IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 May 2024, accepted 14 June 2024, date of publication 20 June 2024, date of current version 28 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3417320

RESEARCH ARTICLE

Joint User Association and Pairing in Multi-UAV-Assisted NOMA Networks: A Decaying-Epsilon Thompson Sampling Framework

BONIFACE UWIZEYIMANA[1], (Member, IEEE), MOHAMMED ABO-ZAHHAD[1,2], (Senior Member, IEEE), OSAMU MUTA[3], (Member, IEEE), AHMED H. ABD EL-MALEK[1], (Member, IEEE), AND MAHA M. ELSABROUTY[1], (Senior Member, IEEE)

[1]Electronics and Communications Engineering, Egypt-Japan University of Science and Technology, New Borg El-Arab City, Alexandria 21934, Egypt
[2]Department of Electrical Engineering, Faculty of Engineering, Assiut University, Asyut 71515, Egypt
[3]Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 812-0053, Japan

Corresponding author: Boniface Uwizeyimana (boniface.uwizeyimana@ejust.edu.eg)

This work was supported in part by the TICAD7 African Scholarship, in part by Egypt-Japan University of Science and Technology (E-JUST) and Egypt Ministry of Higher Education, and in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 24K07490.

ABSTRACT Unmanned aerial vehicles (UAVs) are expected to be integrated into future wireless networks to offer services, especially in unreachable or congested areas. To improve the spectral efficiency, non-orthogonal multiple access (NOMA) scheme can be utilised within the UAV communication to allow more users to be covered and associated. The performance of the NOMA-UAVs network is governed by several factors including power allocation, user association and pairing methods. This paper presents an approach that uses multi-armed bandit (MAB) and two-sided matching frameworks to maximize the throughput of multi-UAV-assisted NOMA networks in a decentralized manner. The approach enables the UAVs to propose to the ground users (GUs) without explicit cooperation among the UAVs while the GUs can accept or reject the proposals. To this end, we propose a modified Thompson sampling algorithm that we named decaying epsilon Thompson sampling (DϵTS) MAB algorithm that is designed to improve the exploration-exploitation tradeoff in the MAB. The performance of the proposed DϵTS MAB algorithm is evaluated against other existing MAB techniques. Simulation results show that the DϵTS algorithm attains faster convergence and improved performance in terms of smaller regret and increased achievable system throughput. The DϵTS MAB algorithm particularly excels in regards of the convergence rate when the number of available action spaces increases.

INDEX TERMS Multi-UAV networks, user association and pairing, multi-armed bandit, decaying-epsilon, Thompson sampling, NOMA, throughput maximization, two-sided matching.

I. INTRODUCTION

The future wireless technologies to be integrated in 6G and beyond networks, will be driven by real-world applications such as telemedicine, smart cities, the Internet of Things (IoT), virtual reality, etc. These niche applications demand high-speed and uninterruptible network services [1], [2]. Learning-based techniques have been increasingly deployed to improve the performance in IoT networks. For example, the work in [3] applies reinforcement learning to maximise throughput of wireless-powered access points.

This is a clear evidence that the technologies to be integrated in the future networks will be those that are optimised for the best performance for real-life applications and can tackle the inherent deployment challenges and changing environment. One major challenge that is encountered with the plethora of IoT applications, is the unprecedented increase in the number of users. This massive number of

The associate editor coordinating the review of this manuscript and approving it for publication was Pietro Savazzi.

users coupled with limited coverage and scarce resources can pose a challenge for future quality of service (QoS) demanding applications. The standard terrestrial networks supported by fixed base stations stand short to meet all these requirements. Consequently, the integration of non-terrestrial networks is intended to enhance the QoS, increase the number of supported devices, and expand the coverage by embedding unmanned aerial vehicles (UAVs), high-altitude platforms (HAPs), or satellite networks [4], [5], [6]. The UAVs which can serve as an aerial base station (ABS) have been seen to be one of the ways to provide network services, especially in areas that are overcrowded and difficult to access [7], [8]. Much attention is required in the deployment of the UAVs depending on the target users and their requirements, energy consumption, and available resources [9], [10], [11]. To improve spectrum efficiency, QoS, and performance in the UAVs' network, non-orthogonal multiple access (NOMA) technology which allows multiple users to access the same resource block, can be applied. More specifically, the network performance is influenced by the NOMA design methods for user association and pairing especially the achievable rate of the system [12]. Traditional user association methods are highly based on matching theory that relies on matching two sides with well-known preference lists [13], [14], [15]. However, it is not always possible for either party to know its preference list. Further studies are needed to improve the throughput of the UAV networks. Integrating the multi-armed bandit (MAB) with matching theory could facilitate the agents that do not have prior preference lists to learn through exploration and exploitation and this can improve the system's capacity [16], [17], [18].

This paper presents the MAB and two-sided matching methods for efficient user association and pairing in multi-UAV-assisted NOMA networks. The agents (UAVs) that do not know their preference lists adaptively learn them by balancing exploration and exploitation trade-offs. Through the two-sided matching, the UAVs propose to the ground users (GUs) using the learned preference lists from MAB and the GUs accept or reject the proposals depending on their channel conditions.

A. RELATED WORKS

Several studies have analyzed the performance of a single UAV serving as a relay node or an ABS [19], [20], [21], [22]. In most cases, the user association design methods affect the performance of the UAV's networks. The work in [23] considered the UAV as a relay node to forward the signal to the ground users by employing the NOMA scheme. Similarly, the outage probability of the UAV's network as ABS was analysed for both NOMA and orthogonal multiple access (OMA) schemes [24]. This study found that the NOMA scheme achieves better performance than OMA. Employing NOMA in multi-UAV networks could indeed improve not only the performance of the system but also spectrum efficiency.

In [22], they proposed a heuristic algorithm for scheduling users to the available subchannels of the UAV networks based on matching and swapping theories. The overall goal was to optimize energy efficiency through optimal resource allocation and trajectory of the UAV. However, this study was done over a single UAV. User scheduling in multiple UAVs is done in [11], which was based on two-sided matching and users are served in the NOMA scheme. However, this work considers the case that the two sides know their preference lists. The case of an unknown preference list of either party was done for example, applicants and colleges, workers and companies in [17], [18], [25], and [26], but its application in wireless networks is not fully studied.

Motivated by these works and the desire to maximize the system throughput of the multi-UAV networks, we propose the MAB and two-sided matching frameworks for user association and pairing whereby the UAVs propose to the GUs aiming to identify the potential ones and the GUs can accept or reject the proposals. In our earlier work in [27], prior analysis and the algorithm for user pairing in multi-UAV networks that apply NOMA was developed. The algorithm was based on the upper confidence bound (UCB) MAB technique and two-sided matching in which the UAVs act as the agents while the users are the actions to be selected. However, there is still a need for extensive study to maximize the system throughput of multi-UAV networks in case all the UAVs start with an unknown prior reward distribution of the actions. This is especially necessary when there are many actions in the network environment. In addition, the algorithm proposed in [27] assumes that once a user accepts the proposal from a certain UAV, it is obligated to retain it. These limitations prompted further research to maximize the throughput of multi-UAV networks.

This current work adopts the system model in [27], and aims to maximize the overall throughput of multi-UAV networks. Every GU can switch to a new proposing UAV if it is highly ranked in its preference list and reject the previous one. In addition, the work in [27] applies the UCB, which uses the deterministic approach to balance exploration and exploitation trade-offs, however, its performance is affected by the number of actions available in the system [28], and it is mostly applied with an initial complete scan of the actions which can be difficult when there are many actions [16], [21], [29], [30].

On the other hand, a more efficient method is presented in the literature known as Thompson sampling (TS). The TS is a probabilistic approach that is particularly suitable for complex and unknown reward distributions within a changing environment [28]. Therefore, in this work, we take advantage of TS to develop the decaying epsilon Thompson sampling (DϵTS) framework.

In [31], the authors proposed the ϵ-exploring Thompson sampling (ϵ-TS), a modified version of the TS algorithm. In the ϵ-TS, arms are exploited greedily based on empirical mean rewards with probability 1-ϵ and based on posterior

samples obtained from TS with probability ϵ (exploration). The ϵ-TS improves computational efficiency compared to standard TS while achieving better regret bounds.

This inspired us to develop the DϵTS algorithm. The DϵTS starts with a relatively large value of ϵ allowing more exploration, which in effect allows faster building and collection of initial beliefs about the actions, but this rate is decreasingly updated to exploit the actions with the highest expected rewards from the learnt information. This improves the achieved rewards by the agents and enables convergence to the best rewards faster than the other MAB algorithms. We integrate the proposed DϵTS within our proposed NOMA-UAV pairing and association framework to improve the performance, especially, when the number of users increases.

B. CONTRIBUTION AND STRUCTURE

The summary of the major contributions of the paper is as follows:

- We investigate a realistic multi-UAV network with limited quotas, multi-ground user pairing and association problem. We consider a realistic aerial-ground channel model with a path-loss model that considers both line and non-line-of-sight probabilities. The interaction between UAVs and GUs aims to maximize the overall system throughput formulated in the optimization problem while considering the GUs' preference as well as giving the GUs the right to switch the association and pairing when a more opportunistic offer from a different UAV is present.
- To solve the optimization problem, we propose an DϵTS MAB algorithm which is based on standard Thompson sampling. However, the standard TS is modified to work with an exploration rate ϵ which gradually decreases as information becomes available. The algorithm allows the agents to select actions with the highest expected rewards in the remaining (1)-ϵ) rate. At every time step, the UAVs make preference lists based on the learned information and enter a two-sided matching game with GUs. Moreover, we consider practical scenarios where the agents start with unknown reward distributions of the available actions in the system and learn through continuous exploration and exploitation.
- To integrate the GUs' preferences, we formulate a two-sided matching game to perform assignments of users to the available UAVs in the network that takes into account the dynamic nature between the UAVs and GUs. At the UAV side, this dynamic nature is addressed by including a learning framework to gather information at every time step aiming to maximize the long-term rewards of the UAV. Therefore, the UAVs build up their preference lists based on the MAB and then propose to the GUs. On the other hand, the GUs can accept or reject the proposal according to their preference lists built based on the channel conditions. Enabling the GUs to reject the previous UAVs if a better UAV is proposing improves the system's achievable rate since the GUs can be matched with their potential best UAVs.
- We evaluate the performance of the proposed DϵTS MAB algorithm against other standard MAB techniques in terms of cumulative regret and achievable rate. The evaluation is done under different conditions and the results obtained from simulation are provided with a detailed discussion. The utilisation of the DϵTS provides improved results especially when the number of GUs, i.e action space increases. It provides an improved convergence time as well as the best reward among other MAB techniques.

