Hybrid UAV-Enabled Secure Offloading via Deep Reinforcement Learning

Seonghoon Yoo\textsuperscript{a}, Graduate Student Member, IEEE, Seongah Jeong\textsuperscript{b}, Member, IEEE, and Joonhyuk Kang\textsuperscript{c}, Member, IEEE

Abstract—In this letter, we consider a secure offloading system consisting of a unmanned aerial vehicle (UAV)-mounted edge server, ground user equipments (UEs) and a malicious eavesdropper UAV. With the aim of maximizing secrecy sum-rate, we propose an adaptation of a helper UAV to switch the mode between jamming and relaying. We jointly optimize the helper UAV’s trajectory and mode and UEs’ offloading decision under energy budget constraints and operational limitations of nodes. The proposed algorithm is developed based on a deep deterministic policy gradient (DDPG)-based method, whose superior performances are verified via numerical results, as compared to other benchmark schemes.

Index Terms—Unmanned aerial vehicle (UAV), offloading, physical-layer security, deep reinforcement learning.

I. INTRODUCTION

Beyond fifth-generation (B5G) and sixth-generation (6G) systems have spawned many new applications, such as eXtend reality (XR) and autonomous driving [1], which require much higher complexity and lower latency than existing applications. However, user equipments (UEs) are in general resource-constrained, which makes it challenging to satisfy the stringent Quality of Services (QoS). To mitigate these issues, mobile edge computing (MEC) has been proposed, and in particular, unmanned aerial vehicle (UAV)-mounted edge server has been actively studied in that computing resources can be provisioned to the desired UEs effectively by leveraging advantages of high mobility and on-demand deployment [2], [3].

However, due to the open nature and the highly probable line-of-sight (LoS) links of aerial wireless channels, security and privacy are of utmost concern in UAV-enabled edge computing systems. Since the cryptography-based security schemes involving authentication and secure transmission to require the high-complexity process of key exchange and management has been challenged in constrained devices, e.g., small-size and battery-powered UAVs, physical-layer security (PLS) techniques have emerged as an alternative.

By leveraging the dynamic features of wireless communications such as fading, interference and noise, the PLS techniques can allow the legitimate UAV to decode the message successfully while preventing the eavesdropper from decoding it [4], [5], [6], [7], [8]. The authors in [4] propose an energy-efficient UAV-assisted secure offloading system in the presence of a ground eavesdropper. In [5], by considering the eavesdropper UAV more prone to intercept the data than the ground eavesdropper, a secure UAV communication system is investigated in the uplink scenario. For further security performance improvement and to relieve the burden of legitimate UAV, the helper UAV can be employed by freely controlling the beam pattern so as to attack the eavesdropper or relay confidential tasks. In [6], a UAV-enabled mobile relaying system is proposed for maximizing the secrecy rate, and [7] investigates the secure UAV systems with a cooperative jammer UAV to maximize energy efficiency. The traditional solutions of PLS systems with the helper UAV [6], [7] are mostly provided by the heuristic alternating algorithm based on convex optimization, which is time-consuming and may depend on the feasible initialization.

Deep reinforcement learning (DRL) has been recently used to address the issue of conventional optimization-based approaches. Among DRL techniques, deep deterministic policy gradient (DDPG) has been actively adopted for UAV systems since it is well-known to work well in systems with high dimensional action space [9]. The authors in [5], [8] propose the DDPG-based design of jointly optimizing UAV trajectory, power allocation or scheduling for secure communication. However, they focus on legitimate UAV design only or helper node in a single fixed mode, which can be further improved.

Motivated by limitations of studies about the DDPG-based secure offloading systems, we propose to consider a hybrid helper UAV to switch the mode between jamming and relaying for maximizing the secrecy sum-rate of offloading systems, where the ground UEs are associated with a legitimate UAV-mounted edge server under a malicious eavesdropper UAV. For the rapid convergence of the proposed algorithm, we...
optimize the helper UAV’s trajectory based on DDPG method with the weight initialization, while its mode selection and UEs’ offloading decision are jointly optimized with the relaxation method. Via simulations, the proposed method is verified to outperform other benchmark schemes with the fixed mode selection and/or trajectory optimization of helper UAV.

