# Protecting Multi-function Wireless Systems From Jammers with Backscatter Assistance: An Intelligent Strategy

Lotfi Ismail, Dusit Niyato, Sumei Sun, Dinh Thai Hoang,
Yonghui Li, and Dong In Kim

***Abstract*—In this paper, we present a novel unified framework to protect multi-function wireless systems from jamming attacks. Examples of such multi-function system include joint radar and communication (JRC) systems and simultaneous wireless information and power transfer (SWIPT) systems. By abstracting the system functionalities as a joint optimization problem of multiple queues, we achieve effective resistance against jammers for the multi-functions simultaneously. We incorporate different anti-jamming techniques into one framework. Deception mechanism is adopted to lure the jammer to attack and make its actions more predictable, and ambient backscatter technology is used to leverage the jamming signals. Since conventional Markov decision process (MDP) has only one decision epoch at every time slot, it cannot be used to model the deception strategy which needs two decision epochs to leverage the jamming signals. We therefore formulate the problem using an advanced two-step MDP. After that, a deep reinforcement learning algorithm with a prioritized double deep Q-Learning architecture is proposed to learn optimal strategies in different system states. We show that by jointly considering the multi-functions of the system with potential jamming attacks during design phase, significant improvement can be achieved for both of the system functionalities.**

***Index Terms*—Jamming attack, multi-function wireless systems, backscatter communication and deception mechanism.**

## I. Introduction

Exploitation of radio signals has significantly shifted modern technology. Traditionally, radio signals have been used for data transmission, radar object detection and ranging, and more recently, for other purposes such as wireless power transfer (WPT) [1], ets. Different spectrum is allocated and used exclusively for different applications. However, spectrum resources are becoming more scarce with growing deployments of wireless-related technologies in various sectors [2]. To overcome these limitations, spectrum sharing has been proposed to allow different systems and applications to share the same spectrum. Different approaches and frameworks have been studied, for example, dynamic spectrum management and cognitive radio for different wireless technologies to share the same frequency resources, joint radar-communications (JRC) for the two different applications to operate in the same frequency bands [2], [3], simultaneous wireless information and power transfer (SWIPT) over the same channel [1], etc. In the cases of JRC and SWIPT, *multi-function* wireless systems are designed [1].

Multi-function wireless systems received considerable attention in the past few years because they allow more efficient resource allocation and reduce implementation costs. For instance, a ship-borne JRC system can be implemented on a battleship to simultaneously detect targets and transmit data packets on the same waves to allies in the nearby, making its defence system more robust [7]. Besides, unmanned aerial vehicle (UAV) base stations are receiving more attention recently for user offloading in dense zones [8]. Therefore, JRC design can be adopted in the design of the autonomous UAV base station to enable simultaneous communication and target detection [9]. However, the joint optimization problem of throughput and radar sensing quality maximization should be designed carefully to meet the system requirements. In addition, SWIPT systems can be also adopted by similar devices used in JRC systems. For example, a UAV base station can use Lidar technology for mapping and geospatial sensing, and instead of transmitting radar signals, it transfers energy to some IoT nodes in the field (e.g., sensor nodes). This makes the optimization problem of data and radar requests in JRC systems similar to the optimization problem of data and energy transfer request in SWIPT systems. Hence, by abstracting these two optimization problems as a joint scheduling problem of two queues, we can develop a unified framework for multi-function wireless systems and then instantiate that solution in the context of each system, i.e., JRC or SWIPT systems.

The benefits of hardware and frequency reuse will only be possible if the multi-function wireless system can operate under different environments and circumstances, fulfilling multiple objectives simultaneously while counteracting deliberate interference and jamming attacks. Jamming attacks on single-function systems, such as wireless communication, radar and WPT systems have been studied extensively in the literature [1], [10], but very few research works studied the

*(Corresponding author: Dong In Kim)*
Lotfi Ismail and Dusit Niyato are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore (e-mail: ismail003@e.ntu.edu.sg, dniyato@ntu.edu.sg). Sun Sumei is with the Institute for Infocomm Research, A*STAR, Singapore (e-mail: sunsm@i2r.a-star.edu.sg). Dinh Thai Hoang is with the School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW, Australia (e-mail: hoang.dinh@uts.edu.au). Yonghui Li is with the School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW, Australia (e-mail: yonghui.li@sydney.edu.au). Dong In Kim is with the Department of Electrical and Computer Engineering, Sungkyunkwan University (SKKU), Suwon, South Korea(e-mail: dikim@skku.ac.kr)

[1] In this paper, we use the term "multi-function wireless systems" to refer to wireless systems that enable the use of the same spectrum or antenna for more than one functionality, e.g., data transmission and radar sensing, or data transmission and wireless power transfer as in [4]–[6].

case of a multi-function wireless systems under hostile jamming attacks. With the increasing demand for multi-function wireless systems, it is important to study the jamming attack models and the mitigation strategies. In JRC systems, when jamming a communication system, the jammer tries to degrade the signal-to-interference plus noise ratio (SINR) at the data receiver, while when jamming a radar system, the jammer tries to degrade the SINR at the radar receiver and prevent it from correctly decoding echos from targets. Similarly, in SWIPT systems, the jammer attacks the communication system by degrading the SINR at the receiver. When attacking the WPT system, the attacker may "steal" transferred energy and use that energy later to attack the communication channel or, prevent the WPT receiver from efficient harvesting of the transferred energy [11]. An important problem in designing multi-function wireless systems arises from the fact that the system needs to satisfy simultaneously multiple utilities. Moreover, a jammer can attack one or multiple functions, making the system design requirements more challenging as multiple attack scenarios have to be considered, and static solutions might not be robust to jointly overcome hostile attacks to different functions and satisfy the system requirements.

In this work, we address the problem of a multi-function wireless system under hostile jamming attack. To address the aforementioned challenges, we propose a novel multi-function wireless system design that cannot only resist jamming attacks but also further improve the system performance by smartly leveraging the jamming signals. In the context of JRC systems, we first make an abstraction of the data transmission and the radar sensing functions to be a joint scheduling problem [2]. Specifically, our objective is to find an optimal strategy to jointly schedule data packets and radar sensing actions requested by applications. To mitigate the jamming attack on the joint system, we explore the fact that the JRC node can perform different actions and the jammer cannot have the same degree of effectiveness on each action. For instance, the JRC node can use different transmission modes such as rate adaptation and backscattering on the jamming signals. However, since the wireless medium is highly dynamic over time, i.e, channel fading and jammer uncertainty, and due to the variety of actions that can be performed by the JRC node, it is hard to design a hand-crafted optimal strategy. Additionally, in multi-function systems, a node may have multiple functions, and thus when a jamming attack happens, it may not know the target of the jammer, e.g., it could be one of the functions or all of functions. Therefore, by deploying a reinforcement learning (RL) algorithm, the JRC node can learn the actual system dynamics, i.e., the jammer's strategy and the channel state, and optimally select the best action to perform at each time slot and hence, minimize the packet loss and increase the probability of target detection [3].

Part of this work was presented in [12] where only JRC systems were studied. In this article, we provide a unified framework for multi-function wireless systems covering JRC and SWIPT systems. In addition, more advanced techniques are proposed to model the system with more experiments and deeper evaluations and analysis. The main contributions of our paper are as follows:

1) We develop a novel framework for multi-function wireless systems where we propose to use the queue management concept for radio resource allocation and scheduling. The proposed abstraction helps designing solutions for both JRC and SWIPT systems at once and smoothly sharing the development from one system to another one. Our proposed framework is resilient and robust against a variety of jamming attacks.
2) We propose an intelligent deception strategy to lure the jammer and utilize its jamming signals to improve the system performance. We subsequently develop a highly-effective reinforcement learning algorithm to help the system obtain the optimal operation policy without prior knowledge about the jamming attack or the wireless channel.
3) We perform intensive simulations to evaluate performance metrics of the system under many scenarios and reveal a number of insights on the joint design of deep reinforcement learning and queue management concept. We show that by misleading the jammer through deception mechanism, we are able to suppress the jammer not to attack continuously, and thus jointly perform radar sensing safely and increase data throughput.

The rest of this paper is organised as follows. Section II reviews related works. Section III and IV describe our system model and the problem formulation, respectively. In section V, we present our proposed deep reinforcement learning algorithm. The evaluation results are then presented in section VI. Section VII concludes the paper and provides several potential research directions.

## II. Related Work

Jamming attacks on wireless communications, radar and WPT systems have been studied extensively in the literature [1], [10]. Nevertheless, most of the existing works investigated these systems separately and only few research works studied the case of a multi-function wireless system, i.e., JRC or SWIPT systems, under hostile attacks. When jamming the communication system, the jammer is trying to degrade the SINR at the receiver, while when jamming the radar system, the jammer is trying to degrade the SINR at the transmitter and prevent it from correctly decoding echos from existing targets. However, jamming WPT systems is quite different because it is difficult to prevent energy harvester (EH) nodes from harvesting energy, but the jammer still can harvest the transferred energy and use it later to attack the network. To the best of our knowledge, most of the present works have only addressed eavesdropping [13], [14] on multi-function wireless systems, but deliberate interference has not been well studied in the literature.

Traditional jamming attacks on wireless communication networks, radar and WPT systems can be easily launched against multi-function wireless systems. However, most of the

[2]In SWIPT systems, radar sensing is replaced with power transfer function.
[3]In SWIPT systems, beside packet loss minimization, it increases harvested energy by legitimate nodes and minimizes harvested energy by the jammer.

existing works only focused on physical layer design without consideration of malicious attacks [1], [2]. An interesting step towards multi-function wireless systems security in the presence of jammers was taken in [15], in which the authors studied a JRC system that uses dual-functional waveform to support target detection and data transmission in the presence of a jammer. Since the dual radar-communication system and the jammer are rivals and have contradictory objectives, the authors in [15] used a game theory model to formulate the problem and derived the Nash equilibrium. However, their solution requires the jamming attack probability in advance, which might not be practical and can change over time. Moreover, game theory studies provide only insights on the system performances for different policy adjustments but do not provide solutions to improve the resiliency of the system against jammers with different objectives and capabilities.

In the following, we present a brief overview of existing works in the literature to cope against communication, radar and WPT jammers.

