## RESEARCH

## **Open Access**

# PPO-based deployment and phase control for movable intelligent reflecting surface



Yikun Zhao<sup>1</sup>, Fanqin Zhou<sup>1</sup>, Huaide Liu<sup>1</sup>, Lei Feng<sup>1</sup> and Wenjing Li<sup>1\*</sup>

### Abstract

Intelligent reflecting surface (IRS) stands as a promising technology to revolutionize wireless communication by manipulating incident signal amplitudes and phases to enhance system performance. While existing research primarily centers around optimizing the phase shifts of IRS, the deployment of IRS on movable platforms introduces a new degree of freedom in the design of IRS-assisted systems. Leveraging flexible deployment strategies for IRS holds the potential to further amplify network throughput and extend coverage. This paper addresses the challenging non-convex joint optimization problem of the movable IRS and proposes a dynamic optimization algorithm based on proximal policy optimization (PPO) for dynamically optimizing the aerial position and phase configuration of IRS. Simulation results show the effectiveness of the proposed approach, demonstrating significant performance improvements compared to communication schemes without IRS assistance and conventional static IRS-assisted methods.

Keywords Intelligent reflecting surface (IRS), Deep reinforcement learning (DRL), Movable deployment

#### Introduction

Intelligent reflecting surface (IRS) is a revolutionary technology that enhances wireless communication performance [1]. Comprising numerous cost-effective, passive, and reflective elements arranged in a planar configuration, IRS serves as a programmable surface capable of reshaping the entire wireless channel environment [2]. By accurately guiding signals to target areas, IRS effectively reduces the power consumption of communication devices [3]. This energy-saving feature holds tremendous potential, especially for mobile devices within cloud computing systems, where prolonged battery life is invaluable [4]. IRS technology boasts versatile applications across various domains, encompassing communication, optical communication, energy systems, military, civilian, and especially in high-energy-efficient communication scenarios such as Internet of Things [5], smart city [6], and industrial automation [7, 8]. Notably, IRS has the potential to optimize cloud computing systems from multiple critical perspectives, including communication quality [9, 10], energy efficiency [11], network performance [12], and security [6], thereby improving the overall availability and reliability of the system.

Researchers have started to explore the design of the IRS-assisted system to address poor coverage and connectivity issues that result in large uploading and downloading delays in cloud computing networks [13, 14]. The work [15] studies the phase optimization of IRS to improve the system sum rate performance. The reference [16] optimizes the phase shift and amplification of IRS to maximize the sum rate of multiple users in the uplink non-orthogonal multiple access system. These works consider deploying IRS in a fixed location, while the recent emergence of aerial base stations has brought new degrees of freedom to the IRS design. By exploiting a movable IRS mounted on an unmanned aerial vehicle (UAV), flexible 3D network coverage can be realized by placing an IRS wherever and whenever is needed [17].



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

<sup>\*</sup>Correspondence:

Wenjing Li

wjli@bupt.edu.cn

<sup>&</sup>lt;sup>1</sup> State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

Rich literature [18–20] has shown that the deployment location of the UAV is an important parameter affecting system communication performance. It is foreseeable that flexible deployment of IRS will further optimize system network throughput and coverage range. However, there is still a lack of research on optimization algorithms for the deployment of movable IRS due to its relatively short development time.

Regarding the deployment of movable IRSs, traditional convex optimization algorithms may struggle to solve the non-convex problems of joint optimization of deployment location and reflection phase [21]. Though some algorithms such as particle swarm optimization [22] promise to find solutions to these complex problems, they still require repeated computations when the network environment changes. Deep learning, an advanced artificial intelligence technique, harnesses multi-layered neural networks to model and learn complex data representations [23]. Deep learning models promise to handle large-scale, high-dimensional data and autonomously discover data representations. This unique advantage enables them to perform a wide range of tasks, including recommendation [24-26], detection [7, 8, 10, 27], and resource optimization [28, 29]. Deep reinforcement learning (DRL) harnesses the ability of deep learning to handle input data and discovers global optima for nonconvex problems through learning and simulating natural evolutionary processes without the need for strict mathematical modeling [30]. Authors in Ref. [31] utilize DRL to jointly design the deployment location and passive beamforming of the IRS. Though results show that DRL can be used to solve the joint optimization problem of IRS, they consider the static deployment strategy during the service. The work [32] considers the combination of IRS and UAV, and designs UAV 3D trajectory and IRS phase shift using DRL algorithm, but they utilize IRS to serve UAVs instead of using UAVs to carry IRS.

