Defeating Jamming Attacks with Ambient Backscatter Communications

Nguyen Van Huynh, Diep N. Nguyen, Dinh Thai Hoang, Eryk Dutkiewicz, Markus Mueck, and Srikathyayani Srikanteswara

***Abstract*—While the existing anti-jamming solutions tend to "escape" the attacks by finding another communication channel or adapting, waiting until the attacks cease, this work proposes an unprecedented method to combat jammers by leveraging the jamming signals to transmit data based on the recent advances in ambient backscatter communication technology. When the jammer attacks the channels, the transmitter modulates the jamming signals to backscatter information to the receiver. To deal with the uncertainty of jamming attacks and environment conditions, we first develop a Markov decision process framework with the Q-learning algorithm to obtain the optimal policy for the system. However, the Q-learning algorithm is widely known for its slow convergence, especially in large system state and action spaces. Hence, we develop a novel deep reinforcement learning algorithm based on the dueling neural network architecture that converges to the optimal policy much faster than the conventional Q-learning. Extensive simulations show that our proposed solution can improve the average throughput up to 426% and reduce the packet loss by 24% compared to other anti-jamming solutions.**

***Index Terms*—Anti-jamming, ambient backscatter, RF energy harvesting, deep dueling, deep reinforcement learning.**

I. Introduction

Due to the broadcast nature, wireless communications are particularly vulnerable to jamming attacks. By injecting interfering signals to the communication channel, a jammer can decrease the signal-to-interference-plus-noise ratio (SINR) at the receiver, thereby interrupting or preventing the legitimate communications of wireless systems. The jamming attacks can be easily launched by using commercial off-the-shelf products [1] and have a significant detriment to wireless applications, especially for low-power communications.

Various anti-jamming solutions have been proposed in the literature. The most common approach is frequency-hopping spread spectrum (FHSS) [2]-[4]. In particular, the FHSS mechanism allows a wireless device to quickly switch its operating frequency to other frequencies by using a shared algorithm implemented at both the transmitter and the receiver. In [3], the authors introduced a stochastic game framework in which the transmitter is equipped with the FHSS technique to cope with jamming attacks. However, the FHSS technique requires more spectrum resources than transmitting in a single frequency. In addition, for powerful jammers which can attack multiple channels simultaneously, FHSS may be less effective. Another popular solution is the rate adaptation (RA) technique [1], [5]. With RA, the transmitter can transmit data at lower rates, adapting with different interference levels from the jammer. In [1], the authors combined the RA and FHSS techniques and formulated a zero-sum game to mitigate attacks from a reactive-sweep jammer. However, the RA technique is not effective on a single channel and under jamming attacks [6].

All the above work and others in the literature tend to "escape" jamming attacks by finding another communication channel or adapting, waiting until the attacks cease. As such, these methods require additional resources, e.g., spectrum bandwidth, transmit power, or hardware capacity. Thus, we develop a novel anti-jamming framework that is extremely efficient in dealing with jamming attacks. In this work, inspired by the state-of-the-art ambient backscatter communication technology [7], instead of escaping the attacks, we leverage jamming signals to transmit data. When the jammer attacks the channels, the transmitter modulates the jamming signals to backscatter information to the receiver. Additionally, the transmitter is also equipped with an energy harvesting circuit to harvest energy from RF signals, and uses the harvested energy to transmit its data. To deal with the uncertainty of jamming attacks and ambient RF signals, we develop the Markov decision process (MDP) framework together with the Q-learning algorithm to obtain the optimal defense policy for the system. However, the Q-learning algorithm is widely known for its slow convergence, especially in large system state and action spaces. Hence, we develop a novel deep reinforcement learning algorithm based on the dueling neural network architecture that converges to the optimal policy much faster than the conventional Q-learning. Extensive simulations show that by leveraging the ambient backscatter communications and the deep dueling neural network architecture, we can improve the average throughput up to 426% and reduce the packet loss by 24% compared to other state-of-the-art anti-jamming solutions. In addition, it is interesting to observe that the transmission rate increases with the jamming power.

