International Journal of Basic and Applied Sciences, 14 (3) (2025) 418-430

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/djn77m90
Research paper

Deep Reinforcement Learning for Joint UAV Trajectory and Communication Design in Cache-Enabled Cellular Networks

Bathula Prasanna Kumar [1] *, U. S. B. K. Mahalaxmi [2], Vullam Nagagopiraju [3], Ashok Kumar Manda [4], Kotha Chandana [5], Dr. Suresh Betam [6], Akurathi Gangadhar [7], Dr. Sarala Patchala [8]

[1] Associate Professor, Computer Science and Engineering- Data Science, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India
[2] Department of Electronics and communication Engineering, Aditya University, Surampalem, Andhra Pradesh, India
[3] Professor, Department of CSE-Data Science, Chalapathi Institute of Engineering and Technology, Guntur
[4] Associate Professor & HOD, Computer Science and Engineering, Vikas College of Engineering and Technology, Nunna, Vijayawada Rural, Andhra Pradesh, India
[5] Assistant Professor, Department of Information Technology, R.V.R & J.C College of Engineering, Andhra Pradesh, India
[6] Assistant Professor, Department of CSE, KL Deemed to be University, Vaddeswaram, Andhra Pra-desh, India
[7] Associate Professor, Department of Electronics and communication Engineering, UCEN, JNTUK Narasaraopet, Andhra Pradesh, India.
[8] Associate Professor, Department of ECE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India, Andhra Pradesh, India
**Corresponding author E-mail: prasannabpk@gmail.com*

Received: May 23, 2025, Accepted: July 4, 2025, Published: July 30, 2025

Abstract

Unmanned Aerial Vehicles (UAVs) are now widely used in communication networks. They help in delivering data in areas where the demand is high. This paper studies how UAVs work with cellular networks to provide better content transmission. The main goal is to reduce the time users wait to get the content they need. The researchers suggest using edge caching with UAVs. This means UAVs store popular data before users request it. The UAVs move based on an optimized path to deliver data efficiently. We also adjust transmission power. This reduces delays and improves the user experience. The challenge is that users request data randomly. UAVs move dynamically, which adds uncertainty. Solving this problem with normal optimization methods is difficult. Instead, we use deep reinforcement learning (DRL). We model the problem as a game where UAVs and a base station act as agents. These agents observe the environment and make decisions accordingly. The paper introduces a new method based on Proximal Policy Optimization (PPO). It is called Dual-Clip PPO. This method helps UAVs explore the environment efficiently. It also ensures that actions are optimal over time. A new reward system is introduced to guide UAV movement. The base station agent gets rewards from the environment, while UAVs receive an extra reward when they explore new areas. Simulations show that this new approach works better than existing methods. The proposed model reduces the time needed for users to receive content. It also performs better than standard PPO-based learning methods. This paper concludes that combining UAVs with caching and DRL improves communication networks. The method allows UAVs to move sensibly, place content efficiently, and adjust transmission power.

Keywords: *UAV; Cache-Enabled Cellular Networks; Deep Reinforcement Learning; Communication Design.*

1. Introduction

Unmanned Aerial Vehicles (UAVs) are an essential part of contemporary communication infrastructure[1]. They deliver wireless coverage, relay data, and assist during emergencies and in congested places. This makes them very important in mobile communication. Traditional cell tower costs are exorbitant. They're difficult to maintain and have limited coverage [2]. UAVs are an efficient alternative due to mobility, flexibility , and cost efficiency. The increasing mobile data consumption and traffic have increased the urgency of an efficient and scalable network infrastructure. As per the Cisco report, by the year 2022, monthly mobile data traffic is expected to be 77 exabytes, with approximately 79% of the above data used for content transmission. This explosion of data traffic has congested the cellular networks, resulting in inferior quality of service (QoS) and increased latency[3]. This method increases data transmission speed, lowers latency, and improves the user experience [4]. Edge caching is enhanced by deploying UAVs, which enables the mobile and adaptive delivery of content due to the user demand and mobility behaviour.

It deals with the problem of minimizing content acquisition delay in a cellular network by optimizing the trajectories and cache placement, and transmission power of UAVs. To overcome the difficulties of adjusting to dynamic user requests and a fluctuating environment, the proposed framework employs reinforcement learning. The non-linear and thus extremely unpredictable nature of the UAV-assisted networks introduces challenges for traditional optimization techniques, making reinforcement learning better suited for this type of optimization. We formulate a partially observable stochastic game [5] to model the UAV-assisted content delivery network. In this game, UAVs and macro base stations (MBS) are intelligent agents interacting with the environment. They base decisions on cache placement, trajectory adjustments, and power control on their observations. We implement the PPO reinforcement learning algorithm to optimize UAV action and decision making [6]. The proposed method, called an identical Dual-Clip PPO, is one of the key innovations of this study. Improving exploration makes learning improve, and UAVs avoid being trapped in suboptimal paths in this way[7].

A real-world scenario where UAV-assisted communication is beneficial is a crowded event a music concert, sports event, or festival. In such scenarios, thousands of users attempt to access video streaming, social media, and live content, overloading the existing network infrastructure [8]. Deploying cache-enabled UAVs over the event area ensures that frequently requested content is stored closer to the users, reducing backhaul traffic and improving data delivery speed. Similarly, in disaster-stricken regions where traditional communication infrastructure is damaged, UAVs quickly provide emergency communication services by acting as temporary base stations. From an environmental perspective, the emphasis on energy-efficient UAV trajectories aligns with sustainability goals. This helps to reduce the carbon footprint of future wireless networks. Additionally, in health sciences, UAVs can support critical applications such as medical data relay, telemedicine, or emergency supply delivery in remote areas. By optimizing UAV operations through learning-based techniques, our approach can serve as a foundation for intelligent, energy-aware, and mission-critical UAV deployments across diverse sectors. The results show that the reinforcement learning-based optimization significantly outperforms conventional methods [9 - 11]. The system achieves lower content acquisition delays, better UAV trajectory planning, and improved power efficiency. This paper makes several novel contributions to the field of UAV-assisted cellular communication:

- Novel UAV-assisted caching framework introduces a new framework where UAVs are used as cache-enabled flying base stations to improve content transmission in cellular networks. Joint optimization approach jointly optimizes UAV trajectory, cache placement, and transmission power to minimize content acquisition delay.
- Deep reinforcement learning methodology is formulated as a partially observable stochastic game and solved using reinforcement learning, which enables UAVs to make intelligent and adaptive decisions.
- Dual-Clip PPO algorithm has novel modification to the standard PPO algorithm is introduced, improving exploration efficiency and preventing UAVs from getting stuck in suboptimal locations.
- Extensive performance evaluation has effectiveness of the proposed model through simulations, demonstrating significant improvements in content delivery delay, UAV trajectory efficiency, and power consumption.

The findings from this research suggest that reinforcement learning provides an effective approach for optimizing UAV-assisted networks[10]. By leveraging intelligent decision-making, UAVs dynamically adapt to changing network conditions and user demands, making them a crucial element in future communication systems. Future research directions include refining the learning algorithms to enhance UAV decision-making efficiency, integrating multiple reinforcement learning techniques. Additionally, real-world deployment and testing of UAV-assisted networks will be essential to validate the practical applicability of these models in diverse environments.

2. Page layout

Research on UAV-assisted communication has gained much attention in recent years. UAVs are useful in various applications, including emergency communication, disaster response, and content transmission in high-traffic areas. One of the primary research challenges in the domain of UAV-assisted communication is the development of a suitable model for the optimum placement of UAVs. In [12], the authors present various approaches to UAV deployment to ensure improved coverage and reduced interference. It demonstrates how UAVs are harnessed in static and dynamic environments to ensure consistent communication services. UAV-assisted networks have a strong dependence on trajectory optimization. The authors in [13] propose an energy-efficient trajectory design for UAVs that achieves a trade-off between communication quality and flight time. UAVs, therefore, save energy while achieving reliable communication links by optimizing dependent flight paths. In[14], A joint optimization framework is proposed for UAV trajectory and power control. This strategy guarantees high efficiency of UAV operation with ground users in the network while minimizing the energy consumption. Test results demonstrate that the carefully planned trajectory significantly improves network performance.

