Deep Reinforcement Learning Based Resource Allocation for Secure RIS-aided UAV Communication

Iqbal, A, Al-Habashna, A, Wainer, G, Bouali, F, Boudreau, G & Wali, K

DOI 10.1109/VTC2023-Fall60731.2023.10333558

Publisher: IEEE

Deep Reinforcement Learning Based Resource Allocation for Secure RIS-aided UAV Communication

Amjad Iqbal
Department of Systems and Computer Engineering, Carleton University,
1125 Colonel By Dr., Ottawa, ON,
K1S 5B6, Canada
amjad.iqbal68a@gmail.com

Ala'a Al-Habashna
Department of Systems and Computer Engineering, Carleton University,
1125 Colonel By Dr., Ottawa, ON,
K1S 5B6, Canada
AlaaAlHabashna@cmail.carleton.ca

Gabriel Wainer
Department of Systems and Computer Engineering, Carleton University,
1125 Colonel By Dr., Ottawa, ON,
K1S 5B6, Canada
gwainer@sce.carleton.ca

Faouzi Bouali
Centre for Future Transport & Cities,
Coventry University Coventry CV1
5FB, United Kingdom,
ad6501@coventry.ac.uk

Gary Boudreau
Ericsson, Canada, 349 Terry Fox Dr.,
Kanata, ON,
K2K 2V6, Canada
gary.boudreau@ericsson.com

Khan Wali
Agricultural Biosystems Engineering
Group, Wageningen University &
Research, The Netherlands
khan.wali@wur.nl

Abstract— **We investigate the use of reconfigurable intelligent surfaces (RISs) in wireless networks to maximize the sum secrecy rate (i.e., the sum maximum rate that can be communicated under perfect secrecy). Specifically, we focus on a network that utilizes RIS-assisted unmanned aerial vehicles (UAVs) under imperfect channel state information (CSI). Our objective is to maximize the sum secrecy rate while dealing with the presence of multiple eavesdroppers. To achieve this, we jointly optimize the active (UAV) and passive (RIS) beamforming together with the UAV's trajectories. The formulated problem is non-convex due to the coupling of CSI with the maneuverability of the UAV. To overcome this challenge, we propose a policy-based deep reinforcement learning (DRL) approach that solves the non-convex optimization problem in a centralized fashion. Finally, simulation results show that our proposed approach significantly improves average sum secrecy rates over conventional approaches**.

Keywords— UAV, RIS, Eavesdropper, Secrecy rate, DRL

I. Introduction

Unmanned aerial vehicles (UAVs) are rapidly gaining popularity across numerous fields due to their cost-effectiveness and superior maneuverability. The high-altitude capability of UAVs provides a unique advantage over traditional wireless networks, enabling them to overcome common bottlenecks, such as building blockages, remote areas, and emergency services. UAV devices have been effectively integrated into various real-world applications, such as surveillance operations [1], geographical exploration [2], disaster response missions [3], and wireless communications [4]. The integration of UAVs is poised to revolutionize the global connectivity landscape, particularly in ensuring the widespread availability of fifth-generation (5G) and beyond networks. With their flexibility and low-cost production, UAVs have emerged as a significant change in the wireless communication industry. By providing wireless connectivity via flying base stations, UAV-assisted networks revolutionize wireless connectivity, considerably expanding network coverage and streamlining information transmission efficiency. In addition, the widespread adoption of UAVs in numerous fields has demonstrated their tremendous potential to revolutionize various industries. With ongoing technological advancements and continuous research, UAVs will play an increasingly important role in shaping the future of many fields, including wireless communication, disaster response, and environmental monitoring.

The emergence of reconfigurable intelligent surfaces (RISs) has paved the way for developing beyond 5G (B5G) and sixth-generation (6G) networks [5]. Utilizing a large number of intelligent reflective elements, RIS efficiently directs the received signal toward the desired destinations. The RIS controller enables dynamic surface adaptation to the propagation environment to fulfill various purposes, such as enhancing the arriving signal and mitigating the eavesdropper effect to ensure secure communication [6]. One of the key advantages of RIS technology is its low-cost hardware production and nearly passive nature, which has enabled its efficient deployment in various settings. Moreover, when combined with unmanned aerial vehicles (UAVs), RIS technology can generate line-of-sight (LOS) signals and improve signal directions, expand coverage areas, reduce the radio frequency chain, and promote energy-saving features [7]. Utilizing the reflective properties of RIS and the high altitude and maneuverability of UAVs can help to overcome various challenges and limitations of existing wireless communication systems, ultimately leading to improved signal quality, increased coverage area, and reduced costs. Overall, incorporating RIS technology in UAV communication holds tremendous potential for enhancing the performance and capabilities of wireless communication systems.

Recent research has explored the potential of combining RIS technology with UAVs to improve wireless communication systems. Although UAVs are able to establish strong connections with their users due to their high altitude, this advantage can be hindered by obstacles such as buildings. RIS can be strategically placed on top of buildings or high places to mitigate signal blockage to reflect the channel between the UAV and users. This approach offers the added benefit of producing fewer intermediate delays and delivering more up-to-date data compared to a mobile active relay. Additionally, RIS technology offers greater convenience in deployment and requires lower energy consumption, making it a promising solution for enhancing wireless communication systems.

