Graph-Based Multi-Agent Proximal Policy Optimization System for Trajectory Optimization and Resource Allocation in UAV-Assisted Emergency Communication Networks

Ahmed Raad Majeed AL Aloosi
Human Resource Departmet
Educational Directorate of Salah AL -Din Salah AL -Din,Tikrit, Iraq
Ahmedraad857@gmail.com

Abstract: The increasing requirement for effective emergency communications creates several difficulties in Unmanned Aerial Vehicles (UAV)-assisted networks, such as efficient coverage, resource allocation, and quality of service. Traditional reinforcement learning techniques frequently struggle with accommodating user mobility, dynamic changes in the environment, and multiple agents. Similarly, current trajectory optimization and resource allocation techniques lack sufficient coordination and efficient spatial representation. In this work, a Graph-based Multi-agent Proximal Policy Optimization with Flash Attention (GMPPO-FA) framework is proposed for joint UAV trajectory optimization and resource allocation. In this framework, multiple Proximal Policy Optimization (PPO) agents collaboratively optimize network operations through coordinated policy learning, while graph-based representations capture spatial and topological relationships among UAVs and users. The integration of Flash Attention enhances feature extraction by efficiently modeling long-range dependencies with reduced computational overhead, enabling faster convergence and improved policy learning. By combining graph learning, multi-agent reinforcement learning, and efficient attention mechanisms, GMPPO-FA adapts effectively to dynamic emergency environments and generates robust strategies for resource allocation and trajectory optimization. Experimental results demonstrate that GMPPO-FA outperforms conventional approaches, achieving superior data transmission rates, network coverage, resource utilization, and overall communication efficiency in UAV-assisted emergency communication networks.

Keywords: Reinforcement Learning; UAV-assisted Networks; UAV Trajectory Optimization and Resource Allocation; Graph-based Multi-agent Proximal Policy Optimization with Flash Attention

Received: June 18, 2026 Revised: June 25, 2026 Accepted: June 30, 2026

Nomenclature

Abbreviation	Description
UAV	Unmanned Aerial Vehicle
GMPPO-FA	Graph-based Multi-agent Proximal Policy Optimization with Flash Attention
GMPPO	Graph-based Multi-agent Proximal Policy Optimization
PPO	Proximal Policy Optimization
RL	Reinforcement Learning
QoS	Quality of Service
LLM	Large Language Model
LPTORA	Large Language Model-enabled Proximal Policy Optimization for Trajectory Optimization and Resource Allocation
NOMA	Non-Orthogonal Multiple Access

1. Introduction

1.1 Background of UAV-Assisted Emergency Communication Networks

Emergency Communication Networks aided by UAVs have emerged as an effective method to provide rapid wireless connectivity in disaster-affected areas when the traditional communication infrastructure is either destroyed or not available [7]. With the help of UAVs acting as temporary aerial base stations, emergency communication networks have been developed that could enable communication and rescue services and help emergency response teams and users [8]. However, efficient coordination among UAVs poses a difficult problem due to factors such as the dynamic nature of the network, constrained energy supply, interference among the UAVs, and continuous change in user requirements. Hence, it is important to optimize the UAVs' trajectories as well as allocate the communication resources in order to ensure better network coverage, improve the quality of service, and utilize the resources efficiently [9]. The recent developments in the field of graph-based learning and multi-agent reinforcement learning provide an important tool to model the interactions between different UAVs and coordinate their decision process [10]. More specifically, GMPPO provides an

efficient mechanism that takes into account the spatiotemporal relationship between UAVs and users and helps in coordinating the trajectory planning and resource allocation policy.

1.2 Challenges in Dynamic Trajectory Optimization and Resource Allocation

Several challenges currently exist in UAV-based emergency communications in terms of optimization of trajectory and resource allocation for efficient functioning of the network [11]. The first challenge is that the majority of the approaches use either static optimization techniques or a centralized decision-making process. As a result, these methods do not adapt well to changes in network conditions and emergencies [12]. Another problem is that most of the existing approaches separately perform trajectory optimization and allocation of resources without realizing the dependency between them. This often leads to ineffective use of resources, an increase in energy consumption, interference in communication, and low Quality of Service (QoS) [13].

Graph-based learning and multi-agent reinforcement learning have gained momentum lately as two effective means to solve these problems through coordination in decision-making among several UAVs [14]. With their help, it is possible to achieve adaptive planning and intelligent resource allocation owing to effective interaction modeling between UAVs and ground users. Nonetheless, there are many existing solutions that are unable to solve both problems simultaneously and maintain stability in a challenging and dynamically changing environment [15][16]. There is a great need for an intelligent solution that will allow combining graph-based learning and cooperation between multiple agents into one unified approach. Inspired by the problems, the GMAPPO framework is proposed.

1.3 Research Contributions

This section explains the major research contributions of the proposed framework.

- To develop a UAV-assisted emergency communication framework that deploys multiple UAVs as aerial base stations and utilizes graph-based multi-agent policy learning to optimize trajectory planning and communication resource allocation. The framework enhances communication coverage, cooperative decision-making, resource utilization, and network reliability while adapting to dynamic emergency scenarios.
- To implement a GMPPO framework integrated with Flash Attention for joint trajectory planning and resource allocation in UAV-assisted emergency communication networks. The graph-based representation effectively models the interactions among UAVs and ground users, while Flash Attention enhances the learning efficiency by focusing on the most relevant spatial features with reduced computational overhead. This enables cooperative decision-making, adaptive trajectory planning, and efficient resource allocation, thereby improving network coverage, communication reliability, and QoS in dynamic emergency environments.

