UAV-Aided 5G Communications with Deep Reinforcement Learning Against Jamming

Xiaozhen Lu*, Liang Xiao*, Canhuang Du*

*Dept. of Communication Engineering, Xiamen Univ., Xiamen, China. Email: lxiao@xmu.edu.cn

Abstract—The 5th generation (5G) systems are vulnerable to jamming attacks, especially smart jammers that choose their jamming policies such as the jamming channel frequencies and power based on the ongoing communication policies and network states. In this article, we present an unmanned aerial vehicle (UAV) aided 5G communication framework against jamming. In this scheme, UAVs use reinforcement learning methods to choose the relay policy for mobile devices in 5G systems, if the serving base station is heavily jammed. More specifically, we propose a deep reinforcement learning based UAV power allocation and relay algorithm to help 5G systems resist smart jamming without being aware of the jamming model and the network model in the dynamic game based on the previous anti-jamming relay experiences and the observed current network status. This algorithm has been proved to achieve the optimal performance and its performance bounds including the bit error rate (BER) and the energy consumption are provided. Simulation results show that this scheme can efficiently reduce the BER of the messages and save the energy consumption of the 5G cellular network, compared with the benchmark existing scheme.

Index Terms—5G, jamming, UAV, reinforcement learning.

I. INTRODUCTION

The 5th generation (5G) cellular communication systems have to support booming computation intensive applications such as augmented reality (AR) games and are vulnerable to jamming attacks due to the high user mobility in large-scale dynamic networks [1], [2]. Jammers send faked or replayed jamming signals to block the ongoing communications, exhaust the battery levels of mobile devices, threaten user privacy, and further perform other attacks such as man-in-the-middle attacks [2]. Jammers can be static or reactive. In particular, smart jammers as an advanced and most dangerous type of reactive jammers use smart radio devices and machine learning techniques to infer the 5G defense policy and then attack it accordingly [3].

Unmanned aerial vehicles (UAVs) can help 5G systems resist jamming due to the high altitude, mobility, and line-of-sight (LoS) channels to the mobile devices [4]– [7]. More specifically, UAVs relay the messages for mobile users if the serving base stations (BSs) are seriously blocked. In the UAV-aided 5G communications, a UAV has to choose its relay policy such as the relay power without being aware of the jamming network and radio channel states of the 5G network. Therefore, the UAV mobility optimization based relay schemes such as [4] cannot be directly applied in the UAV-aided 5G systems. This issue can be addressed by reinforcement learning techniques (RL) such as Q-learning and policy hill climbing (PHC) [2], [3], as the repeated UAV relay process in the dynamic 5G communication game against jamming can be formulated as a Markov decision process (MDP) [9].

Reinforcement learning techniques have been applied in the anti-jamming relay selection and power control in wireless networks [10], [11] and vehicular ad hoc networks (VANETs) [7]. For instance, in the RL-based UAV-aided VANET system as proposed in [7], the UAV applies the PHC algorithm to choose whether to relay a message for an onboard unit in the VANET if the area is seriously jammed or interfered. This scheme can improve the signal-to-interference plus-noise-ratio (SINR) of the onboard unit signals and reduce the bit error rate (BER) of the messages. However, this scheme will suffer from a long learning time and its performance will degrade in the 5G systems, due to the random exploration at the beginning, the estimation error and delay regarding the network state, and reward in the dynamic game.

The user-UAV link and the BS-UAV link usually have better channel states due to the LoS propagation of the UAV, compared with the link of the serving BS and the user at a fixed location that is severely blocked by jamming attacks [7]. In this article, we propose a UAV-aided 5G communication system, in which the UAV applies deep reinforcement learning algorithms to choose its relay policy against jamming attacks, including smart jamming. This system uses fast deep Q-network (DQN) [12] to accelerate the learning speed of the PHC-based anti-jamming UAV relay algorithm in [7]. More specifically, the UAV exploits the deep convolutional neural network (CNN) to compress the high-dimensional state space, applies the experience replay technique to update the CNN parameters, uses transfer learning to initialize the CNN weights with the previous anti-jamming relay experiences in similar scenarios.