### A. Jamming Mitigation in Communication Systems

Mitigating deliberate interference in wireless communications has been extensively studied in the literature and different techniques have been proposed [10]. Several Rate Adaptation Algorithms (RAAs) were adopted to mitigate noise jammers. However, since RAAs cannot distinguish between packet loss due to fluctuations in channel quality and packet loss due to deliberate interference, it has been shown in [16] that RAA turns to become a dangerous vulnerability. A more interesting and widely deployed method to counteract jamming is spread spectrum, such as frequency hopping spread spectrum (FHSS) and direct sequence spread spectrum (DSSS). The key idea is to enable the sender to spread a narrow-band signal over a wide-band of frequencies such that the jammer will not be able to predict it, thus lowering the probability of the signal being jammed. Many works have used spread spectrum techniques in combination with deep reinforcement learning in order to derive an optimal strategy for the transmitter to increase its overall throughput [17], [18]. However, spread spectrum techniques assume initial secret agreement on the spreading codes or frequency hopping pattern between the transmitter and the receiver. Unfortunately, smart jammers can use this condition to increase their jamming performance. Since reactive jammers can sense spectrum before jamming, they cannot only block the target channels with flexible power, but can also eavesdrop the public control channels [19].

More recently, new hybrid approaches have been proposed to mitigate smart and reactive jammers. In [20], authors proposed the use of both RAA and FHSS to counteract jammers. In this way, the transmitter can avoid being jammed by hopping to other channels, adjusting its transmission rates, or both. More interestingly, authors in [21] introduced a novel approach using deep reinforcement learning and ambient backscatter communication (ABC) technology to mitigate jammers. The key idea is to use recent advances in ABC and backscatter modulated jamming signals instead of hiding, thus making the jamming attack ineffective and power-wasting for the jammer. On top of this, the authors added deception mechanism in [22]. The deception mechanism consists of sending fake signals at the beginning of each time slot to mislead the jammer. If the jammer eats the bait and decides to attack, the transmitter uses backscatter technology as previously mentioned. Otherwise, the transmitter can safely transmit data through active transmission for the remaining of the time slot.

### B. Jamming Mitigation in Radar Systems

Jamming radars is a crucial component of any electronic countermeasure (ECM) system. It has been and still an evolving aspect of electronic warfare. Even though many researches in this area are still classified, several works that propose effective solutions against radar jammers are available in the literature. Also, recent years have witnessed more works in this direction because of the advances in vehicular technology which relies heavily on radars [23]. In [24], authors studied how to differentiate legitimate target echo from false echos introduced by a deceptive jammer on a bi-static multiple-input multiple-output (MIMO) radar system. Deceptive jamming consists of re-transmitting original signals after a delay-time which results in an incorrect target range detection and an increase in false alarm rate. They introduced a range deception jamming recognition method through digital signal processing framework and formulated the optimization problem as a weighted least square (WLS) problem. In [25], authors studied a simple anti-jamming scenario in which a spot jammer, i.e., a jammer that focuses all of its power on a single frequency, is attacking a radar system. They adopted a reinforcement learning algorithm to build a system that relies on frequency hopping technique in order to help the radar intelligently select an unoccupied channel to transmit in. However, if the jammer can attack different frequencies simultaneously across the entire bandwidth, frequency hopping becomes ineffective.

### C. Jamming Mitigation in WPT Systems

WPT systems security received more attention compared to JRC systems. Even though interfering RF energy harvesting is not an easy attack to perform, the authors in [11] showed that a jammer can launch a depletion attack on the energy harvester node, which results in a quick drain in device's battery, preventing the node from any further operation in the network. Moreover, if the transmitting device is using beamforming, a *Beamforming Vector Poisoning* attack can be launched as shown in [26], in which the jammer injects signals in the same frequency as the legitimate users resulting in a destructive interference at the receiver. This leads to severe degradation of harvested energy units and increases packet loss. Furthermore, the jammer can also harvest the transferred energy and use it later to attack the network [27]. In [27], authors proposed a deception mechanism to undermine the attack ability of the jammer. They formulated the problem as a throughput maximization problem in a cognitive radio network (CRN) with WPT and EH nodes. However, they did not consider maximizing the transferred energy which might be used for other internal computations by different nodes in the network beside communication purpose. In addition,

jamming signals can also be harvested by EH nodes in order to minimize the efficiency of jamming attacks.

In summary, existing works suffer from the following limitations:

- They only focus on one functionality of a multi-function wireless system, while other functions are not well protected.
- Jamming attacks are still not considered in the design of multi-function wireless systems and no unified framework for these systems exists.
- Most of the existing works consider some prior knowledge about the jamming attacks. However, due to the uncertainty of wireless environment and the jammer's objective, it is hard to obtain this information in real scenarios.

Motivated by the limitations mentioned above, a deep reinforcement learning (DRL) based multi-function wireless system that incorporates a variety of anti-jamming techniques is proposed in this work to improve the resistance of the system against reactive jammers. The proposed solution can defeat jamming attacks and quickly adapt with the environment dynamic. Our system design can learn the properties of the jammer, suppress the jamming attacks and achieve higher performances compared to those of the conventional anti-jamming solutions.

## III. System Model

### A. System Overview

Figure 1 shows an overview of our considered multi-function (MF) wireless system. The MF system supports data transmission, referred to as the data transmission mode, and either radar sensing or power transfer, referred to as the radar sensing mode and power transfer mode, respectively. The MF node can choose between two modes to achieve the required performance level. Either time or frequency based access architecture possesses advantages and disadvantages that make each one preferred over the other for some desired applications. In this work, for the purpose of establishing a robust and unified framework for multi-function wireless systems, we consider that the multi-function wireless system is using time division multiple access (TDMA) method to alternate between different modes. Time division access methods also provide low-complexity design, low implementation cost and help to easily integrate modules sharing same hardware components (e.g., antenna, power, etc.) which is suitable for multi-function wireless systems. For instance, Figure 2 illustrates the use of time division access method in JRC systems, where time is divided into slots of equal duration $T$ and the scenario of a pulsed radar where the basic pulse strength is equal to that used by the data communication system is considered as in [28]. When operating in the radar sensing mode, the JRC node starts by transmitting a pulse $s_1$ at the beginning of the time slot, and then waits for an echo $e_{s1}$ to be received in the remaining time. In data transmission mode, the JRC node continuously transmits data during its allocated time slot [4].

[4]In SWIPT systems, time switching architecture can be adopted similarly [1].

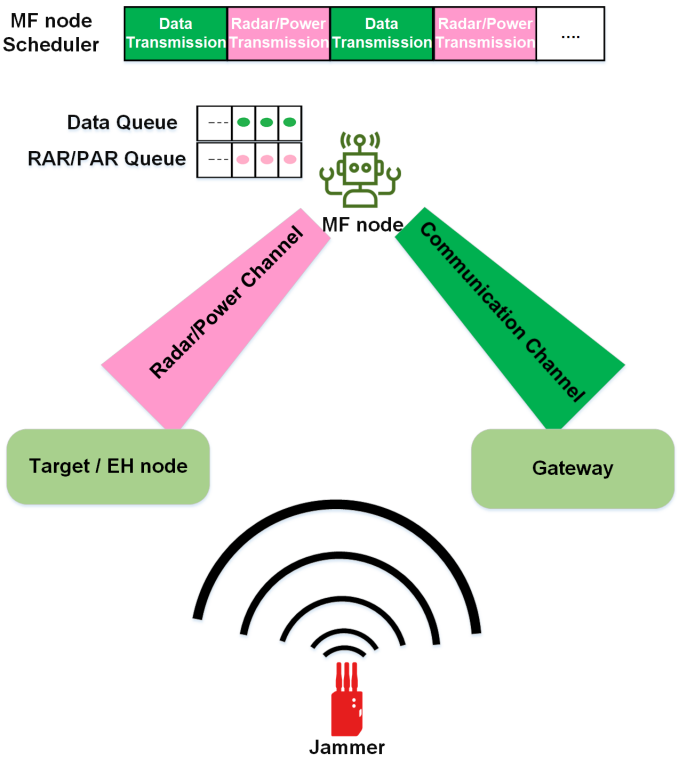


**Figure 1.** System model.

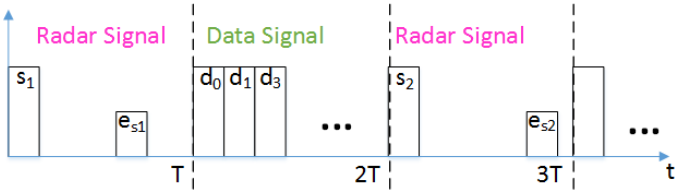


**Figure 2.** The JRC node can choose between data and radar transmissions mode.

The MF node receives different requests based on the supported functionalities. Arriving data packets are stored in the data queue while radar sensing requests are stored in the radar activity request (RAR) queue if the MF node is an instance of JRC system. Similarly, if the MF node is an instance of SWIPT system, power transfer requests are stored in the power activity request (PAR) queue as illustrated in Figure 1. When operating in the radar mode, the MF node is trying to infer information about nearby targets such as angle and distance, while when operating in the power transfer mode, the MF node is transferring energy to the energy harvester (EH) node. Additionally, due to the shared wireless medium, a jammer can disrupt the network performance by introducing hostile interference into the channel upon detection of any type of transmission, i.e., reactive jamming [10]. This will cause a severe degradation of the Quality of Service (QoS) to the multi-function wireless system. Moreover, if the multi-function wireless system is deployed for autonomous vehicles or robot-assisted industrial applications, there may cause severe safety issues [9], [29].

With our proposed abstraction and the proposal of queue concept, the application of our framework to different multi-function wireless systems is straightforward. Accordingly, JRC and SWIPT systems are just instances of a multi-function wireless system. We apply our solution to JRC system, which is an instance of the multi-function wireless system, but the same framework is applicable on SWIPT systems similarly. The only adaptation needed is the actions, ie., instead of having actions of radar sensing as in JRC system, there will be actions of power transfer in SWIPT systems. The rest of the formulation remains the same which shows the power of our framework [5]. Therefore, to avoid ambiguity between JRC and SWIPT systems, we continue the rest of the paper with more emphasis and analysis of the proposed framework on JRC systems.

