Original Article

Check for updates

Enhancing unmanned aerial vehicle communication through distributed ledger and multi-agent deep reinforcement learning for fairness and scalability

Farman Ali[1,2], Muhammad Ahtasham [2], Zahra Anfaal[2]

[1]School of Electronics and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, Jiangsu, China.
[2]Department of Electrical Engineering, Qurtuba University of Science and IT, Dera Ismail Khan 29050, Pakistan.

Correspondence to: Dr. Farman Ali, Department of Electrical Engineering, Qurtuba University of Science and IT, Dera Ismail Khan 29050, Pakistan. E-mail: drfarmanali.optics@qurtuba.edu.pk

Abstract

Unmanned Aerial Vehicles (UAVs) are pivotal in enhancing connectivity in diverse applications such as search and rescue, remote communications, and battlefield networking, especially in environments lacking ground-based infrastructure. This paper introduces a novel approach that harnesses Multi-Agent Deep Reinforcement Learning to optimize UAV communication systems. The methodology, centered on the Independent Proximal Policy Optimization technique, significantly improves fairness, throughput, and energy efficiency by enabling UAVs to autonomously adapt their operational strategies based on real-time environmental data and individual performance metrics. Moreover, the integration of Distributed Ledger Technologies with Multi-Agent Deep Reinforcement Learning enhances the security and scalability of UAV communications, ensuring robustness against disruptions and adversarial attacks. Extensive simulations demonstrate that this approach surpasses existing benchmarks in critical performance metrics, highlighting its potential implications for future UAV-assisted communication networks. By focusing on these technological advancements, the groundwork is laid for more efficient, fair, and resilient UAV systems.

Keywords: Unmanned aerial vehicles-to-ground communication, multi-agent deep reinforcement learning, fairness and throughput, distributed ledger technologies

1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) have significantly revolutionized communication in different areas, including wireless sensor networks (WSN), cellular networks, the Internet of Things (IoT), and Space-Air-Ground Integrated Networks (SAGIN). These flexible platforms facilitate the rapid implementation of communication services to users on the ground in situations when terrestrial infrastructure is not accessible or has been compromised, including in disaster-impacted areas or conflict-ridden battlefields. UAVs are versatile and can be used for various purposes such as improving communication coverage, serving as mobile relays, enabling edge computing, and performing data collection. These applications have advanced UAV-assisted technology to the forefront of research in wireless communications and network sectors[1–4]. The benefits of UAVs are distinctive and can be credited to several significant advancements. Advancements in industrial technology have enabled the downsizing of electronic equipment, enhancing their capacity and facilitating the integration of more advanced modules on UAVs at a lower cost. Furthermore, the great mobility of UAVs allows them to be deployed in difficult terrains such as mountains and rivers, where setting up ground-based infrastructure is not feasible. This guarantees thorough and smooth communication capacities. Thirdly, UAVs provide excellent visibility for ground communication systems, minimizing signal route loss caused by barriers and improving Line-of-Sight (LoS) connections. In recent years, substantial research has been dedicated to enhancing the control systems of UAVs to address the challenges associated with their deployment. Notable among these advancements is the development of distributed adaptive fuzzy formation control for multiple UAVs, which can handle uncertainties and actuator faults while operating under switching topologies. This method utilizes fuzzy logic to adaptively manage the formation of UAVs, ensuring robust performance despite the presence of system uncertainties and potential faults[5]. Additionally, neural adaptive distributed formation control has emerged as a significant approach for managing nonlinear multi-UAV systems with unmodeled dynamics. By leveraging neural networks, this control strategy can adapt to complex, nonlinear interactions within the UAV network, ensuring stable formation control even when the system dynamics are not fully known or are subject to change. These neural adaptive methods provide a high degree of flexibility and robustness, making them suitable for dynamic and uncertain environments[6]. Furthermore, other advanced control techniques, such as Model Predictive Control (MPC) and Reinforcement Learning (RL), have been applied to UAV systems to enhance their autonomous capabilities. MPC allows UAVs to predict and optimize their trajectories based on future states, ensuring efficient navigation and collision avoidance. RL, on the other hand, enables UAVs to learn optimal control policies through interaction with their environment, adapting to new scenarios and improving performance over time[7,8]. These advancements in UAV control systems are critical for enabling UAVs to operate autonomously and efficiently in various applications, from disaster response to commercial delivery services. The integration of these sophisticated control techniques ensures that UAVs can maintain stable formations, handle dynamic changes, and operate reliably even in the presence of uncertainties and external disturbances.

1.1. Related work

Cutting-edge research is being conducted on UAV-assisted communication, fairness, and energy efficiency in wireless networks. These studies are aimed at devising new and inventive solutions and strategies to enhance the performance, reliability, and security of UAV communication systems. The[5] investigated the utilization of UAVs as aerial base stations for Ground Users (GUs) and proposes cooperative jamming by UAV jammers to counter ground eavesdroppers. It leveraged Multi-Agent Deep Reinforcement Learning (MADRL) to optimize UAV trajectories, transmit power, and jamming power. The research[6] has focused on optimizing UAV trajectories, user association, and GUs' transmit power to achieve fairness-weighted throughput.The research has emphasized the importance of balancing fairness and throughput in UAV-Base Stations (BSs)-assisted communication and introduces the UAV-Assisted Fair Communication (UAFC) algorithm based on multi-agent deep reinforcement learning[7]. Moreover, it proposes sharing neural networks to reduce decision-making uncertainty. The authors in have explored energy-efficient UAV trajectories for uplink communication and employ reinforcement learning for load balancing[8]. The challenge of secure communication in the presence

