Collaborative UFH-based Anti-jamming Broadcast with Learning

Canhuang Dai*, Dongjin Xu*, Liang Xiao*, Mugen Peng†, Limin Sun‡

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

† Dept. of Information & Communication Engineering, BUPT, China. Email: pmg@bupt.edu.cn

‡ Institute of Information Engineering, Chinese Academy of Sciences, China. E-mail: sunlimin@iie.ac.cn

***Abstract*—Uncoordinated frequency hopping (UFH) technique has been developed to address smart jamming attacks in wireless networks, in which the receiver does not need to know the pre-shared physical-layer secret keys such as the frequency hopping pattern with the transmitter and thus cannot be efficiently blocked by smart jammers that eavesdrop the public control channel of the network. However, collaborative UFH-based broadcast (CUB) that exploits both the spatial and frequency diversity still suffers from a low communication efficiency, because the probability that a receiver happens to use the same channel with a transmitter with a random channel selection is very low. In this paper, we propose a collaborative UFH-based broadcast scheme based on reinforcement learning to further improve the communication efficiency against smart jamming. More specifically, by applying Q-learning algorithm, a radio node can achieve the optimal transmit strategy via trials in the repeated game without being aware of the jamming model and the network model. Simulation results show that the Q-learning based anti-jamming broadcast can significantly decrease the broadcast delay and reduce the total energy cost, compared with the CUB scheme.**

***Index Terms*—Jamming, reinforcement learning, broadcast, uncoordinated frequency hopping**

I. Introduction

The broadcast of a message to the nodes in a wireless network is vulnerable to jamming attacks [1], [2], especially smart jamming attacks. By applying smart radios such as universal software radio peripherals (USRPs), a smart jammer can eavesdrop the public control channel in the network, and send faked or replayed signals to block the ongoing transmissions according to the estimated physical-layer secrets such as the frequency hopping (FH) pattern. Anti-jamming broadcast is important for the security sensitive applications such as the broadcast of emergency alert and has to address the challenges in mobile networks. As a widely-used anti-jamming technique, frequency hopping requires that all the receivers know the FH pattern to obtain the broadcast signal and the distribution of the pre-shared secret key has to be protected from smart jammers and the compromised receivers that might leak the FH pattern to the jammer in the dynamic network.

The uncoordinated frequency hopping technique has been studied to provide anti-jamming communication without requiring any pre-shared FH pattern [3]–[6]. Both the transmitter and the receiver randomly select their channels from the public channel set, and the packet reception is achieved if the two nodes come across the same unblocked channel. One key vulnerability for this anti-jamming technique is a low communication efficiency due to the large number of channels of UFH and the lack of coordination between the transmitter and the receiver. The collaborative UFH-based broadcast (CUB) scheme allows the nodes having receiving the packet relay it to other nodes to accelerate the broadcast process, and thus improves the communication efficiency and jamming resistance [7]. However, CUB that exploits both the spatial and frequency diversity still suffers from a low communication efficiency, because the probability that a receiver happens to use the same channel with a transmitter with a random channel selection is very low.

The broadcast process can be formulated as a finite Markov decision process (MDP), as a smart jammer chooses its jamming strategy according to the last broadcast policy. Therefore, reinforcement learning (RL) techniques such as Q-learning can be applied for the radio node to achieve the optimal broadcast strategy including the transmission duration and the frequency channel selection via trials-and-errors. In this work, we propose a Q-learning based anti-jamming broadcast scheme to further improve the communication efficiency against smart jamming without being aware of the jamming model and the network model. A radio node observes the last jamming strategy and the number of nodes that have received the packet as its state and select its transmitting channels based on the Q-function. With the algorithm iterates, the radio node can improve its utility, which is considered as the number of successful connections and the energy cost. More specifically, a radio node can achieve the optimal transmit strategy based on Q-learning in the repeated game with probability one, if the number of the anti-jamming broadcast interactions is large enough. Simulation results show that our proposed scheme can significantly decrease the broadcast delay and reduce the total energy cost, compared with the CUB scheme.

The main contributions of this paper are summarized as follows:

- We propose a collaborative UFH-based broadcast scheme based on Q-learning to improve the anti-jamming communication efficiency without being aware of the jamming model and the network model.
- Simulations are performed to evaluate the broadcast performance, showing that our proposed scheme outperforms the CUB in [7] by reducing the energy cost and broadcast delay.

