Anti-Jamming Games in Multi-Channel Cognitive Radio Networks

Yongle Wu, Beibei Wang, *Member, IEEE*, K. J. Ray Liu, *Fellow, IEEE*, and T. Charles Clancy, *Senior Member, IEEE*

***Abstract*—Crucial to the successful deployment of cognitive radio networks, security issues have begun to receive research interests recently. In this paper, we focus on defending against the jamming attack, one of the major threats to cognitive radio networks. Secondary users can exploit the flexible access to multiple channels as the means of anti-jamming defense. We first investigate the situation where a secondary user can access only one channel at a time and hop among different channels, and model it as an anti-jamming game. Analyzing the interaction between the secondary user and attackers, we derive a channel hopping defense strategy using the Markov decision process approach with the assumption of perfect knowledge, and then propose two learning schemes for secondary users to gain knowledge of adversaries to handle cases without perfect knowledge. In addition, we extend to the scenario where secondary users can access all available channels simultaneously, and redefine the anti-jamming game with randomized power allocation as the defense strategy. We derive the Nash equilibrium for this Colonel Blotto game which minimizes the worst-case damage. Finally, simulation results are presented to verify the performance.**

***Index Terms*—Cognitive radio, anti-jamming games, learning schemes, defense strategies.**

I. Introduction

AS A REVOLUTIONARY communication paradigm that enables more efficient and intelligent usage of the spectrum resources, cognitive radio technology has been receiving growing attention in recent years since it was originally proposed [1]. In a cognitive radio network [2], unlicensed users (secondary users) are allowed to access licensed bands on a non-interference basis to legacy spectrum holders (primary users).

Since secondary users usually compete for the limited spectrum resources and are capable of acting intelligently, it is reasonable to assume they are selfish in nature, and hence game theory has been widely applied as a flexible and proper tool to model and analyze their behavior in the network [3] [4]. For example, the spectrum access was formulated into a potential game in [5] where the system equilibrium was approached by iterative updates, and a repeated game framework was applied in [6] for selfish users to share a common spectrum band. Driven by pursuit of higher payoffs, selfish users may not reveal the truth when exchanging private information. To combat the selfish behavior, cheat-proof strategies have been proposed for spectrum sharing in [7], and collusion-resistant strategies have been developed for spectrum auctions in [8][9]. A distributed channel sensing and access policy base on the adversarial bandit problem was presented for cognitive radio under time-varying channels in [10]. Nevertheless, most of these schemes are not immune to malicious attacks.

In fact, cognitive radio networks are extremely vulnerable to malicious attacks, partly because secondary users do not own the spectrum, and hence their opportunistic access cannot be protected from adversaries. Moreover, highly dynamic spectrum availability and often distributed network structures make it difficult to implement effective security countermeasures. In addition, as cognitive radio networks benefit from technology evolution to be capable of utilizing spectrum adaptively and intelligently, the same technologies can also be exploited by malicious attackers to launch more complicated and unpredictable attacks with even greater damage. Therefore, ensuring security is of critical importance to the successful deployment of cognitive radio networks.

However, it was not until recent years that security issues began to receive research interest. For instance, in [11], the primary user emulation attack was described and a transmitter verification scheme was proposed to distinguish a primary user from other sources; [12] discussed the attack where malicious users attempted to mislead the learning process of secondary users; denial-of-service attacks were considered and potential protection remedies were discussed in [13]; in [14], a malicious user reporting false sensing results would be found and excluded from the collaborative spectrum sensing when the calculated "suspicious" level was high; in [15], an information secrecy game was developed to foster collaboration between primary users and secondary users against eavesdroppers.

In this paper, we mainly focus on jamming attacks, one of the major threats to cognitive radio networks, where several malicious attackers intend to interrupt the communications of a secondary user by injecting interference. Because cognitive radio technology enables flexible access to different channels, secondary users are able to transmit information over multiple channels, and may exploit such flexibility as a way to hide from attackers. On the other hand, attackers are also intelligent such that they can come up with efficient attack

Manuscript received 11 January 2011; revised 14 July 2011. This paper was presented in part at the IEEE Global Communications Conference (Globecom).

Y. Wu and B. Wang are with Qualcomm Incorporated, San Diego, CA 92121, USA (e-mail: {yonglew,beibeiw}@qualcomm.com).

K. J. R. Liu is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA (e-mail: kjrliu@umd.edu).

T. C. Clancy is with the Bradley Department of Electrical and Computer Engineering and the Hume Center for National Security and Technology, Virginia Tech, Arlington, VA 22203, USA (e-mail: tcc@vt.edu).

Digital Object Identifier 10.1109/JSAC.2012.120102.

strategies. Therefore, this scenario is modeled as a zero-sum anti-jamming game, in which the two players, namely, the secondary users and the attackers, have opposite objectives.

We first investigate the situation where a secondary user can access only one channel at a time. In order to reduce the probability of being jammed, the defense strategy is to hop across multiple channels. We analyze the first few rounds of the arms race between the secondary user and attackers, and derive a channel hopping strategy based on the Markov decision process (MDP) [16]. We further show that such an MDP-based hopping provides a good approximation to the game equilibrium which is difficult to analyze directly.

Moreover, in order to determine the MDP-based defense strategy, a secondary user needs to know some attacker information which may not be directly available. Hence, the secondary user has to observe and learn from the environment. In this paper, we first propose a learning process where the secondary user estimates the useful parameters based on past observations using maximum likelihood estimation (MLE); then, as an alternative, we apply Q-learning [17] for the secondary user to learn and update the defense strategy without knowing the underlying Markov model.

Finally, we extend our model to the situation where a secondary user is able to access all available channels simultaneously, for example, when the secondary user is equipped with multiple radios. Under such a circumstance, the defense strategy is no longer to hop between channels, but instead, is to allocate power in these channels in a randomized fashion. We show that the game can be formulated as a Colonel Blotto game [18], and derive the equilibrium strategy in terms of probability distribution on allocated power for this game. As shown later in the paper, the defense strategy obtained from the equilibrium can minimize the worst-case damage caused by attackers.

A. Related Works

There have been quite a few papers on jamming attacks in the wireless ad hoc networks, such as [19]–[23]. In [19], a jamming game was formulated with the transmission cost considered, and generalized water-filling was proved to be the unique Nash equilibrium. In [20], the blocking probability was analyzed for different kinds of attack strategies and defense strategies. In [21][22][23], an uncoordinated frequency hopping scheme was developed where a transmitter and several receivers followed their own hopping patterns to mitigate the jamming impact, and a communication link was established when the transmitter and a receiver happened to choose the same unjammed channel.

However, the problem becomes more complicated in a cognitive radio network where primary users' access has to be taken into consideration. Combating the jamming attack was modeled as a dogfight game in [24] and [25], with the assumption of known and unknown channel statistics, respectively. In the former, the Nash equilibrium of a one-shot game was derived where the probability distribution of hopping depended on the quality of different channels, and this equilibrium was further applied to a multi-stage game. In the latter, the algorithm of adversarial bandit problem was adapted to learn the optimal defense strategy using the experience of spectrum access. In [26], the authors considered a malicious user enabled by cognitive radio technology who launched jamming attacks to a multi-channel 802.11 network. The impact of switching delays and jamming durations was evaluated.

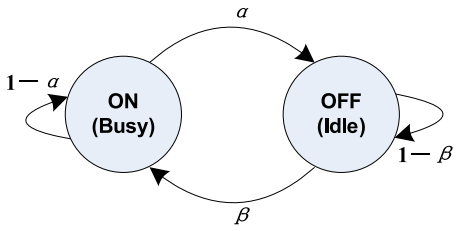


Fig. 1. An ON-OFF model for primary users' spectrum usage.

The rest of this paper is organized as follows. In Section II, the system model is described. The optimal defense strategy given perfect information is derived in Section III, while learning from the environment is proposed in Section IV. Section V extends to the situation where a secondary user is able to access multiple channels simultaneously. Section VI presents some simulation results, and Section VII concludes the paper.

II. System Model

Consider the situation where a secondary user (e.g., a base station for a secondary network) opportunistically accesses the spectrum bands. Assume there are M licensed channels in total, each licensed channel is time-slotted, and the access pattern of primary users can be characterized by an ON-OFF model [27]. As shown in Fig. 1, one channel can either be busy (ON) or idle (OFF) in one time slot, and the state can be switched from ON to OFF (or from OFF to ON) with a transition probability α (or β). We assume all channels share the same model and parameters, but different channels are used by different primary users whose accesses are independent. In order to avoid interference to primary users, a secondary user has to synchronize with the primary network and detect the presence of the primary user at the beginning of each time slot. It is only when the primary user is absent that the secondary user is allowed to access the channel, which is also known as the "listen-before-talk" rule. Meanwhile, there are m malicious attackers intending to jam the secondary user's communications, and they coordinate with each other to maximize the damage. Since attackers are interested in jamming secondary users but not primary users, they have to first listen to the spectrum band to determine the presence of primary users, and then detect the existence of secondary users in that band. They will jam the band if secondary signals are found.

