Signal Processing 230 (2025) 109844

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Maximizing coverage in UAV-based emergency communication networks using deep reinforcement learning

Le Zhao [a,b], Xiongchao Liu [c,*], Tao Shang [b,c]

[a] School of Telecommunication Engineering, Xidian University, Xi'an, 710071, China
[b] State Key Laboratory of Integrated Service Network, Xidian University, Xi'an, 710071, China
[c] Hangzhou Institute of Technology, Xidian University, Hangzhou, 311202, China

ARTICLE INFO

Keywords:
Emergency communication
Deep reinforcement learning (DRL)
UAV trajectory

ABSTRACT

In the optimization of traditional Unmanned Aerial Vehicle (UAV) emergency communication systems in response to natural disasters, existing studies often overlook the simultaneous optimization of time-domain and frequency-domain resources, which leads to low communication efficiency and limited coverage. To address these issues, we propose a UAV-based emergency communication system with a HybridComm architecture. This architecture optimizes the UAV's aerial position, uplink and downlink time slot ratio, and bandwidth allocation based on feedback from transmission rates and channel losses, ensuring optimal resource allocation. Additionally, while ensuring full-duplex communication and a minimum data transmission rate for all nodes, we designed a communication priority mechanism to ensure the communication quality of rescue nodes. Simulation results using deep reinforcement learning show that with a 1.5 Mb/s threshold, the number of ground nodes covered by communication converges to approximately 820 per session, an increase of 170 nodes compared to an amorphous reward function. Furthermore, in systems without priority settings, the number of rescue personnel devices covered per event is nearly zero, while systems with priority settings cover approximately 175 devices per event, thereby ensuring high-quality communication for rescue operations. This study offers an innovative approach to enhancing the efficiency and reliability of disaster response communications.

1. Introduction

The excessive exploitation of Earth's resources has disrupted the ecological balance, leading to frequent natural disasters. These events often damage communication networks, hindering the timely dissemination of rescue information. To address these challenges, UAV equipped with aerial communication systems have emerged, integrating network and image transmission technologies to enhance the interconnectivity of images, voice, and data [1,2]. In August 2023, Beijing, China, suffered from flooding caused by Typhoon Doksuri. The DG-M20 tethered UAV, equipped with Beijing Mobile devices, was rapidly deployed in the Changping District, Beijing. Powered by its tether system and utilizing the Asia-Pacific 6D communication satellite, it was able to cover a 3-square-kilometer area at an altitude of 200 m, supporting 1400 users and enabling all-weather, cross-regional communication. Compared with traditional emergency rescue, it can eliminate the constraints of optical cables and electricity in the original infrastructure and achieve rapid deployment of emergency communications. However, the communication range of tethered UAV is limited by the length of their tether cables.

To overcome these limitations, untethered UAV have garnered increasing attention in the field of emergency communication due to their mobility, rapid deployment capabilities, and scalability [3,4]. The communication performance of UAV-assisted systems is influenced by factors such as UAV trajectory, transmission power, and bandwidth. In the study by Meng et al. the energy consumption model is decomposed into horizontal and vertical flight energy components [5]. Under energy constraints, they optimized the three-dimensional trajectory of UAV in probabilistic line-of-sight (LoS) channels to maximize the minimum expected transmission data for ground devices (GDs). Beyond trajectory optimization, researchers have also jointly optimized transmission power and spectrum to achieve optimal system performance [6]. The optimization problems in UAV-assisted communication systems are generally non-convex, and DRL algorithms are a suitable solution. Extensive research has integrated artificial intelligence algorithms through the establishment of Markov Decision Process (MDP) models, enabling UAV to adapt to more complex scenarios [7]. DRL algorithms can monitor network performance in real time and dynamically adjust network parameters according to evolving disaster

* Corresponding author.
E-mail address: liuxiongchao@xidian.edu.cn (X. Liu).

https://doi.org/10.1016/j.sigpro.2024.109844
Received 13 August 2024; Received in revised form 6 November 2024; Accepted 5 December 2024
Available online 12 December 2024

conditions, thereby optimizing resource allocation. The successful application of UAV in post-disaster rescue operations has significantly accelerated the establishment of communication links, overcoming the limitations of traditional vehicle-based emergency communication systems in terms of flexibility and range. Therefore, researching UAV-based emergency communication systems to enhance communication performance is critically important. However, existing studies still have limitations. Overall, resource optimization has not fully considered the simultaneous optimization of time-domain and frequency-domain resources, indicating potential areas for improvement. Furthermore, there is limited research on role differentiation in emergency communication scenarios, particularly concerning the prioritization of multi-node simultaneous communication under full-duplex conditions.

In this study, we propose an emergency communication system designed for suburban environments following seismic events. Addressing the challenge of swiftly establishing effective emergency communication links between ground equipment and external rescue organizations post-disaster, the system employs emergency communication vehicles to relay information from the backend rescue center to UAV base stations deployed in the air. In order to optimize communication efficiency, we propose HybridComm, a novel communication technology based on a dual-layer architecture, which establishes a data transmission rate model by intelligently allocating uplink and downlink time and spectrum resources. For optimal resource allocation, DRL is introduced, and a Markov decision process model is developed to maximize effective communication coverage while maintaining the lowest data transmission rate. The simulation results indicate that, while ensuring a minimum communication rate of 1.5 Mb/s, the number of ground communication nodes covered throughout the entire rescue operation reaches approximately 820. In this study, our primary contributions are:

1. We have proposed a UAV-based emergency communication system specifically designed for suburban environments, addressing the often-overlooked issues of role distinction and communication prioritization in disaster-stricken areas. The introduced communication prioritization mechanism operates in full-duplex mode, ensuring a minimum communication rate for all nodes while guaranteeing reliable communication quality for high-priority rescue personnel. Our proposed solution optimizes resource allocation to enhance communication coverage while maintaining a high quality of experience (QoE) for users.

2. Our system tackles the often-overlooked issue of simultaneous optimization of uplink–downlink time allocation and bandwidth resources in UAV communications by introducing a dual-layer HybridComm architecture based on DRL. This architecture employs upper-layer Time Division Duplex (TDD) and lower-layer Frequency Division Multiple Access (FDMA) to optimize the UAV's flight paths and resource allocation, maximizing communication coverage while ensuring the minimum transmission rate.

3. To address the limitations of convex optimization methods in MDP models, we adopted the Proximal Policy Optimization (PPO) algorithm and designed an innovative reward function to incentivize greater coverage. Simulation results indicate that this approach significantly enhances coverage performance while maintaining prioritization. With a minimum transmission rate of 1.5 Mb/s, the number of ground nodes covered per episode converged to approximately 820, which represents an increase of about 170 nodes compared to methods without the modified reward function.

The subsequent sections of this paper are structured as follows. Section 2 discusses related work. Section 3 presents the system model employed in our study. Section 4 introduces the UAV coverage maximization problem. In Section 5, we provide pertinent simulation results, validating the system's performance under various conditions. Finally, Section 6 encapsulates the comprehensive summary of our research endeavors.

2. Related works

In recent years, the application of UAV in auxiliary communication has flourished. As previously mentioned, during disasters such as earthquakes and snowstorms, the rapid establishment of emergency communication networks is crucial for post-disaster rescue efforts. Consequently, there has been a growing body of research focused on emergency communication networks within post-disaster rescue scenarios [8–12].

Disasters may occur at sea, in mountainous areas, or in suburbs, and the affected area may range from a few kilometers to tens of kilometers. Therefore, in recent years, optimizing the three-dimensional flight trajectory of UAV to improve service quality has received increasing attention. In [13], Shi et al. proposed a drones-assisted wireless access network architecture that optimizes the 3-D deployment position of drones to maintain drone-to-base (D2B) station link quality while maximizing drone coverage. Many researchers consider communication coverage as an optimization objective, decomposing the three-dimensional positions of UAVs into horizontal and vertical dimensions and employing various algorithms to achieve comprehensive deployment of target nodes [14–16]. In the proposed post-disaster emergency wireless communication network, Amr et al. optimized the flight trajectory of UAV using a dual-cost-aware multi-armed bandit algorithm, achieving the maximum increase in the number of ground users served during the flight, despite the limited available energy for both UAV and ground users [17]. However, their approach dedicated each time slot to a single UE and focused solely on uplink communication quality, which did not maximize communication efficiency. Overall, we observe that these studies primarily focus on optimizing UAV trajectories in physical space while often neglecting the simultaneous optimization in the temporal and frequency domains, leading to suboptimal utilization of communication resources.

