DQN-based Power Control for IoT Transmission Against Jamming

Ye Chen*, Yanda Li*, Dongjin Xu*, Liang Xiao*†
*Department of Communication Engineering, Xiamen University, Xiamen, China. Email: lxiao@xmu.edu.cn
†National Mobile Communications Research Laboratory, Southeast University, Nanjing, China.

***Abstract*—Internet of Things (IoTs) have to address jammers, with goal to interrupt the communication of the energy-constrained IoT devices and sometimes even cause denial-of-service attacks. In this paper, we propose a deep reinforcement learning based power control scheme for IoT devices to improve the transmission efficiency and save energy. This scheme depends on the current IoT transmission status and the jamming strength and applies deep Q-network (DQN) to determine the transmit power without being aware of the IoT topology and the jamming model. This scheme is implemented on the universal software radio peripherals for the anti-jamming communication performance evaluation. Experimental results show that this scheme improves the signal-to-interference-plus-noise of the IoT signals compared with the benchmark Q-learning based power control scheme against jamming.**

***Index Terms*—Jamming, DQN, universal software radio peripherals, IoT, power control.**

I. Introduction

The Internet of Things (IoT) has provided the various applications with the interconnected devices such as sensors, actuators, and mobile phones, which communicate with each other to exchange data and carry out tasks [1]. The security issues of the IoT have been critical to ensure the quality of the communication. However, the IoT is seriously threatened by various attacks, especially the jamming attack [1]–[3], due to the heterogeneous and large-scale nature and the limited energy resource. The jamming attack can be easily launched by the radio devices to interrupt the legitimate communication and even cause the denial-of-service attack [4]–[6].

Anti-jamming techniques such as power control have been studied in the wireless communication [7]. For example, the power control schemes proposed in [8] estimate the channel conditions and adjust the transmit power to override the jamming signal and thus improve the communication quality. However, the IoT device has to avoid high energy consumption regarding its limited battery capacity and transmit power. Moreover, the power control strategy depends on the channel conditions and the jamming model including the jamming channel and the jamming power, which are difficult to estimate accurately with the dynamic topology of the IoT and the fast-varying wireless environments [9]–[11].

This work was supported in part by Natural Science Foundation of China under Grant 61671396, in part by the open research fund of National Mobile Communications Research Laboratory, Southeast University (No. 2018D08) and in part by Science and Technology Innovation Project of Foshan City, China (Grant No. 2015IT100095).

In this paper, we investigate the anti-jamming IoT communication against the jammer who can observe the communication state of the IoT devices and accordingly choose its jamming strategy. The transmit power control policy of the IoT device will impact on the future jamming strategy, thus the interactions between the device and the jammer can be formulated as a finite Markov decision process (MDP). Therefore, reinforcement learning (RL) techniques such as Q-learning can be applied for the IoT devices to achieve the optimal power control strategy without being aware of the channel variation information and the jamming model.

However, the dynamic IoT topology usually results in the large-scale state space for the RL-based power control schemes and thus cause the learning performance degradation, i.e. the learning speed is significantly decreased. Specifically, the widely used model-free algorithm, Q-learning will suffer a long convergence time to achieve the optimal strategy with enormous feasible observation states and even fail to converge [12]. Therefore, we propose a deep Q-network (DQN) based power control scheme for the IoT device, which uses the deep neural network such as convolutional network to accelerate the learning speed.

We implement the DQN-based power control scheme over the universal software radio peripherals (USRPs), which is programmable using GNURadio platform to provide the signal processing and the transmission module. The details of each transmission and computation module are presented in the following. The previous works mainly focus on the theoretical analysis and simulation experiment over personal computer. In contrast, we perform the experiments in the realistic scenarios and demonstrate the effectiveness of the proposed scheme under the hardware constraint. The experiment results show that the proposed DQN-based power control strategy improves the communication efficiency compared with the Q-learning based strategy.

The remainder of this paper is organized as follows. We review related work in Section II, and present the anti-jamming communication game in Section III. We propose the DQN-based power control strategy in Section IV. We provide the experiments results in Section V, and draw conclusions in Section VI.