Therefore, this paper aims to propose an effective method based on deep reinforcement learning to design corresponding solutions. Specifically, this paper considers installing an IRS on a UAV to fully exploit the freedom of deployment and serve multiple users in a specific area. An algorithm based on proximal policy optimization (PPO) is developed for the movable IRS, dynamically optimizing both its position and phase to maximize received power at the user equipments. Simulation results demonstrate the effectiveness of this approach in enhancing network performance, outperforming communication schemes without IRS assistance and traditional static IRS-assisted communication methods. The rest of the paper is organized as follows. The system model and problem formulation are depicted in System model section. Then, we present the details of our proposed algorithm for the joint location and phase optimization for movable IRS in PPO-based joint location and phase optimization algorithm for movable IRS section. Simulation results and discussions section presents the discussions of simulation results. Finally, Conclusion section concludes the paper.

#### System model

We consider a system where there is a single access point (AP), multiple user equipments (UEs), and a single movable IRS mounted on a UAV, as shown in Fig. 1. In a wireless communication system, the IRS can offer substantial flexibility for data transmission, which is also an important factor influencing the performance of cloud computing systems [13]. The AP is equipped with M antennas, and each antenna sends signals  $\omega$  and the transmit power of the AP is limited to below  $P_{max}$ . To reduce the complexity of the problem, it is assumed that the AP transmits at its maximum power in our work. The AP is located in a fixed position and has a certain height, which allows it to cover a large geographical area and provide stable service for users within its coverage area. The movable IRS carried by UAV has the property of 3D coordinates, which allows it to move within a certain height range and horizontal range. Therefore, this article can dynamically adjust the location and direction of IRS according to the communication environment and needs, optimizing the quality of communication. It has two working modes of receiving signal and reflecting signal, which makes it can be flexibly adjusted and applied in different situations. In received signal mode, the IRS receives signals from wireless communication devices for channel estimation and phase adjustment. It is critical to optimize the communication effect. Because it enables the IRS to adjust the signal according to changes of the channel state, which can ensure the quality and stability of the signal. In reflection mode, the IRS will reflect the signal from the access point and send the signal to the client through reflection of certain phase. In this process, the IRS not only acts as a relay but also enhances the signal at the user by adjusting the phase of the reflection.

In our considered movable IRS-assisted wireless communication scenario, multiple active signals are emitted from the AP. A portion of these signals is transmitted directly to multiple UEs via the AP-user channel, while another portion is first transmitted to the IRS via the AP-IRS channel, and then indirectly relayed to UEs after



Fig. 1 A movable IRS-assisted multi-user communication network

reflection by the IRS through the IRS-user channel. This process enhances the received power at UEs. Simultaneously, the IRS is carried by a UAV, initiating movement from a specific location. It continuously searches for the optimal deployment position within a certain range and adjusts reflection phases to optimize the overall network performance of the wireless communication system. The IRS consists of N reflectors,  $N_y$  in the vertical direction and  $N_x$  in the horizontal direction, and  $N = N_x N_y$ . Each reflector can be programmed independently and has an independent phase. This phase angle is controlled by the IRS controller, which can be adjusted as needed to achieve the best network performance.

The channel models among AP, IRS-UAV and user are respectively  $h_d^H \in C^{1*M}$ ,  $h_r^H \in C^{1*N}$ ,  $G \in C^{N*M}$ .  $h_d^H$ represents the channel model between AP and user.  $h_r^H$ represents the channel model between IRS and user. G represents the channel model between IRS and AP.  $C^{a*b}$ represents the complex valued matrix space of a \* b and H represents the conjugate transpose operation. There are multiple user terminal devices. The channel  $h_d^H$ ,  $h_r^H$ and G depend on the distance between AP and user, IRS and user, and IRS and AP, respectively. Denote  $X_{irs}$ ,  $X_{user}$ ,  $X_{ap}$  as the 3D coordinates of the IRS, UE and AP respectively, these distances can calculated as:

$$distance_{irs-user} = \sqrt{\|X_{irs} - X_{user}\|^2},$$
(1)

$$distance_{ap-user} = \sqrt{\left\|X_{ap} - X_{user}\right\|^2},\tag{2}$$

$$distance_{irs-ap} = \sqrt{\left\|X_{irs} - X_{ap}\right\|^2}.$$
(3)

These distances can affect the quality of the UEs' received signal because the signal will have various losses during the propagation process.