The rest of this paper is organized as follows. We review the related work in Section II and present the system model in Section III. We propose the UFH based collaborative broadcast with learning in Section IV and provide simulation results in Section V. Conclusion is drawn for this work in Section VI.

II. Related Work

An erasure coding scheme combined with a one-way authenticator based on bilinear maps as proposed in [4] improves the communication efficiency of UFH. The USD-FH scheme as proposed in [5] further improves the efficiency and robustness by conveying the hopping pattern with UFH and transmitting message with coordinated FH. The collaborative UFH-based broadcast (CUB) as investigated in [7] exploits the node cooperation to enhance the communication efficiency and jamming resistant of the wireless communication. The CUB system was implemented over universal software radio peripherals in [8], and the experiment was conducted to evaluate the anti-jamming performance of the proposed scheme in cognitive radio networks (CRNs). The performance of the UFH used in a large-scale collaborative network is analytically evaluated in [9], demonstrates a significant improvement achieved by cooperative relays and the feasibility of UFH-based schemes in large-scale networks. The power control strategy of an secondary user against a smart jammer under power constraints is formulated as a Stackelberg game in [10], and the Stackelberg equilibrium of the anti-jamming game is derived [10].

Reinforcement learning techniques have been applied to study anti-jamming communications in wireless networks. The reinforcement learning based anti-jamming power control strategy as proposed in [11] improves the SINR of the signals in cooperative CRNs. The Q-learning based spectrum access scheme as investigated in [12] can proactively avoid using the jammed channels in cognitive radio networks. The multi-agent reinforcement learning (MARL) based spectrum sub-band selection policy as designed in [13] improves the communication efficiency in wideband autonomous CRNs. A two-dimensional anti-jamming communication scheme that applies the frequency hopping and user mobility as proposed in [14] uses the deep Q-network algorithm to accelerate the learning speed of the mobile communication system.

III. System Model

In this paper, we consider a wireless network consisting N radio nodes that are all within the transmission range of each other. As shown in Fig. 1, a source node intends to transmit packets to other radio nodes periodically, and the other nodes first enter the receiving mode to receive the packets. Once a radio node receives a packet, it sends an acknowledgement (ACK) signal back and switches into the transmission mode to broadcast the packet to the remaining nodes to accelerate the broadcast process [7]. Each node is labelled with a unique ID and attaches its ID to packets and ACK signals in the packet broadcast process. Without loss of generality, the source node is assumed to be labelled 0. The ID sets of the relays and the receivers at time slot k are denoted by $\Phi^{(k)}$ and $\Psi^{(k)}$, respectively, with $\Phi^{(k)} \cup \Psi^{(k)} = \{1, 2...N-1\}$ and $\Phi^{(k)} \cap \Psi^{(k)} = \emptyset$.

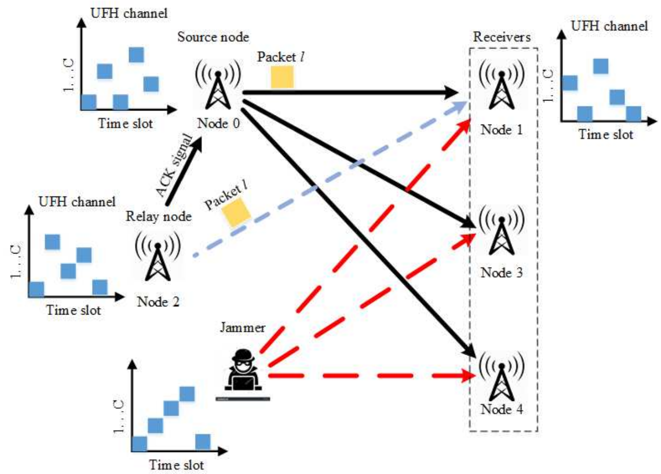


Fig. 1: Illustration of collaborative UFH-based broadcast with learning against jamming attacks.

