“© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Ambient Backscatter: A Novel Method to Defend Jamming Attacks for Wireless Networks

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

***Abstract*—This paper introduces a novel idea to defend jamming attacks for a wireless network. In particular, when the jammer attacks the channel, the transmitter can leverage the jamming signals to transmit data by using ambient backscatter technique or harvest energy from the jamming signals to support its operation. To deal with the uncertainty of the jammer, we propose a reinforcement learning algorithm that allows the transmitter to obtain the optimal operation policy through real-time interaction processes with the attacker. The simulation results show the effectiveness of ambient backscatter in combating jammers, i.e., it enables the transmitter to transmit data even under the jamming attacks. We observe that the more power the jammer uses to attack the channel, the better performance the network can achieve.**

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

I. Introduction

DUE to the broadcast medium, wireless communications are extremely vulnerable to jamming attacks. By injecting interfering signals to a target wireless channel, the jammer can decrease the signal-to-interference-plus-noise ratio (SINR) at the receiver, thereby interrupting or preventing the wireless communications between legitimate devices [1], [2]. Unlike inadvertent interference, jamming signals are usually powerful and can continuously disrupt the legitimate wireless communications.

There are various countermeasures to prevent and mitigate impacts of jamming attacks [4]. One of the first well known methods is regulating the transmission power of transmitters. Specifically, a transmitter can choose to either transmit at a very low power so that the jammer cannot detect its transmission or at very high level to dominate jamming signals. In [6], the authors also find out that by controlling transmission rate and transmission power, the impact of jamming signals can be reduced. However, this method is inefficient especially when the jammer often attacks the channel with high power. Another widely adopted approach is frequency-hopping spread spectrum (FHSS) [4], [5]. The FHSS mechanism allows wireless devices rendezvous on a given channel to communicate (e.g., sharing a predefined hoping pattern) once its current communication channel is attacked. However, the FHSS mechanisms require a set of available communication channels together with a predefined switching algorithms implemented on wireless devices. In addition, in the case if the jammer has sufficient power to attack all channels simultaneously, the communications can be completely disrupted.

In this paper, we introduce an unprecedented method that allows the transmitter to communicate even under jamming attacks. In particular, we adopt the ambient backscatter [7], a novel communication technology that enables two backscatter devices to communicate by leveraging surrounding signals. Specifically, when a transmitter needs to send data, it backscatters the data on ambient signals, e.g., TV or FM signals, to its receiver. The receiver then can decode the data by using the averaging mechanism [7]. Inspired by this idea, when a channel is attacked by a jammer, the transmitter can leverage the jamming signals to backscatter and transmit information. As such, the transmitter not only can avoid the jamming attacks but also leverage the jamming signals for its transmissions. To deal with the uncertainty of jamming attacks, we adopt Markov decision process (MDP) framework and propose a reinforcement learning algorithm to help the transmitter maximize its long-term reward in terms of network throughput and delay. This proposed learning algorithm will allow the transmitter to learn the optimal policy without requiring information about the jammer in advance. Simulation results show that our proposed solution can achieve the best performance (in terms of throughput and packet loss) compared with other optimal solutions that do not consider neither utilizing ambient backscatter technology nor leveraging jamming signals. More interestingly, our results reveal that the more power the jammer uses to attack the channel, the better performance the network can achieve.

II. System Model

We consider a wireless network consisting of a gateway, a transmitter, and a jammer as shown in Fig. 1. The transmitter is equipped with an energy harvesting and a backscatter circuits [7]. The transmitter thus can either harvest energy and use the harvested energy to actively transmit its data, i.e., harvest-then-transmit (HTT) mode, or backscatter data on the jamming signals [7], i.e., backscatter mode.