This section gives an overview of the organization of the work. The literature review and research gap challenges are shown in Section 2. System modeling and network representation are presented in Section 3. Section 4 presents a graph-based multi-agent proximal policy optimization framework, followed by Section 5, which deals with results and discussion, and Section 6 gives the conclusion.

2. Literature Review and Research Gap Analysis

2.1 Related Works on UAV-Assisted Emergency Communication Networks

UAV-assisted emergency communication networks have inspired numerous research studies, which motivated the development of the proposed framework. In 2026, Chu *et al.* [1] proposed a large language model-enabled Proximal Policy Optimization (LPTORA) framework for joint UAV trajectory optimization and resource allocation in emergency communication networks. The study modeled user mobility and service requests using the Maxwell-Boltzmann distribution and Poisson process, respectively, while employing an LLM to automatically generate and refine reward functions for improving policy learning. In the same year, Wang and You [3] developed an effective scheme for the allocation of resources in emergency communication networks through simultaneous optimization of subcarrier assignment and power allocation schemes. An iterative technique with threshold-based rounding was used to enhance the performance of the communication network while ensuring effective resource management. In the same year, Guo *et al.* [6] proposed an emergency rescue communication system using multiple UAVs through simultaneous optimization of UAV scheduling,

transmission power allocation, and trajectory planning. In the year 2025, Ren *et al.* [4] proposed a multi-UAV-enabled NOMA emergency communication system in which both the deployment of UAVs and their transmit powers were jointly optimized for improving the energy efficiency of the system through particle swarm optimization and convex optimization. In 2024, Khan *et al.* [2] examined a multi-UAV-enabled public safety communication network through the joint optimization of UAV placement and communication resource allocation. It was shown that the proposed iterative optimization scheme greatly improved the utility of video transmission from areas affected by disasters. In 2023, Shahzadi *et al.* [5] proposed a resource allocation framework in UAV-enabled heterogeneous networks through optimization. An outer approximation-based optimization scheme was proposed for maximizing the network throughput under transmit power and QoS constraints.

2.2 Research Gaps

The use of UAV-based emergency communication networks is an effective way of delivering rapid and reliable wireless communication support in affected disaster zones when traditional communication facilities are not available. The use of several UAVs as aerial base stations will ensure that the communication range is expanded and that emergency services are supported in a dynamic environment. However, coordinating multiple UAVs becomes quite difficult due to constantly changing user needs, communication limitations, and a dynamic network. The key challenges addressed in this paper are as follows.

- Conventional trajectory planning methods are unable to effectively adapt to rapidly changing user distributions and dynamic network environments. As a result, they often produce inefficient flight paths, reduce network coverage, and increase communication delay during emergencies.
- Existing resource allocation approaches generally optimize communication resources independently of UAV trajectory planning. This separate optimization leads to inefficient bandwidth utilization, higher energy consumption, increased interference, and degraded QoS, particularly in large-scale emergency communication networks.
- Current multi-UAV coordination frameworks have limited capability to capture the complex spatial relationships among UAVs and ground users. Consequently, cooperative decision-making becomes less effective, resulting in slower convergence, increased computational overhead, and reduced adaptability under highly dynamic emergency scenarios.

3. System Modeling And Network Representation

The suggested framework utilizes graph representation learning, Flash Attention, and multi-agent reinforcement learning together for the joint optimization of both UAV trajectory planning and communication resource allocation within dynamic emergency communication networks. With the combination of these components within one framework, GMPPO-FA is able to capture spatial relations, make cooperative decisions, and enhance communication coverage and resource efficiency. Initially, the communication environment is initialized with multiple UAVs, ground users, and network parameters, including UAV transmission power, total wireless access bandwidth, system noise power, SNR threshold, learning rate, discount factor, GAE parameter, clipping parameter, value loss coefficient, entropy regularization coefficient, and gradient norm threshold. The communication network is then represented as a graph, where UAVs and ground users are modeled as interconnected nodes according to their communication relationships. The constructed graph is provided as the input to the proposed GMPPO-FA network. Within the framework, the Flash Attention mechanism efficiently extracts the most relevant spatial features by focusing on significant node interactions, while the MAPPO component utilizes the enhanced graph representations to enable multiple UAV agents to cooperatively learn optimal trajectory planning and communication resource allocation policies. The learning process is guided by a reward function that promotes improved communication coverage, efficient bandwidth utilization, reduced communication latency, and reliable network connectivity. Through continuous policy updates, the proposed GMPPO-FA framework converges toward an effective cooperative strategy for dynamic UAV-assisted emergency communication networks. The pictorial representation of the proposed trajectory optimization and resource allocation framework is illustrated in Fig. 1.

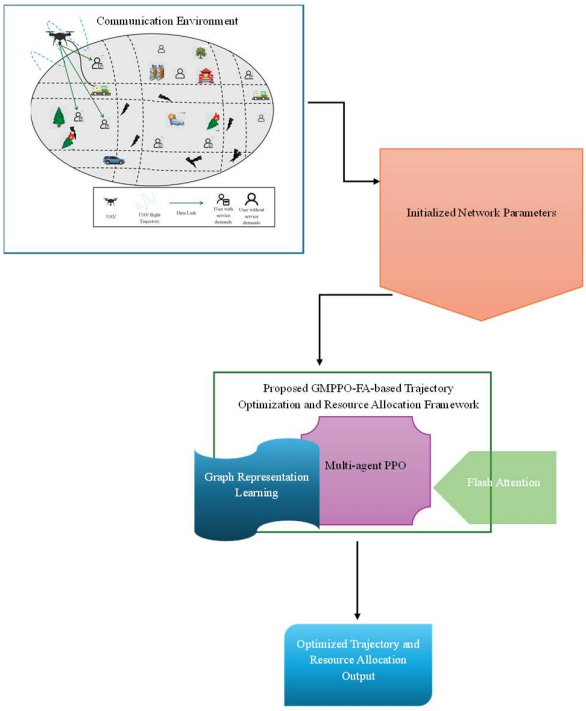


Fig. 1. Overall Architecture of the Proposed Trajectory Optimization and Resource Allocation Network