With the UFH-based broadcast scheme, the packets are transferred over the frequency channels with a high number of repetitions to compensate for the lacking coordination among the transmitters and the receivers. The total number of the frequency channels is denoted by C, with $C > N$. The source node is able to send a packet on multiple channels at a time. The transmission channel selection strategy of the source node at time k is denoted by $\mathbf{c}_0^{(k)} \in S_0 = \{0,1\}^C$, where $\mathbf{c}_0^{(k)}(m) = 1$ if the source node selects the m-th channel, and $\mathbf{c}_0^{(k)}(m) = 0$ otherwise. The source node keeps silence to save energy if $\mathbf{c}_0^{(k)} = \mathbf{0}$. The transmit power of the source node is denoted by P_T. Meanwhile, each node except the source is assumed to send or receive a packet over one channel at a time. The channel selection strategy of other node i, $i \in \{1, 2...N-1\}$, is denoted by $\mathbf{c}_i^{(k)}$, with $||\mathbf{c}_i^{(k)}|| = 1$. The number of the relays changes with time, and is denoted by $\varphi^{(k)}$ at time k, with $0 \leq \varphi^{(k)} \leq N-1$.

The smart jammer is able to block multiple channels simultaneously, and chooses the jamming channels at time k denoted by $\mathbf{c}_J^{(k)} \in S_0$, where $\mathbf{c}_J^{(k)}(m) = 0$ if the m-th channel is blocked, and $\mathbf{c}_J^{(k)}(m) = 1$ otherwise.

According to [7], we adopt the Static Relay Channel selection (StRC) strategy and Adaptive Receiving Channel selection (ARxC) strategy to decide the relay and receiving channel selection of a node, respectively. In the StRC strategy, relay nodes take fixed nonoverlapping channels through the packet broadcast process. We assume that there is a fixed relay channel list chosen for relay nodes, which are known before broadcasting for each radio node. When a relay node starts transmission process, it selects one out of the channel list and sends packets on this channel during current broadcast process. Each receiver is assumed to know the initial relay channel list. After listening to a potential relay channel for a sufficient time, a node can judge whether the channel is clear, active (relaying packets), or jammed. A receiver first continuously sweeps the relay channel list in a random order. When the receiver encounters an active channel, it keeps receiving a packet there until the channel is jammed.

For ease of reference, we summarize the commonly used notation in Table I.

N	Number of nodes
C	Number of frequency channels
P_T	Transmit power
C_h	Frequency hopping loss
$\mathbf{c}_{i/J}^{(k)}$	Channel selection of a node/jammer at time k
$\mathbf{s}^{(k)}$	State at time k
$g_m^{(k)}$	Broadcast gain on channel m at time k
$G^{(k)}$	Total broadcast gain at time k
$u^{(k)}$	Utility of the source node at time k
σ	Weight of the connection gain
$\omega^{(k)}$	Number of received ACKs at time k
α	Learning rate
γ	Discount factor

TABLE I: SUMMARY of SYMBOLS AND NOTATION.

Algorithm 1: Q-learning based channel selection strategy

Initialize α, γ, $Q=\mathbf{0}_{|S_0|(N-1)\times 2|S_0|}$, $V=\mathbf{0}_{|S_0|(N-1)\times 1}$, $\varphi^{(1)}=0$ and $l=1$;
for $k=1,2,\cdots$ **do**
  $\mathbf{s}^{(k)}=[\mathbf{c}_J^{(k-1)},\varphi^{(k)}]$;
  Select the transmitting channels $\mathbf{c}_0^{(k)}$ via Eq. (7);
  for $m=1,2,\cdots C$ **do**
    if $\mathbf{c}_0^{(k)}(m)=1$ **then**
      Transmit Packet p_l on Channel m;
    end
  end
  Observe the IDs of the jammed channels to obtain $\mathbf{c}_J^{(k)}$;
  Obtain $\omega^{(k)}$ ACK signals;
  Obtain $u^{(k)}$;
  Update $Q\left(\mathbf{s}^{(k)},\mathbf{c}_0^{(k)}\right)$ via Eq. (5);
  Update $V\left(\mathbf{s}^{(k)}\right)$ via Eq. (6);
  $\varphi^{(k+1)}=\varphi^{(k)}+\omega^{(k)}$;
  if $\varphi^{(k+1)}=N-1$ **then**
    $l\leftarrow l+1$;
    $\varphi^{(k+1)}=0$;
  end
end

IV. Q-LEARNING BASED COLLABORATIVE BROADCAST USING UFH

In this section, a Q-learning based frequency hopping algorithm is proposed for the source node, different from the mentioned random select strategy in [7]. Without being aware of the jamming model and the relay channel select strategy, a source node can apply Q learning to sense the environment and achieve an optimal frequency hopping strategy via trial-and-error in the dynamic game, as the smart jammer applies the ϵ-greedy policy based on the last transmission of the source node, making a Markov decision process.