We consider that communication and radar circuits are both co-located on the JRC node, share the same frequency band and use a time division access method, and thus there is no need to consider mutual interference between radar and data signals at the JRC node in our system design [30]. The JRC node has a stable energy capacity, e.g., in an autonomous vehicle, and has two functionalities: data transmission and radar sensing. For the data transmission functionality, we consider that data arrival follows a Poisson distribution[6] with mean $\lambda$. At each time slot, if a new packet arrives (or is generated) at the JRC node and the data queue is not full, it will be added to the data queue and can remain in the queue for a predefined threshold $\tau_{data}$. If the data queue is full, the new packet will be discarded. Similarly, the radar activity request (RAR) arrival follows a Bernoulli distribution with parameter $p_{radar}$ related to deployment settings. For example, if the system needs to perform radar sensing more frequently, then $p_{radar}$ is set close to 1. If only few radar sensing operations are required, then $p_{radar}$ is set close to 0. If an RAR remains in the RAR queue for a time exceeding a predefined threshold $\tau_{radar}$, then it is removed. We also define $w = \sum_{wt}$ as the total delay (sum of the waiting time $wt$) of all radar sensing requests in the RAR queue. This parameter, i.e., $w$, is considered later in the reward function for data transmission actions.

In the following, we describe the adversarial model against the presented multi-function wireless system forwarded by our proposed countermeasures.

### B. Channel Model

*1) Communication Channel Model:* The communication channel gain $g_{ab}$ between two devices $a$ and $b$ is composed of the large-scale fading $l_{ab}$ and the small-scale fading $h_{ab}$. The large-scale fading is determined by the distance $d_{ab}$ between the two devices $a$ and $b$. According to Jake's model [32], the small-scale fading can be represented as:

$$h_{ab}^{(t)} = \gamma h_{ab}^{(t-1)} + \zeta^{(t)}, \tag{1}$$

where $\gamma$ $(0 \leq \gamma \leq 1)$ represents the coherent factor and $\zeta^{(t)}$ is a random variable with distribution $\zeta^{(t)} \sim \mathcal{CN}(0, 1-\gamma^2)$. Hence, the channel gain $g_{ab}^{(t)}$ at time $t$ can be given as [32];

$$g_{ab}^{(t)} = l_{ab}^{(t)} |h_{ab}^{(t)}|^2. \tag{2}$$

The SINR at the receiving gateway is then formulated as:

$$SINR_{data} = \frac{g_{data} P_T}{g_{J,d} P_J + \rho^2}, \tag{3}$$

where $P_T$ and $P_J$ are the transmission powers of the JRC node and the jammer, respectively. $g_{data}$ and $g_{J,d}$ are the communication channel gains between the receiving gateway and JRC node and between the receiving gateway and the jammer, respectively. $\rho^2$ is the noise power.

*2) Radar Channel Model:* To be able to detect targets and their distances correctly and accurately, the JRC node sends radar pulses and expects to receive reflections of these pulses such that the SINR at the radar receiver is higher than a predefined threshold. The power density $P_r$ returned by the target to the radar is defined as [33]:

$$P_r = \frac{P_t G_r \sigma A_e}{(4\pi)^2 R^4}, \tag{4}$$

where $P_r$ symbolizes the signal power returned to the radar antenna, $P_t$ is the transmit power by the JRC node, $G_r$ denotes radar antenna gain, $\sigma$ corresponds to Radar Cross Section (RCS) of the target, $R$ is the range to target and $A_e$ is the effective aperture area of the radar antenna. In addition to the signal power $P_r$, thermal noise represented by the variance of additive white Gaussian noise $\rho^2$ and jamming noise $P_J$ are also received at the JRC node. Then the SINR at the JRC node receiver is calculated as:

$$SINR_{radar} = \frac{P_r}{P_J + \rho^2}. \tag{5}$$

*3) Evaluation Metrics:* To evaluate the performance of our system, we consider the channel capacity and rate estimation for data transmission and radar sensing, respectively. Similar to [34], the channel capacity is defined as a function of the SINR as follows:

$$R_{com} = B \log_2(1 + SINR_{data}) \tag{6}$$

where $B$ is the channel bandwidth. The estimation rate of information that can be obtained from the reflected signals of the radar systems is defined as:

$$R_{est} = \frac{1}{2T_{pri}} \log_2(1 + SINR_{radar}) \tag{7}$$

where $T_{pri}$ is the pulse repetition interval of the radar system.

We observe from equations (6) and (7) that data rate and radar estimation rate both increase as the SINR at the communication receiver and the radar receiver increase, respectively. In addition, for radar systems, the accuracy of target parameters inference (e.g., range and angle) is highly influenced by the SINR at the radar receiver. Specifically, the

[5]A practical benefit would be the use of transfer learning to fine tune the learned model into other systems.

[6]Our model can be extended for other distribution or correlated arrival process such as Markovian arrival process (MAP) [31].

measuring errors in range and angle are defined respectively as [35]:

$$\sigma R = \frac{c}{2B\sqrt{2SINR_{radar}}} \quad (8)$$

$$\sigma A = \frac{\theta}{k_M\sqrt{2SINR_{radar}}} \quad (9)$$

where $c$ is the speed of light, $\theta$ is the radar beamwidth in the angular coordinate of the measurement and $k_m$ is the monopulse pattern difference slope. From equations (8) and (9) we observe that the range and angle errors can be minimized by maximizing the SINR at the radar receiver. Therefore, we consider in this work the SINR at the JRC node as the main performance metric for the radar system which is generic for both range and angle estimation, i.e., maximizing the SINR at the JRC node will increase the accuracy of the target parameters. For the communication system, maximizing the SINR at the gateway receiver is considered as the performance metric.

### C. Jamming Attack Model

We consider the jammer to be reactive, i.e., only attacks the channel if it detects some signals from the JRC node on the channel. This makes the attack highly efficient and long-lasting. Note that we consider a smart jammer who can distinguish between data transmission and radar transmission signals, and thus it can focus its attack on each transmission mode separately and effectively [36]. When the jamming signal gets interfered with the reflected target echo during the jamming attack at the radar receiver, in which case the probability of target detection depends on either constructive addition or destructive addition of the two signals, the JRC node experiences less probability of target detection [33]. The JRC node can try to lower the radar threshold [7] to increase the target detection probability, which also increases the probability of false alarm. The continuous adjustment of radar threshold can heavily affect the quality of inferred target parameters, e.g, angle and distance.

The jammer can attack the channel with different discrete jamming powers [37]. Let $\mathbf{P}^J = \{P_0^J, P_1^J, \ldots, P_N^J\}$ denote the available jamming power levels with $P_0^J = 0$ and $\mathbf{x} = \{x_0, x_1, \ldots, x_N\}$ be the probability vector of each jamming power such that $\sum_{i=0}^{N} x_i = 1$. At each time slot $t$, the jammer picks one jamming power $P_J \in \mathbf{P^J}$ according to its associated probability from $\mathbf{x}$. Note that if the jammer chooses $P_0^J$, no attack on the channel is launched. Moreover, the jammer has a limited energy supply. In practice, the jammer has the time-average power constraint defined as follows [20]:

$$\frac{1}{T^J}\sum_{t=1}^{T^J} e_a(t) \leq \hat{E}_J, \quad (10)$$

where $T^J$ is the total time period in which the jammer aims to attack the channel, $e_a(t)$ is the amount of energy used by the jammer to attack at time $t$ and $\hat{E}_J$ is a predefined threshold that specifies the average energy that can be used by the jammer at period $T^J$. The energy constraint is justified by the fact that strong jammers can overheat quickly and thus should not jam at high power continuously during all the time period $T^J$ [33]. Also note that if the jammer attack would occur frequently, its location can be disclosed and the JRC node can easily filter out the jammer's fake echo pulses.

### D. Anti-Jamming Attack Strategy

To overcome the aforementioned jamming attack, we incorporate several techniques into our framework. We first adopt TDMA scheme which helps exploring the benefits of changes of the signal type from communication waveforms to radar waveforms and vice-versa [34]. The changes of the generated signals by the JRC node will force the jammer to change its signal's characteristics to maximize its attack for both cases. Specifically, when communication signals are emitted during communication mode, the jammer needs to inject similar signals in the channel to effectively corrupt the data [38]. Similarly, when radar signals are emitted, the jammer needs to generate signals that are similar to the radar waveforms in order to make its attack to be successful [39], e.g., increase the misdetection rate. Note that since the switching between sensing and data transmission is only every few milliseconds, the target objects will rarely be missed from the detection because of the quiet period.

We then introduce the deception mechanism in the multi-function wireless system design, which consists of transmitting fake signals at the beginning of each time slot and enticing the jammer to attack. Additionally, by implementing an advanced reinforcement learning (RL) algorithm, the jammer's actions become more predictable and can be leveraged through backscatter communication and RAAs techniques. Specifically, since the jammer's actions are more predictable, we can choose to modulate communication bits on the jamming signals instead of active transmission [8]. Therefore, energy consumption by the JRC node is minimized and data throughput is maximized. Moreover, we can choose to transmit radar pulses in time slots where the jamming probability is low, and thereby maximizing the radar sensing functionality.

Upon detection of degradation in the channel quality, the JRC node can adopt rate adaptation techniques to continue transmitting data [20], but with a low data rate. In practice, if the jammer is detected, then based on the jamming power $P_J$, the JRC node can still transmit data through rate adaptation. Several detection techniques exist in the literature to estimate the jammer's state, i.e., current jamming power level, such as energy detection [40]. We then denote $\mathbf{r} = \{r_0, r_1, \ldots, r_i, \ldots, r_M\}$ to be the set of the available $M$ transmission rates supported by the JRC node. For each data rate $r_i$, the JRC node can successfully transmit a maximum of $d_{RA}$ packets.

When jamming signals are detected, the JRC node can still use rate adaptation or transmit its data through backscattering on the jamming signals. The later can achieve higher data rates

[7] Radar threshold refers to the minimum SINR value at the radar receiver to consider the received signals for processing.

[8] This refers to standard radio transmission which is different from backscatter transmission.

than that of rate adaptation technique and less Bit Error Rate (BER) [22]. It has been demonstrated that backscattering on an RF source (e.g., jammer) can increase data throughput as the jamming power increases [22], and thus if the JRC node smartly transmits through backscattering when the jammer is attacking, a high overall data throughput can be achieved. Since the jammer usually attacks with high powers, we use backscatter to modulate on the amplitude of the high-power noise by adjusting the impedance of the backscatter antenna (a wave is reflected when it encounters an antenna with two different impedance) [41]. Specifically, when the input bit to be transmitted is zero, the JRC node switches to non-reflecting state. In contrast, when the input bit to be transmitted is one, it switches to reflecting state [41]. At the receiving node, to extract the backscattered information, averaging methods similar to [22], [41] is used to demodulate reflected signals. The radar signals are desired to have high *time resolution* for accurate time of arrival (ToA) estimation. In this case, they appear to be *wideband*, so that the jamming signals need to be full-band jammers. The latter allows the averaging method to work effectively when backscattered signals are detected [9].