3.1 System Model: UAV-Assisted Emergency Communication Networks

The proposed UAV-assisted emergency communication system consists of multiple UAVs deployed as aerial base stations to provide temporary wireless connectivity for ground users in disaster-affected regions where the terrestrial communication infrastructure is unavailable or severely damaged. Let $v = \{1, 2, .., V\}$ denote the set of UAVs and $h = \{1, 2, ..., H\}$ represent the set of ground users. The ground users are randomly distributed throughout the emergency area, and their communication demands vary dynamically according to the disaster situation.

Each UAV maintains wireless communication channels with ground users around it, while coordinating with its neighboring UAVs to ensure continuous connection in the network. In the process, each UAV adjusts its trajectory according to the varying location of the users, communication needs, and network conditions. At the same time, dynamic allocation of communication resources such as bandwidth and transmit power is made in order to meet the QoS needs of the ground users. In order to accurately characterize such interactions within the network, the UAV-enabled communication network is considered as a graph where UAVs and ground users are the nodes, whereas the communication links established among them form the graph edges. Through this graph-based approach, it becomes feasible to efficiently model the spatial relationships and communication interactions within the network. On the basis of the network state observed, the multiple UAVs jointly decide on their motion control and resource management strategies to increase network coverage, improve communication reliability, optimize energy consumption, and sustain network connectivity. The system model of the UAV-assisted emergency communication networks is depicted in Fig. 2.

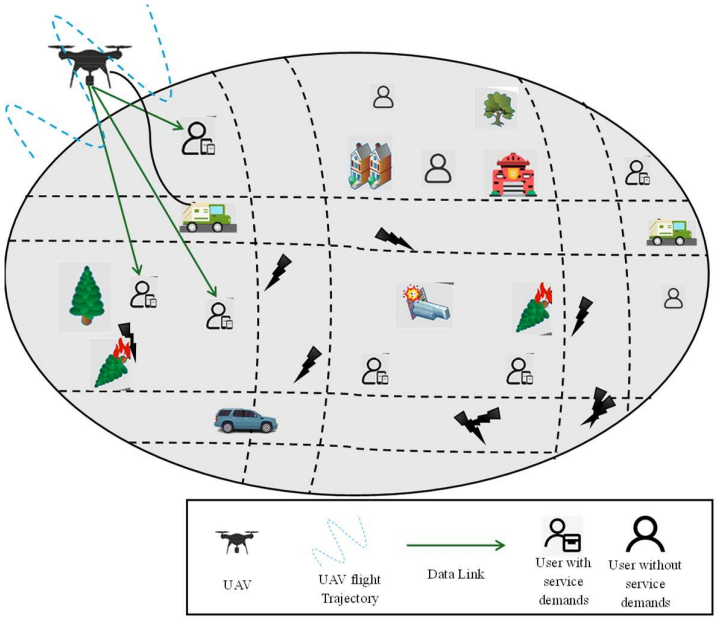


Fig. 2. System Model of the UAV-Assisted Emergency Communication Networks3.2 Problem Formulation

Due to the dynamic characteristics of the emergency communications network, the presented framework jointly optimizes UAV paths and communication resource assignment in order to enhance the performance of the network. The goal is to increase communication efficiency by adjusting to the variations in user distribution and demand while also improving network coverage, reducing communication delays, and making effective use of communication resources. The optimization objective is expressed in Eq. (1).

$$\max_{Q(t),\alpha_{v,k}(t),c_{v,k}(t)} \sum_{v\in V}\sum_{k\in H} \alpha_{v,k}(t) S_{v,k}(t) \tag{1}$$

With subject to,

$$D1: e_k(t) \in \{0,1\}, \quad \forall k \in H, \forall t$$

$$D2: \alpha_{v,k}(t) \in \{0,1\}, \quad \forall v \in V, k \in H, \forall t$$

$$D3: \sum_{v\in V} \alpha_{v,k}(t) \le 1, \quad \forall k, \forall t$$

$$D4: \alpha_{v,k}(t) \le e_k(t), \quad \forall v, k, \forall t$$

$$D5: \Gamma_{v,k}(t) \ge \Gamma_{\min}, \quad \forall v, k, \forall t$$

$$D6: \sum_{k\in H} c_{v,k}(t) \le C, \quad \forall v, \forall t$$

$$D7: \left\| Q_v(t+1) - Q_u(t) \right\| \le W_{\max}\Delta t, \quad \forall v, \forall t$$

The formulated optimization problem is subject to several operational constraints. Constraint $D1$ indicates whether ground user jhas an active communication request during time slot t. Constraint $D2$ defines the association between UAV v and ground user j. Constraint $D3$ ensures that each ground user is served by at most one UAV during a time slot. Constraint $D4$ guarantees that communication resources are allocated only to users with active service demands. Constraint $D5$ maintains reliable communication by satisfying the minimum Signal-to-Noise Ratio (SNR) requirement. Constraint $D6$ limits the total bandwidth allocated by each UAV so that the available communication resources are not exceeded. Constraint $D7$ restricts the UAV movement according to the maximum allowable flight speed, ensuring feasible trajectory planning throughout the emergency communication mission.