By utilizing the distinctive strengths of both UAV and RIS, wireless communication network performance can be significantly elevated, resulting in enhanced receive signal strength and reduced interference. In [8], the authors proposed innovative solutions for UAV-assisted wireless communication systems that incorporate vector beamforming and RIS phase shift optimization algorithms to maximize the received signal on the ground. The UAV flight paths are combined with RIS (passive) beamforming to effectively maximize the network sum rate. At the same time, problems associated with a predetermined UAV trajectory and an optimal phase-shift matrix for RIS have been resolved by employing closed-form solutions and successive convex approximation (SCA) [9]. The work in [9] has been extended to incorporate RIS passive beamforming technology for ultra-reliable and low-latency communication (URLLC) [10]. Specifically, this study optimizes UAV position, RIS passive beamforming, and URLLC block length to minimize total URLLC decoding error rates. Furthermore, the work in [11]-[12] proposes a RIS-assisted UAV communication network to maximize energy efficiency (EE) performance in fixed environments. By exploring various optimization techniques and integrating UAV and RIS technologies, these studies demonstrate the enormous potential of RIS-assisted UAV communications for revolutionizing wireless communication systems. However, the works referred to in [8]-[12] assume perfect channel state information (CSI) between UAVs and legitimate users, which means that the model assumes the communication channel between devices is always perfectly known, and this assumption may not hold in real-world scenarios. Additionally, the works in [8]-[12] use a traditional model-based approach where the network environment is fixed or static, meaning that the network does not change over time. However, this approach may not be practical when users are mobile and the network state changes with each time step. Therefore, it is essential to consider more practical scenarios where the assumptions made in these previous works may not hold. This can be done by exploring more dynamic and adaptable approaches that can handle changes in the network environment and imperfect CSI. Doing so can make the resulting models more versatile and applicable in real-time scenarios, with significant practical benefits.

Recent advancements in machine learning, especially in deep reinforcement learning (DRL) algorithms, have led to the emergence of powerful optimization and decision-making techniques for wireless networks. DRL is highly effective in handling dynamic environments that require continuous actions because it trains an offline neural network to select the right actions in milliseconds or even instantly. Furthermore, DRL is particularly useful in optimizing RIS in wireless networks. For example, recent studies have investigated the efficiency of DRL algorithms for optimizing RIS phase shifts to achieve the optimum signal-to-noise ratio [13], as well as for adjusting UAV altitude and changing the RIS phase to maximize the sum rate [14]. However, it is worth noting that these techniques mostly assume static environment settings, which may not hold in dynamic scenarios. As a result, there is a need to explore more practical and adaptable approaches that can handle wireless networks' complex and dynamic nature under various environmental conditions. The deployment of optimization algorithms utilizing DRL for RIS-assisted UAV communications has been extensively studied in [15]-[16]. However, these works either optimize active UAV beamforming and passive RIS beamforming or UAV trajectory but do not consider joint optimization. To address these limitations, in this paper, we propose a joint optimization approach that considers the imperfect CSI and explicitly considers the maneuverability of the UAV (including beamforming and trajectory) and RIS beamforming to enhance the sum secrecy rate (i.e., the sum maximum rate that can be communicated under perfect secrecy) of the RIS-assisted UAV communication system. Furthermore, to solve the non-convex problem, we proposed a twin delay deep deterministic policy gradient (DDPG) algorithm to optimize the UAV trajectory and active (UAV) and passive (RIS) beamforming, with the goal of maximizing the sum secrecy rate. Specifically, the first DDPG is used to identify the beamforming policies for the UAV and RIS, while the second DDPG specifies the trajectory of the UAV. Compared to a single DDPG structure (baseline), this dual approach offers the advantage of being able to control the trajectory of the UAV, which leads to more secure communication. Finally, simulation results are presented to validate the effectiveness of the proposed method.

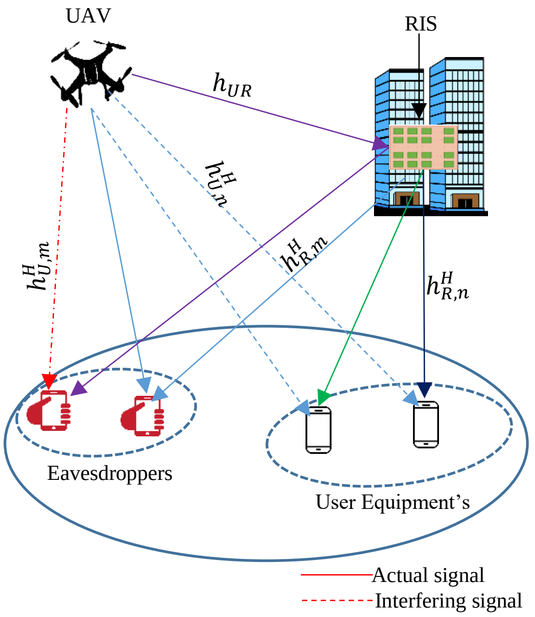

Fig.1. RIS-assisted UAV downlink wireless network

II. SYSTEM MODEL

In this work, we consider a multi-UAV $\mathcal{U} = \{1,2,\dots,U\}$ downlink wireless network supported by RIS, in which RIS facilitates the secure data transmission from the UAVs to N single-antenna user equipment (UE) in the presence of M single-antenna eavesdroppers, as shown in Fig. 1. The UAVs use an A-element uniform linear array (ULA), whereas a uniform planar array (UPA) with $A = a^2$ is used by the RIS (a is an integer). Furthermore, the sets of UEs and eavesdroppers are represented as $\mathbb{N} = \{1,2,\dots,N\}$ and $\mathcal{M} = \{1,2,\dots,M\}$, respectively. All the entities of any RIS and UAV are positioned in a three-dimensional (3D) Cartesian coordinate system, where the RIS is assigned the fixed coordinates at $w_R = (x_R, y_R, z_R)^T$. We consider that the

