & \text{occupied by GB user } n, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

D. Problem Formulation

Our aim is to maximize the EE of the network while satisfying their QoS requirements thereby optimizing the transmit power for GB and GF users, users clustering and selecting appropriate CIoT gateway selection for GF users. The optimization problem can therefore be expressed as

$$\underset{c_{n,s},c_{m,s},b_{m,g},p_{n,s},p_{m,s}}{\text{maximize}} \quad E \quad (9)$$

$$\text{s.t.} \quad (3) \quad (9a)$$
$$(4) \quad (9b)$$
$$p_{n,s}(t) \le P_{max}, \quad \forall n, \forall t, \quad (9c)$$
$$p_{m,s}(t) \le P_{max}, \quad \forall m, \forall t, \quad (9d)$$
$$\sum_{s=1}^{S} c_{n,s}(t) \le 1, \quad \forall n, \forall t, \quad (9e)$$
$$\sum_{s=1}^{S} c_{m,s}(t) \le 1, \quad \forall m, \forall t, \quad (9f)$$
$$\sum_{g=1}^{G} b_{m,s}(t) \le 1, \quad \forall m, \forall t, \quad (9g)$$
$$\sum_{s=1}^{S} R_{n,s}(t) \ge \tau, \quad \forall n, \forall t, \quad (9h)$$
$$\sum_{s=1}^{S} R_{m,s}(t) \ge \bar{\tau}, \quad \forall m, \forall t, \quad (9i)$$
$$\sum_{m\in\mathcal{M}} b_{m,g}(t) \le Z, \quad \forall t, \quad (9j)$$

where the SIC decoding order of GB and GF users is represented by (9a) and (9b), respectively. (9c) provides the maximum transmit power limit for GB user n, while (9d) mentions the maximum transmit power limit for GF user m. Constraint (9e) restricts the GB users to choose a maximum of one sub-channel per time slot t. Constraints (9f) and (9g) limit the GF users to select at most on the relay node and on the sub-channel. (9h) and (9i) are the required data rates for GF and GB users, respectively. (9j) represents the number of GF users that connect with one CIoT relay.

III. Proposed SAMA-DRL Framework for SGF-NOMA Systems

Power control is a critical aspect in SGF-NOMA for several reasons. As different users may have varying QoS requirements, including data rate, latency, and reliability, power control helps allocate power accordingly among these users. By adjusting power levels dynamically, SGF-NOMA can balance the QoS among users with varying channel conditions, ensuring that users with weaker signals still meet their QoS requirements. Moreover, dealing with GF users with sporadic traffic and limited energy makes things more challenging. We leverage ML techniques to solve this challenge due to their excellent capability in resource allocation in wireless networks. We propose single-agent in integration with MA-DRL to solve this two-fold problem. In particular, we adapt SARL on the BS side to strengthen the coordination with GB users and allocate them resources efficiently and intelligently. Due to the non-involvement of the BS in the resource allocation process and the sporadic nature of GF users, we adapted MARL in the second level to find the optimal resources for GF users and also ensure the QoS requirements of GB users. The proposed SAMA-DRL framework is able to overcome the following problems.

1) Scalability and non-stationarity: Scalability is one of the inherent problems in MARL algorithms. As the number of agents increases, the state and action spaces increase exponentially. Moreover, in a multi-agent environment, all agents learn and interact simultaneously. Since a changing environment depends on the combined actions of all agents rather than an agent's own behaviour. Furthermore, state transitions and rewards are no longer stationary for an agent (agents face moving target problems) [26]. As a result, agents need to adapt to

other agents' changing policies. Therefore, we limit the MARL part to GF users only to handle these issues.

2) Expensive learning: It is expensive to learn using single-agent learning [27]. For example, when a single agent (i.e., the BS) is used to find optimal resources for both GB and GF users, the complexity of the BS increases. Therefore, the BS as an agent is responsible to allocate optimal resources to GB users only.

Next, we modelled the environment as MDP for the proposed SAMA-RL framework. An MDP is composed of a tuple $(\mathcal{N}, \mathcal{S}, A, r)$, where $\mathcal{N}$ represents the number of agents in the environment, $\mathcal{S}, A, r$ representing state space, action space and reward, respectively.