Let  $\theta = [\theta_1, \dots, \theta_N] \Theta = diag(\beta e^{j\theta_1}, \dots, \beta e^{j\theta_N})$ , where *j* represents an imaginary unit and *diag* represents the diagonal matrix.  $\Theta$  represents the diagonal phase matrix of the intelligent reflector.  $\theta_n \in [0, 2\pi]$  and  $\beta \in [0, 1]$  respectively correspond to the phase and reflection coefficient. During the experiment, each reflector has to reflect the maximum signal. So the default reflection coefficient  $\beta$  is 1.

$$y = (h_r^H \Theta G + h_d^H)\omega s + z, \tag{4}$$

where  $\theta$  represents the phase set at IRS,  $S_x$ ,  $S_y$ ,  $S_z$  represents the 3D coordinates of IRS-UAV,  $\Theta$  represents the diagonal matrix constructed by  $\theta$ ,  $P_{max}$  represents the maximum power. *s* is an independent and equally distributed random variable with zero mean and unit variance, and *z* represents the additional white Gaussian noise on the user receiver with zero mean and variance  $\sigma^2$ .

Thus, the signal power received by the user is:

$$P = \left| \left( h_r^H \Theta G + h_d^H \right) \omega \right|^2.$$
<sup>(5)</sup>

The value of power is only related to the channel  $h_r^H$ , G,  $h_d^H$  between them, the diagonal matrix  $\Theta$  constructed by the phase angle, and the signal  $\omega$  at the transmitting point.

As indicated in the previous research [33], IRS deployment near the user side or the AP side has a good effect in the case of a single user. However, it was not applicable in the case of multi-users. Considering the coverage relationship between them, the optimization problem in the case of multi-users was mainly constructed. The sum of the power received by all users at a certain time *t* can be expressed as  $\sum_{i}^{n} P_{t_i}$ .

The optimization problem is constructed with the goal of maximizing the sum of the signals received by n UEs over a time period T:

$$\max_{\substack{\theta, S_x, S_y, S_z, \\ \theta, S_x, S_y, S_z, \\ t}} \sum_{t}^{T} \sum_{i}^{n} P_{t_i},$$
s.t  $0 \le \theta < 2\pi,$ 
 $(S_x, S_y, S_z) \in S,$ 
(6)

where  $X_{irs} = (S_x, S_y, S_z)$  is the 3D coordinate of the IRS. The first constraint is the phase angle constraint of the IRS, and the second constraint restricts the movement of the IRS-UAV within a given interval *S* to adapt to the actual situation and avoid unlimited deployment.

# PPO-based joint location and phase optimization algorithm for movable IRS

This paper proposes a joint optimization algorithm for the airborne position and phase of IRS based on PPO to overcome the aforementioned challenges. The optimization problem is formulated as a Markov Decision Process (MDP), with carefully designed states, actions, and rewards to reduce the decision space of the algorithm. By leveraging the framework of deep neural networks (DNN), the IRS-UAV can learn from the environment and select appropriate strategies. Through the convergence of the DNN, the optimal deployment scheme is ultimately obtained.

First of all, we introduce our MDP design. The definitions of its state, action, and reward are as follows:

State: The state space is defined as  $S = [S_x, S_y, S_z, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5],$  where  $(S_x, S_y, S_z)$  represents the 3D dynamic position of the IRS-UAV which limited to a certain range, and  $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$  are the discrete phase shift. Assuming that the IRS has a total of N reflection elements, adjusting the phase angle of these reflection elements independently at the same time will result in excessive overhead in the system state space and action interval. Therefore, we divide the N reflection elements evenly into 5 parts, (i.e.,  $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ ), to avoid excessively large action intervals and high model complexity.

**Action:** The action space is  $A = (a_x, a_y, a_z, a_{\theta_1}, a_{\theta_2}, a_{\theta_3}, a_{\theta_4}, a_{\theta_5})$ , where  $(a_x, a_y, a_z)$  represents the action vectors of the IRS-UAV in the 3D space,  $(a_{\theta_1}, a_{\theta_2}, a_{\theta_3}, a_{\theta_4}, a_{\theta_5})$  represents the phase change of the corresponding reflection element in a step.