TABLE 1. Symbols and notations.

Symbol	Description
$\mathcal{L}$	Set of UAVs.
$\mathcal{M}$	Set of GUs.
q_l	Number of users accepted by UAV l.
$\mathcal{Q}$	Set of quotas of all the UAVs.
$d_{m,l}$	UAV-GU Euclidean distance.
$PL_{m,l}$	UAV-GU propagation path-loss in dB.
$\Phi_{m,l}$	UAV-GU propagation path-loss in linear scale.
$h_{m,l}$	Aerial-ground channel gain coefficient.
γ	Path-loss exponent.
$K_{m,l}$	Rician factor of UAV l and GU m.
P_{LOS}	Power of line-of-sight.
$\theta_{m,l}$	Elevation angle between UAV and GU.
ρ	Transmit signal-to-noise ratio.
$\Omega_{1,q,l}$	Power allocation coefficients of far GU.
$\Omega_{2,q,l}$	Power allocation coefficients of near GU.
$\mathbf{A}_{l,\text{space}}$	Set of all actions in the network.
$\delta_{1,p,l}$	Instantaneous SINR of far GU on pair p of UAV l.
$R^{\text{NOMA}}_{1,p,l}$	Instantaneous rate of far GU on pair p of UAV l.
$\delta_{2,p,l}$	Instantaneous SINR of near GU on pair p of UAV l.
$R^{\text{NOMA}}_{2,p,l}$	Instantaneous rate of near GU on pair p of UAV l.
R^{NOMA}	Overall system throughput of NOMA scheme.
R^{OMA}	Overall system throughput of OMA scheme.
$\mathcal{CN}$	Complex Gaussian distribution.
$\mathcal{NIG}$	Normal-inverse-gamma distribution.
$\mathcal{IG}$	Inverse-gamma distribution.

The structure of this paper is organized as follows. Section II describes the system and aerial channel models. The optimization problem formulation is provided in section III and Section IV details the proposed joint user association and pairing algorithm based on DϵTS MAB and two-sided matching. Section V provides simulation setups, Section VI provides results and discussions, and Section VII concludes the paper.

II. SYSTEM AND CHANNEL MODELS

Table 1 provides a summary of symbols and notations used in this paper to facilitate readability.

A. SYSTEM MODEL

We consider a system of L UAVs acting as ABSs with resource blocks obtained from the backhaul link of the nearby

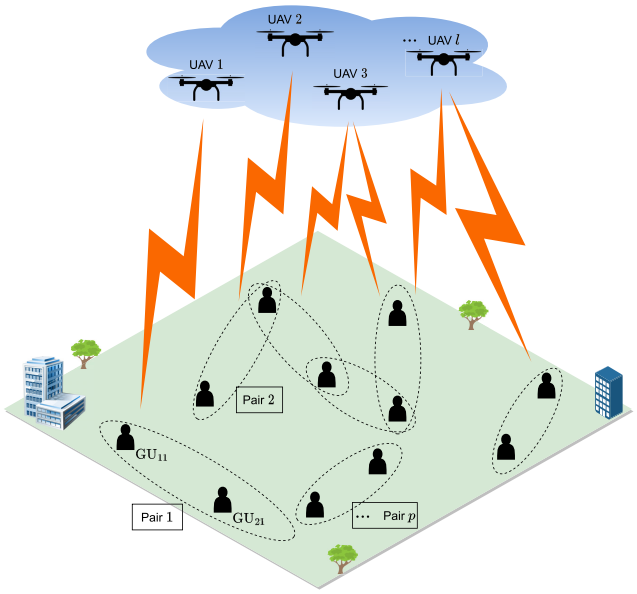


FIGURE 1. The multi-UAV's system model considered in this work.

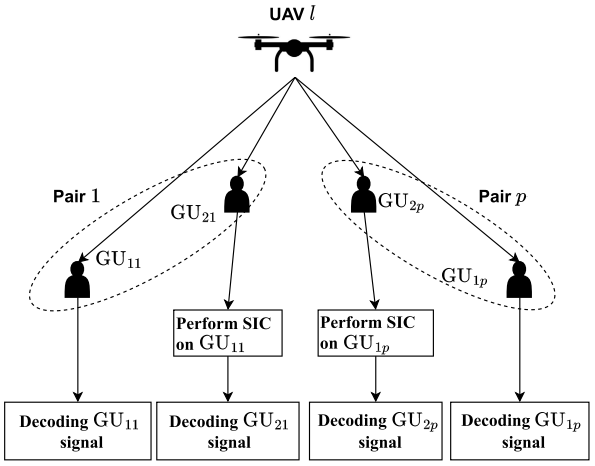


FIGURE 2. Illustration of NOMA principle for the considered system model.

base station and M GUs in a downlink direction as shown in Figure 1. The set of UAVs and GUs are respectively given by $\mathcal{L} = \{1, 2, ..., l, ..., L\}$ and $\mathcal{M} = \{1, 2, ..., m, ..., M\}$. During association, the UAVs seek to identify potential GUs aiming to maximize their achievable throughput. The 3D coordinate of each UAV is given by (x_l, y_l, z_l) where x_l and y_l are the positions at the ground and z_l is the vertical altitude that remains fixed for all UAVs. To reduce the amount of energy consumed by the UAVs, it is assumed that they are launched in positions where each revolves with a constant speed on a small horizontal circle of radius r_l. Additionally, similar to [21] and [27], the GUs are deployed following Random Waypoint, a mobility model widely used to simulate wireless networks. Therefore, during association, a snapshot of the mobility process is taken whereby the 3D coordinate of each GU is given by $(x_m, y_m, 0)$. We have also assumed that each GU can be associated with any of the UAVs. In addition, the UAVs have a limited number of users they can support known as quota defined by set $\mathcal{Q} = \{q_1, q_2, \ldots, q_l, \ldots, q_L\}, \forall l \in \mathcal{L}$. Let $d_{m,l}$ represents the Euclidean distance between the UAV l and GU m, then it is calculated as follows,

$$d_{m,l} = \sqrt{(x_l - x_m)^2 + (y_l - y_m)^2 + z_l^2}. \tag{1}$$

B. CHANNEL MODEL

We adopt type B aerial-to-ground (A2G) channel model detailed in [32] and [33]. The path-loss between UAV l and GU m defined in [33] is used and stated as,

$$PL_{m,l}[\text{dB}] = 10\gamma \log_{10} d_{m,l} + \eta, \tag{2}$$

and its absolute value expression is given by,

$$\Phi_{m,l} = 10^{\frac{PL_{m,l}[\text{dB}]}{10}} = \Gamma_{m,l} d_{m,l}^{\gamma}, \tag{3}$$

where $\Gamma_{m,l} = 10^{\frac{\eta}{10}}$, γ is the path-loss exponent and η is the path-loss at the reference point (1 meter). From [33], the typical values for γ and $\Gamma_{m,l}$ in a free space A2G channel are

$$\gamma = 2, \text{ and } \Gamma_{m,l} = 10^{\frac{B}{10} + \frac{A}{10 + 10\ a' \exp\left(-b'(\theta_{m,l} - a')\right)}}, \tag{4}$$

where the constants A and B are given by $A = \eta_{\text{LOS}} - \eta_{\text{NLOS}}$ and $B = 20 \log_{10}(\frac{4\pi f}{c}) + \eta_{\text{NLOS}}$. The $f = 2$ GHz represents the carrier frequency and $c = 3 \times 10^8$ m/s is the speed of light [32]. The $\theta_{m,l}$ represents the elevation angle between UAV l and user m given by,

$$\theta_{m,l} = \frac{180}{\pi} \arctan\left(\frac{z_l}{d_{m,l}}\right). \tag{5}$$

The a', b', η_{LOS}, and η_{NLOS} are the parameters that depend on the environment; in sub-urban areas, $a' = 5.0188$, $b' = 0.3511$, $\eta_{\text{LOS}} = 0.1$ dB, and $\eta_{\text{NLOS}} = 21$ dB [33].

Therefore, the A2G channel gain coefficient between the UAV l and GU m is given by,

$$h_{m,l} = \frac{g_{m,l}}{\sqrt{\Phi_{m,l}}}, \tag{6}$$

where $g_{m,l} \sim \mathcal{CN}(\mu, 2\sigma^2)$ is the small-scale fading modeled from the Rician distribution since there is a dominant LoS between UAVs and GUs. Complex Gaussian distribution with mean μ and variance $2\sigma^2$, is used to model Rician fading similar to [21]. The mean and variance are given by $\mu = \sqrt{\frac{K_{m,l}}{K_{m,l}+1}} P_{\text{LOS}}$ and $\sigma = \frac{P_{\text{LOS}}}{\sqrt{2(K_{m,l}+1)}}$, where $K_{m,l} = A_1 \exp(A_2 \theta_{m,l})$ is the Rician factor and P_{LOS} is the power of the line-of-sight. The $\theta_{m,l}$ is the elevation angle between user m and UAV l provided in (5). The A_1 and A_2 are the constants depending on the environment. From [21], $A_1 = 5$ and $A_2 = \frac{2}{\pi} \log_{10}(15/A_1)$.