With interference power jammed into spectrum bands, the signal-to-interference-and-noise ratio (SINR) at the secondary user's receiver will be dragged down. We assume that the communication fails (e.g., packets cannot be decoded correctly) when the SINR drops below a certain threshold τ [28]. Denote the power constraint of a secondary user as p^B, and the power constraint of an attacker as i^B. All channel gains are assumed

to be 1 because they can be absorbed into the power constraint term. Furthermore, it is of interest to consider the case that the attacker is stronger than the secondary user, and we limit ourselves to the case $p^B \leq \tau i^B$. For example, when both users allocate all power to the same band, the secondary user always fails to communicate due to the poor SINR $p^B/(i^B+\sigma^2) < \tau$.

In different application scenarios, secondary users may have different capabilities. We first consider the case where a secondary user is equipped with a single radio, and hence can only sense and use one of the M candidate channels at any time slot. Later, we extend it to the case where a secondary user is equipped with multiple radios. Attackers are assumed to be comparable with the secondary user, that is, equipped with a single radio in the first case, and with multiple radios in the second case. In order to improve throughput, in the single-radio case, it is best for the secondary user to pour all power to a single band, and *channel hopping* is the defense strategy. For the multi-radio case, the secondary user could allocate power to several bands, and the defense can be fortified via optimal *power allocation*. The game under these two scenarios will be described in detail in Section III and V, respectively.

A secondary user receives a communication gain R whenever there is a successful transmission. For the single-radio case, channel hopping has some impact on throughput, since after tuning the frequency of transceivers, transmission cannot be started immediately due to the settling time of radio frequency (RF) devices. This cost is denoted by C. In addition, a secondary user suffers from a significant loss L when jammed, since normal communication is interrupted and considerable effort is needed to reestablish the link. As the secondary user aims at maximizing the payoff (communication gains minus cost and loss) but attackers have the opposite objective, this attack-and-defense problem can be formed as a *zero-sum game* between a secondary user and attackers.

In the game, attack and defense should be randomized; otherwise, a fixed pattern of one player will be taken advantage of by the opponent. It is worth pointing out that even if the randomized strategy is adopted, the transmitter and receiver of the secondary user can still stay coordinated (for example, tuning to the same channel after hopping) by initialization with the same random seed.

III. Anti-Jamming Channel Hopping Game

In this section, we investigate the attack and defense problem by modeling it as an anti-jamming game. For this game, it is desirable to know what could be possible attack strategies and what should be the optimal defense strategy. However, an attack-and-defense problem is often like an arms race: when an attacker updates the attack strategy, it is possible for the defender to come up with a new defense strategy that best defeats the new attack strategy, and vice versa. Although the game equilibrium is difficult to analyze, we show how players iteratively update their strategy against the opponent by going through the first several rounds of interaction between the secondary user and attackers. Our main effort is to develop an MDP-based defense strategy, which will be shown as a close approximation for the game equilibrium.

A. Game Formation

Recall that in the single-radio case, the secondary user will fail to meet the SINR constraint when an attacker puts his/her entire jamming power to the same channel. Thus, the secondary user could hop among multiple channels to hide from attackers. Meanwhile, attackers search over multiple channels in order to catch and jam the secondary user. It is inefficient if several malicious attackers tune their radios to the same channel to detect the secondary user; instead, they should coordinate not to overlap, detecting m channels in each time slot.

At the end of each time slot, the secondary user decides either to *stay* or to *hop* for the next time slot, based on the observation of the current and past slots. The secondary user receives an immediate payoff $U(n)$ in the nth time slot, which is the communication gain minus the cost and damage,

$$\begin{aligned} U(n) =& R \cdot \mathbf{1}(\text{Successful transmission}) - L \cdot \mathbf{1}(\text{Jammed}) \\ & - C \cdot \mathbf{1}(\text{Choosing the action 'hop'}), \end{aligned} \tag{1}$$

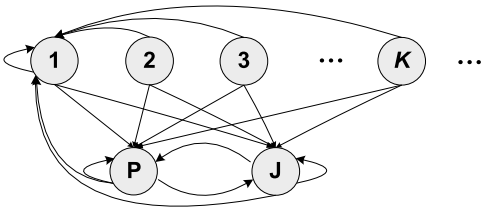
where $\mathbf{1}(\cdot)$ is an indicator function returning 1 when the statement in the parenthesis holds true and 0 otherwise. Because an employed strategy not only affects the current state but also has impact on the future, the payoff of this game $\overline{U}$, which the secondary user wants to maximize but malicious attackers want to minimize, is a discounted sum of payoffs,

$$\overline{U} = \sum_{n=1}^{\infty} \delta^n U(n), \tag{2}$$

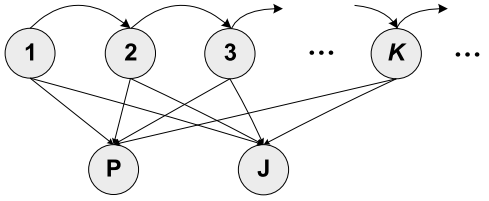
where the discount factor $\delta\,(0 < \delta < 1)$ measures the patience of the secondary user, that is, how much he/she values a future payoff over the current payoff.

Let us start with a naive jamming strategy, the *random attack*, where m attackers randomly choose m channels to detect in each time slot with equal probabilities, regardless of which channels have been detected in the past. Then, the question is what the secondary user should do in face of this random attack. Since every channel is equally probable to be detected by jammers, the secondary user cannot reduce the risk by hopping from one channel to another. Moreover, since channel hopping incurs some cost, the secondary user will be reluctant to tune the radio to another channel. Therefore, the secondary user should use a *minimal hopping* strategy, that is, staying in the same channel until it is unavailable when the primary user reappears.

The next iteration is how attackers would react to the secondary user's minimal hopping strategy. Knowing that the secondary user tends to stay in the same channel, attackers could sweep over all channels in order to find and jam the secondary user as soon as possible, as there is no need to revisit a channel that have been detected recently. Note that sweeping does not necessarily mean from lower frequency bands to higher frequency bands; attackers need to randomize the sweeping order to make it unpredictable to the secondary user. We name it the *sweeping attack*. Specifically, attackers coordinately tune their radios randomly to m undetected channels in each time slot, until this process starts over when



(a) Transition of states when taking the action *'hop'*.



(b) Transition of states when taking the action *'stay'*.

Fig. 2. Markov chains of state transitions when different actions are taken.

either all channels have been sensed or the secondary user has been found and jammed. Then, the next sweeping cycle starts.

Assuming that attackers stick to the sweeping attack, we will derive the optimal strategy that a secondary user should adopt. Note that now the channels being detected are dependent on what channels have been detected in the past, and hence we can model this scenario as a Markov decision process (MDP), from which the defense strategy is obtained.

B. MDP-Based Strategy

At the end of the nth time slot, the secondary user observes the state of the current time slot S_n, and chooses an action a_n, that is, whether to tune the radio to a new channel or not, which takes effect at the beginning of the next time slot. To set a clear distinction, states are denoted by upper-case letters while actions are denoted by lower-case letters. If the primary user occupied the channel or the secondary user was jammed in the nth time slot, denoted by $S_n = P$ and $S_n = J$, respectively, the secondary user has to hop to a new channel, i.e., $a_n = h$; otherwise, the secondary user has transmitted a packet successfully in the time slot, and possible actions are 'to hop' ($a_n = h$) and 'to stay' ($a_n = s$). If this is the Kth consecutive slot with successful transmission in the same channel, the state is denoted by $S_n = K$. For brevity, we will drop the time index n wherever there is no room for ambiguity in the rest of the paper. According to (1), the immediate payoff function depends on both the state and the action, i.e.,

$$U(S,a) = \begin{cases} R, & \text{if } S \in \{1,2,3,\ldots,\}, a = s; \\ R - C, & \text{if } S \in \{1,2,3,\ldots,\}, a = h; \\ -L - C, & \text{if } S = J; \\ -C, & \text{if } S = P. \end{cases} \quad (3)$$

The transition of states can be described by Markov chains, as shown in Fig. 2, where transition probabilities depend on which action has been taken. Hence, we use $p(S'|S,h)$ and $p(S'|S,s)$ to represent the transition probability from the current state S to a new state S' when taking action h and action s, respectively.