**Reward:** The reward is the power *P* received by the system user after executing the corresponding action in the current state state. The reward is the feedback signal obtained by the algorithm after the implementation of the action, which is used to guide the optimization of the model.

In addition, an end state *done* is also defined and *done* is set to *True* when the IRS motion exceeds a certain step, while the default value for others is set to *False*.

We use PPO for agent's behavior learning, which is proposed in Ref. [34] and mainly consists of two parts: actor network and critic network. The actor network is designed in two parts: one part computes the mean and the other part computes the standard deviation. These two parts are combined to obtain the output of a Gaussian distribution. The part that computes the mean consists of multiple fully connected layers, each followed by a Tanh activation function. The purpose of the Tanh activation function is to introduce nonlinearity and enhance the model's expressive power and ability to fit complex nonlinear relationships. The Tanh activation function maps the output to a symmetric S-shaped curve within the range [-1, 1], transforming the linear transformations of the input into a nonlinear space and increasing the model's capacity for nonlinear fitting. The design of the critic network is similar to

the part of the actor network that computes the mean, but with a different output dimension. The final output of the actor network is a vector of the same length as the action space, representing multiple concurrent actions, while the final output of the critic network is a single value representing the evaluation of the model.

Regarding the policy update of the algorithm network, it mainly consists of the two following steps.

1) Calculating the policy loss: First, using the current policy network parameters  $\theta$ , the probability distribution of actions is obtained based on the current interaction with the environment state. Then, using the old policy network parameters  $\theta_k$ , the probability distribution of actions is calculated for the same environment state. The two probability distributions are divided to obtain the probability ratio, denoted as  $r_t(\theta)$ . The policy loss function, denoted as  $\mathcal{L}_{\theta_k}^{CLIP}(\theta)$ , is then defined as:

 $\operatorname{clip}(r_t(\theta), 1 - \varepsilon, 1 + \varepsilon)$  is the clipping function applied to the probability ratio  $r_t(\theta)$ . It ensures that the policy update does not deviate too far from the old policy, and  $\varepsilon$  controls the degree of clipping. The overall objective is to maximize this loss function with respect to the new policy parameters  $\theta$  while ensuring that the policy update remains within a certain range. If the probability ratio exceeds this range, the loss is truncated to limit the magnitude of policy updates. This is implemented to prevent significant changes in the policy network within a single update, avoiding training instability.

2) Updating the network parameters: The parameters of the policy network  $\theta_{k+1}$  are obtained by maximizing the clipped surrogate objective  $\mathcal{L}_{\theta_{L}}^{CLIP}(\theta)$  through techniques like backpropagation and gradient descent:

$$\mathcal{L}_{\theta_{k}}^{CLIP}(\theta) = \mathop{\mathbb{E}}_{\tau \sim \pi_{k}} \left[ \sum_{t=0}^{T} \left[ \min\left( r_{t}(\theta) \hat{A}_{t}^{\pi_{k}}, \operatorname{clip}(r_{t}(\theta), 1-\varepsilon, 1+\varepsilon) \hat{A}_{t}^{\pi_{k}} \right) \right] \right],$$
(7)

$$\theta_{k+1} = \arg \max_{\theta} \mathcal{L}_{\theta_k}^{CLIP}(\theta).$$
(8)

where  $\hat{A}_t^{\pi_k}$  represents the generalized advantage estimation (GAE) of taking action  $a_t$  at time t under the old policy  $\theta_k$ . The advantage is a measure of how much better or worse an action is compared to the average action taken in that state.

After introducing the structural design of the algorithm, the overall process of the PPO-based location and phase optimization algorithm for the movable IRS algorithm is shown in Algorithm 1.