of multiple eavesdroppers is addressed in[9], and the utilization of UAV jammers and artificial noise signals is proposed. The study[10] considered fairness alongside coverage and throughput, introducing the UAFC algorithm for fair throughput optimization. Additionally, the work proposed the sharing of neural networks for distributed decision-making. The study investigated the use of UAVs as mobile relays between GUs and a macro base station, utilizing reinforcement learning for trajectory optimization[11]. The authors in[12] addressed wiretapping by ground eavesdroppers and propose cooperative jamming and multi-agent reinforcement learning as countermeasures. The study[13] has focused on the investigation of UAV-BSs for efficient wireless communication, balancing coverage, throughput, and fairness. A MADRL approach for UAV-BSs serving GUs is proposed in[14], considering fair throughput, coverage, and flight status. Additionally, in the context of integrating UAVs into 6G networks, several key areas of innovation and study have been identified, such as the exploration of UAV capabilities, base station offloading, emergency response, intelligent telecommunication, and mobile edge computing[15–18]. In addition to the current body of knowledge on UAV communication systems and their optimization through deep learning and distributed ledger (DL) technologies, it is pertinent to consider advancements in multi-agent network applications. One notable study in this domain is mentioned in[19]. This research addresses the challenges in multi-agent networks where the ability to perform collective activities is crucial. The study investigates the containment and consensus tracking problems within a network of continuous-time agents characterized by state constraints. These endeavors aim to address the challenges and opportunities associated with UAV-assisted communications, contributing to the development of efficient, secure, and scalable UAV communication systems. Several limitations exist in the current works related to UAV-to-ground communication. These limitations include fairness, complex structure, reliability, security, privacy, and scalability. Furthermore, there are other unresolved issues and limitations in the field of UAV-assisted communications that require additional research. A concern is the lack of emphasis on supporting multiple users in emergency communication situations, requiring equitable service delivery. Innovative methods are required to ensure effective cooperation and connectivity among UAVs that can maintain UAV connectivity without relying on fixed infrastructure. Centralized methodologies now in use pose notable scalability and complexity issues, highlighting the necessity for more adaptable and scalable methods in UAV communication systems. As a result, we have identified a significant research gap that needs to be addressed. To enhance UAV communication systems, we have explored the potential of DL technology and MADRL.

1.2. Major contribution
The major contribution of this paper is summarized as follows.

1. The paper proposed a new system model that utilizes multiple UAVs to provide fair communication services to ground users without ground-based stations. Our approach addresses the challenges of UAV network connectivity and equitable throughput by optimizing fair throughput and energy efficiency.

2. A MADRL framework for UAV communication scenarios is used, with an Independent Proximal Policy Optimization (IPPO) technique. This enables decentralized learning for individual observations, promoting a more in-depth investigation of tactics. The reward system encourages fairness and energy efficiency.

3. The challenges posed by limited load capacities and energy resources in UAVs are addressed through the utilization of MADRL and Distributed Ledger Technologies (DLT). The MADRL framework, specifically employing the IPPO technique, enables UAVs to optimize their energy usage and load management by making efficient, decentralized decisions based on real-time observations. The reward function within the MADRL framework is designed to incentivize energy efficiency, ensuring that UAVs conserve energy during prolonged missions. Additionally, the integration of DLT ensures secure and efficient data management, thereby reducing computational and communication overhead, which, in turn, aids in conserving energy and effectively managing load capacities.
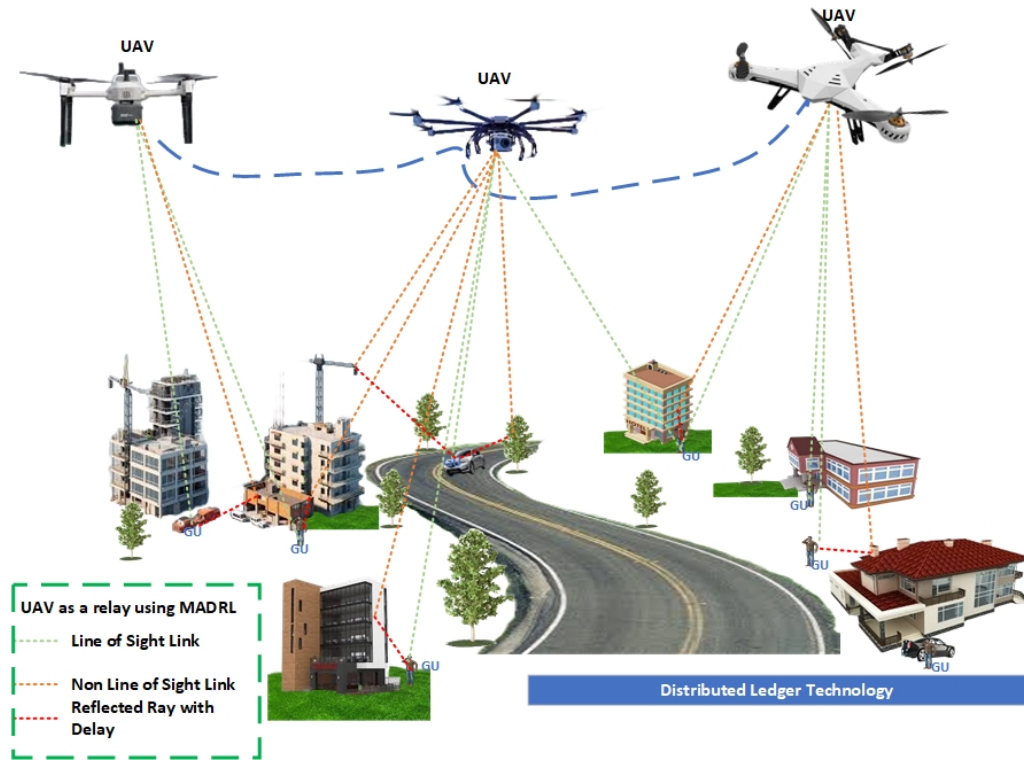
Figure 1. Proposed system model using integrated MADRL and distributed ledger techniques.

4. The MADRL-based solution incorporating DL technology has been rigorously evaluated through experimental validation and compared against traditional benchmarks. The results demonstrate superior performance in terms of fairness, throughput, and energy efficiency. Furthermore, the actual outcomes highlight notable enhancements in scalability, security, and fairness for UAV communications. This presents a substantial advancement in the field, effectively showcasing the technology's capacity to establish new standards for UAV-assisted communication systems.

The rest part of the paper is organized as follows. Section 2 develops the system model; problem formulation and objective function is explained in Section 3; Section 4 discusses the combined features of MADRL DL. Results and discussion are explored in Section 5, and finally, Section 6 provides the conclusion.