III. OPTIMIZATION PROBLEM FORMULATION

In NOMA technology, the signals of multiple users are superposed at the transmitter side and then sent through the

channel as a single signal. It is assumed that the available number of GUs in the system is greater than or equal to the sum of the quotas of all the UAVs and the q_l of UAV l is an even number to facilitate pairing,

$$M \geq \sum_{l=1}^{L} q_l. \tag{7}$$

The q_l GUs associated to UAV l are grouped into pairs of two GUs such that $p_{=}1, \ldots, 2, \ldots q_l/2$. Figure 2 illustrates the UAV serving the GUs in the NOMA scheme for the considered system model. The two GUs which are in the same pair are given a single resource block as in [21]. Hence the signal received at the two GUs in pair p is given by,

$$y_{m,p} = |h_{m,p}|^2 \left(\sum_{n=1}^{2} x_{n,p} \sqrt{\Omega_{n,p} P_t} \right) + \zeta_{m,p}, \tag{8}$$

where $h_{m,p}$ is the channel gain coefficient between the UAV l and the GU m in pair p, $\Omega_{n,p}$ represents the power coefficient allocated to the GUs in pair p, and P_t is the UAV's transmit power. The signal being transmitted is $x_{n,p}$ while $\zeta_{m,p}$ represents the noise term of the channel modeled by Gaussian distribution with zero mean and variance σ^2. We assume that the near user to the UAV has a strong channel while the far user has a weak channel. Therefore, the near user is allocated low power while the far user is allocated high power to maintain the NOMA principle. This requires that the channel gain coefficient of the far user be less than or equal to that of the near user, $|h_{1,p,l}|^2 \leq |h_{2,p,l}|^2$ where $h_{1,p,l}$ and $h_{2,p,l}$ are respectively the channel gains of the far and near users which belong in pair p of UAV l. The instantaneous signal-to-interference-plus-noise ratio (SINR), $\delta_{1,p,l}$, at the far GU in pair p being served by UAV l is calculated as follows,

$$\delta_{1,p,l} = \frac{\rho |h_{1,p,l}|^2 \Omega_{1,p,l}}{\rho |h_{1,p,l}|^2 \Omega_{2,p,l} + 1}, \tag{9}$$

where $\Omega_{1,p,l}$ and $\Omega_{2,p,l}$ are respectively the power allocation coefficients for the far and near GUs that belong to pair p of the UAV l. The $\rho = P_t/N_0^2$ is the transmit SNR of the UAVs which is obtained from transmit power P_t and average noise power spectral density N_0^2.

At the near GUs, the signal of the far user is decoded first and then be removed from the received one by using successive interference cancellation (SIC). Assuming perfect SIC, the instantaneous SINR, $\delta_{2,p,l}$, at the near GUs in pair p being served by UAV l is given by

$$\delta_{2,p,l} = \rho |h_{2,p,l}|^2 \Omega_{2,p,l}. \tag{10}$$

The instantaneous rate achieved by both far and near GUs are respectively given by

$$R_{1,p,l}^{\text{NOMA}} = \log_2 \left(1 + \delta_{1,p,l}\right), \tag{11}$$

and

$$R_{2,p,l}^{\text{NOMA}} = \log_2 \left(1 + \delta_{2,p,l}\right). \tag{12}$$

Therefore, the overall system achievable capacity of all the UAVs with the associated GUs is given by,

$$R^{\text{NOMA}} = \sum_{l=1}^{L} \sum_{p=1}^{q_l/2} \left(R_{1,p,l}^{\text{NOMA}} + R_{2,p,l}^{\text{NOMA}} \right). \tag{13}$$

To ensure the QoS requirements for every NOMA user, orthogonal multiple access (OMA) is also considered similar to [21]. The instantaneous rate of far and near GUs served by UAV l under OMA scheme is calculated as follows,

$$R_{i,p,l}^{\text{OMA}} = \frac{1}{2} \log_2 \left(1 + \rho |h_{i,p,l}|^2\right), i \in \{1, 2\}, \tag{14}$$

where multiplication of factor $\frac{1}{2}$, represents the equal power allocation for the two OMA users [34], [35]. The overall system achievable capacity of all the UAVs in the OMA scheme is given by,

$$R^{\text{OMA}} = \sum_{l=1}^{L} \sum_{p=1}^{q_l/2} \left(R_{1,p,l}^{\text{OMA}} + R_{2,p,l}^{\text{OMA}} \right). \tag{15}$$

We aim to maximize the system throughput of the UAV's network. Referring to [21] and [27], the optimization problem is formulated as follows:

$$\max_{\Omega} \sum_{l=1}^{L} \sum_{p=1}^{q_l/2} \left(R_{1,p,l}^{\text{NOMA}} + R_{2,p,l}^{\text{NOMA}} \right) \tag{16a}$$

$$s.t.: \quad R_{1,p,l}^{\text{NOMA}} \geq R_{1,p,l}^{\text{OMA}}, \tag{16b}$$

$$R_{2,p,l}^{\text{NOMA}} \geq R_{2,p,l}^{\text{OMA}}, \tag{16c}$$

$$\Omega_{1,p,l}, \Omega_{2,p,l} \geq 0, \tag{16d}$$

$$\Omega_{1,p,l} + \Omega_{2,p,l} = 1, \tag{16e}$$

$$\Omega_{1,p,l} \geq \Omega_{2,p,l}. \tag{16f}$$

The conditions specified in (16b) and (16c) are in place to guarantee that the QoS is maintained by ensuring that the rate achieved by both the far and near GUs in the NOMA scheme is greater than or equal to that of OMA [21], [34]. In every NOMA pair, the power coefficients allocated to the far and near GUs denoted by $\Omega_{1,p,l}$ and $\Omega_{2,p,l}$, respectively, must be non-negative (16d) and sum up to one (16e). It is ensured that the far user is allocated higher power than the near user (16f), as per the NOMA principle.

IV. JOINT USER ASSOCIATION AND PAIRING BASED ON DϵTS MAB AND TWO-SIDED MATCHING

This section details the solution approach to the optimization problem in (16a). We propose the DϵTS MAB and two-sided matching frameworks for joint user association and pairing to maximize the overall system throughput.

A. POWER ALLOCATION COEFFICIENTS DESIGN

We first design the adaptive power allocation coefficients for far and near users that satisfy the constraints in (16b)-(16f)

referring to [34]. The relation in constraint (16b) states that

$$\log_2\left(1+\frac{\rho|h_{1,p,l}|^2\Omega_{1,p,l}}{\rho|h_{1,p,l}|^2\Omega_{2,p,l}+1}\right)\geq\log_2\left(1+\rho|h_{1,p,l}|^2\right). \tag{17}$$

By expanding the constraint (16e) which states that $\Omega_{1,p,l} = 1-\Omega_{2,p,l}$ and through algebraic manipulation of (17), the upper bound of $\Omega_{2,p,l}$ is given by,

$$\Omega_{2,p,l}\leq\frac{1}{1+\sqrt{1+\rho|h_{1,p,l}|^2}}. \tag{18}$$

Similarly, from constraint (16c) and (19),

$$\log_2\left(1+\rho|h_{2,p,l}|^2\Omega_{2,p,l}\right)\geq\log_2\left(1+\rho|h_{2,p,l}|^2\right), \tag{19}$$

the lower bound of $\Omega_{2,p,l}$ is calculated as,

$$\Omega_{2,p,l}\geq\frac{1}{1+\sqrt{1+\rho|h_{2,p,l}|^2}}. \tag{20}$$

From (18) and (20), the adaptive power allocation coefficients for the near and far GUs in pair p, served by UAV l are as follows:

$$\Omega_{2,p,l}=\frac{\psi_1}{1+\sqrt{1+\rho|h_{1,p,l}|^2}}+\frac{\psi_2}{1+\sqrt{1+\rho|h_{2,p,l}|^2}}, \tag{21}$$

and

$$\Omega_{1,p,l}=1-\Omega_{2,p,l}, \tag{22}$$

where ψ_1 for far GU and ψ_2 for near GU are respectively the parameters to balance the achievable rate. Both parameters must sum to one, $\psi_1+\psi_2=1$ and $0\leq\psi_i\leq1$ with $i=1,2$. The designed power allocation coefficients $\Omega_{1,p,l}$ and $\Omega_{2,p,l}$ depend on the channel conditions and the transmit SNR. In addition, they ensure that all the constraints in the optimization problem are satisfied.

B. PROPOSED D ϵTS FRAMEWORK

The MAB techniques have generally been used to solve decision-making problems in which agents adaptively learn the dynamic environment to maximize long-term rewards. The joint user association and pairing in UAV-assisted NOMA networks fit the MAB problem where UAVs act as the agents need to identify the potential GUs that can maximize the long-term rewards.

1) ACTION SPACE DESIGN

With the total number of GUs available in the network M, each UAV forms action space, $\mathbf{A}_{l,\text{space}}$ according to its quota q_l. For every UAV, the total number of actions available is calculated as follows,

$$A_l=\frac{M!}{(M-q_l)!(\frac{q_l}{2})!(2^{\frac{q_l}{2}})}, \tag{23}$$

Algorithm 1 Action Space Generator for Every UAV

Input: M number of GUs and quota of each UAV q_l.
Output: $\mathbf{A}_{l,\text{space}}$

Initialization: Set of all possible combinations of q_l out of M GUs, $\mathcal{D}_l$, which results into total number $F_l = \frac{M!}{q_l!(M-q_l)!}$.
Set $\mathbf{A}_{l,\text{space}}=\emptyset$.
for $z=1\longrightarrow F_l$ **do**
  Get row z of $\mathcal{D}_l$, $row_l=\mathcal{D}_l(z,:)$.
  function A_{SET} = generateActions(row_l, q_l)
    Calculate: $W=\frac{q_l!}{2^{q_l/2}(q_l/2)!}$
    Initialize: $A_{\text{SET}}=\text{zeros}(W,q_l)$
    Update: $A_{\text{SET}}(:,1)=row_l(1)$
    for $i=1\longrightarrow q_l-1$ **do**
      $A_{\text{SET}}(i+(i-1)*\frac{W}{q_l-1}:\frac{i*W}{q_l-1},2)=row_l(i+1)$
      if $q_l\neq2$ **then**
        $j=2:q_l$
        $j=j(j\neq(i+1))$
        $A_{\text{SET}}(i+(i-1)*\frac{W}{q_l-1}:\frac{i*W}{q_l-1},3:q_l)=$ generateActions($row_l(j)$).
      end if
    end for
  end function
  $\mathbf{A}_{l,\text{space}}=[\mathbf{A}_{l,\text{space}};A_{\text{SET}}]$
end for

where operation (!) is factorial. Algorithm 1 details the process of generating actions, $\mathbf{A}_{l,\text{space}}$ formulated from M GUs and quota q_l of UAV l. For instance, if $M=8$ and $q_l=4$ for UAV l, the action space could be as follows,

$$\mathbf{A}_{l,\text{space}}=\begin{bmatrix}1&2&3&4\\1&3&2&4\\1&4&2&3\\1&2&3&5\\\cdots&\cdots&\cdots&\cdots\\5&7&6&8\\5&8&6&7\end{bmatrix}, \tag{24}$$

with $A_l=210$ total actions. Let suppose that the action $\mathbf{a}_l=\begin{bmatrix}1&2&3&4\end{bmatrix}$ is selected from $\mathbf{A}_{l,\text{space}}$ by UAV l, then the UAV forms the pairs as follows,

$$\mathbf{P}_{\text{set}}=\begin{bmatrix}1&2\\3&4\end{bmatrix}, \tag{25}$$

where GUs 1 and 2 make the first pair; and 3 and 4 make the second pair.