A. MDP for Single Agent Learning

- Agent: The BS works as an agent to identify the best resources for GB users.
- State Space $\mathcal{S}_n$: The BS obtains the channel gain of GB users as a state.
- Action: The agent selects a combination of power levels for GB users as an action. Power levels determine the total number of actions. For N GB users and P power levels, we have a maximum N^P combinations. Let O represent the total number of combinations and each combination corresponds to an action a, i.e., $O = [a_1, a_2, \cdots, a_O]$. At time step t, the selected action is $a(t) = [p_1, p_2 \cdots, p_N]$.
- Reward: The agent gets GB users' EE as a reward signal, if the agent does not violate the maximum power constraints, otherwise, the agent will get a reward of zero.

B. MDP for Multi-Agent Learning

- Agents: We represent the GF users as agents interacting with the environment.
- State Space: GF users receive their data rates as a state.
- Action: The action for GF users consists of sub-channel selection and transmit power, $a_m = \{p_m, c_m\}$.
- Reward: Reward assignment methods in multi-agent reinforcement learning can be categorized into two distinct approaches: global and local rewards. The global reward approach uniformly allocates a single global reward to all agents, irrespective of their individual contributions. Consequently, this may result in lazy agents receiving disproportionately higher rewards relative to their actual contributions, leading to lack of motivation for optimizing their policies. Conversely, diligent agents may receive lower rewards, even when their actions are beneficial, due to the negative impact of lazy agents on the overall system, causing confusion regarding the optimal policy. In contrast, the local reward approach assigns distinct rewards to each agent based on their individual behavior, thereby discouraging laziness. However, this approach may fail to provide rational incentives for agents to collaborate, potentially leading to the development of selfish and greedy behaviors. Therefore, in our designed reward function, reward of each agent is proportional to its contribution to the total energy efficiency. Thus, the reward for the i^{th} agent can be calculated as:

$$R_i = R_{\text{total}} \times \frac{EE_i}{\sum_{j=1}^{N} EE_j},$$

where R_{total} is the total energy efficiency of the system and EE_i is the EE of i-th agent.

Algorithm 1 SAMA-DRL Based SGF-NOMA Algorithm without KNN

Parameter Setup Phase:
Setup the parameters for the single agent (BS) and multi-agents
Set replay memory for SA and MAs
Initialize Q-network weights and set target weights as primary Q-network (for both SA and MAs)
Training Phase:
for Episode e = 1 to E **do**
  Environment reset
  for Time step t = 1 to T **do**
    Single agent (BS):
    Input state $s_n(t)$
    Choose action $a_n(t)$ using ϵ-greedy policy
    Obtain a new state $s_n(t+1)$ as well as reward $r_n(t)$
    Save the experience in replay memory
    Multi-agents (GF users):
    for each GF agent m **do**
      Input state $s_m(t)$
      Select action $a_m(t)$
    end for
    All agents perform joint actions obtain new state $s(t+1)$ and reward $r(t)$
    for Each IoT agent m **do**
      Save a tuple of $s_m(t), a_m(t), r(t), s_m(t+1)$ to replay memory
    end for
    Single agent (BS) and Multi-agents:
    Select batches equally from memory D
    Using (13), reduce loss between the the primary network and target network using a stochastic gradient
    if e $\%== T_u$ **then**
      Update target Q-network weights
    end if
  end for
end for

To learn optimal action policies for a given environment, an agent receives a state $s(t)$ and chooses action $a(t)$ from the action space, following a policy π. A policy $\pi\big(a(t)|s(t)\big)$ is the mapping from state $s(t)$ to action $a(t)$. The agent receives the next state $s(t+1)$ and reward $r(t)$ for the action taken in the previous time step t. Agent form an experience $e(t+1) = \big(s(t), a(t), s(t+1), r(t)\big)$. The goal of each agent is to find an optimal policy π^* and maximize the discounted long-term reward defined as $R(t) = \sum_{t=1}^{\infty} \gamma^{t-1} r(t)$, where $\gamma \in [0, 1]$ is the discount factor and reflects the importance of future reward as compared to the immediate reward. The classical Q-learning algorithm is based on an action-value function (Q function) to locate the optimal policy π^*. An action-value-function is defined as the expected return after taking action $a(t)$ in a given state $s(t)$, we have

$$Q_\pi\big(s(t), a(t)\big) = \mathbb{E}_\pi\big[R(t)\big|s(t) = s, a(t) = a\big]. \tag{10}$$

