Energy Efficient Relay in UAV Networks Against Jamming: A Reinforcement Learning Based Approach

Weihang Wang*, Xiaozhen Lu*, Sicong Liu*, Liang Xiao*, Bo Yang†
*Dept. of Information and Communication Engineering, Xiamen University, Xiamen, China. Email: liusc@xmu.edu.cn
†Dept. of Automation, Shanghai Jiao Tong University, Shanghai, China.

***Abstract*—Unmanned aerial vehicle (UAV) networks are vulnerable to jamming attacks because of the high mobility, limited battery and scarce spectrum resources of UAVs. In this paper, we propose a reinforcement learning based UAV relay scheme to improve the anti-jamming capability and save energy consumption of the UAV network. Based on the real-time channel conditions and the historical relay experiences, the proposed scheme enables UAVs to improve the policy of relay power and strategies without knowing the UAV network and channel model. Simulation results show that the proposed UAV relay scheme reduces the bit error rate of the messages and reduces the energy consumption of the UAV network compared with the state-of-the-art benchmark.**

***Index Terms*—Jamming, unmanned aerial vehicles, relay, reinforcement learning.**

I. Introduction

Unlike mobile ad hoc networks and vehicular ad hoc networks (VANETs), unmanned aerial vehicle (UAV) networks are more vulnerable to jamming attacks due to the strict power constraints, the time-varying link quality, higher mobility and dynamic topology with two or three spatial degrees of freedom [1]. In particular, jamming attacks in UAV networks result in the transmission outage and severe system performance degradation [2]. For example, a UAV can be controlled by an attacker with jamming and spoofing signals to land in unintended area after its connection to the operator is blocked and replaced by the attacker [3]. Some UAVs deployed as swarms can be employed as mobile relays to address jamming, and fully exploit the line-of-sight links among the UAVs [4].

Frequency hopping as a traditional anti-jamming communication technique has severe performance degradation in UAV networks due to the difficulty distributing and managing the hopping pattern, the limited battery capacity and spectrum resources of the UAV relay against jamming [5]. Power control is critical for UAVs to address jamming attacks. For instance, the UAV power control scheme as proposed in [6] formulates a Bayesian Stackelberg game to maximize the utility in terms of the throughput against jamming attacks. The joint power control and user scheduling scheme as proposed in [7] uses dynamic programming to improve the anti-jamming performance, including the signal-to-interference-plus-noise ratio and the data rate.

Furthermore, the rapidly time-varying channel conditions and the UAV network model bring great challenges to channel modeling, making it difficult for the conventional power control methods to be applied. Fortunately, reinforcement learning (RL)-based anti-jamming methods do not require to be aware of the network and channel model, and thus have been applied in UAV-aided VANETs [8]. However, this scheme cannot be directly applied in a more dynamic UAV relay network since the anti-jamming performance cannot be guaranteed.

To solve the problems of the state-of-the-art methods, a Reinforcement-learning-based Energy-efficient Anti-jamming Relay (REAR) scheme is proposed in this paper, to improve the energy efficiency and communication reliability for the UAV relay networks. Specifically, the proposed scheme enables each UAV relay to determine its transmit power cooperatively and dynamically based on its state consisting of the received signal strength indicator (RSSI) and the BER of the received message, the channel conditions, the battery level and the past experience of the interactions with the jammer, without being aware of the UAV network model and channel model. A hotbooting method [9] is applied to accelerate the learning process by exploiting anti-jamming experiences from similar UAV relay networks. Simulation results show that the REAR scheme can reduce the energy consumption and improve the reliability compared with the benchmark method [7]. The main contributions of this work are twofold as follows:

1) An energy efficient UAV relay scheme against jamming attacks is devised, which enables intelligent adjustment of the UAV relay power in a dynamically varying environment in the presence of a random jammer.
2) An RL-based power control algorithm is proposed to improve the performance of the multi-UAV relay communication system, without being aware of the network model and channel model, which is further accelerated by applying a transfer learning-based hotbooting method.

The rest of this paper is organized as follows. The related works are reviewed in Section II, and the system model is

This work is supported in part by the National Natural Science Foundation of China under grants 61671396, 61971366, 61901403, 61731012, 61573245 and 61871339, and in part by the Natural Science Foundation of Fujian Province of China under grant 2019J05001. (*Corresponding author: Sicong Liu*).