Let C_h denote the loss of frequency hopping for the transmission nodes. The broadcast gain at time k on channel m against jamming is denoted by $g_m^{(k)}$, where $g_m^{(k)}=1$ if the packet is received through channel, and $g_m^{(k)}=0$ otherwise [6]. The broadcast gain can be formulated as

$$g_m^{(k)}=\sum_{u\in\Phi^{(k)}}\sum_{v\in\Psi^{(k)}} f\left(\mathbf{c}_0^{(k)}(m)+\mathbf{c}_u^{(k)}(m)\right)\mathbf{c}_v^{(k)}(m)\mathbf{c}_J^{(k)}(m), \tag{1}$$

where

$$f(x)=\begin{cases}0, & x=0,\\ 1, & \text{o.w.}\end{cases} \tag{2}$$

Furthermore, the total broadcast gain on all frequency channels denoted by $G^{(k)}$ is given by

$$G^{(k)}=\sum_{m=1}^{C} g_m^{(k)}. \tag{3}$$

The utility of the source node at time slot k denoted by $u^{(k)}$ depends on the broadcast gain, the energy loss and the frequency hopping loss. Let σ denote the weight of the broadcast gain. The utility of the broadcast node is given by

$$\begin{aligned} u^{(k)} &= \sigma G^{(k)} \\ &-\sum_{m=1}^{C}\mathbf{c}_0^{(k)}(m)\left(P_T+C_h f\left(\mathbf{c}_0^{(k)}(m)-\mathbf{c}_0^{(k-1)}(m)\right)\right). \end{aligned} \tag{4}$$

At time slot k, the source node determines its channel selection strategy $\mathbf{c}_0^{(k)}$ based on the quantity function or Q-function denoted by $Q(\cdot)$, which describes the expected discounted long-term reward for each state-action pair. The state observed by the source node at time slot k, denoted by $\mathbf{s}^{(k)}$, is considered as the channel selection strategy of the jammer at last time slot, $\mathbf{c}_J^{(k-1)}$ and the number of relays, i.e., $\mathbf{s}^{(k)}=[\mathbf{c}_J^{(k-1)},\varphi^{(k)}]$. The value function $V(\mathbf{s})$ is the maximal Q function over the feasible actions at state $\mathbf{s}$. The Q function and the value function of the source node are updated by:

$$\begin{aligned} Q\left(\mathbf{s}^{(k)},\mathbf{c}_0^{(k)}\right) \leftarrow & (1-\alpha)Q\left(\mathbf{s}^{(k)},\mathbf{c}_0^{(k)}\right) \\ & +\alpha\left(u^{(k)}+\gamma V\left(\mathbf{s}^{(k+1)}\right)\right), \end{aligned} \tag{5}$$

$$V\left(\mathbf{s}^{(k)}\right)=\max_{\mathbf{c}_0\in S_0} Q\left(\mathbf{s}_i^{(k)},\mathbf{c}_0\right), \tag{6}$$

where $\alpha\in(0,1]$ is the learning rate representing the weight of Q value at current state-action pair, and $\gamma\in[0,1]$ is the discount factor indicating the greed of the source node regarding the future utility.

To balance the exploitation and exploration in the learning process, the ϵ-greedy strategy is applied to choose the transmission channels of the source node as

$$\Pr(\mathbf{c}_0^{(k)} = \mathbf{\Theta}) = \begin{cases} 1-\epsilon, & \mathbf{\Theta} = \arg\max_{\mathbf{c}_0 \in S_0} Q\left(\mathbf{s}^{(k)}, \mathbf{c}_0\right), \\ \frac{\epsilon}{|S_0|}, & \text{o.w.} \end{cases} \quad (7)$$

The source node sends packet p_l on the selected channels, receives ω of ACK signals at the same time slot and adds it to the number of accumulated ACK signals, which is also the number of instant relays, φ. Once the source node gets ACK signals from all the other radio nodes, i.e., $\varphi = N-1$, it resets φ to 0 and starts broadcasting next packet p_{l+1}. Meanwhile, all nodes switch into the receiving mode again. The detailed process is shown in algorithm 1.