The achievable maximum and optimal action-value function by a policy π for a given state $s(t)$ and action $a(t)$ can be expressed by the Bellman equation as below,

$$Q^*_{\pi^*}\big(s(t), a(t)\big) = \mathbb{E}_{s(t+1)}\big[r + \gamma \max_{a(t+1)} Q^*\big(s(t+1), a(t+1)\big)|s(t), a(t)\big]. \tag{11}$$

The DRL is the extended version of classical RL where the Q function is approximated by a deep neural network with

Algorithm 2 SAMA-DRL Based SGF-NOMA Algorithm with KNN

Input:
Location or distances of users $D = [d_1, d_2, \cdots, d_m]$, $m \in \mathcal{M}$
Desired number of gateway nodes G
Output:
A set of G clusters, each having a gateway node g
Steps:
Select G data items from D randomly as initial centroids
Repeat
Allocate each user to its closest centroid
For each cluster, compute the new mean
Until convergence criteria met
Place the relay nodes on final centroids
Repeat lines 1-30 of Algorithm 1

TABLE III: Network and Training Parameters

Path loss (α)	3
AWGN(w_0)	-174 dBm
P_{max}	1 W
Required data rate for GB users	10 bps/Hz
Sub-channel bandwidth	10 KHz [18]
Training episodes	400
No. of neurons in each layer	{500, 300, 100}
Discount factor γ	0.9
Update target frequency	1000
Learning rate	0.001

weights θ, known as a deep Q network (DQN). Because the classical Q-learning becomes expensive as the size of state and action spaces increases. Therefore, keeping a large Q table for state action pairs, DQN only memorizes the weights θ of the neural network minimizes computation and memory complexity. The DQN of an agent consists of the primary Q network and target network that produces the actual Q value and target Q value, respectively, to calculate the loss between the actual and predicted Q values. By employing an ϵ-greedy strategy, the agent(s) explore the environment by taking a random action with ϵ , while with a probability of $1-\epsilon$, they opt for the action with the highest Q value. During training, the Q network's weights are updated by selecting a random mini-batch of data from the experience replay. The target value produced by the target Q network can be represented as

$$y(t) = r(t) + \gamma \underset{a(t+1)\in A}{\operatorname{argmax}} \; Q\big(s(t+1), a(t+1); \bar{\theta}\big), \qquad (12)$$

where $\bar{\theta}$ represents the weights parameters of the target Q network which is replaced with the weights θ of the primary Q network after a fixed number of training steps. To train the primary Q network, the loss between the target network values and the primary network Q values can be minimized using

$$L(\theta) = \big(y(t) - Q(t)(s(t), a(t); \theta)\big)^2. \qquad (13)$$

C. Baseline Algorithm

The process of the benchmark algorithm which we considered as a baseline for our proposed work is given in **Algorithm-1**. We defined the training parameters for both levels, i.e., for the single agent (BS) and multi-agents (GF users). First, the single agent receives the state and selects action based on ϵ-greedy policy. After that, all the GF users select actions i.e., transmit power and sub-channel. Both the single agent and multi-agents perform the actions and receive the next state and corresponding reward from the environment. All the agents save the current state, the next state and the reward to their memory. Agents sample random batches from the replay memory and minimize the loss between primary Q values and target Q values using stochastic gradient descent in order to train the primary network. The primary network weights are transferred to the target network weights once a set number of training episodes completed.

TABLE IV: Benchmark schemes and proposed scheme architecture

Scheme	Relay Node	KNN applied	GB user power optimization
Benchmark 1	No	No	No
Benchmark 2	No	No	Yes
Benchmark 3	Yes	No	Yes
Proposed	Yes	Yes	Yes

D. Proposed Algorithm

To minimize the energy consumption of GF users, we placed relay nodes in the cell area using the K-means clustering algorithm, a process given in **Algorithm-2**. We set the value of G and initialize these as centroids and keep iterating until there is no change to the centroids. Placed the relay node on the final centroid. Finally, assign each GF user to the closest relay node. After that, the training phase for GF users begins and they transmit their data to the corresponding relay node.

