

RESEARCH

Open Access

Intelligent cognitive radio jamming - a game-theoretical approach

Kresimir Dabcevic^{*}, Alejandro Betancourt, Lucio Marcenaro and Carlo S Regazzoni

Abstract

Cognitive radio (CR) promises to be a solution for the spectrum underutilization problems. However, security issues pertaining to cognitive radio technology are still an understudied topic. One of the prevailing such issues are intelligent radio frequency (RF) jamming attacks, where adversaries are able to exploit on-the-fly reconfigurability potentials and learning mechanisms of cognitive radios in order to devise and deploy advanced jamming tactics. In this paper, we use a game-theoretical approach to analyze jamming/anti-jamming behavior between cognitive radio systems. A non-zero-sum game with incomplete information on an opponent's strategy and payoff is modelled as an extension of Markov decision process (MDP). Learning algorithms based on adaptive payoff play and fictitious play are considered. A combination of frequency hopping and power alteration is deployed as an anti-jamming scheme. A real-life software-defined radio (SDR) platform is used in order to perform measurements useful for quantifying the jamming impacts, as well as to infer relevant hardware-related properties. Results of these measurements are then used as parameters for the modelled jamming/anti-jamming game and are compared to the Nash equilibrium of the game. Simulation results indicate, among other, the benefit provided to the jammer when it is employed with the spectrum sensing algorithm in proactive frequency hopping and power alteration schemes.

Keywords: Jamming; Anti-jamming; Game theory; Stochastic game; Non-zero-sum game; Fictitious play; Adaptive play; Markov models; Frequency hopping; Power allocation; Smart transmission layer

1 Introduction

Cognitive radio (CR) [1] is a technological breakthrough that - by utilizing concepts of dynamic spectrum access (DSA) and opportunistic spectrum access (OSA) - is expected to bring about means for better radio frequency spectrum utilization.

In order to access spectrum opportunistically, CRs need to be able to acquire information related to the spectrum holes. Currently, there are three established methods allowing the cognitive radios to retrieve the spectrum occupancy information. Among them, spectrum sensing [2] has been given the most focus in the research community.

Different spectrum sensing approaches, such as energy detection [3], matched filters [4], various feature detection methods (e.g., cyclostationary) [5], and hybrid methods [6] have been proposed and analyzed in the past. These are

mainly differentiated by their computational complexity (energy detection being the most computationally efficient, thus simplest to implement), necessity for *a priori* knowledge of the observed signals (matched filters), and the means of extracting features of the recognized signals (feature detectors).

Alternative approach to acquiring spectrum information is utilized by the geolocation/database-driven CRs [7]. This method requires the CRs to have a perfect awareness of their geographical position and to be able to access a database containing the list of the currently available frequencies at a given location.

Another alternative method is the beacon signals method [8], which relies on the usage of beacon rays for providing the prospective CRs information regarding the currently unused channels in their proximity.

As useful as the newly introduced cognitive abilities and on-the-fly reconfigurability perspectives of cognitive radios may be for the functioning of the future wireless communication systems, they also inherently bring

^{*}Correspondence: kresimir.dabcevic@ginevra.dibe.unige.it
DITEN - Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture, University of Genova, Via Opera Pia 11A, Genoa, Italy

a new set of security issues [9] that need to be properly addressed. Among them, primary user emulation attacks [10], spectrum sensing data falsification attacks [11,12], eavesdropping attacks [13], and intelligent jamming attacks [14,15] were given particular attention in the research community.

Radio frequency (RF) jamming attacks may be defined as illicit transmissions of RF signals aimed at disrupting the normal communication on the targeted channels. Adversaries that utilize the CR learning mechanisms to improve their jamming capabilities are considered intelligent. Intuitively, being equipped with such learning mechanisms may also aid the legitimate users in improving their anti-jamming capabilities. The goals of legitimate transceivers and jammers are typically negatively correlated. For this purpose, game theory - a mathematical study of decision-making in situations involving conflicts of interest - has emerged as a tool for formalization of the intelligent jamming problems.

Most of the previous works in the literature on application of game theory to jamming problems consider either channel surfing or the power allocation as anti-jamming strategies. Furthermore, they are mutually differentiated mostly by the objective function subjected to optimization (signal-to-noise ratio, bit error rate, Shannon capacity), various forms of uncertainty (user types, physical presence, system parameters), game formulation (zero-sum vs. non-zero-sum, single-shot vs. dynamic), learning algorithms (Q-learning, SARSA, policy iteration), etc.

In [16], authors have proven the existence and uniqueness of Nash equilibrium for a class of games with transmission cost. In addition, they have derived analytical expressions for the Nash equilibrium and have formulated the jamming game as the generalization of the water-filling optimization problem. Jamming game for OFDM system with 5 channels was analyzed.

Authors in [17] have formulated the problem of jamming in CR networks with primary users as a zero-sum stochastic game, where channel hopping was considered as the anti-jamming scheme and minimax-Q as the learning algorithm. They have compared the performance of the developed stationary policy with the myopic decisioning policy which did not consider the environment dynamics.

The method was extended in [18], comparing the results of Q-learning with those of the policy iteration scheme. The performance of the proposed scheme was evaluated against attackers of varying levels of sophistication.

In [19] and [20], multi-carrier power allocation was considered as an anti-jamming strategy. The games were also formulated as zero-sum.