II. System Model

We consider a wireless system consisting of a gateway, an RF ambient source (e.g., radio or TV towers), a transmitter, and a jammer as illustrated in Fig. 1. The transmitter is equipped with RF energy harvesting and ambient backscatter capabilities to harvest energy and backscatter data through the ambient RF signals and the jamming signals, respectively.

A. Smart and Reactive Jammer with Self-Interference Suppression Capability

We consider a smart and reactive jammer with self-interference suppression capability [1], allowing it to "listen"

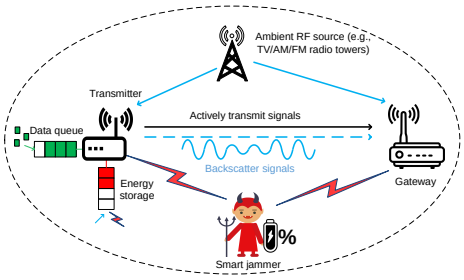


Fig. 1: System model

to the channel while jamming. Doing so, the jammer can discern its jamming outcome and reactively optimize its attack strategy to maximize the disruption of the transmitter. Let $\mathbf{P}_{\mathrm{J}} = \{P_0^{\mathrm{J}}, \ldots, P_n^{\mathrm{J}}, \ldots, P_N^{\mathrm{J}}\}$ denote the vector of discrete jamming power levels, in which P_N^{J} is the peak jamming power $P_{\max}$. In each time slot, the jammer can select a power level P_n^{J} as long as its average power constraint is satisfied. In practice, the jammer can adjust its pulse duty cycle factor to achieve the maximum degradation on the target channel while maintaining a time-average power constraint P_{avg}. Note that P_{avg} should be less than $P_{\max}$ [6]. If we denote $\mathbf{x} \triangleq (x_0, \ldots, x_n, \ldots, x_N)$ as a probability vector, then the strategy space of the jammer can be defined as follows:

$$\mathbf{J}_s \triangleq \Big\{(x_0, \ldots, x_n, \ldots, x_N), \sum_{n=0}^{N} x_n = 1, \\ x_n \in [0,1], \forall n \in \{0, \ldots, N\}, \mathbf{x}\mathbf{P}_{\mathrm{J}}^{\top} \leq P_{\mathrm{avg}}\Big\}. \tag{1}$$

Given a value of P_{avg}, the jammer will find an optimal strategy to attack the channel to maximize its objective. Specifically, the smart jammer can observe the activities and obtain the information of the transmitter, e.g., the number of packets transmitted/backscattered by the transmitter. Then, based on the observations, the jammer can define its objective function to maximize the disruption. We assume that the jammer receives a reward w_n^{J} if it attacks the channel with power level P_n^{J}. w_n^{J} can be referred as the number of packets that have been completely disrupted (i.e., not being successfully received/decoded). Let $\mathbf{w}_{\mathrm{J}} = \{w_0^{\mathrm{J}}, \ldots, w_n^{\mathrm{J}}, \ldots, w_N^{\mathrm{J}}\}$ denote the reward vector of the jammer corresponding the jamming power vector $\mathbf{P}_{\mathrm{J}}$. The optimal attack strategy $\mathbf{x}$ is obtained by using linear programming to solve the following problem.

$$\max_{\mathbf{x}} \ \mathbf{x}\mathbf{w}_{\mathrm{J}}^{\top}, \text{s.t.} \begin{cases} \sum_{n=0}^{N} x_n = 1, \\ x_n \in [0,1], \forall n \in \{0, \ldots, N\}, \\ \mathbf{x}\mathbf{P}_{\mathrm{J}}^{\top} \leq P_{\mathrm{avg}}. \end{cases} \tag{2}$$