If the secondary user hops to a new channel, transition probabilities do not depend on the current state, and furthermore, the only possible new states are P (the new channel is occupied by the primary user), J (transmission in the new channel is detected by an attacker), and 1 (successful transmission begins in the new channel). The probability of entering state P after taking the action h can be approximated by the steady-state probability of the ON-OFF model in Fig. 1, i.e.,

$$p(P|S,h) = \frac{\beta}{\alpha+\beta} \triangleq \gamma, \ \forall S \in \{P, J, 1, 2, 3, \ldots, \}. \quad (4)$$

When there are plenty of channels, the time interval between visiting the same channel will be long, and the steady-state probability will be a good approximation. Provided that the new channel is available, the secondary user will be jammed with the probability m/M, since each attacker detects one channel without overlapping. As a result, transition probabilities are

$$\begin{aligned} p(J|S,h) &= (1-\gamma)\frac{m}{M}, \ \forall S \in \{P, J, 1, 2, 3, \ldots, \}; \\ p(1|S,h) &= (1-\gamma)\frac{M-m}{M}, \ \forall S \in \{P, J, 1, 2, 3, \ldots, \}. \end{aligned} \quad (5)$$

On the other hand, if the secondary user stays in the same channel, the primary user may reclaim the channel with probability β given by the ON-OFF model. With the primary user absent, the state will go to J if the transmission is jammed, and will increase by 1 otherwise. Note that s is not a feasible action when the state is in J or P. At state K, only $\max(M - Km, 0)$ channels have yet to be detected by attackers, but another m channels will be detected in the upcoming time slot; therefore, the probability of jamming conditioned on the absence of a primary user is given by

$$f_J(K) = \begin{cases} \frac{m}{M-Km}, & \text{if } K < \frac{M}{m} - 1; \\ 1, & \text{otherwise.} \end{cases} \quad (6)$$

To sum up, transition probabilities associated with action s are as follows: $\forall K \in \{1, 2, 3, \ldots\}$,

$$\begin{aligned} &p(P|K,s) = \beta, \\ &p(J|K,s) = (1-\beta) f_J(K), \\ &p(K+1|K,s) = (1-\beta)(1-f_J(K)). \end{aligned} \quad (7)$$

If the secondary user stays in the same channel for too long, he/she will eventually be found by an attacker, as it can be seen from (6) and (7) that $p(K+1|K,s) = 0$ if $K > M/m - 1$. Therefore, we can limit the state S to a finite set $\{P, J, 1, 2, 3, \ldots, \bar{K}\}$, where $\bar{K} = \lfloor M/m - 1 \rfloor$ and the floor function $\lfloor x \rfloor$ returns the largest integer not greater than x.

An MDP consists of four important components, namely, a finite set of states, a finite set of actions, transition probabilities, and immediate payoffs. As we have already specified all of them, the defense problem is modeled by an MDP, and the optimal defense strategy can be obtained by solving the MDP.

For an MDP, a *policy* is defined as a mapping from a state to an action, i.e., $\pi : S_n \rightarrow a_n$. In other words, a policy π specifies an action $\pi(S)$ to take whenever the user is in state S. Among all possible policies, the optimal policy is the one that maximizes the expected total discounted payoffs. The value of a state S is defined as the highest expected payoff given

the MDP starts from state S, i.e.,

$$V^*(S) = \max_{\pi} E\left(\sum_{n=1}^{\infty} \delta^n U(n) \middle| S_1 = S\right), \tag{8}$$

where the optimal policy is the optimizer π^*. It is also the optimal defense strategy that the secondary user should adopt since it maximizes the expected payoff. For example, when the secondary user observes the state to be S, the action $\pi^*(S)$ should be taken in order to maximize the payoff.

An important but straightforward idea is that after a first move the remaining part of an optimal policy should still be optimal. Hence, the first move should maximize the sum of immediate payoff and expected payoff conditioned on the current action. This is the well-known Bellman equation [16],

$$\begin{aligned} Q(S,a) &= U(S,a) + \delta \sum_{S'} p(S'|S,a) V^*(S'), \\ V^*(S) &= \max_{a \in \{h,s\}} Q(S,a). \end{aligned} \tag{9}$$

Moreover, as seen from (4) and (5), the transition probabilities associated with action h are independent of the current state. Thanks to this special feature, the solution has a simple structure stated in Proposition 1.

Proposition 1: The optimal policy can be characterized by a single number $K^* \in \{0, 1, \ldots, \bar{K}\}$, i.e.,

$$a^* = \pi^*(S) = \begin{cases} s, & \text{if } S \le K^*; \\ h, & \text{otherwise.} \end{cases} \tag{10}$$

Proof: Using transition probabilities (4) (5) and the definition of $Q(S,a)$ in (9), it is easy to show that $Q(1,h) = Q(2,h) = \cdots = Q(\bar{K},h) \triangleq Q$, and $Q(J,h) = Q - R - L$, $Q(P,h) = Q - R$. Since h is the only action for states J and P, we have $V^*(J) = Q(J,h)$ and $V^*(P) = Q(P,h)$.

According to (7) and (9), $Q(\bar{K},s) - Q(\bar{K}-1,s) = \delta(1-\beta)(1-f_J(\bar{K}-1))(V^*(J) - V^*(\bar{K}))$. Notice that $V^*(\bar{K}) = \max(Q(\bar{K},h), Q(\bar{K},s)) \ge Q(\bar{K},h) = Q > V^*(J)$, and all the other factors are positive. Hence, $Q(\bar{K},s) < Q(\bar{K}-1,s)$ and $V^*(\bar{K}) = \max(Q(\bar{K},h), Q(\bar{K},s)) \le \max(Q(\bar{K}-1,h), Q(\bar{K}-1,s)) = V^*(\bar{K}-1)$.

Similarly, we can show $Q(\bar{K}-1,s) - Q(\bar{K}-2,s) = \delta(1-\beta)[(f_J(\bar{K}-1) - f_J(\bar{K}-2))(V^*(J) - V^*(\bar{K}-1)) + (1-f_J(\bar{K}-1))(V^*(\bar{K}) - V^*(\bar{K}-1))] < 0$, and $V^*(\bar{K}-1) \le V^*(\bar{K}-2)$ follows. The process can go all the way up to $K = 1$, leading to a conclusion that $Q(K,s)$ is a strictly decreasing function of $K \in \{1, 2, \ldots, \bar{K}\}$.

Notice that the optimal action at state K is s if $Q(K,s) \ge Q(K,h)$, and h if $Q(K,s) < Q(K,h)$. Since $Q(K,s)$ is decreasing and $Q(K,h)$ is a constant Q, there must exist a $K^* \in \{1, 2, \ldots, \bar{K}-1\}$ such that $Q(K^*,s) \ge Q > Q(K^*+1,s)$ except two extreme cases. One is $Q(\bar{K},s) \ge Q$ where $K^* = \bar{K}$, and the other is $Q(1,s) < Q$ where we can simply set $K^* = 0$ in (10). This concludes the proof. ■

Intuitively, since the probability of being jammed increases when the secondary user stays in the same channel for a longer time, K^* will be the critical state beyond which the damage overwhelms the hopping cost. If the secondary user stays in the same channel for a short period ($\le K^*$ time slots), he/she should stay to exploit more; otherwise, he/she should proactively hop to another channel since the risk of being jammed becomes significant.

TABLE I
VALUE ITERATION OF THE MDP.

Initialize $V(S)$ arbitrarily. Set a small ε as the stopping criterion.
For $n = 1, 2, 3, \ldots$
  For every state $S \in \{P, J, 1, 2, 3, \ldots, \bar{K}\}$
    $Q(S,a) = U(S,a) + \delta \sum_{S'} p(S'|S,a) V_n(S'),\ a \in \{s,h\}$
    $V_{n+1}(S) = \max\left(Q(S,h), Q(S,s)\right).$
  End For
  If $|V_{n+1}(S) - V_n(S)| < \varepsilon$ for all states
    The outer loop is terminated.
  End if
End For
After convergence,
– $V_n(S)$ is the value of state S;
– $K^* = \max\left\{S \in \{1, 2, 3, \ldots, \bar{K}\} : Q(S,s) \ge Q(S,h)\right\}$.

The value of K^* can be obtained using a standard procedure called *value iteration* [16], which updates the value of every state iteratively according to the Bellman equation, and this iteration is guaranteed to converge to the true value of states. The specific algorithm is summarized in Table-I.

C. The Next Round of the Arms Race

The attack-and-defense problem is like an arms race, and we have shown that the *MDP-based hopping* is the optimal strategy against attackers' sweeping attack. It naturally follows that attackers could further update their strategy to make the attack more efficient against MDP-based hopping secondary users. Taking advantage of the prior knowledge about the secondary user's hopping parameter K^*, attackers can enhance the sweeping attack by keeping only a list of detected channels in the most recent K^* time slots rather than all history. In the sweeping attack, a channel that has been detected by attackers will not be revisited until the next sweeping cycle; however, knowing the secondary user would stay in the same band for up to K^* time slots, attackers could launch a *smarter attack*, i.e., they randomly select m bands out of all the bands that have not been detected in the last K^* time slots, and detect these m bands.

The arms race between the secondary user and attackers could go on and on; however, it becomes too complicated to analyze. Fortunately, we will use simulation results to show that the MDP-based hopping already provides a good approximation to the game equilibrium. Therefore, we propose to use this MDP-based hopping as the defense strategy against jamming attackers for secondary users equipped with a single radio.

IV. THE LEARNING PROCESS

In the previous section, we have derived an MDP-based hopping which requires perfect knowledge such as the number of attackers m. However, in practice, the information is generally not directly available, since the secondary user cannot expect reliable information from adversaries. Both overestimating and underestimating the threat may result in

inappropriate degrees of protection. Therefore, in this section, we propose two learning schemes for the secondary user to learn from environment. The first one is based on maximum likelihood estimation (MLE), while the second one is adapted from Q-learning, a reinforcement learning method.

A. MLE-Based Learning

In this approach, the secondary user has to first go through a learning process to obtain estimates of the parameters, and then after the knowledge is gained, the secondary user updates the critical state K^* accordingly. During the learning period, the secondary user simply sets a value $\hat{K}^*$ as an initial guess of the optimal critical state K^*, and follows the strategy (10) with $\hat{K}^*$. This guess needs not to be accurate, as the goal is merely to observe transitions during the learning period that can be used for estimation of parameters.