II. Related Work

To resist interference from jammers in wireless radio communication, some useful techniques can be applied. For instance, spread spectrum techniques such as frequency hopping and direct-sequence spread spectrum can be applied to address jamming attacks [13]–[17], anti-jamming power control techniques [18], [19] are useful in addressing jamming, and multiple-input multiple-output (MIMO) techniques are also applied to improve the performance [20].

Reinforcement learning algorithm makes it feasible that an agent gradually achieves an optimal policy via trials in Markov decision process over time or steps. It is important and useful in decision-making. The Q-learning is applied in channel allocation scheme in [21] to solve for an optimal channel access strategy in the cognitive radio networks. The multi-agent reinforcement learning is applied in [22] to find the optimal control channel allocation strategy to combat the control channel jamming. The multi-agent reinforcement learning is used in [23] in the power control strategy to achieve the higher learning speed in the energy harvesting communication system to deal with intelligent adversaries.

III. Anti-Jamming Communication System Model

In this paper, we investigate the anti-jamming IoT communication against the jammer who can observe the communication state of the IoT devices and accordingly choose its jamming strategy. At time slot k, the transmitter sends messages to the receiver at the transmit power denoted by $x^{(k)} \in [0, P]$, where P is the maximum power of devices. The channel gain from the transmitter to the receiver is denoted by $h_t^{(k)}$. The cost of the energy loss of transmitting messages is denoted by $C_t^{(k)}$. The utility of the transmitter at time slot k is denoted by $u_t^{(k)}$.

At time slot k, the jammer tries to inject jamming signals at the jamming power denoted by $y^{(k)} \in [0, P]$ to block the ongoing local transmission, resulting in low SINR and a packet loss at the receiver, where P is the maximum power of jammer. The channel gain from the jammer to the receiver is denoted by $h_j^{(k)}$. The cost of the energy loss of injecting jamming signals is denoted by $C_j^{(k)}$. The utility of the jammer at time slot k is denoted by $u_j^{(k)}$.

The noise power at the receiver is denoted by $n^{(k)}$. $\mathrm{SINR}^{(k)}$ denotes the SINR from transmitter to receiver, which is influenced by channel quality and interference of jammer. We can define SINR basing on the signal power, channel gain and the noise power as the following formula,

$$\mathrm{SINR}^{(k)} = \frac{h_t x^{(k)}}{h_j y^{(k)} + n^{(k)}}. \tag{1}$$

The utility of the transmitter and the jammer can be defined based on SINR and the cost of energy loss, as the following formula,

$$u_t^{(k)} = \mathrm{SINR}^{(k)} - C_t^{(k)} x^{(k)}. \tag{2}$$

$$u_j^{(k)} = -\mathrm{SINR}^{(k)} - C_j^{(k)} y^{(k)}. \tag{3}$$

TABLE I: List of Notations

Symbol	Meaning
$x^{(k)}$	Transmit power at time slot k
$y^{(k)}$	Jamming power
$h_t^{(k)}$	Channel power gain of the TX-RX link
$h_j^{(k)}$	Channel power gain of the jammer
$C_t^{(k)}$	Energy cost of the transmitter
$C_j^{(k)}$	Energy cost of the jammer
$\mathrm{SINR}^{(k)}$	SINR of the signal
$n^{(k)}$	Noise power
$u_t^{(k)}$	Utility of the transmitter
$u_j^{(k)}$	Utility of the jammer
$s^{(k)}$	System state
$a^{(k)}$	Transmit power
$\varphi^{(k)}$	Input of the CNN
W	Size of the state-action pairs in $\varphi^{(k)}$
γ	Discount factor
$\theta^{(k)}$	Parameters of the CNN

At each time slot, the jammer chooses the jamming power with the greedy strategy. The jammer observes the SINR last time, that is, $\mathrm{SINR}^{(k-1)}$, then chooses the jamming power which can reduce the SINR and maximize the utility $u_j^{(k)}$. At the same time, the transmitter chooses the transmit power by DQN algorithm to improve SINR and reduce energy loss.