B. System Operation

We denote the probability of the ambient RF source being idle in each time slot by η. Due to the average power constraints P_{avg}, the jammer may attack the channel with different power levels at different time. When the jammer attacks the channel and the ambient RF source is idle, the transmitter can choose one of the following actions: (i) stay idle, (ii) harvest energy, (iii) backscatter data, or (iv) adapt its transmission rate by using RA technique [1], [6]. Depending on the transmit power level P_n^{J} of the jammer, the transmitter can harvest e_n^{J} units of energy or backscatter maximum $\widehat{d}_n^{\mathrm{J}}$ packets through the jamming signals. In practice, the more power the jammer uses to attack the channel, the more energy and the more packets the transmitter successfully harvests and transmits to the gateway, respectively [8]. Let $\mathbf{e} = \{e_0^{\mathrm{J}}, \ldots, e_n^{\mathrm{J}}, \ldots, e_N^{\mathrm{J}}\}$ and $\widehat{\mathbf{d}} = \{\widehat{d}_0^{\mathrm{J}}, \ldots, \widehat{d}_n^{\mathrm{J}}, \ldots, \widehat{d}_N^{\mathrm{J}}\}$ denote the amount of energy that the transmitter can successfully harvest and the number of packets that the transmitter can successfully transmit when the jammer attacks the channel with power level $\mathbf{P}_{\mathrm{J}} = \{P_0^{\mathrm{J}}, \ldots, P_n^{\mathrm{J}}, \ldots, P_N^{\mathrm{J}}\}$, respectively.

In practice, when the jammer attacks the channel and the ambient RF source does not transmit data, the transmitter still can transmit its data by using the RA technique. Specifically, based on jamming power $P_{\mathrm{n}}^{\mathrm{J}}$, the transmitter can actively transmit data at maximum rate r_m. We then denote $\mathbf{r} = \{r_1, \ldots, r_m, \ldots, r_M\}$ as the set of available transmission rates that the transmitter can choose to transmit data when the jammer attacks the channel. At each rate r_m, the transmitter can transmit maximum $\widehat{d}_m^{\mathrm{r}}$ packets. In this work, the arrival data process follows the Poisson distribution with mean rate λ. The maximum data queue size and energy storage capacity are denoted by D and E, respectively. If a packet arrives at the system when the data queue is full, it will be dropped. To consider a low-latency system, if a packet stays in the queue longer than a latency threshold, i.e., t_{th}, it will be discarded. If at least one of the sources (i.e., either the ambient RF source, or the jammer, or both of them) is active, the transmitter can choose to backscatter data or harvest energy. The transmitter then observes the results of the taken action, i.e., the total number of packet backscattered or the total amount of harvested energy, and update the learning function. Based on the states of the ambient RF source and the jammer, the operations of our system can be expressed as follows:

- When the ambient RF source is idle and the jammer is idle, the transmitter can (i) transmit maximum $\widehat{d}_{\mathrm{t}}$ packets if it has enough energy (each packet requires e_{t} units of energy to be successfully transmitted) or (ii) stay idle.
- When the ambient RF source is idle and the jammer attacks with power level P_n^{J}, the transmitter can (i) use the RA technique to transmit maximum $\widehat{d}_m^{\mathrm{r}}$ packets if it has enough energy, (ii) backscatter maximum $\widehat{d}_n^{\mathrm{J}}$ packets, (iii) harvest e_n^{J} units of energy, or (iv) stay idle.
- When the ambient RF source is active and the jammer is idle, the transmitter can (i) backscatter maximum $\widehat{d}_{\mathrm{b}}$ packets, (ii) harvest e_{h} units of energy, or (iii) stay idle.
- When the ambient RF source is active and the jammer attacks with the power level P_n^{J}, the transmitter can (i) backscatter d_{sum} packets with $d_{\min} \leq d_{\mathrm{sum}} \leq d_{\max}$ where $d_{\min} = \min(\widehat{d}_{\mathrm{b}}, \widehat{d}_n^{\mathrm{J}})$ and $d_{\max} = \widehat{d}_{\mathrm{b}} + \widehat{d}_n^{\mathrm{J}}$, (ii) harvest e_{sum} units of energy with $e_{\min} \leq e_{\mathrm{sum}} \leq e_{\max}$ where $e_{\min} = \max(e_{\mathrm{h}}, e_n^{\mathrm{J}})$ and $e_{\max} = e_{\mathrm{h}} + e_n^{\mathrm{J}}$ [9], or (iii) stay idle.