2) DϵTS ALGORITHM

Inspired by [31], we detail the proposed DϵTS MAB algorithm. The authors in [31] proposed an ϵ-explore Thompson sampling (ϵ-TS) where the agents apply the standard Thompson sampling with a fixed small exploration rate ϵ and $(1-\epsilon)$ for playing actions with high expected rewards.

Algorithm 2 Proposed DϵTS and Two-sided Matching Algorithm

Input: Number of GUs M, number of UAVs L, preference lists of GUs $\mathcal{U}_m$, expected rewards $\bar{x}_l$ for all UAVs, and exploration rate ϵ_l.
Output: Assignment matrix $\mathcal{A}_l^{\text{match}}$

function DϵTS($L, M, \mathcal{U}_m, \bar{x}$)
  Initialize: $\mathcal{A}_l^{\text{match}} = \emptyset, \forall l \in \mathcal{L}$, that store the selected action.
  Initialize the set $\mathcal{A}_m^{\text{apt}} = \emptyset$ that stores and keeps track of the UAVs' proposals at the GUs.
  while $\mathcal{A}_l^{\text{match}}(:) = \emptyset$ **do**
    for each UAV $l = 1 \longrightarrow L$ **do**
      Check whether no GU rejected the UAV l in the previously accepted GUs.
      Pick a random number r_l from uniform distribution in [0, 1] for all UAVs.
      if $r_l \leq \epsilon_l$ **then**
        Draw sample means $\hat{\mathcal{M}}_a$ as per (38).
        Choose action with the highest sample mean: $\mathcal{A}_{\text{prop}}(l) = \arg\max_{a \in \mathcal{A}} \hat{\mathcal{M}}_a$.
      else
        Select an action with the highest expected rewards, $\mathcal{A}_{\text{prop}}(l) = \arg\max(\bar{x}_l(:))$.
      end if
    end for
    for each GU $m = 1 \longrightarrow M$ **do**
      Pull the request from the proposing UAVs, $\mathcal{A}_{\text{prop}}(:)$ and store them in set $\sqcup_m$.
      if $\mathcal{A}_m^{\text{apt}} = \emptyset$ **then**
        Accept the high-ranked UAVs in the user's preferences list, $\mathcal{A}_m^{\text{apt}} = \max(\mathcal{U}_m(\sqcup_m))$.
      else if $\mathcal{A}_m^{\text{apt}} < \mathcal{U}_m(\sqcup_m)$ **then**
        Check and accept the high-ranked UAV among the proposing ones and reject the previous acceptance. $\mathcal{A}_m^{\text{apt}} = \max(\mathcal{U}_m(\sqcup_m))$.
      else
        Keep the previous accepted UAV: $\mathcal{A}_m^{\text{apt}}$.
      end if
    end for
    for each UAV $l = 1 \longrightarrow L$ **do**
      Check the feedback from the GUs which is either acceptance or rejection.
      if all $\mathcal{A}^{\text{apt}}(\mathcal{A}_l^{prop}) = l$ **then**
        Keep all the GUs in the selected action: $\mathcal{A}_l^{\text{match}} = \mathcal{A}_l^{prop}(:)$.
      else if any $\mathcal{A}^{\text{apt}}(\mathcal{A}_l^{prop}) \neq l$ **then**
        Keep only the actions with the accepted GUs and remove the ones that have rejected the proposals in the expected rewards of the UAV and continue to the next iteration.
      else if all $\mathcal{A}^{\text{apt}}(\mathcal{A}_l^{prop}) \neq l$ **then**
        Remove the actions that contain the GUs that have rejected the proposals in the expected rewards of the UAV and continue to the next iteration.
      end if
    end for
  end while
end function

With DϵTS, the agents start with an initial large exploration rate $\epsilon = 0.5$ for applying the standard Thompson sampling which gradually decreases as information about the actions becomes available and in the rest $(1 - \epsilon)$, the agents select actions with the highest expected rewards. The main motivation is that starting with high ϵ gives a high chance to apply Thompson sampling which helps in collecting much information about the actions then later as ϵ decreases, the chance of choosing actions with the highest expected rewards increases. This allows enough exploration during the starting and minimizes unnecessary exploration once better actions are identified.

The formula below is used by each UAV to update the rate of applying Thompson sampling, $\epsilon_{t+1,l}$ at every time step,

$$\epsilon_{t+1,l} = \frac{\epsilon_{t,l}}{1 + N_{t,l,\mathbf{a}_l}/T}, \tag{26}$$

where t is the current time step while T is the total time step for the entire experiment. The $N_{t,l,\mathbf{a}_l}$ is the number of times an action $\mathbf{a}_l$ was played by UAV l.

When the agents select to apply Thompson sampling, the details below show how it is applied.

Let the reward received by UAV l on the chosen action $\mathbf{a}_l \in \mathbf{A}_{l,\text{space}}$ at time t be $X_{t,l,\mathbf{a}_l} \in \mathbb{R}$ and defined by,

$$X_{t,l,\mathbf{a}_l} = \frac{R_{t,l,\mathbf{a}_l}^{\text{NOMA}}}{\omega\, q_l}, \forall l \in \mathcal{L}, \forall \mathbf{a}_l \in \mathbf{A}_{l,\text{space}}, \tag{27}$$

which is designed based on the overall rate achieved on the GUs of the selected action, the number of constraints in the optimization problem ω and the quota of the UAV q_l. We assume that the mean $\tau_{l,\mathbf{a}_l}$ and variance $\sigma_{l,\mathbf{a}_l}^2$ of the actions are unknown to the UAVs, then we adopt the use of conjugate priors $\mathcal{NIG}$ as normal-inverse-gamma distribution similar to [36] and [37] as follows,

$$P\left(\tau_{l,\mathbf{a}_l}, \sigma_{l,\mathbf{a}_l}^2 | \tau_{l,\mathbf{a}_l}, \nu_{l,\mathbf{a}_l}, \alpha_{l,\mathbf{a}_l}, \beta_{l,\mathbf{a}_l}\right) \sim \mathcal{NIG}(\tau_{l,\mathbf{a}_l}, \nu_{\mathbf{a}_l}, \alpha_{l,\mathbf{a}_l}, \beta_{l,\mathbf{a}_l}), \forall \mathbf{a}_l \in \mathbf{A}_{l,\text{space}} \tag{28}$$

where $\tau_{l,\mathbf{a}_l}$, $\nu_{l,\mathbf{a}_l}$, $\alpha_{l,\mathbf{a}_l}$, and $\beta_{l,\mathbf{a}_l}$ are the parameters representing the prior mean, prior count, prior shape and prior scale parameters, respectively. Both prior count, shape and scale parameters must be positive, $\nu_{l,\mathbf{a}_l}, \alpha_{l,\mathbf{a}_l}, \beta_{l,\mathbf{a}_l} > 0$. For the UAVs to select the actions with Thompson sampling, they calculate the posterior distributions that also follow the normal-inverse-gamma distribution (details cf. [36], [37]),

$$P\left(\mu_{l,\mathbf{a}_l}, \sigma_{l,\mathbf{a}_l}^2 | D_{t,l,\mathbf{a}_l}\right) \propto \mathcal{N}\left(\mu_{l,\mathbf{a}_l}; \varrho_{t,a}; \varphi_{l,\mathbf{a}_l}^2\right) \mathcal{IG}\left(\sigma_{l,\mathbf{a}_l}^2; \frac{1}{2}N_{t,l,\mathbf{a}_l} + \alpha_{l,\mathbf{a}_l}; \beta_{t,l,\mathbf{a}_l}\right), \tag{29}$$

where

$$\varrho_{t,l,\mathbf{a}_l} = \frac{\nu_{l,\mathbf{a}_l}\tau_{l,\mathbf{a}_l} + N_{t,l,\mathbf{a}_l}\bar{x}_{t,l,\mathbf{a}_l}}{\nu_{l,\mathbf{a}_l} + N_{t,l,\mathbf{a}_l}}, \tag{30}$$

$$\varphi_{l,\mathbf{a}_l}^2 = \frac{\sigma_{l,\mathbf{a}_l}^2}{\nu_{l,\mathbf{a}_l} + N_{t,l,\mathbf{a}_l}}, \tag{31}$$

$$\beta_{t,l,\mathbf{a}_l} = \beta_{l,\mathbf{a}_l} + \frac{1}{2}s_{t,l,\mathbf{a}_l} + \frac{N_{t,l,\mathbf{a}_l}\nu_{l,\mathbf{a}_l}(\bar{x}_{t,l,\mathbf{a}_l} - \tau_{l,\mathbf{a}_l})^2}{2(N_{t,l,\mathbf{a}_l} + \nu_{l,\mathbf{a}_l})}. \tag{32}$$

The $D_{t,l,\mathbf{a}_l}$ is the set that contains all the rewards collected up to time step $t - 1$. The $\bar{x}_{t,l,\mathbf{a}_l}$ and $N_{t,l,\mathbf{a}_l}$ are the average expected rewards and the number of times a UAV l successfully played the action $\mathbf{a}_l$, respectively. The $s_{t+1,l,\mathbf{a}_l}$ is the parameter that depends on the rewards collected and is

updated as follows,

$$s_{t+1,l,\mathbf{a}_l} = s_{t,l,\mathbf{a}_l} + X^2_{t,l,\mathbf{a}_l} + N_{t,l,\mathbf{a}_l}\overline{x}^2_{t,l,\mathbf{a}_l} - N_{t+1,l,\mathbf{a}_l}\overline{x}^2_{t+1,l,\mathbf{a}_l}. \quad (33)$$