In the fast DQN-based UAV relay scheme, the UAV formulates the current state with the SINR of the signals received by the serving BS, that sent from the mobile device to the UAV, that for the signal from the UAV to the backup BS, and the previous BER of the messages by the server. The UAV chooses the relay transmit power based on the current state and the Q-value for each relay policy, which is the output of the CNN. This algorithm enables a UAV to decide the optimal relay policy without knowing the jamming model and the 5G network model in a dynamic anti-jamming communication game.

The DQN-based UAV relay scheme can achieve the optimal relay policy via error-and-trials in the dynamic UAV relay game after a long enough anti-jamming experience. We provide the BER performance bound of the user messages and discuss the computation complexity. Simulation results showing that this UAV relay strategy can efficiently reduce...
the BER of the messages and save the energy consumption for 5G systems to resist smart jamming attacks, compared with existing UAV relay scheme as presented in [7].

The contributions of this work can be summarized as follows:

- We propose a fast DQN-based UAV relay policy to help 5G systems resist smart jamming attacks, without knowing the 5G network model and the jamming model.
- We prove that the relay scheme can achieve the optimal anti-jamming communication performance, and provide the BER performance bound. We also evaluate the computation complexity of the proposed scheme.

This article is organized as follows. We review the related work and present the UAV-aided 5G network model. We then propose a fast DQN-based anti-jamming UAV relay strategy for 5G networks and evaluate its performance. Finally, we conclude this article and identify the future work.

II. RELATED WORK

UAVs can relay messages for ground terminals against jamming attacks in 5G systems. For instance, the mobile relaying systems as proposed in [5] optimize the UAVs transmit power and trajectory to resist jamming and improve the throughput of 5G systems. The 5G communication scheme in [6] uses UAVs to relay the messages for mobile devices to resist jamming and achieve lower time complexity. The UAV-aided VANETs in [7] that presents a hotbooting PHC-based UAV relay algorithm uses UAV to relay the messages for onboard units to resist smart jamming, which can reduce the BER of the messages and thus improves the SINR of the VANET.

UAVs can use reinforcement learning techniques to resist jamming attacks. For instance, the UAV relay scheme in [13] uses the regret-based Q-learning algorithm to choose the transmission duty cycle to relay the messages and increase the communication capacity for 5G systems against jamming attacks. A cache-enabled communication scheme as developed in [12] applies DQN algorithm for interference alignment and user selection to provide user cooperation and resist jamming. The anti-jamming scheme as designed in [8] applies Q-learning algorithm to select the transmit power to resist jamming attacks and achieve higher SINR of the UAV systems. A UAV communication system as developed in [9] uses DQN algorithm to choose the transmit power against subjective smart attackers to improve the safe rate and the secrecy capacity in the UAV communication.

UAVs can also adopt supervised learning techniques such as randomized weighted majority algorithm (WMA) and Bayesian to prevent wireless networks from jamming attacks. For example, the UAV relay scheme as presented in [10] uses the randomized weighted majority algorithm (WMA) to decide whether or not to relay the messages and improve the successful transmission rate for 5G systems against jamming attacks. The Bayesian based UAV relay algorithm as presented in [14] uses UAVs to relay the alarm messages of the lethal jamming attacks to improve the detection accuracy and reduce the energy consumption of VANETs. However, these schemes will suffer from computation and communication costs due to a large amount of training data and complicated features extraction process, as summarized in Table I.

III. SYSTEM MODEL

A. Network Model

As shown in Fig. 1 a user such as a smart-phone with limited caching resources and battery life has to send real-time messages such as videos and AR game information to the servers at the cloud. The current serving BS of the user is denoted by BS0. Each BS at the fixed location is connected via fibers to each other and the core network. A UAV monitors the status of the BSs in the area and helps relay the messages if a BS is seriously blocked by a jammer. The UAV will fly to the heavily jammed area to receive the messages sent by the user and chooses the relay policy to relay the message to
### TABLE I
SUMMARY OF THE LEARNING-BASED ANTI-JAMMING WIRELESS COMMUNICATION METHODS

