Intelligent and Converged Networks ISSN 2708-6240
2021, 2(2): ??–?? DOI: 10.23919/ICN.2021.0002

Distributed reinforcement learning-based framework for energy-efficient UAV relay against jamming

Weihang Wang, Zefang Lv, Xiaozhen Lu, Yi Zhang*, and Liang Xiao

Abstract: Unmanned aerial vehicle (UAV) network is vulnerable to jamming attacks, which may cause severe damage like communication outages. Due to the energy constraint, the source UAV cannot blindly enlarge the transmit power, along with the complex network topology with high mobility, which makes the destination UAV unable to evade the jammer by flying at will. To maintain communication with a limited battery capacity in the UAV networks in the presence of a greedy jammer, in this paper, we propose a distributed reinforcement learning (RL)-based energy-efficient framework for the UAV networks with constrained energy under jamming attacks to improve the communication quality while minimizing the total energy consumption of the network. This framework enables each relay UAV to independently select its transmit power based on historical state-related information without knowing the moving trajectory of other UAVs as well as the jammer. The location and battery level of each UAV need not be shared with other UAVs. We also propose a deep RL based anti-jamming relay approach for UAVs with portable computation equipment like Raspberry Pi to achieve higher and faster performance. We study the Nash equilibrium (NE) and the performance bounds based on the formulated power control game. Simulation results show that the proposed schemes can reduce the bit error rate (BER) and reduce energy consumption of the UAV network compared with the benchmark method.

Key words: unmanned aerial vehicles; relay; jamming; reinforcement learning

1 Introduction

With the fast development and increasing functionality, *unmanned aerial vehicles* (*UAVs*) have become the enablers of more and more advanced applications, such as traffic monitoring and remote sensing[1,2]. Different from mobile ad hoc networks (MANETs) and vehicular ad hoc networks (VANETs), UAV networks are more vulnerable to *jamming attacks* due to the limited battery capacity, time-varying link quality, higher mobility, and dynamic network topology[3]. A jammer may send fake or replayed signals to block ongoing UAV communications and therefore results in transmission outages[4]. The jamming attack also exhausts the battery capacity of the UAVs for retransmissions. Moreover, a UAV may be cheated to land in an unintended spot when its communication link to the operators is blocked by jamming[5].

To overcome possible jamming attacks, *power control* has been applied as an effective technique, which adjusts the transmit power of the device so that the received signal-to-interference-plus-noise ratio (SINR) can be greater or equal to a minimum acceptable threshold. However, most existing solutions[6–8] are not applicable to UAV networks due to the high mobility of the UAVs. Besides, the rapid variant channel conditions and the unknown jamming model lead to great challenges to the premodeling of the channel for

• Weihang Wang, Zefang Lv, Xiaozhen Lu, Yi Zhang, and Liang Xiao are with the Department of Information and Communication Engineering, Xiamen University, Xiamen, China. E-mail: {yizhang, lxiao}@xmu.edu.cn.

• Liang Xiao is also with Beijing Key Laboratory of Mobile Computing and Pervasive Device, Beijing 100190, China.

∗ To whom correspondence should be addressed.

Manuscript received: 2020-12-15; revised: 2021-02-05; accepted: 2021-04-16

power control against jamming.

In the UAV-aided communication network, the UAV acts as a relay to help forward messages to the target device suffering from jamming attacks. Under the unknown network and jamming model, the UAV selects its relay power by adopting smart strategies. *Reinforcement learning* (*RL*), which has been widely used in UAV-aided VANETs[9] and mobile communications[10], enables a UAV to optimize the relay power via trial-and-error without knowing the network topology and jamming model. The RL-based UAV relay system can not only reduce the bit error rate (BER) but also save the energy consumption of message relays. Note that a single UAV may be seriously jammed or possess a limited battery, UAVs deployed as swarms provide great potentials to mitigate the damage of jamming by cooperative message relays. It fully exploits the line-of-sight links between the UAVs and the target device to improve the communication performance[1].

However, most previous works have adopted the centralized framework in the UAV networks, in which a learning center is operated to send control signals to other UAVs. The learning center must deal with huge state spaces when the UAV network expands. To the best of our knowledge, the distributed framework learned individually by each relay UAV is rarely discussed. Compared with the centralized framework, the distributed framework is more robust for a large-scale UAV network. Besides, the UAV under the distributed framework need not share state-related information, i.e., received jamming power and the current battery level, with the learning center, which may cause the leak of privacy and require additional communication overhead.

In this paper, we propose a distributed framework for multi-relay UAV networks against jamming attacks, where each relay UAV independently determines the optimal transmit power for message relays from the source UAV to the destination UAV. The system objective is to maximize the communication performance and meanwhile minimize the overall energy consumption. To handle the external jamming attacks, two RL-based approaches, i.e., the RL-based energy-efficient anti-jamming UAV relay (REAR) approach and the enhanced deep REAR (DREAR) approach, are provided to maximize the expected long-term utility of each relay UAV in terms of Q-value, which depends on the communication performance as well as the energy consumption. The REAR approach exploits historical state-related information to construct a distributed RL model for each relay UAV without knowing the moving trajectory of other UAVs as well as the jammer. In each time slot, each relay UAV selects its optimal transmit power based on the relay policy derived using the proposed RL model and then updates the model parameters after receiving the acknowledgment (ACK) frame from the destination UAV. In particular, to share the UAV relay experiences with other similar UAV-aided communication networks, a transfer learning technique, i.e., hotbooting, is employed to initialize the Q-values in order to accelerate the initial relay exploration. To further enhance the efficiency of the anti-jamming UAV relay, the DREAR approach utilizes two deep neural networks (DNNs), i.e., E-network and T-network, to compress the state space and estimate the Q-value for each UAV relay policy. The E-network outputs the estimated Q-values and the T-network outputs the target Q-values so as to mitigate over-estimation.

Our main contributions are summarized as follows.

(1) We propose a distributed framework for multi-relay UAV networks against jamming attacks. Different from the centralized learning framework, the state space of the proposed distributed approach does not increase with the number of relay UAVs in the network. Moreover, the UAV need not share its location and battery level with a learning center or other UAVs.

(2) An RL-based REAR is provided to enable each relay UAV to independently select its transmit power in a dynamically variant environment without knowing the network topology and jamming strategy. We also propose a DREAR approach to further improve the performance of the UAV network, which adopts the DNN technique to compress the state space and estimate the Q-values for each relay policy. The DREAR approach is suitable for portable computation equipment like Raspberry Pi.

(3) Interaction between the relay UAVs and the jammer is formulated as an anti-jamming power control game. We study the Nash equilibrium (NE) and also

provide the performance bound in terms of the BER, energy consumption, and utility of the overall network. Simulation results show that the proposed schemes can improve energy efficiency and reduce BER compared with the benchmark scheme[6].

The rest of this paper is organized as follows. We review related literature in Section 2 and present the system model in Section 3. We propose two RL-based anti-jamming relay approaches for energy-efficient UAV networks in Section 4. We formulate the anti-jamming relay game in Section 5 and provide the simulation results in Section 6. Finally, the conclusion is drawn in Section 7.

2 Related work

Recently, UAV-aided relay networks have been widely used in many areas recently[11–14]. A two-hop UAV relay network uses a genetic algorithm to determine the data volume and design the trajectory of mobile relays to improve the data downloading rate and reduce the latency[11]. Another cooperative UAV relay scheme for wireless sensor networks[12] optimizes the packet load scheduling strategy via solving a min-max problem to reduce energy consumption with the while guaranteeing BER. A UAV-aided communication system[13] uses a numerical approach to derive the optimal location for UAVs under both static and mobile air-to-ground communication scenario to maximize the communication efficiency in terms of energy consumption and outage probability. A multiple UAV relay network[14] optimizes the UAV placement to increase the transmission quality and compares the performance of a single multi-hop link with that of a multiple dual-hop links in terms of the outage probability and BER.