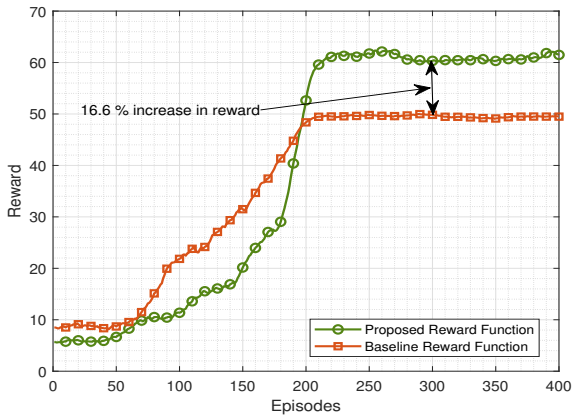
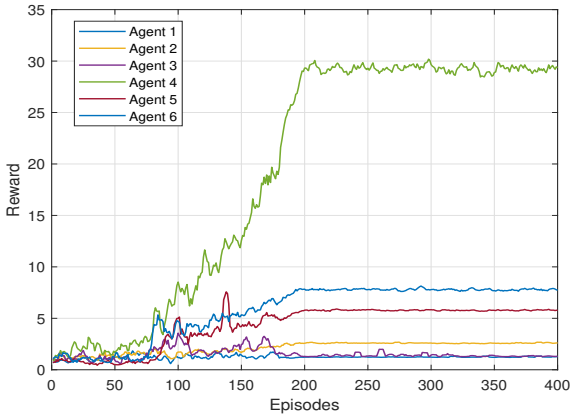


Fig. 3: Reward Comparison

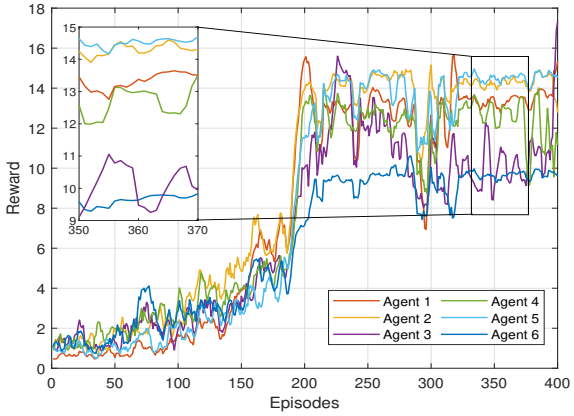
IV. Numerical Results

In this section, we present the numerical results from simulations conducted with the network settings and training hyperparameters configurations shown in Table III. Through simulation results, these parameters are used to evaluate how well the proposed algorithms perform. Analyzing the proposed algorithm's performance in comparison to other algorithms that are currently in the literature, we use the following benchmark schemes, summarized in Table IV:

- **SGF-NOMA without relay node and GB users transmit power optimization:** In this network scenario, GF and GB users transmit their data to the BS directly. Moreover, GB users sends data with a fixed power allocation. Most of the existing works adopted this architecture.
- **SGF-NOMA scheme without relay node and with GB users transmit power optimization:** In this baseline scheme, the transmit power of both types of users is optimized without having a relay node, given in **Algorithm-1**.



(a) Baseline



(b) Proposed

Fig. 4: Individual reward Comparison

- **SGF-NOMA scheme with relay node without KNN and with GB users transmit power optimization:** A relay node is used but at a fixed location, i.e., KNN is not used for finding an optimal location for relay node.

A. Reward Comparison

In this section, we compared the reward obtained by all agents and their contribution to the global reward.

- Combine Reward Comparison: Choosing the appropriate reward function is essential for optimizing the objective function, especially in a multi-agent system. To assess the performance of the agents using our proposed reward function, we compared it with a standard reward function in which all agents obtained an equal reward. The results depicted in Fig. 3 show that our agents achieved a higher reward compared to those using the baseline reward function. This is because in our proposed reward function, each agent receives a distinct reward based on their individual contribution to the overall reward value. On the other hand, when using a uniform reward function, all agents receive a combined and same reward that encourages laziness and discourages efficient and effective exploration of the environment.
- Individual Reward Comparison: Fig. 4 illustrates the individual reward obtained by each agent and its contribution to the objective function. There is a significant disparity in the rewards obtained by each agent in the baseline reward function, given in Fig. 4(a). Only one agent (Agent-4) contributed more to the total reward, while the rest of the agents performed poorly. When agents receive a combined reward without accounting for their individual contributions, they may become lazy and fail to actively explore the environment for optimal states and actions. In contrast, agents in our proposed reward function made significant contributions to the overall system objective, as shown in Fig. 4(b). Moreover, in a system where agents share the same resources, it is necessary to distribute the resources. Therefore, in our proposed reward function, all users in the network can equitably share the limited resources to maintain the fairness.