| Learning techniques | Action                      | Performance                  | Application                  | Ref     |
|---------------------|-----------------------------|------------------------------|------------------------------|---------|
| Q-learning          | Power allocation            | BER                          | VANETs                       | [7]–[9], [12], [13] |
|                     | Relay or not                |                              | UAV systems                  |         |
|                     | Channel selection           | SINR                         | Cellular systems             |         |
|                     | Duty cycle selection        | Capacity                      | Cognitive radio networks     |         |
|                     |                             | Safe rate                    |                              |         |
|                     |                             | Secrecy capacity             |                              |         |
|                     |                             | Percentage of payoff         |                              |         |
| PHC/WoLF-PHC        | Relay or not                | BER                          | VANETs                       | [7], [9] |
|                     | Power allocation            |                              | UAV systems                  |         |
| Randomized WMA      | Transmit or not             | Successful rate              | Wireless networks            | [10]    |
| DQN                 | User selection              | SINR                         | UAV systems                  | [9], [12] |
|                     | Power allocation            |                              | Wireless networks            |         |
|                     | Resource allocation         |                              |                              |         |
| Bayesian            | Monitor or not              | Accuracy rate                | VANETs                       | [14]    |
|                     | Node classification         |                              |                              |         |
|                     |                             | Communication overhead       |                              |         |

a backup BS denoted by BS₁ that is much farther away from the jammer than BS₀. For simplicity, we assume a quadrature phase-shift keying (QPSK) as the digital modulation in the 5G system with an additive white Gaussian noise.

The user sends a message with transmit power $P$ at time slot $k$ to the serving BS connecting to the server in the core network via fibers. The server decodes the message and measures the BER of the user message, and evaluates the SINR of the user-BS₀ signals denoted by $ρ_{1}^{(k)}$. Upon receiving the feedback information from the server, including the BER and the SINR of the user-BS₀ signals, BS₀ notifies the UAV about the SINR of the user-BS₀ signals via a feedback channel, and sends an ACK to the user.

Upon receiving the message sent by the user, the UAV evaluates the SINR of signal denoted by $ρ_{2}^{(k)}$ according to the feedback BER calculated by the user. The UAV chooses its relay power denoted by $x \in \mathcal{A} = [0, P_{U}^{M}]$, where $P_{U}^{M}$ is the maximum UAV transmit power, and $\mathcal{A}$ is the action set. The UAV sends the message with power $x$ and delay cost $C_U$ to the backup BS, i.e., BS₁ in Fig. 1, if $x > 0$.

The backup BS relays the message from the UAV to the server via fibers. The BER of the user message denoted by $P_{c}^{(k)}$ at time slot $k$ depends on the channel qualities of the links of the user and BS₀, the user and the UAV, and the UAV-BS₁ link. The server calculates the BER of the message forwarded by BS₁ and the SINR of the signals denoted by $ρ_{3}^{(k)}$, and sends the SINR of the UAV-BS₁ signals to the UAV via BS₁. For simplicity, we denote $\mathbf{r}^{(k)} = [ρ_{3}^{(k)}]_{1 \leq i \leq 3}$ as the SINR vector.

### B. UAV-Ground Channel Model

The UAV-ground links in 5G systems sometimes have shadow fading due to terrains and buildings and multi-path propagation due to the mountains, ground surface, and foliage [7]. The channel power gain of the user-BS₀ link denoted by $h_{1}^{(k)}$ is usually lower than the user-UAV link denoted by $h_{2}^{(k)}$ due to the LoS propagation of the UAV. In addition, the channel power gain of the UAV-BS₁ link denoted by $h_{3}^{(k)}$ is much higher than that of the user-BS₀ link due to the LoS UAV-BS₁ channel. According to [9], the UAV-ground channel follows a log-normal shadowing to model with constant channel power gains within a time slot.

### C. Jamming Model

A jammer is located close to the current serving BS of the user and sends jamming signals to prevent BS₀ from receiving message from the user. A smart jammer can choose its jamming power $y \in [0, P_{J}^{M}]$, observes the ongoing transmission status, where $P_{J}^{M}$ is the maximum jamming power, with a goal of depleting the energy of the serving BS and the user. For simplicity, we assume that the smart jammer is too far away from the UAV and the backup BS to block them, as shown in Fig. 1.