The jamming attack in UAV network degrades the communication performance and has drawn great research attention on the countermeasures of jamming attacks[6–8, 15]. A cooperative anti-jamming scheme[7] uses a pricing-mechanism-based best-response algorithm to optimize the channel utilization of different users to improve the network throughput. A joint power control and user scheduling scheme against jamming[6] uses dynamic programming to sequentially obtain the optimal power allocation and user scheduling strategy at each slot to improve data rate of the wireless network. A robust anti-jamming beamforming scheme[15] uses linearly constrained optimization to improve the jamming resistance and signal-to-interference-plus-noise ratio (SINR) of navigation signal with a minimum computation load. A cooperative relay scheme against radio-frequency jamming attacks for vehicular networks[8] firstly employs a heuristic selection algorithm to select vehicles outside the jammed area as relays and then exploits the spatial diversity of selected relays to improve the SINR of the messages that the jammed vehicles receive by combining signals from all relays.

Without the stringent requirement of knowing the channel or jamming model, RL-based methods have been widely applied in anti-jamming communications[4, 9, 16–18]. A power control anti-jamming scheme for massive multiple-input multiple-output (MIMO) systems[16] uses a policy hill-climbing (PHC) algorithm to select the transmit power of the base station based on the previous SINR and received jamming power to improve the average SINR and the sum data rate of all user equipments in the system. A deep Q-network-based anti-jamming scheme[17] formulates a Stackelberg dynamic game between an intelligent jammer UAV and mobile users on the ground. It then optimizes user mobility to reduce the received jamming power of the users. A Q-learning-based power control scheme[4] formulates the UAV to the ground channel model and optimizes the UAV transmit power to the base station under the presence of a jammer to improve the SINR. A deep Q-learning-based anti-jamming scheme[18] enables the UAVs to allocate the transmit power over multiple frequency channels based on the received jamming power to improve the secrecy capacity of the UAV system against a smart jammer. A UAV-aided anti-jamming relay scheme in VANETs[9] employs the PHC algorithm to help the UAV relay determine whether to relay the message according to its radio channel condition and previous transmission quality to reduces the BER of the forwarded messages.

3 System model

As shown in Fig. 1, a UAV network mainly consists of a *source UAV*, which intermittently broadcasts messages to its intended *destination UAV* with multiple UAVs as relays. In particular, a malicious *jammer* located near the destination tries to send jamming signals at the same spectrum to interfere with the communication. To cope with the external jamming attacks, a total N *relay UAVs* located between the source and destination can act as relay nodes to help maintain the communication quality of the source-destination link without extra flying.

In the proposed time-slotted system, for simplicity, the source UAV is assumed to only broadcast one message in each time slot. At a time slot k, the source UAV broadcasts a message with the transmit power $p^{(k)}$. It should be noted that both the destination UAV and the relay UAVs may receive this message. For an arbitrary relay UAV, it decodes the message received from the source UAV, which enables the measurement of the received signal strength indicator (RSSI) $r^{(k)}$ and BER $\phi^{(k)}$. Note that the battery power equipped at the UAV is usually limited and crucial. Additionally, the relay UAV must decide its transmit power $x^{(k)}$ for the message relay, which has an upper limit denoted by X. To lower the BER of the target messages, each relay UAV independently determines its relay power with a power constraint. In particular, the relay UAV must ensure whether the remaining energy is sufficient to relay the message or not by observing its current battery level $b^{(k)}$.

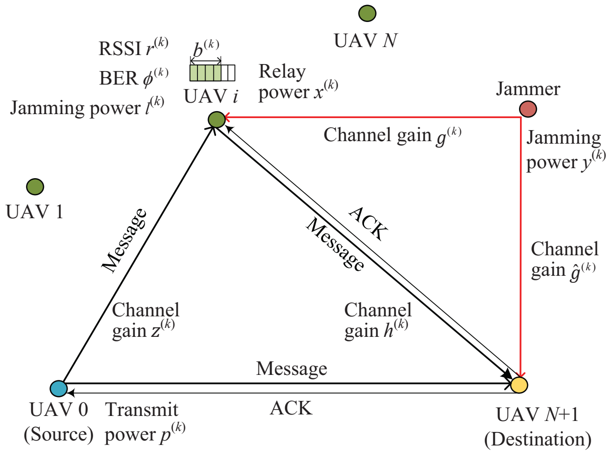


Fig. 1 RL-based UAV-aided wireless relay networks against jamming attacks.

Due to the broadcast nature, the destination UAV may receive multiple copies of the target message at a time slot k, each of which is either directly sent from the source UAV or relayed by some relay UAVs. For each received message, the address of the sender and the corresponding BER $\rho^{(k)}$ can be estimated and further assembled together into one ACK frame. The destination UAV broadcasts this ACK frame as feedback. Successful message delivery to the destination UAV requires at least one message with a BER less than the maximum BER ε. In some cases, the target message may not be successfully received by the destination UAV even with the assistance of those relay UAVs. We introduce a flag denoted by $\omega^{(k)}$ to indicate the message state: if the target message is successfully received, $\omega^{(k)} = 0$; otherwise, $\omega^{(k)} = 1$. The failure of the message delivery is regarded as a punishment in the utility of the relay UAV for learning and decision.

The jammer of the proposed scheme is more smart and detrimental compared with the static jammer with fixed or random jamming power. A greedy jammer selects its moving strategy and the transmit power $y^{(k)}$ within the range of $[0, Y]$, to degrade the UAV communication with less jamming cost. Sending jamming signals for a random period may reduce the energy consumption of the jammer, making the attacks last longer. In the proposed system, *each relay UAV aims to optimize the transmit power to achieve higher energy efficiency when providing message relay against the greedy jammer*. The key notations of this paper are listed in Table 1.

4 RL-based energy-efficient UAV relay against jamming

In this section, we propose the REAR approach, which is independently implemented by each relay UAV, to determine the optimal relay power to mitigate jamming attacks. The proposed REAR approach decreases the BER of the target message and reduces the overall network energy consumption thereby improving communication reliability. The hotbooting method is used in REAR to exploit the UAV relay power control experiences in the similar network model and anti-jamming scenarios to initialize the Q-values for each relay policy, consequently accelerating the initial

Table 1 List of key notations.

Symbol	Description
N	Number of relay UAVs
$x^{(k)} \in [0, X]$	Relay power of UAV i at a time slot k
$y^{(k)} \in [0, Y]$	Jamming power
$p^{(k)} \in [0, P]$	Transmit power of the source
$r^{(k)}$	RSSI of the message received by UAV i
$z^{(k)}$	Channel gain from the source to UAV i
$h^{(k)}$	Channel gain from UAV i to the destination
$\hat{g}^{(k)}, g^{(k)}$	Channel gain from jammer to the {destination UAV, UAV i}
$b^{(k)}, \tilde{b}^{(k)}$	{Measured, estimated} battery level of a relay UAV
ϑ	Battery threshold
ς	Minimum SINR for successful transmission
ε	Maximum BER for successful transmission
$\phi^{(k)}$	BER of the message received by UAV i from the source
$\rho_i^{(k)}$	BER of the message received by the destination from UAV i
$l^{(k)}$	Jamming power received by UAV i
$E^{(k)}$	Energy consumption of the UAV i

exploration. It should be noted that the proposed REAR approach is efficient for those UAVs with limited computing resources. With the development of single-board computers, such as portable Raspberry Pi, some UAVs can learn and optimize its relay policy in more complex environments with higher dimensional states. To enhance the efficiency of the anti-jamming UAV relay, we further propose a DREAR approach by using the DNN technique.