A. Jamming Model

In the system under consideration, the jammer attacks the channel to degrade the effective SINR at the receiver [6]. For that, it usually sets the bandwidth and the center frequency of the noise the same as those of the transmitter. The SINR can be calculated as $\theta = \frac{P_R}{\phi P_J + \rho^2}$, [3], [6], where P_R is the received power from the transmitter at the gateway, P_J is the jamming power transmitted by the jammer, ρ^2 is the variance of additive white Gaussian noise. The jamming power received

N. V. Huynh, D. N. Nguyen, D. T. Hoang, and E. Dutkiewicz are with the School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: huynh.nguyenvan@student.uts.edu.au, {diep.nguyen, hoang.dinh, and eryk.dutkiewicz}@uts.edu.au).

M. Mueck is with Intel Corporation, Neubiberg 85579, Germany (e-mail: markus.dominik.mueck@intel.com).

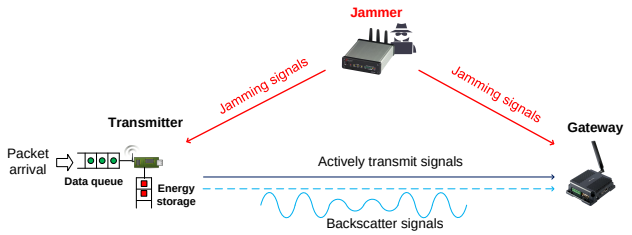


Fig. 1: System Model.

at the gateway is ϕP_J, where $0 \leq \phi \leq 1$ is the attenuation factor. We denote $\mathbf{P}_J = \{P_0^J, \ldots, P_n^J, \ldots, P_N^J\}$ as a vector of discrete jamming power levels from P_0^J to P_N^J. At each time slot, the jammer selects a given jamming power level P_n^J with a given probability x_n as long as its average power constraint is satisfied. Specifically, let $\mathbf{J}_s$ denote the strategy space of the jammer and $\mathbb{X}$ as the attack probability vector, then we have:

$$\mathbf{J}_s = \left\{ \mathbf{x} = (x_0, \ldots, x_n, \ldots, x_N), \sum_{i=0}^{N} x_n = 1. \right\}. \quad (1)$$

Given an average power constraint P_{avg}, $\mathbf{x}\mathbf{P}_J^\top \leq P_{avg}$ [6], the jammer can select an optimal strategy to attack the channel in order to achieve its objective, e.g., minimize the network's throughput. In particular, we consider a smart jammer that would know the information of the transmitter, e.g., how many packets the transmitter can transmit/backscatter and how many packets it can bring down if the jamming is successful. Then, we denote $\mathbf{w}_\mathrm{J} = \{w_0^\mathrm{J}, \ldots, w_n^\mathrm{J}, \ldots, w_N^\mathrm{J}\}$ as the reward vector of the jammer in which w_n^J is the number of packets that have been corrupted if the jamming power is P_n^J. Thus, The optimal attack strategy $\mathbf{x}$ is obtained by using linear programming to solve the following problem.

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

B. Channel Model

When the jammer attacks the channel, the transmitter can either adapt its rate or cease its active transmission and then choose to harvest energy or backscatter data on the jamming signals. Through experiments and analysis on backscatter communication systems, e.g., [7], it can be observed that the stronger the ambient signal (i.e., higher jamming power), the more packets the transmitter can backscatter to the gateway. Note that the relationship between the backscatter rate and the power of the ambient (jamming) signals can be either linear or non-linear and our proposed framework and following analysis are applicable to both. Thus, depending on the transmission power level P_n^J of the jammer, the transmitter can harvest an amount of energy, denoted by e_n^J, or backscatter maximum $\widehat{d}_n^J$ packets based on the jamming signals. We denote $\mathbf{e} = \{e_0^J, \ldots, e_N^J\}$ and $\widehat{\mathbf{d}} = \{\widehat{d}_0^J, \ldots, \widehat{d}_N^J\}$ as the harvested energy amount and backscattered packet vectors of the transmitter, respectively.