978-1-7281-5207-3/20/$31.00 ©2020 IEEE

presented in Section III. The RL-based anti-jamming scheme for UAV relay networks is presented in detail in Section IV. Simulation results are reported in Section V, followed by the conclusion drawn in Section VI.

II. Related Works

Jamming attacks and the countermeasures in UAV networks have drawn plenty of research attention. For example, a cooperative anti-jamming scheme proposed in [10] optimizes the channel utilization subject to jamming by regulating the channel access probability of different users. A joint power allocation and scheduling method proposes a dynamic programming algorithm to achieve optimal power control and scheduling in a jammed wireless network [7]. An anti-jamming adaptive beamforming technique applies linear constrained optimization to help with the UAV navigation against jamming with minimum computation complexity [11]. A cooperative anti-jamming relay selection method employs the spatial diversity of the relays to reduce the outage probability by accumulating all the signals from different relays [12].

Without the stringent requirement of being aware of the network and channel model, RL-based methods have been widely applied in anti-jamming. An RL-based power control scheme is proposed for massive multiple-input multiple-output (MIMO) systems to combat against smart jamming [13]. An anti-jamming power control method in [2] uses Q-learning to improve the transmission quality of the UAV network. An RL-based UAV relay scheme against jamming is proposed for VANETs, which determines the relay strategy according to the radio transmission condition [8]. Reinforcement sparse learning based methods as proposed in [14] and [15] can make use of the sparse measurements to mitigate the NB-IoT jamming and impulsive jamming. A deep RL-based power control method is proposed for MIMO wireless optical communications in the presence of wiretapping attacks [16].

III. System Model

We consider a multi-relay enabled network that consists of N UAVs as relay nodes between the source UAV and the destination UAV. The source broadcasts messages intermittently to the intended destination, whose direct link is impacted by jamming. The N relay UAVs can help forward the messages to the destination.

At time slot k, the source broadcasts a message using power $p^{(k)}$. For simplicity, we assume that, the source only broadcasts one message in each time slot. Both the relay UAVs and the destination may receive the message. Both the system model and the proposed strategy for each UAV relay are identical, so the subscript i for relay indexing is omitted for simplicity without loss of generality, except explicitly stated otherwise. Upon receiving the message from the source, a UAV relay decodes the message, measures the RSSI $r^{(k)}$ of the message and estimates the BER $\varrho^{(k)}$. The jamming power received by the relay $j_{\mathrm{U}}^{(k)}$ and the channel gain of the relay-destination link, $h^{(k)}$, also need to be estimated for the policy decision. Since the energy of the UAVs is limited and crucial, the UAV relay has to observe its current battery level $b^{(k)}$ to determine whether there is sufficient energy left to relay the message. The UAV relay may choose to relay the message from the source to the destination using the power of $x^{(k)}$ ranging from zero to the maximum power $P_{\max}$, which is quantized into $M+1$ discrete levels, i.e., $x^{(k)} \in \mathbf{A} = \{mP_{\max}/M\}_{0\leq m\leq M}$, with $\mathbf{A}$ being the action set. Each relay chooses its own relay power independently to improve the system performance.

Due to the broadcast nature of the messages, the destination can receive multiple copies of the message at time slot k, which may be directly sent from the source or relayed by some relay UAVs. Every message received is decoded upon reception and the BER $\rho^{(k)}$ is estimated and recorded. At last, the destination assembles the source address (the address of the source or the relay, from which the message comes directly) and the corresponding BER during this time slot together into one feedback frame and broadcasts it as a feedback.

A flag denoted by $c^{(k)}$ is used to record and evaluate the message delivery state, i.e., whether the message from the source has been successfully delivered to the destination. Only if the destination successfully receives and decodes at least one message can the corresponding BER be found in the feedback, and in this condition the flag $c^{(k)}$ will be set to 0, indicating a successful message delivery. Otherwise, if both the relays and the source have failed in delivering the message, the flag will be set as $c^{(k)} = 1$ as a punishment in the utility for learning and decision.

Once the feedback of the current message is received, the maximum BER $\rho_{\max}^{(k)}$ and the minimum BER $\rho_{\min}^{(k)}$ in the feedback information frame will be extracted. The BER of the message $\rho^{(k)}$ sent from the very UAV relay itself is also picked out. If the BER of the message sent by the relay is not found in the feedback, $\rho_{\max}^{(k)}$ is adopted instead to serve as a conservative estimate in the learning process, i.e., $\rho^{(k)} = \rho_{\max}^{(k)}$.