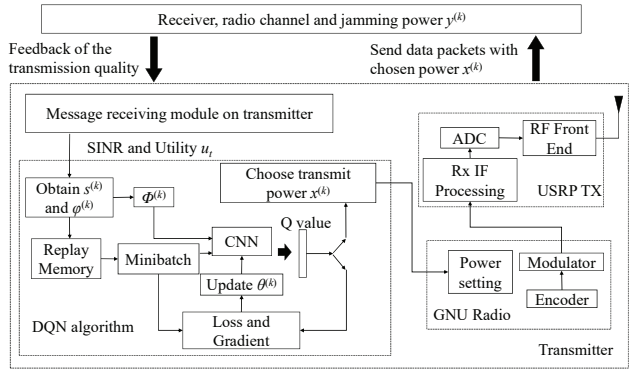
The summary of notations used in this paper is listed in Table I.

IV. DQN-Based Anti-Jamming Communication System

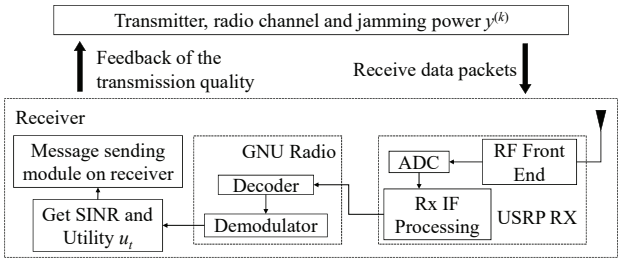
The optimal transmit power at the IoT devices relies on observing the topology structure, the channel model and the jamming model, which are complex and difficult to estimate. Enormous state space results in 'curse of dimensionality' at normal RL algorithms such as Q-learning. Therefore, we apply a power control scheme based on the DQN algorithm for the IoT devices to accelerate the learning speed and address the anti-jamming problem more efficiently and derive the optimal transmit power control scheme without being aware of the channel model and the jamming model.

As is shown in Fig. 1, we implement this DQN-based anti-jamming power control system over USRPs. Fig. 1(a) shows the transmitter on USRP. The transmit data packets are prepared and the transmitter chooses and sets the transmit power $x^{(k)}$ based on DQN algorithm. The data packets are then encoded and modulated by encoder and modulator, and submitted to the IF processing and ADC module, and then transmitted by RF front end of the transmitter with power $x^{(k)}$.

Fig. 1(b) shows the receiver on USRP. The data packets and the jamming signals are received by RF front end of the receiver, via ADC module and IF processing, and submitted



(a) Transmitter



(b) Receiver

Fig. 1: Illustration of the DQN-based IoT power control in the USRP based testbed.

to decoder and demodulator. After that, the SINR and utility are obtained and feed back to the transmitter. The transmitter is able to choose its strategy basing on the feedback next time.

The system state is denoted by $s^{(k)}$ and the action is denoted by $a^{(k)}$ at time slot k. The transmitter judges the current state by observing the feedback from the environment and determines its action.

As is shown in Fig. 1(a), a convolutional neural network (CNN) is required in DQN. System state $s^{(k)}$ at time slot k is the input in CNN. In our model, we set $\mathrm{SINR}^{(k-1)}$ at time slot $k-1$ as the current system state $s^{(k)}$. In order to give more information to the input to speed learning rate up and derive a better policy, the input dimension must be increased. We expand state into state-action pairs sequence, denoted by $\varphi^{(k)}$ at time slot k, which consists of the current state and the last W state-action pairs, i.e.

$$\varphi^{(k)} = (s^{(k-W)}, a^{(k-W)}, ..., s^{(k-1)}, a^{(k-1)}, s^{(k)}). \quad (4)$$

Our CNN consists of two convolutional layers and two fully connected (FC) layers. The first convolutional layer includes 20 filters each with length 3 and stride 1. The second convolutional layer has 40 filters each with length 2 and stride 1. There is activation function, rectified linear unit (ReLU), following both convolutional layers. The first FC layer involves 180 linear units and the second FC layer has $N+1$ units, which represent the action set. The weights of the four layers in the CNN at time slot k are denoted by $\theta^{(k)}$.

Algorithm 1 DQN based power control system.