III. Problem Formulation

To deal with the uncertainty of jamming attacks and the ambient RF signals, we adopt the Markov decision process (MDP) framework.

A. State Space

We define the state space of the system as follows:

$$\mathcal{S} \triangleq \Big\{(c, j, d, e) : c \in \{0, 1\}; j \in \{0, 1\}; \\ d \in \{0, \ldots, D\}; e \in \{0, \ldots, E\}\Big\}, \tag{3}$$

where j represents the state of the jammer, i.e., $j = 1$ when the jammer is active and $j = 0$ otherwise. c represents the state of the ambient RF channel, i.e., $c = 1$ when the ambient RF channel is busy and $c = 0$ otherwise. d and e represent the number of packets in the data queue and the number of energy units in the energy storage of the transmitter, respectively. The system state is then defined as $s = (c, j, d, e) \in \mathcal{S}$.

B. Action Space

The transmitter can perform one of the $(M+4)$ actions, i.e., stay idle, actively transmit data, harvest energy, backscatter data, or actively transmit data when then channel is attacked with one of M transmission rates by using the RA technique. Then, the action space of the transmitter can be defined by $\mathcal{A} \triangleq \{a : a \in \{1, \ldots, M + 4\}\}$, where

$$a = \begin{cases} 1, & \text{stay idle,} \\ 2, & \text{transmit data,} \\ 3, & \text{harvest energy,} \\ 4, & \text{backscatter data,} \\ 4 + m, & \text{adapt to rate } r_m \text{ with } m \in \{1, \ldots, M\}. \end{cases} \tag{4}$$

C. Immediate Reward

We define the reward for the system as the number of packets that are successfully transmitted to the gateway. Thus, the immediate reward of the system after the transmitter makes an action a at state s can be defined as follows:

$$r(s, a) = \begin{cases} d_{\rm t}, & \text{if } c = 0, j = 0, d > 0, e \geq e_{\rm t}, \text{and } a = 2, \\ d_{\rm b}, & \text{if } c = 1, j = 0, d > 0, \text{and } a = 4, \\ d_n^{\rm J}, & \text{if } j = 1, c = 0, d > 0, \text{and } a = 4, \\ d_{\rm sum}, & \text{if } j = 1, c = 1, d > 0, \text{and } a = 4 \\ d_m^{\rm r}, & \text{if } c = 0, j = 1, d > 0, e > 0, \text{and } a = 4 + m, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

In the above, when the ambient RF source is idle, the jammer does not attack the channel, and the number of data and energy units are sufficient for active transmission, the transmitter can actively transmit $0 < d_{\rm t} \leq \widehat{d}_{\rm t}$ packets to the gateway (i.e., $a = 2$). When the ambient RF source is active, the jammer is idle, and the transmitter has data to transmit, it can choose to backscatter $0 < d_{\rm b} \leq \widehat{d}_{\rm b}$ packets (i.e., $a = 4$). Similarly, when the jammer attacks the channel, the RF source is idle, and the transmitter has data to transmit, if it chooses to backscatter, it can transmit maximum $0 < d_n^{\rm J} \leq \widehat{d}_n^{\rm J}$ packets (i.e., $a = 4$). If the ambient RF source is idle, the jammer attacks the channel, and the transmitter has enough energy and data in the queues, it can choose to adapt its rate (i.e., $a = 4 + m; m \in \{1, \ldots, M\}$) and actively transmit $0 < d_m^{\rm r} \leq \widehat{d}_m^{\rm r}$ packets to the gateway. If both the jammer and the RF source are active, the transmitter has data to transmit, and it chooses to backscatter data, it can transmit $d_{\min} \leq d_{\rm sum} \leq d_{\max}$ to the gate way [9]. Finally, the immediate reward is 0 when the transmitter cannot successfully transmit any packet to the gateway. Note that after performing an action, the transmitter will observe the reward, i.e., number of packets that are successfully transmitted, based on ACK messages sent from the gateway. In other words, $d_{\rm t}$, $d_{\rm b}$, $d_n^{\rm J}$, $d_{\rm sum}$, and $d_m^{\rm r}$ are the actually received packets at the gateway. For that, the reward function captures the overall path between the source and the tag-receiver, e.g., fading, SNR, BER, or the packet error rate.

