Learning-based Robust and Secure Transmission for Reconfigurable Intelligent Surface Aided Milimeter Wave UAV Communications

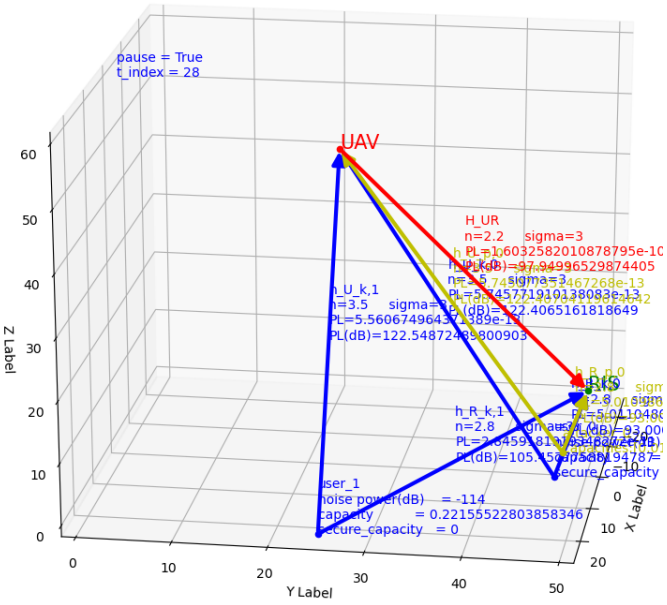

Fig. 1. RIS-aided Milimeter Wave UAV Communications.

***Abstract*—The abstract goes here.**

***Index Terms*—IEEE, IEEEtran, journal, LaTeX, paper, template.**

I. Introduction

THIS demo file is intended to serve as a "starter file" for IEEE journal papers produced under LaTeX using IEEEtran.cls version 1.8b and later. I wish you the best of success.

mds
August 26, 2015

A. Subsection Heading Here

Subsection text here.

1) Subsubsection Heading Here: Subsubsection text here.

II. System Model and Problem Formulation

A. System Model

In this letter, we consider an RIS-aided milimeter wave UAV secure transmission system where a RIS is exploited to assist the secure downlinks from the UAV to K single-antenna legitimate users in the presence of P single-antenna eavesdroppers. Specifically, the UAV is equipped with an A-element uniform linear array (ULA), and the RIS is a uniform planar array (UPA) with $M = m^2$ passive reflecting elements (m should be an integer). The set of the legitimate users and the eavesdroppers are denoted by $\mathcal{K}=\{1,2,...,K\}$,$\mathcal{P}=\{1,2,...,P\}$, respectively. As shown in Fig.1, all entities are placed in the three dimensional (3D) Cartesian coordinate system. Let $\mathbf{w}_k=[x_k,y_k,0]^T, \forall k\in\mathcal{K}$ and $\mathbf{w}_p=[x_p,y_p,0]^T, \forall p\in\mathcal{P}$ denote the legitimate users' and the eavesdroppers' coordinates, respectively. The RIS is fixed at $\mathbf{w}_R=[x_R,y_R,z_R]^T$. In addition, assume that the UAV flies at a fixed altitude in a finite time span which is divided into N time slots, i.e., $T=N\delta_t$, where δ_t is the time slot. Then the coordinate of the UAV at the n-th time slot is denoted by $\mathbf{q}[n]=[x[n],y[n],H_U]^T, n\in\mathcal{N}=\{1,2,...,N\}$, subject to the following mobility constraints:

$$\|\boldsymbol{q}[n+1]-\boldsymbol{q}[n]\|^2 \le D^2, n=1,\dots,N-1, \tag{1a}$$

$$|x[n]|,|y[n]| \le B, n=1,\dots,N-1, \tag{1b}$$

$$\boldsymbol{q}[0] \equiv [0,0,H_U], n=1,\dots,N-1. \tag{1c}$$

Let $\boldsymbol{h}_{U,k}\in\mathbb{C}^{A\times 1}$, $\boldsymbol{h}_{U,p}\in\mathbb{C}^{A\times 1}$, $\boldsymbol{h}_{R,k}\in\mathbb{C}^{M\times 1}$, $\boldsymbol{h}_{R,p}\in\mathbb{C}^{M\times 1}$, $\boldsymbol{H}_{UR}\in\mathbb{C}^{M\times A}$ be the channel gains of the UAV to k-th user, UAV to p-th eavesdropper, RIS to k-th user, RIS to p-th eavesdropper, UAV to RIS links, respectively. All the channels are modeled as milimeter wave channels [1] [2] as