To address the aforementioned issues and achieve an optimal network state, the joint optimization of network resources has gradually deepened. In [18], a linear topology scenario is used as an example, and 1D UAV operation trajectories and wireless resource allocations under transmission schemes such as non-orthogonal multiple access (NOMA), FDMA, and time-division multiple access (TDMA) are jointly allocated to reveal the basic rate limit of UAV-enabled multiple access channel (MAC). In [19], the focus is on the spectrum. To avoid interference with primary users, the UAV employs a non-linear homotopy estimation based hidden Markov model (NLH-HMM) to predict the spectrum state duration at the next location and selects a spectrum with sufficient idle time. Moreover, transmission power is also considered a major optimization parameter [20]. Liu et al. investigated the optimal aerial positioning and resource allocation for UAV base stations in Internet of Things (IoT) networks with the aim of minimizing the total transmission power of IoT devices [21]. They employed the K-Means algorithm for clustering, the modified-Hungarian-based dynamic many–many matching (HD4M) algorithm for sub-channel allocation, and finally used the alternating iteration method to jointly optimize the IoT device transmitter power and UAV height. Similar to [21], Hattab et al. also performed network resource optimization. They proposed that the time division duplex transmission protocol provides shared spectrum access and improves the average allocation of resources between IoT devices and users (UEs) through random geometry protocols, but UEs will suffer additional interference from IoT devices. Therefore, they have optimized the nominal transmit power of IoT devices to maximize their energy efficiency while limiting interference to the UE [22]. DRL is also being increasingly applied in network resource optimization [23–25]. Cao et al. introduces a novel autofocusing heuristic trajectory planning algorithm based on reinforcement learning (AHTP-RL) algorithm to minimize the average delay of different time-sensitive nodes and UAV power consumption during the trajectory optimization process [23]. The study in [24] innovatively optimized the phase shifts of Reconfigurable Intelligent Surfaces (RIS) and the 3D trajectory

of UAV to enhance system energy efficiency and data transmission rates. We found that the literature does not consider the simultaneous optimization of both time and frequency domain resources. To improve this, we propose HybridComm, redefining the optimization goal to maximize the communication range of UAV while ensuring throughput. In addition, the data transmission priority in UAV-assisted communication scenarios has gradually become a focus of researchers. Yang et al. investigated a priority queuing model for heterogeneous traffic with UAV acting as relays. They analyzed the average end-to-end packet delay and buffer overflow probability for both mobile and static trajectories, and validated the accuracy of the analytical model through extensive simulations [26]. Gao et al. dynamically adjusted the priority of each data packet during node transmission based on a weighted combination of the initial priority value, delay factor, and the probability of maintaining link availability, to ensure the maximization of data transmission within the latency requirements [27]. Chen et al. studied power and sub-carrier allocation while setting communication priorities, only guaranteeing the minimum data rate for high-priority user rescue personnel's downlink communication [28]. To simultaneously ensure the upload requirements for rescue information from both rescuers and disaster victims, HybridComm achieves full-duplex communication and mandates that all nodes in the access system meet the minimum threshold.

In addition to the aforementioned solutions for UAV trajectory optimization in both the time and frequency domains, there are also performance optimizations aimed at reducing network latency through mobile edge computing [29] and enhancing network data throughput via opportunistic data transmission methods facilitated by UAV [30]. The distance between disaster areas and the backend rescue command center may exceed the coverage range of UAV airborne base stations, underscoring the necessity of researching emergency communications that utilize UAV as relays to extend communication range [31,32]. A robust relay solution should not only fulfill functional requirements but also meet network performance criteria [33–36].

Through the analysis and summary of the above related work, it has brought new thinking and inspiration to the related work of this study.

3. System model

As illustrated in Fig. 1, a suburban emergency communication network is deployed in response to an earthquake disaster, establishing crucial communication links between the disaster zone and the external environment. UAV is used as temporary base stations. Following the disaster, rescue operations covering an area of 2000 m × 1000 m are conducted by rescue teams, with a remote command center overseeing the entire operation. The remote command center, which orchestrates the overall rescue efforts, requires real-time access to information from within the disaster zone through the emergency communication network. It must maintain constant communication with on-site rescue teams, medical personnel, logistics units, and other essential staff. Due to the obstruction of roads within the disaster area, an emergency communication vehicle is stationed outside the zone to act as a gateway. As rescue operations commence, on-site teams equipped with individual communication devices establish connections with the aerial UAV, enabling them to receive instructions from the remote command center and relay critical on-site updates. Therefore, their communication priority mechanism designed to ensure their highest priority. The deployment of UAV is carefully planned to maximize communication coverage and ensure optimal network QoE. This setup underscores the pivotal role of UAV in facilitating robust and reliable communication during disaster response operations. The UAV's strategic positioning and dynamic network management ensure that all rescue activities are coordinated efficiently, thereby enhancing the effectiveness of post-disaster rescue and recovery efforts.

3.1. UAV trajectory

The flight cycle is divided into T equal time slots, each represented by a duration τ, where $\mathbb{T} = \{1, \ldots, t, \ldots, \mathrm{T}\}$. Given that τ is sufficiently small and equal across all time slots, the UAV's position remains relatively constant within each slot. In the emergency communication network depicted in Fig. 1, rescue equipment is denoted by RE, while disaster-affected personnel are denoted as UE, expressed as $\mathbb{I} = \{1, \ldots, i, \ldots, \mathrm{I}\}$.

The UAV's airborne height is h, so the UAV coordinates of time slot t are expressed as $c^t_{\mathrm{UAV}} = \left(x^t_{\mathrm{UAV}}, y^t_{\mathrm{UAV}}, \mathrm{h}\right), t \in \mathbb{T}$. The speed of the UAV is not fixed, which determines its position and range in the next time slot. The maximum flight speed of UAV is $v_{\max}$, then the maximum flight distance of each time slot t can be expressed as $l_{\max} = v_{\max}\tau$. Assuming the initial coordinates of the UAV is $c^0_{\mathrm{UAV}} = \left(x^0_{\mathrm{UAV}}, y^0_{\mathrm{UAV}}, \mathrm{h}\right)$, the abscissa of time slot t is denoted as $x^t_{\mathrm{UAV}} = x^0_{\mathrm{UAV}} + \sum_{t=1} l_t \cos\left(\phi^t_{\mathrm{UAV}}\right), \forall t \in \mathbb{T}$, and the ordinate as $y^t_{\mathrm{UAV}} = y^0_{\mathrm{UAV}} + \sum_{t=1} l_t \sin\left(\phi^t_{\mathrm{UAV}}\right), \forall t \in \mathbb{T}$. Where ϕ is the flight angle, with a range of values from $[0, 2\pi]$. Hence, the flight coordinate constraints are as follows:

$$x^{t+1}_{\mathrm{UAV}} \le x^t_{\mathrm{UAV}} + l_{\max}\ \cos\left(\phi^{t+1}_{\mathrm{UAV}}\right), \forall t \in \mathbb{T} \tag{1}$$

$$y^{t+1}_{\mathrm{UAV}} \le y^t_{\mathrm{UAV}} + l_{\max}\ \sin\left(\phi^{t+1}_{\mathrm{UAV}}\right), \forall t \in \mathbb{T} \tag{2}$$

3.2. Channel model

The emergency communication network shown in Fig. 1 primarily operates on an air-to-ground (A2G) channel. The suburban setting comprises a limited number of buildings and vegetation that may obstruct the communication link between UAV and ground users. Therefore, the channel model encompasses both line of sight (LoS) and non-line of sight (NLoS) conditions.

The probabilities of LoS and NLoS connections between UE_i and UAV at time slot t are expressed as follows:

$$P^t_{\mathrm{UE}_i,\mathrm{UAV},\mathrm{LoS}} = 1/\left(1 + \alpha \exp\left(-\beta\left(\theta^t_{\mathrm{UE}_i,\mathrm{UAV}} - \alpha\right)\right)\right), \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{3}$$

$$P^t_{\mathrm{UE}_i,\mathrm{UAV},\mathrm{NLoS}} = 1 - P^t_{\mathrm{UE}_i,\mathrm{UAV},\mathrm{LoS}}, \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{4}$$

where α and β are environmental constants, $\theta^t_{\mathrm{UE}_i,\mathrm{UAV}}$ is the elevation angle between UE_i and UAV at time slot t.

The LoS and NLoS path loss between UE_i and UAV at time slot t is:

$$L^t_{\mathrm{UE}_i,\mathrm{UAV},\mathrm{LoS}} = 20\log\left(4\pi f d^t_{\mathrm{UE}_i,\mathrm{UAV}}/c\right) + \eta_{\mathrm{LOS}}\ , \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{5}$$

$$L^t_{\mathrm{UE}_i,\mathrm{UAV},\mathrm{NLOS}} = 20\log\left(4\pi f d^t_{\mathrm{UE}_i,\mathrm{UAV}}/c\right) + \eta_{\mathrm{NLoS}}, \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{6}$$

where η_{LoS} is the additional loss in the LoS path, and η_{NLoS} is the additional loss in the NLoS path. f is the channel transmission frequency from UAV to UE_i, measured in Hz; c is the speed of light, measured in m/s; $d_{\mathrm{UE}_i,\mathrm{UAV}}$ is the distance between UAV and UE_i in time slot t, measured in m.