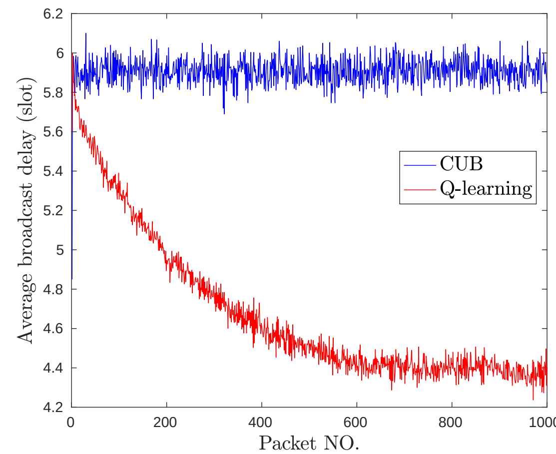
V. Simulation Results

Simulations have been performed to evaluate the performance of the Q-learning based channel selection strategy in a collaborative broadcast network with $N = 8$ and $C = 12$. If not specified otherwise, we set $P_T = 10$ mW, $C_h = 0.2$ mW and $\sigma = 20$. As a benchmark, we consider the CUB scheme proposed in [7], in which the source node randomly chose its channels and the other radio nodes used the same StRC and ARxC strategies for relaying and receiving.

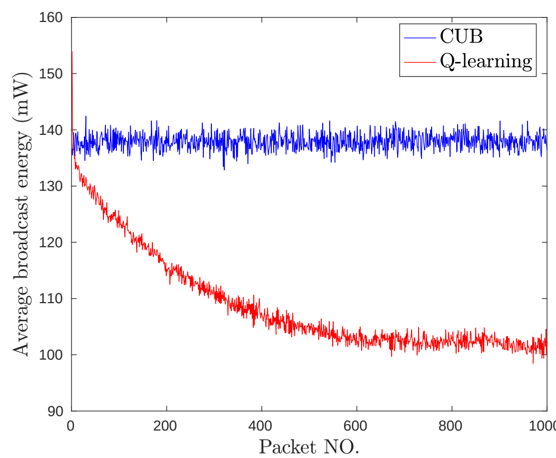
As shown in Fig. 2, the proposed Q-learning based frequency hopping strategy outperforms the CUB strategy with less broadcast delay and energy cost and higher utility. For example, as shown in Fig. 2(a), the Q-learning based strategy decreases the average delay for broadcasting a packet by 20.3% after 400 packets. The optimal broadcast delay of the Q-learning based strategy is 26.7% lower than the CUB strategy, because the transmitters and the receivers can come across on the same unjammed channel with a higher probability. In Fig. 2(b), due to the optimal transmission duration control based on the Q-learning strategy, the energy cost of the source node decreases by 20.2% after 400 packets. And its optimal energy cost is 23.8% lower than the CUB strategy. At the 5000-th time slot, the utility of the source node with Q-learning based strategy increases 64.2%, compared with the CUB scheme, as shown in Fig. 2(c).

VI. Conclusion

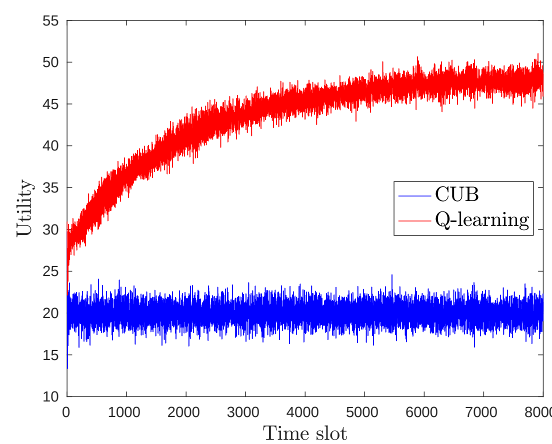
In this paper, we have proposed a collaborative UFH-based broadcast with Q-learning to address smart jammers that can eavesdrop the public control channel. The source node in the collaborative broadcast uses Q-learning to achieve the optimal transmission duration and frequency channel and this scheme is applicable to the dynamic game network with time-varying jamming strategy to improve its communication performance. Simulation results show that the proposed broadcast scheme reduces the broadcast delay and energy cost by 26.7% and 23.8%, respectively, compared with the benchmark CUB strategy, in a network with 8 nodes against the ϵ-greedy policy based jamming policy.