$$\boldsymbol{h}_{U,i}=\sqrt{\frac{1}{L_{UK}}}\sum_{l=1}^{L_{UK}} g_{i,l}^{u}\mathbf{a}_L\left(\theta_{i,l}^{AoD}\right), \forall i\in\mathcal{K}\cup\mathcal{P}, \tag{2a}$$

$$\boldsymbol{h}_{R,i}=\sqrt{\frac{1}{L_{RK}}}\sum_{l=1}^{L_{RK}} g_{i,l}^{r}\mathbf{a}_P\left(\theta_{i,l}^{AoD},\phi_{i,l}^{AoD}\right), \forall i\in\mathcal{K}\cup\mathcal{P}, \tag{2b}$$

$$\boldsymbol{h}_{UR}=\sqrt{\frac{1}{L_{RK}}}\sum_{l=1}^{L_{RK}} g_{l}^{ur}\mathbf{a}_P\left(\theta_{l}^{AoA},\phi_{l}^{AoA}\right)\mathbf{a}_L\left(\theta_{l}^{AoD}\right)^H. \tag{2c}$$

In (2), the large-scale fading coefficients defined by $g\in\{g_{i,l}^u, g_{i,l}^r, g_l^{ur}\}$ follow a complex Gaussian distribution as $\mathcal{CN}(0,10^{\frac{PL}{10}})$, where PL(dB)$=-\mathrm{C}_0-10\alpha\log_{10}(\mathrm{D})-\mathrm{PL}_s$, C_0=61 dB is the path loss at a reference distance of one meter, D (meters) is the link distance, α denotes the path-loss exponent, and $\mathrm{PL}_s\sim\mathcal{CN}(0,\sigma_s^2)$ is the shadow fading component. The steering vector of the ULA is denoted by $\mathbf{a}_L(\theta)=\left[1,e^{j\frac{2\pi}{\lambda_c}d\sin(\theta)},\dots,e^{j\frac{2\pi}{\lambda_c}d(N-1)\sin(\theta)}\right]^{\mathrm{H}}$ [3], where θ stands for the azimuth angle-of-departure(AoD) $\theta_{i,l}^{AoD}$ and θ_l^{AoD}, d is the antenna inter-spacing, and λ_c is the carrier wavelength.

The steering vector of the UPA is denoted by $\mathbf{a}_P(\theta,\phi) = \left[1,\ldots,e^{j\frac{2\pi}{\lambda c}d(p\sin(\theta)\sin(\phi)+q\cos(\theta)\sin(\phi))},\ldots\right]^{\mathrm{H}}$ [3],where $0 \leq p,q \leq m-1$, $\theta(\phi)$ is the azimuth(elevation) AoD $\theta_{i,l}^{AoD}(\phi_{i,l}^{AoD})$ and the angle-of-arrival (AoA) $\theta_l^{AoA}(\phi_l^{AoA})$.

The cascaded channel from the UAV to the i-th user or eavesdropper can be written as $\boldsymbol{H}_{C,i} = \mathrm{diag}(\boldsymbol{h}_{R,i}^H)\boldsymbol{h}_{UR}, \forall i \in \mathcal{K}\cup\mathcal{P}$. The passive beamforming matrix of the RIS [4] is defined as $\boldsymbol{\Theta} = \mathrm{diag}\left(\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, ..., \beta_M e^{j\theta_M}\right)$, where $\theta_m \in [0, 2\pi)$ and $\beta_m \in [0,1]$ represent the phase shift and amplitude reflection coefficient of the m-th RIS reflection element, respectively. For feasibility, the amplitude reflection coefficient subjects to unit-modulus constraints, i.e., β_m=1. Let $\boldsymbol{\Psi}$=vec$(\boldsymbol{\Theta})^{\mathbf{T}}$ denote the vectorized passive beamforming matrix. Then, the received signal at the i-th user or eavesdropper from the UAV can be formulated as