As far as the jamming model is concerned, a random jammer sending jamming signals in the same frequency as that of the UAVs is considered. The jamming power $j^{(k)}$ is in the range of $[0, j_{\max}]$. Different from a static jammer that makes the jamming power fixed, a random jammer is smarter and more detrimental. Besides, sending jamming signals using random power for a random period may reduce the energy consumption of the jammer, making the attacks last longer.

IV. RL-Based Energy Efficient UAV Relay Scheme Against Jamming

In this section, we present an RL-based energy efficient UAV relay scheme to combat against random jamming attacks. In the framework of RL, the proposed scheme is able to optimize the relay policy for better energy efficiency and reliability of the UAV relay system. We also apply a transfer learning technique, i.e., hotbooting, to further accelerate the learning process.

The pseudocode of the proposed scheme is presented in **Algorithm 1**. At time slot k, the source broadcasts a message. Upon receiving the message, the UAV relay determines a proper power to forward the message to the destination.

Algorithm 1: RL-based energy efficient anti-jamming scheme for UAV relay

```
Initialize parameters: $\mathbf{A}$, $\rho^{(0)}$, $\rho_{\min}^{(0)}$ and $c^{(0)}$
Obtain $\tilde{\mathbf{Q}}$ from similar scenarios based on transfer learning
Initialize Q-function as $\mathbf{Q}$=$\tilde{\mathbf{Q}}$
for $k = 1, 2, \cdots$ do
    UAV relay receives a message from the source
    Measure and observe $\varrho^{(k)}$, $r^{(k)}$, $j_{\mathrm{U}}^{(k)}$, $b^{(k)}$ and $h^{(k)}$
    Formulate $\boldsymbol{s}^{(k)}$ via (1)
    Select $x^{(k)} \in \mathbf{A}$ by $\epsilon$-greedy method
    Estimate $\tilde{b}^{(k)}$ based on $x^{(k)}$ and $b^{(k)}$
    if $\tilde{b}^{(k)} <= \vartheta$ then
        $x^{(k)} = 0$ (insufficient power, mute relaying)
    else
        Relay the message with power $x^{(k)}$
    end
    if Receive the feedback for message then
        $c^{(k)} = 0$ (transmission successful)
        if $\rho^{(k)}$ is contained in the feedback then
            Extract the minimum BER
            in the feedback $\rho_{\min}^{(k)}$
        else
            Extract the maximum and minimum
            BER $\rho_{\max}^{(k)}$ and $\rho_{\min}^{(k)}$
            $\rho^{(k)} = \rho_{\max}^{(k)}$
        end
    else
        $c^{(k)} = 1$ (transmission failed)
    end
    Calculate $u^{(k)}$ via (2)
    Update $Q\left(\boldsymbol{s}^{(k)}, x^{(k)}\right)$ via the Bellman iterative equation
end
```

The destination broadcasts a feedback frame containing the information of the source address and the corresponding BER after decoding the message. Each UAV relay determines its relay power independently based on **Algorithm 1**, so any of the relay UAVs can be investigated to show our proposed method, without loss of generality. Thus, the UAV relay index i is omitted for simplicity in **Algorithm 1**.

Specifically, the metrics reflecting the transmission quality of the UAV network, including the BER of the message $\varrho^{(k)}$ and the RSSI of the message $r^{(k)}$, is observed before relaying. The current received jamming power $j_{\mathrm{U}}^{(k)}$ and the battery level of the UAV relay $b^{(k)}$ are also observed. The channel gain $h^{(k)}$ from the UAV relay to the destination is estimated based on the preambles of the messages [17]. The parameters of the previous time slot including the BER of the relay message $\rho^{(k-1)}$ and the transmission state flag $c^{(k-1)}$ can also be obtained, from the feedback.