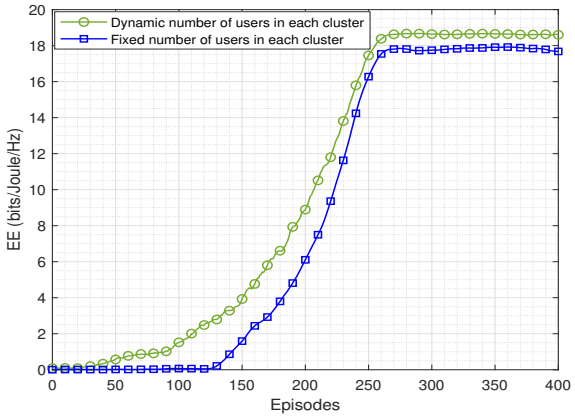


Fig. 5: Performance with fixed and dynamic users in each cluster.

B. Impact of Users in Each Cluster

Fig. 5 presents a comparison of our proposed approach to both fixed and dynamic clustering. In fixed clustering, the number of users per RB is fixed, whereas in dynamic clustering, users select the RB and transmission power based on the DRL algorithm. The second case, dynamic clustering with DRL, achieved the best results in terms of EE. By allowing the number of users per RB to vary, the DRL algorithm utilizes the spectrum more efficiently. The formation of clusters is based on current network conditions, resulting in optimal load distribution across the RBs.

C. EE Performance Comparison

The proposed algorithm demonstrates a remarkable improvement in energy efficiency relative to the baseline algorithms, as depicted in Fig. 6. This improvement can be attributed to the strategic placement of relay nodes, optimized transmit power, and efficient user clustering for both GF and GB users. The implementation of the K-Nearest Neighbors algorithm facilitates the optimal positioning of relay nodes, allowing GF users to transmit data over shorter distances at lower power levels. Additionally, the BS, acting as an agent, fine-tunes the power of GB users to attain optimal EE. The GF users also play a critical role as agents in determining

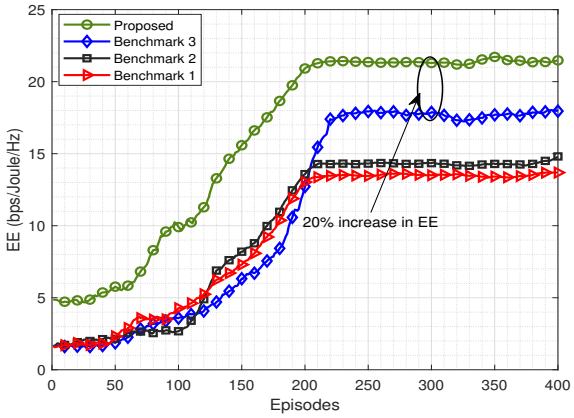


Fig. 6: EE comparison of the proposed scheme with benchmark schemes.

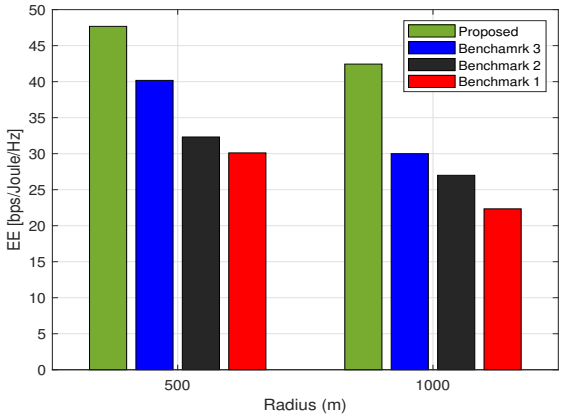


Fig. 7: Comparison of EE for different radii.

the optimal power settings and forming NOMA clusters to minimize interference. In contrast, Benchmark 3 lacks the KNN-based placement of relay nodes. Benchmark 2 omits the use of relay nodes entirely, resulting in direct transmissions to the central BS by GF users. Furthermore, Benchmark 1 not only excludes relay nodes but also maintains a constant transmit power for GB users, resulting in the least energy-efficient algorithm among the compared methods.