II. SYSTEM MODEL

We consider a hybrid UAV-enabled secure offloading system as shown in Fig. 1, in which one legitimate UAV is employed as an edge server for $U$ ground UEs with the presence of a malicious eavesdropper UAV, while a helper UAV with a half-duplex operation is adopted as a hybrid node to switch the role between relaying and jamming. For simplicity, all the nodes are assumed to be equipped with a single antenna, and we focus on the uplink scenario. By applying the centralized training process [10], a high altitude platform (HAP)-mounted server can be considered to distribute optimization results to each UAV node. For relay mode of helper UAV to convey the received data from UEs, the decode-and-forward (DF) method [11] is considered. In jamming mode, the helper UAV generates the artificial noise against the eavesdropper UAV. To this end, the information of the eavesdropper is assumed to be available, e.g., via radio frequency (RF) sensing, cameras or radar [12], which enables to collect data sets of the eavesdropper UAV for the learning process. Here, the legitimate UAV receives the offloaded data from both helper UAV and UEs in relay mode, or from only UEs in jamming mode, and computes the received offloaded data. In the following, we denote the legitimate UAV as $L$, the helper UAV as $H$, and the eavesdropper UAV as $E$.

For the system stability, we assume that both legitimate UAV and helper UAV hover at a fixed altitude $h_e > h$ to keep undiscovered. For tractability of analysis, the time horizon $N$ is divided into $T$ time slots (TSs) as in Fig. 2, each of which has $\Delta$ seconds small enough that the location of UAVs can be approximately considered fixed during each TS, satisfying $N = T\Delta$. Also, the orthogonal multiple access is assumed, for which $\Delta/U$ seconds of each TS is allocated to each UE. In TS $t$, the mode of the helper UAV is denoted as $k(t) = \{0, 1\}$, with $k(t) = 0$ indicating jamming mode, and $k(t) = 1$ indicating relay mode, which is subject to be optimized for each TS. We define the offloading decision variable for UE $u \in U$ as $z_u(t) = \{0, 1\}$ to be optimized, with $z_u(t) = 0$ indicating the local execution and $z_u(t) = 1$ indicating the offloading, where the set of $U$ UEs is represented as $U \triangleq \{1, 2, \ldots, U\}$. Moreover, the helper UAV flies at the velocity variable $v_H(t) = \{v_x(t), v_y(t)\}$ with the horizontal velocity $v_x(t)$ and the vertical velocity $v_y(t)$ in TS $t$, both of which are limited by the maximum velocity constraint $v_{\text{max}}$. Accordingly, its horizontal coordinates can be expressed as $x_H(t) = x_H(0) + \sum_{t'=1}^{t} v_x(t') \Delta$ and $y_H(t) = y_H(0) + \sum_{t'=1}^{t} v_y(t') \Delta$. The legitimate UAV, the eavesdropper UAV and the UE $u$ are assumed to be fixed at $(x_L, y_L)$, $(x_E, y_E)$ and $(x_u, y_u)$, respectively.

For the ground-to-air (G2A) channel, Rician fading is adopted [13], and the channel power gain between UE $u$ and UAV $i$ in TS $t$ can be written as

$$g_{u,i}(t) = \frac{\beta_0}{(h_{u,i})^2 + (D_{u,i})^2} \gamma_{\text{G2A}}(t),$$

(1)

for $\forall u \in U$ and $i \in \{H, L, E\}$, where $D_{u,i}$ represents the Euclidean horizontal distance between UE $u$ and UAV $i$ on the $xy$-plane as $D_{u,i} = \sqrt{(x_u - x_i)^2 + (y_u - y_i)^2}$; $h_{u,i}$ represents the altitude of UAV $i$, and becomes $h$ and $h_E$ when $i \in \{H, L\}$ and $i = E$, respectively; and $\beta_0$ denotes the received power at the reference distance $d_0 = 1\text{ m}$ of the G2A link. $\gamma_{\text{G2A}}(t)$ in (1) is a small scale fading component in the G2A environment with $K_{\text{G2A}}$ factor defined as