The idea of switching between different transmission techniques is inspired from frequency hopping strategy. Instead of switching between frequencies, we switch between different actions (transmission techniques). The effectiveness of this method is argued by the fact that the attacker cannot expect our next action, and since the jamming attack cannot have the same effect on all the set of actions chosen by the JRC node over time (e.g., the JRC node can achieve higher throughput if using backscatter technique compared to rate adaptation), we can improve the system performance by intelligently choosing the best transmission technique at every time slot. Note the joint scheduling problem of data and radar queues is highly coupled and the channel state is dynamic over time. Additionally, the JRC node needs to accurately predict the next jamming attack and choose the best action to perform to satisfy both of the system functions requirements. Therefore, beside the idea of switching between different transmission techniques, we develop an RL algorithm to learn an optimal transmission strategy for the JRC node to execute at every state.

Note that if the jammer attacks the channel with very dynamic signals that fluctuates in both frequency and power, then ambient backscatter communication might not work well. However, in our proposed framework, ambient backscatter is just one option among others to deal with communication jamming attack. In case if ambient backscatter does not work, the JRC node can use rate adaptation to transmit data but with a lower rate. Another option would be staying idle for some time slots to deceive the jammer to stop its attack for a while then start transmitting again. It is important to note that in our work we use DRL algorithm to learn the best transmitting strategy in the presence of the jammer. Therefore, the JRC node can automatically find the best anti-jamming strategy in cases if one of options is not effective.

[9]Note that the rate differentiation, where a low-rate data is embedded into the high-rate jamming signals, can be utilized for the averaging method.

### E. Other Considerations

In this paper, time is slotted and at the end of each time slot, the JRC node observes the environment and evaluates its previous action. If operating in the data transmission mode, the JRC node can detect that data packets have been jammed when not receiving an acknowledge message from the gateway [22]. Otherwise, if the JRC node is operating in the radar sensing mode, it can detect that it is under a radar jamming attack by not receiving correct reflected pulse back. Specifically, the JRC node observes its expected SINR from signals reflected by the target. If the SINR is less than a predefined threshold, insufficient information can be obtained from the reflected pulses, and thus the JRC node considers that it is under a radar jamming attack [33], [42].

Unlike traditional wireless networks, multi-function wireless systems are composed of systems that have different goals, metrics and operators. Thus, system performance needs to be optimized jointly with careful attention to evaluation metrics and units. Communication systems are traditionally evaluated in terms of throughput and WPT systems can be evaluated in terms of harvested energy units. However, in radar systems, several evaluation metrics can be required, e.g, angle and distance. To allow our study and analysis being useful for different multi-function wireless systems, we adopt the radar capacity from [43] in which it is defined in a binary fashion, i.e., "1" implies that target is present and "0" implies that target is absent. This helps us to extract important insights about the system dynamics when integrating DRL in JRC systems and guides us towards possible improvements. Additionally, inspired by the theoretical framework proposed in [34] to jointly analyze the JRC system performances, we define the following evaluation metrics to assess our proposed framework:

1) *Successful packet transmission ratio:* the number of packets successfully transmitted, i.e., through active transmission, backscattering and rate adaptation, over all the number of generated packets.
2) *Successful RAR ratio:* the number of RAR successfully performed over all the number of RARs generated. This metric reflects the sensing rate of radar results.
3) *Data throughput:* the number of packets successfully transmitted, i.e., through active transmission, backscattering and rate adaptation, per time unit.
4) *RAR throughput:* the number of RAR successfully performed per time unit (excluding miss detection and false alarm).

In the following sections, we present our problem formulation and show how the proposed system design can help the JRC node find the best operation mode at each time slot.

## IV. Problem Formulation

We develop an advanced two-step version of the MDP framework to model the jamming mitigation problem in multi-function wireless systems. Specifically, the MDP is defined by a tuple $< \mathcal{S}, \mathcal{A}, r >$ where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space and $r$ is the immediate reward that the multi-function node receives after performing action $a$ at state $s$ [44]. In

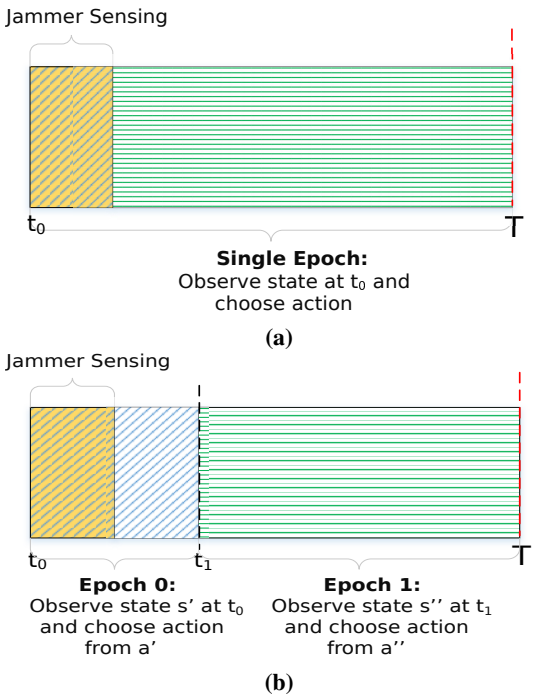


**Figure 3.** (a) Conventional MDP model, (b) Two-step MDP model.

conventional MDP model, the multi-function node observes the environment at the beginning of the time slot and then chooses an action to perform for the rest of the decision epoch. This would limit the performance of our system as the deception mechanism cannot be captured by this form of MDP. In particular, the reactive jammer does not attack the channel if it does not detect any signals in the channel, and consequently the JRC node will never be able to backscatter on the jamming signals or perform radar sensing safely. To overcome this limitation, we propose an advanced version of conventional MDP model as illustrated in Figure 3(b).

Specifically, there will be two decision epochs at each iteration. Since the jammer has a time-average power constraint, the jammer cannot attack continuously. The JRC node can make use of that and undermine the jammer's effectiveness through deception technique. At the beginning of each time slot $T$, i.e., at $t_0$ $(Epoch\ 0)$, the JRC node can inject data signals or radar pulses on the channel to mislead the jammer to perform the attack. Upon detection of these signals on the air, and based on the objective of the jammer, i.e., attacking all types of signals, only data signals or only radar signals, the jammer then decides to attack the channel or to stay idle for the rest of the time slot $T$. At the end of the deception period, which is considered to be longer than the detection time of the jammer, if the JRC node detects the jammer's presence at $t_1$, which marks the beginning of $Epoch\ 1$, it can either use rate adaptation technique to actively transmit data, but with a low data rate, or it may leverage the jamming signals by using backscatter technology. In cases that the jammer does not attack the channel, the JRC node can perform active transmission or radar sensing with a high chance of success. If the JRC node does not choose the deception action at $t_0$, then the action taken at the first decision epoch will continue until the end of the current time slot $T$ with the jamming risk remaining.

Note that in our work, the jamming sensing period is considered to be fixed but it can be optimized in advance before the learning process of the anti-jamming strategy begins. Specifically, the JRC node can observe the jammer's attacks for a period of time and find an optimal time for this period. Then our proposed DRL algorithm can be deployed to learn the optimal ant-jamming strategy. Several works have been proposed to optimize the sensing time [45], [46], which can be applied into our model straightforwardly.

*1) State Space:* The state space of the system is defined as:

$$\mathcal{S} \triangleq \Big\{(l,c,d,w) : l \in \{0,1,2\}; c \in \{0,1\}; \\ d \in \{0,\dots,D\}; w \in \{1,\dots,M\}\Big\}, \tag{11}$$

where $l$ represents deception action of the JRC node, i.e., $l = 1$ when the data deception is performed, $l = 2$ when the radar deception is performed and $l = 0$ otherwise. $c$ represents the state of the channel (presence of jamming signals or not), i.e., $c = 1$ if the channel is under attack and $c = 0$ otherwise. $d$ and $D$ represent the number of packets in the data queue and the maximum data queue size of the JRC node, respectively. Finally, $w$ represents the total time for the RARs (radar activity requests) waiting in the RAR queue (radar activity request queue) with $M$ being the maximum value of $w$. The system state is then defined as a composite variable $\mathbf{s} \triangleq (l,c,d,w) \in \mathcal{S}$.

Note that the states of our MDP can be obviously observed and obtained. Specifically, from equation (11), the first element of the state tuple $l$ is an internal information of the JRC node and can be always observed accurately. Similarly, the third and fourth elements of the tuple, which are the data queue and radar activity request queue, are also internal information and always observable. The second element of the system state space, which is the jammer state, is an external information for the JRC node but still can be inferred with high accuracy using several sensing techniques such as energy detection [40].

*2) Action Space:* The action space is defined by: $\mathcal{A} \triangleq \Big\{(a', a'') : a' \in \{0,\dots,5\}, a'' \in \{11,\dots,15\}\Big\}$. At the beginning of each time slot, we have the following actions:

$$a' = \begin{cases} 0, & \text{the JRC node stays idle,} \\ 1, & \text{the JRC node performs data deception,} \\ 2, & \text{the JRC node performs radar deception,} \\ 3, & \text{the JRC node actively transmits data,} \\ 4, & \text{the JRC node adapts its transmission rate,} \\ 5, & \text{the JRC node emits radar pulses,} \end{cases} \tag{12}$$

if $a' = 1$ or $a' = 2$ is chosen (data deception or radar deception), then we have the following actions for the rest of the time slot:

$$a'' = \begin{cases} 11, & \text{the JRC node actively transmits data, if } c = 0, \\ 12, & \text{the JRC node adapts its transmission rate, if } c = 1, \\ 13, & \text{the JRC node emits radar pulses, if } c = 0, \\ 14, & \text{the JRC node stays idle,} \\ 15, & \text{the JRC node backscatters data, if } c = 1, \end{cases} \tag{13}$$

if $a' = 1$ and $a' = 2$ are not chosen, we set $a'' = 0$.

*3) Immediate Reward:* We define the reward value as a function of successful data transmission and successful radar target detection (excluding miss detection and false alarm). For instance, if the JRC node chooses not to perform deception ($l = 0$) and actively transmits data ($a' = 3$), and if the jammer does not attack the channel ($c = 0$), then the JRC node will receive a reward $r_{act}$.