The path loss of the jammer-UAV link is much higher than that of the jammer-BS₀ link due to a longer distance. Thus, the channel power gain of the jammer-BS₀ link denoted by $h_{4}^{(k)}$ is much larger than the jammer-UAV link denoted by $h_{5}^{(k)}$. For simplicity, the channel power gain vector is denoted by $h^{(k)} = [h_{1}^{(k)}, h_{2}^{(k)}, h_{3}^{(k)}, h_{4}^{(k)}, h_{5}^{(k)}]$, the jamming cost is denoted by $C_J$, and the receiver noise power is denoted by $σ$.

### IV. FAST DQN-BASED UAV RELAY ALGORITHM

In this article, we present a UAV-aided anti-jamming communication scheme for 5G systems, in which the UAV uses fast DQN, a combination of reinforcement learning, deep learning and transfer learning techniques to decide the relay power control policy in each time slot. The fast DQN-based UAV...
Algorithm 1: Fast DQN based UAV relay algorithm

1: Initialize $A$, $S$, $s^{(0)}$, $E$, $M$, and $D$
2: Initialize the CNN weights with previous anti-jamming relay experiences
3: for $k = 1, 2, \ldots$ do
4: $s^{(k)} = [s^{(k-1)}, P_e^{(k-1)}]$
5: if $k \leq E$ then
6: Choose $x^{(k)} \in A$ at random
7: else
8: Input $\varphi^{(k)}$ to the CNN with parameters
9: Update Q-value with the outputs of the CNN
10: Choose $x^{(k)} \in A$ via the $\varepsilon$-greedy algorithm
11: end if
12: if $x^{(k)} > 0$ then
13: Relay the user message to BS$_1$ with a transmit power $x^{(k)}$
14: end if
15: Receive the SINR $\rho_1, \rho_2, \rho_3$, and the BER $P_e$ from core network
16: Evaluate utility $u^{(k)}$
17: $\varphi^{(k+1)} = \{s^{(k-E+2)}, x^{(k-E+2)}, \ldots, x^{(k)}, s^{(k+1)}\}$
18: $D \leftarrow \{\varphi^{(k)}, x^{(k)}, u^{(k)}, \varphi^{(k+1)}\} \cup D$
19: for $m = 1, 2, \ldots, M$ do
20: Choose $\{\varphi^{(m)}, x^{(m)}, u^{(m)}, \varphi^{(m+1)}\} \in D$ randomly
21: end for
22: Update the CNN weights with $\theta^{(k)}$ via SGD
23: end for

The relay algorithm determines the relay power $x^{(k)}$ based on the state denoted by $s^{(k)}$. The state includes the previous SINR vector $r^{(k-1)}$ and the previous BER $P_e^{(k-1)}$ sent by the core network server, with $s^{(k)} = [x^{(k-1)}, P_e^{(k-1)}] \in S$, where $S$ is the feasible state space. As the next state observed by the UAV $s^{(k+1)}$ is independent of the previous state and actions, the optimal relay policy $x^{(k)}$ can be viewed as a MDP and the UAV can use the reinforcement learning technique such as DQN to achieve the optimal communication performance of the 5G system without being aware of the jamming model and the 5G network model.

In this scheme, the UAV obtains the feedback messages from the backup BS that consist of the SINR of the signals received by the serving BS $\rho_1^{(k)}$, the SINR of the signals from the UAV to the backup BS $\rho_2^{(k)}$, and the BER of the user message $P_e^{(k)}$ calculated by the server. The UAV evaluates the SINR of the signal sent from the user to the UAV (i.e., $\rho_3^{(k)}$) with the feedback BER sent by the user.

This UAV relay algorithm uses CNN to compress the state space and applies the hotbooting technique as presented in (7) to initialize the CNN weights denoted by $\theta^{(k)}$ and learning parameters such as the learning rate with previous anti-jamming UAV relay experiences denoted by $\epsilon^{(k)}$. The state sequence denoted by $\varphi^{(k)}$ consists of the previous $E$ states $\{s^{(k-E+2)}, \ldots, s^{(k+1)}\}$ and $E - 1$ relay policies $\{x^{(k-E+2)}, \ldots, x^{(k)}\}$. As shown in Fig. 2 the UAV reshapes the state sequence $\varphi^{(k)}$ into a $n_1 \times n_1$ matrix and input it to the CNN. More specifically, the CNN includes two convolutional (Conv) layers and two fully connected (FC) layers, in which the first Conv layer has $f_1$ filters each with size $n_2 \times n_2$ and stride 1, and the second Conv layer has $f_2$ filters each with size $n_3 \times n_3$ and stride 1. The two Conv layers use the rectified linear units (ReLUs) as the activation function, and the FC layers involve $r_1$ and $r_2$ ReLUs, respectively.