2. SYSTEM MODEL

The framework illustrated in Figure 1 presents an advanced system for communication that consists of a group of UAVs, each equipped with DLT and MADRL capabilities. The purpose of this design is to overcome the limitations of traditional communication networks, particularly in areas where conventional infrastructure is insufficient or degraded. In this innovative network, UAVs act as independent aerial relay stations, forming a robust mesh network to provide connectivity to GUs across different terrains. Each UAV m from the fleet M has a position $\mathbf{W}_{u_m}(t)$ at time t, which includes altitude H. The movement of UAVs between time slots is limited by their maximum speed V_{max} and the time slot duration δ[20,21].

$$\|\mathbf{W}_{u_m}(t) - \mathbf{W}_{u_m}(t-1)\|_2 \leq (V_{max} \cdot \delta), \quad \forall m \in M, t \in T \tag{1}$$

The UAVs manage communication links with GUs using a Time Division Multiple Access (TDMA) scheme. The binary indicator variables $\alpha_{src_{m,k}}(t)$ and $\beta_{dst_{m,k}}(t)$ denote whether a source k_{src} or destination k_{dst} GU is connected to UAV m during time slot t [22,23].

$$\alpha_{src_{m,k}}(t), \beta_{dst_{m,k}}(t) \in \{0,1\}, \quad \forall m \in M, k \in K, t \in T \tag{2}$$

The constraints C1 and C2 for access control are expressed as

$$\sum_{m \in M} \alpha_{src_{m,k}}(t) \le 1, \quad \forall k \in K, t \in T \tag{3}$$

$$\sum_{m \in M} \beta_{dst_{m,k}}(t) \le 1, \quad \forall k \in K, t \in T \tag{4}$$

The integration of DLT ensures a secure and unchangeable exchange of data, while MADRL enables UAVs to make informed decisions for enhancing the network's performance. In MADRL, the Q-function is used to measure the effectiveness of a particular action taken in a specific state, represented as $\Theta(s, a)$, where s denotes the state and a represents the action. The primary goal is to optimize the policy parameters Θ such that the total future rewards are maximized. This optimization process involves using an update function U, which modifies Θ iteratively according to a specific update rule.

$$U(\Theta|s, a) = \Theta(s, a) + \Delta\Theta \tag{5}$$

The change $\Delta\Theta$ in the policy parameters is defined as

$$\Delta\Theta = \alpha \cdot \left(\mathcal{R}(s, a) + \gamma \cdot \max_{a' \sim \pi(\cdot|s')} \Theta(s', a') - \Theta(s, a) \right) \tag{6}$$

In this particular formulation, the learning rate α represents the extent to which new information overrides old information. The function $\mathcal{R}(s, a)$ provides the immediate reward after taking an action a in a particular state s. The discount factor γ balances the significance of immediate and future rewards. The policy $\pi(\cdot|s')$, which can potentially be an ϵ-greedy strategy, guides the selection of the subsequent action a' in the next state s', and the operation $\max_{a' \sim \pi(\cdot|s')}$ identifies the action that maximizes the expected Q-value given the policy π and the new state s'. By means of this mechanism, the algorithm gradually approaches an optimal policy that prescribes the best action to take in any given state to maximize long-term rewards. Dual transmitters on UAVs and Orthogonal Frequency-Division Multiple Access (OFDMA) can reduce interference with a factor η that mitigates its impact in UAV-to-UAV communication.

$$I_{total} = \sum_{i=1}^{N} \eta \cdot I_i \tag{7}$$

where I_{total} represents the overall interference experienced by a GU. The value of N refers to the total number of interfering signals, and the value of I_i is the interference generated by the $i - th$ signal. The term η (where $0 < \eta < 1$) indicates the effectiveness of interference reduction techniques.

2.1. UAV to UAV channel

In the system model illustrated in Figure 1, the communication channel between UAVs plays a crucial role in ensuring efficient data transfer in scenarios where ground infrastructure is lacking or non-existent, such as in search and rescue missions or in remote communication setups. This model addresses the challenges of providing equal access and uninterrupted connectivity among UAVs, while also taking into account the

limitations imposed by their load capacities and energy resources. The channel model is based on the principles of free space propagation, adapted to the dynamic and versatile nature of UAVs enabled by MADRL[24–26].

The channel gain or path loss between two UAVs, m_i and m_j, at a particular time slot t, is characterized by the following equation, which adheres to the free space propagation model.

$$g_{m_i,m_j}(t) = \frac{p_0}{\|\mathbf{W}_{m_i}(t) - \mathbf{W}_{m_j}(t)\|^2} \tag{8}$$

where p_0 is the channel gain at a reference distance of one meter under ideal conditions, $\mathbf{W}_{m_i}(t)$ and $\mathbf{W}_{m_j}(t)$ represent the position vectors of UAV m_i and m_j at time slot t, respectively. The expression $\|\mathbf{W}_{m_i}(t) - \mathbf{W}_{m_j}(t)\|$ stands for the Euclidean distance between two UAVs at time t, which reflects the spatial connection impacting signal quality. The signal-to-noise ratio (SNR) at UAV m_j caused by the transmission from UAV m_i is determined to handle communication links and potential interference when many UAVs are operating close to each other.

$$\gamma_{m_i,m_j}(t) = \frac{P_{max} \cdot g_{m_i,m_j}(t)}{n_0 \cdot b_u} \tag{9}$$

where P_{max} is the maximum transmit power of the UAV, determining the strength of the transmitted signal; n_0 is the noise spectral density, reflecting the background noise level that can impact signal reception and b_u signifies the bandwidth allocated for the UAV-to-UAV communication link, which influences the data rate.

To maintain the quality of the communication link, the communication strategy must keep the SNR above a predetermined threshold. This creates a communication range R_c that serves as a key distance for UAVs to communicate efficiently. When the distance between two UAVs exceeds R_c, signal quality may drop, affecting the reliability of the communication link.

Using MADRL, each UAV independently changes its position and communication settings to optimize the network's total data transfer rate while ensuring equal treatment of all GUs being serviced. The MADRL architecture, via its IPPO technique, enables individual observation-based learning to tackle the non-convex nature and hybrid variable difficulties found in traditional systems. Integrating DLT into this architecture enhances security and immutability by openly and permanently recording all communication transactions.