Thus, the state $\boldsymbol{s}^{(k)}$ for the UAV relay can be formulated as given by

$$\begin{aligned}\boldsymbol{s}^{(k)} = \Big[&\varrho^{(k)}, r^{(k)}, \rho_{\min}^{(k-1)}, \rho^{(k-1)}, c^{(k-1)},\\ &h^{(k)}, b^{(k)}, j_{\mathrm{U}}^{(k)}\Big],\end{aligned} \tag{1}$$

which contains the BER $\varrho^{(k)}$ and the RSSI $r^{(k)}$ of the message received by the UAV relay, the estimated channel gain $h^{(k)}$, the battery level $b^{(k)}$, the jamming power imposed on the UAV relay $j_{\mathrm{U}}^{(k)}$, together with the parameters from the feedback of the destination at the previous time slot $k-1$ including the minimum BER $\rho_{\min}^{(k-1)}$, the BER of the relayed message $\rho^{(k-1)}$, and the flag $c^{(k-1)}$.

Based on the state $\boldsymbol{s}^{(k)}$, the UAV relay chooses its transmit power $x^{(k)} \in \mathbf{A}$ according to the Q-function table. Specifically, if the UAV chooses to help relay the message, then $x^{(k)} > 0$, otherwise $x^{(k)} = 0$ if relay is denied. During the learning process, the ϵ-greedy method is used to balance exploration and exploitation and prevent stopping at a local minima, where ϵ denotes the probability of exploration. After the relay power $x^{(k)}$ is chosen, the feasibility of the action has to be evaluated before relaying. The remaining battery level $\tilde{b}^{(k)} = b^{(k)} - x^{(k)}$ should be calculated after consuming the relay power of $x^{(k)}$. If the remaining battery level can not support the UAV to keep working normally, i.e., $\tilde{b}^{(k)} < \vartheta$, where ϑ is the minimum battery threshold, then $x^{(k)}$ will be reset to 0 to prevent the relay. Only if the remaining battery level is sufficient will the UAV finally forward the message to the destination.

If the feedback of the current message is received, the flag is set as $c^{(k)} = 0$ to indicate a successful relay. In this case, the BER $\rho^{(k)}$ of the message sent by the UAV as well as the minimum BER $\rho_{\min}^{(k)}$ in the feedback will be recorded. On the other hand, if the corresponding BER is not contained in the feedback, e.g., when the UAV has denied to relay the message or the message it relays fails to reach the destination due to jamming, the minimum BER $\rho_{\min}^{(k)}$ and the maximum BER $\rho_{\max}^{(k)}$ will be recorded, and the maximum BER $\rho_{\max}^{(k)}$ will be regarded as the actual BER for that message as a conservative estimate. If no feedback is received, the flag will be set to $c^{(k)} = 1$ as a punishment to the utility to reflect a transmission outage. Besides, the UAV relay has to measure its energy consumption $E^{(k)}$ by observing the battery level $b^{(k+1)}$ at the end of the current time slot to evaluate the energy efficiency.

Let ω_{b} denote the weight of the BER on the utility function for the learning process, and ω_{c} can be set to an empirically large number as a punishment of transmission outage. The utility $u^{(k)}$ is evaluated upon receiving the feedback from the destination as given by:

$$u^{(k)} = -E^{(k)} - \omega_{\mathrm{b}}\rho^{(k)} - \omega_{\mathrm{c}}c^{(k)}. \tag{2}$$