$$\gamma_{\text{G2A}}(t) = \sqrt{K_{\text{G2A}}/(K_{\text{G2A}} + 1)} + \sqrt{1/(K_{\text{G2A}} + 1)},$$

where $\gamma$ denotes the deterministic LOS component with $|\gamma| = 1$, and $\gamma$ is a circularly symmetric complex Gaussian distribution. Similarly, for the air-to-air (A2A) channel gain between the helper UAV and UAV $i \in \{L, E\}$, we define $g_{H,i}(t) = \gamma_{\text{A2A}}(t)$ as

$$g_{H,i}(t) = \frac{\beta_1}{(h_{H,i})^2 + (D_{H,i}(t))^2} \gamma_{\text{A2A}}(t),$$

(2)

for $i \in \{L, E\}$, where $D_{H,i}(t)$ is the horizontal distance between helper UAV and legitimate or eavesdropper UAV; $h_{H,i}$ is the altitude difference between two UAV nodes and becomes $0$ if $i = L$, or $h - h_E$ if $i = E$; $\beta_1$ denotes the reference channel power gain of the A2A link; and $\gamma_{\text{A2A}}(t)$ indicates the small scale fading component with $K_{\text{A2A}}$ factor [14]. Since the UE $u$ needs to be supported with the legitimate UAV and also can be additionally supported by the helper UAV in relay mode within their coverage area, we have

$$z_u(t)(k(t)D_{u,i} + (1 - k(t))D_{u,L}) \leq D_{\text{max}},$$

(3)

for $\forall u \in U$ and $i \in \{H, L\}$, where $D_{\text{max}}$ denotes the radius of their coverage area.

A. Communication Model

In this section, we provide the communication model required for the secure offloading procedure according to the different mode of helper UAV. As in Fig. 2, for relaying operation of helper UAV, the time division manner is considered due to its half-duplex limitation [11]. In particular, the time fraction $\Delta/U$ allocated to each UE is divided into two parts of equal size, the first of which is used for each UE to transmit the data to both legitimate UAV and helper UAV, while the remainder is adopted for the helper UAV to relay the received data to the legitimate UAV. In jamming mode, the entire $\Delta/U$ is consumed for transmission from each UE to the legitimate UAV, while the helper UAV generates the jamming signal. Depending on the helper UAV’s operation mode, the achievable data rates $R_{u}^d(k(t), v_H(t))$ and...
TABLE I
THE ACHIEVABLE RATES AT LEGITIMATE UAV AND EAVESDROPPER UAV ACCORDING TO THE DIFFERENT MODE OF HELPER UAV

| Mode                  | Relay mode                                                                 | Jamming mode                                                                 |
|-----------------------|-----------------------------------------------------------------------------|------------------------------------------------------------------------------|
| \( R^E_v(k(t), v_H(t)) \) | \[ \frac{1}{2} \min \left\{ \log_2 \left( 1 + \frac{p_H(k(t)) g_{H,L}(v_H(t)) + p_u g_{u,L}(t)}{\sigma^2} \right) \right\} \] | \[ \frac{1}{2} \log_2 \left( 1 + \frac{p_u g_{u,L}(t)}{p_H(k(t)) g_{H,L}(v_H(t)) + \sigma^2} \right) \] |
| \( R^C_v(k(t), v_H(t)) \) | \[ \frac{1}{2} \log_2 \left( 1 + \frac{p_H(k(t)) g_{H,L}(v_H(t)) + p_u g_{u,L}(t)}{\sigma^2} \right) \] | \[ \log_2 \left( 1 + \frac{p_u g_{u,L}(t)}{p_H(k(t)) g_{H,L}(v_H(t)) + \sigma^2} \right) \] |