The anti-jamming relay experience is $e^{(k)} = \{\varphi^{(k)}, x^{(k)}, u^{(k)}, \varphi^{(k+1)}\}$ is stored in the memory pool denoted by $D$, with $D = \{e^{(1)}, \ldots, e^{(k)}\}$ for $k \geq 1$. The UAV applies a experience replay technique to extract the anti-jamming relay experience $e^{(k)}$ in a memory pool at each time slot. Upon the updated memory pool, the UAV randomly selects $M$ experiences from the memory pool $D$ in the experience replay, where $M$ is the size of the minibatch. The weights of the CNN $\theta^{(k)}$ are updated according to the stochastic gradient descent (SGD) algorithm, and the minibatch $M$ is chosen similar to (9).

The UAV chooses its relay policy $x^{(k)}$ with the $\varepsilon$-greedy strategy, which depends on the Q-value of each relay policy for the state sequence $\varphi^{(k)}$ from the output of the CNN. According to the selected relay policy $x^{(k)}$, the UAV relays the message of the user with a transmit power $x^{(k)}$ and relay cost $C_U$, to the backup BS, if $x^{(k)} > 0$. The current utility denoted by $u^{(k)}$ depends on the estimated SINR vector $r^{(k)}$ of the message received by the core network server and the relay cost $C_U$, with $u^{(k)} = \max(\rho_1^{(k)}, \min(\rho_2^{(k)}, \rho_3^{(k)})) - x^{(k)}C_U$, and saves such relay experience $e^{(k)}$ in the memory pool $D$, as shown in Algorithm 1.

V. PERFORMANCE ANALYSIS

The performance of the UAV-aided 5G communication systems based on fast DQN can be evaluated in a dynamic UAV relay game, in which the jammer chooses its jamming power $y \in [0, P_{max}^{j}]$, and the UAV decides the relay transmit power $x \in [0, P_{U}^{m}]$. The utility of the UAV depends on the estimated SINR vector of the message received by the server and the relay cost given by

$$u^{(k)} = \max \left( \frac{P_{h_1}^{(k)}}{\sigma + y^{(k)}h_3^{(k)}}, \min \left( \frac{P_{h_2}^{(k)}}{\sigma + y^{(k)}h_4^{(k)}}, \frac{x^{(k)}h_5^{(k)}}{\sigma} \right) \right) - x^{(k)}C_U \tag{1}$$

where the first term is the SINR of the signal received by BS$_1$, and the second term is the SINR of the weaker signal on the user-UAV link and the UAV-BS$_1$ link.

The performance lower bound of Algorithm 1 can be proved to be given by the Nash equilibrium in the UAV relay game. The UAV using the fast DQN-based relay scheme in the dynamic game can reach the performance after sufficiently long time slots. The time index $k$ is omitted in the superscript if no confusion occurs in this section.

Theorem 1. The UAV applying Algorithm 1 in the dynamic UAV relay game under high jamming cost and transmission
cost does not relay the message and the BER of the message with QPSK is given by

\[ P_e = 2 \operatorname{erfc} \left( \frac{P h_1}{\sigma} \right). \]  

**Proof.** Similar to Theorem 1. \( \square \)

**Remark:** The UAV does not relay the user message to the backup BS with high relay cost and good user-BS0 link. On the other hand, the jammer chooses the maximum jamming power to block the communication between BS0 and the user under low jamming cost.