$$y_i = (\boldsymbol{h}_{U,i}^H + \boldsymbol{\Psi}^H \boldsymbol{H}_{C,i})\mathbf{G}\boldsymbol{s} + n_i, \forall i \in \mathcal{K}\cup\mathcal{P}, \tag{3}$$

where s_k with $E[|s_k|^2] = 1$ and $\mathbf{G} \in \mathbb{C}^{\mathrm{A}\times\mathrm{K}}$ represent the transmitted symbol and the beamforming matrix at the UAV, respectively, and it is assumed that $n_i \sim \mathcal{N}(0,\sigma_n), \forall i \in \mathcal{K}\cup\mathcal{P}$. Let $\boldsymbol{g}_k$ be the k-th column of the beamforming matrix $\mathbf{G}$. Then, the the achievable unsecured rate of the k-th user is given by

$$R_k^u = \log_2\left(1 + \frac{|(\boldsymbol{h}_{U,k}^H + \Psi^H \boldsymbol{H}_{C,k})\boldsymbol{g}_k|^2}{\sum_{k'\in\mathcal{K}\setminus k}|\boldsymbol{h}_{U,k}^H + \Psi^H \boldsymbol{H}_{C,k})\boldsymbol{g}_{k'}|^2 + n_k^2}\right). \tag{4}$$

If the p-th eavesdropper aims to eavesdrop the signal of the k-th user, its achievable rate can be denoted by

$$R_{p,k}^e = \log_2\left(1 + \frac{|(\boldsymbol{h}_{U,p}^H + \Psi^H \boldsymbol{H}_{C,p})\boldsymbol{g}_k|^2}{\sum_{k'\in\mathcal{K}\setminus k}|\boldsymbol{h}_{U,p}^H + \Psi^H \boldsymbol{H}_{C,p})\boldsymbol{g}_{k'}|^2 + n_p^2}\right). \tag{5}$$

The achievable individual secrecy rate from the UAV to the k-th user [5] can be expressed by

$$R_k^{\mathrm{sec}} = \left[R_k^{\mathrm{u}} - \max_{\forall p} R_{p,k}^{\mathrm{e}}\right]^+ \tag{6}$$

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

It is worth noting that the outdated CSI will lead to substantial performance loss in practical systems. According to [6], the outdated CSI can be expressed as statistical CSI error model. Furthermore, let T_d be the delay between the outdated CSI and the real-time CSI. The relation between the outdated channel vector $\boldsymbol{h}(t)$ and the real-time channel vector $\boldsymbol{h}(t+T_d)$ can be expressed as [7]

$$\boldsymbol{h}\left(t+T_{\mathrm{d}}\right) = \rho\boldsymbol{h}(t) + \sqrt{1-\rho^2}\boldsymbol{e}, \tag{7}$$

where $\boldsymbol{e}$ is independent identically distributed with $\boldsymbol{h}\left(t+T_{\mathrm{d}}\right)$ and $\boldsymbol{h}(t)$, ρ is the autocorrelation function of the channel gain $\boldsymbol{h}(t)$, given by the zeroth-order Bessel function of the first kind as

$$\rho = J_0\left(2\pi f_D T_{\mathrm{d}}\right), \tag{8}$$

where f_D is the Doppler spread which is expressed as $f_D = vf_c/c$, where v, f_c, c represent the velocity of the transceivers, the carrier frequency and the speed of light, respectively.

Then, the actual channel coefficients can be rewritten as

$$\begin{aligned}
\boldsymbol{h}_{\mathrm{U},i} &= \rho\tilde{\boldsymbol{h}}_{\mathrm{U},i} + \Delta\boldsymbol{h}_{\mathrm{U},i}, \forall k \in \mathcal{K}\cup\mathcal{P},\\
\boldsymbol{h}_{\mathrm{R},i} &= \rho\tilde{\boldsymbol{h}}_{\mathrm{R},i} + \Delta\boldsymbol{h}_{\mathrm{R},i}, \forall k \in \mathcal{K}\cup\mathcal{P},\\
\boldsymbol{h}_{\mathrm{UR}} &= \rho\tilde{\boldsymbol{h}}_{\mathrm{UR}} + \Delta\boldsymbol{h}_{\mathrm{UR}}.
\end{aligned} \tag{9}$$