4. Graph-based Multi-Agent Proximal Policy Optimization Framework

The developed GMPPO-FA framework integrates graph-based representation learning, Flash Attention, and multi-agent reinforcement learning for UAV-assisted emergency communication networks. The communication environment is modeled as a graph to capture the interactions

among UAVs and ground users, while Flash Attention extracts important spatial features. These features are then utilized by the multi-agent PPO framework for cooperative trajectory planning and adaptive resource allocation under dynamic emergency conditions.

4.1 Multi-Agent Proximal Policy Optimization

Multi-Agent Proximal Policy Optimization (MAPPO) [18] is used to support collaboration amongst a group of UAVs in emergency communication systems. A centralized training and decentralized execution approach is utilized, where all UAVs make decisions based on their observations while using the knowledge of the entire system during the training process. This approach uses an Actor-Critic structure, where the Actor network determines the optimal policy of navigation and communication resource distribution, while the Critic network assesses the state value associated with the respective policy.

For a given system state s_t, the actor selects an action according to the policy, which is expressed in Eq. (2).

$$t_b \sim \pi_\theta\left(t_b \mid t_s\right) \tag{2}$$

Here, the policy parameterized by θ is denoted as π_θ, and the critic estimates the state-value function as in Eq. (3).

$$W_\phi\left(t_s\right) = F_\pi\left[\sum_{l=0}^{\infty}\gamma^l t_{r+l} \mid t_s\right] \tag{3}$$

Here, the term t_r is the reward at timestep t, γ is the discount vector, and ϕ represents the critic parameters. The objective of MAPPO is to maximize the expected cumulative reward, which is expressed in Eq. (4).

$$K\left(\theta\right) = F\left[\sum_{t=0}^{\infty}\gamma^t t_s\right] \tag{4}$$

The policy parameters are updated using the policy gradient, which is formulated in Eq. (5).

$$\Delta K\left(\theta\right) = F\left[\Delta \log \pi_\theta\left(t_a \mid t_s\right) R\left(t_s, t_a\right)\right] \tag{5}$$

Here, the action-value function is indicated as $R\left(t_s, t_a\right)$. To reduce the variance during training, the advantage function is computed in Eq. (6).

$$B\left(t_s, t_a\right) = R\left(t_s, t_a\right) - W\left(t_s\right) \tag{6}$$

The actor network is optimized using the PPO clipped objective, which is formulated in Eq. (7).

$$M_{act}\left(\theta\right) = F\left[\min\left(t_r\left(\theta\right) t_B, clip\left(t_r\left(\theta\right), 1-\varepsilon, 1+\varepsilon\right) t_B\right)\right] \tag{7}$$

Where t_r is refined in Eq. (8).

$$t_r\left(\theta\right) = \frac{\pi_\theta\left(t_a \mid t_s\right)}{\pi_{\theta_{old}}\left(t_a \mid t_s\right)} \tag{8}$$

Here, t_r is the probability ratio between the updated and previous policies. Finally, the Actor and Critic networks are iteratively updated through gradient descent until convergence. This enables multiple UAVs to cooperatively learn efficient trajectory planning and adaptive communication resource allocation policies under dynamic emergency communication scenarios.

4.2 GMPPO-FA: Trajectory Optimization and Resource Allocation Network

An innovative GMPPO-FA technique is presented to enhance trajectory optimization and resource allocation in UAV-aided emergency communication networks. The technique increases the spatial modeling capability, improves multi-agent coordination, and facilitates rapid decision-making based on context awareness in highly dynamic scenarios. Through a combination of graph-based modeling, cooperative PPO agents, and Flash Attention for effective long-term dependency learning, GMPPO-FA speeds up convergence, maximizes resource efficiency, and provides more dependable communication solutions.

Novelty: The GMPPO-FA approach uniquely leverages graph-based modeling along with multi-agent PPO and Flash Attention to address trajectory planning and resource allocation of UAVs in emergencies. The combination helps ensure better adaptability and precision even in dynamic situations and, at the same time, saves on computational costs and ensures quick convergence. This results in better coverage, faster data transfer, and more efficient use of resources than existing approaches.

Implementation Steps: At first, the communication environment is initialized with several UAVs, ground users, and several system parameters that include the transmission power of UAVs, wireless access bandwidth, noise power of the system, signal-to-noise ratio threshold, learning rate, discount factor, generalized advantage estimation parameter, clipping parameter, value loss coefficient, entropy regularization coefficient, and gradient norm threshold. The communication network is represented in the form of a graph in which the UAVs and ground users act as graph nodes, whereas the communication connections among them are modeled as edges in the graph. The communication graph is represented as in Eq. (9).

$$H = (W, F) \tag{9}$$

Here, the term W denotes the set of graph nodes corresponding to UAVs and ground users, and Erepresents the communication links among them. The constructed communication graph is then provided as the input to the GMPPO-FA framework, where it guides cooperative learning among multiple UAVs. Within the framework, the MAPPO component leverages this graph data to enable collaboration between agents. During the central training phase, UAVs collectively learn a shared policy through interactions in the communication environment. In decentralized execution, each UAV applies this policy to make independent decisions about its trajectory and resource allocation based on local observations. The Actor network generates the optimal actions for each UAV, while the Critic network evaluates the corresponding state values. Through continuous interaction with the environment and ongoing policy updates, the MAPPO element learns cooperative strategies that expand coverage, improve bandwidth efficiency, and reduce communication delays in emergency scenarios. To further enhance the quality of the learned graph representations, the proposed GMAPPO-FA framework integrates the Flash Attention mechanism.