D. Performance with Different Radii

Fig. 7 demonstrates that our proposed NOMA algorithm outperforms conventional benchmarks in terms of energy efficiency across different cell radii. This effectiveness is attributed to the strategic placement of relay nodes and optimized transmit power for both GB and GF users. Notably, the EE advantage is maintained even as the cell radius increases, which typically leads to degraded performance because of increased path loss effects. The proposed scheme's resilience is partly due to the reduced transmission distance for GF users, which mitigates the impact of path loss and thus improves EE. In contrast, Benchmark 1 exhibits the lowest efficiency because it depends on direct transmissions to the BS by GF users and fixed-power transmissions by GB users, without taking into account channel conditions. Our proposed method offers a more practical approach for real-world IoT applications. It minimizes transmission distances and adjusts power levels based on individual user channel gains, in contrast to traditional methods that assume long-distance direct transmission to the central BS. It is worth noting that IoT devices have limited processing capabilities and cannot transmit over long distances. Therefore, our proposed scheme is more practical compared to existing methods, as it eliminates the need for users to transmit their data directly to the central BS.

E. Impact of the Proposed QoS-Based SIC Order

Fig. 8 shows a comparison of the performance of different SIC ordering strategies for the given CIoT setup. The battery level of the user in both SIC ordering initially decreases drastically due to the exploration phenomena. Since the users act as agents and initially choose random actions (power levels), they may opt for high transmit power levels for data transmission.

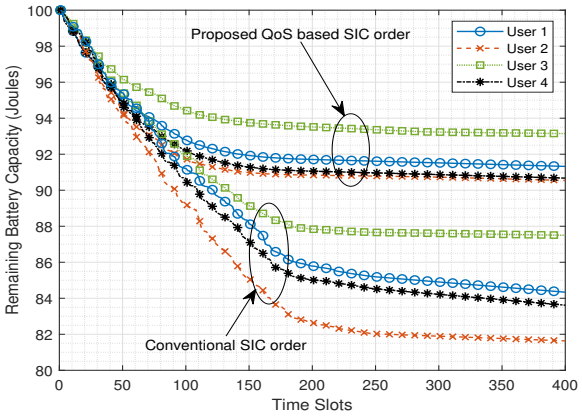


Fig. 8: Comparison of battery life of users.

In the proposed QoS-based SIC scenario, the user with the highest battery level is given priority in decoding, which is a strategy that prioritizes device longevity and energy efficiency. It can be observed that the proposed QoS-based SIC decoding order outperforms the conventional SIC decoding order. The conventional SIC decoding order prioritizes users based on the highest received power level, while in our proposed SIC order, users with the lowest battery level are decoded in the final stage of SIC, thereby avoiding interference from other users and achieving the required QoS with the lowest possible transmit power, which enhances the lifespan of users with low battery levels. However, this ordering results in a slight delay due to the SIC decoding order. In the conventional SIC decoding order, users with the highest received power level are decoded first, regardless of their battery level. The users decoding in the first stage of SIC face interference from other users in the same NOMA cluster, and therefore, to achieve the required QoS, they must transmit with a high power level. Consequently, the conventional SIC ordering is the least energy-efficient, which can degrade the network lifetime.

F. System Performance with Increasing Eavesdroppers

Fig. 9 compares the EE of the proposed algorithm with benchmark schemes with increasing number of eavesdroppers. it is evident that the proposed algorithm outperforms the

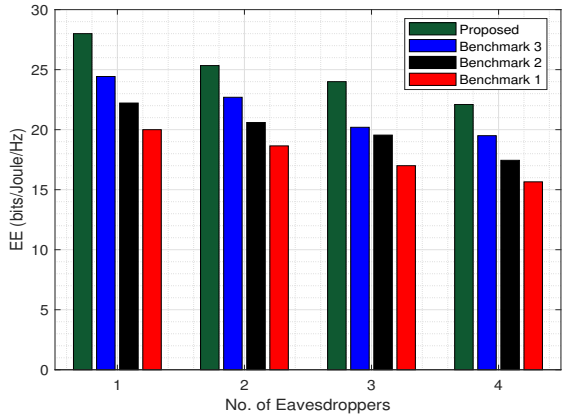


Fig. 9: Impact of increasing No. of eavesdroppers.

benchmark schemes in every case. To ensure that the intended recipient receives a stronger signal than eavesdroppers, users may need to increase their transmission power. Moreover, multiple eavesdroppers can increase interference in the communication channels, which may necessitate legitimate users to retransmit data or boost their transmission power to overcome interference. The EE of all schemes falls as the number of eavesdroppers rises for the reasons mentioned above.