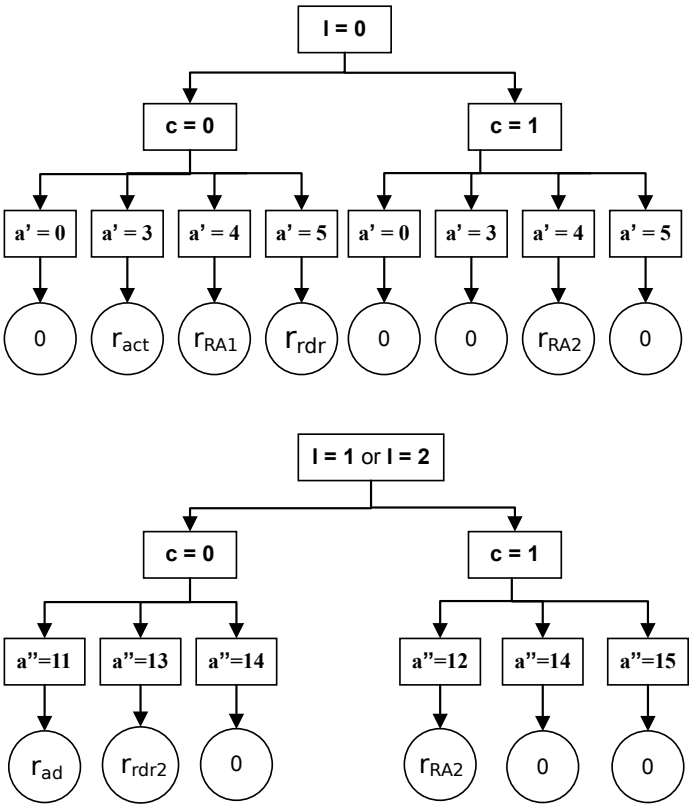


**Figure 4.** Immediate reward function.

Importantly, as shown in Table I, we set an adaptive reward function for data transmission actions which is dependant on the total waiting time in the RAR queue, i.e., if RAR queue is not full, the action related to data transmission receives a high reward. However, if the RAR queue is full, less reward is given for data transmission action. This is set to help the algorithm learn that RARs are very sensitive for delay and should not be deferred for long time.

Note that having the radar reward $r_{rdr}$ to be an adjustable hyper-parameter of the model gives a high flexibility for the system to be deployed in different environments. Moreover, we can perform several tests during the online learning before deployment to derive the optimal value of $r_{rdr}$.

TABLE I: Utilities for different situations

| Reward Notation | Value | Description |
|---|---|---|
| $r_{act}$ | $\frac{d_{active}}{w}$ | $d_{active}$: number of packet successfully transmitted with no deception. |
| $r_{ad}(< r_{act})$ | $\frac{d_{ad}}{w}$ | $d_{ad}$: number of packet successfully transmitted after deception. |
| $r_{RA1}$ | $\frac{d_{RA1}}{w}$ | $d_{RA1}$: number of packet successfully transmitted with rate adaptation and no deception. |
| $r_{RA2}(< r_{RA1})$ | $\frac{d_{RA2}}{w}$ | $d_{RA2}$: number of packet successfully transmitted with rate adaptation after deception. |
| $r_{rdr}$ | $d_{radar}$ | $d_{radar}$: number of radar equivalent packet. |
| $r_{rdr2}$ | $d_{radar} - 1$ | $d_{radar} - 1$: number of radar equivalent packet after deception. |
| $r_b$ | $\frac{d_b}{w}$ | $d_b$: number of packet successfully backscattered. |

*4) Optimization Formulation:* We then aim to find an optimal policy $\pi^*$ that has the best mapping from the state space to the action space which maximizes the average long-term reward. The optimization problem can be formulated as follows:

$$\max_{\pi} \qquad \mathcal{R}(\pi) = \lim_{\Upsilon \to \infty} \frac{1}{\Upsilon} \sum_{t=1}^{\Upsilon} \mathbb{E}(r_t(s_t, \pi(s_t))), \tag{14}$$

where $r_t(s_t, \pi(s_t))$ is the immediate reward under policy $\pi$ at time $t$ and $\mathcal{R}(\pi)$ is the long-term average reward of the JRC node under policy $\pi$.

## V. Optimal Defense Strategy With Deep Reinforcement Learning

In this work, we choose to solve the problem using a model-free RL algorithm. The main advantage of model-free over model-based RL algorithms is that it has a constant time policy for each state, i.e., constant time to compute the optimal action and no further thinking/computation is required for any state [47]. Therefore, this characteristic is very efficient for wireless systems where the time duration is very short for each state (typically in milliseconds), enabling the agent to react instantly. Additionally, in model-free RL, we only need to fine tune the hyper parameters of the network, but in model based RL, we need also to find the appropriate model which is hard to derive for many problems [44].

Several DRL algorithms have been proposed with some algorithms working better in some domains than others [47]. In our problem, since it is hard to collect millions of episodes for training during deployment, we choose to use a deep Q-Learning (DQL) based algorithm as it is known for its fast convergence speed with small training episodes [44]. In the following, we present the Q-Learning algorithm followed by the proposed DQL-based solution.

### A. Q-Learning based Approach

Based on the current state, i.e., the jammer state, the number of requests in the queues and their total delay $w$, the multi-function node follows its current policy to take the best action to maximize the long-term average reward. Q-Learning learns the action-value function $\mathcal{Q}(s,a)$, i.e., how good to take an action $a$ at a particular state $s$. In Q-Learning, a memory table $\mathcal{Q}[s,a]$ is built in order to store Q-values for all possible combinations of states and actions. However, Q-Learning algorithm suffers from a long learning time to derive the optimal defense policy and might be trapped in a local optimum, which cannot be tolerated in security system where high accuracy and fast adaptation are required. Next, we develop a deep Q-Learning based algorithm which allows the multi-function node to obtain an optimal solution quickly through utilizing advantages of the neural network architecture and replay memory.

### B. Deep Reinforcement Learning based Approach

By incorporating deep Q-Learning into RL, we can approximate the values of $\mathcal{Q}^*$ more efficiently and accelerate the system convergence speed [48]. First, the deep Q-Learning Algorithm (DQLA) adopts the experience replay mechanism in which a memory pool $\mathbb{M}$ of capacity $N$ stores a set of transitions $(s_t, a_t, r_t, s_{t+1})$ obtained when interacting with the environment at time $t$. Then, the DQLA randomly samples batches from the memory pool to train the deep neural network. This helps the algorithm to learn from previous transitions several times and reduce the correlations between experiences. The second step is to use two neural networks with the same structure to approximate Q-values. The first Q-network $\mathcal{Q}$ has parameters $\Theta$ and refers to the actual predictions of the Q-values. Since neural networks can overfit quickly and to avoid destabilizing the learning process [48], we use the second Q-network $\hat{\mathcal{Q}}$ with parameter $\Theta^-$ to refer to the maximum possible values of the next state. Specifically, the weights in $\hat{\mathcal{Q}}$ are set to be fixed temporally and not updated for $C$ steps, while the weights in $\mathcal{Q}$ are updated at every iteration. The goal is to increase the stability of the target Q-value $(r_j + \gamma \max_{a_j} \hat{\mathcal{Q}}(s_j, a_j; \Theta^-))$ when deriving the temporal difference (TD) error which is calculated by taking the difference between the target network $\hat{\mathcal{Q}}$ and the current network $\mathcal{Q}$, i.e.,

$$\delta_i = (r_t + \gamma \max_{a \in \mathcal{A}} \hat{\mathcal{Q}}(s_{t+1}, a; \Theta^-) - \mathcal{Q}(s_t, a_t; \Theta)). \quad (15)$$

In fact, the experience replay can be further improved by using the prioritized experience replay (PER) technique [49]. Some experiences might be more important than others but occurs less frequently. Therefore, instead of uniformly sampling from the replay memory $\mathbb{M}$, in PER we sample experiences that are more important to learn more efficiently. This can help to reduce the correlation between states and avoid forgetting important experiences. The idea is to give higher scores for experiences that can help to reduce the TD error. The importance score $p_i$ is calculated using the following expression:

$$p_i = \frac{(\delta_i + \varepsilon)^\alpha}{\sum_{k=1}^{N} (\delta_k + \varepsilon)^\alpha}, \quad (16)$$

where $\varepsilon$ is a small value to ensure the edge transitions can still be visited when their TD error is zero, $\alpha$ is a constant exponent between zero and one. The denominator is a normalization term by all priority values $N$ in the replay memory. Since transitions with high probability will be chosen frequently, the network might be biased towards experiences with high priority. This bias is eliminated through importance sampling which reduces the weights of often seen samples, i.e., $w_i = \left(\frac{1}{N} \cdot \frac{1}{p_i}\right)^\beta$, where $\beta$ is an increasing variable from zero to one. Finally, this weight $w_i$ is multiplied by the TD error as shown in Algorithm 1.

The overestimation of the target Q-values can be further reduced by adopting double DQN approach [50]. By using two networks to decouple the action selection from the action evaluation, where the first network $\mathcal{Q}$ is used to derive the best action to take for the next state and the target network $\hat{\mathcal{Q}}$ is used to calculate the target Q-value of taking that action at the next state, we can then reformulate (15) as follows:

$$\delta_i = (r_t + \gamma \hat{\mathcal{Q}}(s_{t+1}, \operatorname*{argmax}_{a \in \mathcal{A}} \mathcal{Q}(s_{t+1}; a; \Theta); \Theta^-) - \mathcal{Q}(s_t, a_t; \Theta)). \quad (17)$$

Therefore, by reducing the overestimation of the Q-values, we can train the network faster and stabilize the learning process. Algorithm 1 summarizes the resulting algorithm, namely Prioritized Double Deep Q-Learning (PDDQL) algorithm.