Note that the system only has access to the estimated CSI $\tilde{\boldsymbol{h}} \in \{\tilde{\boldsymbol{h}}_{\mathrm{U},i}, \tilde{\boldsymbol{h}}_{\mathrm{R},i}, \tilde{\boldsymbol{h}}_{\mathrm{UR}}\}$, which are outdated, to generate active and passive beamforming and UAV trajectory. And the actual CSI $\boldsymbol{h} \in \{\boldsymbol{h}_{\mathrm{U},i}, \boldsymbol{h}_{\mathrm{R},i}, \boldsymbol{h}_{\mathrm{UR}}\}$ is employed to calculate achievable secrecy rate of each user which has been expressed in (4), (5), (6).

B. Problem Formulation

In this letter, we aim to maximize the sum secrecy rate $\sum_{k=1}^K R_k^{\mathrm{sec}}$ by jointly optimizing the UAV's trajectory $\mathbf{Q} \triangleq \{\mathbf{q}[\mathbf{n}], \mathbf{n} \in \mathcal{N}\}$ and the active (passive) beamforming matrix $\mathbf{G}(\boldsymbol{\Theta})$. The optimization problem is formulated as

$$\max_{\mathbf{Q},\mathbf{G},\boldsymbol{\Theta}} \quad \sum_{k\in\mathcal{K}} R_k^{\mathrm{sec}} \tag{10a}$$

$$s.t. \quad (1), \tag{10b}$$

$$\Pr\left\{R_k^{\mathrm{sec}} \geq R_k^{\mathrm{sec,th}}\right\} \geq 1-\rho_k, \forall k \in \mathcal{K}, \tag{10c}$$

$$\mathrm{Tr}\left(\mathbf{G}\mathbf{G}^{\mathbf{H}}\right) \leq P_{\max}, \tag{10d}$$

$$\theta_m \in [0, 2\pi), \forall m \in \mathcal{M}, \tag{10e}$$

where the rate outage constraint in (10c) guarantees that the probability that each legitimate user can successfully decode its message at a data rate of $R_k^{\mathrm{sec,th}}$ is no less than $1-\rho_k$. (10b) models the UAV mobility constraint. (10d) restricts the transmission power at UAV side does not exceed the maximum value. (10e) constrains the reflection coefficients of the RIS reflecting elements. Obviously, problem (10), due to the non-convex objective function and constraints, is a non-convex problem that is rather challenging to solve. To tackle this challenging problem, the sum rate optimization problem is formulated in the context of DRL based method to obtain the feasible $\mathbf{G}$, $\boldsymbol{\Theta}$ and $\mathbf{Q}$.

III. DRL-BASED SOLUTION

To solve the non-convex problem in (10), we propose a twin-DDPG deep reinforcement learning framwork, instead of using single agent to find the optimal $\mathbf{G}$, $\boldsymbol{\Theta}$ and $\mathbf{Q}$, since $\mathbf{Q}$ would be highly coupled with large scale CSI using single agent which is actually irrelevant. As shown in Fig.2, the first network takes CSI as state to obtain the optimal $\mathbf{G}$ and $\boldsymbol{\Theta}$, while the second network takes UAV position as state to obtain the movement $\mathbf{q}[\mathbf{n}+\mathbf{1}] - \mathbf{q}[\mathbf{n}]$ of UAV at n-th time slot. Both network take the sum secrecy data rate as reward. The overall algorithm for solving problem (10) is summarized in Algorithm 1.