2.1.1 Handling NLoS conditions
In urban environments where obstacles are prevalent, Non-Line-of-Sight (NLoS) conditions are common and significantly influence UAV-to-UAV communication. To address this, our model incorporates both LoS and NLoS conditions by adapting the path loss model accordingly. The probability of LoS $P_{LoS_{m_i,m_j}}$ and NLoS $P_{NLoS_{m_i,m_j}}$ conditions is modeled based on environmental factors such as building density and height. The average path loss $PL_{m_i,m_j}(t)$ considering both LoS and NLoS conditions is given by

$$PL_{m_i,m_j}(t) = P_{LoS_{m_i,m_j}} \times PL_{LoS_{m_i,m_j}} + P_{NLoS_{m_i,m_j}} \times PL_{NLoS_{m_i,m_j}} \tag{10}$$

where $P_{LoS_{m_i,m_j}}$ is the probability of LoS condition, $P_{NLoS_{m_i,m_j}} = 1 - P_{LoS_{m_i,m_j}}$ is the probability of NLoS condition, $PL_{LoS_{m_i,m_j}}$ is the path loss under LoS condition, and $PL_{NLoS_{m_i,m_j}}$ is the path loss under NLoS condition.

The path loss values for LoS and NLoS conditions are calculated based on empirical models suitable for urban environments. The LoS probability $P_{LoS_{m_i,m_j}}$ is influenced by the elevation angle θ_{m_i,m_j} and the density of obstacles, modeled as:

$$P_{LoS_{m_i,m_j}} = \frac{1}{1 + a\exp(-b(\theta_{m_i,m_j} - a))} \tag{11}$$

where a and b are constants that depend on the urban environment.

2.2. UAV-to-ground channel

The UAV-to-ground channel in the proposed model includes the probability of LoS communication, which is essential for signal strength and reliability and is influenced by the environment. Urban environments with many buildings tend to have a lower LoS probability compared to open rural areas. The probability is modeled as

$$P_{LoS_{k(\cdot),m(t)}} = \frac{1}{1 + a\exp(-b(\theta_{k(\cdot),m[t]} - a))} \tag{12}$$

where a and b are environment-dependent constants, and θ represents the elevation angle between the GU and the UAV. The average path loss considers both LoS and NLoS conditions, and is calculated as

$$PL_{k(\cdot),m(t)} = P_{LoS_{k(\cdot),m(t)}} \times PL_{LoS_{k(\cdot),m(t)}} + (1 - P_{LoS_{k(\cdot),m(t)}}) \times PL_{NLoS_{k(\cdot),m(t)}} \tag{13}$$

The path loss for both LoS and NLoS scenarios is calculated based on the carrier frequency f_1, the distance $d_{k(\cdot),m(t)}$, and the speed of light c. The average path losses for LoS and NLoS situations are denoted by η_{LoS} and η_{NLoS} correspondingly.

$$PL_{LoS_{k(\cdot),m(t)}} = 20\log\left(\frac{4\pi f_1 d_{k(\cdot),m(t)}}{c}\right) + \eta_{LoS} \tag{14}$$

$$PL_{NLoS_{k(\cdot),m(t)}} = 20\log\left(\frac{4\pi f_1 d_{k(\cdot),m(t)}}{c}\right) + \eta_{NLoS} \tag{15}$$

The SNR for the uplink from GU to UAV and the signal-to-noise interference ratio (SINR) for the downlink from UAV to GU are computed considering the transmit power of the GU (p_G), the channel gain (h), the bandwidth allocated to each GU (b_u), and the interference from other UAVs. The SNR for uplink from GU to UAV is defined as

$$\gamma_{kG2U_{src,m1(t)}} = \frac{p_G \cdot h_{k_{src},m1(t)}}{n_0 \cdot b_u} \tag{16}$$

where $h_{k_{src},m1(t)} = 10^{-PL_{k_{src},m(t)}/10}$ is the channel gain.

The SINR for downlink from UAV to GU takes into account the interference from other UAVs.

$$\gamma_{U2G_{m2,k_{dst}(t)}} = \frac{p_U \cdot h_{m2,k_{dst}(t)}}{I_{m2,k_{dst}(t)} + n_0 \cdot b_u} \tag{17}$$

The transmission rate for uplink and downlink is calculated using the logarithmic function of $1 + \gamma$, where γ is the SNR or SINR, as appropriate.

It is assumed that the bottleneck in transmission rate is either on the uplink or downlink, not on the UAV to UAV link, due to high-quality links between UAVs. The instant transmission rate between the kth pair is determined by the minimum of the uplink or downlink transmission rates, considering the hovering time ($\tau_{m_{hover}(t)}$) and the ratio of source (q_1) and destination (q_2) allocation factors.

3. PROBLEM FORMULATION AND OBJECTIVE FUNCTION

We define the accumulative throughput for the k-th user pair at time slot t as

$$TP_k(t) = \sum_{r=1}^{t} r_k(r) \tag{18}$$

Here, $r_k(r)$ represents the throughput for user pair k at time slot r.

Subsequently, the sum accumulative throughput for all user pairs is given by

$$SUM\_TP(t) = \sum_{k \in K} TP_k(t) \tag{19}$$

To ensure fairness, we utilize Jain's index [26] to measure fair throughput

$$f(t) = \frac{(\sum_{k \in K} TP_k(t))^2}{K \sum_{k \in K} (TP_k(t))^2} \tag{20}$$

The parameter e represents the energy consumed by each UAV m at each time slot t. where the fairness index $f(t)$ ranges between $\frac{1}{K}$ and 1.

The access policies of GUs and the locations of UAVs at time slot t are represented by $A = \{\alpha_{m,k}^{src}(t), \beta_{m,k}^{dst}(t)\}_{m \in M, k \in K, t \in T}$ and $W = \{w_u^m(t)\}_{m \in M, t \in T}$, respectively.