## VI. Performance Evaluation

### A. Simulation Settings

In all the simulations, unless otherwise stated, we set the data packet arrival to follow the Poisson distribution with mean $\lambda = 3$ and set that of the radar activity requests (RARs) to follow a Bernouli distribution with $p_{radar} = 0.5$. The data queue of the JRC node can store up to 50 packets while RAR queue is set to store up to 5 RARs. The latency thresholds $\tau_{data}$ and $\tau_{radar}$ are set to 5 and 3 time units, respectively. The jammer has two transmit power levels, i.e $\mathbf{P}^J = \{0W, 10W\}$ [10]. To emulate the jammer's energy constraint, we set the jamming probability to 0.5, i.e., the jammer attacks and stays idle with equal probabilities. When the jammer is detected, the JRC node can either backscatter 3 packets on the jamming signals or use rate adaptation to transmit 1 packet. If the jammer is not detected, then the JRC node can actively transmit 3 packets or perform 1 radar sensing operation. If the JRC node chooses not to perform deception and the jammer does not attack the channel, it can actively transmit 4 packets, use rate adaptation to transmit 2 packets or perform 1 radar sensing operation [22], [36], [41].

[10] Note that the constraint on the power levels of the jammer can be relaxed to take values from the continuous space. A common method to extend Q-learning to work in continuous spaces is the use of actor-critic approach based on deep deterministic policy gradient (DDPG) [51], which we leave for the future work.

**Algorithm 1:** PDDQL Based Optimal Defense Algorithm

Initialize: replay memory $\mathbb{M}$ to capacity $N$, $p_1$=1, replay period $K$, mini-batch $k$, $\Delta = 0$, step-size $\eta$, $\alpha$, $\beta$;
Initialize $\mathcal{Q}$ with random weights $\Theta$ and $\hat{\mathcal{Q}}$ with weights $\Theta^- = \Theta$;
**for** $t = 1, 2, \ldots$ **to** *convergence* **do**
  With probability $\epsilon$ perform any feasible action $a_t$, otherwise perform $a_t = \underset{a \in \mathcal{A}}{\operatorname{argmax}}\ \mathcal{Q}^*(s_t, a)$;
  Perform action $a_t$ and get next state $s_{t+1}$;
  Store transition $(s_t, a_t, r_t, s_{t+1})$ in $\mathbb{M}$ with maximal priority $p_t = \max_{i<t} p_i$;
  **if** $t \equiv 0 \ mod \ K$ **then**
    **for** $j = 1$ **to** $k$ **do**
      Sample transition $j$ from $\mathbb{M}$;
      Compute importance-sampling weight $w_j$ using (16);
      Compute TD error $\delta_j$ using (17);
      Update transition priority: $p_j \leftarrow |\delta_j|$;
      Perform gradient descent and accumulate weights:
      $\Delta = \Delta + w_j \delta_j \nabla \mathcal{Q}(s_{j-1}, a_{j-1}; \Theta)$
    **end**
    Update weights: $\Theta \leftarrow \Theta + \eta\Delta$, reset $\Delta$;
    Every $C$ steps, reset $\Theta^- = \Theta$;
  **end**
**end**

Besides, we set the minimum thresholds for the SINR at the data receiver and radar receiver to be 10 $dB$ and 15 $dB$, respectively. Therefore, if the actual SINR is higher than the corresponding threshold, the executed action is then considered to be successful and a reward is obtained for the state-action pair $(s, a)$ [11]. The JRC node is operating at the $f_c = 5.9\,GHz$ frequency band and uses 50 $MHz$ bandwidth. We consider one fixed target located randomly in a $[5, 100]$ meter range with nonfluctuating radar cross section.

We adopt two baseline algorithms to analyze the performances of our proposed reinforcement learning based solutions, namely, Fixed with No Deception (FND) and Fixed With Deception (FWD):

- FND: This algorithm takes actions based on a fixed probability vector for two solely actions: active transmission with the probability of 0.6 and radar sensing with the probability of 0.4. We set the active transmission probability to be higher to go inline with our initial assumptions in the simulation settings, where data arrival rate is set to be higher than RAR arrival rate.
- FWD: In this algorithm, in addition to the actions of active transmission and radar sensing, now the algorithm can also perform data or radar deception at the beginning of the time slot with the probabilities of 0.4 and 0.3, respectively. The probabilities of active transmission and radar sensing actions are now set at 0.1 and 0.2, respectively. With deception implemented and upon detection of jammer's presence, the JRC node switches directly to data transmission with backscattering. If no jammer is detected, the JRC node continues to perform the desired action initially intended, i.e., active transmission or radar sensing.

[11] These SINR values are chosen to guarantee an acceptable data and secrecy rates for data communication and minimize radar false alarm rate. Designing the optimal values of these thresholds is beyond the scope of our work. Further details can be found in [13].

### B. Performance Results

*1) **Convergence of Deep Reinforcement Learning Approaches**:* We first start by showing the convergence speeds of different RL algorithms in the considered system. As illustrated in Figure 5(a), the PDDQL algorithm is able to converge faster to an optimal solution in the first $10^5$ iterations compared to other RL algorithms, which affirms the outstanding performance of the proposed PDDQL algorithm. The DQL algorithm is also able to converge to a similar optimal solution but its convergence speed is slower than that of PDDQL by a factor of 5. This shows the effectiveness of the introduced modifications on DQL algorithm, i.e, prioritized experience replay and double Q-network. However, the Q-Learning algorithm is not able to reach an equivalent performance even after $10^6$ iterations which is due to the slow convergence problem when the state space is very large. Even though Q-Learning is proved theoretically to reach an optimal solution [44], the large state and action spaces of our MDP make it hard to fine tune the learning parameters to reach the optimal solution in practise as illustrated in Figure 5(a). As such, since the proposed PDDQL algorithm can learn a better strategy over a huge state space, the multi-function node can be resilient to a variety of jamming attacks.

Figure 5(b) shows the learning time required by each algorithm to reach its optimal solution. We observe that Q-Learning is super slow and requires a huge amount of time with no final convergence. We also observe that PDDQL does not require too much time to converge even though it uses double neural networks and prioritized memory. This is explained by the fact that PDDQL algorithm requires less number of episodes to converge compared to that of the DQL algorithm which compensates the learning time required for each episode and makes PDDQL algorithm the fastest among other algorithms. Note that since the performance difference between DQL and PDDQL is related to convergence speed and not average reward, and thus both of the algorithms will converge to an equivalent optimal policy, we omit the DQL algorithm from our following comparisons and use the PDDQL to represent DRL solutions.

*2) **Optimal Policy under Different Jamming Strategies**:* Next, we analyze how the PDDQL algorithm performs with two different jamming cases: attacking all types of signals and attacking only data signals. We show how the system can adapt its learned policy, i.e., the actions taken at different states. First, we define 6 root states from our system space $\mathcal{S}$ necessary for our analysis as follows:

- $S1 : \{l = 1, c = 1\}$= {Data Deception, Jamming}.

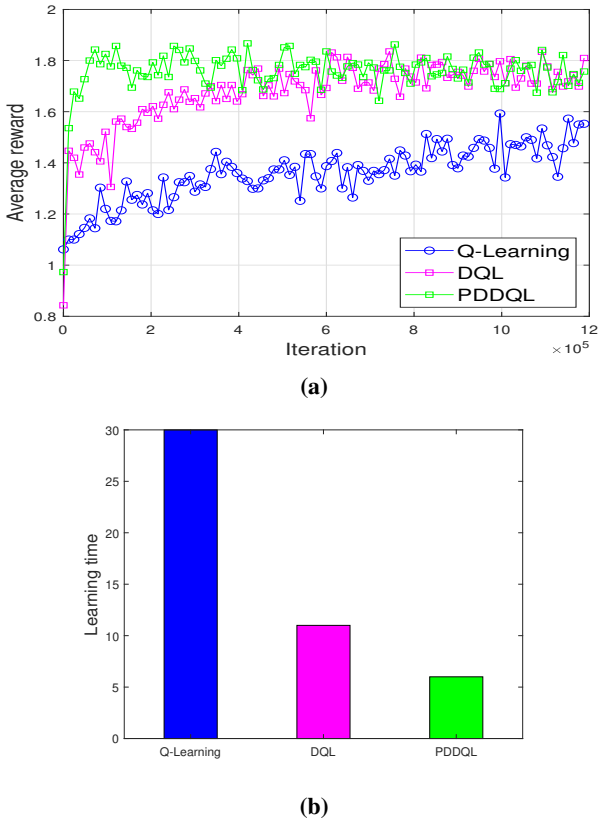


**Figure 5.** Convergence rate and learning time of various RL algorithms.

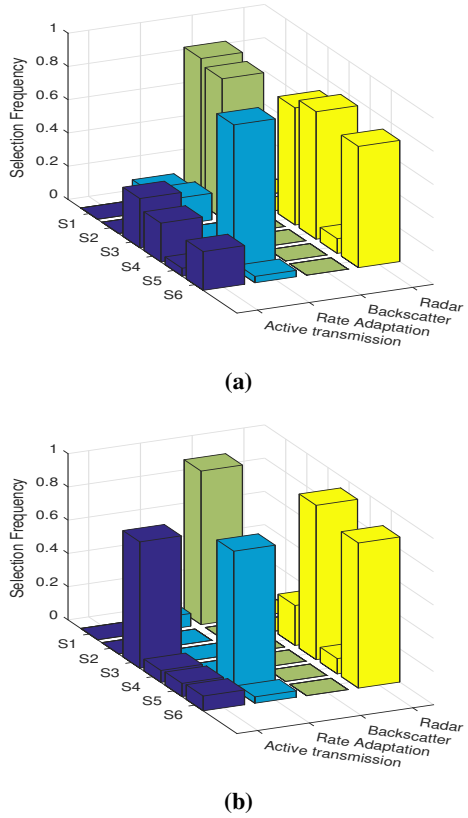


**Figure 6.** Selection frequency when: (a) attacking all types of signals, (b) attacking data signals only.

- $S2 : \{l = 2, c = 1\}$= {Radar Deception, Jamming}.
- $S3 : \{l = 1, c = 0\}$= {Data Deception, No Jamming}.
- $S4 : \{l = 2, c = 0\}$= {Radar Deception, No Jamming}.
- $S5 : \{l = 0, c = 1\}$= {No Deception, Jamming}.
- $S6 : \{l = 0, c = 0\}$= {No Deception, No Jamming}.

where $S1$, for example, is the root state for all states with all possible values for data and radar queue. Specifically, the JRC node is performing data deception and the jammer is attacking the channel. Other root states are defined similarly to cover all possible combinations. Note that these root states are a combination of the jammer state and the initial action of the JRC node at the beginning of the time slot. As such, the frequency of visiting each state is an important metric which reflects how the JRC node learns to perform deception or not. By analyzing the decision making process of our system, we can assert a high level of trustworthiness of the proposed DRL algorithm.