Initialize θ, γ, W, $x^{(0)}$, $y^{(0)}$;
Obtain $\mathrm{SINR}^{(0)}$ via Eq. (1) as $s^{(1)}$;
for $k = 1, 2, 3, ...$ **do**
  if $k \leqslant W$ **then**
    Set transmit power $x^{(k)} \in \{0, 1, 2, ..., N\}$;
  else
    Set $\varphi^{(k)}$ via Eq. (4) as input of CNN;
    Get output of CNN $Q(\varphi^{(k)}, x^{(k)}, \theta^{(k)})$ as the estimated Q values;
    Choose transmit power $x^{(k)}$ with ϵ-greedy strategy;
  end if
  Send data packets with power $x^{(k)}$;
  Obtain $\mathrm{SINR}^{(k)}$ via Eq. (1) as $s^{(k+1)}$;
  Evaluate the utility of the transmitter $u_t^{(k)}$ via Eq. (2);
  Update $\theta^{(k)}$ by minimize Eq. (5);
end for

At each step, the current system state is input into the CNN and then the output from CNN is obtained, i.e. an array of the estimated Q values for each action, denoted by $Q(\varphi^{(k)}, x, \theta^{(k)})$.

Notice that the one who applies DQN algorithm is the devices. The devices have decided its action, that is, transmit power at this time slot. Executing the chosen action, the SINR at this time slot will be measured and calculated at the end of this time slot.

The mean-squared error between estimated Q value and excepted Q value, i.e. the loss function chosen, is minimized with minibatch updates as following

$$L(\theta^{(k)}) = (Q' - Q(\varphi^{(k)}, x, \theta^{(k)}))^2. \quad (5)$$

where Q' is the excepted Q function given by

$$Q' = \mathrm{SINR}^{(k)} + \gamma \max_{x'} Q(\varphi^{(k+1)}, x', \theta^{(k)}). \quad (6)$$

The second term on the right-hand-side of Eq. (6) means that, when next state is input, estimated Q value for each action x is calculated and the highest returned, with corresponding action x. After that, the parameters $\theta^{(k)}$ of the CNN are updated and optimized at the end of each time slot by stochastic gradient descent algorithm.

V. Experiment Results

We implement this anti-jamming system over USRPs to evaluate the performance of it. As is shown in Fig. 2. Our devices consist of several laptops with Ubuntu14.04 and GNU Radio, three USRP boards with two antennas at each USRP board, a transmitting antenna TX and a receiving antenna

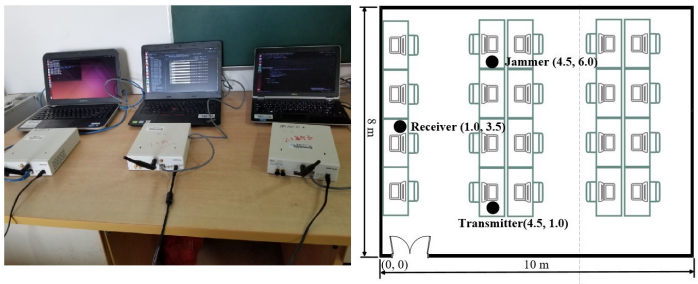


Fig. 2: Settings in the experiments

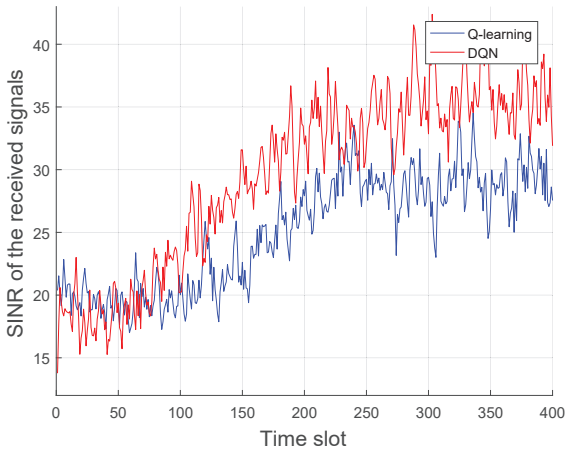
RX. The message transmitter, receiver and the jammer are respectively set on one USRP.

The reactive jammer calculates its utility to choose the jamming power based on the last transmit power of the transmitter. At each time slot, the jammer observes $\mathrm{SINR}^{(k-1)}$, then chooses the jamming power which can maximize $u_j^{(k)}$ with the greedy strategy.At each time slot, the transmitter chooses and sets the transmit power and sends data packets on the radio channel. As a typical example, the transmit power is discretized from 0 dBm to 15 dBm out of $N = 50$, with cost of energy loss $C_t = C_j = 1$ and discount factor $\gamma = 0.5$. Q-learning algorithm and DQN algorithm are respectively implemented at this system.