Therefore, the average path loss between UE_i and UAV in time slot t can be expressed as:

$$\begin{aligned} g^t_{\mathrm{UE}_i,\mathrm{UAV}} &= P^t_{\mathrm{UE}_i,\mathrm{UAV},\mathrm{LOS}} * L^t_{\mathrm{UE}_i,\mathrm{UAV},\mathrm{LoS}} + P^t_{\mathrm{UE}_i,\mathrm{UAV},\mathrm{NLOS}} * L^t_{\mathrm{UE}_i,\mathrm{UAV},\mathrm{NLOS}} \\ &= \left(\eta_{\mathrm{LOS}} - \eta_{\mathrm{NLaS}}\right) / \left(1 + \alpha \exp\left(-\beta\left(\theta^t_{\mathrm{UE}_i,\mathrm{UAV}} - \alpha\right)\right)\right) \\ &\quad + 20\log\left(4\pi f d^t_{\mathrm{UE}_i,\mathrm{UAV}}/c\right) + \eta_{\mathrm{NLos}}, \forall i \in \mathbb{I}, t \in \mathbb{T} \end{aligned} \tag{7}$$

In the same way, the average path loss between UAV and RE in time slot t can be express as:

$$\begin{aligned} g^t_{\mathrm{RE,UAV}} &= P^t_{\mathrm{RE,UAV,LOS}} * L^t_{\mathrm{RE,UAV,LoS}} + P^t_{\mathrm{RE,UAV,NLOS}} * L^t_{\mathrm{RE,UAV,NLOS}} \\ &= \left(\eta_{\mathrm{LOS}} - \eta_{\mathrm{NLaS}}\right) / \left(1 + \alpha \exp\left(-\beta\left(\theta^t_{\mathrm{RE,UAV}} - \alpha\right)\right)\right) \\ &\quad + 20\log\left(4\pi f d^t_{\mathrm{RE,UAV}}/c\right) + \eta_{\mathrm{NLos}}, \forall i \in \mathbb{I}, t \in \mathbb{T} \end{aligned} \tag{8}$$

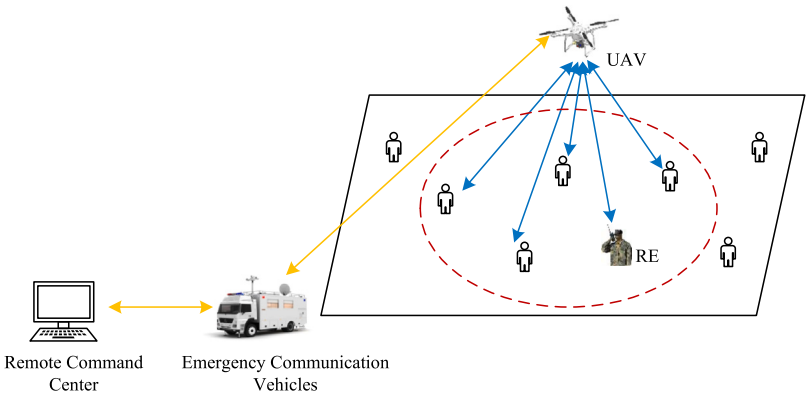


Fig. 1. Post-disaster emergency communication network system model.

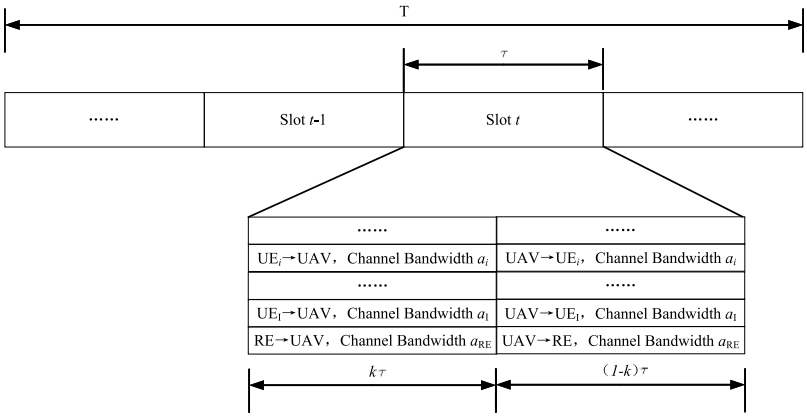


Fig. 2. HybridComm communication architecture.

3.3. Ground communication

According to the aforementioned description, ground communication encompasses two primary components: disaster victims and rescuers. During rescue operations, the efficient transmission of critical information heavily relies on rescuers, thus prioritizing the immediacy and reliability of communication over that of disaster victims. To address this, we have implemented a communication priority mechanism within our framework, ensuring that rescuers consistently receive higher communication priority. Essentially, this system mandates that the communication quality for rescuers remains consistently above a specified threshold. The threshold in this context refers to the minimum transmission rate for uplink and downlink communication.

We investigate a HybridComm communication architecture that addresses often simultaneously overlooked issues of uplink–downlink timeslot ratio and bandwidth allocation. To achieve duplex communication between UAV and both the UE and RE within time slot t, the uplink-to-downlink ratio k is dynamically adjusted according to the volume of uplink and downlink data in time slot $t-1$ using TDD technology. In time slot t, the UAV, UE and RE access via FDMA technology. The total system bandwidth B is divided into I+1 non-overlapping sub-bands, which are allocated to I disaster-affected personnel and a rescue equipment in time slot t. The non-overlapping sub-bands are expressed as $\mathbb{W} = \{a_1\mathrm{w}, \ldots, a_i\mathrm{w}, \ldots, a_1\mathrm{w}, a_{\mathrm{RE}}\mathrm{w}\}$, as shown in Fig. 2. Since both rescue personnel and disaster victims operate at the same hierarchical level within the system, their communication mathematical models are identical.

According to Fig. 2, the uplink data transmission rate during slot t can be determined as follows:

$$r^t_{\mathrm{UE}_i,\mathrm{UAV}} = k\tau a_i \mathrm{w} \log\left(1 + p_{\mathrm{UE}_i,\mathrm{UAV}} g^t_{\mathrm{UE}_i,\mathrm{UAV}} / \sigma_0^2\right), \forall i \in \mathbb{I}, t \in \mathbb{T}, a_i\mathrm{w} \in \mathbb{W} \tag{9}$$

Downlink data transmission rate during slot t:

$$r^t_{\mathrm{UAV},\mathrm{UE}_i} = (1-k)\tau a_i \mathrm{w} \log\left(1 + p_{\mathrm{UAV},\mathrm{UE}_i} g^t_{\mathrm{UAV},\mathrm{UE}_i} / \sigma_0^2\right), \forall i \in \mathbb{I}, t \in \mathbb{T}, a_i\mathrm{w} \in \mathbb{W} \tag{10}$$

where $r^t_{\mathrm{UE}_i,\mathrm{UAV}}$ is the uplink data transmission rate between UE_i and UAV during time slot t, $r^t_{\mathrm{UAV},\mathrm{UE}_i}$ is the downlink data transmission rate between UAV and UE_i during time slot t; k is the uplink–downlink ratio, with $k\tau$ indicating the transmission duration of the uplink phase and $(1-k)\tau$ indicating the transmission duration of the downlink phase; w is the total system frequency bandwidth, and the allocation ratio of sub-bandwidth for the ith UE is denoted as a_i; $p_{\mathrm{UE}_i,\mathrm{UAV}}$ is the uplink transmission power between UE_i and the UAV, $p_{\mathrm{UAV},\mathrm{UE}_i}$ is the downlink transmission power from the UAV to UE_i, and σ_0^2 is the noise power.