4.1 REAR approach for UAV relay

In the proposed REAR approach, we define a state vector $\boldsymbol{s}^{(k)}$ at a time slot k for each UAV relay as follows:

$$\boldsymbol{s}^{(k)}=[\phi^{(k)},r^{(k)},h^{(k)},b^{(k)},l^{(k)},\min_{0\leqslant i\leqslant N}\rho_i^{(k-1)},\rho^{(k-1)},\omega^{(k-1)}] \tag{1}$$

The BER of the message $\phi^{(k)}$ and the RSSI of its signal $r^{(k)}$ are the key metrics measured using the relay UAV to reflect the transmission quality of the UAV network. The channel gain of the relay-destination link $h^{(k)}$ can be estimated based on the preambles of the messages[19]. The current received jamming power is calculated using $l^{(k)}=y^{(k)}g^{(k)}$. From the ACK frame delivered by the destination UAV, the BER $\rho^{(k-1)}$ and the message state $\omega^{(k-1)}$ of the relay message at the previous time slot $k-1$ are known. Specifically, the minimum BER $\min_{0\leqslant i\leqslant N}\rho_i^{(k-1)}$ is recorded. The measurements in the state metric are quantized into limited discrete levels, in which the BER $\phi^{(k)}$ and $\min_{0\leqslant i\leqslant N}\rho_i^{(k-1)}$ are quantized by the exponential region and others are quantized uniformly in the range of possible values. The granularity of quantization will influence the size of the state space. Therefore, when it becomes finer, the algorithm needs more exploration steps before convergence. However, the algorithm is more likely to obtain the theoretical optimal value. Basing on the current state vector $\boldsymbol{s}^{(k)}$, the relay UAV selects its transmit power $x^{(k)}$ according to the Q-function table, which is quantized in the range $[0, X]$ with a total of $M+1$ discrete levels, i.e., $x^{(k)} \in \{mX/M : m \in \{0,\dots,M\}\}$.

If a relay UAV determines to help forward the message, $x^{(k)} > 0$ should be guaranteed; otherwise, $x^{(k)} = 0$. The relay UAV uses ϵ-greedy method to select the action, i.e., relay power, with ϵ probability for randomly choosing to avoid stopping at a local optimal policy during the learning process. To ensure the feasibility of the selected relay power $x^{(k)}$, it should be satisfied that the remaining battery level after this message relay should be greater or equal to a minimum battery threshold ϑ, i.e., $b^{(k)} - x^{(k)} \geqslant \vartheta$. If the remaining battery level is not sufficient, the relay power should then be set to zero, that is, the relay UAV denies this message relay.

When an ACK frame of the current message is received, the minimum BER $\min_{0\leqslant i\leqslant N}\rho_i^{(k)}$ will be calculated and recorded. If $\min_{0\leqslant i\leqslant N}\rho_i^{(k)} \leqslant \varepsilon$, the flag of the message state is set as $\omega^{(k)} = 0$ to indicate a successful transmission, else, set the flag $\omega^{(k)} = 1$. In this case, if the corresponding BER is not contained in the ACK frame, e.g., the relay UAV i has denied forwarding the message or the relay message fails to reach the destination UAV due to the jamming, the minimum BER $\min_{0\leqslant i\leqslant N}\rho_i^{(k)}$ and the maximum BER $\max_{0\leqslant i\leqslant N}\rho_i^{(k)}$ should be recorded, where the maximum BER is regarded as a conservative estimation of the actual BER for this message. If no ACK frame is received, the flag is set as $\omega^{(k)} = 1$. It should be noted that the flag $\omega^{(k)}$ is utilized as a punishment in the utility of the relay UAV to represent the transmission outages.

Furthermore, the utility of the relay UAV denoted by $u^{(k)}$ is evaluated upon receiving the ACK frame from the destination UAV:

$$u^{(k)} = -E^{(k)} - c_1 \min_{0 \leqslant i \leqslant N} \rho_i^{(k)} - c_2 \omega^{(k)} \tag{2}$$

where c_1 and c_2 are the weights of the minimum BER received by the destination and transmission outage punishment, respectively, and c_2 is determined to be an empirically large number. The energy consumption $E^{(k)}$ is measured to evaluate the energy efficiency by observing the battery level $b^{(k+1)}$ at the end of the relay, i.e., $E^{(k)} = b^{(k)} – b^{(k+1)}$. The Q-function is exploited to obtain the optimal transmit power for the relay UAV and is updated using the Bellman iterative equation based on the current relay experience and the utility with learning rate α and the discount factor γ.

The pseudocode for the proposed REAR approach is presented in Algorithm 1. We can observe that a transfer-learning-based hotbooting method is used to improve the efficiency of exploration at the beginning of the learning process. This method initializes the Q-values for each relay policy with the anti-jamming UAV relay experiences randomly selected from several similar UAV relay scenarios.

4.2 Enhanced DREAR approach for UAV relay

In the enhanced DREAR approach, two isomorphic fully connected DNNs are used to compress the state space of the UAV relays, i.e., E-network and a T-network. As shown in Fig. 2, each network consists of an input layer, two hidden layers with f_1 and f_2 units, and an output layer with $M + 1$ units. All of them use a leaky rectified linear unit as the activation function. The two networks decouple the action selection and the computation of target Q-value[20]. In particular, the E-network outputs the maximum estimated Q-value of each state with weights $\boldsymbol{\theta}_{\mathrm{E}}^{(k)}$ while the T-network outputs the target Q-value with weights $\boldsymbol{\theta}_{\mathrm{T}}^{(k)}$.

When a relay UAV receives the message from the source at a time slot k, it measures the BER of the message $\phi^{(k)}$, the RSSI of its signal $r^{(k)}$, the current battery level $b^{(k)}$, and the received jamming power $l^{(k)}$. These parameters formulate the current state $\boldsymbol{s}^{(k)}$ as Eq. (1), which is treat as the input of the E-network. The transmit power is the output of the E-network to maximize the Q-value. The estimated long-term utility is expressed as

$$x_{\max}^{(k)}(\boldsymbol{s}^{(k)}, \boldsymbol{\theta}_{\mathrm{E}}^{(k)}) = \arg\max_{x\prime} Q(\boldsymbol{s}^{(k)}, x'; \boldsymbol{\theta}_{\mathrm{E}}^{(k)}) \tag{3}$$