UAVs fly at a fixed altitude over K finite time slots, i.e., $T = K t_k$, where t_k is the time slot. At the k-th time slot, the coordinates of the UAVs and the coordinates of the UEs and the eavesdroppers can be denoted as $q[k] = (x_u[k], y_u[k], H_u)^T$ and $w_i[k] = (x_i[k], y_i[k], z_i[k])^T$, $\forall_i \in \mathbb{N} \cup \mathcal{M}$, respectively. The location information at the k-th time slot can be represented as $W = \{q[k]\} \cup \{w_i[k],\ \forall_i \in \mathbb{N} \cup \mathcal{M}\}$. The UAVs move within a certain area and cannot fly above a maximum height D_{max}. Additionally, it is assumed that UAVs are able to detect and avoid obstacles. Therefore, the mobility constraints of UAVs can be formulated as in eq. (1):

$$q[0] = (0,0,H_U) \tag{1a}$$

$$B \geq \max(x[k], y[k])\,, k = 1, \dots K \tag{1b}$$

$$D_{max} \geq \sqrt{\left||q[k+1] - q[k]|\right|^2}, k = 1, \dots K-1 \tag{1c}$$

Eq. (1a) represents the initial coordinates for the UAVs, whereas in Eq. (1b) and (1c), B and D_{max} indicate the UE moving boundaries and UAV's maximum distance at each time step, respectively. Let the channel gain from the u-th UAV to the RIS, from the u-th UAV to the m-th eavesdropper, from the u-th UAV to the n-th user, from the RIS to the n-th user, and from the RIS to the m-th eavesdropper be represented as $h_{U,R} \in \mathbb{C}^{M \times A}$, $H_{U,m} \in \mathbb{C}^{A \times 1}$, $h_{U,n} \in \mathbb{C}^{A \times 1}, h_{R,n} \in \mathbb{C}^{N \times 1}$, and $h_{R,m} \in \mathbb{C}^{M \times 1}$, respectively. These channels can be modeled according to the 3D saleh-valenzuela (SV) channel model [16]:

$$h_{U,i} = \sqrt{\frac{1}{L_{UN}}} \sum_{l=1}^{L_{UN}} g_{i,l}^{u} q_L(\varphi_{i,l}^{AoD}), \forall_i \in \mathbb{N} \cup \mathcal{M} \tag{2a}$$

$$h_{R,i} = \sqrt{\frac{1}{L_{RN}}} \sum_{l=1}^{L_{RN}} g_{i,l}^{r} q_M(\varphi_{i,l}^{AoD}, \vartheta_{i,l}^{AoD}), \forall_i \in \mathbb{N} \cup \mathcal{M} \tag{2b}$$

$$h_{U,R} = \sqrt{\frac{1}{L_{RN}}} \sum_{l=1}^{L_{RN}} g_{l}^{ur} q_M(\varphi_{l}^{AoA}, \vartheta_{l}^{AoA}) q_L(\varphi_{l}^{AoD})^H, \forall_i \in \mathbb{N} \cup \mathcal{M} \tag{2c}$$

where $g \in \{g_{i,l}^{u}, g_{i,l}^{r}, g_{l}^{u\,r}\}$ is the large-scale fading coefficient formulated as a complex Gaussian distribution, i.e., $\mathcal{CN}\left(0, 10^{\frac{PL}{10}}\right)$, such that $PL(dB) = -\mathcal{C}_0 - 10\alpha \log_{10}(D) - PL_s$, where $\mathcal{C}_0$, D and α indicate the reference distance of path-loss of one meter, link distance (in meters), and path-loss exponent, respectively, whereas, $PL_s \sim \mathcal{CN}(0, \sigma_s^2)$ represents the shadow fading coefficient. According to [17], q_L known is the steering vector of the ULA and can be expressed as:

$$q_L(\varphi) = \left[1, e^{j\frac{2\pi}{\lambda c} d \sin(\varphi)}, \dots, e^{j\frac{2\pi}{\lambda c} d (A-1) \sin(\varphi)}\right]^H \tag{3}$$