- 1: Input: Number of user terminals n and their respective positions  $d_{user}$ , maximum training steps M, memory capacity;
- 2. Output: Optimized deployment location and phase of the movable IRS
- 3: Initialize:
- Import and initialize the running environment and parameters 4:
- 5: for i = 1 : M do
- 6: Reset the state  $S = S_0$
- 7: Initialize the locations of UEs
- 8: while  $done = True \ do$
- 9: Agent selects actions based on the current state S, obtains the distribution of actions, and samples them:
- 10: Agent interacts the action with the IRS-assisted communication environment to obtain the next state, reward, end condition *done*, and store them in the replay buffer;
- 11: If the number of counts in the replay buffer is equal to batch size:
- 12: Calculate the advantage function  $\hat{A}_t^{\pi_k}$
- 13: //Adopt small batch K-step updates;
- 14: for k = 1 : K do
- 15:
- $\begin{array}{l} \text{Calculate the loss function } \mathcal{L}_{\theta_k}^{CLIP}(\theta); \\ \text{Update model parameters } \theta_{k+1} = \arg\max_{\theta} \mathcal{L}_{\theta_k}^{CLIP}(\theta); \end{array}$ 16:
- 17: end for
- end while 18:
- 19: end for



#### Table 1 Main simulation parameters

Parameters	Values
Number of AP antennas	8
Number of IRS reflection units	50
Height of base station (meters)	2
actor learning rate	$3 \times 10^{-4}$
critic learning rate	$3 \times 10^{-4}$
Discount factor $\gamma$	0.99
GAE parameters $\lambda$	0.95
Renewal amplitude $arepsilon$	0.2
Hidden layer size	64
Maximum training times	$3 \times 10^{6}$
Batch size	2048
Mini batch size	64

#### Simulation results and discussions

The location of the AP is set to (0,0,2), that is, the origin of the coordinates of the horizontal position. The intelligent reflector reflection units are divided into 5 groups for phase adjustment. The total number of users is between 5 and 10, and the user locations are relatively clustered. In terms of channel, the intelligent reflector channel and base station channel are considered as uniform rectangular array and uniform linear array respectively. The signal attenuation of all channels is 30 dB in the range of reference distance 1. The corresponding channel matrix *G* has rank 1 and the row and column vectors are linearly dependent. The AP-user (direct) and AP-IRS-user channels are set to have 10 dB penetration losses, as well as their independent Rayleigh fading and path loss indices of 3. Set the signal gain at the user and AP to 0 dBi, and each reflector signal gain to 5 dBi. For all simulations, the information transmission scenario is considered and the power size and signal-to-noise ratio

The specific parameters are listed in Table 1. To further optimize the model, several common optimization techniques, including regularization, gradient clipping, orthogonal initialization, and learning rate decay, were employed in this work. Specifically, to further optimize the model and reduce its complexity, enhance generalization, reduce overfitting, and improve stability, this study applies advantage function regularization. After calculating the advantages for a batch using GAE, the mean and standard deviation of advantages for the entire batch are computed, and each advantage value is then normalized by subtracting the mean and dividing by the standard deviation. State and reward regularization is performed to ensure that states and rewards are within a consistent scale, preventing large or small rewards from adversely affecting model training, especially when

of the user's receiver are used as performance indicators.



Fig. 2 Convergence of the proposed algorithm

computing value functions. To prevent and address gradient explosion issues, gradient clipping was applied during practical training. Specifically, after calculating the loss, before updating the actor and critic networks, a threshold was set to control the magnitude of gradients, ensuring they do not become too large and preventing gradient explosions. This clipping ensures that gradient values are truncated within a reasonable range. Additionally, orthogonal initialization was introduced to further mitigate gradient-related problems. During neural network training, the learning rate is gradually reduced as training progresses. This gradual reduction in learning rate as the total training steps increase helps reduce fluctuations in later stages of training, enhancing model stability and accelerating convergence.

Then we analyze the convergence performance of the proposed algorithm through simulations, paying special attention to the impact of the optimization methods mentioned above on the algorithm performance. The result is shown in Fig. 2, and the horizontal axis is the number of algorithm training steps and the vertical axis is the real-time reward obtained after training and evaluating the current model. The curve in blue is the proposed PPO-based location and phase optimization algorithm with optimization techniques. Before  $5 \times 10^5$  steps, it is still in the exploration stage and PPO algorithm has not yet converged, which is highly volatile. After the execution of  $5 \times 10^5$  steps, the PPO algorithm basically converges and only fluctuates up and down within a limited

range, which indicates that the optimal deployment position and phase have been reached, and the dynamic deployment of movable IRS has been realized. The curve in gray is the PPO-based algorithm without using the aforementioned optimization measures, and the convergence rate is slow and the effect is poor, which proves the effectiveness of the optimization measures proposed above. Results show that the optimization approaches we adopted can significantly improve the convergence speed and performance of the algorithm.