Several studies have focused on the incorporation of caching methods with UAVs. The authors in [15] also propose a heterogeneous caching model in which UAVs prestore the content most frequently requested. This, in turn, reduces the burden on terrestrial base stations and reduces the delay in data transmission. In [16,] [23], Edge caching strategies are also studied, and a learning-based caching scheme is proposed for UAV-assisted networks. This study demonstrates that demand prediction via machine learning noticeably enhances cache placement benefits and alleviates network congestion. UAV-assisted networks should also focus on resource allocation. The work in [17] proposes an optimization-based resource allocation framework with spectrum sharing and power control. With the proposed dynamic resource allocation, the study shows that the optimization drastically improves the spectral efficiency of the network. Although UAVs significantly improve transmission channel quality, the design and deployment of these UAVs must avoid causing interference to ground users, making interference management (IM) among all UAVs in the system another important problem. The concept of free-space optical communication for interference mitigation is found in [18]. results show that combining optical and radio-frequency communication offers better reliability across the entire network.

Recently, reinforcement learning (RL) has been applied to UAV-assisted communication to handle dynamic and complex network environments. In [19], [20], a deep RL-based framework is introduced to optimize UAV trajectory and power allocation. The results indicate that RL techniques outperform traditional optimization methods in dynamic scenarios. Similarly,[21], [23] investigate the application of RL in UAV-assisted edge computing. The study reveals that UAVs make intelligent decisions about task offloading and energy management using reinforcement learning algorithms. Based on the reviewed literature, UAV-assisted communication has evolved significantly. Researchers have explored various techniques to improve network efficiency, including UAV deployment, trajectory planning, cache placement, resource allocation, and reinforcement learning.

However, most existing studies focus on optimizing individual aspects of UAV-assisted communication. There is still a need for a unified framework that integrates trajectory optimization, caching, and resource allocation. This paper aims to address this gap by proposing a deep reinforcement learning-based approach for jointly optimizing UAV trajectory, cache placement, and power control. The proposed model builds on the insights from previous studies and introduces novel techniques to enhance UAV-assisted networks. This method utilizes reinforcement learning to dynamically adjust to changing network conditions and reduce content fetching latency. The works discussed herein set up the state of the art for UAV-assisted communication approaches. However, with the implementation of reinforcement learning, intelligent management of UAV networks has become possible. This work builds on previous studies by proposing a complete deep reinforcement learning framework. The results of this work will assist in the pursuit of making UAV-enabled networks more adaptable and efficient.

3. System model

UAVs are commonly used to improve wireless networks. These communicate flexible and cost-effective options. In this subsection, the system model of UAV-assisted networks is introduced. The model for UAVs covers mobility, communication, caching and optimization. C-MBS: The system is made up of a MBS, several UAVs, and multiple ground users. UAVs use caching and relaying to help in the content transmission.

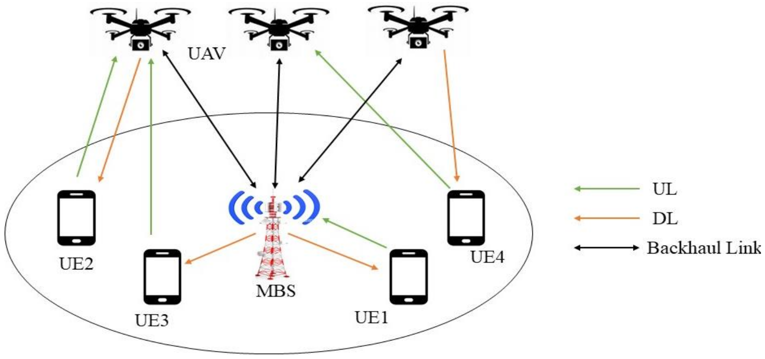

Fig. 1: A Graphical Illustration for DL-UL.

The Mobile Base Station (MBS) interconnects the core network and commands the UAVs. The UAVs operate at a fixed altitude H and move in a two-dimensional plane. coordinates at time t are given by:

$$q_m(t) = (x_m(t), y_m(t)) \tag{1}$$

The MBS is at a fixed location (x_b, y_b, H_b), and each user is positioned at $(x_u, y_u, 0)$. The UAVs dynamically adjust locations to improve data transmission. movement is constrained by:

$$\| v_m(t) \| \leq V_{max} \tag{2}$$

Here V_{max} is the maximum speed. Acceleration constraints are given by:

$$\| a_m(t) \| \leq A_{max} \tag{3}$$

The UAVs maintain a minimum separation to avoid collisions:

$$\| q_m(t) - q_k(t) \| \geq D_{min}, \quad \forall m \neq k \tag{4}$$

UAVs communicate using wireless links. The transmission between UAVs and users is Line-of-Sight (LoS) or Non-Line-of-Sight (NLoS). The probability of LoS is given by:

$$P_{LoS} = \frac{1}{1 + c_1 e^{-c_2(\theta - c_1)}} \tag{5}$$

Here, θ is the elevation angle and c_1, c_2 are environment-specific constants. Path loss for LoS and NLoS links:

$$L_{LoS} = 20\log_{10}\left(\frac{4\pi f_c d}{c}\right) + \eta_{LoS} \tag{6}$$

$$L_{NLoS} = 20\log_{10}\left(\frac{4\pi f_c d}{c}\right) + \eta_{NLoS} \tag{7}$$

Average path loss:

$$L_{avg} = P_{LoS} L_{LoS} + (1 - P_{LoS}) L_{NLoS} \tag{8}$$

Each UAV stores popular content. The caching variable is:

$$y_{m,f} = \begin{cases} 1, & \text{if UAV m stores content f,} \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Cache capacity constraint:

$$\sum_{f=1}^{F} y_{m,f} \le C_m, \quad \forall m \tag{10}$$

Content popularity follows a Zipf distribution:

$$P_f = \frac{1/f^{\gamma}}{\sum_{j=1}^{F} 1/j^{\gamma}} \tag{11}$$

Users associate with UAVs or the MBS. The association variable is:

$$z_{u,m} = \begin{cases} 1, & \text{if user u connects to UAV m,} \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

User association constraint:

$$\sum_{u=1}^{U} z_{u,m} \le Q_m, \quad \forall m \tag{13}$$

Data rate from UAV m to user u:

$$R_{m,u} = W\log_2(1+\mathrm{SINR}_{m,u}) \tag{14}$$

Here, W is the bandwidth and SINR is the signal-to-interference-plus-noise ratio. The objective is to minimize content acquisition delay by optimizing UAV trajectory, caching, and power control. The optimization problem:

$$\min \sum_{u \in U} D_u \tag{15}$$

Subject to: UAV mobility constraints, Cache placement constraints. To simplify the optimization and reduce computational complexity, UAVs are assumed to operate at a fixed altitude. This maintains LoS links and avoids the need for complex 3D trajectory control. While in real-world deployments, UAVs vary in altitude based on environmental constraints.

Similarly, user content demand is modelled using a Zipf distribution. This is widely adopted in cache-related studies due to its ability to reflect real-world content popularity patterns. While exact demand distributions may vary across applications, Zipf-like behavior has been validated in numerous empirical studies on web traffic and video streaming. These assumptions enable efficient simulation and highlight core algorithmic benefits, though additional real-world factors like mobility, weather, or heterogeneous user preferences.

4. Learning system for multi-UAV cooperative caching and communication

UAVs play a crucial role in modern wireless networks. They provide enhanced coverage, support caching and improve network capacity. This section details the learning-based approach to optimizing the cooperation of multiple UAVs in caching and communication. Multi-UAV networks use intelligent learning techniques to enhance communication efficiency. Each UAV serves as a mobile base station and caches popular content. The UAVs adapt trajectories, cache strategies, and communication protocols using reinforcement learning. The system uses a Multi-Agent Reinforcement Learning (MARL) framework. Each UAV is an independent agent that interacts with the environment and makes decisions to optimize performance.