Our objective function aims to maximize

$$S_{max} = \frac{\sum_{t \in T} f(t) \cdot SUM\_TP(t)}{\sum_{t \in T} \sum_{m \in M} e_m[(t)} \tag{21}$$

Subject to constraints (1)–(4), where $C1$ enforces a safe distance d_{safe} between any two UAVs to avoid collisions, for any m, n in M and t in T, and $C2$ secures the network's algebraic connectedness with a positive measure λ_2. The target function aims to enhance equity and efficiency, as well as decrease the energy usage of all UAVs, in accordance with system needs. Constraint $C1$ controls the spatial mobility of UAVs, while constraints (2)–(4) manage the access control for GUs. Constraint $C2$ is based on the second smallest eigenvalue of the Laplacian matrix to guarantee that the UAVs sustain a linked network topology.

4. COMBINED MADRL AND DISTRIBUTED LEDGER

The IPPO method improves trajectory optimization for many UAVs by employing a decentralized approach. Every UAV functions within a partially observable Markov Decision Process (MDP) framework, making decisions based on its own observations rather than a common global state. The system's state is determined by

aggregating all UAV observations, while actions are decided independently in each location.

1. **Observation Space** Every UAV has a specific observation area that contains crucial environmental data, including the locations of all UAVs and GUs, total throughput, satisfaction levels, and access regulations. This extensive observation space is created to avoid execution problems in simulation and acts as a placeholder because of the fixed input dimensions needed by neural networks. Let O_i represent the observation space for UAV i. The information can be expressed as a vector containing environmental data.

$$O_i = [\mathbf{p}_1, \ldots, \mathbf{p}_N, \mathbf{g}_1, \ldots, \mathbf{g}_M, \tau, \sigma, \alpha] \tag{22}$$

where $\mathbf{p}_j$ is the location of UAV j, $\mathbf{g}_k$ is the position of GU k, τ is the accumulative throughput, σ is satisfaction levels, and α is access policies.

2. **Action Space** A UAV's action space is defined by the direction and distance of movement. The collective action space for all UAVs is the combination of individual actions, with a dimensionality equal to twice the number of UAVs. The action space for UAV i, represented as $\mathcal{A}_i$, comprises movement direction θ_i and distance d_i. The communal action space $\mathcal{A}$ is determined by

$$\mathcal{A} = \bigtimes_{i=1}^{N} \mathcal{A}_i = \bigtimes_{i=1}^{N} (\theta_i, d_i) \tag{23}$$

where N is the number of UAVs.

3. **Reward Function** The incentive function integrates a cooperative element shared by all UAVs and an individual element that encompasses penalties for border breaches, unsafe distances, connectivity problems, and energy usage. The incentive function design ensures that UAVs work together to optimize fair throughput while minimizing energy consumption and penalty infractions. The reward function for UAV i is denoted as R_i.

$$R_i(O_i, \mathcal{A}_i) = w_c C + w_i I - w_b B - w_d D - w_e E \tag{24}$$

where C is the cooperative component, I is the individual throughput, B represents border violation penalties, D represents unsafe distance penalties, E represents energy consumption, and w_c, w_i, w_b, w_d, w_e are the corresponding weights.

4. **Actor-Critic Networks** This methodology maintains distinct actor-critic networks for each UAV, updating them exclusively based on individual observations, in contrast to classic centralized training and exploration (CTDE) systems that could restrict the diversity of agent policies. This encourages a wider range of actions and decreases policy similarity among actors. Each UAV i has actor π_i and critic Q_i networks, which are defined by parameters θ_i and ϕ_i correspondingly.

$$\pi_i(a|O_i; \theta_i) \tag{25}$$

$$Q_i(O_i, a; \phi_i) \tag{26}$$

The networks are updated using gradients from the collected experiences:

$$\Delta\theta_i \propto \nabla_{\theta_i} \log \pi_i(a|O_i; \theta_i) A_i \tag{27}$$

$$\Delta\phi_i \propto \nabla_{\phi_i}(R_i - Q_i(O_i, a; \phi_i))^2 \tag{28}$$

where A_i is the advantage function for UAV i.

5. Algorithms The training algorithm requires UAVs to observe the state, make judgments using the actor network, execute actions, receive rewards, and update their networks with knowledge stored in memory. IPPO differs from MAPPO in that it utilizes only the buffered information of each agent during the update phase, instead of aggregating information from all agents. The IPPO algorithm is defined by the following iterative phases.

Algorithm
- Observe state O_i for UAV i.
- Choose action a_i using actor network: $a_i \sim \pi_i(\cdot|O_i; \theta_i)$.
- Execute action a_i and observe reward r_i and new state O_i'.
- Store transition (O_i, a_i, r_i, O_i') in memory.
- Sample random mini-batch from memory and update θ_i and ϕ_i using stochastic gradient descent.

The key distinction of IPPO in the update phase is in the use of the experiences:

$$\text{For each UAV } i : \theta_i \leftarrow \theta_i + \eta\Delta\theta_i(O_i, a_i, R_i, O_i') \tag{29}$$

where η is the learning rate.

4.1. Computational complexity analysis

The computational complexity of the proposed MADRL algorithm with the IPPO technique involves several components, including observation space processing, action space exploration, reward calculation, and the update of actor-critic networks.

1. Observation Space Processing: Each UAV has an observation space of size $O(N + M)$, where N is the number of UAVs and M is the number of GUs. This space includes the positions, battery levels, and other relevant state information of all UAVs and GUs. The processing of the observation space, which involves gathering and updating this information, is $O(N + M)$ per UAV.

2. Action Space Exploration: The action space for each UAV includes movement direction and distance. The exploration of the action space, which involves selecting an action based on the current policy, is $O(1)$ per UAV since the action space's dimensionality remains constant regardless of the number of UAVs and GUs.

3. **Reward Calculation:** The reward function involves cooperative components, individual throughput, border violations, unsafe distances, and energy consumption. These calculations involve linear operations on the observations and actions. Therefore, the complexity for calculating the reward function is $O(N + M)$ per UAV.

4. **Update of Actor-Critic Networks:** The update of the actor and critic networks involves backpropagation through neural networks. Let L be the number of layers and D be the maximum number of neurons per layer. The complexity of updating each network is $O(LD^2)$. Given that each UAV has its separate networks, the total complexity for updating all UAVs is $O(NLD^2)$.