To analyze the solving efficiency of the algorithm, we compare the proposed algorithm with the method based on mathematical optimization, as shown in Fig. 3. The horizontal axis is the number of users, that is, the problem size of the task being solved. The vertical axis is the solution time required. As can be seen from the figure, it can be seen that the method based on mathematical solving experiences a rapid escalation in solving time as the number of users grows, while the algorithm based on PPO has little change in solving time for the increase in the number of users within the same range. It can also be noticed that for a given problem scale, the proposed PPO-based algorithm exhibits notable efficiency advantages compared to the math method. Moreover, as the problem scale expands, these efficiency advantages become increasingly pronounced. It reflects the advantages of deep reinforcement learning in dealing with complex, high-dimensional and nonlinear problems.



Fig. 3 Comparison of algorithm solving efficiency



Fig. 4 Comparison of algorithm under different user number



Fig. 5 Comparison of algorithm under different AP transmission power

The algorithm performance under different number of users and different AP transmit power is shown in Figs. 4 and 5, respectively. It can be seen from the figure that with the increase of the number of users, the received power increases, and higher transmitting power leads to exponential growth in received power at the user end. We compare our proposed scheme with the other three schemes. In the ideal scheme, the result is obtained by mathematical optimization. The IRS without optimization scheme represents the case where the position of the IRS is not optimized but fixed in a random place and only the phase shift of IRS is optimized. In the without IRS scheme, IRS is not employed and AP transmits its signal to users. It can be seen that the result of the proposed algorithm is very close to the result of mathematical optimization under various network conditions, which performs much better than the other two cases. It can also be seen from the figure that the case without IRS achieves the worst performance. This outcome indicates that IRS can enhance communication performance, with even greater improvements when using movable IRS, thanks to its flexibility in deployment, which introduces new performance gains.

#### Conclusion

This paper addresses the joint optimization problem of phase shift and the location of a movable IRS which is equipped on a UAV in an IRS-assisted multi-user wireless communication system. A PPO-based joint dynamic optimization algorithm is designed for controlling the aerial position and phase of IRS. The simulation results show that the proposed scheme can improve the network performance of the system compared to communication schemes without IRS assistance and traditional static IRS-assisted communication schemes. Additionally, the proposed algorithm also has good performance in both convergence and solving efficiency. In future work, we will consider the coordinated deployment of multiple movable IRSs to accommodate scenarios with dispersed user distributions.

#### Authors' contributions

Yikun Zhao proposed the main idea, designed the algorithms and experiment schemes and drafted the technical part. Fanqin Zhou guided the design of the algorithms and experiment and prepared the final manuscript for submission. Huaide Liu was responsible for experiment environment setup and data visualization. Lei Feng refined the whole text of the manuscript and helped with preparing the final manuscript for submission. Wenjing Li investigated the research background and related research part of the manuscript. All the authors reviewed the manuscript.

#### Funding

This research was funded by the National Natural Science Foundation of China (No. 61971053), BUPT Excellent Ph.D. Students Foundation (No. CX2022223) and BUPT Innovation and Entrepreneurship Support Program (2023-YC-A131).

Not applicable.

#### Declarations

**Ethics approval and consent to participate** Not applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 8 September 2023 Accepted: 11 October 2023 Published online: 01 December 2023