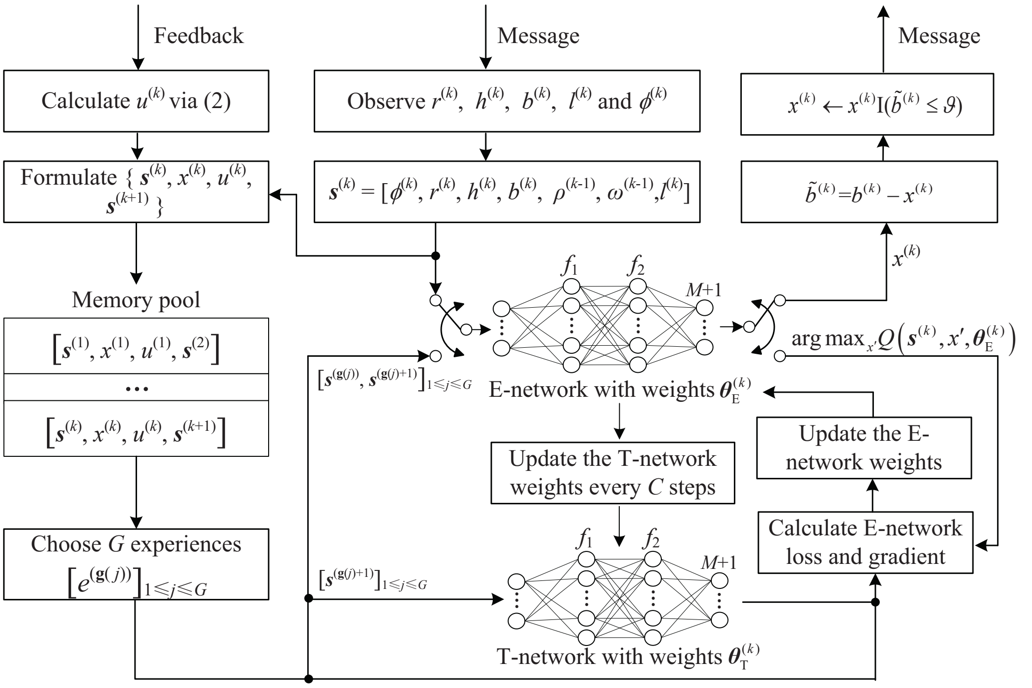

Feedback
Message
Message
Calculate $u^{(k)}$ via (2)
Observe $r^{(k)}$, $h^{(k)}$, $b^{(k)}$, $l^{(k)}$ and $\phi^{(k)}$
$x^{(k)} \leftarrow x^{(k)}\mathrm{I}(\tilde{b}^{(k)} \leq \vartheta)$
Formulate { $\boldsymbol{s}^{(k)}, x^{(k)}, u^{(k)}, \boldsymbol{s}^{(k+1)}$ }
$\boldsymbol{s}^{(k)} = [\phi^{(k)}, r^{(k)}, h^{(k)}, b^{(k)}, \rho^{(k-1)}, \omega^{(k-1)}, l^{(k)}]$
$\tilde{b}^{(k)} = b^{(k)} - x^{(k)}$
$x^{(k)}$
f_1
f_2
M+1
Memory pool
$\left[\boldsymbol{s}^{(1)}, x^{(1)}, u^{(1)}, \boldsymbol{s}^{(2)}\right]$
…
$\left[\boldsymbol{s}^{(k)}, x^{(k)}, u^{(k)}, \boldsymbol{s}^{(k+1)}\right]$
$\arg\max_{x'} Q\left(\boldsymbol{s}^{(k)}, x', \boldsymbol{\theta}_{\mathrm{E}}^{(k)}\right)$
$\left[\boldsymbol{s}^{(\mathbf{g}(j))}, \boldsymbol{s}^{(\mathbf{g}(j)+1)}\right]_{1 \leqslant j \leqslant G}$
E-network with weights $\boldsymbol{\theta}_{\mathrm{E}}^{(k)}$
Update the T-network weights every C steps
Update the E-network weights
Calculate E-network loss and gradient
Choose G experiences $\left[e^{(\mathbf{g}(j))}\right]_{1 \leqslant j \leqslant G}$
$\left[\boldsymbol{s}^{(\mathbf{g}(j)+1)}\right]_{1 \leqslant j \leqslant G}$
f_1
f_2
M+1
T-network with weights $\boldsymbol{\theta}_{\mathrm{T}}^{(k)}$

Fig. 2 Illustration of DREAR for UAV networks.

Algorithm 1 REAR approach for UAV relay

```
Initialize parameters: $\min_{0\leqslant i\leqslant N} \rho_i^{(0)}$ and $\omega^{(0)}$
Obtain $\tilde{\boldsymbol{Q}}$ from similar scenarios based on hotbooting
Initialize Q-function as $\boldsymbol{Q} = \tilde{\boldsymbol{Q}}$
for $k = 1, 2, \ldots$ do
    Relay UAV receives a message from the source UAV
    Measure $\phi^{(k)}$, $r^{(k)}$, $l^{(k)}$, $h^{(k)}$ and observe $b^{(k)}$
    Formulate $\boldsymbol{s}^{(k)}$ by (1)
    Select $x^{(k)} \in \{mX/M : m \in \{0, \ldots, M\}\}$ via $\epsilon$-greedy method
    if $b^{(k)} - x^{(k)} \geqslant \vartheta$ then
        Relay the message with power $x^{(k)}$
    else
        Set $x^{(k)} = 0$ (insufficient power, mute relaying)
    end if
    if Receive the ACK frame then
        if $\rho^{(k)}$ is contained then
            Calculate the minimum BER $\min_{0\leqslant i\leqslant N} \rho_i^{(k)}$
        else
            Calculate $\max_{0\leqslant i\leqslant N} \rho_i^{(k)}$ and $\min_{0\leqslant i\leqslant N} \rho_i^{(k)}$
            Set $\rho^{(k)} = \max_{0\leqslant i\leqslant N} \rho_i^{(k)}$
        end if
        if $\min_{0\leqslant i\leqslant N} \rho_i^{(k)} \leqslant \varepsilon$ then
            Set $\omega^{(k)} = 0$ (successful transmission)
        else
            Set $\omega^{(k)} = 1$ (failed transmission)
        end if
    else
        Set $\omega^{(k)} = 1$ (failed transmission)
    end if
    Calculate $u^{(k)}$ by Eq. (2)
    Update $Q(\boldsymbol{s}^{(k)}, x^{(k)})$ by the Bellman iterative equation
end for
```

based on which the relay UAV selects its transmit power $x^{(k)} \in \{mX/M : m \in \{0, \ldots, M\}\}$ using ϵ-greedy method at the current time slot k, on the other hand, the output transmit power is further used as the input of the T-network for Q-value calculation. Similar to Algorithm 1, the relay UAV estimates the battery power after relaying the message and then calculates the utility $u^{(k)}$ using Eq. (2) based on the received ACK frame.

Specifically, there is a memory pool $\mathcal{D}$ to store the experiences of the relay UAV at a time slot k denoted by $e^{(k)} = \{\boldsymbol{s}^{(k)}, x^{(k)}, u^{(k)}, \boldsymbol{s}^{(k+1)}\}$. In this way, the UAV can randomly and uniformly sample total G experiences from $\mathcal{D}$ to formulate a minibatch $\mathcal{B}$.

Here, $\boldsymbol{g}(j)_{j\in[1,G]}$ represents the serial number of selected experiences, i.e., $\boldsymbol{g}(\cdot) \sim U(1, k)$. Furthermore, the weights of E-network $\boldsymbol{\theta}_{\mathrm{E}}^{(k+1)}$ are updated using Eq. (4) in order to minimize the mean squared error (MSE) between the target and the estimated Q values. The weights of the E-network $\boldsymbol{\theta}_{\mathrm{E}}^{(k+1)}$ are also utilized to update the weights of the T-network $\boldsymbol{\theta}_{\mathrm{T}}^{(k+1)}$ of each $\mathcal{C}$ time slot. The overall description of the proposed DREAR approach is summarized in Algorithm 2.

$$\boldsymbol{\theta}_{\mathrm{E}}^{(k+1)} \leftarrow \arg\min_{\boldsymbol{\theta}_{\mathrm{E}}^{*}} \mathbb{E}_{e(\boldsymbol{g}(j))\in\mathcal{B}} \cdot \Bigg[\bigg(u^{(\boldsymbol{g}(j))} + \gamma Q\bigg(\boldsymbol{s}^{(\boldsymbol{g}(j)+1)}, \arg\max_{x'} Q\Big(\boldsymbol{s}^{(\boldsymbol{g}(j)+1)}, x'; \boldsymbol{\theta}_{\mathrm{E}}^{*}\Big); \boldsymbol{\theta}_{\mathrm{T}}^{(k)} - Q\left(\boldsymbol{s}^{(\mathrm{g}(j))}, x^{(\mathrm{g}(j))}; \boldsymbol{\theta}_{\mathrm{E}}^{*}\right)\bigg)^{2}\Bigg] \tag{4}$$

5 Equilibrium analysis for anti-jamming power control game

So far, we have introduced the proposed REAR and DREAR approaches for optimal relay power control against jamming attacks. Note that the jammer expects to degrade the UAV communication by selecting its jamming power while the relay UAVs must optimize the transmit power for successful message relay and improved energy efficiency. In this section, we model the interactions between the relay UAVs and the jammer as an anti-jamming power control game and study its NE.