The complexity of the fast DQN-based UAV relay scheme as shown in Algorithm 1 is dependent on the CNN complexity. According to [15], the total complexity of the CNN is \( O \left( \sum_{i=1}^{n} f_i n_i^2 m_i \right) \), where \( f_0 \) is the number of input channels for the CNN, and \( m_i \) is the spatial size of the output feature map of the Conv layer \( i \). As shown in Fig. 2, the first Conv layer has a single input channel and \( f_1 \) filters. Each filter has size \( n_2 \times n_2 \) inputs a \( n_1 \times n_1 \) matrix, and outputs a \((n_1 - n_2 + 1)^2\) feature map. Similarly, the second Conv layer has \( f_1 \) filters each with size \( n_3 \times n_3 \). Each filter in Conv. 2 inputs a \( n_2 \times n_2 \) matrix and outputs a \((n_1 - n_2 + 1)^2\) feature map. Thus, the CNN computation complexity is \( O \left( f_1 (n_2^2 (n_1 - n_2 + 1)^2 + f_2 n_3^2 (n_1 - n_2 - n_3 + 2)^2) \right) \). According to [9], we have \( n_2^2 (n_1 - n_2 + 1)^2 \ll f_2 n_3^2 (n_1 - n_2 - n_3 + 2)^2 \). Thus the computation complexity of the CNN in Algorithm 1 is \( O \left( f_1 f_2 n_3^2 (n_1 - n_2 - n_3 + 2)^2 \right) \).

**VI. SIMULATION RESULTS**

The performance of the UAV-aided 5G system is evaluated in the simulation to match the analysis results in Theorem 1 and Theorem 3, where the transmit power of the user is 50 mW, the noise power is 2 mW, the unit relay cost of the UAV is 0.5 mW, and the unit jamming cost is 1 mW. The CNN
parameters are setting as $n_1 = 8$, $n_2 = 6$, $n_3 = 5$, $f_1 = 20$, $f_2 = 40$, $r_1 = 1000$, $r_2 = 31$, $E = 13$, and the size of the minibatch is $16$ in Algorithm 1 to reduce the BER of the user messages and energy consumption of the 5G system.

As shown in Fig. 3(a), the fast DQN-based UAV relay scheme has lower BER compared with that of the Q-learning and hotbooting PHC-based algorithms. For example, the fast DQN-based UAV relay policy reduces the BER of the user messages by 46.6 percent and 99.7 percent at the 1000-th time slot, compared with hotbooting PHC-based UAV relay scheme and the Q-learning based scheme, respectively. That’s because the fast DQN-based UAV relay scheme uses CNN to compress the high-dimensional state space and applies the experience replay technique to update the CNN parameters in the learning process and retrieve successful anti-jamming relay experiences and thus accelerate the learning process in the dynamic UAV relay game. In addition, our proposed UAV relay scheme converges to the NE of the theoretical results in Theorem 1 and Theorem 3, if the dynamic UAV relay game is long enough.

Our proposed UAV relay scheme achieves less convergence time and lower energy consumption compared with that of the Q-learning and hotbooting PHC-based algorithms, as shown in Fig. 3(b). For instance, the fast DQN-based UAV relay scheme takes about 200 time slots to converge to the optimal policy, which is 83.3% and 88.9% less than that of the hotbooting PHC-based and Q-learning based relay scheme, respectively. In addition, the proposed scheme saves the energy consumption by 33.6 percent and 45.4 percent compared with the hotbooting PHC-based scheme and the Q-learning scheme, respectively. Our proposed scheme accelerates the learning speed of the dynamic UAV relay game, yielding a lower energy consumption compared with the benchmarks.

VII. CONCLUSIONS AND FUTURE WORK

In this article, we have proposed a UAV relay scheme to help 5G resist jamming attacks. We have presented a fast DQN-based relay power allocation algorithm for the UAV to achieve the optimal relay policy without knowing the jamming and 5G network model in the dynamic UAV relay game and developed a hotbooting-PHC based relay algorithm for the UAV without enough computation resources to support deep learning. We have provided the performance bound of the scheme in terms of the BER and the energy consumption and evaluated its computation complexity. These analysis results have been verified via simulations, showing that this relay algorithm can efficiently improve the jamming resistance of 5G systems. For instance, the fast DQN-based relay algorithm reduces the BER by 44.6 percent and saves the energy consumption by 33.6 percent compared with the hotbooting PHC-based UAV relay scheme as presented in [7].