Combining all components, the overall computational complexity per time step per UAV is written as

$$O(N + M) + O(1) + O(N + M) + O(LD^2) = O(N + M) + O(LD^2) \tag{30}$$

Given that the number of layers L and the number of neurons D per layer are typically constants for a fixed neural network architecture, the primary factors influencing computational complexity are the number of UAVs N and the number of GUs M. In scenarios with a large number of UAVs and GUs, the complexity scales linearly with $N + M$. However, the quadratic term LD^2 related to the neural network updates can become significant if the network architecture is very deep or wide. Thus, the overall computational complexity of the MADRL algorithm with the IPPO technique for the entire system (considering N UAVs) is:

$$O(N \cdot (N + M + LD^2)) = O(N^2 + NM + NLD^2) \tag{31}$$

5. RESULTS AND DISCUSSION

UAV-assisted communication network simulation is a technology that aims to connect locations without ground-based infrastructure. It uses a 10 km x 10 km 3D simulated environment to replicate both rural and urban terrains. UAVs fly between 100 and 300 meters to communicate directly with ground users for optimal transmission. To simulate real-world user dispersal, 50 GUs are randomly dispersed over the simulation territory. The MADRL algorithm is used to balance current and future rewards with a learning rate of 0.01 and a discount factor of 0.95. The communication technicalities include a noise spectral density of -174 dBm/Hz and a transmit power of 20 dBm for UAVs and 23 dBm for GUs. The implementation details of the IPPO algorithm describe the MADRL model: Each UAV agent's policy and value networks are updated depending on observations and rewards to imitate learning. Each UAV contains a blockchain module that imitates DL technology for secure and immutable communication transactions. The networks have densely linked layers with Restricted Linear Unit (ReLU) activation and a softmax layer for action probability. The observation space contains important data such as the UAV's location, battery level, adjacent GUs, and network throughput. Specific actions describe UAV direction and distance motions for the next time slot, defining the action space. The incentive function encourages GUs to improve throughput, energy efficiency, and fairness.

The simulation results depicted in Figure 2 illustrate the comparative performance of four UAV communication strategies across a sequence of time slots. The primary strategy, MADRL with DL, is shown to achieve the highest throughput, a testament to the efficacy of integrating DL technology with advanced RL algorithms. This strategy exhibits a rapid ascent in throughput that stabilizes as it approaches the system's capacity limit, forming an S-shaped logistic curve that represents a realistic growth pattern in network throughput. The second strategy, standalone MADRL, although effective, does not reach the peak performance achieved by its
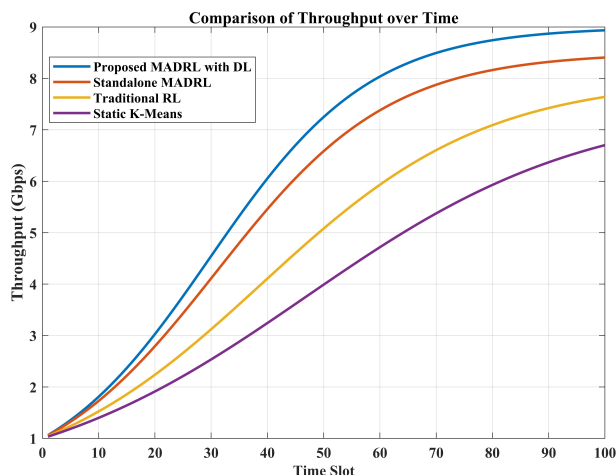
Figure 2. Comparison investigation proposed framework based on MADRL with DL, standalone MADRL, traditional RL and static K-means.

DL-enhanced counterpart, suggesting that while MADRL is a robust approach, its integration with DL technologies offers significant improvements. This curve parallels the top performer at a slightly subdued growth rate, demonstrating substantial but not maximal efficiency. The traditional RL strategy curve rises at a more modest pace, reflecting a lower growth rate and capacity. This indicates that traditional RL, while capable, falls short of the more sophisticated MADRL techniques, particularly in environments that demand dynamic and complex decision-making. Lastly, the Static K-Means strategy lags behind, with the least growth rate and capacity, suggesting its relative inadequacy in adapting to the evolving demands of UAV communication networks. Its performance trajectory, while still positive, is the most gradual and plateaus at the lowest throughput level, underscoring the limitations of less dynamic optimization methods. The results collectively encapsulate the cumulative throughput achievements of the UAV networks over time, measured in Gbps.