5.1 Power control game

There are two sides to the proposed power control game: one side is the N relay UAVs, the other side is the greedy jammer. To against the jamming attack, each relay UAV optimizes its relay power in the range $[0, X]$. Therefore, the power control strategy of the total N relay UAVs is represented by a vector $\boldsymbol{x}^{(k)} = \{[x_i^{(k)}]_{1\leqslant i\leqslant N} | 0 \leqslant x_i^{(k)} \leqslant X\}$. On the contrary, the greedy jammer adjusts its transmit power $y^{(k)} \in [0, Y]$ to minimize the utility of the relay UAVs and reduce its energy, where Y is the maximum transmit power of the jammer.

In the multi-relay-enabled UAV network, there are a total of N possible paths from the source UAV to the destination UAV. The SINR of the i-th path is constrained by the lower one between the source-to-relay-i link and relay-i-to-destination link[9]. As we

Algorithm 2 DREAR approach for UAV relay

```
Initialize $\min_{0\leqslant i\leqslant N}\rho_i^{(0)}$, $\omega^{(0)}$, $\boldsymbol{\theta}_{\mathrm{E}}^{(k)}=\boldsymbol{\theta}_{\mathrm{E}}^{*}$ and $\boldsymbol{\theta}_{\mathrm{T}}^{(1)}=\boldsymbol{\theta}_{\mathrm{E}}^{(1)}$
for $k=1,2,\ldots,$ do
    Relay UAV receives a message from the source UAV
    Measure and observe $\phi^{(k)}$, $r^{(k)}$, $l^{(k)}$, $h^{(k)}$, and $b^{(k)}$
    Formulate $\boldsymbol{s}^{(k)}$ by Eq. (1)
    Input $\boldsymbol{s}^{(k)}$ to the E-network
    E-network outputs $x_{\max}^{(k)}(\boldsymbol{s}^{(k)},\boldsymbol{\theta}_{\mathrm{E}}^{(k)})$
    Select $x^{(k)}\in\{mX/M:m\in\{0,\ldots,M\}\}$ by $\epsilon$-greedy method
    if $b^{(k)}-x^{(k)}\geqslant\vartheta$ then
        Relay the message with power $x^{(k)}$
    else
        Set $x^{(k)}=0$ (insufficient power, mute relaying)
    end if
    if Receive the feedback for message then
        Calculate the minimum BER $\min_{0\leqslant i\leqslant N}\rho_i^{(k)}$
        if $\min_{0\leqslant i\leqslant N}\rho_i^{(k)}\leqslant\varepsilon$ then
            Set $\omega^{(k)}=0$ (successful transmission)
        else
            Set $\omega^{(k)}=1$ (failed transmission)
        end if
        if $\rho^{(k)}$ is contained in the feedback then
            Calculate $\max_{0\leqslant i\leqslant N}\rho_i^{(k)}$ and $\min_{0\leqslant i\leqslant N}\rho_i^{(k)}$
            Set $\rho^{(k)}=\max_{0\leqslant i\leqslant N}\rho_i^{(k)}$
        end if
    else
        Set $\omega^{(k)}=1$ (failed transmission)
    end if
    Calculate $u^{(k)}$ by Eq. (2)
    Formulate $\boldsymbol{s}^{(k+1)}$ by Eq. (1)
    $\mathcal{D}\leftarrow\mathcal{D}\cup\{\boldsymbol{s}^{(k)},x^{(k)},u^{(k)},\boldsymbol{s}^{(k+1)}\}$
    for $j=1,2,\ldots,G$ do
        $e^{(\boldsymbol{g}(j))}=\mathcal{D}(\boldsymbol{g}(j))$
    end for
    Update $\boldsymbol{\theta}_{\mathrm{E}}^{(k+1)}$ via Eq. (2)
    Update $\boldsymbol{\theta}_{\mathrm{T}}^{(k+1)}$ with $\boldsymbol{\theta}_{\mathrm{E}}^{(k+1)}$ every $C$ time slots
end for
```

have discussed, the destination UAV only selects the message with the minimum BER, namely, the message with the maximum SINR. When the direct channel from the source to the destination experiences an outage, the SINR of the UAV network is determined by the maximum SINR of those possible paths:

$$\xi_{\mathrm{S,D}}=\max_{1\leqslant i\leqslant N}\left\{\min\left\{\frac{pz_i}{\sigma^2+y^{(k)}g_i},\frac{x_i^{(k)}h_i}{\sigma^2+y^{(k)}\hat{g}}\right\}\right\} \tag{5}$$

Considering the quadrature phase-shift keying modulation, the corresponding minimum BER of the UAV network can be represented by

$$\min_{1\leqslant i\leqslant N}\rho_i=\frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{\xi_{\mathrm{S,D}}}{2}}\right) \tag{6}$$

where erfc$(\cdot)$ represents the complementary error function.

The objective of the UAV network is to improve the network utility, which depends on the minimum BER of the target message and the total energy consumption. Similar to the utility function Eq. (4), we use the maximum SINR from the source UAV to the destination UAV $\xi_{\mathrm{S,D}}^{(k)}$ instead of the minimum BER $\min_{1\leqslant i\leqslant N}\rho_i^{(k)}$ for simplicity. Further, the maximum BER for successful transmission ε is also converted to the minimum SINR ς. Besides, to omit the transmission outage punishment $\omega^{(k)}$ in the theoretical analysis, we have to consider a successful transmission constrain, i.e., $\xi_{\mathrm{S,D}}^{(k)}>\varsigma$. In this way, the utility of the UAV network is calculated as

$$u_{\mathrm{R}}^{(k)}=c_3\xi_{\mathrm{S,D}}^{(k)}-\sum_{1\leqslant i\leqslant N}E_i^{(k)} \tag{7}$$

$$\text{s.t.}\quad \xi_{\mathrm{S,D}}^{(k)}\geqslant\varsigma \tag{8}$$

where c_3 represents the weight of the maximum SINR $\xi_{\mathrm{S,D}}^{(k)}$. Accordingly, the utility of the greedy jammer, which aims to interrupt the message transmission of the UAV network, is expressed as the negative of the UAV network utility u_{R} minus the jamming power consumption,

$$u_{\mathrm{J}}=-u_{\mathrm{R}}^{(k)}-c_4y^{(k)} \tag{9}$$

where c_4 is an invariable cost coefficient.

5.2 Equilibrium analysis

To study NE, we assume that all channel gains are fixed in the rest of this section, i.e., g_i, $\hat{g}$, h_i, z_i. We also assume that the transmit power of the source UAV p and the transmission delay T are constant.