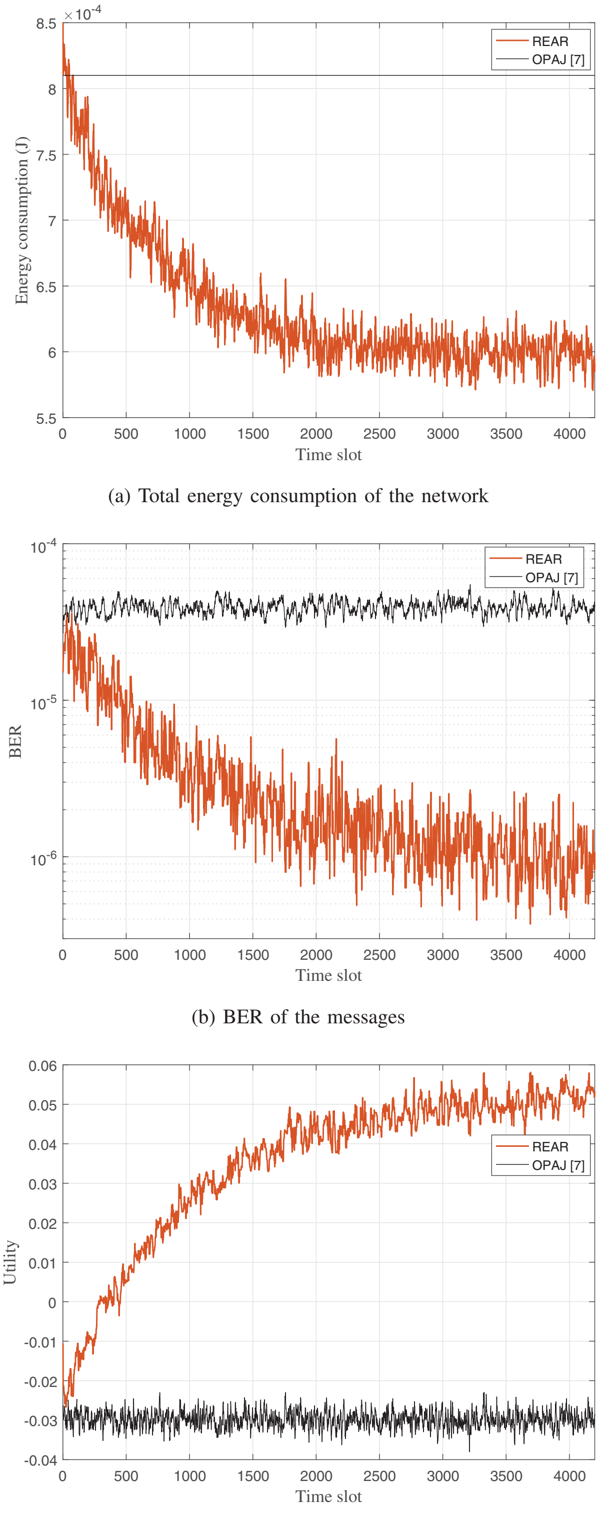
As shown in **Algorithm 1**, the Q-function is exploited and updated via the Bellman iterative equation with the learning rate α and the discount factor γ to maximize the long-term utility. A transfer learning-based hotbooting method is applied to accelerate the learning process, which is implemented by initializing the Q-values with the randomly selected anti-jamming UAV relay experiences $\tilde{\mathbf{Q}}$ from a number of similar UAV relay scenarios.

V. Simulation Results

Simulations are conducted in the scenario including one source UAV, one destination UAV, three relay UAVs, and one moving jammer. The transmission power of the source is set as 0.1 W. The relay UAVs can choose their relay power in the range of [0, 0.1] W which is quantized into 11 discrete levels. Since the air-to-air channel in the system model evaluated in this work is dominated by free-space propagation and less multi-path fading is present compared to the air-to-ground channel [18], we can calculate the path loss using the channel model presented in [19]. There is a jammer moving stochastically near the destination. The jamming power can be changed randomly. Specifically, the jamming power imposed on the destination can be randomly changing among 2 dBm, 6 dBm and 7 dBm. Relays are subject to the jamming power stochastically changing among -7 dBm, -5 dBm and -3 dBm. The system performance can be evaluated by the utility of the network, which is calculated by the minimum BER of the messages received by the destination and the total energy consumption of the UAV network. The learning rate is set as α = 0.5 and the discount factor is set as γ = 0.7. The method of optimal power control against jamming (OPAJ) using fixed optimal relay power [7] is evaluated as a benchmark.

The performance of the RL-based energy efficient UAV relay scheme against jamming is reported in Fig. 1. It is shown that with the proposed REAR scheme, the relay UAVs can successfully learn to improve the energy efficiency of the network. Specifically, the relay UAVs in bad channel conditions select a lower relay power or keep silent to save energy, while other UAVs in better channel conditions choose to relay with higher power. Thus the total energy consumption of the UAV network decreases over time and converges to $6{\times}10^{-4}$ J after 4300 time slots, which saves about 25% energy compared with the start of the learning process. It can also be noted that the communication reliability is significantly improved as the BER decreases to less than 1% of the start of the learning process after 4300 time slots. Moreover, the utility of the UAV relay network is improved by 3 times after 4300 time slots.

Furthermore, it is also shown by Fig. 1 that, the proposed RL-based energy efficient UAV relay scheme significantly outperforms the benchmark OPAJ [7]. For instance, the energy consumption is about 26% lower than that of the benchmark, as shown in Fig. 1(a). The BER of the RL-based scheme is about two orders of magnitude lower than the benchmark, as shown in Fig. 1(b). The energy consumption of the benchmark remains invariant due to using fixed relay power, while the proposed method will significantly decrease the energy consumption with the process of learning. Finally, the utility of the UAV relay network using the proposed scheme is about



(a) Total energy consumption of the network

(b) BER of the messages

(c) Utility of the network

Fig. 1. Performance of the RL-based energy efficient UAV relay scheme averaged over 50 episodes for the UAV network with 3 relays in the presence of a random jammer. The jamming power imposed on the destination changes among 2dBm, 6dBm and 7dBm each time slot.