**Proof**

\( R^E_v(k(t), v_H(t)) \) at legitimate UAV and eavesdropper UAV are calculated as in Table I. In Table I, \( p_u \) is the transmit power of UE \( u \), \( \sigma^2 \) is the noise power, and \( p_H(k(t)) \) is the transmit power of the helper UAV, where \( p_H(k(t)) = p_R \) in relay mode, otherwise \( p_H(k(t)) = p_J \). It is noted in Table I that the achievable data rate at legitimate UAV in relay mode is calculated as the minimum of data rates obtained in two time fractions, while the eavesdropper UAV can overhear the data via both UE- legitimate UAV link and helper UAV- legitimate UAV link, with 1/2 scaling in DF relaying system [15]. In jamming mode, the artificial interference from the helper UAV is factored into the data rate. Consequently, the secrecy sum-rate of secure offloading system can be written as

\[
C(k(t), z(t), v_H(t)) = \sum_{u \in \mathcal{U}} z_u(t) \left[ R^d_u(k(t), v_H(t)) - R^E_v(k(t), v_H(t)) \right] + \text{, (4)}
\]

where \( [x]^+ \triangleq \max(x, 0) \) and \( z(t) = \{ z_u(t) \}_{u \in \mathcal{U}} \).

**B. Computing Model**

For computing model, we define \( S_u(t) \) as the data size of the task and \( F_u(t) \) as the number of CPU cycles per bit. In the local execution, the task needs to be computed within \( \Delta \), and hence the CPU frequency \( f_u(t) \) of the UE \( u \) is determined as \( f_u(t) = \frac{S_u(t)}{T} \). When offloading, the total data received at legitimate UAV in the previous TS \( t-1 \) is assumed to be computed in TS \( t \), and the CPU frequency of the legitimate UAV, \( f_L(z(t)) \), is then calculated as \( f_L(z(t)) = \sum_{u \in \mathcal{U}} z_u(t-1) F_u(t-1) / \Delta \). By [16], the computation energy at UE or legitimate UAV is given by \( E^C_v(z(t)) = \kappa f^2 L \), where \( \kappa \) denotes the power consumption coefficient, and \( f_L(z(t)) \) if \( i = L \), respectively.

**III. PROPOSED DDPG-BASED METHOD**

In this letter, we aim to maximize the secrecy sum-rate by jointly optimizing the helper UAV’s mode \( k(t) \) and velocity \( v_H(t) \) and the UE’s offloading choice \( z(t) \) for all \( t \). To endow the formulation, we optimize the problem as

\[
\max_{k(t), z(t), v_H(t)} \quad C(k(t), z(t), v_H(t)) \tag{5a}
\]

s.t. \( k(t) = \{0,1\} \), \( z_u(t) = \{0,1\} \), \( \forall u \in \mathcal{U} \),

\[ -0.5 \leq z_H(v_H(t)) - z_H(v_H(t)) \leq 0.5 \tag{5b} \]

\[ z_u(t) k(t) D_{u,i} + (1-k(t)) D_{u,L} \leq D_{\max} \tag{5c} \],

\[ i \in \{H,L\}, \forall u \in \mathcal{U} \],

\[ z_u(t) p_u \left( k(t) \frac{\Delta L}{U} + (1-k(t)) \frac{\Delta U}{U} \right) \]

\[ + (1-z_u(t)) E^C_u(t) \big| \leq E_u, \forall u \in \mathcal{U} \tag{5e} \]

where (5b) is a binary variable constraint pertaining to the helper UAV’s mode and offloading decision; (5c) ensures that the helper UAV travels within a \( \lambda_{\text{max}} \)-side-length square; (5d) restricts legitimate UAV and helper UAV to hover within their coverage area; and (5e)-(5g) represent the energy budget constraints ofUEs, legitimate UAV and helper UAV, respectively. For helper UAV in (5g), the energy constraint \( E_H \) concerns for transmission, and flying, while regarding the other energy consumption as the additional term, where \( E^C_v(z(t), v_H(t)) = k(t) \sum_{u \in \mathcal{U}} z_u(t) p_R \Delta (2U) + (1-k(t)) p_J \Delta \) and \( E^C_v(v_H(t)) = 0.5 M \Delta (v_H(t))^2 + (v_H(t))^2 \) are the energy consumption for transmission and flying [2], respectively, with \( M \) being the mass of the UAV.

To find the optimal design of \[ \{ k(t), z(t), v_H(t) \} \] in (5), we propose the DDPG-based framework as in Algorithm 1.