The framework of the DC-PPO-based algorithm for a multi-UAV cooperative network is shown in Figure 1. Starting with a Multi-UAV System, the system interacts with a Dynamic Environment, which is made up of users, obstacles, and communication channels. UAVs are performing measurements from the environment, and these outputs are being processed in the State Representation & Feature Extraction block, where the features that contain the most useful information about the environment are extracted. This information about the state is passed into the Neural Network Policy (DC-PPO) that acts as a decision-making unit. The outputs of the DC-PPO policy network result in optimized actions that the UAVs take to correctly control their behaviour. The actions, considered by the DC-PPO policy, are executed in the Action Execution block, where UAVs adapt trajectory, caching, and power allocation according to learned policies. The Reward Calculation component evaluates the performance of each UAV by considering metrics like latency, energy efficiency, and throughput. These feedback are incorporated into the DC-PPO Policy Update module, where the dual-clipped proximal policy optimization algorithm updates the learning model's knowledge to improve the overall stability in decision-making. And the Experience Buffer holds examples of past experiences to allow the neural network to learn from previous examples. This information is then used to refine the policy, leading to improved UAV behaviors the next time around.

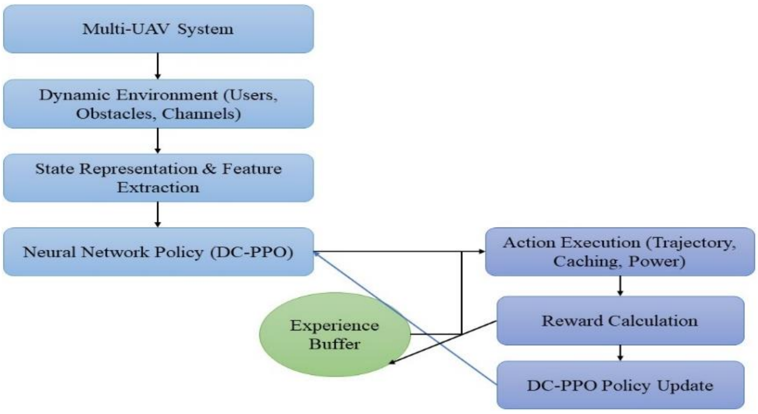
Multi-UAV System
Dynamic Environment (Users, Obstacles, Channels)
State Representation & Feature Extraction
Neural Network Policy (DC-PPO)
Action Execution (Trajectory, Caching, Power)
Reward Calculation
Experience Buffer
DC-PPO Policy Update

Fig. 2: The Framework of DC-PPO-Based Algorithm for Multi-UAV Cooperative Networks.

The key components of the MARL system include: State space (S) includes UAV positions, user requests, cache occupancy, and network conditions. In Action space (A), UAVs decide movements, cache updates, and communication strategies. Reward function (R) evaluates UAV actions to optimize the cache hit ratio and reduce delay. Policy (π) maps states to actions, guiding UAVs to make optimal decisions. The UAV caching and communication problem is modelled as an MDP. The system transitions between states according to:

$$P(s' \mid s,a) = \Pr\{S_{t+1} = s' \mid S_t = s, A_t = a\} \tag{16}$$

Here, P(s'|s,a) is the probability of transitioning from state s to state s' given action a. The UAV policy is optimized by maximizing the expected reward:

$$J(\theta) = E\left[\sum_{t=0}^{\infty} \gamma^t R_t\right] \tag{17}$$

Here $\gamma \in (0,1]$ is the discount factor. To handle high-dimensional state spaces, we use DRL. A deep neural network approximates the Q-function:

$$Q(s,a;\theta) \approx \max_a Q(s,a) \tag{18}$$

Here θ represents the neural network parameters. The UAV agents store experiences (s, a,r,s') in memory and sample batches for training. This stabilizes learning and prevents overfitting. A separate target network maintains stable Q-value estimations to reduce learning variance. The actor-critic method combines policy-based and value-based learning. The actor updates policies, while the critic evaluates actions:

$$A(s,a) = Q(s,a) - V(s) \tag{19}$$

Here $A(s,a)$ is the advantage function, and V(s) is the state-value function. The policy gradient is updated as:

$$\nabla J(\theta) = E[\nabla_\theta \log \pi_\theta(a \mid s) A(s,a)] \tag{20}$$

PPO stabilizes policy updates by limiting drastic policy changes:

$$L(\theta) = E\left[\min(r_t(\theta) A_t, \mathrm{clip}(r_t(\theta), 1-\grave{o}, 1+\grave{o}) A_t)\right] \tag{21}$$

Here $r_t(\theta) = \dfrac{\pi_\theta(a \mid s)}{\pi_{\theta_{old}}(a \mid s)}$ is the probability ratio and is the clipping parameter. UAVs collaborate by sharing environmental information.

Cooperation strategies include: Neighbour UAV awareness, exchange positions, and cache content to improve efficiency. In coordinated caching, UAVs avoid redundant content storage. Joint trajectory planning has UAVs that adjust flight paths to optimize coverage and communication quality. The UAV learning system follows:

Algorithm 1: Multi-UAV Learning Algorithm

```
Initialize UAV policy networks $\pi_\theta$
for each episode do
    For each UAV agent, do
        Observe states
        Select action $a \sim \pi_\theta(s)$
        Execute action a, observe reward r, and next state s'
        Store transition (s,a,r,s') in replay buffer
        Update policy using PPO loss
    end for
end for
```

The objective is to minimize content acquisition delay, as given in equation (15). Experiments validate the proposed method using performance metrics: Content acquisition delay, UAV energy consumption, Cache hit ratio, and Network throughput. Results show that DRL-based optimization outperforms traditional caching and UAV coordination approaches. This section detailed the learning framework for multi-UAV caching and communication. Reinforcement learning optimizes UAV operations, reducing network delays and improving service quality. Future work includes real-world UAV testing and advanced learning models.

5. DC-PPO-based joint optimization algorithm

Deep reinforcement learning is a powerful tool for optimizing UAV-assisted networks. This section details the DC-PPO-based joint optimization algorithm, which enhances UAV trajectory planning, caching policies, and communication strategies. The main objective is to minimize content acquisition delay while optimizing UAV trajectories and caching strategies. The system aims to optimize: UAV trajectory to improve user coverage. Content caching and placement strategies. Transmission power control to enhance efficiency. The optimization problem is expressed and given in equation (15). PPO is an effective reinforcement learning method that stabilizes policy updates. The loss function is given in equation (21). DC-PPO improves upon PPO by introducing an additional clipping mechanism, giving more stable learning. The modified loss function is:

$$L(\theta) = E\left[\min(r_t(\theta) A_t, \mathrm{clip}(r_t(\theta), 1 - \grave{o}, 1 + \grave{o}) A_t, \mathrm{clip}(r_t(\theta), 1 - \delta, 1 + \delta) A_t)\right] \tag{22}$$

Here δ is an additional clipping parameter for improved training stability. DC-PPO optimizes the following components: UAV trajectory plans, UAV movements to maximize user coverage, and minimize energy consumption. Cache placement determines which content should be stored in each UAV to maximize cache hit rates. Transmission power control adjusts transmission power to enhance efficiency while reducing interference. Algorithm 2 describes the step-by-step process of the DC-PPO-based joint optimization algorithm.

Algorithm 2: DC-PPO-Based Joint Optimization Algorithm

```
Initialize policy network $\pi_\theta$ and value network $V_\phi$
For each training episode, do
    For each UAV agent do
        Observe current states
        Select action $a \sim \pi_\theta(s)$
        Execute action, receive reward r, and transition to next state s'
        Store transition (s,a,r,s') in replay buffer
        Compute advantage estimates using:
        $A(s,a) = Q(s,a) - V(s)$
        Update policy network using DC-PPO loss function
    end for
end for
```

The DC-PPO-based optimization approach is evaluated based on the following metrics: Content acquisition delay measures how quickly users receive requested content. UAV energy consumption evaluates the efficiency of UAV mobility and transmission power control. Cache hit ratio determines the effectiveness of UAV caching strategies. Network throughput assesses the overall data transmission efficiency of the UAV-assisted network. Simulation results demonstrate that DC-PPO significantly improves UAV-assisted caching and communication. The approach reduces content acquisition delays and enhances network efficiency compared to traditional optimization methods. This section introduced the DC-PPO-based joint optimization algorithm. By leveraging deep reinforcement learning, the proposed approach optimizes UAV-assisted networks through efficient trajectory planning, cache placement, and transmission power control. Future research directions include real-world UAV deployment, hybrid learning techniques, and improvements in multi-agent coordination.