With full history available including states and actions, the secondary user is able to count the occurrences of transitions given either action. For example, the notation $N^{(h)}_{S,S'}$ gives the total number of transitions from S to S' with action h taken, whereas $N^{(s)}_{S,S'}$ is the total number of transitions with action s taken. We define $K_L \triangleq \max\{K : N^{(s)}_{K,K+1} > 0\}$, $\mathbb{H} \triangleq \{P, J, K_L+1\}$, and $\mathbb{S} \triangleq \{1, 2, \ldots, K_L\}$. Given the sequence of transitions in history, the likelihood that such a sequence has occurred can be written as a product over all feasible transition tuples $(S, a, S') \in \{P, J, 1, 2, 3, \ldots, K_L+1\} \times \{s, h\} \times \{P, J, 1, 2, 3, \ldots, K_L+1\}$,

$$\Lambda = \prod_{(S,a,S'):\, p(S'|S,a)>0} \left(p(S'|S,a)\right)^{N^{(a)}_{S,S'}}. \tag{11}$$

Moreover, if we define $\rho \triangleq m/M$ and relax it to any real number, Proposition 2 gives the MLE of the parameters β, γ, and ρ. In the proof, we use the fact that the number of transitions into a state equals the number of transitions out of that state except the beginning and ending states,

$$\begin{aligned}&\sum_{S\in\mathbb{H}} N^{(h)}_{S,1} = N^{(s)}_{1,2} + N^{(s)}_{1,P} + N^{(s)}_{1,J},\\ &N^{(s)}_{K-1,K} = N^{(s)}_{K,K+1} + N^{(s)}_{K,P} + N^{(s)}_{K,J},\ \forall K \geq 2, K \in \mathbb{S}.\end{aligned} \tag{12}$$

If the beginning state and the ending state are not the same, there will be a difference of one transition in the above equations, but the impact could be negligible when the learning period is long enough.

Proposition 2: Given $N^{(h)}_{S,S'}$, $S \in \mathbb{H}$ and $N^{(s)}_{S,S'}$, $S \in \mathbb{S}$ counted from history of transitions, the MLE of primary users' parameters are

$$\beta_{\text{ML}} = \frac{\sum_{K\in\mathbb{S}} N^{(s)}_{K,P}}{\sum_{K\in\mathbb{S}} \left(N^{(s)}_{K,P} + N^{(s)}_{K,J} + N^{(s)}_{K,K+1}\right)}, \tag{13}$$

$$\gamma_{\text{ML}} = \frac{\sum_{S\in\mathbb{H}} N^{(h)}_{S,P}}{\sum_{S\in\mathbb{H}} \left(N^{(h)}_{S,P} + N^{(h)}_{S,J} + N^{(h)}_{S,1}\right)}, \tag{14}$$

and the MLE of attackers' parameters ρ_{ML} is the unique root within an interval $(0, 1/(K_L+1))$ of the following (K_L+1)-order polynomial of ρ,

$$\frac{1}{\rho}\left(\sum_{S\in\mathbb{H}} N^{(h)}_{S,J} + \sum_{K\in\mathbb{S}} N^{(s)}_{K,J}\right) = \sum_{K\in\mathbb{S}} \frac{N^{(s)}_{K,P}}{\frac{1}{K}-\rho} + \frac{N^{(s)}_{K_L,K_L+1}}{\frac{1}{K_L+1}-\rho}. \tag{15}$$

Proof: With transition probabilities specified in (4) – (7), the likelihood of the observed sequence of transitions (11) can be written as,

$$\begin{aligned}\Lambda = &\prod_{S\in\mathbb{H}} \gamma^{N^{(h)}_{S,P}} ((1-\gamma)\rho)^{N^{(h)}_{S,J}} ((1-\gamma)(1-\rho))^{N^{(h)}_{S,1}}\\ &\cdot \prod_{K\in\mathbb{S}} \beta^{N^{(s)}_{K,P}} \left(\frac{(1-\beta)\rho}{1-K\rho}\right)^{N^{(s)}_{K,J}} \left(\frac{(1-\beta)(1-K\rho-\rho)}{1-K\rho}\right)^{N^{(s)}_{K,K+1}}.\end{aligned} \tag{16}$$

Thanks to (12), it can be further decoupled and simplified into a product of three terms $\Lambda = \Lambda_\beta \Lambda_\gamma \Lambda_\rho$, where

$$\begin{aligned}\Lambda_\beta =&\beta^{\sum_{K\in\mathbb{S}} N^{(s)}_{K,P}} (1-\beta)^{\sum_{K\in\mathbb{S}} \left(N^{(s)}_{K,J} + N^{(s)}_{K,K+1}\right)},\\ \Lambda_\gamma =&\gamma^{\sum_{S\in\mathbb{H}} N^{(h)}_{S,P}} (1-\gamma)^{\sum_{S\in\mathbb{H}} \left(N^{(h)}_{S,J} + N^{(h)}_{S,1}\right)},\\ \Lambda_\rho =&\rho^{\sum_{S\in\mathbb{H}} N^{(h)}_{S,J} + \sum_{K\in\mathbb{S}} N^{(s)}_{K,J}} \cdot (1-(K_L+1)\rho)^{N^{(s)}_{K_L,K_L+1}}\\ &\cdot \prod_{K\in\mathbb{S}} (1-K\rho)^{N^{(s)}_{K,P}}.\end{aligned} \tag{17}$$

The MLE of β is derived from

$$\frac{\partial \ln \Lambda_\beta}{\partial \beta} = 0, \tag{18}$$

which yields (13). Similarly, we get equations (14) and (15), by differentiating $\ln \Lambda_\gamma$ and $\ln \Lambda_\rho$, and equating them to 0.

To ensure that the likelihood is positive, ρ has to lie in the interval $(0, 1/(K_L+1))$. Within this interval, the left-hand side of equation (15) decreases monotonically and approaches positive infinity as ρ goes to 0, whereas the right-hand side increases monotonically and approaches positive infinity as ρ goes to $1/(K_L+1)$. Therefore, there must be a unique value of $\rho \in (0, 1/(K_L+1))$ which is both the root of the equation and the MLE ρ_{ML}. ■

After the learning period, the secondary user rounds $M \cdot \rho_{\text{ML}}$ to the nearest integer as an estimate of m, and calculate the optimal strategy using the MDP approach described in the previous section.

B. Q-Learning

The previous approach is to estimate the parameters based on the Markov model; however, the model is not always accurate. For example, it is possible that not all the attackers are able to coordinate and their targeted channels may overlap. Therefore, the alternative approach is to learn the optimal policy without explicitly knowing the model. This is known as Q-learning [17] in the reinforcement learning literature.

The intuition behind Q-learning is to approximate the unknown transition probability in (9) by the empirical distribution of states that have been reached as the game unfolds.

Specifically, (9) is replaced by an iterative process

$$Q_n(S,a) = (1-\mu_n)Q_{n-1}(S,a) + \mu_n(U(S,a)+\delta V_n(S')),$$
$$V_{n+1}(S) = \max_{a\in\{h,s\}} Q_n(S,a), \tag{19}$$

where the Q-value of a state-action pair (S,a) is updated based on the observed new state S', the frequency of which represents the empirical distribution of the transition from state S with action a. μ_n is the learning rate decreasing in time, and we set

$$\mu_n = \frac{1}{1+\text{number of updates for } Q(S,a)}. \tag{20}$$

Since the learning rate takes values $1, \frac{1}{2}, \frac{1}{3}, \ldots$ for each update, according to Proposition 3 (Theorem 7.4.2 in [17]), the convergence is guaranteed.

Proposition 3: Q-learning converges to the optimal policy with probability 1, provided that each state-action pair is encountered infinitively, and the learning rate obeys $0 \le \mu_n < 1$, $\sum_{n=1}^{+\infty} \mu_n = \infty$, and $\sum_{n=1}^{+\infty} \mu_n^2 < \infty$.

In general, at time n, only the Q-value of the current state-action pair (S_n, a_n) is updated, whereas the Q-values of all the other pairs remain unchanged. However, the transition probabilities associated with action h are independent of the current state in this anti-jamming problem. Once a new state S' is reached after action h, Q-values of all state-action pairs (S,h) can be updated, since they share the same underlying transition probability. By doing this, the observations are fully utilized, and the convergence is made faster.

Another issue is choosing an action for a given state. One may choose $a_n = \pi(S_n)$, but the problem is $\pi(S)$ during learning may not be the true optimal policy, and always following $a_n = \pi(S_n)$ may enhance the false impression and prevent the truth from being discovered. Thus, the secondary user should deviate from $\pi(S_n)$ with a small probability η to exploit the state-action pairs that have been rarely visited. Finally, the Q-learning method for this anti-jamming game is summarized in Table-II. It usually takes longer for the Q-learning process to converge than the MLE-based learning scheme, but the Q-learning method has the advantage that it does not require any explicit modeling of the underlying Markov chains.

TABLE II
Q-LEARNING IN THE PROPOSED ANTI-JAMMING GAME.