φ indicated the azimuth angle of departure (AoD) for $\vartheta_{i,l}^{AoD}$ and ϑ_{l}^{AoD}, and λc, d represent the carrier wavelength and inter-spacing for the antenna, respectively. Similarly, the UPA steering vector can be represented as $q_M(\varphi, \vartheta) = \left[1, \dots, e^{\frac{j2\pi}{\lambda c} d(i\ \sin(\varphi)\sin(\vartheta) + j\cos(\varphi)\sin(\vartheta))}, \dots\right]^H$, where $\varphi(\vartheta)$ is the angle of arrival (AoA), $\varphi_l^{AoA}(\vartheta_l^{AoA})$, and AoD, $\varphi_{i,l}^{AoD}(\vartheta_{i,l}^{AoD})$, of the azimuth (elevation), respectively, and $0 \leq i, j \leq a-1$. The LOS components between the trajectories of user's and UAV's for each link can be determined as $\varphi(\vartheta)_{l=1}^{AoA(AoD)}$, which enables the coupling of CSI and optimization variable Q. In the SV channel model, AoA/AoDs vary according to propagation paths. Hence, the assumption that LOS components only depend on the location of the UAV is not valid [17]. Thus, the LOS components for each link $\varphi(\vartheta)_l^{AoA(AoD)}, l \neq 1$ can be expressed further as [18]:

$$\varphi(\vartheta)_l^{AoA(AoD)} = \varphi(\vartheta)_{l=1}^{AoA(AoD)} + \Phi(\Lambda)_l^{AoA(AoD)}, l = 2, \dots, L \tag{4}$$

such that $\Phi(\Lambda)_l^{AoA(AoD)}$ is known as a spreading factor [19]. The channel from UAVs to users or the eavesdropper can be represented as $H_{C,i} = \mathrm{diag}(h_{R,i}^H) h_{UR}, \forall i \in \mathbb{N} \cup \mathcal{M}$. The RIS (passive) beamforming matrix is represented as $\Phi = \mathrm{diag}(\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, \dots \beta_A e^{j\theta A})$, where $\beta_a \in [0,1]$, $a = \{1, 2, \dots, A\}$, $\theta \in [0, 2\pi)$ represents the amplitude reflection and phase shift of the a-th RIS reflection elements, respectively. We assume a constant value of $\beta_a = 1$ for all elements so that the reflecting signal has maximum power. Let the channel gains from a UAV to all receivers be combined as:

$$H_C = \{h_{U,i}^H + \psi^H H_{C,i} | \forall_i \in \mathbb{N} \cup \mathcal{M}\} \tag{5}$$

ψ indicates the passive beamforming matrix for RIS that can be vectored as $\psi = \mathrm{vec}(\Phi)$. Finally, the signal received at the i-th user or eavesdropper from each UAV can be expressed as:

$$y_i = (h_{U,i}^H + \psi^H H_{C,i}) \boldsymbol{W} \boldsymbol{b} + o_i, \forall_i \in \mathbb{N} \cup \mathcal{M} \tag{6}$$

where $\boldsymbol{W} \in \mathbb{C}^{A \times N}$ and $\boldsymbol{b} \in \mathbb{C}^{N \times 1}$ with $E[|b_n|^2]$ specifies the beamforming matrix and transmitted symbol at the UAV, respectively. Therefore, the signal-to-interference-plus-noise-ratio (SINR) of the n-th UE at time slot t can be represented as:

$$SINR_n^u(t) = \frac{\left(|h_{U.n}^H + \psi^H H_{C,n}| w_n\right)^2}{\sum_{n' \in \mathbb{N} \backslash n} \left(|h_{U.n}^H + \psi^H H_{C,n}| w_{n'}\right)^2 + \sigma_n^2} \tag{7}$$

o_i denotes the background noise and is defined as $o_i \sim \mathcal{N}(0, \sigma_n), \forall_i \in \mathbb{N} \cup \mathcal{M}$. Thus, the total achievable data rate at the end of n-th UE can be expressed as:

$$\partial_n^u = \log_2(1 + SINR_n^u(t)) \tag{8}$$

Similarly, the SINR of the m-th eavesdropper signal to the n-th UE at time slot t can be expressed as:

$$SINR_{m,n}^e(t) = \frac{\left(|h_{U.m}^H + \psi^H H_{C,m}| w_n\right)^2}{\sum_{n' \in \mathbb{N} \backslash n} \left(|h_{U.m}^H + \psi^H H_{C,m}| w_{n'}\right)^2 + \sigma_m^2} \tag{9}$$

whereas the achievable rate from the m-th eavesdropper signal to the n-th user can be represented as:

$$\partial_{m,n}^e = \log_2\left(1 + SINR_{m,n}^e(t)\right) \tag{10}$$

According to [19], the individual secrecy rate from the UAV to the n-th users can be expressed as follows:

$$\partial_n^{\mathrm{sec}} = \left[\partial_n^u - \max_{\forall_m} \partial_{m,n}^e\right]^+ \tag{11}$$

where $[j]^+ = \max(0, j)$.

A. Problem Formulation

The objective of this work is to maximize the sum secrecy rate $\partial_n^{\sec}$ of the UEs' at each time slot by jointly optimizing the UAVs trajectory $\mathbf{q}$ and active and passive beamforming matrices ($\boldsymbol{\Phi}, \boldsymbol{W}$). The UAV will select the appropriate coordinate to direct the transmit signal to RIS. The RIS will then choose the phase-shift value according to the local information it receives from the environment at each step t to the UEs.