In the same way, the uplink data transmission rate between UAV and RE in time slot t can be express as:

$$r^t_{\mathrm{RE},\mathrm{UAV}} = k\tau a_{\mathrm{RE}} \mathrm{w} \log\left(1 + p_{\mathrm{RE},\mathrm{UAV}} g^t_{\mathrm{RE},\mathrm{UAV}} / \sigma_0^2\right), \forall i \in \mathbb{I}, t \in \mathbb{T}, a_{\mathrm{RE}}\mathrm{w} \in \mathbb{W} \tag{11}$$

Downlink data transmission rate during slot t:

$$r^t_{\mathrm{UAV,RE}} = (1-k)\tau a_{\mathrm{RE}} \mathrm{w} \log\left(1 + p_{\mathrm{UAV,RE}} g^t_{\mathrm{UAV,RE}} / \sigma_0^2\right), \forall i \in \mathbb{I}, t \in \mathbb{T}, a_{\mathrm{RE}}\mathrm{w} \in \mathbb{W} \tag{12}$$

Due to factors such as distance and environmental conditions, UAV experiences varying network throughputs when establishing communication links with different UE, directly impacting the quality of network communication. In this study, communication decision is denoted by $e^t_{\mathrm{UE}_i}, \forall i \in \mathbb{I}, t \in \mathbb{T}$. If $e^t_{\mathrm{UE}_i} = 1, \forall i \in \mathbb{I}, t \in \mathbb{T}$, the network throughput of the UE_i communication link exceeds or equals a certain threshold, the communication link is established; if $e^t_{\mathrm{UE}_i} = 0, \forall i \in \mathbb{I}, t \in \mathbb{T}$, the network throughput of the UE_i communication link falls below a certain threshold, the communication link is not established. The communication decision can be expressed as

$$e^t_{\mathrm{UE}_i} = \begin{cases} 0, & r^t_{\mathrm{UE}_i,\mathrm{UAV}} < r_{\mathrm{UE}_i,u,\min} \ \textit{or}\ r^t_{\mathrm{UAV,UE}_i} < r_{\mathrm{UE}_i,d,\ \min} \\ 1, & r^t_{\mathrm{UE}_i,\mathrm{UAV}} \geq r_{\mathrm{UE}_i,u,\min} \ \textit{and}\ r^t_{\mathrm{UAV,UE}_i} \geq r_{\mathrm{UE}_i,d,\ \min} \end{cases}, \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{13}$$

where $r_{\mathrm{UE}_i,u,\min}, \forall i \in \mathbb{I}, t \in \mathbb{T}$ is the minimum threshold value for the uplink rate of the UE, and $r_{\mathrm{UE}_i,d,\min}, \forall i \in \mathbb{I}, t \in \mathbb{T}$ is the minimum threshold value for the downlink rate of the UE.

4. Proposed algorithm

4.1. Problem formulation

In this study, our primary focus lies in optimizing the total number of UE covered, a critical metric contingent upon the data transmission rate between UAV and UE. Achieving maximum UE coverage necessitates a holistic approach, wherein we concurrently optimize several key parameters. These include the trajectory of UAV motion, the uplink and downlink ratio within TDD technology, and the sub-bandwidth allocation ratio within FDMA technology. $\mathbb{C} = \{c^t\}, \forall t \in \mathbb{T}$ represents the UAV motion trajectory, $\mathbb{K} = \{k^t\}, \forall t \in \mathbb{T}$ represents the uplink and downlink ratio of TDD technology, $\mathbb{A} = \{a_1, \ldots, a_i, \ldots, a_{\mathrm{I}}, a_{\mathrm{RE}}\}$ represents the allocation ratio of each sub-bandwidth within FDMA technology.

$$\mathrm{OP}: \max_{\mathbb{A},\mathbb{C},\mathbb{K}} \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \sum_{i=1}^{\mathrm{I}} \left(e^t_{\mathrm{UE}_i}\right), \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{14}$$

$$\text{s.t.} \quad e^t_{\mathrm{UE}_i} \in \{0,1\}, \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{14a}$$

$$\sum_{i=1} a^t_i + a^t_{RE} = 1, \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{14b}$$

$$\begin{cases} r^t_{\mathrm{RE,UAV}} \geq r_{\mathrm{RE},u,\min} \\ r^t_{\mathrm{UAV,RE}} \geq r_{\mathrm{RE},d,\ \min} \end{cases}, \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{14c}$$

$$0 \leq \phi_{\mathrm{UAV}} \leq 2\pi \tag{14d}$$

$$0 \leq l \leq l_{\max} \tag{14e}$$

$$\begin{cases} 0 \leq x^{t+1}_{\mathrm{UAV}} \leq x^t_{\mathrm{UAV}} + l_{\max} \cos\phi_{\mathrm{UAV}} \\ 0 \leq y^{t+1}_{\mathrm{UAV}} \leq y^t_{\mathrm{UAV}} + l_{\max} \sin\phi_{\mathrm{UAV}} \end{cases}, \forall t \in \mathbb{T} \tag{14f}$$

In the preceding description of the problem formulation, (14a) indicates whether UE_i is covered, while (14b) ensures that the sum of FDMA sub-bandwidth allocation ratios equals 1. In this system, we have established a communication priority mechanism, wherein (14c) ensures that rescue personnel in disaster-stricken areas are given top priority in communication, thereby guaranteeing a minimum data transmission rate. (14d)~(14f) delineate the constraints governing the trajectory of the UAV. Specifically, (14d) imposes limitations on the UAV's flight angle, (14e) caps the maximum flight distance of the UAV within period, and (14f) restricts the abscissa and ordinate coordinates of the UAV's position in the next slot. This problem is a complex non-convex optimization issue with nonlinear constraints, which increases the difficulty of finding solutions and makes traditional methods inadequate for handling such problems. DRL excels in addressing these complex optimization challenges by learning and exploring different decision paths to gradually optimize solutions. It is particularly well-suited for such problems due to its ability to effectively manage multiple local optima and intricate constraints.

4.2. PPO algorithm

In DRL, the agent optimizes its strategy through continuous trials and feedback to maximize expected rewards using policy gradients. This process is computationally expensive due to the need for constant interaction with the environment to gather data for updates. PPO is derived from Trust Region Policy Optimization (TRPO) and enhances training stability by limiting policy updates' disparity, offering a more simplified algorithm. Empirical studies show PPO's learning rate is comparable to that of TRPO, making it a preferred choice in the DRL community. Among its variants, PPO-Clip, known for its efficiency and performance [37], is chosen in this study to meet our research objectives.

The fundamental concept underpinning the PPO-Clip algorithm revolves around maintaining policy stability through the constraint of policy update magnitudes. To realize this objective, the PPO-Clip algorithm employs a near-end ratio clipping loss mechanism, defined as follows:

$$L^{clip}(\theta) = \mathbb{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, \mathrm{clip}\left(r_t(\theta), 1-\varepsilon, 1+\varepsilon\right)\hat{A}_t\right)\right] \tag{15}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\mathrm{old}}}(a_t|s_t)}$ is the magnitude of policy update, depicting the ratio between the likelihood of the current policy selecting action a_t within state s_t and that of the previous policy undertaking the action a_t within the state s_t. A larger $r_t(\theta)$ implies a heightened likelihood of the current policy selecting action a_t within state s_t, thereby resulting in a more substantial update magnitude for the previous policy. The $r_t(\theta)$ is constrained within the interval [0,1] by $\mathrm{clip}\left(r_t(\theta), 1-\varepsilon, 1+\varepsilon\right)$, thereby ensuring that the discrepancy between the two strategies remains within acceptable bounds. Fig. 3 shows the $\mathrm{clip}\left(r_t(\theta), 1-\varepsilon, 1+\varepsilon\right)$.

As shown in Fig. 3(a), $r_t(\theta)$ is represented by the green line, and $\mathrm{clip}\left(r_t(\theta), 1-\varepsilon, 1+\varepsilon\right)$ by the blue line. We aim to find the minimum value between the green and blue lines. Assuming the coefficient A is greater than 0, the resulting minimum value is indicated by the red line. Conversely, as depicted in Fig. 3(b), if A is less than 0, the minimum value again results in the red line.

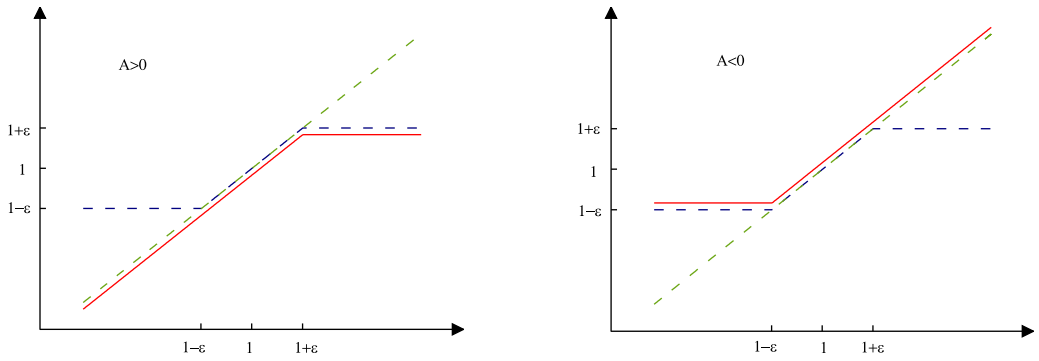
$\hat{A}_t$ is the advantage function, which is utilized to calculate the clipping magnitude in the near-end ratio clipping loss. The advantage function aids the agent in evaluating the advantage of a particular action relative to others, thereby guiding policy updates. It is defined as follows:

$$\hat{A}_t = Q_{\pi_{\theta_{\mathrm{old}}}}\left(s_t, a_t\right) - V_{\pi_{\theta_{\mathrm{old}}}}\left(s_t\right) \tag{16}$$

where $Q_{\pi_{\theta_{\mathrm{old}}}}\left(s_t, a_t\right)$ is the value of taking action a_t in state s_t, while $V_{\pi_{\theta_{\mathrm{old}}}}\left(s_t\right)$ is the average value in state s_t. A higher value of the advantage function indicates that the current state–action pair is more advantageous, thus warranting a larger reward.