Figure 3 presents a comparison of four UAV communication optimization techniques across four metrics, including cumulative rewards, fairness index, cumulative throughput, and energy consumption. After 5000 episodes, the proposed MADRL with DL approach demonstrates the highest rewards in the cumulative rewards graph, indicating its superior performance in accumulating benefits over time. The data analysis shows a steady and significant rise in rewards, suggesting that this strategy is highly effective in achieving the desired outcomes in the simulated scenario. The fairness index graph evaluates the fair distribution of resources among users or agents. The proposed MADRL with DL consistently exhibits good performance, with a fairness index close to one, indicating optimal fairness. This method effectively distributes resources in a fair manner, guaranteeing that no individual or group is given special treatment. The cumulative throughput graph demonstrates the amount of data successfully transmitted over the network. The proposed MADRL with DL approach achieves the highest throughput, indicating efficient network traffic management. This technique shows a consistent and substantial increase in throughput, indicating an effective optimization strategy for maximizing data flow in the network. The energy consumption graph compares the energy needed by each technique to achieve its objectives. The proposed MADRL with DL technique is notable for its minimal energy usage, emphasizing its effectiveness. It is essential to focus on energy efficiency for UAV operations to ensure sustainability and cost-effectiveness. The proposed technique is designed to be high-performing and energy-efficient. The proposed MADRL with DL approach outperforms the standalone MADRL, traditional RL, and static K-Means strategies in all performance metrics. These findings highlight the advantages of incorporating deep learning with DL technology in UAV communication systems, resulting in improved performance, fairness, throughput, and energy efficiency. Utilizing effective analysis rather than relying on a curve can highlight the enhanced benefits of using advanced algorithms to oversee intricate communication networks.
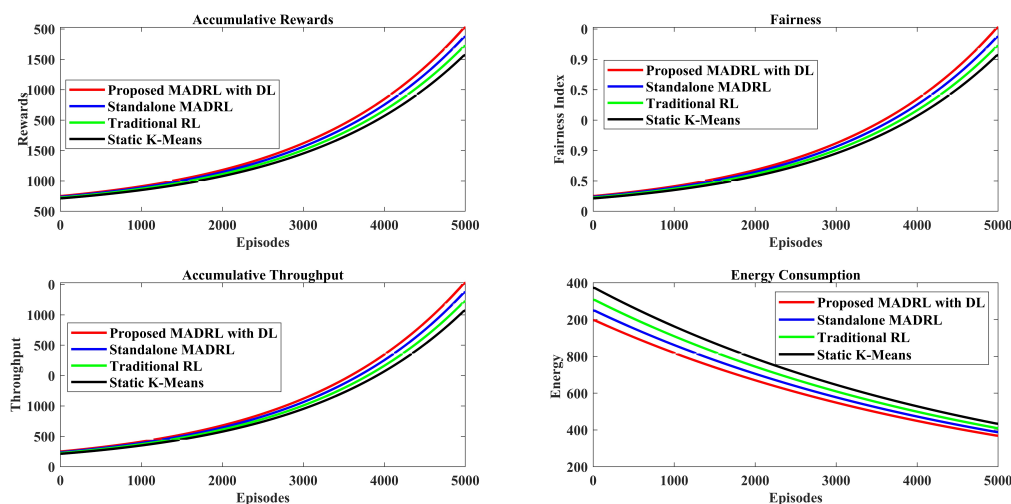
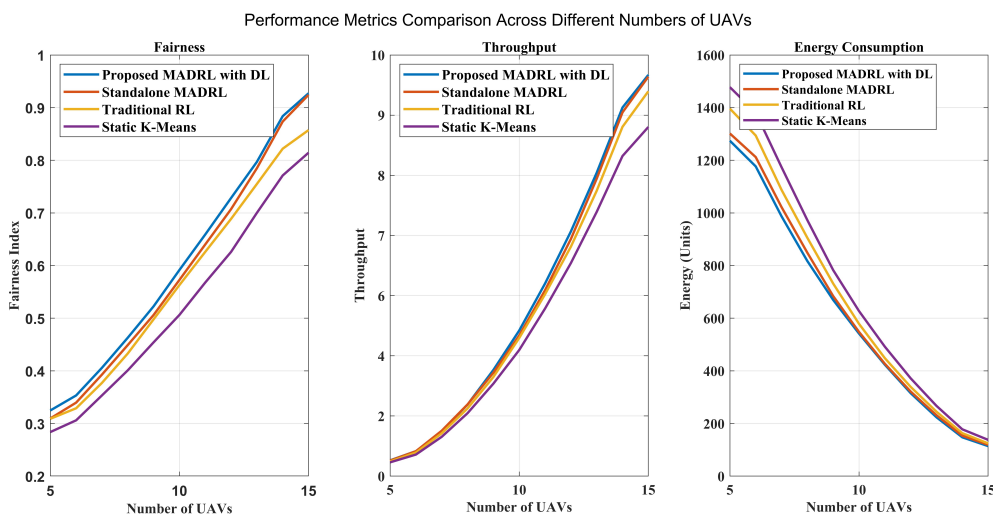**Figure 3.** Throughput, rewards, fairness index and energy estimations.

Figure 4. Number of UAVs evaluation in terms of energy, throughput and fairness index.

Figure 4 compares the performance of various communication mechanisms using different numbers of UAVs. The methods evaluated are the proposed MADRL with DL, Standalone MADRL, Traditional RL, and Static K-Means. The fairness index is used to assess the equitable allocation of resources among UAVs or network users. The MADRL with DL proposal shows the highest fairness index, indicating its effectiveness in ensuring equitable conditions throughout the network. As the number of UAVs increases, fairness grows, demonstrating the effectiveness of the proposed MADRL with Deep Learning as the UAV fleet size scales up. The Standalone MADRL has a higher fairness index compared to the Traditional RL and Static K-Means, indicating a more equitable distribution of resources. The throughput represents the data transmission rate controlled by each approach. The graph shows that as the number of UAVs increases, all techniques enhance throughput, but the Proposed MADRL with DL performs better than the other methods, showcasing its superior capacity to efficiently manage more intricate connections and larger data transfers. The Standalone MADRL outperforms Traditional RL, while Static K-Means has the lowest throughput, indicating that advanced dynamic techniques can effectively utilize more UAVs to enhance network capacity. Similarly, energy consumption shows the

energy usage for each technique. Reduced energy usage is favorable as it indicates a more effective use of resources. The Proposed MADRL with DL consumes more energy compared to previous options, but it offers a trade-off for increased fairness and throughput. The Standalone MADRL has somewhat greater efficiency compared to the Proposed MADRL with DL, while Traditional RL consumes less energy than both MADRL approaches. The Static K-Means algorithm, while highly energy-efficient, may exhibit lower performance compared to other algorithms, as seen by preceding measures. The data indicates that the Proposed MADRL with DL method achieves superior performance in fairness and throughput, albeit with increased energy usage. The Standalone MADRL method is characterized by a balanced combination of moderate energy consumption and performance. Traditional RL and Static K-Means are more energy-efficient but lag in network performance, showing a trade-off between energy efficiency and operational efficacy. This research can help identify the optimal technique for UAV network communication based on the specific needs for fairness, throughput, and energy efficiency.

6. CONCLUSIONS

This research marks a significant advancement in UAV communication systems, particularly beneficial in environments lacking ground-based infrastructure. By integrating MADRL with DLT, we have significantly enhanced the performance, fairness, and energy efficiency of UAV networks. The proposed system architecture leverages a cluster of UAVs to deliver equitable communication services, striking an optimal balance between network connectivity and resource utilization. The decentralized learning strategy, based on the IPPO technique, enables UAVs to tailor their operational strategies based on individual observations, fostering a dynamic and adaptive communication environment. While our findings demonstrate that the MADRL with DLT approach substantially outperforms traditional methods across various metrics, this study also opens several avenues for future research. Future work could explore the integration of more complex adaptive algorithms to further enhance the system's responsiveness to changing environmental conditions and user demands. Additionally, further research is needed to address potential challenges in scalability and manageability as the size and complexity of UAV networks increase. The complexities involved in real-world implementations, such as regulatory hurdles and varying environmental conditions, also present substantial challenges that require innovative solutions. Moreover, as UAV technologies and the frameworks for their operation evolve, continuous improvements in security measures will be essential to safeguard against increasingly sophisticated cyber threats. The adoption of newer cryptographic techniques and advanced security protocols will be crucial to ensure the integrity and confidentiality of the data transmitted within these networks.