A. Active and Passive Beamforming

Inspired by the work of [5], a DDPG-based network is employed to learn the optimal policy in terms of the UAV's beamforming matrix $\mathbf{G}$ and the RIS's reflecting beamforming

Algorithm 1 Twin-DDPG Deep Reinforcement Learning Algorithm

Initialize the 1-st network with initial Q-function $Q_1(s,a;\boldsymbol{\theta}_1)$;
Initialize the 2-nd network with initial Q-function $Q_2(s,a;\boldsymbol{\theta}_2)$;
for Episode $n_{ep}=1,2,...,N_{ep}$ **do**
  Reset the positions of the UAV and users;
  Reset the active and passive beamforming matrix $\mathbf{G}$, $\boldsymbol{\Theta}$;
  for Step $n=1,2,...,N_{step}$ **do**
    Observe all CSI as $s_{n,1}$;
    Observe the UAV position as $s_{n,2}$;
    Select actions $a_{n,1}, a_{n,2}$ with a gaussian action noise n_a with variance σ_a :

$$\begin{aligned} a_{n,1} &= \arg\max_{a_{n,1}\in\mathcal{A}} Q_1\left(s_{n,1}, a_{n,1}; \theta_{n,1}\right) + n_a \\ a_{n,2} &= \arg\max_{a_{n,2}\in\mathcal{A}} Q_2\left(s_{n,2}, a_{n,2}; \theta_{n,2}\right) + n_a \end{aligned} \tag{11}$$

    Execute action $a_{n,1}, a_{n,2}$, receive an immediate reward $r_{n,1}$ using Eq. (12) and new states $s_{n+1,1}, s_{n+1,2}$. Note that $r_{n,1}=r_{n,2}$;
    Store the transitions $[s_{n,1}, a_{n,1}, r_{n,1}, s_{n+1,1}]$ and $[s_{n,2}, a_{n,2}, r_{n,2}, s_{n+1,2}]$ in the two networks' memory queues, respectively;
    Sample a mini-batch of transitions in memory queue randomly to update both networks using proper loss function and policy gradient function [8];
  end for
end for

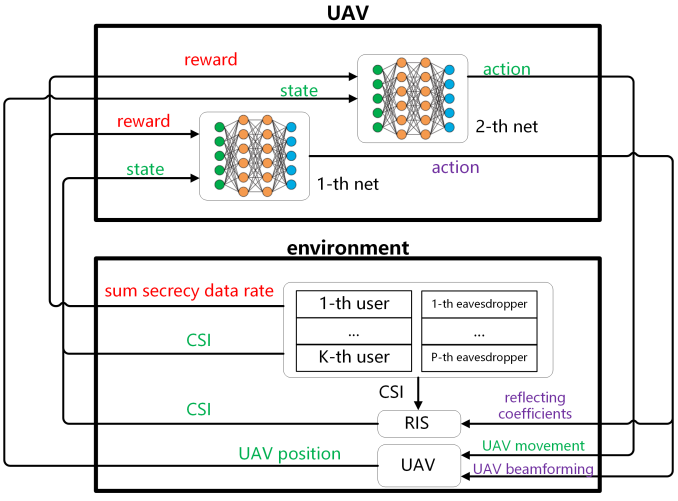

Fig. 2. structure of proposed twin-DDPG framwork.

matrix $\boldsymbol{\Theta}$ by interacting with the whole system. Each episode is defined as a time span T, where each step is defined as a time slot δ_n. In order to maximize the sum secrecy rate, the state $s_{n,1}$, the action $a_{n,1}$, the reward $r_{n,1}$ in n-th time slot of the first agent is defined as follows:

1) State $s_{n,1}$: the state of the first agent in n-th time slot contains the estimated comprehensive CSI from the UAV to all legitimate users and eavesdroppers, i.e., $\{\mathbf{h}_{\mathbf{U,i}}^{\mathbf{H}} + \boldsymbol{\Psi}^{\mathbf{H}}\mathbf{H}_{\mathbf{C,i}}\}, \forall \mathbf{i} \in \mathcal{K}\cup\mathcal{P}$.
2) Action $a_{n,1}$: we define the phase shift of all RIS reflecting elements $\theta_n, \forall n \in \mathcal{N}$ and the transmit beamforming matrix G as action. It is worth noting that $\mathbf{G}=\mathbf{Re}\{\mathbf{G}\}+\mathbf{Im}\{\mathbf{G}\}$ are separated as real part and imaginary part to tackle with the real input problem.
3) Reward $r_{n,1}$: the reward function is defined as:

$$r_{n,1} = tanh(\sum_{k=1}^{K} R_k^{\text{sec}} - p_r - p_m), \tag{12}$$

where p_r is the penalty if the outage constraint (10c) is not satisfied, and p_m is the the penalty when the UAV flies out of the target area. The hyperbolic tangent function $tanh(\cdot)$ is exploited to limite reward in range of $(-1,1)$ for better convergence.