In the PPO algorithm, neural networks estimate the state-value function. Using this estimate and the current reward, we compute the advantage function. Accurate estimation of the advantage function improves action evaluation, guiding strategy updates and enhancing the algorithm's efficiency and performance, as detailed in Algorithm 1.

The code framework appears straightforward. Lines 3–5 handle the initialization for each epoch. Line 6 updates the parameters of the



(a) A>0 (b) A<0

Fig. 3. Clipping function for policy probability ratio.

Algorithm 1 PPO-Clip

1: **Input:** initial policy parameters θ_0, initial value function parameters ϕ_0.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Collect set of trajectories $\mathcal{D}_{\hat{k}} = \{\tau_i\}$ by running policy $\pi_k = \pi(\theta_k)$ in the environment.
4: Compute reward-to-go $\hat{R}_t$.
5: Compute advantage estimates $\hat{A}_t$:

$$\hat{A}_t = Q_{\pi_{\theta_{\text{old}}}}(s_t, a_t) - V_{\pi_{\theta_{\text{old}}}}(s_t)$$

6: Update the policy by maximizing the PPO-Clip objective:

$$\theta_{k+1} = \arg\max_{\theta} \left(1/\left(|D_k|T\right) \sum_{\tau \in D} \sum_{t=0}^{T} \min\left(\pi_\theta(a_t \mid s_t)/\pi_{\theta_i}(a_t \mid s_t) A^{\pi_{\theta_k}}(s_t, a_t), g\left(\epsilon, A^{\pi_{\theta_k}}(s_t, a_t)\right)\right)\right)$$

typically via stochastic gradient ascent with Adam.
7: Fit value function by regression on mean-squared error:

$$\phi_{k+1} = \arg\min_{\phi} \left(1/\left(|D_k|T\right) \sum_{\tau \in D_k} \sum_{t=0}^{T} \left(V_\phi(s_t) - \hat{R}_t\right)^2 \right)$$

typically via some gradient descent algorithm.
8: **end for**
9: **Output:** Optimized policy parameters θ, Optimized value function parameters ϕ

Actor network using the Adam optimizer, while Line 7 updates the parameters of the Critic network.

4.3. MDP build

In this study, the UAV acts as an intelligent agent that collects state parameters of the network environment in real-time. We have discovered that both the movement of the UAV and the wireless channel fading exhibit Markov properties. Consequently, we apply a DRL algorithm within the framework of a MDP. A key feature of MDP in DRL is that the immediate reward and state transition probabilities depend only on the current system state and chosen actions, without being influenced by past behaviors.

A complete MDP consists of four essential components, represented as the tuple (S, A, R, P). Here, S represents the state set, A represents all feasible actions within each state, and P represents the probability of transitioning from state S to state S' upon taking action a. Given the unknown nature of state transition probabilities in this study, our MDP aligns with the model-free paradigm, which is expressed as follows:

State space: Within the context of our research scenario, the state space encapsulates crucial variables the uplink/downlink data rate of UE and RE, alongside the channel status during time slot t. Thus, the state space, denoted as S^t, is comprehensively expressed as follows:

$$S^t = \left\{ \left\{ \left(r^t_{\text{UE}_i,\text{UAV}}\right), \left(r^t_{\text{UAV,UE}_i}\right) \right\}, \left\{ \left(r^t_{\text{RE,UAV}}\right), \left(r^t_{\text{UAV,RE}}\right) \right\}, \left\{ \left(g^t_{\text{UE}_i,\text{UAV}}\right), \left(g^t_{\text{RE,UAV}}\right) \right\} \right\}, \quad \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{17}$$

where $r^t_{\text{UE}_i,\text{UAV}}, \forall t \in \mathbb{T}$ and $r^t_{\text{UAV,UE}_i}, \forall t \in \mathbb{T}$ represent the uplink and downlink data rates of the link between the UE_i and the UAV, respectively. $r^t_{\text{RE,UAV}}, \forall t \in \mathbb{T}$ and $r^t_{\text{UAV,RE}}, \forall t \in \mathbb{T}$ denote the uplink and downlink data rates of the link between the RE and the UAV, respectively. $g^t_{\text{UE}_i,\text{UAV}}, \forall t \in \mathbb{T}$ is the average path loss between the UE_i and the UAV, while $g^t_{\text{RE,UAV}}, \forall t \in \mathbb{T}$ is the average path loss between the RE and the UAV.

Action space: In this study, we employ a HybridComm architecture. The action space enompasses various parameters, including the UAV flight angle and distance, as well as the uplink and downlink ratio, and sub-bandwidth ratio during slot t. Thus, the action space A^t can be expressed as:

$$A^t = \left\{ \phi^t_{\text{UAV}}, l^t_{\text{UAV}}, \left\{ a^t_i, a^t_{RE} \right\}, k^t \right\}, \forall i \in \mathbb{I}, t \in \mathbb{T} \tag{18}$$

where ϕ^t_{UAV} is the movement angle of the UAV at time slot t, l^t_{UAV} is the movement distance of the UAV at time slot t, a^t_i and a^t_{RE} represent the allocation rates for each sub-band, and k denotes the proportion of uplink and downlink timeslots.

Reward function: The reward serves as a metric for assessing the efficacy of actions taken. In each period, the flying position of the UAV, the ratio of uplink and downlink, and the ratio of sub-bandwidth are all random. Under the premise of ensuring priority mechanisms, the most direct unshaped reward function is the number of ground nodes covered by UAV communication, while ensuring the priority of communication for rescue personnel. This can be expressed as:

$$reward\_unshaped = \begin{cases} -\left|N_t - N_{t-1}\right|, & \text{if } e^t_{\text{RE}} = 0 \\ N_t, & other \end{cases} \tag{19}$$

However, the unshaped reward often results in low training efficiency for the agent. To address this issue, we developed a shaped reward function as an enhancement of the unshaped reward. Specifically, we grant a reward when the coverage rate increases in consecutive time slots and impose a penalty when it decreases. If the coverage rate remains unchanged, the reward provided is smaller than that given for an increase. Additionally, to ensure continuous communication for ground rescue personnel, we apply the same penalty when there is no coverage.

$$reward\_shaped = \begin{cases} -\left|N_t - N_{t-1}\right|, & \text{if } e^t_{\text{RE}} = 0 \\ N_t - N_{t-1}, & \text{if } N_t - N_{t-1} < 0 \text{ and } e^t_{\text{RE}} = 1 \\ N_t, & \text{if } N_t - N_{t-1} = 0 \text{ and } e^t_{\text{RE}} = 1 \\ 10^* N_t, & \text{if } N_t - N_{t-1} > 0 \text{ and } e^t_{\text{RE}} = 1 \end{cases} \tag{20}$$

Table 1
Simulation parameters.

Parameter	Value
Target area range	2000 m × 1000 m
Number of ground nodes	25
Number of rescue nodes	1
Number of UAV	1
Height of UAV (h)	80 m
Speed of UAV (v_{max})	10 m/s
Transmit power ($p_{UE_i,UAV}/p_{UAV,UE_i}$)	1 W
Bandwidth (w)	5 MHz
Length of time slot (τ)	1s
Noise (σ_0^2)	1e–14w
Carrier frequency (f)	2.4 GHz
$(a, b, \eta_{LOS}, \eta_{NLOS})$	(4.88,0.43,0.1,21)

where N_t is the number of communication coverage for time slot t, N_{t-1} is the number of communication coverage for time slot $t-1$, e^t_{RE} indicates whether the rescuers' communication devices are covered: $e^t_{RE} = 1$ means covered, $e^t_{RE} = 0$ means not covered.

Overall, the UAV collects state space parameters through sensors or communication modules, which are then fed into the policy network. The network processes these inputs through multiple neural layers to produce an action probability distribution. Acting as an intelligent agent, the UAV selects and executes the optimal action based on this distribution, transitioning to a new state and receiving a corresponding reward. The DRL algorithm utilizes state transitions and reward information to optimize the policy network. Through continuous iterations, it gradually enhances the UAV's decision-making quality, thereby optimizing communication efficiency and stability.

5. Performance evaluation

In this section, we rigorously evaluate the performance of the proposed post-disaster emergency communication network system based on the PPO algorithm through comprehensive simulation analyses. The focus of our simulations includes the initialization positions of UAV, the action space, and the minimum communication rate thresholds. The simulations were executed on the Windows 10 operating system with i9 CPU and 32-GB RAM, utilizing Python 3.7 and PyTorch 1.5.0. The actor and critic networks of the PPO algorithm each have two hidden layers, with the first hidden layer containing 256 neurons and the second layer containing 128 neurons.