Theorem 1 Given a fixed jamming power y, the optimal solution to the relay power control is

$$\boldsymbol{x}^*=\left[\underbrace{0,\ldots,0}_{i'-1},x_{i'}^*,\underbrace{0,\ldots,0}_{N-i'}\right] \tag{10}$$

where $x_{i'}^*$ is the optimal relay power determined by the relay UAV i', which is expressed as follows.

$$i' = \arg \max_{1 \leqslant i \leqslant N} \left\{ \min \left\{ \frac{pz_i}{\sigma^2 + yg_i}, \frac{x_i h_i}{\sigma^2 + y\hat{g}} \right\} \right\} \quad (11)$$

The utility achieved by the UAV network could be reduced to

$$\hat{u}_{\mathrm{R}} = c_3 \min \left\{ \frac{pz_{i'}}{\sigma^2 + yg_{i'}}, \frac{x_{i'} h_{i'}}{\sigma^2 + y\hat{g}} \right\} - Tx_{i'} \quad (12)$$

Proof. We denote the index of the relay UAV with maximum SINR by i' according to Eq. (11). It should be noted that the destination UAV only selects the message with the minimum BER, namely, the message forwarded by the relay UAV i'. The best strategy of the other relay UAVs except UAV i' is to keep silent in order to avoid unnecessary energy consumption. In this way, the optimal solution to the relay power control can be represented by $\boldsymbol{x}^* = \Big[\underbrace{0,\ldots,0}_{i'-1}, x_{i'}^*, \underbrace{0,\ldots,0}_{N-i'}\Big]$, where $x_{i'}^*$ is the optimal relay power determined by the relay UAV i'. According to the definition in Eq. (7), we have

$$u_{\mathrm{R}} = c_3 \min \left\{ \frac{pz_{i'}}{\sigma^2 + yg_{i'}}, \frac{x_{i'} h_{i'}}{\sigma^2 + y\hat{g}} \right\} - \sum_{1 \leqslant i \leqslant N} Tx_i \leqslant$$

$$c_3 \min \left\{ \frac{pz_{i'}}{\sigma^2 + yg_{i'}}, \frac{x_{i'} h_{i'}}{\sigma^2 + y\hat{g}} \right\} - Tx_{i'} = \hat{u}_{\mathrm{R}} \quad (13)$$

Therefore, u_{R} can be reduced to $\hat{u}_{\mathrm{R}}$ when the UAV network tries to maximize its utility. ■

Accordingly, given the optimal power control in Eq. (10), the utility function of the jammer can be reduced to

$$\hat{u}_{\mathrm{J}} = Tx_{i'} - c_3 \min \left\{ \frac{pz_{i'}}{\sigma^2 + yg_{i'}}, \frac{x_{i'} h_{i'}}{\sigma^2 + y\hat{g}} \right\} - c_4 y \quad (14)$$

We observe that different network environments may lead to different NEs basing on Eqs. (12) and (14). Hence, we divide the power control game into two cases:

Case 1. The SINR of the source-relay-i' link is greater than the relay-i'-destination link, that is,

$$\frac{pz_{i'}}{\sigma^2 + yg_{i'}} > \frac{x_{i'} h_{i'}}{\sigma^2 + y\hat{g}} \quad (15)$$

In this way, the transmit power of the relay UAV i' should satisfy the following condition:

$$x_{i'} < \frac{pz_{i'}(\sigma^2 + y\hat{g})}{h_{i'}(\sigma^2 + yg_{i'})} \quad (16)$$

Basing on Eq. (15), we have

$$\hat{u}_{\mathrm{R}} = x_{i'} \left(\frac{c_3 h_{i'}}{\sigma^2 + y\hat{g}} - T \right) \quad (17)$$

$$\hat{u}_{\mathrm{J}} = -x_{i'} \left(\frac{c_3 h_{i'}}{\sigma^2 + y\hat{g}} - T \right) - c_4 y \quad (18)$$

The first- and second-order derivative of $\hat{u}_{\mathrm{J}}$ with respect to y will be

$$\frac{\partial \hat{u}_{\mathrm{J}}}{\partial y} = \frac{c_3 x_{i'} h_{i'} \hat{g}}{(\sigma^2 + y\hat{g})^2} - c_4 \quad (19)$$

$$\frac{\partial^2 \hat{u}_{\mathrm{J}}}{\partial^2 y} = -\frac{2c_3 x_{i'} h_{i'} \hat{g}^2}{(\sigma^2 + y\hat{g})^3} \quad (20)$$

Obviously, $\hat{u}_{\mathrm{J}}$ is a concave function of y because its second derivative is always negative when $x_{i'} \neq 0$. Because y should be within the range $[0, Y]$, the optimal jamming power will be

$$y^* = \begin{cases} \hat{y}, & 0 < \hat{y} < Y; \\ 0, & \hat{y} \leqslant 0; \\ Y, & \text{otherwise} \end{cases} \quad (21)$$

where $\hat{y}$ is derived from $\partial \hat{u}_{\mathrm{J}} / \partial y = 0$ and is expressed as follows:

$$\hat{y} = \sqrt{\frac{c_3 x_{i'} h_{i'}}{c_4 \hat{g}}} - \frac{\sigma^2}{\hat{g}} \quad (22)$$

Besides, the first derivative of $\hat{u}_{\mathrm{R}}$ of $x_{i'}$ is

$$\frac{\partial \hat{u}_{\mathrm{R}}}{\partial x_{i'}} = \frac{c_3 h_{i'}}{\sigma^2 + y\hat{g}} - T \quad (23)$$

When the condition $c_3 h_{i'} \geqslant T(\sigma^2 + y^*\hat{g})$ is satisfied, $\hat{u}_{\mathrm{R}}$ will be an increasing function of $x_{i'}$. Note that $x_{i'}$ should be ranged within $[0, X]$, the optimal relay power $x_{i'}^*$ is X. Combining with Eqs. (8) and (16), we achieve

$$\max \left\{ \frac{T(\sigma^2 + y\hat{g})}{c_3}, \frac{\varsigma \left(\sigma^2 + y\hat{g}\right)}{X} \right\} \leqslant h_{i'} \leqslant \frac{pz_{i'}(\sigma^2 + y\hat{g})}{X(\sigma^2 + yg_{i'})} \quad (24)$$

Otherwise, $\hat{u}_{\mathrm{R}}$ become an decreasing function of $x_{i'}$ with the constrain in Eq. (8). Moreover, $x_{i'}^* = \left(\varsigma \left(\sigma^2 + y\hat{g}\right)\right) / h_{i'}$ is the optimal relay power if the following condition is guaranteed:

$$0 < h_{i'} < \frac{T(\sigma^2 + y\hat{g})}{c_3} \quad (25)$$

Case 2. The SINR of the source-relay-i' link is lower or equal to the relay-i'-destination link:

$$\frac{pz_{i'}}{\sigma^2 + yg_{i'}} \leqslant \frac{x_{i'} h_{i'}}{\sigma^2 + y\hat{g}} \quad (26)$$

Similar to Case 1, the optimal jamming power will be

$$y^* = \begin{cases} \hat{y}, & 0 < \hat{y} < Y; \\ 0, & \hat{y} \leqslant 0; \\ Y, & \text{otherwise} \end{cases} \quad (27)$$

$$\hat{y} = \sqrt{\frac{c_3 pz_{i'}}{c_4 g_{i'}}} - \frac{\sigma^2}{g_{i'}} \quad (28)$$

Specifically, the first derivative of $\hat{u}_{\mathrm{R}}$ of $x_{i'}$ is

$$\frac{\partial \hat{u}_{\mathrm{R}}}{\partial x_{i'}} = -T < 0 \quad (29)$$

which means that $\hat{u}_{\rm R}$ monotonically decreases with $x_{i'} \in [0, X]$. Basing on Eq. (26), the optimal relay power should be

$$x_{i'}^* = \frac{p z_{i'}(\sigma^2 + y\hat{g})}{h_{i'}(\sigma^2 + y g_{i'})} \tag{30}$$

We utilize the derived conditions above and have the following theorem:

Theorem 2 The power control solution $(\boldsymbol{x}^*, y^*) = \big([0,\ldots,0,x_{i'}^* = X,0,\ldots,0],0\big)$ is an NE of the anti-jamming power control game when

$$\max\left\{\frac{T\sigma^2}{c_3},\frac{\sigma^2\varsigma}{X}\right\} \leqslant h_{i'} \leqslant \min\left\{\frac{p z_{i'}}{X},\frac{\sigma^4 c_4}{c_3 X \hat{g}}\right\} \tag{31}$$