Several challenges have to be addressed to implement the reinforcement learning based UAV relay scheme in practical 5G systems:

The deep learning techniques such as DQN require a UAV to try all the policies in the learning process, and “bad” policies that sometimes cause network disasters for 5G systems. The dangerous UAV exploration can lead to the failure to send critical information for the users and to satisfy the quality of the service by the users. This issue can be addressed by transfer learning and data mining, which explore the anti-jamming communication defense experiences to reduce the random exploration and save the risks of trying dangerous UAV policies at the beginning of the learning process. Backup anti-jamming communication protocols have to be designed and incorporated with the learning based UAV relay scheme to provide reliable and secure 5G communications.

Another important issue for the fast DQN based relay scheme is the state estimation error and delay of the UAV. The current analysis assumes that the UAV can accurately estimate the transmission performance of the 5G systems and evaluate the immediate utility in time, which does not hold for practical UAV-aided 5G systems. Therefore, our future work is focus on the design of the deep reinforcement learning based UAV algorithms that are robust against the state observation delay and the utility evaluation errors to resist jamming attacks for 5G systems.

REFERENCES

[1] Y. Zeng, R. Zhang, and T. J. Lim, “Wireless communications with unmanned aerial vehicles: opportunities and challenges,” IEEE Commun. Magazine, vol. 54, no. 5, pp. 36–42, May 2016.
[2] S. Zhang, Y. Zeng, and R. Zhang, “Cellular-enabled UAV communication: Trajectory optimization under connectivity constraint,” arXiv preprint arXiv:1710.11619, Oct. 2017.

[3] Y. Huo, et al., “Jamming strategies for physical layer security,” IEEE Wireless Commun., vol. 25, no. 1, pp. 148–153, Oct. 2017.

[4] C. Dixon and E. W. Frew, “Optimizing cascaded chains of unmanned aircraft acting as communication relays,” IEEE JSAC, vol. 30, no. 5, pp. 883–898, June 2012.

[5] Y. Zeng, R. Zhang, and T. J. Lim, “Throughput maximization for UAV-enabled mobile relaying systems,” IEEE Trans. Commun., vol. 64, no. 12, pp. 4983–4996, Sept. 2016.

[6] J. Feng, W. E. Dixon, and J. M. Shea, “Positioning helper nodes to improve robustness of wireless mesh networks to jamming attacks,” in Proc. IEEE Global Commun. Conf., pp. 1–6, Singapore, Dec. 2017.

[7] L. Xiao, et al., “UAV relay in VANETs against smart jamming with reinforcement learning,” IEEE Trans. Vehi. Tech., in press.

[8] S. Lv, et al., “Anti-jamming power control game in unmanned aerial vehicle networks,” in Proc. IEEE Global Commun. Conf., pp. 1–6, Singapore, Dec. 2017.

[9] L. Xiao, et al., “User-centric view of unmanned aerial vehicle transmission against smart attacks,” IEEE Trans. Vehi. Tech., vol. 67, no. 4, pp. 3420–3430, Apr. 2018.

[10] J. Dams, M. Hoefler, and T. Kesselheim, “Jamming-resistant learning in wireless networks,” IEEE/ACM Trans. Networking, vol. 24, no. 5, pp. 2809–2818, Oct. 2015.

[11] Y. Wu, B. Wang, and T. J. Lim, “Anti-jamming games in multi-channel cognitive radio networks,” IEEE JSAC, vol. 30, no. 1, pp. 4–15, Dec. 2011.

[12] Y. He, Z. Zhang, and F. R. Yu, “Deep-reinforcement-learning-based optimization for cache-enabled opportunistic interference alignment wireless networks,” IEEE Trans. Vehi. Tech., vol. 66, no. 11, pp. 10433–10445, Sept. 2017.

[13] D. Athukoralage, et al., “Regret based learning for UAV assisted LTE-U/WiFi public safety networks,” in Proc. IEEE Global Commun. Conf., pp. 1–7, Washington, DC, Dec. 2016.

[14] H. Sedjelmaci, S. M. Senouci, and N. Ansari, “Intrusion detection and ejection framework against lethal attacks in UAV-aided networks: A bayesian game-theoretic methodology,” IEEE Trans. Intelligent Transportation Systems, vol. 18, no. 5, pp. 1143–1153, May 2017.

[15] K. He and J. Sun, “Convolutional neural networks at constrained time cost,” in IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 5353–5360, Boston, MA, June 2015.