#### References

- Wu Q, Zhang S, Zheng B, You C, Zhang R (2021) Intelligent reflecting surface-aided wireless communications: A tutorial. IEEE Trans Commun 69(5):3313–3351. https://doi.org/10.1109/TCOMM.2021.3051897
- Dai Y, Guan YL, Leung KK, Zhang Y (2021) Reconfigurable intelligent surface for low-latency edge computing in 6G. IEEE Wirel Commun 28(6):72–79. https://doi.org/10.1109/MWC.001.2100229
- Wu Q, Zhang R (2020) Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network. IEEE Commun Mag 58(1):106–112. https://doi.org/10.1109/MCOM.001.1900107
- Gopu A, Thirugnanasambandam K, AlGhamdi AS, Alshamrani SS, Maharajan K, Rashid M (2023) Energy-efficient virtual machine placement in distributed cloud using NSGA-III algorithm. J Cloud Comput 12(1):124
- Xu X, Jiang Q, Zhang P, Cao X, Khosravi MR, Alex LT, Qi L, Dou W (2022) Game theory for distributed IoV task offloading with fuzzy neural network in edge computing. IEEE Trans Fuzzy Syst 30(11):4593–4604. https:// doi.org/10.1109/TFUZZ.2022.3158000
- Xu X, Fang Z, Zhang J, He Q, Yu D, Qi L, Dou W (2021) Edge content caching with deep spatiotemporal residual network for IoV in smart city. ACM Trans Sen Netw 17(3). https://doi.org/10.1145/3447032
- Yang Y, Yang X, Heidari M, Khan MA, Srivastava G, Khosravi M, Qi L (2022) ASTREAM: Data-stream-driven scalable anomaly detection with accuracy guarantee in IIoT environment. IEEE Trans Netw Sci Eng 1–1. https://doi. org/10.1109/TNSE.2022.3157730
- Qi L, Yang Y, Zhou X, Rafique W, Ma J (2022) Fast anomaly identification based on multiaspect data streams for intelligent intrusion detection toward secure Industry 4.0. IEEE Trans Ind Inform 18(9):6503–6511. https://doi.org/10.1109/TII.2021.3139363
- Shrivastav K, Yadav R, Jain K (2021) Joint MAP channel estimation and data detection for OFDM in presence of phase noise from free running and phase locked loop oscillator. Digit Commun Netw 7(1):55–61. https://doi.org/10.1016/j.dcan.2020.09.007
- Dai H, Yu J, Li M, Wang W, Liu AX, Ma J, Qi L, Chen G (2023) Bloom filter with noisy coding framework for multi-set membership testing. IEEE Trans Knowl Data Eng 35(7):6710–6724. https://doi.org/10.1109/TKDE. 2022.3199646
- 11. Su Y, Pang X, Chen S, Jiang X, Zhao N, Yu FR (2022) Spectrum and energy efficiency optimization in IRS-assisted UAV networks. IEEE Trans Commun 70(10):6489–6502. https://doi.org/10.1109/TCOMM.2022.3201122
- 12. Dong L, Li R (2022) Optimal chunk caching in network coding-based qualitative communication. Digit Commun Netw 8(1):44–50. https://doi. org/10.1016/j.dcan.2021.06.002
- Li W, Zhang J, Guan D, Cui B, Zheng Z, Feng G, Wang H, Zhang L (2023) Latency minimization for intelligent reflecting surface-assisted cloudedge collaborative computing. In: 2023 15th International Conference on Computer Research and Development (ICCRD). pp 51–56. https://doi. org/10.1109/ICCRD56364.2023.10080403
- Abed GA, Jaleel IF (2023) Enhancement of spectral efficiency in intelligent reflecting surfaces (IRS's) over distributed and cloud-computing systems. In: 2023 Second International Conference on Electrical, Electronics,

Information and Communication Technologies (ICEEICT). pp 1–7. https://doi.org/10.1109/ICEEICT56924.2023.10157483