Thus, the optimization problem of all UEs subject to UAV's trajectory and beamforming matrix can be formulated as:

$$\text{(P1):} \quad \max_{\boldsymbol{q},\boldsymbol{\Phi},\boldsymbol{W}} \sum\nolimits_{n\in N} \partial_n^{\sec} \tag{11}$$

$$s.t. \qquad (1) \tag{11a}$$

$$P_r\{R_n^{sec} \geq R_n^{sec,th}\} \geq 1 - \rho_n, \forall\, n \in N \tag{11b}$$

$$0 \leq \theta_m \leq 2\pi, \;\; m = 1, \ldots. M \tag{11c}$$

$$T_r(\boldsymbol{W}\boldsymbol{W}^{\boldsymbol{H}}) \leq P_{max} \tag{11d}$$

Constraint (11b) denotes the guarantee of QoS of n-th user with a probability of at least $1 - \rho_n$. Constraint (11c) indicates the RIS angle for each element and should be bounded between 0 and 2π , and (11d) represents the beamforming evaluations. It is hard to find an optimal global solution to the problem (P1) due to the non-convex nature of (11b), (11c), and (11d) and the CSI coupling at each time step. To solve the (P1), we propose a policy-based DRL-based framework.

III. PROPOSED POLICY-BASED SOLUTION

In this section, we proposed the policy-based twin DDPG algorithm to solve the problem (P1), which enables the agent to learn beamforming policies and UAV trajectory without having any prior background knowledge. Due to the high degree of coupling between CSI and the UAV trajectory Q, optimizing all variables at once is difficult, which may result in inadequate performance and convergence. Instead of using one agent as in the conventional DRL-based network, we construct two DDPG networks to address this issue. Specifically, the first DDPG stipulates the best policy for selecting the beamforming (active and passive), whereas the second DDPG obtains the policy for UAV trajectories. In the end, the reward function is shared by both networks. Before developing the twin DDPG, we defined our proposed framework's state space, action space, and reward function.

A. *Active and Passive Beamforming*

In the first DDPG agent, the CSI is taken into account when generating optimal beamforming policies. The generated beamforming is then used to generate an action that is fed into the environment. The state/action spaces and reward function can be formulated based on a Markov decision process (MDP) as follows:

a) State space (s_1): At each time step, the agent of DDPG 1 predicts the CSI from UAV to all UEs and eavesdroppers based on the received signal strengths.

b) Action space (a_1)*:* At each time step, the agent of DDPG 1 performs an action as active and passive beamforming, i.e., Φ, W. These actions can be used to evaluate the environment and decide the best course of action. The agent of DDPG 1 learns complex tasks in an automated fashion due to its ability to quickly learn from interactions with the environment. To address the complex-valued input, the beamforming values are separated into real and imaginary parts as $G = Re\{\Phi\} + Img\{\Phi\}$ and $\theta = Re\{W\} + Img\{W\}$.

Table.1. Algorithm 1: Twin DDPG-Based Framework

```
Input: CSI for DDPG1 and local information for DDPG2
Output: Maximize Average secrecy rate of combined network
Initialize the actor $\pi_1(.)$, critic $\aleph_1(.)$ and target actor $\pi_1(.)$, target
   critic network $\aleph_1(.)$ for DDPG1.
Initialize the actor $\pi_2'(.)$, critic $\aleph_2'(.)$ and target actor $\pi_2'(.)$, target
   critic network $\aleph_2'(.)$ for DDPG2.
Initialize the experience replay memory Ɗ
for Episode= $1,2, \ldots. \mathbb{N}^{eps}$ of DDPG2 do
  t=0
  Reset the UAV and all UE positions
   for Step $n = 1,2, \ldots N_{step}$, do
    Initialize the state for DDPG1 and DDPG2 , i.e., (CSI and local
    information)
    Select action for DDPG1 and DDPG2 with a Gaussian noise $g_a$ and
    variance $\wp_a$;  $a_1 = \pi_1(.) + g_a$, $a_2 = \pi_2(.) + \wp_a$
    Execute action $a_1$ and $a_2$ obtained at $s_1$ and $s_2$ from the
    environment and received a reward according to (12)
    Store the transition $[s_1, a_1, r_1, s_1']$ and $[s_2, a_2, r_2, s_2']$ into experience
    reply memory Ɗ
    Random sample $\mathcal{M}_B$ mini batch transitions $[s^i, a^i, r^i, s^{i+1}]$ from
    experience reply memory Ɗ, $i \in \{1,2\}$
    Update the target actor network for DDPG1 and DDPG 2
  end for
end for
```

c) Reward (r_1)*:* This work aims to maximize the sum secrecy rate defined in (11). This can be achieved once the agent receives the best optimal action values. We define the reward function as:

$$r_t = \tanh\left(\sum_{n=1}^{N} \partial_n^{\sec} - c_1\tau_m - c_2\tau_r - c_3\tau_g\right) \tag{12}$$

where tanh denotes the hyperbolic tangent function.

In this case, τ_m, τ_r and τ_g represent penalties imposed when constraints (11b), (11c), and (11d) are not met. To balance these penalties and the sum secrecy rate, the coefficients $c_i, i = \{1,2,3\}$ are used.