Initialize $Q_0(S,a)$ and $V_0(S)$ to be 0, $\pi(K) = s, K \ge 1$.
Fix $\pi(S) = h$, for $S = J$ and P.
For $n = 1, 2, 3, \ldots$
  `%% Play the game`
  Observe state S_n.
  If $S_n \neq J, P$, with a small probability η,
  take an action other than $\pi(S_n)$;
  take action $\pi(S_n)$ in all the other cases.
  Denote the action actually taken as a_n.
  The state will transit to a new state S_{n+1}.
  `%% Update the Q-functions`
  Determine the learning rate μ_n according to (20).
  If $a_n = s$
    $Q_n(S_n, s) = (1-\mu_n)Q_{n-1}(S_n, s) + \mu_n(U(S_n, s)+\delta V(S_{n+1}))$.
    $Q_n(S,a) = Q_{n-1}(S,a)$, for all the other (S,a) pairs.
  End If
  If $a_n = h$
    $Q_n(S,h) = (1-\mu_n)Q_{n-1}(S,h) + \mu_n(U(S,h) + \delta V(S_{n+1}))$,
    and $Q_n(S,s) = Q_{n-1}(S,s)$, for all states S.
  End If
  `%% Update the policy`
  For every state $K \ge 1$
    $V(K) = \max\left(Q_n(K,h), Q_n(K,s)\right)$,
    $\pi(K) = \text{argmax}_{a\in\{s,h\}} Q_n(K,a)$.
  End For
  $V(S) = Q_n(S,h)$ and $\pi(S) = h$, for $S = J$ and P.
End For

V. ANTI-JAMMING POWER ALLOCATION GAME

It is not always the case that secondary users have the single-radio constraint. In some scenarios, secondary users do have the capability of accessing multiple bands at the same time, and we need to consider the anti-jamming defense under this multi-radio assumption. In this section, we extend the anti-jamming game to the scenario where a secondary user is equipped with multiple radios and is able to access all the available channels simultaneously with a limited power budget. Each attacker is also assumed to be able to inject interference to all channels, and thus all attackers can be viewed as a single super attacker whose power budget is the sum of individual budgets.

A. Game Reformulation

In this case, the defense strategy is not to hop between channels, but to randomly allocate power in different channels. Whether the attackers can successfully jam communications in one particular channel will depend on how much power the secondary user and attackers allocate on that channel. Therefore, we have to redefine the game to reflect the changes.

The secondary user still adopts the "listen-before-talk" rule, that is, sensing for spectrum opportunities at the beginning of a time slot. Recall that transmitters have power constraints. On finding M_0 available channels out of the M total channels, the secondary user allocates power p_k to the kth available channel such that $\sum_{k=1}^{M_0} p_k = p^B$. At the same time, the attacker injects i_k to the kth available channel such that $\sum_{k=1}^{M_0} i_k = i^B$. The power allocation vectors $\mathbf{p} = (p_1, p_2, \ldots, p_{M_0})$ and $\mathbf{i} = (i_1, i_2, \ldots, i_{M_0})$ are actions. If the received SINR exceeds the minimum requirement τ, i.e.,

$$\frac{p_k}{i_k + \sigma_k^2} \ge \tau, \tag{21}$$

packets can be transmitted successfully on that channel. σ_k^2 is the noise variance of channel k, which we assume is the same for all channels, i.e., $\sigma_k^2 = \sigma^2$. Because each successful transmission yields a communication gain R, the secondary user's payoff is defined as the number of successful

transmissions, i.e.,

$$U(\mathbf{p},\mathbf{i})=\sum_{k=1}^{M_0}\mathbf{1}\left(\frac{p_k}{i_k+\sigma^2}\geq\tau\right), \tag{22}$$

where $\mathbf{1}(\cdot)$ is the indicator function, and the attackers' payoff is the opposite. In order to hide the allocation strategy from attackers, the secondary user has to randomize the power allocation, and the strategy is characterized by a probability distribution function $F(\mathbf{p})$. Similarly, attackers will employ a random strategy characterized by $H(\mathbf{i})$. The expected payoff is to average (22) over the distribution of $F(\mathbf{p})$ and $H(\mathbf{i})$, i.e., $\bar{U}(F(\mathbf{p}),H(\mathbf{i}))=\iint U(\mathbf{p},\mathbf{i})\,\mathrm{d}F(\mathbf{p})\,\mathrm{d}H(\mathbf{i})$.

Different from the single-radio case, we do not need to consider the arms race in this multi-radio case. Assuming perfect knowledge, we are able to derive the Nash equilibrium of this game, which is the best response given the other player sticks to the equilibrium strategy. Furthermore, since it is a zero-sum game, the Nash equilibrium $(F^*(\mathbf{p}),H^*(\mathbf{i}))$ also provides the minimax strategy [29] such that $F^*(\mathbf{p})$ is a maximizer to $\min_H \bar{U}(F(\mathbf{p}),H(\mathbf{i}))$. This property is of great interest. If capable of learning the secondary user's strategy $F(\mathbf{p})$, attackers can always come up with a strategy $H(\mathbf{i})$ tailored to $F(\mathbf{p})$, which minimizes the secondary user's expected payoff and maximizes the damage. Therefore, the secondary user should choose the strategy $F^*(\mathbf{p})$ to maximize the worst-case expected payoff.

To simplify the game, we define $j_k=\tau(i_k+\sigma^2)$ with the constraint $\sum_{k=1}^{M_0}j_k=\tau(i^B+M_0\sigma^2)\triangleq j^B$. Then, the condition of a successful transmission becomes $p_k\geq j_k$. This game falls into the category of Colonel Blotto games where two opponents distribute limited resources over a number of battlefields with a payoff equal to the sum of outcomes from individual battlefields [18]. However, the difference is that j_k has to be lower bounded by $\tau\sigma^2$, since attackers only have control over the i_k part. In this new game, the attackers' strategy is also given by a joint distribution function, denoted by $G(\mathbf{j})$.

B. Nash Equilibrium

We first derive the necessary condition of the Nash equilibrium (NE) in terms of marginal distribution functions $F_1(p_1),F_2(p_2),\ldots,F_{M_0}(p_{M_0})$, $G_1(j_1),G_2(j_2),\ldots,G_{M_0}(j_{M_0})$.

Notice that the probability of a successful transmission is $Pr(p_k\geq j_k)=G_k(p_k)$, and the payoff of the secondary user is $\sum_{k=1}^{M_0}G_k(p_k)$ when he/she fixes the power allocation as $(p_1,p_2,\ldots,p_{M_0})$. When the player employs a randomized strategy, the expected payoff becomes

$$\sum_{k=1}^{M_0}\int_0^\infty G_k(p_k)\mathrm{d}F_k(p_k), \tag{23}$$

and the necessary condition of the total power constraint becomes

$$p^B=E\left(\sum_{k=1}^{M_0}p_k\right)=\sum_{k=1}^{M_0}\int_0^\infty p_k\mathrm{d}F_k(p_k). \tag{24}$$

If we introduce a Lagrangian multiplier λ_P, the optimization problem of the secondary user can be formulated as

$$\max_{\{F_k(p_k)\}}\sum_{k=1}^{M_0}\int_0^\infty\left(G_k(p_k)-\lambda_P p_k\right)\mathrm{d}F_k(p_k)+\lambda_P p^B. \tag{25}$$

Similarly, we can derive the optimization problem for attackers who attempt to maximize $\sum_{k=1}^{M_0}\mathbf{1}(p_k<j_k)$. As shown later in Proposition 4, at the equilibrium the amount of power allocated in the kth channel p_k is a random variable with a discrete part at 0 and a continuous part elsewhere. Hence, the event $p_k=j_k$ happens with probability 0, and $Pr(p_k<j_k)=Pr(p_k\leq j_k)=F_k(j_k)$. Therefore, from the attackers' point of view, the optimization problem is

$$\max_{\{G_k(j_k)\}}\sum_{k=1}^{M_0}\int_{\tau\sigma^2}^\infty\left(F_k(j_k)-\lambda_J j_k\right)\mathrm{d}G_k(j_k)+\lambda_J j^B, \tag{26}$$

where λ_J is the Lagrangian multiplier for attackers.

For the secondary user, he/she can either decide not to access channel k (i.e., $p_k=0$) or decide to access that channel with some power lower bounded by $\underline{p}_k$ and upper bounded by $\bar{p}_k$ (i.e., $p_k\in[\underline{p}_k,\bar{p}_k]$). Apparently, $\underline{p}_k\geq\tau\sigma^2$, because if p_k is chosen in the open interval $(0,\tau\sigma^2)$, the secondary user will always fail in that channel, and it is better not to allocate power at all. When the equilibrium strategy is a mixed strategy over the domain $0\cup[\underline{p}_k,\bar{p}_k]$, according to game theory, the player must be indifferent among these values [29], namely, $G_k(p_k)-\lambda_P p_k=constant$ for $p\in 0\cup[\underline{p}_k,\bar{p}_k]$. In particular, since $G_k(0)=0$, we can further have

$$G_k(p_k)-\lambda_P p_k=0,\ \text{for } p_k\in 0\cup[\underline{p}_k,\bar{p}_k]. \tag{27}$$

The similar argument can be applied to attackers who allocate power $j_k\in[\underline{j}_k,\bar{j}_k]$ and has to be indifferent among the values, namely,

$$F_k(j_k)-\lambda_J j_k=constant,\ \text{for } j_k\in[\underline{j}_k,\bar{j}_k]. \tag{28}$$

Proposition 4: For the NE strategy, bounds are determined as $\bar{p}_k=\bar{j}_k=\min(1/\lambda_P,1/\lambda_J)$, and $\underline{p}_k=\underline{j}_k=\tau\sigma^2$. Moreover, $Pr(j_k=\tau\sigma^2)=\lambda_P\tau\sigma^2$, and $Pr(p_k=\tau\sigma^2)=0$; the probability distribution function $F_k(p_k)$ is continuous in the range $(\tau\sigma^2,\bar{p}_k]$, and so is $G_k(j_k)$.