Once a UAV takes an action, the posterior parameters are updated as follows:

$$\overline{x}_{t+1,l,\mathbf{a}_l} \longleftarrow \overline{x}_{t,l,\mathbf{a}_l} + \frac{X_{t,l,\mathbf{a}_l} - \overline{x}_{t,l,\mathbf{a}_l}}{N_{t+1,l,\mathbf{a}_l}}, \quad (34)$$

$$N_{t+1,l,\mathbf{a}_l} = N_{t,l,\mathbf{a}_l} + 1, \quad (35)$$

where $\overline{x}_{t+1,l,\mathbf{a}_l}$ is the updated expected rewards of the played action at time step t. The $N_{t+1,l,\mathbf{a}_l}$ is the updated number of times the action $\mathbf{a}_l$ has been selected by UAV l. The posterior variance, $\hat{\sigma}^2_{l,\mathbf{a}_l}$ and posterior mean, $\hat{\mathcal{M}}_{l,\mathbf{a}_l}$ are respectively updated as follows,

$$\hat{\sigma}^2_{l,\mathbf{a}_l} \sim \mathcal{IG}\left(\frac{1}{2}N_{t,l,\mathbf{a}_l} + \alpha_{l,\mathbf{a}_l}, \beta_{t,l,\mathbf{a}_l}\right), \quad (36)$$

$$\hat{\mathcal{M}}_{l,\mathbf{a}_l} \sim \mathcal{N}\left(\varrho_{t,l,\mathbf{a}_l}, \frac{\hat{\sigma}^2_{l,\mathbf{a}_l}}{\left(N_{t,l,\mathbf{a}_l} + \nu_{l,\mathbf{a}_l}\right)}\right), \quad (37)$$

where $\mathcal{IG}(.)$ and $\mathcal{N}(.)$ are inverse-gamma and normal distribution, respectively. Therefore, once a UAV applies Thompson sampling, it proposes to the GUs that are in the action which has the highest sample means as follows,

$$\mathbf{a}_l = \underset{\mathbf{a}_l \in \mathbf{A}_{l,\text{space}}}{\arg\max}\ \mathbb{E}\left[X_{t,l,\mathbf{a}_l} | \hat{\mathcal{M}}_{l,\mathbf{a}_l}, \hat{\sigma}^2_{l,\mathbf{a}_l}\right] = \underset{\mathbf{a}_l \in \mathbf{A}_{l,\text{space}}}{\arg\max}\ \hat{\mathcal{M}}_{l,\mathbf{a}_l}. \quad (38)$$

The summary of the proposed DϵTS MAB and two-sided matching is provided in Algorithm 2. The inputs to the algorithm are the hyper-parameters of Thompson sampling, a set of UAVs, a set of GUs, the probability to apply Thompson sampling ϵ_l, and the expected rewards of all UAVs over all actions. At the start, each UAV picks a random number, r from a uniform distribution and compares it with the ϵ_l. If $r \leq \epsilon_l$, the UAV chooses an action with Thompson sampling, otherwise, it chooses an action with the highest expected rewards. Each UAV starts to communicate with the GUs in its selected action offering proposals containing the promised throughput to those GUs. Note that all the UAVs select the actions and propose to the GUs simultaneously which leads to conflict on the user side. If multiple UAVs attempt to choose the same GU, the GU accepts the proposal of the UAV which is highly ranked in its preference lists. Therefore, the following cases are considered for the UAV to decide on the feedback received from the GUs.

i) When all GUs involved in the selected action $\mathbf{a}_l$ accept the proposal made by UAV l. In this case, the UAV has successfully found the GUs to serve. However, it continues to check whether no GU rejects the previous acceptance. If it happens, the UAV resumes proposing to the GUs in the next better action excluding all the actions which have any of the GUs that rejected the previous proposals.

Algorithm 3 Rewards Calculation for DϵTS MAB Algorithm

Input: Number of GUs M, number of UAVs L, and total time steps T.

Output: Association matrix

1: Obtain A_{space} from Algorithm 1, for all UAVs.
2: Initialize the UAV's expected rewards, $\overline{x}_{t,l,\mathbf{a}_l}$, to random numbers in the range 0 and 1 from the uniform distribution.
3: Set $N_{l,\mathbf{a}_l} = 1$, and $\epsilon_l = 0.5, \forall l \in \mathcal{L}, \forall \mathbf{a}_l \in \mathbf{A}_{l,\text{space}}$.
4: **for** $t = 1 \longrightarrow T$ **do**
5: Obtain an association matrix $\mathcal{A}^{\text{match}}_l$ that contains the selected GUs by running Algorithm 2.
6: Calculate rewards $X_{t,l,\mathbf{a}_l}$ as per (27).
7: Update $\overline{x}_{t,l,\mathbf{a}_l}$ and $N_{t,l,\mathbf{a}_l}$ as per (34) and (35) respectively.
8: Update $s_{t,l,\mathbf{a}_l}$, $\varrho_{t,l,\mathbf{a}_l}$, $\tau_{t,l,\mathbf{a}_l}$, and $\beta_{t,l,\mathbf{a}_l}$.
9: **end for**

ii) When some GUs in the selected action $\mathbf{a}_l$ accept the proposal while others reject it. This happens when a GU gets proposals from multiple UAVs. In this case, the GU accept the UAV's proposal which is highly ranked in its preference list. Thus, each UAV keeps the actions that contain only the accepted users and chooses from these actions in the next matching iteration.

iii) When all GUs in the selected action $\mathbf{a}_l$ reject the proposal of the UAV l. In this instance, the UAV selects from only the actions that do not contain any of the GUs that rejected the previous proposal. This means that the UAV removes all actions that contain the rejected GUs from the pool of available actions and moves to the next matching iteration to choose again.

In contrast to the approach proposed in [27], where it is assumed that a GU must reject all subsequent proposals if it accepts an association with one UAV, this work allows the GUs to switch to a new proposing UAV if it is high ranked than the previous one. This could enhance the system throughput since the GUs will be matched with the potential UAVs in their preference lists that have ever proposed. Note that the matching process has to run until all the UAVs are matched according to their quotas. In addition, this matching does not allow a GU to be matched with more than one UAV.

3) REWARDS CALCULATION

Once a UAV gets the GUs to serve from the chosen action, the reward obtained is designed based on the overall rate achieved, the number of constraints in the optimization problem ω and the quota of the UAV q_l as per (27). Similarly, the expected reward of each UAV is updated as per (34). Algorithm 3 summarizes the pseudo-code for joint user association and pairing in multi-UAV-assisted NOMA networks based on DϵTS. To understand clearly the association and pairing process, a flowchart is provided

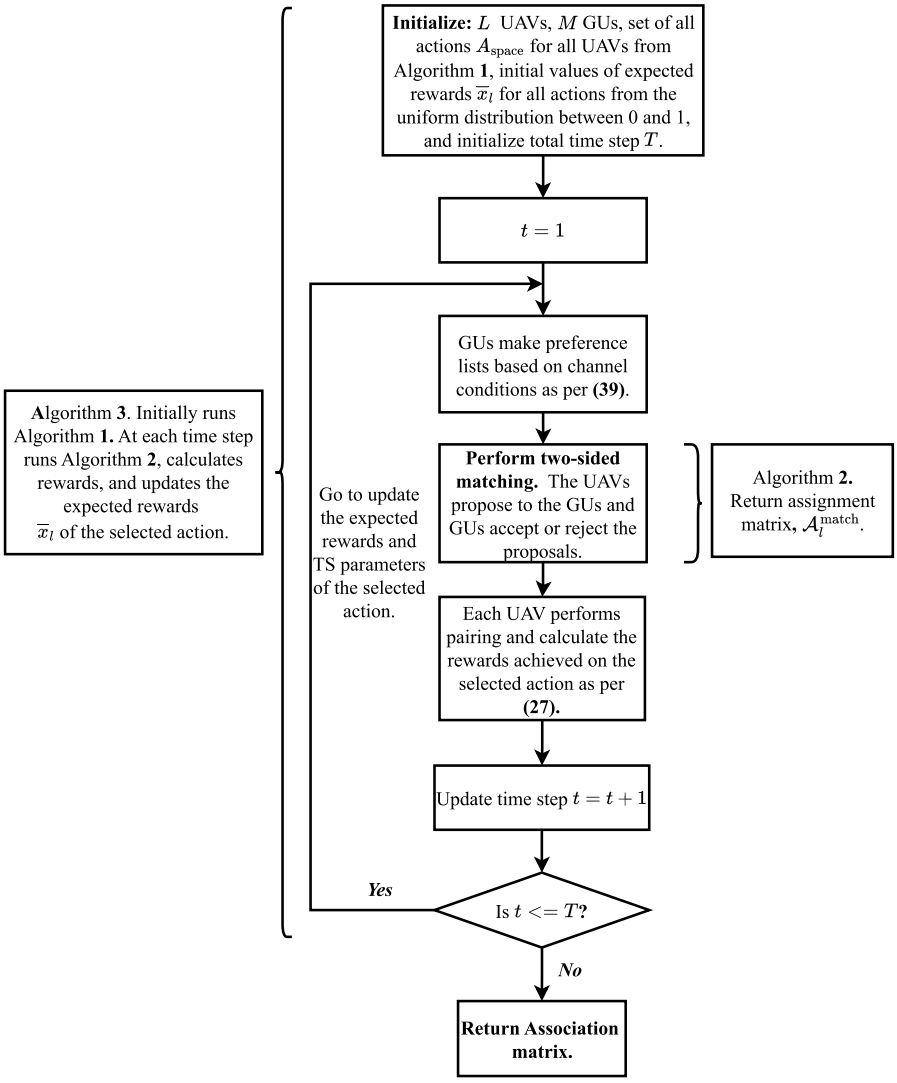


FIGURE 3. Flowchart illustrating the proposed DϵTS method for user association and pairing in the considered system model.

in Figure IV-B2. It shows where each of the mentioned algorithms has to be applied.

4) PREFERENCE LISTS OF GUS

Referring to [27], [38], and [39], the GUs make their preference lists according to the channel conditions as follows,

$$\mathcal{U}_m = sort(h(:, m), \text{“}Descend\text{”}), \tag{39}$$

where $h(:, m)$ represents the channel gains between GU m and all the UAVs, and *sort* is the function used for sorting. The *Descend* operation means that the sorting is done in descending order.

To evaluate the performance of the proposed algorithms, regret analysis is considered. The regret of a particular UAV is calculated by comparing the rewards that would have been received if the best action was selected and the rewards it has received from its current action [21]. Mathematically, it is defined as,

$$Regret_l = \sum_{t=1}^{T} \left(\max_{\mathbf{a}_l} \overline{x}(l, :) - E[\overline{x}_{t,l,\mathbf{a}_l}(t)] \right), \mathbf{a}_l \in \mathbf{A}_{l,\text{space}}, \tag{40}$$

where the $E[.]$ operation denotes the statistical average. Regret measures an algorithm's efficiency, specifying how far the UAVs deviate from the optimal selection.