B. UAV Trajectory:

In addition to the first DDPG agent, the second DDPG agent is utilized to determine the UAV's optimal trajectory, $\mathbf{q}$, based on the local information. The UAV trajectory can be optimized as an MDP by taking into consideration the following states, actions, and rewards:

a) State space (s_2): A large amount of CSI is coupled with the trajectory of the UAV. Therefore, the agent of the second DDPG decouples these variables by using only the location information as an input.

b) Action space (a_2)*:* The agent of the second DDPG generates the 3D Cartesian flying direction $d[t]$ for each time slot t. Based on the $d[t]$, the next UAV coordinate can be collected as $\mathbf{q}[t] = \boldsymbol{q}[t-1] + \boldsymbol{d}[t]$. The complete UAV trajectory after T time slot can be represented as $\boldsymbol{Q} = \{\boldsymbol{q}[0], \boldsymbol{q}[1], \ldots, \boldsymbol{q}[t]\}$. This model allows the UAV to successfully navigate to its destination without any external control or guidance.

c) Reward (r_2)*:* Both networks are trained for the same reward function as both networks aim to maximize the same utility function, i.e., the sum secrecy rate. The reward

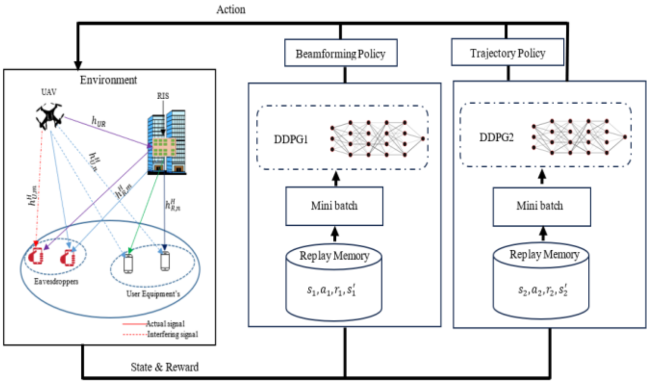

Fig. 2. Proposed System Model

function for both networks is the same as in (12). After training, the two networks output the best secrecy rate.

As the training process nears convergence, the first DDPG algorithm attains the optimal active (UAV) and passive (RIS) beamforming strategies, while the second DDPG algorithm excels in generating the best UAV trajectory. Through sharing reward functions and environmental information, these two DDPG networks can learn a favorable policy by cooperating with each other. Thus, the optimal beamforming matrix and the UAV trajectory can be yielded according to the proposed twin DDPG algorithm. The detailed pseudo-code for the proposed algorithm is presented in Table. 1. The proposed system model is depicted in Fig. 2.

IV. PERFORMANCE EVALUATION

We implement our proposed model using Python 3.6 with a Pytorch 1.10.0 framework. For deploying of proposed twin DDPG network, we used four fully connected (FC) hidden layers with neurons of [512,256,128, and 64] for DDPG1 and DDPG2. In order to train the actor and critic network of DDPG1 and DDPG2, we consider the learning rate is 10^{-3} with an Adam optimizer. Furthermore, we train the proposed network on 500 episodes, each with 100 steps, each corresponding to a one-time slot. The network is trained to predict the future demand for each time slot based on the previous time slots. In addition, the starting positions of UAV, RIS, UE, and Eavesdropper are placed at (25m, 0m, 50m), (0m, 50m, 12.5m) (25m,25m, 0m) and (47m,-4m, 0m), respectively [20]. The other system parameters are set as, D_{max} =0.25m, $t_k =$ 0.1s, T_d =1s, f_c =28GHz, C_0 =61dB, P_{max} =30dBm, σ_n =-114dBm, σ_m = 5dB, g_u =3.5, g_{ur}=2.2, $g_r =$ 2.8, Φ_l^{AoA} ∈{30,45,60}, Φ_l^{AoD} ∈{5,10,15,25}, Λ_l^{AoD}∈{1,3,5} , Λ_l^{AoA}∈{5,10,15} (degrees) and L=3 as defined in [20]. In order to make a better comparison, we compared our proposed algorithm with three benchmark algorithms, i.e., *Baseline, Random, and Greedy*. In the case of *Baseline*, a single DDPG agent is used to find the optimal policy for joint optimization. For the *Random*, the UAV randomly chooses the flying direction and distance for each time slot. For the *Greedy*, the UAV moves to the place where it is most likely to maximize the reward function defined in Eq. (12) at each time slot.

A. Average Sum Secrecy Rate

In this particular scenario, we analyze the impact of the number of RIS elements and the UAV altitude on the average sum secrecy rate. Fig. 3 showcases the average sum secrecy rate plotted against the number of RIS elements. It is evident from Fig.3, that the average sum secrecy rate exhibits a linear increase with a growing number of RIS elements. Compared to the other approaches, the proposed method achieves a significant increase (i.e., up to 35%) in the average sum secrecy rate depending on the number of RIS elements. This is because the proposed method is more capable of finding the best policy to direct the radio single strength toward the UEs and reduces the effect of eavesdroppers and penalties. The *Baseline* approach performs better than *Greedy* and *Random* because it can effectively learn from past experience and find the best possible RIS beamforming to direct the radio signal toward the UEs. In contrast, *Greedy* gives reasonably higher performance than *Random* because it controls where the UAV flies to achieve the best average sum secrecy rate. The flexibility and varied locations of the UAV play a significant role in enhancing the overall network performance.