D. Optimization Formulation

We formulate an optimization problem to obtain the optimal policy, denoted by π^*, that maximizes the average long-term reward for the system. Specifically, the optimal policy is a mapping from a state to an action taken by the transmitter. The optimization problem is then expressed as follows:

$$\max_{\pi} \qquad \mathcal{R}(\pi) = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}\left(r_k(s_k, \pi(s_k))\right), \tag{6}$$

where $\mathcal{R}(\pi)$ is the average reward under the policy π, and $r_k(s_k, \pi(s_k))$ is the immediate reward under policy π at time step k. From a given state, the process can go to any other states after k steps. Thus, the average throughput $\mathcal{R}(\pi)$ is well defined and does not depend on the initial state [10].

IV. Defeating Jammer with Deep Dueling Neural Network Architecture

A. Q-learning based Algorithm

In our system, the transmitter cannot obtain the information about the jammer in advance to find the optimal policy. Thus, this section introduces Q-learning algorithm [11] to help the transmitter find the optimal policy without requiring prior information about the jammer as well as the channel. In particular, given a current state s_t, the algorithm selects an action a_t based on the ϵ-greedy method. After performing action a_t, the algorithm observes immediate reward r_t and next state s_{t+1}, and updates the Q-values as follows [11]:

$$\mathcal{Q}_{t+1}(s_t, a_t) = \mathcal{Q}_t(s_t, a_t) + \tau_t \Big[r_t(s_t, a_t) + \\ \gamma \max_{a_{t+1}} \mathcal{Q}_t(s_{t+1}, a_{t+1}) - \mathcal{Q}_t(s_t, a_t)\Big], \tag{7}$$

where γ is the discount factor which represents the importance of long-term reward [11] and τ_t is the learning rate corresponding to the impact of new information to the existing value. To guarantee the convergence for the algorithm, $\tau_t \in [0, 1), \sum_{t=1}^{\infty} \tau_t = \infty$, and $\sum_{t=1}^{\infty} (\tau_t)^2 < \infty$. Based on (7), the transmitter can employ the Q-learning algorithm to obtain the optimal defense policy. However, for complicated systems, the convergence rate of the Q-learning algorithm is usually slow. That makes the Q-learning algorithm practically inapplicable. In the following, we introduce the deep dueling algorithm to overcome this issue by leveraging the deep Q-learning and novel dueling architecture.

B. Deep Dueling Neural Network Architecture

We propose a deep dueling based anti-jamming algorithm [12] to improve the system's convergence speed. Similar to conventional deep reinforcement learning approaches, the deep dueling uses a deep neural network to estimate the value of the Q-function. However, the key idea making the deep dueling superior to conventional approaches is its novel neural network architecture. Clearly, in many states, it is unnecessary to estimate the value of corresponding actions as the choice of these actions has no repercussion on what happens [12]. Hence, the algorithm divides the Q-function into the value function and the advantages function. These two function are then separately estimated by two stream of fully connected layers. As such, the deep dueling algorithm can achieve more robust estimates of state value, and significantly improve its convergence rate as well as stability. In the following, we present details of the value and advantage functions.

Given a stochastic policy π, the values of state-action pair (s, a) and state s are expressed as $\mathcal{Q}^{\pi}(s,a) = \mathbb{E}\big[r_t | s_t = s, a_t = a, \pi\big]$ and $\mathcal{V}^{\pi}(s) = \mathbb{E}_{a\sim\pi(s)}\big[\mathcal{Q}^{\pi}(s,a)\big]$, respectively. The advantage function of actions can be expressed as $\mathcal{G}^{\pi}(s,a) = \mathcal{Q}^{\pi}(s,a) - \mathcal{V}^{\pi}(s)$. Specifically, the value function $\mathcal{V}$ corresponds to how *good it is to be in a particular state* s [12] and the advantage function $\mathcal{G}$ measures the importance of each action. To estimate values of $\mathcal{V}$ and $\mathcal{G}$ functions, we use a dueling neural network in which one stream of fully-connected layers outputs a scalar $\mathcal{V}(s;\beta)$ and the other stream estimates an $|\mathcal{A}|$-dimensional vector $\mathcal{G}(s,a;\alpha)$, where α and β are the parameters of fully-connected layers. These two sequences are then combined at the output layer by Eq. (8).