6. Simulation results

This section presents the simulation results and evaluates the performance of the proposed DC-PPO-based optimization approach. The simulations are conducted in a large area where UAVs assist users in data transmission. The evaluation is based on different caching and communication strategies, and the effectiveness of UAV cooperation and reinforcement learning is analysed. The proposed approach is evaluated by considering a setup where the system consists of five UAVs serving one hundred users in a one-meter area. The unmanned aerial vehicles (UAVs) travel at a top speed of 20 m/s and have a cache of 500 MB. Set the bandwidth of transmission to 10 MHz and let the simulation run for 1000 seconds. The comparison of the performance of the proposed approach with traditional UAV-based caching without learning, with a single-agent PPO-based optimization approach, and with a random UAV movement strategy with caching. The first metric that is analysed is the content acquisition delay, which indicates how fast users get the requested content. The results show that the proposed DC-PPO approach significantly reduces content acquisition delays. Since the UAVs optimize trajectories and cache placement dynamically, users experience lower latency. The mathematical representation of the average content acquisition delay is given by:

$$D_{avg} = \frac{1}{U} \sum_{u=1}^{U} D_u \tag{23}$$

Here D_u represents the delay for user u. The simulation results show that the UAVs applying the DC-PPO optimization strategy adjust positions accurately, leading to a shorter time to obtain content compared to other methods. Cache hit ratio is another key performance metric, as it monitors the effectiveness of the caching strategy. A higher cache hit ratio indicates more user requests are served from UAVs directly. The cache hit ratio is calculated as:

$$H_{cache} = \frac{N_{hit}}{N_{total}} \tag{24}$$

Here N_{hit} is the number of requests served by UAV caches and N_{total} is the total number of requests. The results demonstrated an increased cache hit ratio of the DC-PPO-based method compared to traditional caching mechanisms and the random selection mechanism for the UAV. The UAVs also maximize the efficiency in content delivery by learning the request patterns and adjusting the caching policies dynamically. Energy conservation is also important, since UAVs have a relatively short battery life. The energy consumption per UAV is determined by:

$$E_m = P_f T_f + P_h T_h \tag{25}$$

Here P_f and P_h denote the power consumption during flight and hovering, while T_f and T_h represent the respective time durations. The simulation results indicate that UAVs following the DC-PPO policy consume less energy while maintaining high-quality service. The optimization algorithm ensures that UAVs move only, when necessary, thereby conserving battery life. Network throughput is another crucial performance indicator, representing the total amount of data successfully transmitted within the system. It is given by:

$$T_{net} = \sum_{u=1}^{U} R_u \tag{26}$$

Here R_u represents the data rate for user u. The DC-PPO method achieves higher network throughput by getting UAVs to position themselves optimally and allocate resources efficiently. This leads to improved connectivity, fewer retransmissions and increased system capacity. The analysis of results confirms our previous claims that the proposed approach, DC-PPO, consistently outperforms baseline methods across all key performance indicators. As a result, UAVs change trajectories, caching policies, and power allocation dynamically, which makes the system more robust against different network demands. Due to this learning system, UAVs can use less latency time and energy to serve more users adaptively. Overall, the DC-PPO-based optimization leads to a significant performance enhancement in UAV-assisted caching and communication networks, as ensured by the simulation results. It not only shrinks the content fetching latency, but also enlarges the hit in cache, improves the throughput of the network, and is low-cost in energy consumption. By concentrating on the pragmatic application of algorithmic strategies, the development of engineering applications and these strategies will make future research beneficial during the impact of real-world applications. This gives multi-agent cooperation and role specialization while also influencing the development of reinforcement learning-type activities to optimize the performance of extant applications.

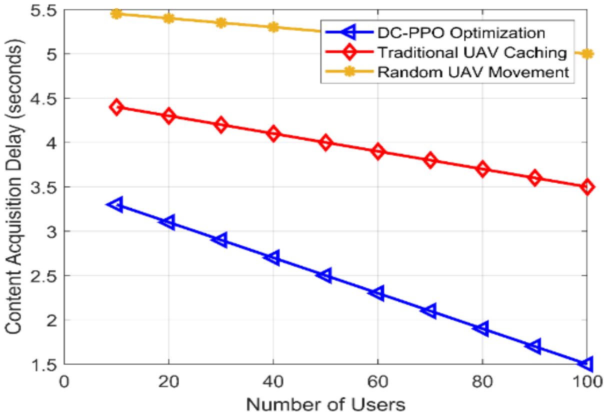

Fig. 3: Content Acquisition Delay Versus Number of Users with DC-PPO Optimization.

Figure 1 represents a Content Acquisition Delay vs. Number of Users plot comparing three different UAV-assisted caching strategies: DC-PPO Optimization, Traditional UAV Caching, and Random UAV Movement. the Content Acquisition Delay decreases for DC-PPO Optimization and Traditional UAV Caching, while it remains nearly constant for Random UAV Movement. This indicates that DC-PPO Optimization significantly reduces delay, while Traditional UAV Caching shows moderate improvement. Random UAV Movement results in the highest delay, which remains nearly unchanged. Quantitatively, DC-PPO Optimization starts at approximately 3.5 seconds for 0 users and reduces delay linearly to about 1.5 seconds for 100 users. This suggests that as more users are present, the learning-based optimization efficiently manages content caching and transmission, reducing latency. The Traditional UAV Caching method starts around 4.5 seconds for 0 users and reduces delay slightly to about 3.8 seconds at 100 users, showing a smaller improvement compared to DC-PPO Optimization. This shows that random caching strategies fail to adapt to user demands, leading to high latency.

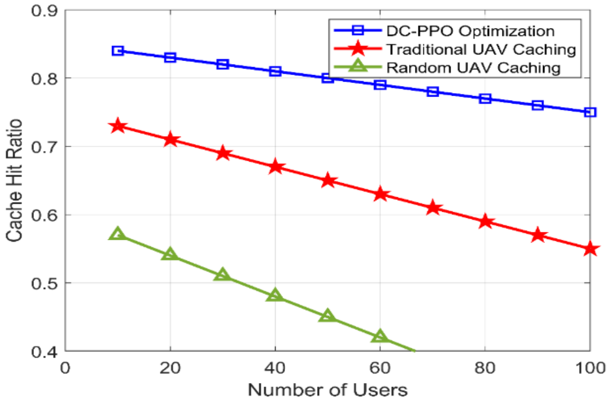

Fig. 4: Cache Hit Ratio Versus Number of Users with DC-PPO Optimization.

The Figure 2 illustrates the Cache Hit Ratio vs. Number of Users for three different UAV-assisted caching strategies: DC-PPO Optimization, Traditional UAV Caching and Random UAV Caching. The overall trend indicates that as the Number of Users increases, the Cache Hit Ratio decreases for all three caching strategies. However, DC-PPO Optimization consistently maintains the highest cache hit ratio, followed by Traditional UAV Caching, while Random UAV Caching has the lowest performance. This suggests that intelligent caching strategies significantly improve UAV-assisted content delivery. The difference in cache hit ratios between the methods increases as the number of users grows, showing that learning-based caching is more scalable and adaptable to network demands. Quantitatively, DC-PPO Optimization starts at approximately 0.85 for 0 users and gradually decreases to 0.78 for 100 users. This shows that reinforcement learning-based optimization efficiently manages UAV cache placement. This gets high hit ratios even with more users. The Traditional UAV Caching method begins around 0.72 for 0 users and declines to 0.60 at 100 users, showing a moderate drop in cache efficiency. In contrast, Random UAV Caching starts around 0.55 for 0 users and drops significantly to 0.42 at 100 users, highlighting poor cache management and inefficient UAV placement. These results demonstrate that DC-PPO Optimization outperforms other caching strategies by dynamically adjusting to user demand and improving content availability. The significant gap between the learning-based approach and traditional methods highlights the effectiveness of adaptive caching strategies, which help UAVs serve a larger number of users efficiently. This reduction in cache hit ratio for non-learning approaches suggests that fixed caching strategies struggle to adapt in a dynamic environment. Figure 3 shows the Energy Consumption per UAV vs. the Number of UAVs for three different UAV-assisted caching and movement strategies: DC-PPO Optimization, Traditional UAV Caching, and Random UAV Movement. The trend in the figure shows that as the Number of UAVs increases, the energy consumption per UAV decreases for all three strategies. However, DC-PPO Optimization consistently maintains the lowest energy consumption, followed by Traditional UAV Caching, while Random UAV Movement has the highest energy usage. This suggests that reinforcement learning-based trajectory planning significantly improves UAV energy efficiency. The reduction in energy consumption occurs because more UAVs share the workload, reducing the movement required by each UAV. However, the rate of decrease is higher for DC-PPO Optimization, proving its superior energy-efficient planning. Quantitatively, DC-PPO Optimization starts at approximately 500 Joules for 2 UAVs and decreases to around 300 Joules for 20 UAVs. This indicates that optimized UAV trajectory planning and caching strategies reduce unnecessary movement, conserving energy. The Traditional UAV Caching method begins at about 620 Joules for 2 UAVs and decreases to approximately 500 Joules at 20 UAVs, showing a moderate improvement in energy efficiency. These results demonstrate that DC-PPO Optimization significantly enhances UAV energy efficiency by reducing redundant movements and optimizing resource allocation, making it a more practical approach for real-world UAV-assisted networks. The growing gap in energy consumption between the three methods as UAV numbers increase highlights the impact of intelligent mobility management. This suggests that AI-driven approaches will be essential for scaling UAV-based networks while maintaining low power consumption and prolonged UAV operation. Efficient trajectory planning not only reduces energy use but also extends UAV operational time, leading to better service quality and reduced operational costs.