- Zhang P, Wang X, Feng S, Sun Z, Shu F, Wang J (2022) Phase optimization for massive irs-aided two-way relay network. IEEE Open J Commun Soc 3:1025–1034. https://doi.org/10.1109/OJCOMS.2022.3185463
- Chen CW, Tsai WC, Wu AY (2022) Low-complexity two-step optimization in active-irs-assisted uplink NOMA communication. IEEE Commun Lett 26(12):2989–2993. https://doi.org/10.1109/LCOMM.2022.3204749
- Xiao Y, Tyrovolas D, Tegos SA, Diamantoulakis PD, Ma Z, Hao L, Karagiannidis GK (2023) Solar powered UAV-mounted RIS networks. IEEE Commun Lett 27(6):1565–1569. https://doi.org/10.1109/LCOMM.2023.3264493
- Deng D, Li X, Menon V, Piran MJ, Chen H, Jan MA (2022) Learning-based joint UAV trajectory and power allocation optimization for secure IoT networks. Digit Commun Netw 8(4):415–421. https://doi.org/10.1016/j. dcan.2021.07.007
- Zhang S, Zhang L, Xu F, Cheng S, Su W, Wang S (2023) Dynamic deployment method based on double deep Q-network in UAV-assisted MEC systems. J Cloud Comput 12(1):1–16. https://doi.org/10.1186/ s13677-023-00507-6
- Zhao Y, Zhou F, Feng L, Li W, Yu P (2023) MADRL-based 3D deployment and user association of cooperative mmWave aerial base stations for capacity enhancement. Chin J Electron 32(2):283–294. https://doi.org/10. 23919/cje.2021.00.327
- Truong TP, Tuong VD, Dao NN, Cho S (2023) FlyReflect: Joint flying IRS trajectory and phase shift design using deep reinforcement learning. IEEE Internet Things J 10(5):4605–4620. https://doi.org/10.1109/JIOT.2022. 3218740
- Lu Y, Liu L, Gu J, Panneerselvam J, Yuan B (2022) EA-DFPSO: An intelligent energy-efficient scheduling algorithm for mobile edge networks. Digit Commun Netw 8(3):237–246. https://doi.org/10.1016/j.dcan.2021.09.011
- Wang Y, Wang J, Zhang W, Zhan Y, Guo S, Zheng Q, Wang X (2022) A survey on deploying mobile deep learning applications: A systemic and technical perspective. Digit Commun Netw 8(1):1–17. https://doi.org/10. 1016/j.dcan.2021.06.001
- Liu Y, Zhou X, Kou H, Zhao Y, Xu X, Zhang X, Qi L (2023) Privacy-preserving point-of-interest recommendation based on simplified graph convolutional network for geological traveling. ACM Trans Intell Syst Technol. https://doi.org/10.1145/3620677
- Liu Y, Wu H, Rezaee K, Khosravi MR, Khalaf OI, Khan AA, Ramesh D, Qi L (2023) Interaction-enhanced and time-aware graph convolutional network for successive point-of-interest recommendation in traveling enterprises. IEEE Trans Ind Inform 19(1):635–643. https://doi.org/10.1109/ TII.2022.3200067
- Qi L, Liu Y, Zhang Y, Xu X, Bilal M, Song H (2022) Privacy-aware point-ofinterest category recommendation in Internet of things. IEEE Internet Things J 9(21):21398–21408. https://doi.org/10.1109/JIOT.2022.3181136
- Xu X, Tian H, Zhang X, Qi L, He Q, Dou W (2022) DisCOV: Distributed COVID-19 detection on X-Ray images with edge-cloud collaboration. IEEE Trans Serv Comput 15(3):1206–1219. https://doi.org/10.1109/TSC.2022. 3142265
- Jia Y, Liu B, Dou W, Xu X, Zhou X, Qi L, Yan Z (2022) CroApp: A CNN-based resource optimization approach in edge computing environment. IEEE Trans Ind Inform 18(9):6300–6307. https://doi.org/10.1109/TII.2022.3154473
- Zhu D, Xu Z, Xu X, Zhao Q, Qi L, Srivastava G (2021) Cognitive analytics of social media services for edge resource pre-allocation in industrial manufacturing. IEEE Trans Comput Soc Syst 8(2):500–511. https://doi.org/ 10.1109/TCSS.2021.3052231
- Huang Y, Feng B, Cao Y, Guo Z, Zhang M, Zheng B (2023) Collaborative on-demand dynamic deployment via deep reinforcement learning for IoV service in multi edge clouds. J Cloud Comput 12(1):1–18. https://doi. org/10.1186/s13677-023-00488-6
- Liu X, Liu Y, Chen Y, Poor HV (2021) RIS enhanced massive non-orthogonal multiple access networks: Deployment and passive beamforming design. IEEE J Sel Areas Commun 39(4):1057–1071. https://doi.org/10.1109/JSAC. 2020.3018823
- Mei H, Yang K, Liu Q, Wang K (2022) 3D-trajectory and phase-shift design for RIS-assisted UAV systems using deep reinforcement learning. IEEE Trans Veh Technol 71(3):3020–3029. https://doi.org/10.1109/TVT.2022. 3143839
- Mu X, Liu Y, Guo L, Lin J, Schober R (2021) Joint deployment and multiple access design for intelligent reflecting surface assisted networks. IEEE

Trans Wirel Commun 20(10):6648–6664. https://doi.org/10.1109/TWC. 2021.3075885

 Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimization algorithms. arXiv:1707.06347

#### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- Open access: articles freely available online
- ► High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com