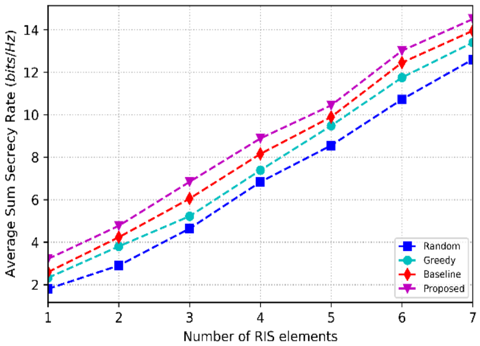

Fig. 3. Average Sum Secrecy rate vs Number of RIS Elements

In Fig. 4, we assess the impact of the UAV altitude on the average sum secrecy rate. It is evident that the twin DDPG approach outperforms the other three methods. Furthermore, the observed behavior demonstrates that increasing the altitude of the UAV leads to an improvement in the average sum secrecy rate. The maximum average sum secrecy rate can be achieved when the UAV flies at 20 meters. Beyond that point, the average sum secrecy rate starts to decline with increasing altitude. This observation can be attributed to two factors. Firstly, elevating the altitude within the range of 10 to 20 meters reduces interference between UAVs and UEs that utilize the same frequency spectrum**.** Additionally, it ensures that the air-to-ground channel's path loss is low enough to keep a strong-received signal. Secondly, when the altitude exceeds 20 meters, there is a significant decrease in the channel gain between UAVs and UEs. This reduction severely impacts the sum-rate of UEs, subsequently lowering the average sum secrecy rate. Consequently, this diminishes the quality-of-service (QoS) experienced by UEs, thereby reducing the usability of UAV networks. Moreover, higher altitudes also result in increased power consumption by UAVs, which affects their endurance and service coverage area.

B. UAV Trajectory

In Fig. 5, we depict the process of exploring the agent of DDPG2 to find the optimal trajectory for a UAV in a highly dynamic scenario considering two UEs. UAVs normally approach RISs while moving away from eavesdroppers in the beginning. With increasing distances between RIS and UEs, the UAV follows the UEs and moves toward the midpoint between them. This intriguing phenomenon can be ascribed

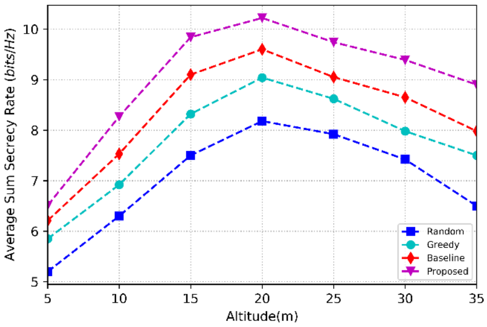

Fig. 4. Average Sum Secrecy Rate vs. the UAV altitude

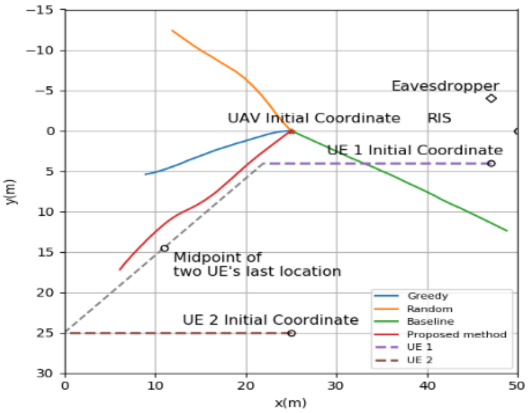

Fig. 5. UAV Trajectory by using Twin DDPG

to the gradual attenuation of the cascaded UAV-RIS-UE links as the distance between them progressively increases. Consequently, the direct links become dominant for transmission, and the UAV strives to serve the two UEs as equitably as possible. Moreover, our proposed method demonstrates its capability to adapt to environmental variations compared to other approaches when considering two UEs. This implies that the proposed method effectively positions the UAV in close proximity to both the RIS and the UE's. Enlightened by these results, it becomes evident that the twin DDPG approach enables the UAV to adjust flexibly to dynamic environments, leading to improved system performance through jointly optimizing the UAV trajectory and beamforming.

V. CONCLUSION

In this paper, we investigate the joint optimization of the active (UAV) and passive (RIS) beamforming and the UAV trajectory in a UAV communication system assisted by RIS. The objective is to maximize the average sum secrecy rate for all UEs served by UAV communications subject to imperfect channel state information (CSI). To address this challenge, we propose a twin DDPG approach. Through extensive simulations, we demonstrate the effectiveness of our proposed method by achieving superior performance compared to random, greedy, and single-DDPG baseline approaches. Our results highlight the benefits of jointly optimizing UAV trajectory and active (UAV) and passive (RIS) beamforming for improved system performance. In the future, we will extend this work to maximize the sum of secrecy energy efficiency in a highly dynamic scenario.

ACKNOWLEDGMENT

This work is funded by Ericsson Canada and the Natural Sciences and Engineering Research Council of Canada (NSERC)

REFERENCES

[1] S. Shakoor, Z. Kaleem, M. I. Baig, O. Chughtai, T. Q. Duong, and L. D. Nguyen, "Role of UAVs in public safety communications: Energy efficiency perspective," *IEEE Access*, vol. 7, pp. 140665–140679, 2019.