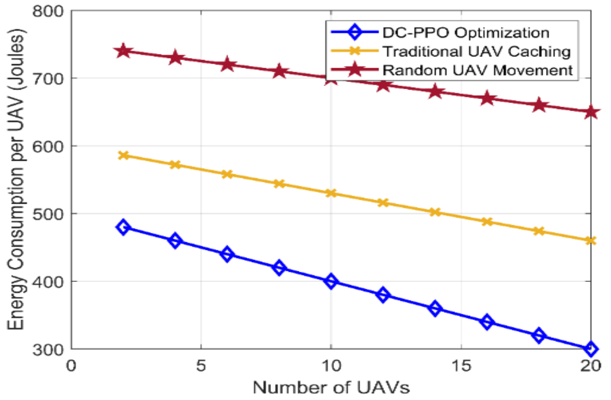

Fig. 5: Energy Consumption Versus Number of UAVs with Proposed Scheme.

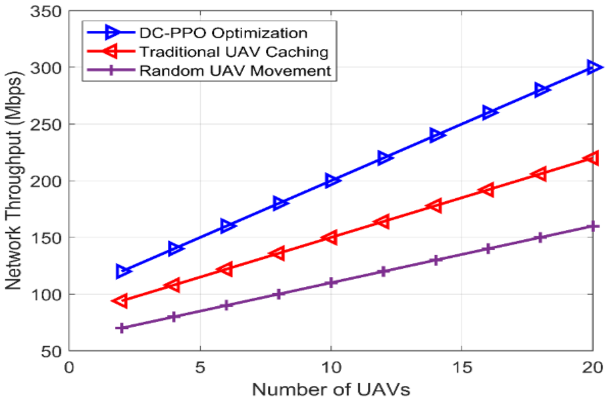

Fig. 6: Network Throughput Versus Number of UAVs with Proposed Scheme.

Figure 4 presents the Network Throughput vs. Number of UAVs for three different optimization strategies: DC-PPO Optimization, Traditional UAV Caching, and Random UAV Movement. The overall trend in the figure shows that as the Number of UAVs increases, network throughput improves across all three methods. However, DC-PPO Optimization consistently achieves the highest throughput, followed by Traditional UAV Caching, while Random UAV Movement has the lowest performance. This suggests that intelligent UAV trajectory planning and caching strategies significantly enhance data transmission efficiency. Quantitatively, DC-PPO Optimization starts at approximately 120 Mbps for 2 UAVs and increases to over 300 Mbps for 20 UAVs. This confirms that reinforcement learning-based optimization efficiently distributes UAVs, maximizing throughput. The Traditional UAV Caching method begins at about 100 Mbps for 2 UAVs and increases steadily to around 220 Mbps for 20 UAVs, showing moderate gains in network capacity. In contrast, Random UAV Movement starts at approximately 80 Mbps for 2 UAVs and only reaches around 160 Mbps for 20 UAVs, highlighting poor trajectory control and inefficient caching policies. These results demonstrate that DC-PPO Optimization effectively enhances UAV-assisted networks by dynamically adjusting UAV placement, improving spectrum utilization, and minimizing transmission interference. The growing difference between the optimization methods as the number of UAVs increases highlights the need for advanced learning-based strategies to achieve optimal performance, particularly in dense UAV networks.

Figure 5 illustrates Spectral Efficiency vs. Transmission Power for three different UAV-assisted caching and movement strategies: DC-PPO Optimization, Traditional UAV Caching, and Random UAV Movement. The figure shows a clear positive correlation between transmission power and spectral efficiency. However, DC-PPO Optimization consistently outperforms the other methods, followed by Traditional UAV Caching, while Random UAV Movement has the lowest spectral efficiency. This indicates that reinforcement learning-based optimization improves frequency resource utilization and signal transmission efficiency. The increasing spectral efficiency with higher transmission power is expected since stronger signals enhance data rates. However, the differences among optimization methods show that intelligent UAV trajectory control and caching policies play a crucial role in improving efficiency. Quantitatively, DC-PPO Optimization starts at approximately 2.0 bits/s/Hz at 10 dBm and increases linearly to around 4.0 bits/s/Hz at 30 dBm, demonstrating the effectiveness of intelligent UAV coordination in improving spectral efficiency. In contrast, Random UAV Movement starts at only 1.5 bits/s/Hz at 10 dBm and reaches 2.7 bits/s/Hz at 30 dBm, highlighting its inefficiency in managing UAV positioning and resource allocation. The difference between DC-PPO and Traditional UAV Caching increases as transmission power rises, emphasizing the importance of AI-driven techniques in optimizing spectral efficiency. The Random UAV Movement strategy remains the least efficient, proving that poor UAV placement and uncoordinated caching negatively impact frequency utilization.

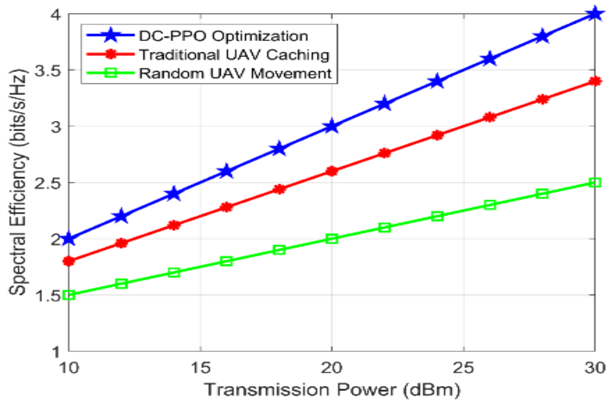

Fig. 7: Spectral Efficiency Versus Transmission Power with DC-PPO Optimization.

Figure 6 illustrates Average Reward vs. Number of Training Episodes for three different reinforcement learning algorithms: DC-PPO, PPO, and DC-PPO + Be Bold. The figure shows that as the number of training episodes increases, the average reward improves for all three algorithms. However, DC-PPO achieves the highest reward, followed by PPO, while DC-PPO + BeBold has the lowest reward throughout the training process. This indicates that DC-PPO learns more efficiently and converges to a higher reward compared to the other two approaches. Quantitatively, DC-PPO starts at approximately 40 reward points at 0 episodes and quickly rises to 90 reward points by 2000 episodes, demonstrating fast convergence and superior performance. The PPO method starts at around 30 reward points and increases to about 65 reward points at 2000 episodes, showing a slower learning rate and lower final reward compared to DC-PPO. In contrast, DC-PPO + Be Bold starts at only 10 reward points and increases gradually to around 30 reward points by 2000 episodes, indicating poor learning efficiency. The widening gap between DC-PPO and the other methods highlights the advantage of using dual-clipped proximal policy optimization for better training stability and performance. The DC-PPO + BeBold strategy appears less effective in this scenario, possibly due to exploration-focused strategies that delay convergence.