V. SIMULATION SETUPS

The summary of simulation setups is provided in Table 2. It consists of 4 UAVs available in the system and various numbers of GUs. The ground coverage considered is 1000 m x 1000 m while the heights of every UAV are considered to be [50, 100, 150, 200] m. Following Random Waypoint mobility model, the GUs are distributed in the coverage area and change direction in every sample. The simulation is

TABLE 2. Simulation parameters.

Parameters	Value
Ground coverage	1000 m x 1000 m
Number of UAVs	4
UAVs' heights	[50, 100, 150, 200] m
UAVs' radius	20 m
ψ_1	0.8
ψ_2	0.2
ω	5.0
a'	5.0188
b'	0.3511
η_{LOS}	0.1 dB
η_{nLOS}	21 dB
$\tau_{l,\mathbf{a}_l}$	0
$v_{l,\mathbf{a}_l}$	10.0
$\alpha_{l,\mathbf{a}_l}$	1.0
$\beta_{l,\mathbf{a}_l}$	1.0
Small scale fading	Rician distribution: $K = 15$ and $P_{\text{LOS}} = 1.0$
Path-loss exponent, γ	2.0
Total time steps, T	2000
Total number of samples	100
Carrier frequency	2 GHz
Noise power, N_0^2	-100 dBm

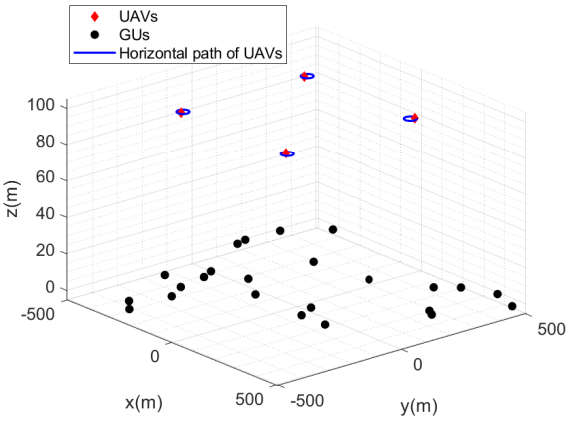


FIGURE 4. 3D coordinate plane of UAVs and GUs.

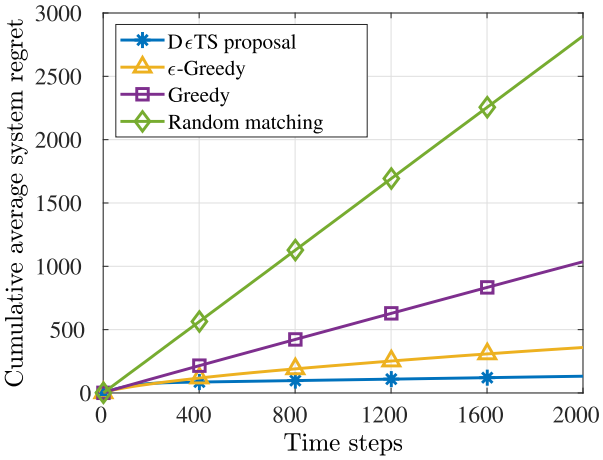


FIGURE 5. Cumulative average system regret versus time steps (P_t 20 dBm, 4 UAVs, 24 GUs, quota = 4 GUs and height = 100 m for each UAV).

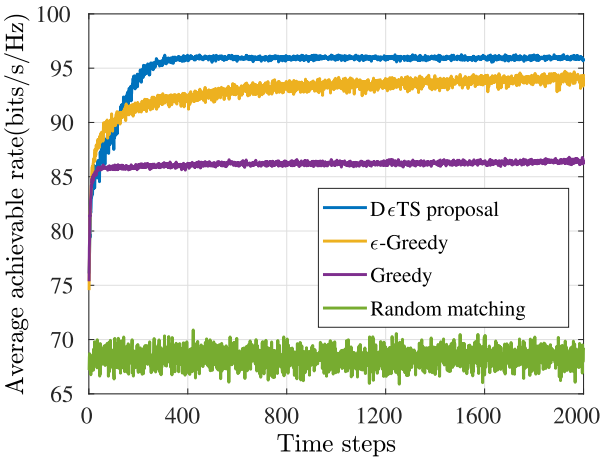


FIGURE 6. Average system sum rate versus time steps (P_t 20 dBm, 4 UAVs, 24 GUs, quota = 4 GUs and height = 100 m for each UAV).

done by running 100 Montecarlo samples, and UAVs run the algorithms for time steps $T = 2000$ in every sample. The $\gamma = 2.0$ has been set for the path-loss exponent [33]. In every NOMA pair of GUs associated with each UAV, parameters for balancing the user's rate are $\psi_1 = 0.8$ and $\psi_2 = 0.2$ in respect of far and near GUs chosen similar to [21]. The average noise power used is $N_0^2 = -100$ dBm as in [40]. In addition, to start with unknown reward distributions of every action, we have set the UAVs to initialize them at random following a uniform distribution in the range between 0 and 1. Afterwards, we evaluate the performance of the proposed DϵTS against the TS, ϵ-Greedy, Greedy, and random matching methods.

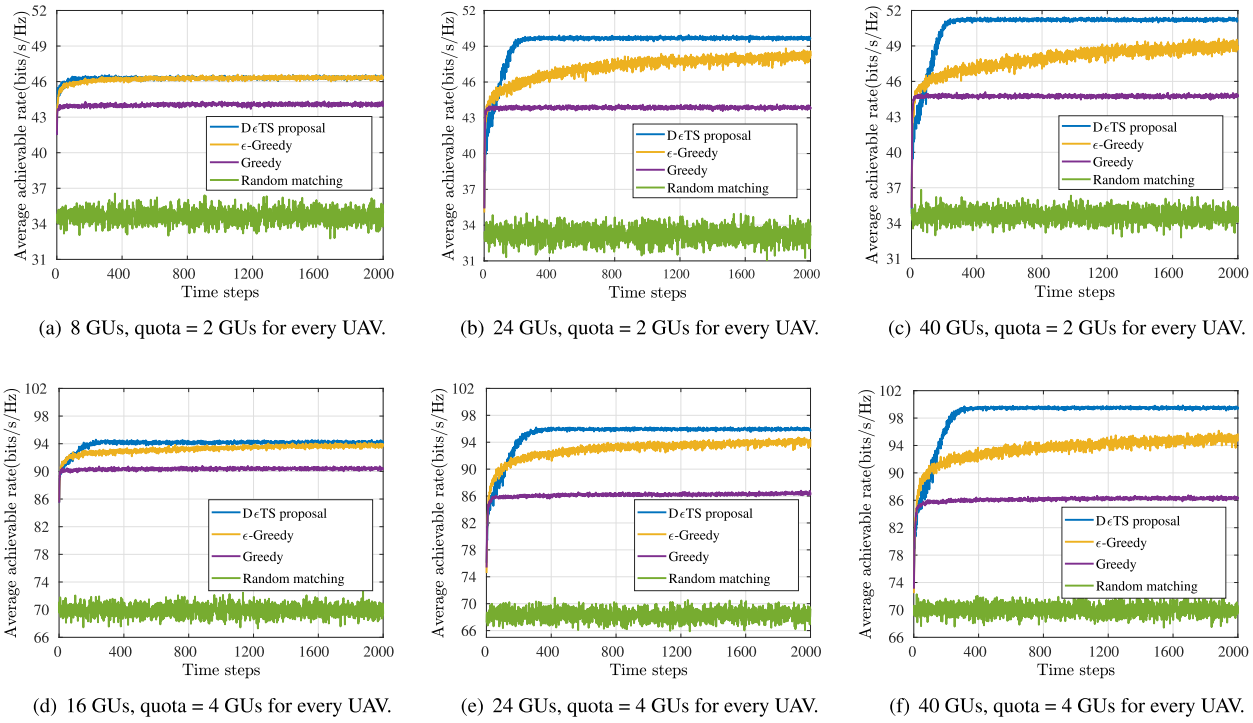
VI. RESULTS AND DISCUSSIONS

The simulation was done in MATLAB R2023b software and the results obtained are as follows: The 3D coordinate plane demonstrating the positions of UAVs and GUs is shown in Figure 4. The UAVs are deployed in space at a fixed altitude where they revolve around a small horizontal circle for energy saving while the GUs are deployed following the Random Waypoint mobility model [21].

To evaluate the performance of the proposed DϵTS and two-sided matching algorithm, a system consisting of 4 UAVs, 24 GUs, quota = 4 GUs each UAV, and transmit power $P_t = 20$ dBm, is taken into account. This is considered for Figure 5 and 6.

Note that the UCB method requires the agents to make a complete scan of all actions for initialization [21], [27], [29], [30]. Since there is a larger number of actions in the considered system model than the running time steps, the agents can not go through all actions. Therefore, it is not considered in the simulation results.

Figure 5 depicts the average cumulative regret of the system versus time steps. This figure shows how far the algorithm is from choosing the optimal actions. The proposed DϵTS method demonstrates the ability to identify potential actions that can maximize the system throughput, resulting in smaller regret compared to the other MAB algorithms.



(a) 8 GUs, quota = 2 GUs for every UAV. (b) 24 GUs, quota = 2 GUs for every UAV. (c) 40 GUs, quota = 2 GUs for every UAV.

(d) 16 GUs, quota = 4 GUs for every UAV. (e) 24 GUs, quota = 4 GUs for every UAV. (f) 40 GUs, quota = 4 GUs for every UAV.

FIGURE 7. Average system sum rate versus time steps for different number of GUs and quotas (P_t = 20 dBm, 4 UAVs, height = 100 m for each UAV).

Figure 6 provides the performance of DϵTS algorithm along with other MAB methods in terms of the achievable rate of the system versus time steps. From this figure, it is clear that the DϵTS can learn faster and converge to maximum than the other MAB algorithms. The reason for the better performance of the DϵTS is that the exploration rate is initialized at a large number $\epsilon_l = 0.5$, which allows more chances of applying TS by enabling the UAV to play the action with the highest sample mean. This is a part of the learning process to identify and build beliefs about the potential actions. Subsequently, as the UAVs obtain information about actions, the ϵ_l is decreased leading to a reduction of unnecessary exploration which minimizes the time for learning. This ensures a focus on actions that have the highest expected rewards.