Flash Attention (FA) [17] is incorporated to efficiently learn the most informative spatial relationships among UAVs and ground users while reducing memory consumption and computational complexity. First, the features in the graph are transformed into C, S and K matrices. Next, the attention scores are calculated in order to discover the important spatial relations among the nodes of the network. The process of attention is defined in Eq. (10).

$$Attention(C, S, K) = Soft\max\left(\frac{CS^{N}}{\sqrt{e}}\right)K \tag{10}$$

Here, the query, key, and value matrices are indicated as $C, S,$ and K , and e represents the feature dimension. In order to increase efficiency in the computational process, FA splits the attention matrices into smaller blocks and computes them inside fast, high-speed on-chip memory. Block computing helps to process the graph representation efficiently and, at the same time, maintain the accuracy of attention scores. This block-wise computation enables efficient processing of large graph representations while preserving the accuracy of the learned attention scores. By integrating graph-based representation learning, FA, and MAPPO, the proposed GMPPO-FA framework achieves efficient trajectory optimization and adaptive communication resource allocation in UAV-assisted emergency communication networks. The framework enables cooperative policy learning, captures complex spatial relationships, and supports reliable decision-making under dynamic emergency conditions. Consequently, it improves communication coverage, bandwidth utilization, resource efficiency, and overall network performance. The systematic architecture of the proposed GMPPO-FA-based trajectory optimization and resource allocation framework is illustrated in Fig. 3.

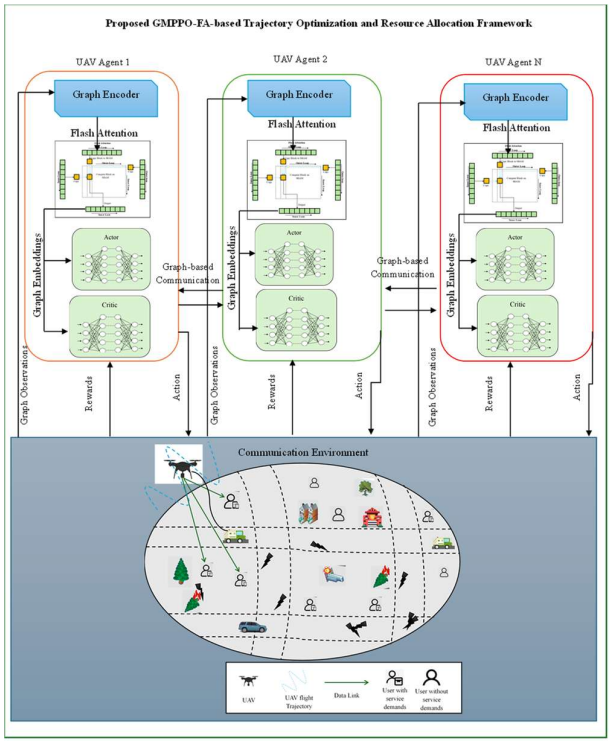


Fig. 3. Systematic Representation of the Proposed GMPPO-FA-based Trajectory Optimization and Resource Allocation Framework

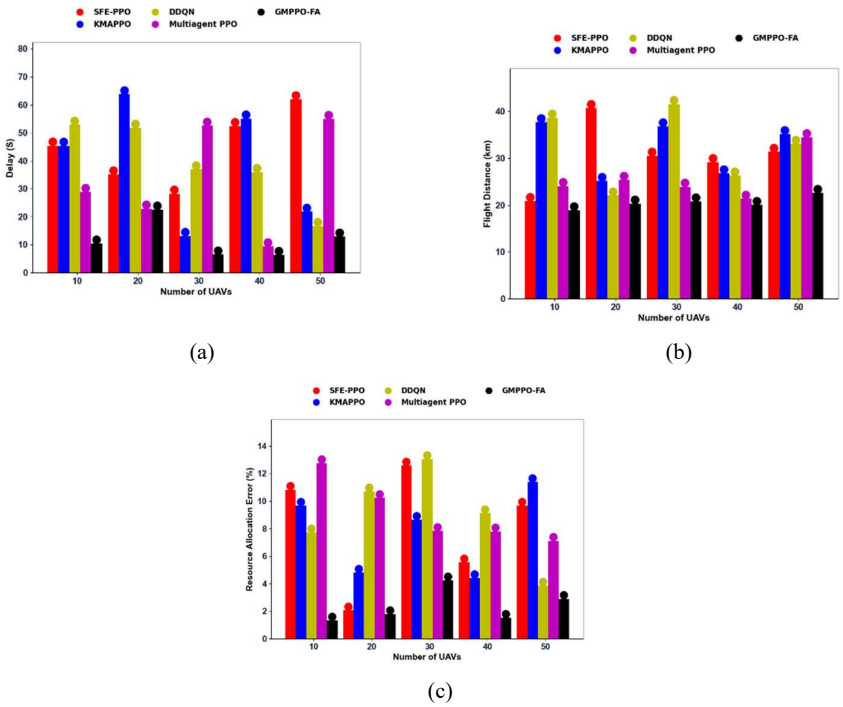
5. Results and Discussion

5.1 Simulation Setup

The proposed framework was implemented and evaluated using a Python-based simulation environment for UAV-assisted emergency communication networks. The performance of the proposed framework was compared to some of the advanced reinforcement learning frameworks such as SFE-PPO [19], KMAPPO [20], DDQN [21], and Multi-Agent PPO [22]. The comparison was performed using the metrics of data transmission rate, communication coverage, resource utilization, communication latency, energy efficiency, rate of convergence, and network throughput. The simulation results show that the proposed GMAPPO-FA framework performs better in optimizing cooperative trajectory and communication resource allocation.