[2] A. Vacca, H. Onishi, and F. Cuccu, "Drones: Military weapons, surveillance or mapping tools for environmental monitoring? Advantages and challenges. A legal framework is required," *Transp. Res. Procedia*, vol. 25, pp. 51–62, 2017.

[3] T. Q. Duong, L. D. Nguyen, H. D. Tuan, and L. Hanzo, "Learning-aided real-time performance optimisation of cognitive UAV-assisted disaster communication," *2019 IEEE Glob. Commun. Conf. GLOBECOM 2019 - Proc.*, pp. 1–6, 2019.

[4] K. K. Nguyen, N. A. Vien, L. D. Nguyen, M. T. Le, L. Hanzo, and T. Q. Duong, "Real-Time Energy Harvesting Aided Scheduling in UAV-Assisted D2D Networks Relying on Deep Reinforcement Learning," *IEEE Access*, vol. 9, pp. 3638–3648, 2021.

[5] Y. C. Liang *et al.*, "Reconfigurable intelligent surfaces for smart wireless environments: channel estimation, system design and applications in 6G networks," *Sci. China Inf. Sci.*, vol. 64, no. 10, pp. 1–21, 2021.

[6] H. Yang *et al.*, "Intelligent Reflecting Surface Assisted Anti-Jamming Communications: A Fast Reinforcement Learning Approach," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 3, pp. 1963–1974, 2021.

[7] Z. Wei *et al.*, "Sum-Rate Maximization for IRS-Assisted UAV OFDMA Communication Systems," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 4, pp. 2530–2550, 2021.

[8] L. Ge, P. Dong, H. Zhang, J. B. Wang, and X. You, "Joint beamforming and trajectory optimization for intelligent reflecting surfaces-assisted UAV communications," *IEEE Access*, vol. 8, pp. 78702–78712, 2020.

[9] S. Li, B. Duo, X. Yuan, Y. C. Liang, and M. DI Renzo, "Reconfigurable Intelligent Surface Assisted UAV Communication: Joint Trajectory Design and Passive Beamforming," *IEEE Wirel. Commun. Lett.*, vol. 9, no. 5, pp. 716–720, 2020.

[10] A. Ranjha and G. Kaddoum, "URLLC Facilitated by Mobile UAV Relay and RIS: A Joint Design of Passive Beamforming, Blocklength, and UAV Positioning," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4618–4627, 2021.

[11] M. DIamanti, M. Tsampazi, E. E. Tsiropoulou, and S. Papavassiliou, "Energy Efficient Multi-User Communications Aided by Reconfigurable Intelligent Surfaces and UAVs," *Proc. - 2021 IEEE Int. Conf. Smart Comput. SMARTCOMP 2021*, pp. 371–376, 2021.

[12] D. Ma, M. Ding, and M. Hassan, "Enhancing Cellular Communications for UAVs via Intelligent Reflective Surface," *IEEE Wirel. Commun. Netw. Conf. WCNC*, vol. 2020-May, 2020.

[13] K. Feng, Q. Wang, X. Li, and C. K. Wen, "Deep Reinforcement Learning Based Intelligent Reflecting Surface Optimization for MISO Communication Systems," *IEEE Wirel. Commun. Lett.*, vol. 9, no. 5, pp. 745–749, 2020..

[14] C. Huang, R. Mo, and C. Yuen, "Reconfigurable Intelligent Surface Assisted Multiuser MISO Systems Exploiting Deep Reinforcement Learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, 2020.

[15] B. Sheen, J. Yang, X. Feng, and M. M. U. Chowdhury, "A Deep Learning Based Modeling of Reconfigurable Intelligent Surface Assisted Wireless Communications for Phase Shift Configuration," *IEEE Open J. Commun. Soc.*, vol. 2, no. February, pp. 262–272, 2021.

[16] M. Samir, M. Elhattab, C. Assi, S. Sharafeddine, and A. Ghrayeb, "Optimizing Age of Information through Aerial Reconfigurable Intelligent Surfaces: A Deep Reinforcement Learning Approach," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3978–3983, 2021.

[17] B. Liao and S. C. Chan, "Adaptive beamforming for uniform linear arrays with unknown mutual coupling," *IEEE Antennas Wirel. Propag. Lett.*, vol. 11, pp. 464–467, 2012.

[18] G. Zhou, C. Pan, H. Ren, K. Wang, M. Elkashlan, and M. Di Renzo, "Stochastic Learning-Based Robust Beamforming Design for RIS-Aided Millimeter-Wave Systems in the Presence of Random Blockages," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 1057–1061, 2021.

[19] Y. Zhu, G. Zheng, and M. Fitch, "Secrecy rate analysis of UAV-enabled mmWave networks using Matérn hardcore point processes," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1397–1409, 2018.

[20] Tham, Mau-luen; Yi Jie, Wong; Iqbal, Amjad; Ramli, Bin Nordin; Zhu, Yongxu; Dagiuklas, "Deep Reinforcement Learning for Secrecy Energy- Efficient UAV Communication with Reconfigurable Intelligent Surface," *IEEE Wirel. Commun. Netw. WCNC 2003*, pp. 1–6, 2023.