As aforementioned, under jamming, the transmitter still can (actively) transmit its data by adapting its rate with the jamming power, i.e., rate adaptation (RA). The criterion for rate adaptation is derived from the optimal strategy that maximizes the long-term average throughput. Let $\mathbf{r} = \{r_1, \ldots, r_m, \ldots, r_M\}$ denote 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. Note that, for $m = 1, \ldots, M$, when $\gamma_{m-1} \leq \theta < \gamma_m$ with γ_m is the value of SINR, the gateway only can decode packets sent at rates $r_0, r_1, \ldots, r_{m-1}$, and the packets sent at rate r_m or higher is lost [3] (i.e., not successfully decoded). The transmitter can detect if the jammer is active or idle, but not the specific jamming power level. The proposed algorithm below allows the transmitter to learn the information about the jamming power level and the associated likelihood, then adapt its defense strategy to maximize the average long-term throughput. The data arrival at the transmitter is assumed to follow the Poisson distribution with mean rate λ. If a packet stays in the queue longer than a latency threshold, i.e., t_{th}, it will be discarded. In summary, when the jammer is idle, the transmitter can (i) actively transmit maximum $\widehat{d}_t$ packets (each packets requires e_t units of energy to be successfully transmitted) or (ii) stay idle. When the jammer is active, the transmitter can leverage the jamming signals to (i) backscatter maximum $\widehat{d}_n^J$ packets, (ii) harvest e_n^J units of energy, (iii) transmit maximum $\widehat{d}_m^\mathrm{r}$ packets using the RA technique or (iv) stay idle.

III. Problem Formulation

A. State Space and Action Space

We define the state space of the system as follows:

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

where j represents the state of the jammer, i.e., $j = 1$ when the jammer attacks and $j = 0$ otherwise. d and e represent the numbers of packets in the data queue and the energy units in the energy storage of the transmitter, respectively. D and E are the maximum sizes of the data queue and the energy storage, respectively. The state of the system is then defined as a composite variable $s = (j, d, e) \in \mathcal{S}$. Note that, to highlight the effect of the backscattering and energy harvesting from the jamming signals, we assume that the transmitter has zero energy at the beginning (i.e., the worst case in which the transmitter has no energy and being attacked). If the transmitter is powered with a battery or if the energy is not of concern, our model is still applicable by setting the transmitter with a non-zero or infinite initial energy level, respectively.

The transmitter can perform one of the $4 + M$ actions, i.e., stay idle, actively transmit data, harvest energy from the jamming signals, backscatter data using the jamming signals, and reduce the transmission rate to one of M rates by using the RA technique when the jammer attacks the channel.

Then, the action space of the transmitter can be defined by $\mathcal{A} \triangleq \{a : a \in \{1, \ldots, 4+M\}\}$, where

$$a=\begin{cases}1, & \text{stays idle,}\\ 2, & \text{actively transmits data,}\\ 3, & \text{harvest energy from the jamming signals,}\\ 4, & \text{backscatter data using the jamming signals,}\\ 4+m, & \text{the transmitter adapts its transmission to}\\ & \quad \text{rate } r_m \text{ with } m \in \{1, \ldots, M\}.\end{cases} \tag{4}$$

B. Reward Function

The reward of the transmitter is defined as the number of packets it can send to the gateway. Thus, the immediate reward of the transmitter after an action a is executed at state s is defined as follows:

$$\mathcal{T}=\begin{cases} d_{\mathrm{t}},\ (j=0, d>0, e \geq e_t, a=2; 0 < d_{\mathrm{t}} \leq \widehat{d_{\mathrm{t}}}),\\ d_n^J, (j=1, d>0, a=4; 0 < d_n^J \leq \widehat{d_n^J}),\\ d_m^r, (j=1, d>0, e \geq e_t, a=4+m; 0 < d_m^r \leq \widehat{d_m^r}),\\ 0, \quad \text{otherwise.}\end{cases} \tag{5}$$

In this paper, we aim to find the optimal policy for the transmitter, denoted by Ω^*, to maximize its long-term average reward defined as follows:

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

where $\mathcal{R}(\Omega)$ is the long-term average throughput under the policy Ω, and $\mathcal{T}_k(\Omega)$ is the immediate throughput under policy Ω at time step k. The optimal policy Ω^* will allow the transmitter to make its optimal decisions based on its current states, i.e., data queue, energy queue, and channel state.