DECLARATIONS

Authors' contributions
Made substantial contributions to conception and design of the study and performed data analysis and interpretation: Ali F
Performed data acquisition and provided administrative, technical, and material support: Ahtasham M, Anfaal Z

Availability of data and materials
The data can be provided as per request.

Financial support and sponsorship
None.

Conflicts of interest
Ali F is a Guest Editor of the journal *Complex Engineering Systems*, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2024.

REFERENCES

1. Mahen MA, Anirudh SA, Chethana HD, Shashank AC. Design and development of amphibious quadcopter. *Int J Mech Prod Eng* 2014;2:30-4. DOI

2. Zhang D, Li C, Zhang Y. Dual-hand gesture controlled quadcopter robot. In: 2017 36th Chinese Control Conference (CCC). Dalian, China, 2017, pp. 6869-74. DOI

3. Marwan M, Han M, Dai Y, Cai M. The impact of global dynamics on the fractals of a quadrotor unmanned aerial vehicle (QUAV) chaotic system. *Word Sci* 2024;32:2450043. DOI

4. Elmas EE, Alkan M. Bir İnsansız Hava Aracı Sisteminin Tasarımı, Benzetimi ve Gerçekleştirilmesi. *Politeknik Dergisi* 2023;26:929-40. DOI

5. Zhang Y, Mou Z, Gao F, Jiang J, Ding R, Han Z. UAV-enabled secure communications by multi-agent deep reinforcement learning. *IEEE Trans Veh Technol* 2020;69:11599-611. DOI

6. Lu J, Guo X, Huang T, Wang Z. Consensus of signed networked multi-agent systems with nonlinear coupling and communication delays. *App Math Comput* 2019;350:153-62. DOI

7. Zhou Y, Jin Z, Shi H, et al. UAV-assisted fair communication for mobile networks: a multi-agent deep reinforcement learning approach. *Remote Sens* 2022;14:5662. DOI

8. Abohashish SMM, Rizk RY, Elsedimy EI. Trajectory optimization for UAV-assisted relay over 5G networks based on reinforcement learning framework. *J Wireless Com Network* 2023;55:2023. DOI

9. Li H, Li J, Liu M, Gong F. UAV-assisted secure communication for coordinated satellite-terrestrial networks. *IEEE Commun Lett* 2023;27:1709-13. DOI

10. Luo X, Xie J, Xiong L, Wang Z, Liu Y. UAV-assisted fair communications for multi-pair users: a multi-agent deep reinforcement learning method. *Comput Netw* 2024;242:110277. DOI

11. Agrawal N, Bansal A, Singh K, Li CP, Mumtaz S. Finite block length analysis of RIS-assisted UAV-based multiuser IoT communication system with non-linear EH. *IEEE Trans Commun* 2022;70:3542-57. DOI

12. Sun G, Zheng X, Sun Z, et al. UAV-enabled secure communications via collaborative beamforming with imperfect eavesdropper information. *IEEE Trans Mobile Comput* 2024;23:3291-308. DOI

13. Li J, Liu A, Han G, Cao S, Wang F, Wang X. FedRDR: federated reinforcement distillation-based routing algorithm in UAV-assisted networks for communication infrastructure failures. *Drones* 2024;8:49. DOI

14. Zhang Z, Liu Q, Wu C, Zhou S, Yan Z. A novel adversarial detection method for UAV vision systems via attribution maps. *Drones* 2023;7:697. DOI

15. Zhang Y, Zhuang Z, Gao F, Wang J, Han Z. Multi-agent deep reinforcement learning for secure UAV communications. In: 2020 IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea (South); 25-28 May 2020. DOI

16. Tang, D, Zhang Q. UAV 5G: enabled wireless communications using enhanced deep learning for edge devices. *Wireless Netw* 2023. DOI

17. Oubbati OS, Atiquzzaman M, Baz A, Alhakami H, Ben-Othman J. Dispatch of UAVs for urban vehicular networks: a deep reinforcement learning approach. *IEEE Trans Veh Technol* 2021;70:13174-89. DOI

18. Ansari S, Taha A, Dashtipour K, Sambo Y, Abbasi QH, Imran MA. Urban air mobility-a 6G use case? *Front Comms Net* 2021;2:729767. DOI

19. Shang Y. Consensus tracking and containment in multiagent networks with state constraints. *IEEE Trans Syst Man Cybern Syst* 2023;53:1656-65. DOI

20. Zhang G. 6G enabled UAV traffic management models using deep learning algorithms. *Wireless Netw* 2023;1-11. DOI

21. Elamin A, El-Rabbany A. UAV-based multi-sensor data fusion for urban land cover mapping using a deep convolutional neural network. *Remote Sens* 2022;14:4298. DOI

22. Kuutti S, Fallah S, Katsaros K, Dianati M, Mccullough F, Mouzakitis A. A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications. *IEEE Int Things J* 2018;5:829-46. DOI

23. Al-Hourani A, Kandeepan S, Lardner S. Optimal LAP altitude for maximum coverage. *IEEE Wireless Commun Lett* 2014;3:569-72. DOI

24. Ge J, Zhang S. Adaptive inventory control based on fuzzy neural network under uncertain environment. *Complexity* 2020;2020:1-10. DOI

25. Sun Q, Ren J, Zhao F. Sliding mode control of discrete-time interval type-2 fuzzy Markov jump systems with the preview target signal. *Appl Math Comput* 2022;435:127479. DOI

26. Zhong W, Xu L, Zhu Q, Chen X, Zhou J. MmWave beamforming for UAV communications with unstable beam pointing. *China Commun*

2019;16:37-46 DOI