Proof: According to the definition of the NE, no single player can be better off by deviating unilaterally from the NE strategy. In what follows, we give a proof mainly by contradiction.

From optimization problems (25) and (26), it is clear that $p_k\leq 1/\lambda_P$ and $j_k\leq 1/\lambda_J$ have to be satisfied to avoid negative payoffs. $\bar{p}_k=\bar{j}_k$ can be proved by contradiction. If $\bar{p}_k\neq\bar{j}_k$, say $\bar{p}_k<\bar{j}_k$, attackers are better off by moving $\bar{j}_k$ to $(\bar{p}_k+\bar{j}_k)/2$, as $F_k(\bar{j}_k)-\lambda_J\bar{j}_k=1-\lambda_J\bar{j}_k<1-\lambda_J(\bar{p}_k+\bar{j}_k)/2=F_k\left((\bar{p}_k+\bar{j}_k)/2\right)-\lambda_J(\bar{p}_k+\bar{j}_k)/2$. The analysis is similar for the case $\bar{p}_k>\bar{j}_k$.

Next, we prove $\underline{p}_k=\underline{j}_k$ by contradiction. If $\underline{p}_k\neq\underline{j}_k$, say $\underline{p}_k<\underline{j}_k$, the secondary user is better off by moving $(\underline{p}_k+\underline{j}_k)/2$ to $\underline{p}_k$, since power can be saved without affecting the winning probability. The analysis is similar for the case $\underline{p}_k>\underline{j}_k$. According to (27), $Pr(j_k=\underline{j}_k)=G(\underline{j}_k)=\lambda_P\underline{j}_k$. Because $p_k\geq\underline{j}_k$ always holds for $p_k\in[\underline{p}_k,\bar{p}_k]$,

by contradiction, if $\underline{j}_k > \tau\sigma^2$, attackers will be better off by moving $\underline{j}_k$ to $\tau\sigma^2$. Therefore, $\underline{p}_k = \underline{j}_k = \tau\sigma^2$, and $Pr(j_k = \tau\sigma^2) = \lambda_P \tau\sigma^2$.

Then, if $Pr(p_k = \tau\sigma^2) > 0$, attackers can change the probability mass from $\tau\sigma^2$ to $\tau\sigma^2 + \epsilon$ where ϵ is an arbitrary small number, and can increase the jamming probability by $\lambda_P \tau\sigma^2 \cdot Pr(p_k = \tau\sigma^2)$ with only negligible power increase. This cannot be an NE, and as a result, $Pr(p_k = \tau\sigma^2) = 0$.

Finally, we show that $F_k(p_k)$ cannot have discontinuous points in the interval $(\tau\sigma^2, \bar{p}_k]$. By contradiction, assume there is at least one discontinuous point, denoted by p^o, and thus $Pr(p_k = p^o) > 0$. Then, attackers can move the neighborhood $(p^o - \epsilon, p^o)$ to $(p^o, p^o + \epsilon)$ to increase the jamming probability by $Pr(p_k = p^o) \cdot Pr(j_k \in (p^o - \epsilon, p^o))$ with only negligible power increase when ϵ is an arbitrary small number. Similar arguments can be made to prove $G_k(j_k)$ cannot have discontinuous points in the interval $(\tau\sigma^2, \bar{j}_k]$ either. This concludes the proof. ∎

Based on Proposition 4 and necessary conditions (27)(28), in Proposition 5, we derive the marginal distribution of the NE under the condition $p^B \leq \tau i^B$.

Proposition 5: Under the condition $p^B \leq \tau i^B$, there exists a unique Nash equilibrium whose marginal distributions for the secondary user and attackers are given by

$$F_k^*(p_k) = \begin{cases} 0, & p_k < 0, \\ 1 - \lambda_J/\lambda_P + \lambda_J \tau\sigma^2, & p_k \in [0, \tau\sigma^2), \\ 1 - \lambda_J/\lambda_P + \lambda_J p_k, & p_k \in [\tau\sigma^2, 1/\lambda_P], \end{cases} \tag{29}$$

and

$$H_k^*(i_k) = \begin{cases} 0, & i_k < 0, \\ \lambda_P \tau(\sigma^2 + i_k), & i_k \in [0, 1/(\tau\lambda_P) - \sigma^2], \end{cases} \tag{30}$$

where $\lambda_J = M_0 p^B / ((j^B)^2 - \tau^2 {M_0}^2 \sigma^4 + j^B \sqrt{(j^B)^2 - \tau^2 {M_0}^2 \sigma^4})$ and $\lambda_P = M_0/(j^B + \sqrt{(j^B)^2 - \tau^2 {M_0}^2 \sigma^4})$.

Proof: Define $\bar{p}_k = \bar{j}_k = \min(1/\lambda_P, 1/\lambda_J) \triangleq \bar{p}$ which is independent on k. According to Proposition 4, $F_k(p_k)$ is continuous in the interval $[\tau\sigma^2, \bar{p}]$, and therefore, we can take the derivative of (28)

$$\mathrm{d}F_k(x) = \lambda_J \mathrm{d}x, \; x \in [\tau\sigma^2, \bar{p}], \tag{31}$$

and substitute it to the power constraint (24),

$$p^B = \sum_{k=1}^{M_0} \int_0^{\bar{p}} p_k \mathrm{d}F_k(p_k) = M_0 \int_{\tau\sigma^2}^{\bar{p}} \lambda_J p_k \mathrm{d}p_k = \frac{M_0}{2} \lambda_J (\bar{p}^2 - \tau^2\sigma^4). \tag{32}$$

Similar derivation can be applied to attackers' power constraint except that $G_k(j_k)$ is discontinuous at $j_k = \tau\sigma^2$,

$$j^B = M_0 \left(\tau\sigma^2(\lambda_P \tau\sigma^2) + \frac{1}{2} \lambda_P (\bar{p}^2 - \tau^2\sigma^4) \right). \tag{33}$$

If $1/\lambda_P \leq 1/\lambda_J$, then $\bar{p} = 1/\lambda_P$ and (33) becomes a quadratic equation of the variable $1/\lambda_P$, two roots of which are given by

$$\left(\frac{1}{\lambda_P}\right)_{1,2} = \frac{1}{M_0} \left(j^B \pm \sqrt{(j^B)^2 - \tau^2 {M_0}^2 \sigma^4} \right). \tag{34}$$

However, only the root with the plus sign is valid since the other root is smaller than $\tau\sigma^2$. Then, $1/\lambda_J$ can be solved from (32) accordingly,

$$\frac{1}{\lambda_J} = \frac{(j^B)^2 - \tau^2 {M_0}^2 \sigma^4 + j^B \sqrt{(j^B)^2 - \tau^2 {M_0}^2 \sigma^4}}{M_0 p^B}. \tag{35}$$

When the condition $p^B \leq \tau i^B$ holds, it is easy to verify that $1/\lambda_P \leq 1/\lambda_J$.

The pair of Lagrangian multipliers have been uniquely determined by (34) and (35). Since at least one mixed-strategy NE exists in a game [29], we can safely draw a conclusion that this characterizes the unique NE in the anti-jamming game. With parameters known, it is straightforward to write down the marginal distribution. For instance, according to (27),

$$G_k(j_k) = \begin{cases} 0, & j_k < \tau\sigma^2, \\ \lambda_P j_k, & j_k \in [\tau\sigma^2, 1/\lambda_P], \end{cases} \tag{36}$$

which can be further mapped back to the original domain $H_k(i_k)$ (30) using $j_k = \tau(i_k + \sigma^2)$. Similarly, marginal distribution $F_k(p_k)$ given by (29) can be derived from (31). ∎

So far, we have known the existence of the NE and the formula of marginal distribution functions; however, it still remains a question to find the specific NE strategy determined by the joint probability distribution function. We have followed the procedure in [18] to construct one kind of joint distribution that matches desired marginal distribution and meets the total power restriction. With this procedure, we can finally characterize the NE strategy for the anti-jamming game.

Similar to the single-radio case, the secondary user still needs to learn about the information of the opponent; specifically, the power budget of the opponent should be known before the optimal power allocation can be carried out. It is fairly easy in the multi-radio case, because the secondary user can simply shut down his/her own transmission in some time slots, and estimate the jamming power in each band. The sum would be an estimate of the attacker's total power budget.

VI. Simulation Results

In this section, we present some simulation results to evaluate the proposed defense strategies against jamming attacks. We first consider the scenario with the single-radio secondary user, whose defense strategy is proactive hopping among multiple channels. In the simulation, we fix a set of parameters to gain some insight of the defense strategy. The parameters are as follows: the communication gain $R = 5$, the hopping cost $C = 1$, the total number of channels $M = 60$, the discount factor $\delta = 0.95$, and the primary users' access pattern $\beta = 0.01, \gamma = 0.1$.