IV. Q-Learning Algorithm

To deal with the uncertainty of jamming attacks, we develop a reinforcement learning algorithm based on the Q-learning algorithm [8] to help the transmitter find the optimal policy without requiring any information about the jammer in advance. In particular, the Q-learning algorithm constructs a Q-table to store values of all state-action pairs. Given the current state, the transmitter performs an action based on its current policy and updates the Q-table based on its observations, i.e., immediate reward and the next state. In this paper, we aim to find the optimal policy, i.e., a mapping from the state space to the action space, $\pi^* : \mathcal{S} \to \mathcal{A}$ for the transmitter to maximize its long-term average reward. The expected value function obtained by policy π from state $s \in \mathcal{S}$ is as follows:

$$\begin{aligned}\mathcal{V}^{\pi}(s) &= \mathbb{E}_{\pi}\Big[\sum_{t=0}^{\infty} \gamma^t r_t(s_t, a_t) | s_0 = s\Big]\\ &= \mathbb{E}_{\pi}\Big[r_t(s_t, a_t) + \gamma \mathcal{V}^{\pi}(s_{t+1}) | s_0 = s\Big],\end{aligned} \tag{7}$$

where $0 \leq \gamma < 1$ is the discount factor that determines the importance of long-term reward [8]. In particular, if γ is close to 0, the transmitter will prefer to select actions to maximize its short-term reward. In contrast, if γ approaches 1, the transmitter will make actions to maximize its long-term reward. $r_t(s_t, a_t)$ is the immediate reward achieved by taking action a_t at state s_t. At each state s, an optimal action determined through the optimal value function is expressed as $\mathcal{V}^*(s) = \max_{a_t}\Big\{\mathbb{E}_{\pi}[r_t(s_t, a_t) + \gamma \mathcal{V}^{\pi}(s_{t+1})]\Big\}, \quad \forall s \in \mathcal{S}$. The optimal Q-functions for state-action pairs are denoted as $\mathcal{Q}^*(s,a) \triangleq r_t(s_t, a_t) + \gamma \mathbb{E}_{\pi}[\mathcal{V}^{\pi}(s_{t+1})], \quad \forall s \in \mathcal{S}$. Then, $\mathcal{V}^*(s)$ can be expressed as $\mathcal{V}^*(s) = \max_a\{\mathcal{Q}^*(s,a)\}$. By making samples iteratively, the problem is reduced to determining $\mathcal{Q}^*(s,a)$ for all state-action pairs. Intuitively, the Q-function is updated to find the temporal difference between the predicted Q-value and its current value as follows:

$$\begin{aligned}\mathcal{Q}_t(s_t, a_t) &= \mathcal{Q}_t(s_t, a_t) +\\ &\alpha_t\Big[r_t(s_t, a_t) + \gamma \max_{a_{t+1}} \mathcal{Q}_t(s_{t+1}, a_{t+1}) - \mathcal{Q}_t(s_t, a_{s_t})\Big],\end{aligned} \tag{8}$$

where α_t is the learning rate that determines the impact of new information to the existing value. Moreover, to guarantee the convergence for the Q-learning algorithm, the learning rate α_t is deterministic, nonnegative, and satisfies (9) [8].

$$\alpha_t \in [0,1), \sum_{t=1}^{\infty} \alpha_t = \infty, \text{ and } \sum_{t=1}^{\infty} (\alpha_t)^2 < \infty. \tag{9}$$