Figure 6(a) shows the frequency distribution of each action in different root states, which is defined as the number of times of taking one action over the total time period. When the jammer is present and either data deception $(S1)$ or radar deception $(S2)$ is performed, the JRC node chooses to backscatter on the jammer's signals for 85% of the time. However, if deception is not performed and the jammer attacks, as in $S5$, the JRC node chooses to transmit data using rate adaptation for 86% of the time. It is the best action to take in this case because the JRC node does not have any prior information that the jammer will attack during this time slot and cannot use backscattering. Moreover, the JRC node performs radar sensing less than 10% in $S5$, which indicates that it is able to learn that when no deception is performed, it is better not to perform radar sensing as there is a high risk of being jammed. When deception is performed and the jammer does not attack the channel, i.e., $S3$ and $S4$, the JRC node dedicates 75% of the time slots for radar sensing action and only 25% for active transmission. This is explained by the fact that RARs cannot tolerate delay and should be performed as soon as they arrive.

We then analyze how the system behaves if the jammer attacks only data transmission. From Figure 6(b) we can see that in $S3$, the JRC node is able to learn that the jammer is attacking only data transmission, and thus the JRC node is giving more time slots for active transmission (more than 75%) compared to 25% in the case where the jammer attacks all types of signals. Moreover, when radar deception is performed and no jamming signals are detected in the channel as in $S4$, the JRC node is giving 90% of time slots for radar sensing and this state is now more frequently visited compared to the scenario when the jammer attacks all types of signals as shown in Figure 7. The JRC node is able to learn to perform radar sensing safely after radar deception, but give 10% of the time in this state to data transmission. We also see from Figure 6(b) that the JRC node is able to learn that it is safer to perform radar sensing when not performing deception as the jammer

is only interested in jamming data signals. In addition, we observe from Figure 6(b) and Figure 7 that most of radar sensing actions are done in $S4$ and $S6$ and in all other states, the JRC node tends to perform data transmission. Therefore, we can conclude that the deep reinforcement learning solution can enable the JRC node to achieve the effective strategy of switching between the initial goals of data and radar deception when facing smart attacks with different objectives.

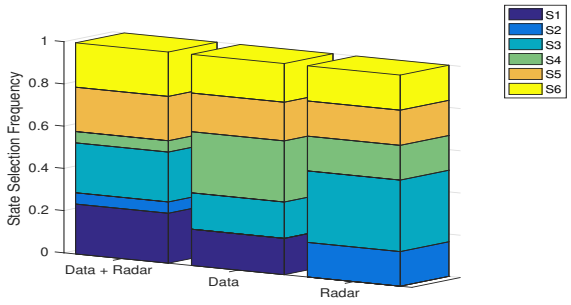


**Figure 7.** State selection frequency for different attack scenarios.

*3) **Performance Evaluation under Different Scenarios**:*

*a) Varying radar reward:* First, we fix the jamming probability at 0.5 and vary $r_{rdr}$, i.e., the reward value for taking the action "radar sensing". This is to observe how our proposed model performs when the radar sensing has a higher priority than data transmission and vice versa. Figure 8 is the result of running the PDDQL algorithm. The Q-Learning and DQL algorithms also give similar results, but those are omitted here for brevity.

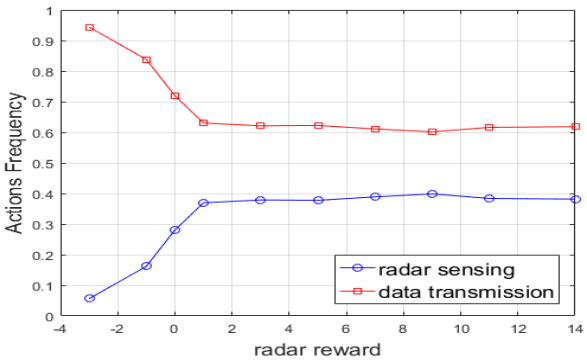


**Figure 8.** Ratio of taking different actions when the radar reward is varied.

Figure 8 presents the action frequency, which indicates the devoted time to radar and communication operations. We observe that with negative rewards, most of the actions taken are related to data transmission (more than 80%) while radar sensing is performed only 20% of the time. However, when increasing the reward value $r_{rdr}$ of the radar sensing action, the system starts giving more time slots for radar sensing. Unexpectedly, the system does not dedicate more time slots for radar sensing when increasing its reward value. This is due to the fact that we have more data transmission requests than radar sensing requests, i.e., the queue of RAR becomes empty quickly. Our proposed model can successfully learn this without prior knowledge.

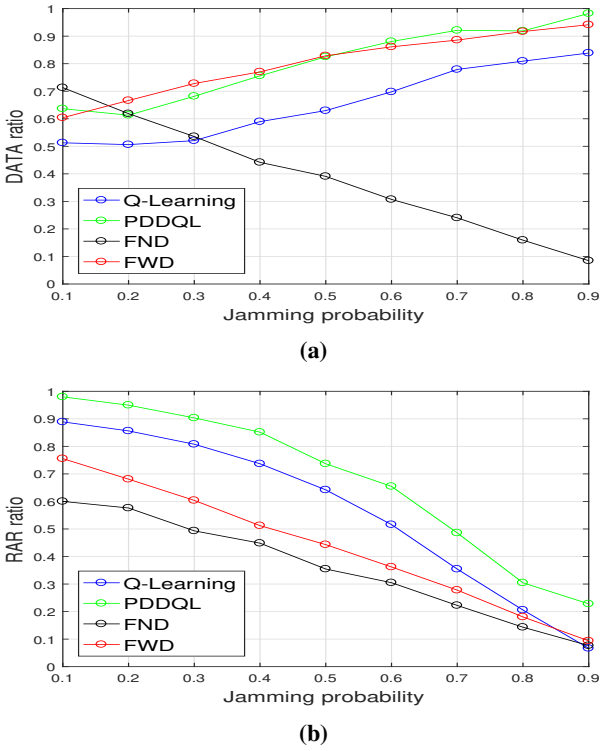


**Figure 9.** Data and RAR ratio versus jamming probability.

*b) Varying jamming probability:* We then vary the jamming probability and observe the performance of the system using different algorithms. As shown in Figure 9, as the jamming probability increases, successful data transmission increases for all the algorithms except for FND. This is due to the low proportion of time given for data transmission in FND algorithm. Interestingly, we can observe that the performance of PDDQL and FWD are similar. This is due to the hyperparameters of FWD which use deception strategy and give most of the time slots for backscattering if jamming signals are detected.

The performance of radar sensing decreases for all algorithms. This is due to the fact that there is no alternative solution that can make the JRC node to perform radar sensing when the spectrum is full of noise introduced by the jammer. We also observe that the performance of all fixed solutions are worst than those of the DRL solutions for radar sensing. This is also due to the hyperparameters settings of fixed solutions which cannot change over time, showing the importance of dynamic approaches. Moreover, the gap between the performance in data transmission and radar sensing is very high for FWD, which makes it not useful in practical applications, e.g, autonomous driving. The Q-Learning based solutions are more preferred as the gap is reduced and more balance is achievable between data transmission and radar sensing.

### C. Discussions

We examined in this section the performance of our proposed system design with focus on JRC systems under different scenarios and attacks. We highlight that the incorporation of different techniques: queuing concepts, ambient backscatter technology, deception strategy and rate adaptation algorithms can significantly help building a more robust wireless system. This strategy is straightforwardly applicable to other multi-function systems. An important takeaway from our results is that simple anti-jamming techniques, e.g, frequency hopping, are no more favorable as dedicated attacks can significantly reduce their effectiveness. Instead, by designing more complex systems and integrating different techniques jointly with DRL, we can build more resilient systems.

The proposed DRL algorithms were trained through online simulation. An important step before deployment is to train these algorithms on real systems, where the accuracy of observations and reward shaping plays an important role on the system performance. During the training phase we can provide the agent with real target information, e.g, distance and angle, and jamming power to compute the loss. The reward values can be then computed based on the accuracy of the estimated parameters. The reward values related to data transmission can be exchanged using acknowledge messages from the gateway. In SWIPT systems, the number of harvested energy units can be exchanged through communication channel.

Another motivation to design a unified framework for JRC and SWIPT systems is to minimize training costs. Specifically, instead of training a model for each system, we train the model on one system and then use transfer learning to fine tune the learned model into other systems. The process of transfer learning significantly reduces the amount of data and time required for training and hence, is practical and efficient in real deployment of wireless systems.

## VII. Conclusion and Future Works

A novel architecture has been proposed to counteract reactive and smart jammers in multi-function wireless systems. The optimization problem of the system functionalities is formulated as a joint scheduling problem of multiple queues. As different techniques are merged together in the proposed framework, an advanced two-step MDP architecture is proposed to accurately model the system dynamics. A deep reinforcement learning solution based on prioritized replay memory and double Q-Learning is then developed to derive an optimal strategy for the multi-function node. The proposed abstraction shows interesting results when evaluated on JRC systems and is straightforwardly applicable for SWIPT systems. In particular, the framework is shown to be able to handle a variety of jamming attacks while satisfying the joint system requirements. With deception strategy and DRL, we are able to suppress the jammer from attacking continuously and thus perform radar sensing safely. Moreover, the use of ambient backscatter technology helps leveraging the jamming signals significantly.

The in-depth analysis of the system under different scenarios and variations reveals the proposed deep reinforcement learning algorithm has a good level of trustworthiness, but more advanced techniques for deep learning models interpretability should be considered in future works. Another extension of this research is more than two functions supported in the considered wireless systems.

## References

[1] M. A. Hossain, R. Md Noor, K. A. Yau, I. Ahmedy, and S. S. Anjum, "A survey on simultaneous wireless information and power transfer with cooperative relay and future challenges," *IEEE Access*, vol. 7, pp. 19 166–19 198, Jan. 2019.

[2] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3834–3862, Jun. 2020.

[3] P. Kumari, J. Choi, N. Gonzalez-Prelcic, and R. W. Heath, "IEEE 802.11ad-based radar: An approach to joint vehicular communication-radar system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3012–3027, Apr. 2018.

[4] D. Garmatyuk, J. Schuerger, and K. Kauffman, "Multifunctional software-defined radar sensor and data communication system," *IEEE Sensors Journal*, vol. 11, no. 1, pp. 99–106, Jan. 2011.

[5] S. Ouedraogo, I. D. H. Saenz, R. Guinvarc'h, and R. Gillard, "Design and experimental validation of multifunction antenna with direct modulation for radar and communication," *Progress In Electromagnetics Research*, vol. 164, pp. 17–25, 2019.