V. Conclusions and Future Research Directions

In this article, we have investigated the problem of energy-efficient resource allocation in the SGF-NOMA based CIoT network in the presence of untrusted users. We minimize the overall energy consumption by jointly optimizing the resource selection decisions, transmit power, subchannel assignment, and relay node selection using SAMA-DRL. We take into account the appropriate position of the relay node and use KNN for this purpose to enhance coverage for GF users. We have employed collaborative contribution reward function to avoid agents' laziness and utilized QoS based SIC decoding order. Our study compares the performance of our proposed SAMA-DRL based SGF-NOMA CIoT network with various baseline algorithms. Simulation results demonstrate that our approach using SAMA-DRL has enhanced rewards and outperformed baseline algorithms in terms of secrecy EE across different network parameters. We will use lifelong learning to further improve the performance of SGF-NOMA CIoT networks with multiple antenna system in our future work.

References

[1] G. Dhiman and N. S. Alghamdi, "SMoSE: Artificial intelligence-based smart city framework using multi-objective and IoT approach for consumer electronics application," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 3848–3855, 2024.

[2] D. Minoli, K. Sohraby, and B. Occhiogrosso, "IoT considerations, requirements, and architectures for smart buildings—energy optimization and next-generation building management systems," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 269–283, 2017.

[3] L. Chettri and R. Bera, "A comprehensive survey on internet of things IoT toward 5G wireless systems," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 16–32, 2019.

[4] C. K. Wu, C.-T. Cheng, Y. Uwate, G. Chen, S. Mumtaz, and K. F. Tsang, "State-of-the-art and research opportunities for next-generation consumer electronics," *IEEE Trans. Consum. Electron.*, vol. 69, no. 4, pp. 937–948, 2023.

[5] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, 2015.

[6] 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, June 2019.

[7] M. B. Shahab, S. J. Johnson, M. Shirvanimoghaddam, and M. Dohler, "Enabling transmission status detection in grant-free power domain non-orthogonal multiple access for massive internet of things," *Trans. Emerging Telecommun. Technol.*, p. 4565–4591, 2022.

[8] Z. Yang, P. Xu, J. Ahmed 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, Oct. 2020.

[9] 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, Apr. 2021.

[10] C. Zhang, Y. Liu, and Z. Ding, "Semi-grant-free NOMA: A stochastic geometry model," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

[11] X. Bai and X. Gu, "NOMA assisted semi-grant-free scheme for scheduling multiple grant-free users," *IEEE Sys. Journal*, vol. 17, no. 2, pp. 3294–3305, 2023.

[12] H. Lei, F. Yang, H. Liu, I. S. Ansari, K. J. Kim, and T. A. Tsiftsis, "On secure NOMA-aided semi-grant-free systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 74–90, 2024.

[13] F. Fang, Y. Xu, Q.-V. Pham, and Z. Ding, "Energy-efficient design of IRS-NOMA networks," *IEEE Trans. Veh. Tech.*, vol. 69, no. 11, pp. 14 088–14 092, 2020.

[14] M. Zeng, N.-P. Nguyen, O. A. Dobre, Z. Ding, and H. V. Poor, "Spectral- and energy-efficient resource allocation for multi-carrier uplink NOMA systems," *IEEE Trans. Veh. Tech.*, vol. 68, no. 9, pp. 9293–9296, 2019.

[15] X. Wang, Y. Zhang, R. Shen, Y. Xu, and F.-C. Zheng, "DRL-based energy-efficient resource allocation frameworks for uplink NOMA systems," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7279–7294, 2020.

[16] M. Shirvanimoghaddam, M. Condoluci, M. Dohler, and S. J. Johnson, "On the fundamental limits of random non-orthogonal multiple access in cellular massive IoT," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2238–2252, 2017.

[17] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "A novel analytical framework for massive grant-free NOMA," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2436–2449, 2018.