$$\mathcal{Q}(s,a;\alpha,\beta) = \mathcal{V}(s;\beta) + \mathcal{G}(s,a;\alpha). \tag{8}$$

Note that adding a constant to $\mathcal{V}(s;\beta)$ and subtracting the same constant from $\mathcal{G}(s,a;\alpha)$ result in the same Q-value. Therefore, Eq. (8) is unidentifiable resulting in poor performance. To address this problem, the combining module of the network is implemented the following mapping:

$$\mathcal{Q}(s,a;\alpha,\beta) = \mathcal{V}(s;\beta) + \big(\mathcal{G}(s,a;\alpha) - \max_{a\in\mathcal{A}} \mathcal{G}(s,a;\alpha)\big). \tag{9}$$

In this way, the advantage function estimator has zero advantage when choosing action. Intuitively, given $a^* = \arg\max_{a\in\mathcal{A}} \mathcal{Q}(s,a;\alpha,\beta) = \arg\max_{a\in\mathcal{A}} \mathcal{G}(s,a;\alpha)$, we have $\mathcal{Q}(s,a^*;\alpha,\beta) = \mathcal{V}(s;\beta)$. Hence, we convert (9) into a simple form by replacing the max operator with an average as follows:

$$\mathcal{Q}(s,a;\alpha,\beta) = \mathcal{V}(s;\beta) + \big(\mathcal{G}(s,a;\alpha) - \frac{1}{|\mathcal{A}|}\sum_{a} \mathcal{G}(s,a;\alpha)\big). \tag{10}$$

Based on (10), the advantage and value functions are combined at the output layer to obtain the optimal policy for the system.

V. Performance Evaluation

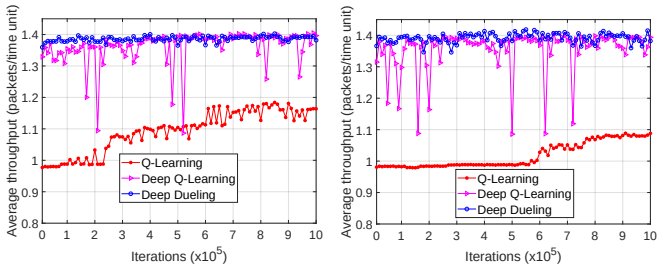
A. Parameter Setting

In our system, the jammer has four transmit power levels, i.e., $\mathbf{P}_{\mathrm{J}}$ = {0W, 7W, 15W, 21W}, with $P_{\max}$= 21W [15]. As explained in the Section II, as the jamming power increases, the transmitter can successfully harvest more energy or transmit more packets by backscattering jamming signal, and thus we set $\mathbf{e} = \{0, 2, 3, 4\}$ and $\widehat{\mathbf{d}} = \{0, 1, 2, 3\}$. In addition, when the jammer attacks the channel and the rate adaption technique is implemented, the transmitter can transmit d_m^{r} ={2, 1, 0} packets when the jammer attacks under power levels P_n^{J}={7W, 15W, 21W}, respectively. If both the jammer and the ambient RF source are active, the total number of backscattered packets d_{sum} and the total amount of harvested energy e_{sum} follow the Poisson distribution with the means of $(d_{\min} + d_{\max})/2$ and $(e_{\min} + e_{\max})/2$, respectively as defined in Section II-B. The data queue of transmitter can store up to 20 packets with the packet size set at 300 bits [13]. The energy storage capacity is set to be 20 units. The fundamental energy unit is 60 μJ [14]. Other parameters are provided in Table I. To evaluate the proposed solution, we compare its performance with two other schemes, i.e., HTT and WTJ. For the HTT, the transmitter only implements harvest-then-transmit protocol without considering ambient backscatter communication technology. For the WTJ, the transmitter can implement both harvest-then-transmit protocol and ambient backscatter communication technology only for the ambient RF signals. This scheme evaluates the system performance without leveraging the jamming signal.