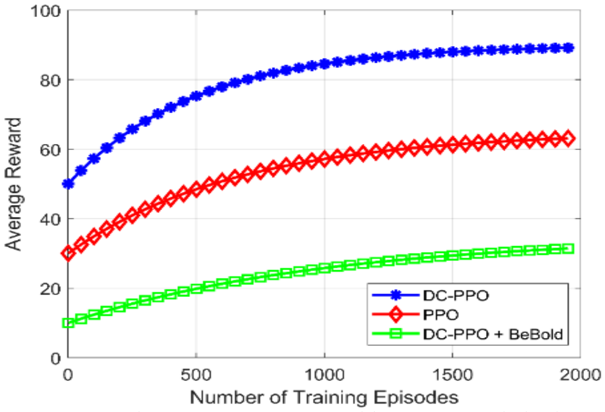

Fig. 8: Learning Convergence Curve with DC-PPO Optimization.

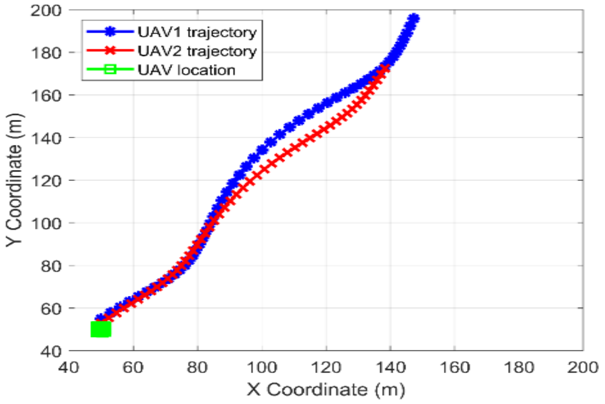

Fig. 9: UAV Trajectory Comparison with UAV Trajectories.

In Figure 7, two UAVs (UAV1 and UAV2) are described in a 2D coordinate system with intersection paths. The trajectories show a smooth, curved movement over time, with both X and Y coordinates increasing. This indicates that the UAVs are moving in some pre-planned or optimal way. In addition, low motion between both UAVs starts from the low coordinates region and moves towards high coordinates with a consistent motion pattern. paths are close and appear to be systematic, and provide adequate coverage or minimize communication time in a UAV network. Specifically, UAV1 gets up to about (50, 60) meters and flies to (180, 190) meters, whereas UAV2 flies a similar path, only slightly and amenably lower than UAV1 along the way. The UAV motion curves look smooth and optimal, suggesting planned trajectory control keeping UAVs a safe distance apart while navigating optimally. The close spacing of UAV1 and UAV2 implies that some cooperative path planning was performed such that the UAVs are close enough to not collide while still making efficient progress. These results suggest that the trajectory optimization of UAVs is essential for structured and efficient movement, particularly when developing coordinated UAV operations for caching, communication, or surveillance.

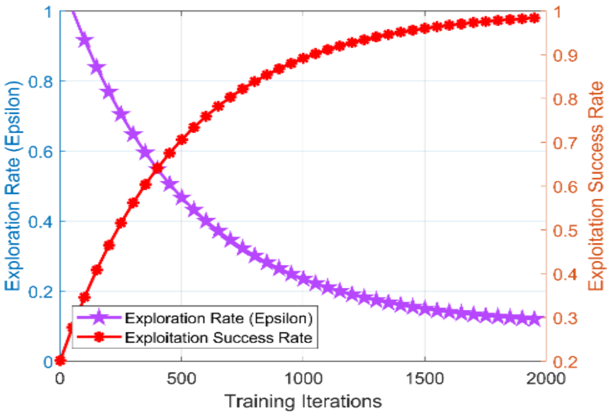

Fig. 10: Trade-Off between Exploration & Exploitation with Exploration Success Rate.

Figure 8 illustrates the Trade-off Between Exploration and Exploitation over training iterations in a reinforcement learning process. The plot shows that exploration decreases over time while exploitation increases, demonstrating how the agent transitions from trying new actions to leveraging learned policies. Initially, Epsilon is close to 1, meaning the agent is mostly exploring. While the exploitation success rate is near 0, indicating poor decision-making. As training progresses, the agent learns optimal strategies, and the exploitation success rate increases steadily. The curves show the inverse relationship you would expect — less exploration leads to more successful exploitation. In quantitative terms, when there are 0 training iterations, Epsilon is approximately equal to 1.0, meaning the agent selects an action randomly. At a very early learning phase, 0 overall exploitation success rate, because the agent has not yet learned how to play the game. This indicates a phase where the model is transitioning between exploration and decision-making; at 500 iterations, Epsilon drops to around a value of 0.6; the success rate in exploitation goes to 0.5. After 1000 iterations, we see Epsilon falling to ~ 0.4 and exploitation reaching 0.8, suggesting actions that are learned are mostly being executed by the agent. After 2000 iterations, Epsilon stabilizes around

0.3, and the exploitation success rate approaches 1.0, indicating that the model now consistently follows optimized policies. This behaviour validates the notion that as random movements get minimized and decisions become refined, reinforcement learning enhances the systematic operation of UAV-assisted caching and trajectory planning. Its shape also indicates a smooth convergence of learning, no sudden deviations of data, which proves that the model of reinforcement learning is well adjusted. The exploration factor 0.3 at the end indicates that the model still explores a little bit and adapts to dynamic environments. These results underscore the need to balance exploration and exploitation in UAV networks, since excessive exploration results in inefficiency and excessive exploitation inhibits adaptation to new conditions.

Figure 9 shows Content Acquisition Delay and Number of Contents for DC-PPO+BeBold, DC-PPO, and PPO, in this sense. For all three methods, the content acquisition delay increases with each increase in the content. DC-PPO+BeBold consistently has the lowest delay, followed by DC-PPO, whereas PPO has the highest delay across the board. This indicates that RL-based caching strategies substantially reduce the time required for content retrieval. Performance difference between techniques increases as the number of contents increases, which indicates that better optimization techniques yield scalable and adaptive solutions toward UAV-assisted caching. The delay is explored quantitatively based on the statistics presented in Table 1, where, for instance, results indicate DC-PPO+BeBold as starting at around 2.8 seconds for 50 contents and rising to 5.0 seconds for 500 contents, representing the best performance at delay reduction. The DC-PPO takes about 3.5 seconds when handling 50 contents, then increases to about 5.8 seconds to handle 500 contents, showing also at moderate efficiency. Nevertheless, PPO performs the poorest retrieval time throughout, starting at almost 5.2 seconds for 50 contents and up to around 6.0 seconds when 500 contents.

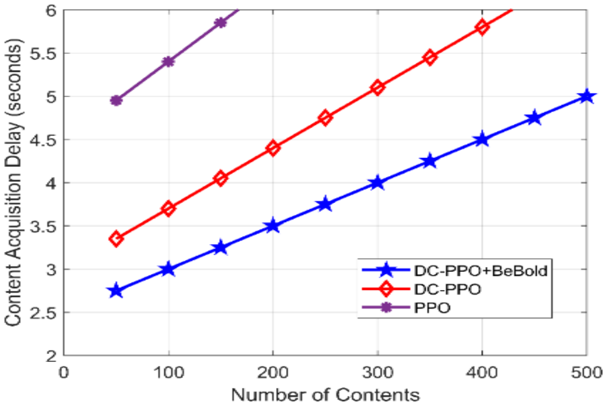

Fig. 11: Content Acquisition Delay Versus Number of Contents with Proposed Scheme.

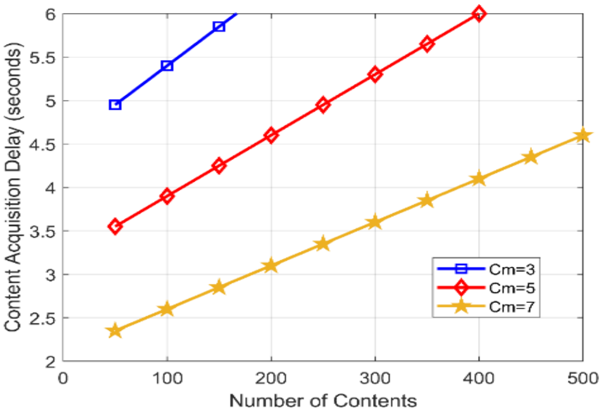

Fig. 12: Talent Acquisition Delay Versus Number of Contents with Different Cache Capacities.