[18] J. Zhang, X. Tao, H. Wu, N. Zhang, and X. Zhang, "Deep reinforcement learning for throughput improvement of the uplink grant-free NOMA system," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6369–6379, 2020.

[19] Y. Xu, J. Wang, Q. Wu, J. Zheng, L. Shen, and A. Anpalagan, "Dynamic spectrum access in time-varying environment: Distributed learning beyond expectation optimization," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5305–5318, 2017.

[20] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 2, pp. 257–265, 2018.

[21] 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.

[22] D.-D. Tran, S. K. Sharma, V. N. Ha, S. Chatzinotas, and I. Woungang, "Multi-agent DRL approach for energy-efficient resource allocation in URLLC-enabled grant-free NOMA systems," *IEEE Open Journal Commun. Society*, vol. 4, pp. 1470–1486, 2023.

[23] Z. Ding, R. Schober, and H. V. Poor, "A new QoS-guarantee strategy for NOMA assisted semi-grant-free transmission," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7489–7503, 2021.

[24] Y. Guo, M. Lin, H. Kong, M. Cheng, and W.-P. Zhu, "NOMA assisted semi-grant-free transmission scheme in satellite systems," *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 2122–2126, 2023.

[25] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[26] A. Wong, T. Bäck, A. V. Kononova, and A. Plaat, "Deep multiagent reinforcement learning: Challenges and directions," *Artificial Intelligence Review*, pp. 1–34, 2022.

[27] D. Lee, N. He, P. Kamalaruban, and V. Cevher, "Optimization for reinforcement learning: From a single agent to cooperative agents," *IEEE Signal Processing Mag.*, vol. 37, no. 3, pp. 123–135, 2020.

Sohail Abbas received PhD degree in wireless network security from Liverpool John Moores University, UK in 2011. Currently, he is working as an Associate Professor in the Department of Computer Science, College of Computing and Informatics, University of Sharjah, UAE. He has been involved in academia for more than 20 years and in research for more than 16 years.

His research interests are focused on security issues including intrusion detection, identity-based attacks, and trust in wireless networks, such as mobile ad hoc networks, wireless sensor networks, and the Internet of Things. Dr. Sohail is a member of various technical program committees, including IEEE CCNC, IEEE VTC, IEEE ISCI, IEEE ISWTA, etc. He is also serving various prestigious journals as a reviewer, such as Security and Communication Networks, IET Wireless Sensor Systems, Mobile Networks and Applications, International Journal of Electronics and Communications, International Journal of Distributed Sensor Networks.

Muhammad Fayaz (S'20-M'23) received his Ph.D. degree in computer science from Queen Mary University of London, U.K., in 2023. He is currently a lecturer in the department of Computer Science and IT, University of Malakand, Pakistan.

His research interests include Artificial Intelligence for Wireless Systems, Beyond 5G Wireless Networks and Internet of Things (IoT).

Abdulrahman Ghandoura received the Ph.D. degrees in Electrical and Computer Engineering from Southern Illinois University, Carbondale, IL USA in 2018. He joined Umm-Alqura University, Mecca, KSA, where he is currently an Assistant Professor with the Department of Engineering and Science, Applied Collage. His research interests include Network Architecture and Wireless Sensor Networks, 6G and beyond wireless communication technologies, Internet of Things (IoT) also Internet of Everything (EoT) Applications.

Muhammad Zahid Khan received the B.C.S. degree (Hons.) in computer science from the University of Peshawar and the Ph.D. degree from the School of Computing and Mathematical Sciences, Liverpool John Moores University, U.K., in 2013. He is currently an Assistant Professor with the Department of Computer Science and Information Technology, University of Malakand (UoM), Khyber Pakhtunkhwa, Pakistan. He is also leading the Network Systems and Security Research Group, Department of Computer Science and Information Technology, UoM. He has over 18 years of teaching experience and a good track record of research and publications in international academic journals.

His current research interests include WSNs, WBAN, VANET, the IoT, network security, and cyber security. He is a Higher Education Commission's Pakistan Approved Supervisor.

Ateeq ur Rehman (Member, IEEE) is currently working as a Lecturer at the Department of Computing, Staffordshire University, UK. He received his Ph.D. degree from the University of Southampton, UK in 2017. His area of research includes cyber security, blockchain, and privacy-preserved machine learning, particularly in healthcare and smart cities.