The performance bound under this NE is given by

$$\xi_{\rm S,D} = \frac{X h_{i'}}{\sigma^2} \tag{32}$$

$$\sum_{1\leqslant i\leqslant N} E_i = XT \tag{33}$$

$$\min_{1\leqslant i\leqslant N} \rho_i = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{X h_{i'}}{2\sigma^2}}\right) \tag{34}$$

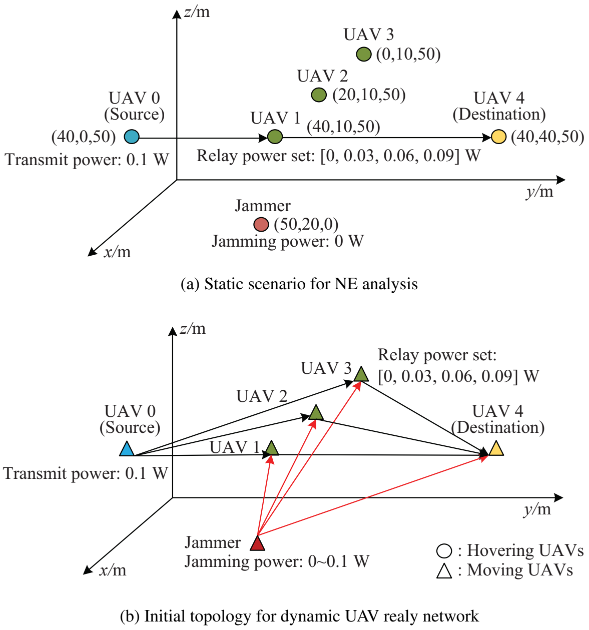
$$u_{\rm R} = \frac{c_3 X h_{i'}}{\sigma^2} - XT \tag{35}$$

That is, only the relay UAV i' with the best channel condition selects to relay the message with the maximum transmit power X and the others keep silent. The jammer selects to stop jamming to save energy.

6 Simulation result

In the performance evaluation, we consider one source UAV, one destination UAV, three relay UAVs, and a jammer on the ground, as shown in Fig. 3. The first simulation is conducted under the static theoretical analysis scenario in Section 5 with $N = 3$ relays and a mute jammer which satisfies the condition given by Eq. (31), as shown in Fig. 3a. The simulation results in Fig. 4 show that the proposed DREAR approach can converge to the optimal relay strategy after around 2500 time steps. For example, the BER, the SINR, the total network energy consumption, and the overall utility almost converge to the performance bound given by Eqs. (32)–(35).

Another simulation is conducted using a dynamic network model with a moving jammer (Fig. 3b). The transmit power of the source is 0.1 W. The relay UAVs select the relay power among the set, which is discretized uniformly among $[0, 0.9]$ W with total 4 levels. The jamming power received by the destination changes among 9.0 dBm, 9.5 dBm, and 10.0 dBm. Relays are also influenced by the jammer, the jamming power received by each relay is different and changes from 3.0 dBm, 6.0 dBm, and 7.0 dBm.



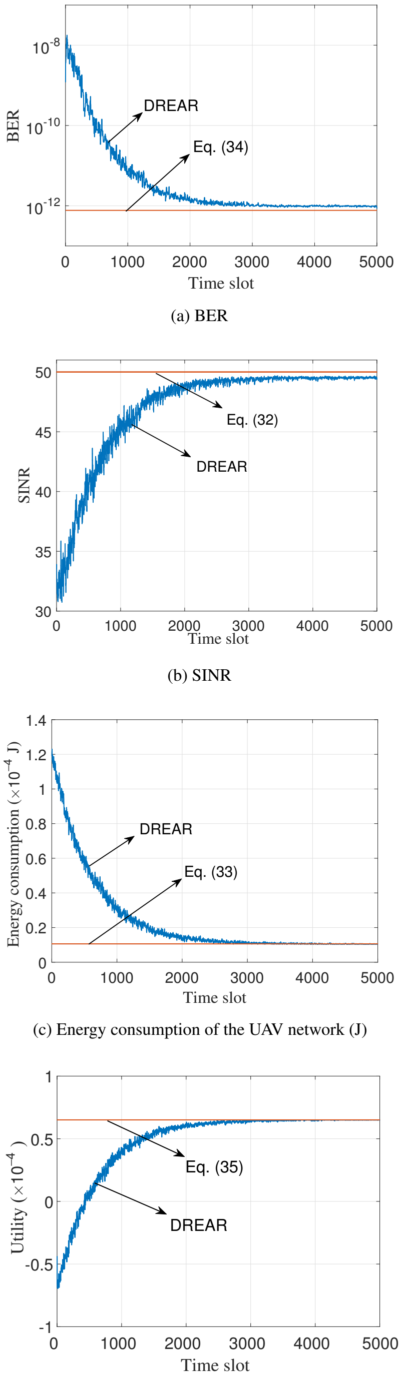
(a) Static scenario for NE analysis

(b) Initial topology for dynamic UAV realy network

Fig. 3 Simulation settings for performance evaluations.

The system performance is evaluated using the minimum BER of the received relay messages by the destination, total energy consumption of 3 relays, and utility of the network is calculated based on the minimum BER and the total energy consumption. We use the optimal power control against jamming (OPAJ) algorithm[6] with fixed optimal relay power as the benchmark algorithm. To control that the relay power is the only variable in the comparison between the proposed algorithm and the benchmark, the BER used to represent the performance of the overall system is selected from the relay message of the same relay in the two algorithms at every time slot.

Figure 5 shows the performance of the proposed schemes in the dynamic UAV relay network against jamming with a learning rate α of 0.5 and a discount factor γ of 0.7. Due to the fixed relay power of the benchmark scheme[6], the energy consumption of it remains unchanged, but the BER of it decreases a little because that it is selected from the relay message of the same relay as in the proposed algorithm. Results show that both the REAR and the DREAR approaches improve



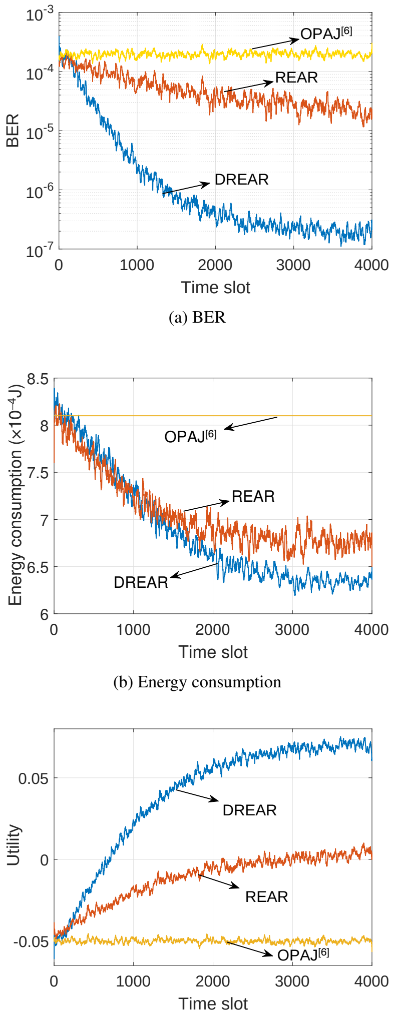
(a) BER

(b) SINR

(c) Energy consumption of the UAV network (J)

(d) Utility of the UAV network

Fig. 4 Performance of the deep RL-based energy-efficient UAV relay scheme averaged over 50 episodes for the UAV network with 3 relays in the theoretical analysis scenario compared with the performance bound.