The simulation environment was defined as a 2000 m × 1000 m suburban area, incorporating 25 randomly selected disaster-affected ground network nodes and one ground rescue network node. The detailed parameters for the simulation are presented in Table 1.

To illustrate the effectiveness of the PPO algorithm, we compare its simulation results with those of the Soft Actor-Critic (SAC) algorithm. The analysis focuses on two main aspects: the number of UAV communication coverage nodes and high-priority performance for disaster relief personnel. In addition to adjusting algorithm architectures and hyperparameters, we also examine the impact of initial UAV positions: the upper right, center, and centroid of the target area. Notably, the centroid position is determined using the k-means algorithm. This approach ensures that the centroid is strategically calculated, thus allowing a more rigorous evaluation of the algorithm's performance across different starting conditions. The results are shown in Figs. 4 and 5. Fig. 4 shows that, except for the SAC algorithm with the UAV initially positioned at the upper right, the number of communication coverage nodes generally converges between 550 and 620, with the PPO algorithm starting at the centroid performing best. Meanwhile, the SAC algorithm with the UAV initially at the upper right shows the best convergence, reaching approximately 760. However, as shown in Fig. 5, the number of rescue nodes covered per episode by the SAC algorithm is significantly lower than that of the PPO algorithm. Across

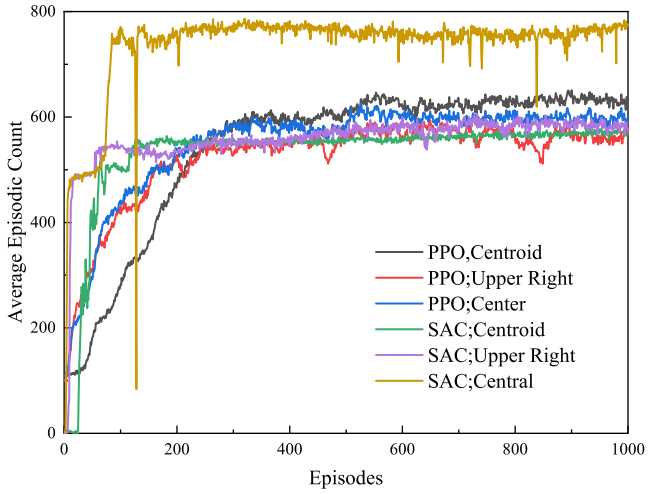


Fig. 4. Communication coverage comparison of UAV at different initial positions.

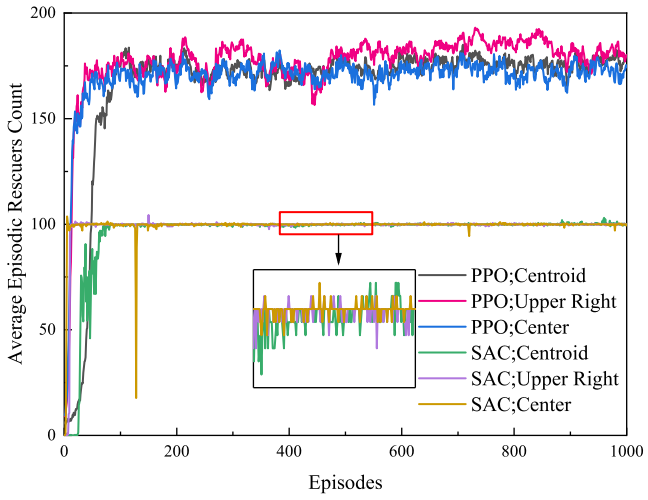


Fig. 5. Rescuer node coverage comparison of UAV at different initial positions.

the three initial positions, the SAC algorithm covers about 100 nodes per episode, while the PPO algorithm covers approximately 175 nodes. From the perspective of convergence, we found that both algorithms exhibit an upward trend during the training process and gradually converge, indicating continuous improvement in the strategies. After convergence, they demonstrate a degree of stability without significant further enhancement. However, in Figs. 4 and 5, some curves of the SAC algorithm exhibit a sharp decline after convergence, indicating significant volatility. Since we ultimately calculate the average total number of nodes covered per episode for each batch, the convergence value is in the hundreds, significantly exceeding the 25 distressed nodes and the single rescue node. Comprehensive analysis indicates that the PPO algorithm is superior in terms of convergence and optimization effects. To maximize the final algorithm performance, our subsequent simulation analysis will use the centroid of the target area as the initial UAV position.

A significant aspect of this study involves the shaping of the reward function to achieve optimal solutions. To demonstrate the extent of the optimization brought by the reshaped reward function, we conducted simulations comparing it with the unshaped reward function. The comparison results are shown in Fig. 6. We found that when the threshold is 1.5 Mb/s, the number of nodes covered per episode with the reshaped reward function is approximately 820, which is about

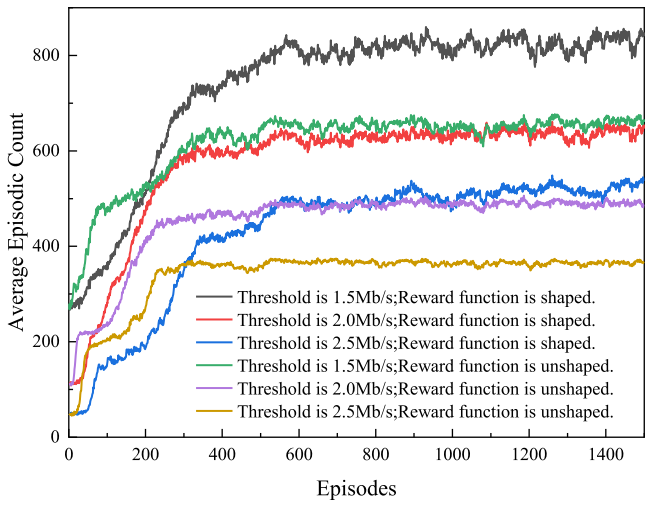


Fig. 6. Communication coverage comparison of different reward function.

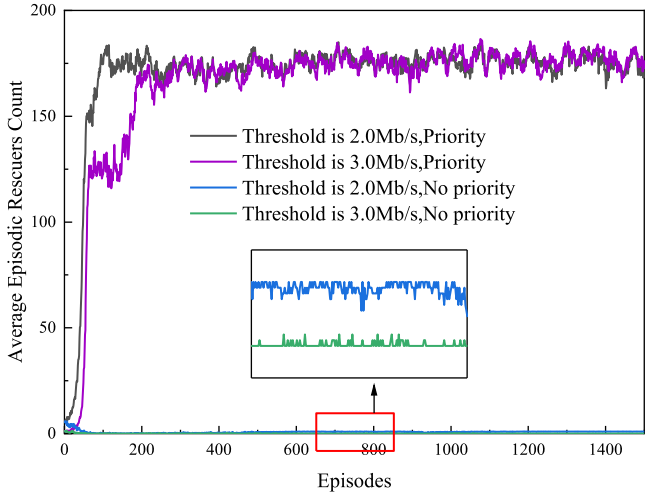


Fig. 7. Rescuer node coverage comparison of priority.

170 more than that with the unshaped reward function. Similarly, the reshaped reward function exhibits advantages at other thresholds as well: at a threshold of 2 Mb/s, it covers approximately 150 additional nodes; at 2.5 Mb/s, it covers about 160 more nodes. We found that all curves, whether shaped or unshaped, demonstrate a certain degree of stability after convergence. The slight fluctuations are a typical characteristic of DRL algorithms, indicating good convergence properties. In designing the reward function, we adopted the principle of providing significant rewards for increased coverage and penalties for decreased coverage, continuously guiding the agent toward the optimal value. Naturally, this approach outperforms the reward function that solely uses coverage as the reward value.

In the scenario designed for this study, the most crucial ground communication equipment is that used by rescue personnel. Therefore, a communication priority mechanism was established to ensure the real-time connectivity of rescue personnel's communication devices. To demonstrate the effectiveness of this mechanism, we conducted simulations comparing the system with and without priority settings, as shown in Fig. 7. We found that, the effectiveness of the communication priority mechanism is not significantly influenced by the threshold value. Regardless of whether the threshold was set at 2 Mb/s or 3 Mb/s, the number of rescue personnel devices covered per episode in the system without priority settings was almost zero. In contrast, with the priority mechanism in place, the coverage per episode was around 175 devices, ensuring high-quality communication for rescue personnel and various organizations.