[6] P. Lu, C. Song, and K. M. Huang, "A two-port multipolarization rectenna with orthogonal hybrid coupler for simultaneous wireless information and power transfer (SWIPT)," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 10, pp. 6893–6905, Oct. 2020.

[7] X. Wang and J. Xu, "Co-design of joint radar and communications systems utilizing frequency hopping code diversity," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–6.

[8] C. Lai, C. Chen, and L. Wang, "On-demand density-aware uav base station 3d placement for arbitrarily distributed users with guaranteed data rates," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 913–916, Feb. 2019.

[9] Y. Zeng, Y. Ma, and S. Sun, "Joint radar-communication with cyclic prefixed single carrier waveforms," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4069–4079, Apr. 2020.

[10] K. Grover, A. Lim, and Q. Yang, "Jamming and anti-jamming techniques in wireless networks: a survey," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 17, no. 4, pp. 197—215, Dec. 2014.

[11] A. Mauro, D. Papini, and N. Dragoni, "Security challenges for energy-harvesting wireless sensor networks," in *International Conference on Pervasive Embedded Computing and Communication Systems (PECCS)*, Feb. 2012, pp. 422–425.

[12] L. Ismail, D. Niyato, S. Sun, D. T. Hoang, D. I. Kim, and Y.-C. Liang, "Jamming mitigation in jrc systems via deep reinforcement learning and backscatter-supported intelligent deception strategy," *2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS)*, pp. 1053–1058, 2021.

[13] A. Deligiannis, A. Daniyan, S. Lambotharan, and J. A. Chambers, "Secrecy rate optimizations for MIMO communication radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 5, pp. 2481–2492, Oct. 2018.

[14] N. Zhao, Y. Cao, F. R. Yu, Y. Chen, M. Jin, and V. C. M. Leung, "Artificial noise assisted secure interference networks with wireless power transfer," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1087–1098, Feb. 2018.

[15] A. Garnaev, W. Trappe, and A. Petropulu, "Optimal design of a dual-purpose communication-radar system in the presence of a jammer," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.

[16] G. Noubir, R. Rajaraman, B. Sheng, and B. Thapa, "On the robustness of IEEE 802.11 rate adaptation algorithms against smart jamming," in *Proceedings of the fourth ACM conference on Wireless network security - WiSec '11*, 2011, pp. 97–108.

[17] L. Kong, Y. Xu, Y. Zhang, X. Pei, M. Ke, X. Wang, W. Bai, and Z. Feng, "A reinforcement learning approach for dynamic spectrum anti-jamming in fading environment," in *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, Oct. 2018, pp. 51–58.

[18] X. Liu, Y. Xu, L. Jia, Q. Wu, and A. Anpalagan, "Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach," *IEEE Communications Letters*, vol. 22, no. 5, pp. 998–1001, May 2018.

[19] L. Xiao, *Anti-Jamming Transmissions in Cognitive Radio Networks*. Springer International Publishing, 2015.
[20] M. K. Hanawal, M. J. Abdel-Rahman, and M. Krunz, "Joint adaptation of frequency hopping and transmission rate for anti-jamming wireless systems," *IEEE Transactions on Mobile Computing*, vol. 15, no. 9, pp. 2247–2259, Sep. 2016.
[21] N. V. Huynh, D. N. Nguyen, D. T. Hoang, and E. Dutkiewicz, ""jam me if you can:" defeating jammer with deep dueling neural network architecture and ambient backscattering augmented communications," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 11, pp. 2603–2620, Nov. 2019.
[22] D. T. Hoang, D. N. Nguyen, M. A. Alsheikh, S. Gong, E. Dutkiewicz, D. Niyato, and Z. Han, "Borrowing arrows with thatched boats: The art of defeating reactive jammers in IoT networks," *IEEE Wireless Communications*, vol. 27, no. 3, pp. 79–87, Jun. 2020.
[23] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Access*, vol. 8, pp. 58 443–58 469, Mar. 2020.
[24] Y. Guo, G. Liao, J. Li, and H. Kang, "An improved range deception jamming recognition method for bistatic MIMO radar," *Digital Signal Processing*, vol. 95, p. 102578, Dec. 2019.
[25] S. Ak and S. Brüggenwirth, "Avoiding jammers: A reinforcement learning approach," in *2020 IEEE International Radar Conference (RADAR)*, Apr. 2020, pp. 321–326.
[26] Q. Liu, K. S. Yildirim, P. Pawełczak, and M. Warnier, "Safe and secure wireless power transfer networks: challenges and opportunities in rf-based systems," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 74–79, Sep. 2016.
[27] D. T. Hoang, D. Niyato, P. Wang, and D. I. Kim, "Performance analysis of wireless energy harvesting cognitive radio networks under smart jamming attacks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 2, pp. 200–216, Jun. 2015.
[28] B. Li, A. P. Petropulu, and W. Trappe, "Optimum co-design for spectrum sharing between matrix completion based MIMO radars and a MIMO communication system," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4562–4575, Sep. 2016.
[29] N. Cao, Y. Chen, X. Gu, and W. Feng, "Joint bi-static radar and communications designs for intelligent transportation," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 060–13 071, 2020.
[30] Y. L. Sit, B. Nuss, and T. Zwick, "On mutual interference cancellation in a MIMO OFDM multiuser radar-communication network," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3339–3348, Apr. 2018.
[31] P. Buchholz, "An EM-algorithm for MAP fitting from real traffic data," in *Computer Performance Evaluation. Modelling Techniques and Tools*. Springer Berlin Heidelberg, 2003, pp. 218–236.
[32] T. Kim, D. J. Love, and B. Clerckx, "Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems?" *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1654–1669, Apr. 2011.
[33] Naval Air Warfare Center Weapons, *Electronic Warfare and Radar Systems Handbook: Engineering Handbook*. Washington, DC , USA: Avionics Department, 2013.
[34] A. R. Chiriyath, B. Paul, and D. W. Bliss, "Radar-communications convergence: Coexistence, cooperation, and co-design," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 1, pp. 1–12, Mar. 2017.
[35] G. R. Curry, *Radar System Performance Modeling*, 2nd ed. Artech House Radar Library, 2005.
[36] P. Ren, A. Munari, and M. Petrova, "Performance tradeoffs of joint radar-communication networks," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 165–168, Feb. 2019.
[37] K. Firouzbakht, G. Noubir, and M. Salehi, "On the capacity of rate-adaptive packetized wireless communication links under jamming," in *Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks - WISEC*, Apr. 2012, pp. 3–14.
[38] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '05*. ACM Press, 2005, p. 46–57.
[39] X. Chang and C. Dong, "A barrage noise jamming method based on double jammers against three channel SAR GMTI," *IEEE Access*, vol. 7, pp. 18 755–18 763, 2019.
[40] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 21–24, Jan. 2007.
[41] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: Wireless communication out of thin air," in *Proceedings of the ACM SIGCOMM*, Aug. 2013, pp. 39–51.
[42] J. Moghaddasi and K. Wu, "Multifunctional transceiver for future radar sensing and radio communicating data-fusion platform," *IEEE Access*, vol. 4, pp. 818–838, Feb. 2016.
[43] J. R. Guerci, R. M. Guerci, A. Lackpour, and D. Moskowitz, "Joint design and operation of shared spectrum access for radar and communications," in *2015 IEEE Radar Conference (RadarCon)*, May 2015, pp. 0761–0766.
[44] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.
[45] F. Kong, Z. Jin, J. Cho, S. Jeon, and S. Lee, "Optimizing spectrum sensing time for energy-efficient crsns," in *2016 13th International Conference on Embedded Software and Systems (ICESS)*, 2016, pp. 1–6.
[46] Y. Pan, X. Da, and H. Hu, "Joint optimization of sensing time and power allocation for UAV cognitive radio systems," in *Proceedings of the 5th International Conference on Communication and Information Processing*, 2019, p. 254–258.
[47] H. nan Wang, N. Liu, Y. yun Zhang, D. wei Feng, F. Huang, D. sheng Li, and Y. ming Zhang, "Deep reinforcement learning: a survey," *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 12, pp. 1726–1744, Oct. 2020.
[48] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[49] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," 2016. [Online]. Available: https://arxiv.org/abs/1511.05952
[50] H. v. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Feb. 2016, p. 2094–2100.
[51] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proceedings of the International Conference on Learning Representations*, 2016.

**Lotfi Ismail** is working towards his Ph.D in the School of Computer Science and Engineering at Nanyang Technological University, Singapore. Ismail received the B.Eng degree from Oran University of Science and Technology, Algeria, in 2015 and the master degree in Emerging Networks, Security and Multimedia from Oran-1 University, Algeria in 2017. His main research interests are in the area of resource management and optimization in communication networks.

**Dusit Niyato** (Fellow, IEEE) received the B.Eng. degree from the King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1999 and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests are in the area of energy harvesting for wireless communication, Internet of Things (IoT), and sensor networks.

**Sumei Sun** (Fellow, IEEE) is currently a Principal Scientist and the Head of the Communications and Networks Department with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), and an Adjunct Professor with the National University of Singapore. Her current research interests include cognitive communications and networks, next-generation wireless communications, and the industrial Internet of Things.

**Dinh Thai Hoang** (Member, IEEE) received the Ph.D. degree in computer science and engineering from Nanyang Technological University, Singapore, in 2016. He is currently a Faculty Member at the School of Electrical and Data Engineering, University of Technology Sydney, Australia. His research interests include emerging topics in wireless communications and networking, such as ambient backscatter communications, vehicular communications, cybersecurity, the IoT, and 5G networks.

**Yonghui Li** (Fellow, IEEE) received the Ph.D. degree from the Beijing University of Aeronautics and Astronautics in 2002. Since 2003, he has been with the Centre of Excellence in Telecommunications, The University of Sydney, Australia, where he is currently a Professor with the School of Electrical and Information Engineering. His current research interests are in the area of wireless communications, with a particular focus on MIMO, machine to machine communications, and cooperative communications.

**Dong In Kim** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Southern California, USA in 1990. He was a tenured Professor with the School of Engineering Science, Simon Fraser University, Canada. Since 2007, he has been an SKKU-Fellowship Professor with the College of Information and Communication Engineering, Sungkyunkwan University (SKKU), Suwon, South Korea. He is also a Fellow of the Korean Academy of Science and Technology.