5.2 Performance Analysis of the Developed Model

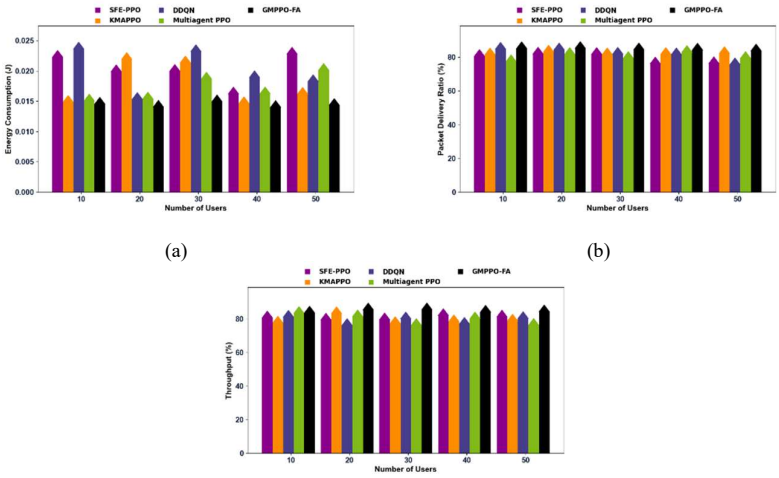
Figs. 4 and 5 present the comparative performance of the proposed framework by varying the number of UAVs and users with SFE-PPO, KMAPPO, DDQN, and Multi-Agent PPO. The proposed framework achieves the lowest communication delay of approximately 15 s, the shortest flight distance of about 21 m, and the minimum resource allocation error of nearly 2%. Moreover, it records the lowest energy consumption of around 0.015 J, while attaining a packet delivery ratio of approximately 89% and a throughput of about 89%, outperforming the existing approaches. Furthermore, the proposed framework reduces the resource allocation error to nearly 2%, demonstrating its superiority in trajectory optimization, resource allocation, and overall communication performance under dynamic emergency communication scenarios.



(a) (b)

(c)

Fig. 4.Performance Analysis of the Developed Model by varying the Number of UAVs based on (a) Delay, (b) Flight Distance, and (c) Resource Allocation Error



(a) (b)

(c)

Fig. 5. Performance Analysis of the Developed Model by varying the Number of Users based on (a) Energy Consumption, (b) Packet Delivery Ratio, and (c) Throughput

5.3 Performance Assessment of the Designed Approach by varying the Learning Rate

Table 1 presents the performance assessment of the designed framework under different learning rates. The results indicate that the proposed framework consistently achieves superior performance across all evaluation metrics compared with SFE-PPO, KMAPPO, DDQN, and Multiagent PPO. Specifically, GMPPO-FA attains a maximum network coverage of 89.901%, resource utilization of 89.922%, service success ratio of 93.607%, packet delivery ratio of 89.806%, QoS of 89.972%, fairness index of 0.976, and throughput of 89.644%. Furthermore, it records the lowest energy consumption of 0.015 J, communication delay of 11.317 s, flight distance of 20.591 m, and resource allocation error of 2.021%, demonstrating its effectiveness in improving communication reliability, resource efficiency, and trajectory optimization under varying learning rates.

Table 1: Performance Assessment of the Designed Approach by Varying Learning Rate

Learning Rate	SFE-PPO [19]	KMAPPO [20]	DDQN [21]	Multiagent PPO [22]	GMPPO-FA
Network Coverage					
0.011	87.992	83.396	85.933	82.704	88.368
0.022	83.365	88.967	81.746	84.372	89.901
0.033	87.546	87.877	87.756	85.415	89.620
0.044	80.795	83.282	81.646	82.110	89.848
0.055	86.921	88.211	86.498	80.080	89.571
Resource Utilization					
0.011	86.831	83.167	85.196	81.724	89.922
0.022	81.926	85.928	81.093	84.547	87.265
0.033	85.416	82.266	86.679	84.783	87.984
0.044	85.270	81.171	86.239	86.516	89.471
0.055	81.598	81.546	83.679	85.410	86.907
Service Success Ratio					
0.011	90.446	93.117	89.606	90.330	93.298
0.022	91.027	92.826	89.232	92.470	93.413
0.033	90.063	90.089	88.617	90.734	92.160
0.044	88.369	89.789	91.875	90.375	93.286
0.055	88.842	90.242	90.431	89.726	93.607
Packet Delivery Ratio					
0.011	86.983	82.955	89.187	87.199	89.370
0.022	88.377	84.819	89.764	81.496	89.806
0.033	84.887	82.587	81.445	87.151	89.402
0.044	85.384	81.796	84.902	85.657	88.120
0.055	81.522	82.704	85.473	85.339	87.309
Energy Consumption					
0.011	0.019	0.019	0.020	0.023	0.015
0.022	0.018	0.019	0.015	0.025	0.015
0.033	0.019	0.023	0.020	0.021	0.016
0.044	0.021	0.021	0.017	0.021	0.015
0.055	0.022	0.021	0.022	0.023	0.015
QoS					
0.011	81.152	81.457	83.196	87.075	88.607
0.022	82.049	85.232	87.830	81.632	89.087
0.033	86.557	83.775	86.391	87.496	88.717
0.044	88.753	88.941	83.384	83.536	89.239
0.055	83.748	85.830	87.977	83.850	89.972
Fairness Index					
0.011	0.535	0.394	0.797	0.814	0.954
0.022	0.241	0.391	0.629	0.461	0.924
0.033	0.711	0.667	0.476	0.537	0.786
0.044	0.494	0.246	0.543	0.583	0.976
0.055	0.699	0.713	0.552	0.853	0.874
Delay					
0.011	70.905	63.448	22.508	57.609	12.640
0.022	34.555	40.107	44.466	14.753	11.317
0.033	76.549	77.255	53.312	44.947	28.714
0.044	69.640	46.604	76.795	61.131	14.945
0.055	33.955	28.170	71.136	64.007	12.395
Flight Distance					
0.011	34.794	40.259	37.846	30.586	21.570
0.022	28.384	38.799	34.482	43.505	22.819
0.033	37.996	35.672	37.463	32.332	20.591
0.044	42.027	25.465	44.337	35.997	21.032
0.055	36.781	29.324	44.509	44.758	23.967
Resource Allocation Error					
0.011	14.071	10.114	14.450	8.486	3.934
0.022	14.179	10.669	8.801	12.574	2.424
0.033	12.695	12.906	5.020	7.076	2.810
0.044	4.586	11.696	14.998	14.022	2.021
0.055	14.779	6.173	10.083	11.007	6.138
Throughput					
0.011	84.020	88.651	87.013	82.799	89.077
0.022	83.676	82.101	82.506	83.927	85.498
0.033	86.537	83.712	80.469	82.975	89.183
0.044	89.486	82.186	82.416	81.261	89.644
0.055	82.920	84.636	87.563	80.531	88.636