B. UAV Trajectory

The second DDPG is exploited to simultaneously obtain the optimal movement $\mathbf{q}[\mathbf{n+1}]-\mathbf{q}[\mathbf{n}]$ with $\mathbf{G}$ and $\boldsymbol{\Theta}$. It is feasible to utilize a single DDPG network to tune all parameters $\mathbf{G}, \boldsymbol{\Theta}, \mathbf{Q}$, which have been done by most works. But in this letter, UAV's trajectory is rarely relevant to the large amount of CSI, leading to instability and divergence by connecting irrelevant actions and feedbacks using a single network. The state $s_{n,2}$, the action $a_{n,2}$, the reward $r_{n,2}$ in n-th time slot of the second agent is defined as follows:

1) State $s_{n,2}$: as mentioned before, UAV's trajectory is rarely relevant to the large amount of CSI. So the second network only takes the UAV's position $\mathbf{q}[\mathbf{n}]$ as state.
2) Action $a_{n,2}$: the action contains the UAV's flying distance $\mu[n]$ and the direction $\psi[n]$. Then, the movement of UAV can be expressed as:

$$\mathbf{q}[\mathbf{n}+\mathbf{1}] - \mathbf{q}[\mathbf{n}] = \mu[\mathbf{n}](\mathbf{cos}\psi[\mathbf{n}]\mathbf{e_x} + \mathbf{sin}\psi[\mathbf{n}]\mathbf{e_y}) \tag{13}$$

3) Reward $r_{n,2}$: the same reward function in (12) is employed, since both network have the same objective to maximize the sum secrecy rate.

IV. Simulation Results

In this section, numerical results are presented to characterize the performance of our proposed solution. For the first DDPG-based network, we deploy four fully-connected hidden layers with [800,600,512,256] neurons in both actor and critic networks and the AdamOptimizer is used to train the actor network with learning rate 0.0001 and critic network with learning rate 0.001. The second net has the same structure as the first net, but with different number of four layers [400,300,256,128]. The initial coordinates of UAV and RIS are set as [0,25,50], [0,50,12.5]. The eavesdropper is placed at [47,-4,0], while we model two legitimate users' movement as uniform motion in a straight line as shown in Fig.3. More parameters are shown in Tabel. I.

Fig.3 illustrates the optimized trajectory, which eventually converges as the learning procedure is over. It can be observed that the UAV tends to move away from the eavesdropper.

TABLE I
MAIN PARAMETERS.

Parameter	Value
UAV antennas number	$\mathbf{A} = 4$
eavesdropper number	$\mathbf{P} = 1$
legitimate user number	$\mathbf{K} = 2$
step number	$N_{step} = 100$
episode number	$N_{ep} = 100$
carrier frequency	$f_c = 28$ GHz
max transmission power	$P_{max} = 30$ dBm
noise power	$\sigma_n = -114$ dBmW
path loss factor [9]	$\alpha_{ur} = 2.2, \alpha_u = 3.5, \alpha_r = 2.8$
shadow fading factor	$\sigma_s = 3$ dB

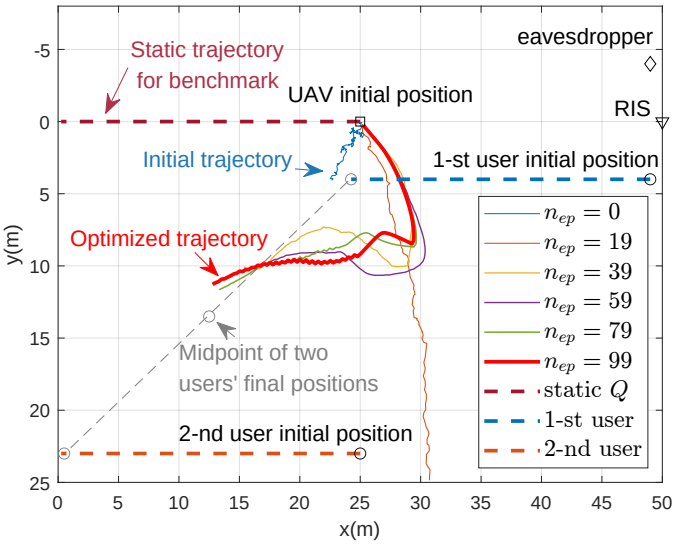

Fig. 3. Converged trajectory of the UAV.