(a) BER

(b) Energy consumption

(c) Utility

Fig. 5 Performance of the RL-based energy-efficient UAV relay schemes averaged over 50 episodes for the UAV network with 3 relays against a greedy jammer.

the UAV communication performance and reduce the relay energy consumption compared with the benchmark scheme. For instance, the REAR approach decreases the BER by an order of magnitude and reduces energy consumption by 17.3% compared with OPAJ after 3500 time slots. The DREAR approach further decreases the BER by three orders of magnitude, i.e., to 1.5×10^{-7} and minimizes energy consumption by 22.8%, i.e., to 0.625 mJ.

7 Conclusion

In this paper, we propose a distributed framework for energy-efficient UAV relay networks, which aims at maintaining the transmission quality in the presence of a jammer. The relay UAVs can help relay messages cooperatively without sharing their real-time location and battery level with each other. The proposed RL-based approaches enable each relay UAV to select the optimal relay strategies derived based on the game theory without knowing the moving trajectory of other UAVs as well as the jammer. The performance bounds including the BER, and total network energy consumption and utility, are obtained by equilibrium analysis and are also verified via simulations in a static scenario. Simulation results show that the proposed approaches can improve the transmission quality of the UAV relay network in terms of the transmission BER while reducing the total energy consumption. For instance, the DREAR approach reduces energy consumption by 22.8% and decreases the BER by three orders of magnitude compared with the benchmark scheme.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 61971366 and 61731012) and the Fundamental Research Funds for the central universities (No. 20720200077).

References

[1] M. Mozaffari, W. Saad, M. Bennis, Y. H. Nam, and M. Debbah, A tutorial on UAVs for wireless networks: Applications, challenges, and open problems, *IEEE Commun. Surv. Tuts.*, vol. 21, no. 3, pp. 2334–2360, 2019.

[2] Y. Zeng, R. Zhang, and T. J. Lim, Wireless communications with unmanned aerial vehicles: Opportunities and challenges, *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, 2016.

[3] L. Gupta, R. Jain, and G. Vaszkun, Survey of important issues in UAV communication networks, *IEEE Commun. Surv. Tuts.*, vol. 18, no. 2, pp. 1123–1152, 2016.

[4] S. C. Lv, L. Xiao, Q. Hu, X. S. Wang, C. Z. Hu, and L. M. Sun, Anti-jamming power control game in unmanned aerial vehicle networks, in *Proc. IEEE Global Communications Conf.*, Singapore, 2017, pp. 1–6.

[5] H. Sedjelmaci, S. M. Senouci, and N. Ansari, A hierarchical detection and response system to enhance security against lethal cyber-attacks in UAV networks, *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 48, no. 9, pp. 1594–1606, 2018.

[6] S. D'Oro, E. Ekici, and S. Palazzo, Optimal power allocation and scheduling under jamming attacks, *IEEE/ACM Trans. Netw.*, vol. 25, no. 3, pp. 1310–1323, 2017.

[7] L. Y. Zhang, Z. Y. Guan, and T. Melodia, United against the enemy: Anti-jamming based on cross-layer cooperation in wireless networks, *IEEE Trans. Wirel. Commun.*, vol. 15, no. 8, pp. 5733–5747, 2016.

[8] P. Gu, C. Q. Hua, R. Khatoun, Y. Wu, and A. Serhrouchni, Cooperative antijamming relaying for control channel jamming in vehicular networks, *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7033–7046, 2018.

[9] L. Xiao, X. Z. Lu, D. J. Xu, Y. L. Tang, L. Wang, and W. H. Zhuang, UAV relay in VANETs against smart jamming with reinforcement learning, *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4087–4097, 2018.

[10] L. Xiao, D. H. Jiang, D. J. Xu, H. Z. Zhu, Y. Y. Zhang, and H. V. Poor, Two-dimensional antijamming mobile communication based on reinforcement learning, *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9499–9512, 2018.

[11] K. Anazawa, P. Li, T. Miyazaki, and S. Guo, Trajectory and data planning for mobile relay to enable efficient internet access after disasters, in *Proc. IEEE Global Communications Conf.*, San Diego, CA, USA, 2015, pp. 1–6.

[12] K. Li, W. Ni, X. Wang, R. P. Liu, S. S. Kanhere, and S. Jha, Energy-efficient cooperative relaying for unmanned aerial vehicles, *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1377–1386, 2016.

[13] Y. F. Chen, W. Feng, and G. Zheng, Optimum placement of UAV as relays, *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 248–251, 2018.

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

[15] L. Zhang, L. Huang, B. Li, M. Huang, J. W. Yin, and W. M. Bao, Fast-moving jamming suppression for UAV navigation: A minimum dispersion distortionless response beamforming approach, *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7815–7827, 2019.

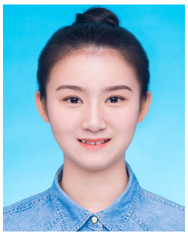
[16] Z. C. Xiao, B. Gao, S. C. Liu, and L. Xiao, Learning based power control for mmWave massive MIMO against jamming, in *Proc. IEEE Global Communications Conf.*, Abu Dhabi, United Arab Emirates, 2018, pp. 1–6.

[17] N. Gao, Z. J. Qin, X. J. Jing, Q. Ni, and S. Jin, Anti-intelligent UAV jamming strategy via deep Q-Networks, *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 569–581, 2020.

[18] L. Xiao, C. X. Xie, M. H. Min, and W. H. Zhuang, User-centric view of unmanned aerial vehicle transmission against smart attacks, *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3420–3430, 2018.

[19] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[20] H. van Hasselt, A. Guez, and D. Silver, Deep reinforcement learning with double Q-learning, in *Proc. 30th AAAI Conf. Artificial Intelligence*, Phoenix, Arizona, 2016, pp. 2094–2100.



Weihang Wang received the BS degree in electronic and information engineering from Hefei University of Technology, China in 2018. She is currently working toward the MS degree at the Department of Communication Engineering, Xiamen University, China. Her research interests include network security, privacy, and wireless communications.

Zefang Lv received the BS degree at the School of Mathematics from Shandong University in 2016 and the MS degree in the School of Mathematics and Physics from North China Electric Power University in 2020. She is currently pursuing the PhD degree at the Department of Information and Communication Engineering, Xiamen University, China. Her research interests include network security and wireless communications.

Xiaozhen Lu received the BS degree in communication engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2017. She is currently pursuing the PhD degree at the Department of Information and Communication Engineering, Xiamen University, China. Her research interests include network security and wireless communications.

Yi Zhang received the BS degree in software engineering from Xiamen University in 2014. He received the MS and PhD degrees in communication engineering from National Taiwan University in 2016 and 2020, respectively. He was with Quanzhou Institute of Equipment Manufacturing, Chinese Academy of Sciences from 2016 to 2017. He is currently an assistant professor at the Department of Information and Communication Engineering, Xiamen University. His research interests include mobile and wireless networking, fog/edge computing, and game theoretical models for communications networks.

Liang Xiao is currently a professor at the Department of Information and Communication Engineering, Xiamen University, China. She has served as an associate editor of *IEEE Trans. Information Forensics and Security* and guest editor of *IEEE Journal of Selected Topics in Signal Processing*. She is the recipient of the best paper award for 2016 INFOCOM Big Security WS and 2017 ICC. She received the BS degree in communication engineering from Nanjing University of Posts and Telecommunications in 2000, the MS degree in electrical engineering from Tsinghua University in 2003, and the PhD degree in electrical engineering from Rutgers University in 2009. She was a visiting professor with Princeton University, Virginia Tech, and University of Maryland, College Park.