TABLE I: PARAMETER SETTING

Symbol	e_{h}	$\widehat{d_{\mathrm{b}}}$	$\widehat{d_{\mathrm{t}}}$	e_{t}	t_{th}	η	P_{avg}
Value	2	1	4	1	3	0.5	7

B. Simulation Results

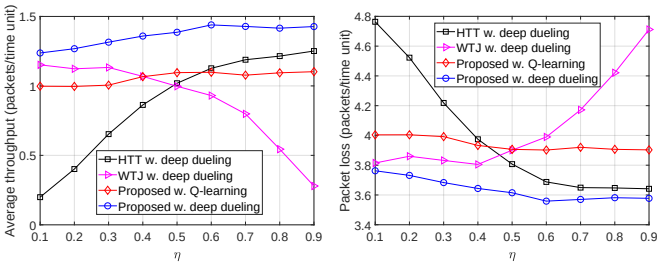


(a) $D = E = 10$ (b) $D = E = 20$

Fig. 2: Convergence rates

a) Convergence of Deep Reinforcement Learning Approaches: We first show the learning process and the convergence of the proposed deep reinforcement learning algorithms in several scenarios. As shown in Fig. 2(a) and Fig. 2(b), when $D = 10$ and $D = 20$, respectively, after 10^6 iterations, the average throughput obtained by the Q-learning algorithm is much lower than those of the deep reinforcement learning algorithms, especially in the first 10^5 iterations. This implies that as the system state space increases, the Q-learning algorithm requires more time to be converged. Note that the performance obtained by Deep Q-learning algorithm is as close as that of Deep Dueling algorithm, however the average throughput obtained by the Deep-Q learning algorithm is very fluctuated compared with that of the Deep Dueling algorithm. The reason is that the Deep-Q learning algorithm requires more time to be converged compared with that of the Deep Dueling algorithm.

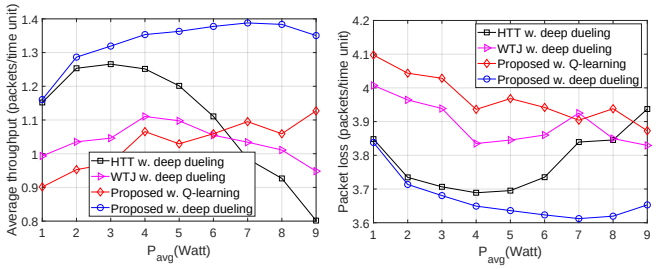
b) Performance Evaluation: Next, we perform simulations to evaluate and compare the performance of proposed solutions with those of the HTT and WTJ schemes. For the HTT and WTJ schemes, we adopt the Deep Dueling algorithm (with 4×10^4 iterations) to obtain the optimal policy for the transmitter. For the proposed solutions, we recruit both the Deep Dueling (with 4×10^4 iterations) and Q-learning algorithms (with 10^6 iterations).



(a) (b)

Fig. 3: (a) Average throughput and (b) Packet loss vs. η.

In Fig. 3, we vary the idle channel probability of the ambient RF source η. As shown in Fig. 3(a), when η increases, the throughput of the WTJ policy decreases. The reason is that the WTJ has less opportunities to harvest energy and backscatter data from the ambient signal when the ambient RF source is likely to be idle. In contrary, the average throughputs obtained by the HTT policy and the proposed solution increase and their packet loss will be reduced when the idle channel probability increases. This is from the fact that the transmitter has more opportunities to harvest energy from the jamming signal to support its transmissions. Moreover, the proposed solution can also backscatter data through both the jamming and ambient signals, thereby its throughput is considerably higher than that of the HTT scheme. Note that the Q-learning algorithm cannot obtain the optimal policy in the first 10^6 iterations. Thus, the performance derived by the Q-learning algorithm is much lower than that of the Deep Dueling algorithm.