We show the critical state K^* obtained from the value iteration of the MDP, when we change the value of damage L and the number of attackers m. We assume that the secondary user has perfect knowledge of the environment. As shown in Fig. 3, if the damage from each jamming L is fixed, say $L = 10$ for example, the critical state K^* decreases from 11 to 3 when the number of attackers m increases from 2 to 6. Similarly, when the number of attackers m is fixed, the critical state K^* also decreases as the value of L increases.

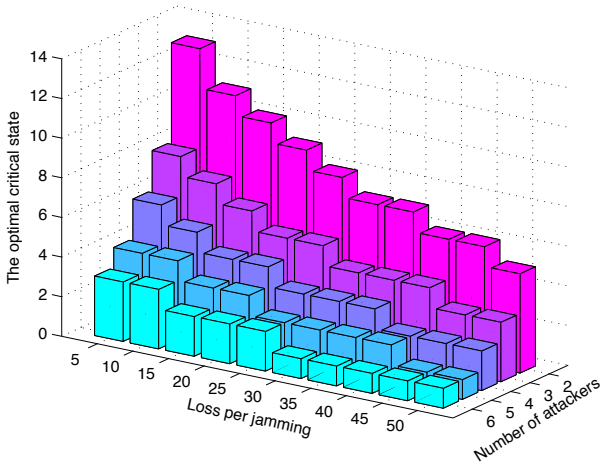


Fig. 3. The critical state K^* with different attack strengths and damages.

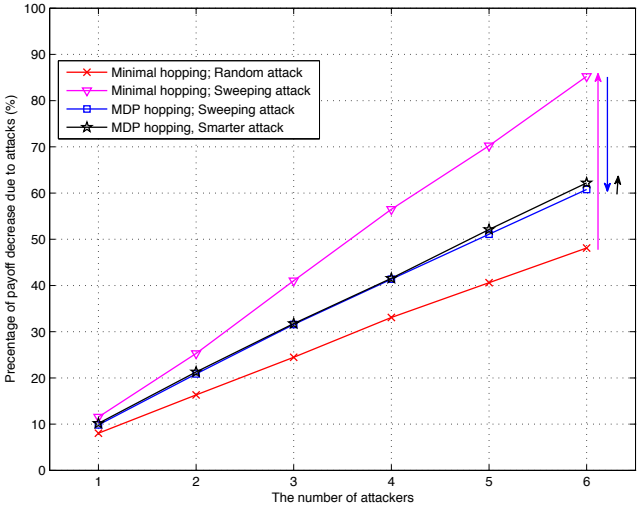


Fig. 4. The percentage of payoff decrease due to jamming attacks during the first few iterations of the arms race.

The reason is that the secondary user should proactively hop more frequently (i.e., K^* is smaller) to avoid being jammed when the threat from attackers are stronger (more attackers and/or more severe damage if jammed).

In Fig. 4, we illustrate the arms race between the secondary user and attackers by presenting the payoffs with different combinations of the attack and the defense strategy. We plot the percentage of payoff loss compared with a network without malicious attackers, and the damage L is set to 20 in this simulation. As discussed in Section III.A, the iteration starts with the random attack strategy for attackers and the minimal hopping for the secondary user. Then, as shown in the figure, attackers can significantly intensify the jamming damage by launching the sweeping attack when the secondary user sticks to the minimal hopping strategy. To counterbalance the impact, the secondary user could update the strategy to the MDP-based hopping, and the payoff loss would be reduced. Recall that in Section III.C, attackers could further adopt the smarter attack to increase the damage, but the simulation result shows very little difference. Since jammers' updated attack strategy would move the curve upwards while the secondary user' updated defense strategy would move the curve downwards, we expect that payoff curves with consequent iterations will move down and up alternatively, but they will stay in the region between the curves of the last two iterations shown in the figure. Therefore, the MDP-based hopping provides a very close approximation to the game equilibrium.

We also compare our algorithm with other existing algorithms; for example, Fig. 5 presents the comparison with the dogfight game equilibrium in [24]. We simulate the case with two and four attackers, and vary the channel hopping cost C. We still use the percentage of payoff loss as the performance metric. Because the hopping cost is not taken into consideration in the dogfight game model, our algorithm outperforms the dogfight equilibrium especially when the hopping cost is not negligible. In practice, the hopping cost may come from the throughput loss due to the radio tuning time. When it takes longer for the RF hardware to settle down after tuning, the hopping cost will be larger, and our

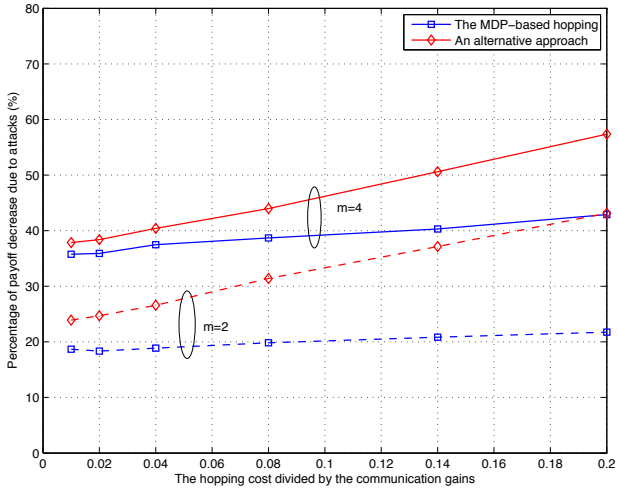


Fig. 5. Performance comparison between the proposed MDP-based hopping and the dogfight game equilibrium.

algorithm has more advantage since it maintains a balance between hopping cost and jamming loss.

We evaluate the MLE learning algorithm by showing the variance of estimation errors $M\rho_{\text{ML}}-m$ from 100 independent simulation runs with certain lengths of learning period. The learning curves are plotted in the upper figure of Fig. 6. As the learning period lengthens, the variance decreases which means a more accurate estimate. The accuracy degrades slightly when there are more attackers in the network. Recall that the last step of learning is rounding $M\rho_{\text{ML}}$ to the nearest integer, which could further reduce the estimation errors. In the lower figure of Fig. 6, we show the percentage of trials that the estimated number of attackers is exactly the true value. From the figure, we can see the percentage of exact estimation grows fast and approaches to one hundred percent with increasing learning periods.

The anti-jamming game with the multi-radio secondary user who employs randomized power allocation strategy is also

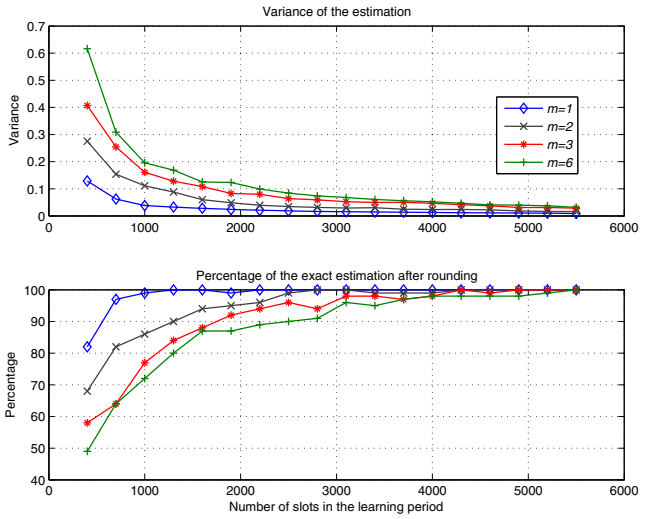


Fig. 6. Learning curves of the MLE learning process.

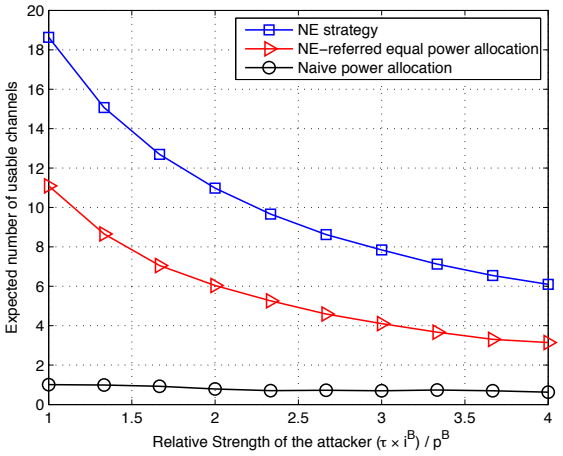


Fig. 7. The average number of channels that meet the SINR requirement when different strategies are adopted by the secondary user.

presented. In order to show that for the secondary user, the NE strategy is a minimax strategy such that the worst possible damage is minimized, we have run simulations with two other possible strategies considered: one decides the number of channels to access according to the NE strategy but allocates power equally, and the other allocates power based on a naive assumption that the jammer would inject equal interference to each channel. They are referred to as "NE-referred equal power allocation" and "naive power allocation", respectively. Fig. 7 provides the average number of channels that meet the SINR requirement when the secondary user adopts these strategies. When attackers are more powerful with a higher interference budget i^B, fewer usable channels can be expected for all three strategies. However, it is clear that the NE strategy performs much better than the other two strategies, and the secondary user has to choose it as the optimal power allocation strategy against malicious jamming attacks.