The performance evaluation of the DϵTS against other methods in terms of system achievable rate versus time steps is done by varying the system size as shown in Figure 7. In Figures 7.(a), (b), and (c), the quota was fixed to 2 GUs for each UAV and changing the number of GUs. Clearly, the DϵTS have better performance than others in all cases. Similarly, once quota is set to 4 GUs, in Figures 7.(d), (e), and (f), the DϵTS still outperforms others in identifying the potential GUs that maximize the system sum rate. Increasing the number of users in the system also allows the UAVs to find the best users according to their quotas. We can simply see that the rate achieved by the DϵTS during learning in Figure 7.(d) is lower than that of Figures 7.(e) and (f) but still outperforms other MAB methods.

Figure 8 and Table 3 demonstrate the performance of the proposed DϵTS in comparison to other Thompson sampling

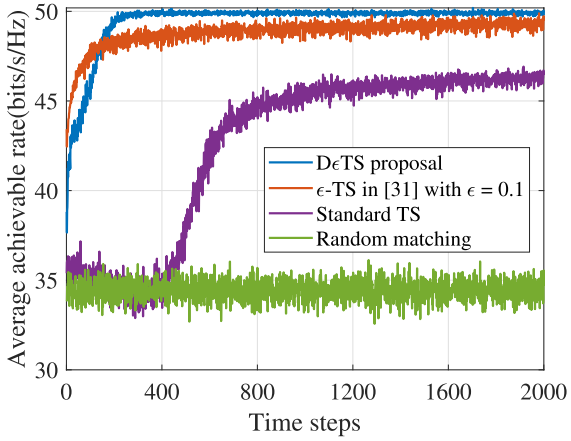


FIGURE 8. Average system sum rate versus time steps: different Thompson sampling methods (P_t = 20 dBm, 4 UAVs, 24 GUs, quota = 2 GUs and height = 100 m for each UAV).

methods. In Figure 8, the DϵTS can successfully identify the GUs that achieve higher throughput and quickly converge to the maximum while taking shorter average running time as shown in Table 3. The ϵ-TS method proposed in [31] is applied by taking a small rate ϵ_l to apply Thompson sampling, which remains fixed throughout the total time steps. As a result, it continues to apply Thompson sampling with this rate even after acquiring enough information to determine the potential actions. This results into much convergence time and reduced performance in finding suitable associations. The $\epsilon_l = 0.1$ [31] was considered to simulate the ϵ-TS method. The average running time shown in Table 3 was

TABLE 3. Average running time of different Thompson sampling algorithms over total time steps T = 2000 (P_t $\mathcal{D}$ 20 dBm, 4 UAVs, 24 GUs, quota = 2 GUs and height = 100 m for each UAV) in seconds.

System size	Proposed DϵTS	ϵ-TS [31] with $\epsilon = 0.1$	Standard-TS
4 UAVs, 8GUs, quota = 2 GUs each	**2.0600**	3.085	10.8099
4 UAVs, 16 GUs, quota = 4 GUs each	**31.328**	67.679	575.4955
4 UAVs, 24 GUs, quota = 4 GUs each	**111.3201**	290.793	2.8195e+03

TABLE 4. Average system sum rate at time step T = 2000 for various MAB methods (P_t $\mathcal{D}$ 20 dBm, 4 UAVs, height = 100 m for each UAV) in bits/s/Hz.

Quotas	Number of GUs	Proposed DϵTS	ϵ-Greedy	Greedy	Random matching
2 GUs	8	**46.98**	46.67	44.41	35.06
	24	**49.67**	48.18	44.03	33.22
	40	**51.17**	49.19	44.69	34.61
4 GUs	16	**94.71**	93.97	90.55	70.57
	24	**96.10**	94.14	86.28	68.11
	40	**99.11**	95.19	86.11	70.01

recorded and averaged over the total time steps $T = 2000$, under different system sizes. It has been found that the DϵTS takes the shortest time to run compared to the ϵ-TS and standard TS. When the standard Thompson sampling is applied for both exploration and exploitation, the UAVs require a significant amount of time to create the distribution beliefs of all actions. This is due to the large number of actions available in the considered system model. Therefore, its performance behaves like Random matching in the starting but later improves.

To further discuss the performance of the proposed DϵTS method, the average system sum rate is recorded at T = 2000 and summarized in Table 4. The findings clearly show that over different number of GUs available in the system, the DϵTS method outperforms the other MAB techniques in terms of the rate achieved by the entire system. Table 5 provides the average system sum rate achieved by changing the value of the vertical altitude of the UAVs. The proposed method achieves better performance than other MAB techniques when considering different heights and quotas of UAVs. We have also noticed that when the vertical height of the UAVs is increased, the maximum value of the achievable rate of the system falls due to the change in distances between UAVs and GUs which affect the channel conditions.

Further validation of the proposed DϵTS method is done by considering the system throughput versus the transmit power of the UAVs as shown in Figure 9. Both the Exhaustive search and random matching methods were simulated and compared with the DϵTS. The DϵTS has shown a noteworthy performance that is much closer to the Exhaustive search, which was simulated over all the available actions and returns the optimal one. In addition, random matching could not achieve good performance because it does not perform any learning.

TABLE 5. Average system sum rate versus heights of UAVs (P_t $\mathcal{D}$ 20 dBm, 4 UAVs, 24 GUs) in bits/s/Hz.

Heights (m)	Quotas	Proposed DϵTS	ϵ-Greedy	Greedy	Random matching
50	2 GUs	**52.28**	51.05	46.01	24.25
	4 GUs	**100.94**	100.08	87.72	54.55
100	2 GUs	**50.27**	48.08	44.57	33.30
	4 GUs	**98.58**	97.45	83.37	69.03
150	2 GUs	**47.54**	46.25	43.57	37.830
	4 GUs	**94.67**	93.35	85.49	75.22
200	2 GUs	**45.18**	44.21	42.04	37.99
	4 GUs	**89.75**	88.82	84.20	74.21

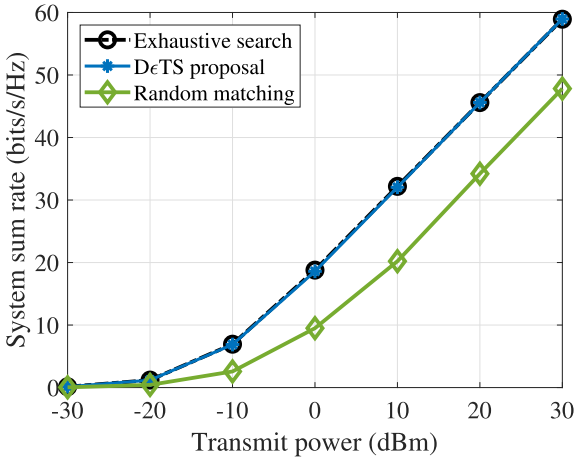


FIGURE 9. System sum rate versus transmit power (4 UAVs, 8 GUs, quota = 2 GUs and height = 100 m for each UAV).

VII. CONCLUSION

In this paper, the system throughput of multi-UAV-assisted NOMA networks has been maximized through joint user associations and pairing with no cooperation among the UAVs. The UAVs with quotas aim to identify the potential GUs that achieve better throughput while the GUs seek the UAVs with better channel conditions. Therefore, an optimization problem was formulated and analysed as an MAB and two-sided matching problem. Moreover, a DϵTS MAB algorithm was proposed and its performance evaluated against other state-of-the-art MAB techniques such as standard Thompson sampling, ϵ-Greedy, and Greedy algorithms. The simulation results are conducted under a realistic channel model. The simulations took into account different numbers of GUs available in the system as well as different quotas for the UAVs. The results consistently demonstrated the superiority of the DϵTS method in achieving faster learning and convergence, minimum system regret as well as the best reward among other MAB techniques. Additionally, the DϵTS achieves a performance much closer to the exhaustive search method and is more suitable when the number of GUs increases.

REFERENCES

[1] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nature Electron.*, vol. 3, no. 1, pp. 20–29, Jan. 2020.

[2] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, J. S. Thompson, E. G. Larsson, M. D. Renzo, W. Tong, P. Zhu, X. Shen, H. V. Poor, and L. Hanzo, "On the road to 6G: Visions, requirements, key technologies and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, Feb. 2023.

[3] K. Zheng, R. Luo, Z. Wang, X. Liu, and Y. Yao, "Short-term and long-term throughput maximization in mobile wireless-powered Internet of Things," *IEEE Internet Things J.*, vol. 11, no. 6, pp. 10575–10591, Mar. 2024.

[4] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul. 2019.

[5] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Netw.*, vol. 35, no. 2, pp. 244–251, Mar. 2021.

[6] D. Zhou, M. Sheng, J. Li, and Z. Han, "Aerospace integrated networks innovation for empowering 6G: A survey and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 975–1019, Feb. 2023.

[7] X. Jiang, M. Sheng, N. Zhao, C. Xing, W. Lu, and X. Wang, "Green UAV communications for 6G: A survey," *Chin. J. Aeronaut.*, vol. 35, no. 9, pp. 19–34, Sep. 2022.

[8] G. Geraci, A. Garcia-Rodriguez, M. M. Azari, A. Lozano, M. Mezzavilla, S. Chatzinotas, Y. Chen, S. Rangan, and M. D. Renzo, "What will the future of UAV cellular communications be? A flight from 5G to 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1304–1335, 3rd Quart., 2022.

[9] Y. Li, S. Xu, Y. Wu, and D. Li, "Network energy-efficiency maximization in UAV-enabled air–ground-integrated deployment," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13209–13222, Aug. 2022.

[10] X. Liu, B. Lai, B. Lin, and V. C. M. Leung, "Joint communication and trajectory optimization for multi-UAV enabled mobile Internet of Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15354–15366, Sep. 2022.

[11] H. Zhang, J. Zhang, and K. Long, "Energy efficiency optimization for NOMA UAV network with imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2798–2809, Dec. 2020.

[12] A. Sayed-Ahmed and M. Elsabrouty, "User selection and power allocation for guaranteed SIC detection in downlink beamforming non-orthogonal multiple access," in *Proc. Wireless Days*, Mar. 2017, pp. 188–193.

[13] M. Sami and J. N. Daigle, "User association and power control for UAV-enabled cellular networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 267–270, Mar. 2020.

[14] K. Wang, Y. Liu, Z. Ding, A. Nallanathan, and M. Peng, "User association and power allocation for multi-cell non-orthogonal multiple access networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5284–5298, Nov. 2019.