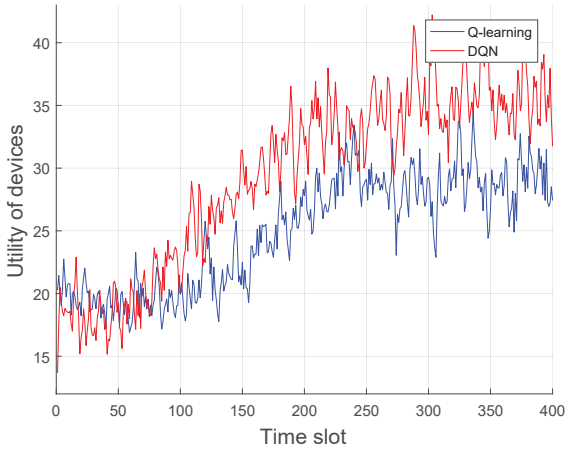
The performances of the anti-jamming system based on different power control algorithm is shown in Fig. 3. The DQN-based power control scheme outperforms that on Q-learning, with higher utility and faster learning speed. For instance, the SINR increases from 19.5 at the beginning to 29.6 by 51.8% at approximate time slot 240 at the Q-learning system, while the SINR increases from 18.8 at the beginning to 34.1 by 81.4% at approximate time slot 210 at the DQN-based system, which is 15.2% higher than that on Q-learning with faster convergence rate. The utility of devices increases from 18.5 at the beginning to 28.8 and at approximate time slot 240 at the Q-learning system, while the utility increases from 17.2 at the beginning to 33.9 at approximate time slot 210 at the DQN-based system, which is 17.7% higher than the Q-learning-based scheme. The utility of jammer decreases from -18.1 at the beginning to -29.8 at the Q-learning system, while the utility decreases from -17.6 at the beginning to -35.2 at the DQN-based system, which is 18.1% lower than that on Q-learning.

VI. Conclusion

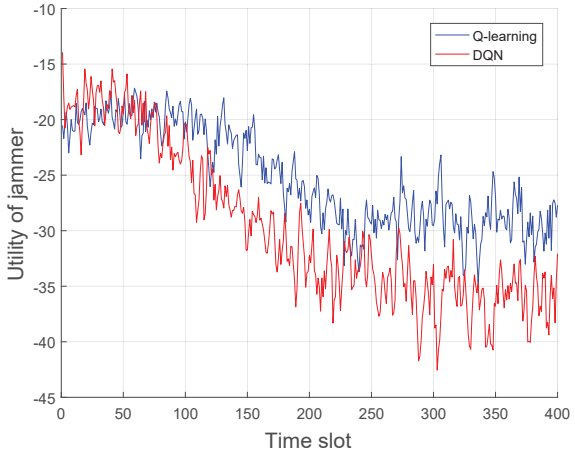
In this paper, we have presented an anti-jamming IoT power control scheme against jamming, which applied a DQN to accelerate the learning speed for the case with a large number of SINR quantization levels and jamming levels. Experiments based on the test bed using USRPs have been performed to evaluate the anti-jamming transmission performance, showing that this scheme can improve the average SINR of the IoT signals against jamming compared with the standard Q-learning-based strategy. For instance, this scheme increases the SINR



(a) SINR of the received signals



(b) Utility of devices



(c) Utility of jammer

Fig. 3: Performances of the IoT power control against reactive jamming in an experiment as shown in Fig. 2, with $N = 50$, $C_t = C_j = 1$ and $\gamma = 0.5$.

of the signals by 51.8% and the utility the IoT devices by 81.4% after 240 time slots, respectively, which are 15.2% and 17.7% higher than the benchmark scheme.

REFERENCES

[1] N. Namvar, W. Saad, N. Bahadori, *et al.*, "Jamming in the internet of things: A game-theoretic perspective," in *IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–6, Washington, DC, Dec. 2016.

[2] J. Gubbi, R. Buyya, S. Marusic, *et al.*, "Internet of things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, Sept. 2013.

[3] G. Chen, Y. Li, L. Xiao, *et al.*, "Collaborative anti-jamming broadcast with uncoordinated frequency hopping over USRP," in *Proc. IEEE Vehicular Technology Conf. (VTC Spring)*, pp. 1–6, Glasgow, UK, May 2015.