5.4 Episode-based Performance Evaluation of the Proposed Framework

Fig. 6 illustrates the performance of the proposed framework under varying training episodes in terms of average service distance, data transmission rate, and service ratio. As the number of training episodes increases, the proposed model consistently outperforms the existing SFE-PPO, KMAPPO, DDQN, and Multiagent PPO approaches. After 1000 training episodes, the proposed GMPPO-FA reduces the average service distance to approximately 176 m, whereas Multiagent PPO, DDQN, KMAPPO, and SFE-PPO achieve nearly 191 m, 205 m, 224 m, and 253 m, respectively. Similarly, the proposed framework attains a maximum data transmission rate of about 1.0 Mbps, while Multiagent PPO, DDQN, KMAPPO, and SFE-PPO achieve approximately 0.91 Mbps, 0.79 Mbps, 0.67 Mbps, and 0.48 Mbps, respectively. Moreover, the service ratio of GMPPO-FA reaches nearly 100%, outperforming Multiagent PPO of 99%, DDQN of 93%, KMAPPO of 85%, and SFE-PPO of 71%. These results demonstrate that the proposed framework achieves faster convergence, higher communication efficiency, and more reliable UAV-assisted emergency communication than the existing reinforcement learning methods.

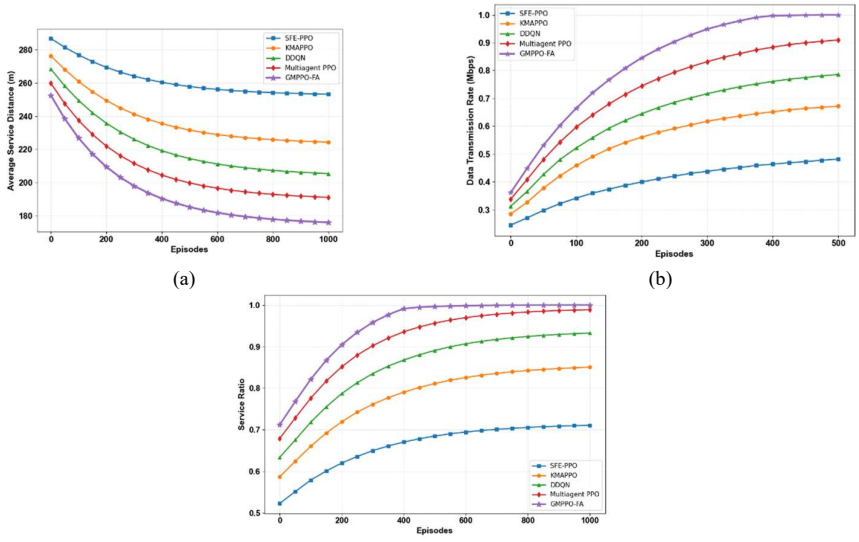


Fig. 6.Episode-based Performance Evaluation of the Proposed Framework in terms of (a) Average Service Distance, (b) Data Transmission Rate, and (c) Service Ratio

6. Conclusion

An innovative deep reinforcement learning-based framework was developed for trajectory optimization and resource allocation in UAV-assisted emergency communication networks. The proposed architecture successfully resolved the issues related to user mobility, limited communication resources, and UAV cooperation. In the first place, the communication network was formulated as a graph, in which the connections among UAVs and ground users were formulated using graph nodes and edges. After that, the Flash Attention algorithm was applied to graph attributes to efficiently capture important spatial dependencies. Finally, by applying the MAPPO approach, multiple UAV agents learned to optimize their trajectories and allocate communication resources through interacting with the environment. The experimental results clearly showed that the suggested algorithm performed better than SFE-PPO, KMAPPO, DDQN, and Multiagent PPO in different performance parameters. In particular, the values obtained by the GMPPO-FA approach were 89.9% for network coverage, 89.9% for resource utilization, 93.6% for the service success rate, 89.8% for the packet delivery ratio, 89.6% for throughput, 0.976 for the fairness index, 11.3 s for communication delay, 0.015 J for energy consumption, 20.6 m for flight distance, and 2.0% for resource allocation error. Thus, the presented GMPPO-FA framework can provide efficient trajectory planning and effective communication resource allocation for dynamic UAVs-supported emergency communication networks.

Drawbacks and Future Directions: The developed framework required high computational resources and longer training time, which may limit its deployment in large-scale

emergency communication networks. Future work can focus on lightweight learning models and online learning techniques to improve scalability, adaptability, and real-time performance in dynamic UAV-assisted communication environments.