Based on (8), the transmitter can employ the Q-learning algorithm to obtain the optimal policy. Specifically, the algorithm first initializes the table entry $\mathcal{Q}(s,a)$ arbitrarily, e.g., zero, for each state-action pair (s,a). Given current state s, the algorithm performs action a through ϵ-greedy algorithm. In particular, the Q-learning algorithm selects a random action with probability ϵ or selects an action that maximizes the $\mathcal{Q}(s,a_s)$ with probability $1-\epsilon$. Thus, the algorithm can explore the whole state space. The algorithm then determines the next state and immediate reward after performing action a and updates the Q-table based on (8). The algorithm is terminated when all Q-values converge, or after a finite number of iterations. This algorithm yields the optimal policy indicating an action to be taken at each state such that $\mathcal{Q}(s,a)$ is maximized for all states in the state space, i.e., $\pi^*(s) = \arg\max_a \mathcal{Q}^*(s,a)$. Under (9), it is proved in [8] that the Q-learning algorithm will converge to the optimal solution with probability one. Note that the Q-learning algorithm is feasible to deploy on the transmitter as it just needs to store a Q-table with a few hundreds of state-action pairs and perform basic calculations.

V. Performance Evaluation

A. Experiment Setup

In our system, the data queue can store up to 10 packets with a packet size set at 300 bits [9]. The energy storage capacity is set at 10 units. The fundamental energy unit is 60 μJ [10]. λ is set at 3 packets/time unit. If the transmitter performs active transmission, it can successfully transmit up to 4 packets. Each transmitted packet requires 1 unit of energy. The jammer has four transmission power levels, i.e., $\mathbf{P}_{\mathrm{J}} = \{0W, 5W, 15W, 20W\}$ with $P_{\max} = 20W$ [12]. As the numbers of harvested energy units and backscattered packets increase when the jammer's transmission power increases [7], we set $\mathbf{e} = \{0,1,2,3\}$ and $\widehat{\mathbf{d}} = \{0,1,2,3\}$. When the jammer attacks the channel with power level $\mathbf{P}_{\mathrm{J}} = \{5W, 15W, 20W\}$, the transmitter can use the RA technology to transmit $\mathrm{r} = \{2,1,0\}$ packets accordingly.

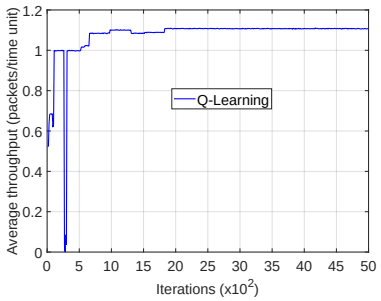
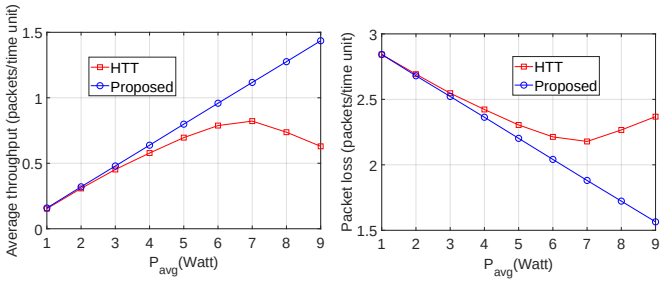
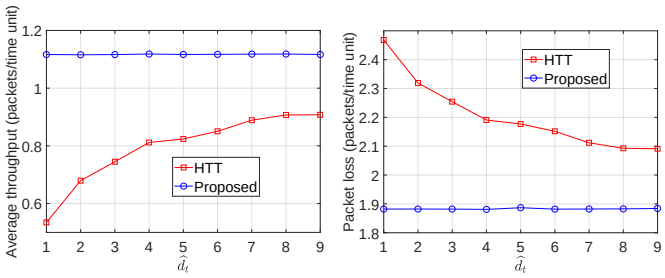


Fig. 2: Convergence of the Q-learning algorithm.



(a) (b)

Fig. 3: (a) Average throughput (packets/time unit) and (b) packet loss (packets/time unit) vs. P_{avg}.



(a) (b)

Fig. 4: (a) Average throughput (packets/time unit) and (b) packet loss (packets/time unit) vs. $\widehat{d}_t$.