[4] L. Xiao, *Anti-Jamming Transmissions in Cognitive Radio Networks*. Springer, 2015.

[5] C. Li, H. Dai, L. Xiao, *et al.*, "Communication efficiency of anti-jamming broadcast in large-scale multi-channel wireless networks," *IEEE Trans. Signal Processing*, vol. 60, no. 10, pp. 5281–5292, Oct. 2012.

[6] X. Lu, D. Xu, L. Xiao, *et al.*, "Anti-jamming communication game for UAV-Aided VANETs," in *IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–6, Singapore, Dec. 2017.

[7] L. Xiao, T. Chen, J. Liu, *et al.*, "Anti-jamming transmission stackelberg game with observation errors," *IEEE Commun. Letters*, vol. 19, no. 6, pp. 949–952, June 2015.

[8] L. Xiao, J. Liu, Q. Li, *et al.*, "User-centric view of jamming games in cognitive radio networks," *IEEE Trans. Information Forensics and Security*, vol. 10, no. 12, pp. 2578–2590, Dec. 2015.

[9] Y. Xiao, V. Rayi, B. Sun, *et al.*, "A survey of key management schemes in wireless sensor networks," *J. Computer Commun.*, vol. 30, pp. 2314–2341, Sept. 2007.

[10] X. Du, Y. Xiao, M. Guizani, *et al.*, "An effective key management scheme for heterogeneous sensor networks," *Elsevier Ad Hoc Networks*, vol. 5, no. 1, pp. 24–34, Jan. 2007.

[11] H. Lu, J. Li, and M. Guizani, "Secure and efficient data transmission for cluster-based wireless sensor networks," *IEEE Trans. Parallel and Distributed Systems*, vol. 25, pp. 750–761, Mar. 2014.

[12] V. Mnih, K. Kavukcuoglu, D. Silver, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[13] M. Strasser, C. Pöpper, S. Capkun, *et al.*, "Jamming-resistant key establishment using uncoordinated frequency hopping," in *Proc. IEEE Symposium Security and Privacy*, pp. 64–78, Oakland, CA, May 2008.

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

[15] L. Xiao, G. Han, D. Jiang, H. Zhu, Y. Zhang, and H. V. Poor, "Two-dimensional anti-jamming mobile communication based on reinforcement learning," *arXiv preprint arXiv:1712.06793*, 2017.

[16] C. Dai, D. Xu, L. Xiao, *et al.*, "Collaborative UFH-based anti-jamming broadcast with learning," in *Proc. IEEE/CIC Int. Conf. Commun.*, Qingdao, China, Oct. 2017.

[17] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxylakis, "A survey on security threats and detection techniques in cognitive radio networks," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 1, pp. 428–445, Jan. 2013.

[18] L. Xiao, Y. Li, J. Liu, *et al.*, "Power control with reinforcement learning in cooperative cognitive radio networks against jamming," *J. Supercomputing*, vol. 71, no. 9, pp. 3237–3257, Apr. 2015.

[19] L. Xiao, Q. Li, T. Chen, *et al.*, "Jamming games in underwater sensor networks with reinforcement learning," in *IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–6, San Diego, CA, Dec. 2015.

[20] Q. Yan, H. Zeng, T. Jiang, *et al.*, "MIMO-based jamming resilient communication in wireless networks," in *IEEE Int. Conf. Computer Commun. (INFOCOM)*, pp. 2697–2706, Toronto, Canada, May 2014.

[21] Y. Gwon, S. Dastangoo, C. Fossa, *et al.*, "Competing mobile network game: Embracing antijamming and jamming strategies with reinforcement learning," in *Proc. IEEE Conf. Commun. and Network Security (CNS)*, pp. 28–36, Washington, DC, Oct. 2013.

[22] B. F. Lo and I. F. Akyildiz, "Multiagent jamming-resilient control channel game for cognitive radio ad hoc networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1821–1826, Ottawa, Canada, Jun. 2012.

[23] X. He, H. Dai, and P. Ning, "Faster learning and adaptation in security games by exploiting information asymmetry," *IEEE Trans. Signal Processing*, vol. 64, no. 13, pp. 3429–3443, July 2016.