(a) Average broadcast delay for each packet



(b) Average energy cost for each packet



(c) Utility

Fig. 2: Performance of the anti-jamming collaborative broadcast in the dynamic network with $N = 8$, $C = 12$, $P_T = 10$ mW, $C_h = 0.2$ mW and $\sigma = 20$.

VII. Acknowledgement

This work was supported in part by the National Science Foundation of China (61671396).

References

[1] W. Xu, W. Trappe, Y. Zhang, and T. Wood, “The feasibility of launching and detecting jamming attacks in wireless networks,” in *Proc. ACM Int. Symp. Mobile Ad Hoc Networking and Computing (MobiHoc)*, pp. 46–57, Urbana Champaign, IL, May 2005.

[2] M. Li, I. Koutsopoulos, and R. Poovendran, “Optimal jamming attack strategies and network defense policies in wireless sensor networks,” *IEEE Transactions on Mobile Computing*, vol. 9, no. 8, pp. 1119–1133, Aug. 2010.

[3] M. Strasser, C. Pöpper, S. Capkun, and M. Cagalj, “Jamming-resistant key establishment using uncoordinated frequency hopping,” in *Proc. IEEE Symp. Security and Privacy*, pp. 64–78, Oakland, CA, May 2008.

[4] M. Strasser, C. Pöpper, and S. Čapkun, “Efficient uncoordinated fhss anti-jamming communication,” in *Proc. ACM Int. Symp. Mobile Ad Hoc Networking and Computing (MobiHoc)*, pp. 207–218, New Orleans, LA, May 2009.

[5] A. Liu, P. Ning, H. Dai, and Y. Liu, “Usd-fh: Jamming-resistant wireless communication using frequency hopping with uncoordinated seed disclosure,” in *Proc. IEEE Int. Conf. Mobile Ad-Hoc and Sensor Systems*, pp. 41–50, San Francisco, CA, Nov. 2010.

[6] Q. Wang, P. Xu, K. Ren, and X. Y. Li, “Towards optimal adaptive ufh-based anti-jamming wireless communication,” *IEEE J. Selected Areas in Communications*, vol. 30, no. 1, pp. 16–30, Jan. 2012.

[7] L. Xiao, H. Dai, and P. Ning, “Jamming-resistant collaborative broadcast using uncoordinated frequency hopping,” *IEEE transactions on Information Forensics and Security*, vol. 7, no. 1, pp. 297–309, Feb. 2012.

[8] G. Chen, Y. Li, L. Xiao, and L. Huang, “Collaborative anti-jamming broadcast with uncoordinated frequency hopping over usrp,” in *IEEE Vehicular Technology Conference*, pp. 1–6, Glasgow, UK, May 2015.

[9] C. Li, H. Dai, L. Xiao, and P. Ning, “Communication efficiency of anti-jamming broadcast in large-scale multi-channel wireless networks,” *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5281–5292, 2012.

[10] L. Xiao, T. Chen, J. Liu, and H. Dai, “Anti-jamming transmission stackelberg game with observation errors,” *IEEE Communications Letters*, vol. 19, no. 6, pp. 949–952, 2015.

[11] L. Xiao, Y. Li, J. Liu, and Y. Zhao, “Power control with reinforcement learning in cooperative cognitive radio networks against jamming,” *J. Supercomputing*, vol. 71, no. 9, Apr. 2015.

[12] F. Slimeni, B. Scheers, Z. Chtourou, and V. Le Nir, “Jamming mitigation in cognitive radio networks using a modified Q-learning algorithm,” in *IEEE Int. Conf. Military Communications and Information Systems (ICMCIS)*, pp. 1–7, Cracow, Poland, May 2015.

[13] M. A. Aref, S. K. Jayaweera, and S. Machuzak, “Multi-agent reinforcement learning based cognitive anti-jamming,” in *IEEE. Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, San Francisco, CA, Mar. 2017.

[14] G. Han, L. Xiao, and H. V. Poor, “Two-dimensional anti-jamming communication based on deep reinforcement learning,” in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2087–2091, New Orleans, LA, Mar. 2017.