In the Markov Decision Process (MDP), the agent has a state \( s_t \), and takes action \( a_t \) in TS \( t \). Given \( s_t \) and \( a_t \) as the agent proceeds with the various interactions in the environment, the agent obtains a reward \( r_t \) and the next state \( s_{t+1} \). The policy \( \pi \) is designed to maximize the accumulated reward \( R_t = \sum_{i=t}^T \gamma^{(i-t)} r_i \), where \( \gamma \in [0,1] \) is the discount factor.

The critic network learns the action-value function \( Q(s_t, a_t) = E_{a \sim \pi} [R_t | s_t, a_t] \) using Bellman’s equation in Q-learning, and proceeds to minimize the loss function \( L(\gamma) \), which is defined as

\[
L(\theta^Q) = \mathbb{E} \left[ (Q(s_{t+1}, a_{t+1} - \gamma Q(s_t, a_t | \theta^Q)) - y_t )^2 \right] \tag{6}
\]

where \( \theta^Q \) is the weight of the critic network and \( y_t = r_t + \gamma Q(s_{t+1}, a_{t+1} | \theta^Q) \). The actor network updates with the policy gradient method to maximize the expected reward \( J = E_{a \sim \pi} [R_t] \), and uses a policy function approximator, which follows

\[
\nabla_{\theta^a} J \approx E\left[ \nabla_a Q(s, a \theta^Q) \big| s=s_t, a=a_t(s_t) \right] \nabla_{\theta^a} \mu(s \theta^a)|_{s_t} \tag{7}
\]

where \( \theta^a \) is the weight of the actor network. It is noted that, the DDPG algorithm improves the update stability by using the target networks \( \theta' \) and \( \theta'' \) identical to those of the critic network and the actor network, and these target networks are updated by the soft update method.

In order to optimize the helper UAV’s trajectory based on DDPG method, we define the state, action and reward function in TS \( t \) as follows:

**State:** Let \( S \) denote the system state space as \( S = \{ s_t | s_t = (x_H(v_H(t)), y_H(v_H(t)), k(t), \{ D_H,i(v_H(t)) \}_{i \in \{L,E \}, t \in \{1,2,...,T \}} \) \}

**Action:** Let \( A \) denote the system action space as \( A = \{ a_t | a_t = \{ v_{x,t}(t), v_{y,t}(t) \}_{t \in \{1,2,...,T \}} \) \).
The Proposed DDPG-Based Method

Initialize: Replay buffer $B$, actor $\mu$, critic $Q$ and target network $\mu'$ and $Q'$ with $\theta$, $Q$, $\theta'$, $\theta$ and $\theta'$ $\leftarrow \theta$;

for Episode = 1, 2, ..., $e_{\text{max}}$ do

for $T$ in $D$ do

Set $\{z_{\text{u}}(t) = 0\} \forall t \in U$ and $C_0 = C_1 = 0$;

Calculate $s_f$ following “State” step;

Execute action $a_t = \mu(s_f; \theta_t) + \varepsilon$;

Calculate $C_t = C(k(t), z(t), \psi_H(t))$

end for

for $i \in \{0, 1\}$ do

$k(t) = i$ and obtain $z(t)$ by (8);

Calculate $C_t = C(k(t), z(t), \psi_H(t))$

end for

$k(t) = \arg \max_{i \in \{0, 1\}} C_i$ and obtain $r_t$ and $s_{t+1}$;

Store transition $(s_t, a_t, r_t, s_{t+1})$ in $B$;

Sample a random mini-batch of $K$ transitions;

Update critic, actor network by using (6) and (7);

Update target networks as $\theta_{Q^+} \leftarrow \theta_{Q^+} + (1 - \tau)\theta_{Q'}$ and $\theta_{\mu^+} \leftarrow \theta_{\mu^+} + (1 - \tau)\theta_{\mu}$;

end for

end for

Algorithm 1: The Proposed DDPG-Based Method

Input: Structures of actor, critic and target network. Output: $k(t), z(t), \psi_H(t)$.

Reward: We define $r_t = C_t - r_{\text{om}}$ as a reward function, where $C_t$ is the secrecy sum-rate according to mode $i$; and $r_{\text{om}}$ is the penalty value to avoid the case, when the helper UAV is out of the given map. When the helper UAV goes off the map, the helper UAV is set to be located at the previous location.