Moreover, the UAV is inclined to chase and follow the midpoint of two users position, while keeping relatively close distance to the RIS. This implies the UAV trajectory is jointly optimized with the active and passive beamforming, and the proposed algorithm can adapt to dynamic conditions brought by the users' mobility.

Fig.4 plots the average sum secrecy rate by different benchmarks which all increase with traning episode n_{ep}. It is found that the proposed solution achieves the best performance under imperfect CSI. However, the proposed solution performs slightly better under the perfect CSI, which implies the proposed solution has good robustness. Thus, the secrecy rates of legitimate users in our proposed system under imperfect CSI can be maximized leveraging RIS and UAV by DDPG-based algorithm.

V. CONCLUSION

In this letter, we investigated robust and secure transmission for RIS-aided mmWave UAV communications. To maximize the secrecy rates of legitimate users, we proposed a DDPG-baesd optimization algorithm. The simulation results validated that by jointly optimizing UAV trajectory and active (passice) beamforming, the best performance can be achieved under imperfect CSI compared with other benchmarks.

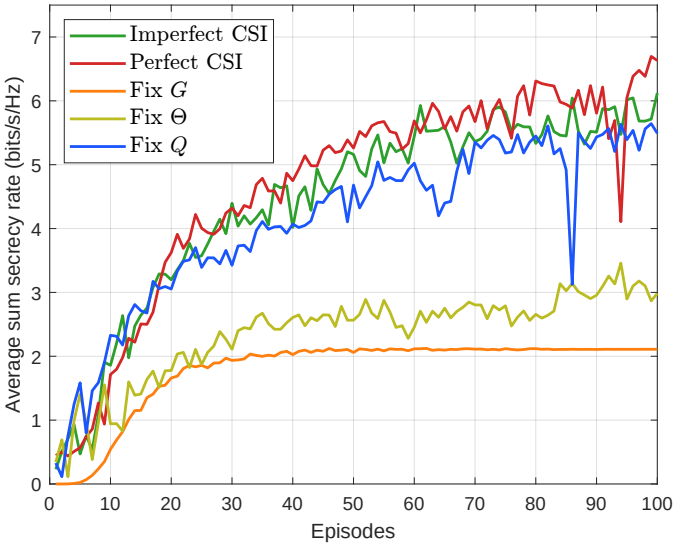

Fig. 4. Accumulated reward performance versus episodes under different RIS elements number.

REFERENCES

[1] 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," *arXiv preprint arXiv:2009.09716*, 2020.

[2] D. Zhao, H. Lu, Y. Wang, H. Sun, Y. Gui, and J. Wu, "Joint power allocation and user association optimization for irs-assisted mmwave systems," *arXiv preprint arXiv:2010.11713*, 2020.

[3] S. Yong and J. Thompson, "A three-dimensional spatial fading correlation model for uniform rectangular arrays," *IEEE Antennas and Wireless Propagation Letters*, vol. 2, pp. 182–185, 2003.

[4] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Two-timescale beamforming optimization," *IEEE Transactions on Wireless Communications*, 2020.

[5] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning based intelligent reflecting surface for secure wireless communications," *IEEE Transactions on Wireless Communications*, 2020.

[6] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for irs-aided miso communications with imperfect cascaded channels," *arXiv preprint arXiv:2001.07054*, 2020.

[7] Y. Huang, F. Al-Qahtani, C. Zhong, Q. Wu, J. Wang, and H. Alnuweiri, "Performance analysis of multiuser multiple antenna relaying networks with co-channel interference and feedback delay," *IEEE Transactions on Communications*, vol. 62, no. 1, pp. 59–73, 2013.

[8] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[9] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in uav-ris wireless networks," *IEEE Journal on Selected Areas in Communications*, 2020.