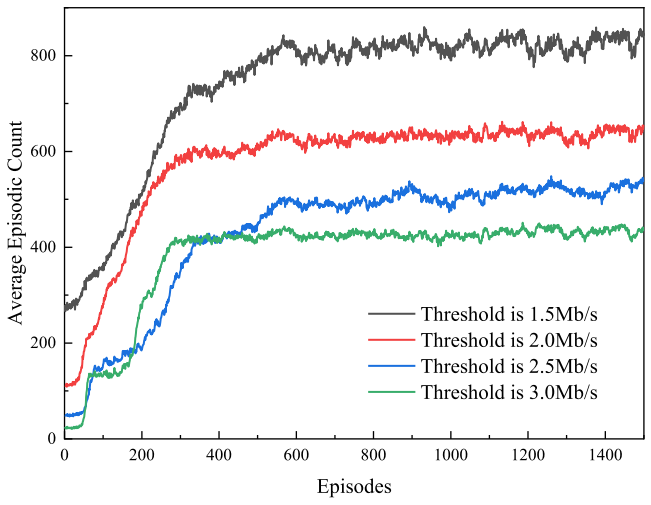


Fig. 8. Communication coverage comparison of different threshold.

To verify the effectiveness of the algorithm in the proposed scenario and the HybridComm framework, we conducted simulations with different thresholds, using the centroid of the target area as the initial position of the UAV. The verification results are presented in Figs. 8–11. Fig. 8 illustrates the average number of communication coverages per episode in each batch. We observed that as the threshold decreases, the communication coverage area gradually increases. When the threshold is set to the lowest value of 1.5 Mb/s, the number of ground nodes covered converges to about 820 per episode, achieving the highest coverage rate. Conversely, when the threshold increases to a maximum of 3 Mb/s, the coverage area converges to about 420 ground nodes per episode. Therefore, a lower threshold allows for greater communication coverage, which explains why the initial value of the curve with a threshold of 1.5 Mb/s in Fig. 8 is the highest. To further confirm the communication quality of the proposed scheme, we plotted the number of communication coverages per episode in Fig. 9, and the simulation results are consistent with those in Fig. 8. Figs. 10 and 11 show the uplink and downlink transmission rates of UAV-covered communication nodes, respectively. We observe that as the threshold increases, the communication rate also increases. This is because each node stops increasing once it meets the threshold requirement and reallocates communication resources to other nodes to achieve the optimization goal of maximizing the number of UAV communication coverage nodes.

To evaluate the performance of our algorithm from various perspectives, we conducted simulation analyses on UAV scenarios with different flight altitudes and transmission powers. The results are shown in Figs. 12 and 13. Fig. 12 illustrates that as the UAV flight altitude increases, the number of ground nodes covered decreases from approximately 640 to 430. Fig. 13 shows that as the transmission power increases, the number of covered nodes increases from 560 to about 700. This phenomenon can be attributed to the fact that greater transmission distances result in higher signal attenuation and lower signal-to-noise ratios (SNR). In contrast, higher transmission power leads to higher SNR. Since the SNR is logarithmically related to channel capacity, the simulation results align well with our expectations.

6. Conclusion

In this study, we propose a UAV-based emergency communication system to address the communication challenges following a disaster.

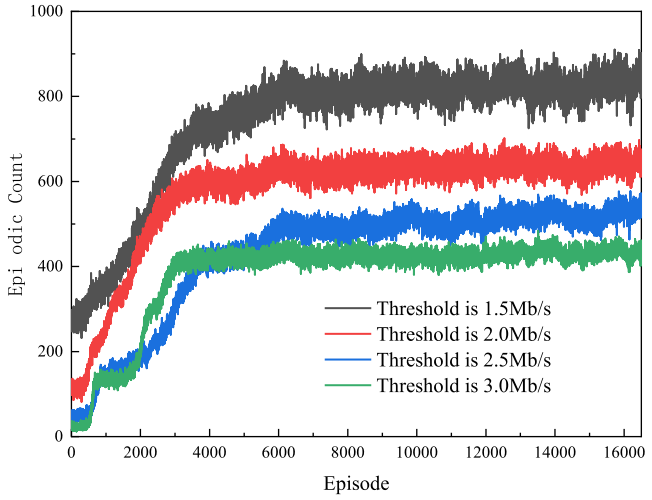


Fig. 9. Communication coverage comparison of different threshold in a single episode.

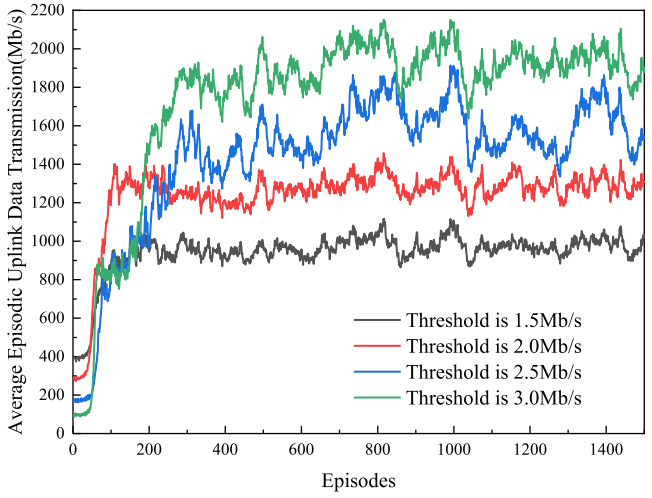


Fig. 10. Average episodic uplink data transmission.

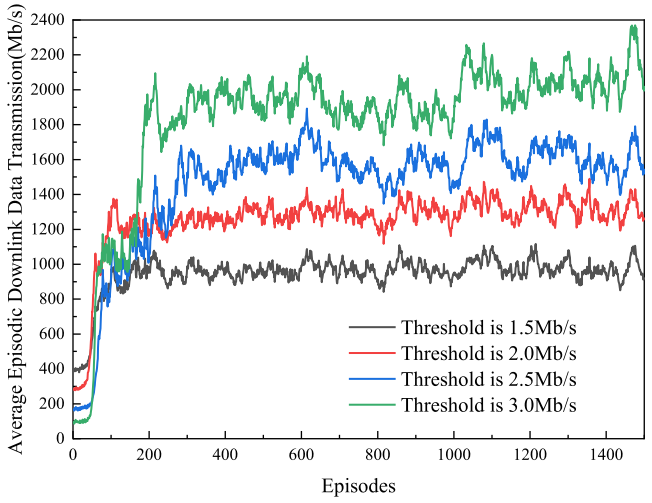


Fig. 11. Average episodic downlink data transmission.

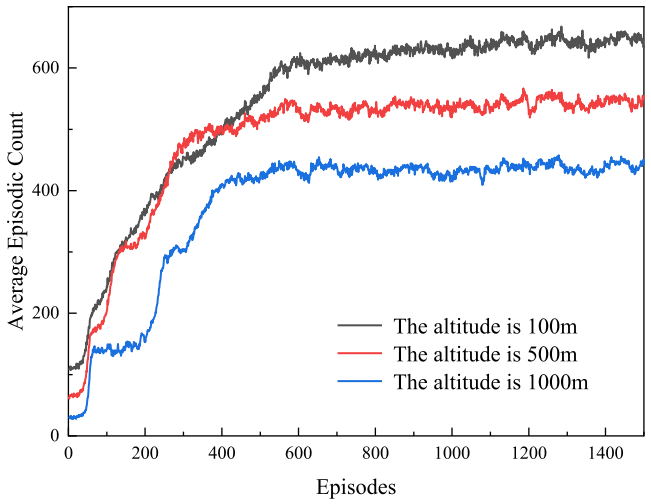


Fig. 12. Communication coverage comparison of UAV at different altitude.

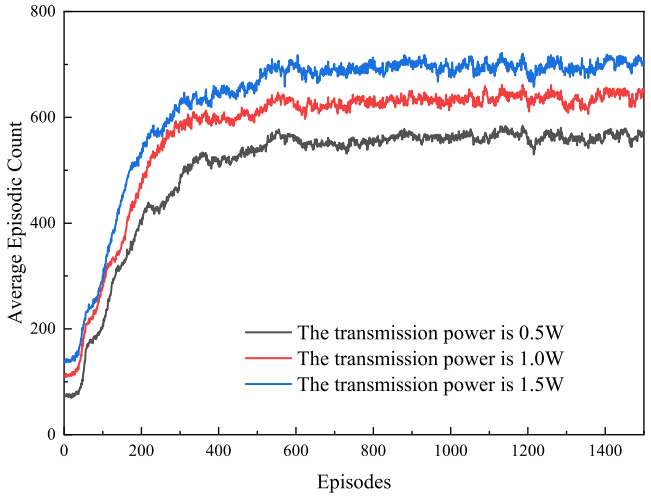


Fig. 13. Communication coverage comparison of different transmission power.

The ground network nodes establish communication with emergency rescue vehicles near the disaster area via the system's aerial UAV, facilitating information transmission between the back-end rescue center and the disaster site. Within this emergency communication system, we introduce the HybridComm framework, which addresses the often simultaneously overlooked issues of uplink–downlink timeslot ratio and bandwidth allocation, thereby enabling duplex communication and multi-node network access. We designed the rescue communication node priority, defined the communication link establishment rules, and ultimately optimized UAV communication coverage using a DRL algorithm.