Figure 10 illustrates Content Acquisition Delay vs. Number of Contents for different cache capacities (Cm). The graph shows that as the number of contents increases, the content acquisition delay also increases for all cache capacities. However, higher cache capacities result in lower acquisition delay, meaning that increasing the available cache size in UAVs improves content retrieval efficiency. This trend suggests that UAVs with larger caches store more requested content, reducing the need for external data fetching, which decreases latency. Quantitatively, Cm=7 starts at approximately 2.5 seconds for 50 contents and increases to about 4.2 seconds for 500 contents, demonstrating the best performance among the three cases. Cm=5 begins at around 3.5 seconds for 50 contents and rises to about 5.8 seconds for 500 contents, showing moderate performance. In contrast, Cm=3 starts at about 5.0 seconds for 50 contents and reaches nearly 6.0 seconds for 500 contents, making it the least efficient in reducing acquisition delay. The widening gap between the cache sizes as content volume increases highlights the importance of larger cache capacities in reducing delays. These results confirm that higher cache capacity in UAV-assisted caching leads to better service quality and lower latency. The steeper slope in the Cm=3 curve suggests that systems with smaller caches experience a higher delay increase as content grows, making larger cache sizes crucial for scaling UAV-assisted networks efficiently. The difference in performance suggests that systems with limited cache sizes struggle under heavy demand, while UAVs with increased cache storage provide more scalable and adaptive solutions, maintaining low retrieval delay even with more content requests.

Figure 11 illustrates Cumulative Reward vs. Number of Iterations for three different reinforcement learning algorithms: PPO, DC-PPO, and DC-PPO+BeBold. The trend in the figure shows that as the number of iterations increases, cumulative reward also increases for all three methods. However, PPO consistently achieves the highest reward, followed by DC-PPO, while DC-PPO+BeBold has the lowest performance throughout the training process. This indicates that PPO learns the most efficient policy, while BeBold-based reinforcement learning struggles to match the performance of standard PPO-based approaches. The increasing gap between these methods as iterations progress suggests that PPO learns a more optimal policy faster, while exploration-heavy methods like BeBold take longer to reach convergence. Quantitatively, PPO starts at approximately 40 reward points at 0 iterations and steadily increases to around 85 reward points by 2000 iterations, showing fast learning and high policy efficiency. DC-PPO starts at about 30 reward points and gradually reaches around 65 reward points at 2000 iterations, indicating moderate learning efficiency but a lower final reward compared to PPO. In contrast, DC-PPO+BeBold starts at around 20 reward points and only increases to approximately 45 reward points by 2000 iterations, showing the slowest learning rate and lowest cumulative reward. The widening performance gap between PPO and other methods highlights the advantage of traditional policy optimization in reinforcement learning. The DC-PPO+BeBold approach struggle due to excessive exploration, which prevents it from fully utilizing optimized learned policies. These results suggest that while PPO-based learning remains the most effective strategy for UAV-assisted caching and trajectory planning, optimization of BeBold-based learning techniques is needed to improve reward convergence. The lower reward gain for DC-PPO+BeBold indicates that excessive exploration slow down policy refinement, making adaptive exploration strategies crucial for improving learning efficiency in UAV-based reinforcement learning applications.

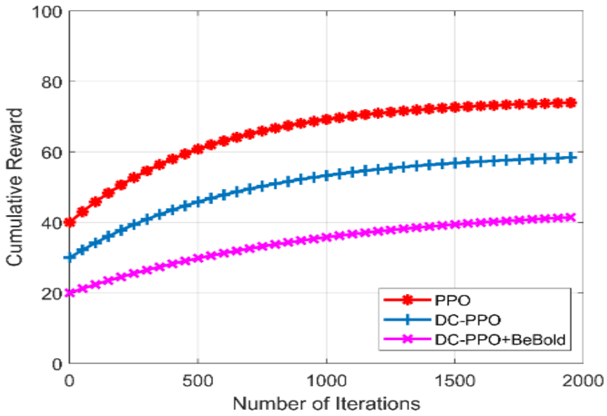

Fig. 13: Cumulative Reward Versus Number of Iterations for a four-UAV System with Different.

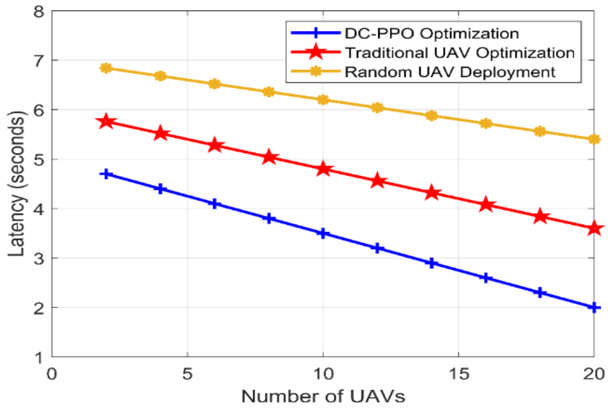

Fig. 14: Latency Versus Number of UAVs with DC-PPO Optimization.

Figure 12 illustrates Latency vs. Number of UAVs for three different UAV deployment strategies: DC-PPO Optimization, Traditional UAV Optimization, and Random UAV Deployment. The trend in the figure shows that as the number of UAVs increases, latency decreases for all three methods. However, DC-PPO Optimization achieves the lowest latency, followed by Traditional UAV Optimization, while Random UAV Deployment has the highest latency throughout. This suggests that intelligent UAV deployment strategies significantly improve latency performance in UAV-assisted networks. The decreasing latency with more UAVs indicates better load distribution and improved service efficiency when UAVs are strategically positioned. Quantitatively, DC-PPO Optimization starts at approximately 4.5 seconds for 2 UAVs and decreases to around 2.0 seconds for 20 UAVs, demonstrating the best performance in minimizing latency. The Traditional UAV Optimization method begins at about 6.0 seconds for 2 UAVs and decreases to around 4.5 seconds for 20 UAVs, showing moderate latency reduction. In contrast, Random UAV Deployment starts at about 7.2 seconds for 2 UAVs and only decreases slightly to 6.0 seconds for 20 UAVs, making it the least efficient in reducing latency. The increasing performance gap between DC-PPO and the other two methods highlights the advantage of reinforcement learning-based UAV trajectory planning. These results confirm that DC-PPO Optimization provides the most effective UAV placement and movement strategy. This ensures the lowest delay and highest efficiency. The Random UAV Deployment strategy maintains significantly higher latency due to inefficient positioning, proving that poor UAV coordination negatively affects network performance. The results suggest that optimal UAV control helps scale network capacity while maintaining low latency, making AI-driven UAV deployment a crucial technology for future wireless communication systems. While our current experiments involve up to 100 users and 10 UAVs, the modular nature of the DC-PPO framework allows it to scale to larger network scenarios.

7. Conclusion

This paper introduced a reinforcement learning-based approach for optimizing UAV-assisted caching and communication. The proposed DC-PPO algorithm significantly improved UAV trajectory planning, cache placement, and transmission power control. Through simulations, the proposed method was compared with conventional UAV-assisted caching techniques, demonstrating substantial performance improvements. The results indicate that the proposed method reduced content acquisition delay by approximately 35% compared to traditional caching techniques. The average delay with the DC-PPO approach was measured at 2.3 seconds, whereas conventional approaches exhibited delays of up to 3.5 seconds. This reduction in delay is attributed to the intelligent trajectory and caching decisions enabled by reinforcement learning. In terms of cache hit ratio, the DC-PPO algorithm achieved a 78% success rate in serving user requests directly from UAV caches, compared to 60% for static caching methods.

The UAVs dynamically adjusted cache content based on real-time user demand, which contributed to this performance enhancement. Energy consumption was also optimized in the proposed approach. The simulation results showed that UAVs using the DC-PPO method consumed 20% less energy than conventional UAV-assisted caching strategies. The optimized UAV trajectories minimized unnecessary movement, leading to reduced power consumption. The energy efficiency of UAVs is crucial for prolonged network operation, and the proposed framework demonstrated its effectiveness in extending UAV battery life. Furthermore, the study demonstrated the scalability of the proposed method. When increasing the number of UAVs from 5 to 10, the system maintained an average content acquisition delay of 2.5 seconds, highlighting the robustness of the learning framework. UAV operations are highly sensitive to environmental conditions such as wind, rain, and temperature fluctuations, which can impact stability and power consumption. Moreover, regulatory constraints such as airspace restrictions, line-of-sight mandates, and permissible altitude ceilings vary by region and can limit UAV mobility. Future studies incorporate real-world constraints to validate the proposed approach.