VII. Conclusions

In this paper, we have investigated the anti-jamming defense in a cognitive radio network with multiple available channels, by modeling the interaction between a secondary user and attackers as anti-jamming games and studying the optimal strategy and the equilibrium of the games. In the scenario where both the secondary user and attackers are equipped with a single radio and access only one channel at any time, the secondary user hop proactively between channels as the defense strategy. We show that the MDP-based hopping is a good approximation to the game equilibrium. Moreover, in order to gain knowledge about the adversaries, learning schemes are proposed for the secondary user based on maximum likelihood estimation and Q-learning. Extending the anti-jamming problem to the scenario where the multi-radio secondary user can access multiple channels simultaneously, we redefine the game with randomized power allocation as the defense strategy. The defense strategy obtained from the Nash equilibrium is optimal in the sense that it minimizes the worst-case damage caused by attackers.

References

[1] J. Mitola III, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2000.

[2] K. J. R. Liu and B. Wang, Cognitive Radio Networking and Security: A Game Theoretical View, Cambridge University Press, 2010.

[3] B. Wang, Y. Wu, and K. J. R. Liu, "Game theory for cognitive radio networks: An overview," *Computer Networks*, vol. 54, no. 14, pp.2537–2561, Oct. 2010.

[4] Z. Han and K. J. R. Liu, *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications*, Cambridge Univ Press, 2008.

[5] J. Neel, J. Reed, and R. Gilles, "The role of game theory in the analysis of software radio networks," in *SDR Forum Technical Conference*, San Diego, Nov. 2002.

[6] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 517–528, Apr. 2007.

[7] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "Repeated open spectrum sharing game with cheat-proof strategies," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1922–1933, Apr. 2009.

[8] Z. Ji and K. J. R. Liu, "Multi-stage pricing game for collusion-resistant dynamic spectrum allocation," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 182–191, Jan. 2008.

[9] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "A scalable collusion-resistant multi-winner cognitive spectrum auction game," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3805–3816, Dec. 2009.

[10] S. Gao, L. Qian, D. R. Vaman, and Z. Han, "Distributed cognitive sensing for time varying channels: Exploration and exploitation," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Sydney, Australia, Apr. 2010.

[11] R. Chen, J.-M. Park, and J. H. Reed, "Defense against primary user emulation attacks in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 25–37, Jan. 2008.

[12] T. C. Clancy and N. Goergen, "Security in cognitive radio networks: threats and mitigation," in *Proc. International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, Singapore, May 2008.

[13] T. X. Brown and A. Sethi, "Potential cognitive radio denial-of-service vulnerabilities and protection countermeasures: A multi-dimensional analysis and assessment," *Mobile Networks and Applications*, vol. 13, no. 5, pp. 516–532, Oct. 2008.

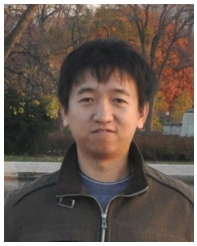
[14] W. Wang, H. Li, Y. Sun, and Z. Han, "CatchIt: Detect malicious nodes in collaborative spectrum sensing," in *Proc. IEEE Global Communications Conference (Globecom)*, Hawaii, Dec. 2009.

[15] Y. Wu and K. J. R. Liu, "An information secrecy game in cognitive radio networks", *IEEE Trans. Inf. Forens. Security*, vol. 6, no. 3, pp. 831–842, Sept. 2011.

[16] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, 1994.

[17] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*, Cambridge University Press, 2008.

[18] B. Roberson, "The Colonel Blotto game," *Economic Theory*, vol. 29, no. 1, pp. 1–24, Sept. 2006.
[19] E. Altman, K. Avrachenkov, and A. Garnaev, "A jamming game in wireless networks with transmission cost," in *Proc. of NET-COOP 2007. Lecture Notes in Computer Science*, vol. 4465, pp. 1–12, 2007.
[20] S. Khattab, D. Mosse, and R. Melhem, "Jamming mitigation in multi-radio wireless networks: Reactive or proactive?" in *Proc. of International Conference on Security and Privacy in Communication Netowrks (SecureComm 2008)*, Istanbul, Turkey, Sept. 2008.
[21] M. Strasser, S. Capkun, C. Popper, and M. Cagalj, "Jamming-resistant key establishment using uncoordinated frequency hopping," in *Proc. IEEE Symposium on Security and Privacy*, Oakland, CA, May 2008.
[22] D. Slater, P. Tague, R. Poovendran, and B. J. Matt, "A coding-theoretic approach for efficient message verification over insecure channels," in *Proc. ACM conference on Wireless network security (WiSec)*, Zurich, Switzerland, Mar. 2009.
[23] Q. Wang, P. Xu, K. Ren,and X.-Y. Li, "Delay-bounded adaptive UFH-based anti-jamming wireless communication", in *Proc. IEEE International Conference on Computer Communications (Infocom)*, Shanghai, China, Apr. 2011.
[24] H. Li and Z. Han, "Dogfight in spectrum: Combating primary user emulation attacks in cognitive radio systems, Part I: Known channel statistics," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3566–3577, Nov. 2010.
[25] H. Li and Z. Han, "Dogfight in spectrum: Combating primary user emulation attacks in cognitive radio systems, Part II: Unknown channel statistics," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 274–283, Jan. 2011.
[26] A. Sampath, H. Dai, H. Zheng, and B. Y. Zhao, "Multi-channel jamming attacks using cognitive radios," in *Proc. International Conference on Computer Communications and Networks (ICCCN)*, pp. 352–357, Hawai'i, Aug. 2007.
[27] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.
[28] M. Takai, J. Martin, and R, Bagrodia, "Effects of wireless physical layer modeling in mobile ad hoc networks," in *Proc. ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc)*, Long Beach, CA, Oct. 2001.
[29] D. Fudenburg and J. Tirole, *Game Theory*, MIT Press, 1993.



Yongle Wu received the Ph.D. degree in Electrical and Computer Engineering from University of Maryland, College Park in 2010. He received the B.S. (with highest honor) and M.S. degrees in Electronic Engineering from Tsinghua University, Beijing, China, in 2003 and 2006, respectively.

Dr. Wu is currently a senior engineer with Qualcomm Incorporated, San Diego, CA. His research interests are in the areas of wireless communications and networks, including cognitive radio techniques, dynamic spectrum access, network security, and MIMO-OFDM communication systems.

Dr. Wu received the Graduate School Fellowship from the University of Maryland in 2006, the Future Faculty Fellowship in 2009 and the Litton Industries Fellowship in 2010, both from A. James Clark School of Engineering, University of Maryland, and the Distinguished Dissertation Fellowship from Department of Electrical and Computer Engineering, University of Maryland in 2011.

Beibei Wang (S'07-M'11) received the B.S. degree in electrical engineering (with the highest honor) from the University of Science and Technology of China, Hefei, in 2004, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park in 2009. From 2009 to 2010, she was a research associate at the University of Maryland. Currently, she is a senior engineer with Corporate Research and Development, Qualcomm Incorporated, San Diego, CA.

Her research interests include wireless communications and networking, including cognitive radios, dynamic spectrum allocation and management, network security, and multimedia communications. Dr. Wang was the recipient of the Graduate School Fellowship, the Future Faculty Fellowship, and the Deans Doctoral Research Award from the University of Maryland, College Park. She is a coauthor of *Cognitive Radio Networking and Security: A Game-Theoretic View*, Cambridge University Press, 2010.

K. J. Ray Liu (F'03) is named a Distinguished Scholar-Teacher of University of Maryland, College Park, in 2007, where he is Christine Kim Eminent Professor of Information Technology. He serves as Associate Chair of Graduate Studies and Research of Electrical and Computer Engineering Department and leads the Maryland Signals and Information Group conducting research encompassing broad aspects of wireless communications and networking, information forensics and security, multimedia signal processing, and biomedical engineering.

Dr. Liu is the recipient of numerous honors and awards including IEEE Signal Processing Society Technical Achievement Award and Distinguished Lecturer. He also received various teaching and research recognitions from University of Maryland including university-level Invention of the Year Award; and Poole and Kent Senior Faculty Teaching Award and Outstanding Faculty Research Award, both from A. James Clark School of Engineering. An ISI Highly Cited Author in Computer Science, Dr. Liu is a Fellow of IEEE and AAAS.

Dr. Liu is President-Elect and was Vice President - Publications of IEEE Signal Processing Society. He was the Editor-in-Chief of IEEE Signal Processing Magazine and the founding Editor-in-Chief of EURASIP Journal on Advances in Signal Processing.

T. Charles Clancy (S'02-M'06-SM'10) is Associate Professor in the Bradley Department of Electrical and Computer Engineering at Virginia Tech where he is Director of the Ted and Karyn Hume Center for National Security and Technology. Prior to joining Virginia Tech, Dr. Clancy was a senior advisor to the US military in Baghdad, Iraq, where he led successful efforts to establish Baghdad's first commercial international fiber-optic Internet connectivity. Prior to Iraq, Dr. Clancy was a senior scientist with the Laboratory for Telecommunications Sciences, a federal research lab at the University of Maryland, where he led programs in RF and signal processing research. He received his MS in Electrical Engineering from the University of Illinois, and PhD in Computer Science from the University of Maryland. Dr. Clancy is author to over 60 peer-reviewed technical papers, nine Internet standards, holds one patent, and is a Senior Member of the IEEE. His research interests are in wireless and spectrum security.