Through simulations involving various starting positions, we observed that the centroid position determined by the k-means algorithm exhibited superior performance. Consequently, we based our subsequent simulations on this centroid position. Additionally, to demonstrate the significance of the reward function shaping, we compared results and found a substantial increase in coverage, encompassing

hundreds of additional nodes. Finally, our simulations reveal that a system incorporating a priority mechanism can achieve coverage of up to 175 rescue nodes per episode.

CRediT authorship contribution statement

Le Zhao: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Conceptualization. **Xiongchao Liu:** Writing – review & editing, Methodology, Conceptualization. **Tao Shang:** Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] A. Bouguettaya, H. Zarzour, A.M. Taberkit, A. Kechida, A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms, Signal Process. 190 (2022) 108309.

[2] M. Zhang, S. Li, F. Yu, X. Tian, Image fusion employing adaptive spectral-spatial gradient sparse regularization in UAV remote sensing, Signal Process. 170 (2020) 107434.

[3] Z. Fan, H. Sun, L. Wang, Research of the classification model based on dominance rough set approach for China emergency communication, Math. Probl. Eng. (2015) 2015428218.1-428218.8.

[4] W. Liang, Y. Zeng, W. Zhu, Z. Shi, X. Li, Y. Liang, Research on Beidou new generation E mergency group communication decision-making mechanism, in: 2022 IEEE 10th International Conference on Informations, Communication and Networks, ICICN, IEEE, 2022, pp. 15–19.

[5] A. Meng, X. Gao, Y. Zhao, Z. Yang, Three-dimensional trajectory optimization for energy-constrained UAV-enabled IoT system in probabilistic LoS channel, IEEE Internet Things J. 9 (2) (2021) 1109–1121.

[6] B. Zhao, X. Zhao, Deep reinforcement learning resource allocation in wireless sensor networks with energy harvesting and relay, IEEE Internet Things J. 9 (3) (2021) 2330–2345.

[7] J. Zhang, Y. Guo, L. Zheng, Yang. Qiming, G. Shi, Y. Wu, Real-time UAV path planning based on LSTM network, J. Syst. Eng. Electron. (2024).

[8] Z. Xu, Application research of tethered UAV platform in marine emergency communication network, J. Web Eng. 20 (2) (2021) 491–511.

[9] A. Ranjan, B. Panigrahi, H.K. Rath, P. Misra, A. Simha, H.B. Sahu, A study on pathloss model for UAV based urban disaster and emergency communication systems, in: 2018 Twenty Fourth National Conference on Communications, NCC, IEEE, 2018, pp. 1–6.

[10] C. Zhang, M. Dong, K. Ota, Heterogeneous mobile networking for lightweight UAV assisted emergency communication, IEEE Trans. Green Commun. Netw. 5 (3) (2021) 1345–1356.

[11] J. Sun, Z. Sheng, A.A. Nasir, Z. Huang, H. Yu, Y. Fang, Energy efficiency maximization for WPT-enabled UAV-assisted emergency communication with user mobility, Phys. Commun. 61 (2023) 102200.

[12] Z. Yao, W. Cheng, W. Zhang, H. Zhang, Resource allocation for 5G-UAV-based emergency wireless communications, IEEE J. Sel. Areas Commun. 39 (11) (2021) 3395–3410.

[13] W. Shi, J. Li, W. Xu, H. Zhou, N. Zhang, S. Zhang, X. Shen, Multiple drone-cell deployment analyses and optimization in drone assisted radio access networks, IEEE Access 6 (2018) 12518–12529.

[14] X. Liu, X. Wang, M. Huang, J. Jie, N. Bartolini, Q. Li, D. Zhao, Deployment of UAV-BSs for on-demand full communication coverage, Ad Hoc Netw. 140 (2023) 103047.

[15] R. Gao, X. Wang, Rapid deployment method for multi-scene UAV base stations for disaster emergency communications, Appl. Sci. 13 (19) (2023) 10723.

[16] Z. Qian, Y. Jia, C. Li, L. Liu, 3D deployment of UAV-BSs for effective communication coverage, IEEE Internet Things J. (2024).

[17] A. Amrallah, E.M. Mohamed, G.K. Tran, K. Sakaguchi, UAV trajectory optimization in a post-disaster area using dual energy-aware bandits, Sensors 23 (3) (2023) 1402.

[18] L. Peiming, X. Jie, Fundamental rate limits of UAV-enabled multiple access channel with trajectory optimization, IEEE Trans. Wireless Commun. 19 (1) (2020) 458–474.

[19] S. Luo, T. Zhou, Y. Xiao, R. Lin, Y. Yan, Predicting spectrum status duration using non-linear homotopy estimation based HMM for UAV communications, Signal Process. 212 (2023) 109131.

[20] G. Park, K. Lee, Optimization of the trajectory, transmit power, and power splitting ratio for maximizing the available energy of a UAV-aided SWIPT system, Sensors 22 (23) (2022) 9081.

[21] Y. Liu, K. Liu, J. Han, L. Zhu, Z. Xiao, X. Xia, Resource allocation and 3D placement for UAV-enabled energy-efficient IoT communications, IEEE Internet Things J. PP (99) (2020) 1-1.

[22] G. Hattab, D. Cabric, Energy-efficient massive IoT shared spectrum access over UAV-enabled cellular networks, IEEE Trans. Commun. PP (99) (2020) 1-1.

[23] H. Cao, W. Zhu, Z. Chen, Z. Sun, D.O. Wu, Energy-delay tradeoff for dynamic trajectory planning in priority-oriented UAV-AidedIoT networks, IEEE Trans. Green Commun. Netw. 7 (1) (2022) 158–170.

[24] Y. Wang, Y. Deng, L. Kang, F. Jiang, F. Jiang, Reinforcement learning-based energy efficiency optimization for RIS-assisted UAV hybrid uplink and downlink system, Comput. Netw. 245 (2024) 110390.

[25] S. Ahmad, J. Zhang, A. Nauman, A. Khan, K. Abbas, B. Hayat, Deep-EERA: DRL-based energy-efficient resource allocation in UAV-empowered beyond 5G networks, Tsinghua Sci. Technol. (2024).

[26] H. Yang, W. Ruby, K. Wu, Delay performance of priority-queue equipped UAV-based mobile relay networks: Exploring the impact of trajectories, Comput. Netw. 210 (2022) 108856.

[27] M. Gao, B. Zhang, L. Wang, A dynamic priority packet scheduling scheme for post-disaster UAV-assisted mobile ad hoc network, Comput. Netw. 210 (2022) 108856.

[28] T. Chen, J. He, H. Zhu, L. Cai, P. Yue, J. Wang, Resource allocation in drone-assisted emergency communication systems, in: IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS, IEEE, 2020, pp. 132–137.

[29] Y. Liu, J. Yan, X. Zhao, Deep reinforcement learning based latency minimization for mobile edge computing with virtualization in maritime UAV communication network, IEEE Trans. Veh. Technol. 71 (4) (2022) 4225–4236.

[30] Y. Liu, J. Yan, X. Zhao, Deep-reinforcement-learning-based optimal transmission policies for opportunistic UAV-aided wireless sensor network, IEEE Internet Things J. 9 (15) (2022) 13823–13836.

[31] L. Xiaochen, M. Weidong, Z. Rui, A new store-then-amplify-and-forward protocol for UAV mobile relaying, IEEE Wirel. Commun. Lett. 9 (5) (2019) 1-1.

[32] G. Cheng, G. Chongtao, Z. Shengli, D. Zhiguo, UAV-enabled NOMA networks analysis with selective incremental relaying and imperfect CSI, IEEE Trans. Veh. Technol. 69 (12) (2020) 16276–16281.

[33] P. Cunhua, R. Hong, D. Yansha, M. Elkashlan, A. Nallanathan, Joint blocklength and location optimization for URLLC-enabled UAV relay systems, IEEE Commun. Lett. 23 (3) (2019) 498–501.

[34] X. Jiang, Z. Wu, Z. Yin, Z. Yang, Joint power and trajectory design for UAV-relayed wireless systems, IEEE Wirel. Commun. Lett. 8 (3) (2019) 697–700.

[35] Z. Shuhang, Z. Hongliang, H. Qichen, B. Kaigui, S. Lingyang, Joint trajectory and power optimization for UAV relay networks, IEEE Commun. Lett. 22 (1) (2018) 161–164.

[36] T. Matsuda, M. Kaneko, T. Hiraguri, K. Nishimori, T. Kimura, A. Nakao, Adaptive direction control for UAV full-duplex relay networks using multiple directional antennas, IEEE Access (2020) 885083-85093.

[37] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, IEEE J. Sel. Areas Commun. (2017) arxiv preprint arxiv:1707.06347.