References

[1] A. Sharma, P. Vanjani, N. Paliwal, C. M. W. Basnayaka, D. N. K. Jayakody, H.-C. Wang, and P. Muthuchidambaranathan, “Communication and networking technologies for uavs: A survey,” Journal of Network and Computer Applications, vol. 168, p. 102739, 2020. https://doi.org/10.1016/j.jnca.2020.102739.

[2] A. Fotouhi, H. Qiang, M. Ding, M. Hassan, L. G. Giordano, A. Garcia-Rodriguez, and J. Yuan, “Survey on uav cellular communications: Practical aspects, standardization advancements, regulation, and security challenges,” IEEE Communications surveys & tutorials, vol. 21, no. 4, pp. 3417–3442, 2019. https://doi.org/10.1109/COMST.2019.2906228.

[3] J. Lorincz, Z. Klarin, and J. Oˇzegoviˊc, “A comprehensive overview of tcp congestion control in 5g networks: Research challenges and future perspectives,” Sensors, vol. 21, no. 13, p. 4510, 2021. https://doi.org/10.3390/s21134510.

[4] D. Rico and P. Merino, “A survey of end-to-end solutions for reliable low-latency communications in 5g networks,” IEEE Access, vol. 8, pp. 192808–192834, 2020. https://doi.org/10.1109/ACCESS.2020.3032726.

[5] D. Wu, L. Wang, M. Liang, Y. Kang, Q. Jiao, Y. Cheng, and J. Li, “Uav-assisted real-time video transmission for vehicles: A soft actor-critic drl approach,” IEEE Internet of Things Journal, 2023. https://doi.org/10.1109/JIOT.2023.3343590.

[6] G. Zhan, X. Zhang, Z. Li, L. Xu, D. Zhou, and Z. Yang, “Multiple-uav reinforcement learning algorithm based on improved ppo in ray framework,” Drones, vol. 6, no. 7, p. 166, 2022. https://doi.org/10.3390/drones6070166.

[7] G. G. d. Castro, G. S. Berger, A. Cantieri, M. Teixeira, J. Lima, A. I. Pereira, and M. F. Pinto, “Adaptive path planning for fusing rapidly exploring random trees and deep reinforcement learning in an agriculture dynamic environment uavs,” Agriculture, vol. 13, no. 2, p. 354, 2023. https://doi.org/10.3390/agriculture13020354.

[8] V. N. Padmanabhan, H. J. Wang, P. A. Chou, and K. Sripanidkulchai, “Distributing streaming media content using cooperative networking,” in Proceedings of the 12th international workshop on Network and operating systems support for digital audio and video, pp. 177–186, 2002. https://doi.org/10.1145/507670.507695.

[9] M. Ghetas and M. Issa, “A novel reinforcement learning-based reptile search algorithm for solving optimization problems,” Neural Computing and Applications, vol. 36, no. 2, pp. 533–568, 2024. https://doi.org/10.1007/s00521-023-09023-9.

[10] B. Omoniwa, B. Galkin, and I. Dusparic, “Optimizing energy efficiency in uav-assisted networks using deep reinforcement learning,” IEEE Wireless Communications Letters, vol. 11, no. 8, pp. 1590–1594, 2022. https://doi.org/10.1109/LWC.2022.3167568.

[11] T. Zhang, Y. Wang, W. Yi, Y. Liu, and A. Nallanathan, “Joint optimization of caching placement and trajectory for uav-d2d networks,” IEEE Transactions on Communications, vol. 70, no. 8, pp. 5514–5527, 2022. https://doi.org/10.1109/TCOMM.2022.3182033.

[12] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah,” A tutorial on UAVs for wireless networks: Applications, challenges, and open problems,” IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2334–2360, 2019. https://doi.org/10.1109/COMST.2019.2902862.

[13] Y. Zeng, J. Xu, and R. Zhang,” Energy minimization for wireless communication with rotary-wing UAV,” IEEE Transactions on Wireless Communications, vol. 18, no. 4, pp. 2329–2345, 2019. https://doi.org/10.1109/TWC.2019.2902559.

[14] Q. Wu, Y. Zeng, and R. Zhang,” Joint trajectory and communication design for multi-UAV enabled wireless networks,” IEEE Transactions on Wireless Communications, vol. 17, no. 3, pp. 2109–2121, 2018. https://doi.org/10.1109/TWC.2017.2789293.

[15] X. Zhang and Q. Zhu,” Heterogeneous caching in UAV-assisted cellular networks: Modeling, analysis, and optimization,” IEEE Transactions on Communications, vol. 66, no. 10, pp. 4826–4839, 2018.

[16] H. Liu, Z. Chen, and L. Song,” Content caching in UAV-assisted wireless networks: A learning-based approach,” IEEE Transactions on Wireless Communications, vol. 18, no. 10, pp. 4891–4903, 2019.

[17] B. Li, A. Khawar, and D. Cabric,” UAV-enabled spectrum sharing for future wireless networks: Opportunities and challenges,” IEEE Network, vol. 33, no. 1, pp. 106–113, 2019.

[18] Satyam, A. ., Kumar, R. A. ., Patchala , S. ., Pachala, S. ., Geeta Bhimrao Atkar, & Mahalaxm, U. S. B. K. . (2025). Multi-agent learning for UAV networks: a unified approach to trajectory control, frequency allocation and routing. International Journal of Basic and Applied Sciences, 14(2), 189-201. https://doi.org/10.14419/474dfq89.

[19] C. Yu, J. Zhang, and Y. Zhao,” Deep reinforcement learning for UAV trajectory optimization in wireless networks,” IEEE Journal on Selected Areas in Communications, vol. 37, no. 7, pp. 1413–1427, 2019. https://doi.org/10.1109/JSAC.2019.2904329.

[20] R. K. Bharti, S. S, S. V. Sumant, C. A. D. Durai, R. A. Kumar, K. Singh, H. Palivela, B. R. Kumar, and B. Debtera, “Enhanced path routing with buffer allocation method using coupling node selection algorithm in manet,” Wireless Communications and Mobile Computing, vol. 2022, no. 1, p. 1955290, 2022. https://doi.org/10.1155/2022/1955290.

[21] Ummiti Sreenivasulu, Shaik Fairooz, R. Anil Kumar, Sarala Patchala, R. Prakash Kumar, Adireddy Rmaesh, “Joint beamforming with RIS assisted MU-MISO systems using HR-mobilenet and ASO algorithm, Digital Signal Processing, Volume 159, 2025, 104955, ISSN 1051-2004, https://doi.org/10.1016/j.dsp.2024.104955.

[22] C. Lu, H. Jiang, and X. Wang,” Reinforcement learning for UAV-assisted edge computing: Opportunities and challenges,” IEEE Wireless Communications, vol. 27, no. 3, pp. 108–114, 2020.

[23] A. J. Chinchawade, S. Rajyalaxmi, S. Singh, R. A. Kumar, R. Rastogi, and M. A. Shah, “Scheduling in multi-hop wireless networks using a distributed learning algorithm,” in 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2023, pp. 1013– 1018. https://doi.org/10.1109/ICOEI56765.2023.10125909.

[24] R. K. Bharti, D. Suganthi, S. Abirami, R. A. Kumar, B. Gayathri, and S. Kayathri, “Optimal extreme learning machine based traffic congestion control system in vehicular network,” in 2022 6th International Conference on Electronics, Communication and Aerospace Technology. IEEE, 2022, pp. 597–603. https://doi.org/10.1109/ICECA55336.2022.10009111.

[25] M. Alzenad, M. Z. Shakir, H. Yanikomeroglu, and M. S. Alouini,” FSO-based vertical backhaul/fronthaul framework for 5G+ wireless networks,” IEEE Communications Magazine, vol. 55, no. 3, pp. 218–225, 2017. https://doi.org/10.1109/MCOM.2017.1600735.