With the optimized helper UAV’s trajectory, we develop a relaxation method to jointly optimize the offloading decision $z(t)$ and the helper UAV’s operation mode $k(t)$ so as to decrease the training complexity. The offloading decision variable $z(t)$ is designed as

$$z(t) = \begin{cases} 1, & \text{if } R_0^q(k(t), \psi_H(t)) - R_0^q(k(t), \psi_H(t)) > \varepsilon, \\
0 & \text{otherwise}, \end{cases}$$

so that the secrecy sum-rate greater than the minimum limit $\varepsilon$ and (3) are both satisfied, after action $a_t$ is performed. According to the offloading decision $z(t)$, the helper UAV’s mode $k(t)$ is selected as $k(t) = \arg \max_{i \in \{0, 1\}} C_i$ and the reward $r_t$ is redefined as $r_t = \max_{i \in \{0, 1\}} C_i - r_{\text{om}}$. Based on the discussions above, the proposed algorithm is provided in Algorithm 1. To ensure the convergence and increase the convergence speed of Algorithm 1, the initial weights are set experimentally based on the previous step with the higher reward. For the helper UAV, action $a_t$ is generated by the actor network $\mu$, and a noise process $N$ is added for exploration. Then, we obtain the reward $r_t$ and the next state $s_{t+1}$, while storing the transition into its finite-sized buffer $B$.

IV. Simulation Results

In this section, we present the numerical results to evaluate the performance of the proposed Algorithm 1 compared to the other benchmark schemes. For simulations, we refer to the 3GPP NTN standard [17], and consider the parameter setting provided in Table II by following [16] for convergence and fine tuning. For the energy budget constraints, we set $E_u = 0.025J$, $E_L = 243J$ and $E_H = 3.9kJ$ with excluding hovering energy of UAVs, and set 1000 episodes in the training stage. The capacity of the replay buffer is 8000, and the mini-batch size is 70. We consider the noise process $N$ to follow $\mathcal{N}(0, 0.6)$ with the decaying rate of 0.999. The networks have three fully-connected hidden layers with $[300, 100, 100]$ neurons, which are trained at the learning rate of $10^{-4}$. The activation function is used as tanh function, and the network is updated using the AdamOptimizer. For references, we consider four types of benchmark schemes with the fixed mode and/or trajectory optimization of helper UAV; i) Re-LT scheme with the helper UAV in relay mode flying along a linear trajectory to reach the midpoint between the legitimate UAV and UEs at TS $T$ based on proposed offloading decision, ii) Ja-LT scheme with the helper UAV in jamming mode flying along a linear trajectory to reach the point, where the eavesdropper UAV exists at TS $T$ based on proposed offloading decision, iii) Re-OT scheme with the helper UAV in jamming mode flying along a linear trajectory to reach the point, where the eavesdropper UAV exists at TS $T$ based on proposed offloading decision, iv) Ja-OT scheme with the helper UAV in jamming mode, where its trajectory and offloading decision are optimized by Algorithm 1.

Fig. 3 shows the accumulated reward $R_t$ of the proposed algorithm as the number of training episodes. It is observed that the proposed method converges after 600 episodes, and the proposed method achieves the higher accumulated reward than both Re-OT and Ja-OT schemes by further optimizing the operation mode of the helper UAV. Secrecy performance is analyzed based on the dataset after convergence when the highest accumulated reward is obtained. In Fig. 4, the optimal trajectory obtained by the proposed method is illustrated for different UE deployments. In Fig. 4(a), we consider the case with a single UE cluster, where the UEs are randomly distributed around the legitimate UAV. In Re-OT scheme, the
of ground users, wherein the helper UAV can adaptively switch the roles between relaying and jamming based on its location. We jointly optimize the helper UAV’s mode selection and trajectory along with the users’ offloading decisions based on the DDPG method, whose superior performance is verified via simulation results compared to benchmark methods. Interesting open problems concern the generalization and the robustness of the optimization studied here to multiple cells with multiple helpers and eavesdroppers with imperfect channel state information.