(a) (b)

Fig. 4: (a) Average throughput and (b) Packet loss vs. P_{avg}.

In Fig. 4, we vary P_{avg} to evaluate the average throughput and packet loss of the system. Clearly, when P_{avg} increases from 1W to 3W, the throughputs of the HTT and WTJ policies increase. The reason is that the transmitter has more chances to harvest energy from the strong jamming signal to support its active transmissions when the jammer and the ambient RF source are idle. However, when P_{avg} is large (e.g., higher than 3W), i.e., the jammer is likely to attack the channel, the throughput of these policies decreases as the transmitter has less chance to actively transmit data to the gateway. However, the throughput achieved by the proposed solution increases as it allows the transmitter to switch to the backscatter mode when the jammer is likely to attack the channel. Consequently, the proposed solutions achieve the best performance in terms of packet loss as shown in Fig. 4(b). Again, the performance of the Q-learning algorithm is not as good as the Deep Dueling algorithm due to the slow-convergence problem.

VI. Conclusion

In this paper, we have developed the anti-jamming framework which allows the transmitter to effectively defeat jamming attacks. In particular, with the ambient backscatter capability, while being attacked, the transmitter can either backscatter its data to the gateway through the jamming signal or harvest energy from the jamming signal to support its operations. To obtain the optimal defense policy under the uncertainty of the jammer and the channel, we have proposed the deep dueling algorithm with a novel deep neural network architecture. Via extensive simulations, it is interesting to observe that, using the proposed framework, the transmission rate increases with the jamming power.

References

[1] M. K. Hanawal *et al.*, "Joint adaptation of frequency hopping and transmission rate for anti-jamming wireless systems," *IEEE Trans. Mobile Comput.*, vol. 15, no. 9, Sept. 2016, pp. 2247-2259.

[2] A. Sabharwal *et al.*, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, Jun. 2014, pp. 1637-1652.

[3] B. Wang *et al.*, "An anti-jamming stochastic game for cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, Apr. 2011, pp. 877-889.

[4] X. Liu *et al.*, "Anti-Jamming Communications Using Spectrum Waterfall: A Deep Reinforcement Learning Approach," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 998-1001, May 2018.

[5] K. Pelechrinis, I. Broustis, S. V. Krishnamurthy, and C. Gkantsidis, "Ares: An anti-jamming reinforcement system for 802.11 networks," *ACM CoNEXT*, Rome, Italy, Dec. 2009.

[6] K. Firouzbakht *et al.*,"On the capacity of rate-adaptive packetized wireless communication links under jamming," *ACM WISEC*, Tucson, AZ, USA, 2012, pp. 3-14.

[7] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: Wireless communication out of thin air," *ACM SIGCOMM*, Hong Kong, China, Aug. 2013.

[8] N. V. Huynh *et al.*, "Ambient Backscatter Communications: A Contemporary Survey," *IEEE Commun. Surveys Tuts.*, vol. 20 , no. 4 , Fourthquarter 2018, pp. 2889-2922.

[9] C. Yang *et al.*, "Riding the airways: Ultra-wideband ambient backscatter via commercial broadcast systems," *IEEE INFOCOM*, Atlanta, GA, USA, May 2017.

[10] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. Springer Press, 1997.

[11] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, no. 34, pp. 279292, 1992.

[12] Z. Wang *et al.*, "Dueling network architectures for deep reinforcement learning," [Online]. Available: arXiv:1511.06581.

[13] P. Blasco *et al.*, "A learning theoretic approach to energy harvesting communication system optimization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, Apr. 2013, pp. 1872-1882.

[14] G. Papotto *et al.*, "A 90-nmCMOS 5-Mbps crystal-Less RF-powered transceiver for wireless sensornetwork nodes," *IEEE J. Solid-State Circuits*, vol. 49, no. 2, Feb. 2014, pp. 335-346.

[15] 21W jammer [Online]. Available: http://drone-jammers.com/shop/21w-jammer-with-8-antennas-for-blocking-cdma-2g-3g-4g-lte-wimax-wifi-2-4ghz-uhf-vhf-rc-gps-lojack/