References

[1] Chu, C., Zhang, J., Zhang, Y., He, P., Chen, Y., and Liu, Q., "Large Language Model Enabled DRL for Trajectory Optimization and Resource Allocation in UAV-assisted Emergency Communication Networks", Physical Communication, pp. 103177, 2026.

[2] Khan, N., Ahmad, A., Wakeel, A., Kaleem, Z., Rashid, B., and Khalid, W., "Efficient UAVs deployment and resource allocation in UAV-relay assisted public safety networks for video transmission", IEEE Access, vol. 12, pp. 4561–4574, 2024.

[3] Wang, T., and You, C., "Joint subcarrier assignment and power allocation for UAV-assisted full-duplex NOMA emergency communication networks", Journal of Communications and Networks, 2026.

[4] Ren, Y., Wang, J., Xiao, Y., Zhang, X., Jiang, F., Wang, J., Jiang, J., and Lu, G., "Energy Efficiency Optimization for UAV Distribution and Resource Allocation in NOMA and Multi-UAV Assisted Wireless Networks", IEEE Open Journal of the Communications Society, 2025.

[5] Shahzadi, R., Ali, M., and Naeem, M., "Combinatorial resource allocation in UAV-assisted 5G/B5G heterogeneous networks", IEEE Access, vol. 11, pp. 65336–65346, 2023.

[6] Guo, S., Li, Y., Song, Z., Liu, L., and Zhou, L., "Resource optimization and trajectory planning approach for multi-UAV assisted heterogeneous emergency rescue communication networks", Chinese Journal of Aeronautics, pp. 104256, 2026.

[7] Ji, B., Huo, Z., Li, B., Li, Y., Fan, H., and Li, Z., "Resource Allocation for UAV-Assisted Vehicular Networks with Integrated Sensing, Communication, and Computing in Emergency Scenarios", Physical Communication, pp. 103209, 2026.

[8] Liu, F., Gao, Y., and Du, R., "Joint optimization of resource allocation and position deployment in UAV swarm emergency communication networks under interference", Ad Hoc Networks, pp. 104152, 2026.

[9] Zhou, S., Niu, X., Lu, S., Xiang, Y., and Pu, C., "Self-organized anti-jamming reinforcement learning for resource allocation in UAV-assisted networks", Signal Processing, pp. 110346, 2025.

[10] Liu, Z., Liu, X., Liu, Y., Leung, V. C. M., and Durrani, T. S., "UAV-assisted integrated sensing and communications for Internet of Things: 3D trajectory optimization and resource allocation", IEEE Transactions on Wireless Communications, vol. 23, no. 8, pp. 8654–8667, 2024.

[11] Zhang, Z., Wang, Y., Luo, Y., Zhang, H., Zhang, X., and Ding, W., "Iterative trajectory planning and resource allocation for UAV-assisted emergency communication with user dynamics", Drones, vol. 8, no. 4, pp. 149, 2024.

[12] Qiu, W., "Intelligent Optimization of UAV Swarm-Assisted Emergency Communication Networks: Resource Allocation and Trajectory Planning", 2025.

[13] Pan, Y., Li, R., Da, X., Hu, H., Zhang, M., Zhai, D., Cumanan, K., and Dobre, O. A., "Cooperative trajectory planning and resource allocation for UAV-enabled integrated sensing and communication systems", IEEE Transactions on Vehicular Technology, vol. 73, no. 5, pp. 6502–6516, 2023.

[14] Xie, P., Chen, Q., Wu, J. Y., Gao, X., Xing, L., Zhang, Y., and Sun, H., "NOMA-based intelligent resource allocation and trajectory optimization for multi-UAVs assisted semantic communication networks", Ad Hoc Networks, vol. 171, pp. 103762, 2025.

[15] Meem, T. A., Yu, H., and Ju, M., "Joint Optimization of UAV Trajectory and Resource Allocation in Multi-UAV-enabled Wireless Powered Communication Networks Under Max-Min Criterion", IEEE Access, 2025.

[16] Wang, J., Zhang, H., Zhou, X., Liu, W., and Yuan, D., "Joint resource allocation and trajectory design for energy-efficient UAV-assisted networks with user fairness guarantee", IEEE Internet of Things Journal, vol. 11, no. 13, pp. 23835–23849, 2024.

[17] Yang, Q., Bai, Y., Liu, F., and Zhang, W., "Integrated visual transformer and flash attention for lip-to-speech generation GAN", Scientific Reports, vol. 14, no. 1, pp. 4525, 2024.

[18] Gao, Z., Hao, H., Gao, F., and Zhao, R., "Behavior-constrained multi-agent proximal policy optimization for cooperative vehicle control at future intersections", Automotive Innovation, vol. 8, no. 4, pp. 896–912, 2025.

[19] Chu, C., Zhang, J., He, P., Zhang, Y., Ouyang, M., Wan, F., Liu, Q., and Chen, Y., "Trajectory Optimization and Resource Allocation for UAV-Assisted Emergency Communication Networks", Drones, vol. 10, no. 4, pp. 233, 2026.

[20] Guan, Y., Zou, S., Peng, H., Ni, W., Sun, Y., and Gao, H., "Cooperative UAV trajectory design for disaster area emergency communications: A multiagent PPO method", IEEE Internet of Things Journal, vol. 11, no. 5, pp. 8848–8859, 2023.

[21] Xu, J., "Efficient trajectory optimization and resource allocation in UAV 5G networks using dueling-Deep-Q-Networks", Wireless Networks, vol. 30, no. 7, pp. 6687–6697, 2024.

[22] Rehman, A. U., Sualiheen, S., and Chang, K., "Adaptive resource allocation in 6G A2G-TN integrated system: A cross-layer multi-agent PPO approach", Computer Networks, pp. 111655, 2025.