[15] C. Xu, G. Zheng, and L. Tang, "Energy-aware user association for NOMA-based mobile edge computing using matching-coalition game," *IEEE Access*, vol. 8, pp. 61943–61955, 2020.

[16] Y. M. Waheidi, M. Jubran, and M. Hussein, "User driven multiclass cell association in 5G HetNets for mobile & IoT devices," *IEEE Access*, vol. 7, pp. 82991–83000, 2019.

[17] L. T. Liu, H. Mania, and M. I. Jordan, "Competing bandits in matching markets," 2019, *arXiv:1906.05363*.

[18] L. T. Liu, F. Ruan, H. Mania, and M. I. Jordan, "Bandit learning in decentralized matching markets," *J. Mach. Learn. Res.*, vol. 22, no. 1, pp. 9612–9645, Jan. 2021.

[19] H. Jia, Y. Wang, Y. Chen, M. Liu, and Z. Li, "Sum rate maximization for multi-UAV enabled space-air-ground wireless powered communication networks," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Mar. 2021, pp. 1–6.

[20] K. Lu, F. Wu, L. Xiao, Y. Liang, and D. Yang, "Sum-rate maximization for UAV-enabled two-way relay systems," *Digit. Commun. Netw.*, vol. 8, no. 6, pp. 1105–1114, Dec. 2022.

[21] B. K. S. Lima, R. Dinis, D. B. da Costa, R. Oliveira, and M. Beko, "User pairing and power allocation for UAV-NOMA systems based on multi-armed bandit framework," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13017–13029, Dec. 2022.

[22] Y. Li, H. Zhang, K. Long, C. Jiang, and M. Guizani, "Joint resource allocation and trajectory optimization with QoS in UAV-based NOMA wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6343–6355, Oct. 2021.

[23] D.-T. Do, A.-T. Le, Y. Liu, and A. Jamalipour, "User grouping and energy harvesting in UAV-NOMA system with AF/DF relaying," *IEEE Trans. Veh. Technol.*, vol. 70, no. 11, pp. 11855–11868, Nov. 2021.

[24] M. N. Kharil, M. S. Johal, F. Idris, and N. Hashim, "UAV-enabled communications using NOMA for 5G and beyond: Research challenges and opportunities," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 31, no. 3, p. 1420, Sep. 2023.

[25] F. Kong, J. Yin, and S. Li, "Thompson sampling for bandit learning in matching markets," 2022, *arXiv:2204.12048*.

[26] Z. Wang, L. Guo, J. Yin, and S. Li, "Bandit learning in many-to-one matching markets," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manag.* New York, NY, USA: Association for Computing Machinery, Oct. 2022, pp. 2088–2097.

[27] B. Uwizeyimana, O. Muta, A. H. Abd El-Malek, M. Abo-Zahhad, and M. Elsabrouty, "A multi-agent multi-armed bandit approach for user pairing in UAV-assisted NOMA-networks," in *Proc. Int. Symp. Netw., Comput. Commun. (ISNCC)*, Oct. 2023, pp. 113–118.

[28] L. Zanzi, V. Sciancalepore, A. Garcia-Saavedra, H. D. Schotten, and X. Costa-Pérez, "LACO: A latency-driven network slicing orchestration in beyond-5G networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 667–682, Jan. 2021.

[29] E. M. Mohamed, S. Hashima, K. Hatano, E. Takimoto, and M. Abdel-Nasser, "Load balancing multi-player MAB approaches for RIS-aided mmWave user association," *IEEE Access*, vol. 11, pp. 15816–15830, 2023.

[30] S. Hashima, K. Hatano, E. Takimoto, and E. Mahmoud Mohamed, "Neighbor discovery and selection in millimeter wave D2D networks using stochastic MAB," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1840–1844, Aug. 2020.

[31] T. Jin, X. Yang, X. Xiao, and P. Xu, "Thompson sampling with less exploration is fast and optimal," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2023, pp. 15239–15261.

[32] A. H. A. El-Malek, M. A. Aboulhassan, A. M. Salhab, and S. A. Zummo, "Performance analysis and optimization of UAV-assisted networks: Single UAV with multiple antennas versus multiple UAVs with single antenna," *IEEE Syst. J.*, vol. 17, no. 3, pp. 3468–3479, Jun. 2023.

[33] Y. Chen, N. Zhao, Z. Ding, and M.-S. Alouini, "Multiple UAVs as relays: Multi-hop single link versus multiple dual-hop links," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6348–6359, Sep. 2018.

[34] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.

[35] B. Lima, N. Fachada, R. Dinis, D. B. D. Costa, and M. Beko, "Uavnoma: A UAV-NOMA network model under non-ideal conditions," *J. Open Res. Softw.*, vol. 10, no. 1, p. 9, 2022.

[36] F. Banaeizadeh, M. Barbeau, J. Garcia-Alfaro, V. S. Kothapalli, and E. Kranakis, "Uplink interference management in cellular-connected UAV networks using multi-armed bandit and NOMA," in *Proc. IEEE Latin-American Conf. Commun. (LATINCOM)*, Nov. 2022, pp. 1–6.

[37] G. J. J. Van den Burg. (2020). *An Exploration of Thompson Sampling*. [Online]. Available: https://gertjanvandenburg.com/blog/thompsonsampling/

[38] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, "Matching theory for future wireless networks: Fundamentals and applications," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 52–59, May 2015.

[39] Y. Li, H. Zhang, K. Long, S. Choi, and A. Nallanathan, "Resource allocation for optimizing energy efficiency in NOMA-based fog UAV wireless networks," *IEEE Netw.*, vol. 34, no. 2, pp. 158–163, Mar. 2020.

[40] G. Qu, A. Xie, S. Liu, J. Zhou, and Z. Sheng, "Reliable data transmission scheduling for UAV-assisted air-to-ground communications," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13787–13792, May 2023.

BONIFACE UWIZEYIMANA (Member, IEEE) received the B.Sc. degree (Hons.) in electronics and telecommunication engineering from the University of Rwanda, Rwanda, in 2019, and the M.Sc. degree in electronics and communication engineering from Egypt-Japan University of Science and Technology (E-JUST), Egypt, in 2024. His research interests include 5G, 6G, and beyond wireless technologies, machine learning for wireless communication, cognitive radio technology, non-terrestrial networks, and digital signal processing.

MOHAMMED ABO-ZAHHAD (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Assiut University (AU), Egypt, in 1979 and 1983, respectively. In 1988, and the Ph.D. degree from the University of Kent at Canterbury, U.K., and AU, Egypt. He was promoted to the rank of professor of wireless communications and multimedia processing in January 1999. He was elected to the position of Vice-Dean for graduate studies from August 2006 to June 2012. Moreover, he was the Chair of the Department of Electrical and Electronics Engineering from November 2013 to December 2016, both in the Faculty of Engineering, AU. During 1989 to 1992, he was with the Department of Telecommunications and Media Informatics, Budapest Technical University as a Research Fellow. From 1996 to 2003, he joined the Department of Electronics at Yarmouk University, Jordan. He was a Former Dean of the Faculty of Electronics, Communications, and Computer Engineering at the Egypt-Japan University of Science and Technology (E-JUST), Egypt, from January 2017 to March 2022. Currently, he is a Professor of Wireless Communications and Multimedia Processing at E-JUST. He also acted as the Director of the Information and Communication Technology Centers at AU and E-JUST. He has published more than 230 papers in international conferences and high-ranked journals. His research interests include biomedical and genomic signal processing, multimedia processing, wireless sensor nodes, massive MIMO and millimeter wave communications, and the Internet of Medical Things (IoMT). He has earned many national and international research awards, among which is the Encouragement State Award in Engineering from the Egyptian Academy of Scientific Research and Technology (ASRT). He is also a member of the National Electronics and Communication Promotion Committee and an accredited reviewer of the National Quality Assurance and Accreditation Authority (NAQQA), Egypt. He served as a referee for IEEE, IET Journals, and Transactions. In addition, he was the chair of the last five rounds (2017–2021) of the international Japan-Africa conference on electronics, communications, and computations (JAC-ECC), organized in cooperation between E-JUST and Kyushu University, Japan. He has also served as an editor of the *Journal of Engineering Sciences* (JES), Egypt, and as an associate editor of the *Journal of Engineering and Applied Science* (JEAS), Springer, and the *Future Internet Journal* published by MDPI.

OSAMU MUTA (Member, IEEE) received the Associate B.E. degree from Sasebo Institute of Technology, in 1994, the B.E. degree from Ehime University, in 1996, the M.E. degree from Kyushu Institute of Technology, in 1998, and the Ph.D. degree from Kyushu University, Japan, in 2001. In 2001, he joined the Graduate School of Information Science and Electrical Engineering, Kyushu University, as an Assistant Professor. From 2010 to 2023, he was an Associate Professor with the Center for Japan-Egypt Cooperation in Science and Technology, Kyushu University. Since 2023, he has been a Professor with the Faculty of Information Science and Electrical Engineering, Kyushu University. His research interests include signal processing techniques for wireless communication and power line communication, MIMO techniques, interference coordination techniques, low-power wide-area networks, and nonlinear distortion compensation techniques for high-power amplifiers. He is a Senior Member of the Institute of Electronics, Information, and Communication Engineering (IEICE). He was a recipient of the 2005 Active Research Award from the IEICE Radio Communication Systems Technical Committee, the Chairperson's Award for Excellent Paper from the IEICE Communication Systems Technical Committee, in 2014, 2015, and 2017, the 2020 IEICE Communications Society Best Paper Award, and the International Symposium on Computing and Networking 2022 (CANDAR-22) Best Paper Award.

AHMED H. ABD EL-MALEK (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Alexandria University, Egypt, in 2007 and 2010, respectively, and the Ph.D. degree from the Electrical Engineering Department, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia, in 2016. He is currently an Associate Professor with the Electronics and Communications Engineering Department, Egypt-Japan University of Science and Technology. His research interests include cognitive radio, design and analysis of wireless networks, network coding, physical layer security, and interference cancellation.

MAHA M. ELSABROUTY (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electrical, electronics, and communication engineering from Cairo University, Egypt, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Ottawa. She is currently a Professor of wireless communications and signal processing with the Department of Electronics and Communications Engineering, Egypt-Japan University of Science and Technology (E-JUST) University. Her current research interests include massive MIMO techniques, interference management in HetNets, cognitive radio, intelligent techniques for wireless communications, and green communication systems.