2 times higher than that of the benchmark, as shown in Fig. 1(c).

VI. Conclusion

In this paper, we have presented an RL-based energy efficient relay scheme for UAV networks in the presence of jamming. This proposed scheme uses a hotbooting method to accelerate the UAV relay learning process. Simulation results for the network that consists of 5 UAVs against a random jammer show that our proposed scheme can improve the utility of the UAV relay network, such as increasing the communication reliability, i.e., decreasing the BER by two orders of magnitude, and reducing the energy consumption by 26% compared with the benchmark OPAJ.

References

[1] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1123–1152, Nov. 2015.

[2] S. Lv, L. Xiao, Q. Hu, X. Wang, C. Hu, and L. Sun, "Anti-jamming power control game in unmanned aerial vehicle networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017.

[3] H. Sedjelmaci, S. M. Senouci, and N. Ansari, "A hierarchical detection and response system to enhance security against lethal cyber-attacks in UAV networks," *IEEE Trans. Systems, Man, and Cybernetics: Systems*, vol. 48, no. 9, pp. 1594–1606, Sept. 2018.

[4] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, Mar. 2019.

[5] L. Zhang, Z. Guan, and T. Melodia, "Cooperative anti-jamming for infrastructure-less wireless networks with stochastic relaying," in *Proc. IEEE Conf. on Computer Commun. (INFOCOM)*, Toronto, Canada, May 2014.

[6] Y. Xu, G. Ren, J. Chen, Y. Luo, L. Jia, X. Liu *et al.*, "A one-leader multi-follower Bayesian-Stackelberg game for anti-jamming transmission in UAV communication networks," *IEEE Access*, vol. 6, pp. 21 697–21 709, Apr. 2018.

[7] S. D'Oro, E. Ekici, and S. Palazzo, "Optimal power allocation and scheduling under jamming attacks," *IEEE/ACM Trans. Networking*, vol. 25, no. 3, pp. 1310–1323, Nov. 2016.

[8] L. Xiao, X. Lu, D. Xu, Y. Tang, L. Wang, and W. Zhuang, "UAV relay in VANETs against smart jamming with reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4087–4097, May 2018.

[9] L. Xiao, Y. Li, C. Dai, H. Dai, and H. V. Poor, "Reinforcement learning-based NOMA power allocation in the presence of smart jamming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3377–3389, Apr. 2018.

[10] L. Zhang, Z. Guan, and T. Melodia, "United against the enemy: Anti-jamming based on cross-layer cooperation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5733–5747, May 2016.

[11] L. Zhang, L. Huang, B. Li, M. Huang, J. Yin, and W. Bao, "Fast-moving jamming suppression for UAV navigation: A minimum dispersion distortionless response beamforming approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7815–7827, Aug. 2019.

[12] P. Gu, C. Hua, R. Khatoun, Y. Wu, and A. Serhrouchni, "Cooperative antijamming relaying for control channel jamming in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7033–7046, Aug. 2018.

[13] Z. Xiao, B. Gao, S. Liu, and L. Xiao, "Learning based power control for mmwave massive MIMO against jamming," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018.

[14] S. Liu, L. Xiao, Z. Han, and Y. Tang, "Eliminating NB-IoT interference to LTE system: A sparse machine learning based approach," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6919–6932, Aug. 2019.

[15] S. Liu, L. Xiao, L. Huang, and X. Wang, "Impulsive noise recovery and elimination: A sparse machine learning based approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2306–2315, Mar. 2019.

[16] L. Xiao, G. Sheng, S. Liu, H. Dai, M. Peng, and J. Song, "Deep reinforcement learning enabled secure visible light communication against eavesdropping," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6994–7005, Oct. 2019.

[17] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996, vol. 2.

[18] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[19] Y. Chen, N. Zhao, Z. Ding, and M.-S. Alouini, "Multiple UAVs as relays: Multi-hop single link versus multiple dual-hop links," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6348–6359, Aug. 2018.