The latency threshold t_{th} is 3 time units and P_{avg} = 7W. Note that the Q-learning algorithm does not require any information about the jammer in advance. It can learn the jammer's strategy to obtain the optimal defense policy. To evaluate the proposed solution, we compare its performance with the HTT scheme in which the transmitter only performs harvest-then-transmit protocol [11]. Moreover, in this scheme, the transmitter can also adapt its data rates when the channel is under attack. For the Q-learning algorithm, the learning rate and discount factor are set at 0.1 and 0.9, respectively.

B. Simulation Results

In Fig. 2, we first show the convergence and learning process of the Q-learning algorithm. Clearly, the Q-learning algorithm can converge to the optimal policy after 2,000 iterations. As such, our proposed solution is applicable to deploy in practice as it can quickly obtain the optimal defense strategy for the system. In Fig. 3, we vary P_{avg} to evaluate the performance of the system. Clearly, the throughput of the HTT scheme increases when P_{avg} increases from 1W to 7W. The reason is the transmitter has more opportunities to harvest RF energy from the jamming signals. However, when P_{avg} > 7W, the transmitter has less chance to actively transmit data, and thus the throughput of the HTT scheme decreases. By switching between backscattering data and harvesting energy from the the jamming signals, the average throughput of the proposed solution increases and the packet loss decreases (as shown in Fig. 3(b)) when the jammer is likely to attack the channel. This is a very interesting finding as the transmitter considers the jammer as an RF source to leverage for its operations. Next, in Fig. 4, we vary the maximum number of packets $\widehat{d}_t$ that can be actively transmitted. Clearly, when the $\widehat{d}_t$ increases, the throughput of the HTT mode increases and remains the same when the $\widehat{d}_t \geq 7$. This is stemmed from the fact that the amount of the harvested energy of the transmitter is limited. In contrast, by balancing between the harvesting and backscattering time, the proposed solution achieves the highest throughput and lowest packet loss among the two schemes.

VI. Summary

In this letter, we have proposed an anti-jamming approach which allows the transmitter to backscatter data based on the jamming signals. To deal with the uncertainty of the jammer, we have formulated the optimization problem based on the MDP framework and developed the Q-learning algorithm to obtain the optimal solution. The simulation results interestingly showed that the legitimate transmitters can attain higher throughput and less packet loss with higher jamming power.

References

[1] S. G. Hong *et al.*, "Game-theoretic modeling of backscatter wireless sensor networks under smart interference," *IEEE Commun. Lett.*, vol. 22, no. 4, Apr. 2018, pp. 804-807.

[2] L. Xiao *et al.*, "Reinforcement learning-based NOMA power allocation in the presence of smart jamming," *IEEE Trans. Veh. Techno*, vol. 67, no. 4, Apr. 2018, pp. 3377-3389.

[3] 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.

[4] A. Mpitziopoulos *et al.*, "A survey on jamming attacks and countermeasures in WSNs," *IEEE Commun. Surv. Tutor.*, vol. 11, no. 4, Dec. 2009, pp. 42-56.

[5] L. Jia *et al.*, "Stackelberg game approaches for anti-jamming defence in wireless networks," *IEEE Wireless Commun.*, vol. 25, no. 6, Dec. 2018, pp. 120-128.

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

[7] V. Liu *et al.*, "Ambient backscatter: Wireless communication out of thin air," *ACM SIGCOMM*, Hong Kong, China, Aug. 2013.

[8] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, no. 3-4, pp. 279-292, 1992.

[9] P. Blasco, D. Gunduz, and M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, Apr. 2013, pp. 1872-1882.

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

[11] N. Zhao *et al.*, "Exploiting interference for energy harvesting: A survey, research issues, and challenges," *IEEE Access*, vol. 5, May 2017, pp. 10403-10421.

[12] 2.4Ghz WIFI 20W Powerful Jammer [Online]. Available: http://jammers4u.com/2.4ghz-wifi-powerful-jammer