V. CONCLUSION

In this letter, we have proposed a hybrid UAV-enabled secure offloading system to maximize the secrecy sum-rate of ground users, wherein the helper UAV can adaptively switch the roles between relaying and jamming based on its location. We jointly optimize the helper UAV’s mode selection and trajectory along with the users’ offloading decisions based on the DDPG method, whose superior performance is verified via simulation results compared to benchmark methods. Interesting open problems concern the generalization and the robustness of the optimization studied here to multiple cells with multiple helpers and eavesdroppers with imperfect channel state information.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.

[2] S. Jeong, O. Simeone, and J. Kang, “Mobile edge computing via a UAV-mounted cloudlet: Optimization of bit allocation and path planning,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2049–2063, Mar. 2018.

[3] Z. Yang, C. Pan, K. Wang, and M. Shikh-Bahaei, “Energy efficient resource allocation in UAV-enabled mobile edge computing networks,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4576–4589, Sep. 2019.

[4] T. Bai, J. Wang, Y. Ren, and L. Hanzo, “Energy-efficient computation offloading for secure UAV-edge-computing systems,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6074–6087, Jun. 2019.

[5] C. Wen, Y. Fang, and L. Qu, “Securing UAV communication based on multi-agent deep reinforcement learning in the presence of smart UAV eavesdropper,” in *Proc. IEEE Wireless Commun. Netw. Conf.* (WCNC), 2022, pp. 1164–1169.

[6] Q. Wang, Z. Chen, W. Mei, and J. Fang, “Improving physical layer security using UAV-enabled mobile relaying,” *IEEE Wireless Commun. Netw. Conf.* (WCNC), 2022, pp. 310–313, Jun. 2017.

[7] Y. Cai, Z. Wei, R. Li, D. W. K. Ng, and J. Yuan, “Joint trajectory and resource allocation design for energy-efficient secure UAV communication systems,” *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4536–4553, Jul. 2020.

[8] H. Kang, X. Chang, J. Mišić, V. B. Mišić, J. Fan, and J. Bai, “Improving dual-UAV aided ground-UAV bi-directional communication security: Joint UAV trajectory and transmit power optimization,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 10570–10583, Oct. 2022.

[9] C. Huang et al., “Multi-hop RIS-empowered Terahertz communications: A DRL-based hybrid beamforming design,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1663–1677, Jun. 2021.

[10] Y. Zhang, Z. Mou, F. Gao, J. Jiang, R. Ding, and Z. Han, “UAV-enabled secure communications by multi-agent deep reinforcement learning,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11599–11611, Oct. 2020.

[11] S. Yin, Z. Qu, and L. Li, “Uplink resource allocation in cellular networks with energy-constrained UAV relay,” in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Porto, Portugal, Jun. 2018, pp. 1–5.

[12] X. Sun, D. W. K. Ng, Z. Ding, Y. Xu, and Z. Zhong, “Physical layer security in UAV systems: Challenges and opportunities,” *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 40–47, Oct. 2019.

[13] D. W. Matolak and R. Sun, “Air–ground channel characterization for unmanned aircraft systems—Part III: The suburban and near-urban environments,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6607–6618, Aug. 2017.

[14] M. Al-Jarrah, A. Al-Dweik, E. Alsausa, Y. Iraqi, and M.-S. Alouini, “On the performance of IRS-assisted multi-layer UAV communications with imperfect phase compensation,” *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8531–8568, Dec. 2021.

[15] J. Mo, M. Tao, and Y. Liu, “Relay placement for physical layer security: A secure connection perspective,” *IEEE Commun. Lett.*, vol. 16, no. 6, pp. 878–881, Jun. 2012.

[16] L. Wang, K. Wang, C. Pan, W. Xu, N. Aslam, and L. Hanzo, “Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 1, pp. 73–84, Mar. 2021.

[17] X. Lin, S. Rommer, S. Ender, E. A. Yavuz, and R. S. Karlsson, “5G from space: An overview of 3GPP non-terrestrial networks,” *IEEE Commun. Stand. Mag.*, vol. 5, no. 4, pp. 147–153, Dec. 2021.