In [21], a study of the performance of fictitious play as the learning algorithm in intelligent jamming games was performed. In this work, we extend upon

the aforementioned ideas. A formulation of a proactive jamming/anti-jamming game with intelligent players in an increased action space created by combining frequency hopping and power alteration is considered. Due to hopping and transmission costs, the game is formulated as non-zero-sum. Simulation results are used for finding near-optimal strategies for a game with incomplete information on an opponents' payoffs and strategy. The results are compared to the Nash equilibrium of the game. Oftentimes, there exists a significant gap between the theoretical contributions and the practical aspects of the radio systems. In order to infer the parameters relevant for the modelled game, thus bridging this gap, a set of experiments is performed on the real-life software-defined radio SDR/CR test bed.

We summarize the contributions and novelties of this paper with respect to the state-of-the-art papers on the application of game theory to intelligent jamming scenarios as follows:

- We present the ideas of learning algorithms that correspond to CRs with and without spectrum sensing capabilities, comparing their performance.
- We compare the performance of the considered learning algorithms for the modeled game with the Nash equilibrium of the game.
- We consider an increased action space created by combining two anti-jamming tactics.
- We use a real-life SDR/CR platform to infer parameters that allow modeling the game in a more realistic manner.

The remainder of the paper is structured as follows: Section 2 describes the system model. Game formulation along with equilibrium analysis and the description of considered learning algorithms and decisioning policies is presented in Section 3. SDR/CR test bed setup is described in Section 4, with experimental results presented in Section 5. Application to the modelled game and the simulation results are given in Section 6, whereas conclusions are drawn in Section 7.

2 System model

Consider a simplistic two-way transmitter-receiver communication occurring over one of the n_f pre-defined channels and a malicious user (jammer) that is trying to disrupt the communication by creating narrowband interference. Transmitter and receiver are considered the primary users over all of the considered channels and are able to tune to the same channel at a given time instance. Without the loss of generality, all of the channels are modelled with the same parameters; however, it will become obvious that the proposed anti-jamming techniques would be able to indirectly infer different channel parameters

and fit these inferences into their decision-making process.

Jammer is able to create a narrowband interference on a single channel at a time, causing the deterioration of the signal to interference plus noise ratio (SINR) and subsequently increase of the bit error rate (BER) on that channel. It is assumed that the jamming attack is the only possible reason for the deterioration of the channel quality, neglecting other possible sources of interference, as well as the time-varying nature of channels, including effects of the multipath propagation.

To mitigate the jamming effects and increase the SINR at the receiver side over the threshold needed for successful decoding, the transmitter may deploy a combination of channel hopping and increasing its transmission power (power alteration).

Both transmitter and jammer are able to make use of the on-the-fly reconfigurability as well as the learning prospectives of the cognitive radio technology. In different studied scenarios, both transmitter and jammer may have different spectrum sensing capabilities. Following that, two different learning algorithms are studied: payoff-based adaptive play (PBAP), where players are not necessarily embodied with spectrum sensing, and fictitious play, where players are able to infer the actions of the opponent in each step as a result of the deployed spectrum sensing scheme. In addition, performance of the proposed jamming/anti-jamming schemes is evaluated against static, non-learning types of opponents.

Other assumptions and abstractions that were taken in order to take a game-theoretical approach to jamming/anti-jamming problems are given as follows:

- Considered channels are perfectly orthogonal and non-overlapping, with frequency spacing between them large enough to make any energy spillover negligible.
- A discrete number of transmission powers were considered for both the transmitter and the jammer.
- Following the previous assumption, an occurrence of jamming is modelled as a discrete event, i.e., it always occurs with success or failure, disregarding the typical stochastic processing involved with the occurrence of jamming^a.
- Both transmitter and jammer are in continuous transmission mode, i.e., they always have packets ready to send.
- Jammer is available to create interference powerful enough to successfully jam communications when the transmitter is transmitting with its maximum transmission power (provided that they are both transmitting on the same channel at the time).
- All players maintain their relative positions as well as antenna orientations with respect to each other.

3 Game formulation

The attack and defense problem is modelled as a multi-stage proactive jamming/anti-jamming stochastic game. A stochastic game [22] is played in a sequence of steps, where at the end of each step, every player receives a payoff for the current step and chooses an action for the next step that is expected to maximize his payoff. A player's payoff in each step is determined not only by his action but also by the actions of all the other players in the game. Collection of all of the actions that a player can take comprise his (finite) action set. The distribution of a player's choices of actions constitute his strategy. The strategy may be fixed or may be updated according to the deployed learning algorithm.

The proposed game is an extension of Markov decision process (MDP), whose state transition probabilities may be depicted as finite Markov chains.

The modelled game consists of two players: transmitter T and jammer J . At the end of each step, every player observes his payoff for the given step and decides either to continue transmitting with the same power and at the same frequency or to change one of them, or both. The payoff consists of a summation of reward for the successful transmission (jamming), penalty for the unsuccessful transmission (jamming), and negative values related to cost of transmission (jamming) and cost of frequency hopping. Transmission (jamming) cost is related to the power spent by the user for transmitting (jamming) in a given step. Hopping cost may be explained by the fact that, after changing the channel of the transceiver pair (jammer), a certain time elapses before the communication may be resumed (interference created) due to the settling time of the radios or by other hardware constraints.

A generalized payoff at the end of the step s for transmitter T is expressed as (1). Here, R^T denotes the reward for successful transmission, X^T is the sustained fixed penalty for the unsuccessful transmission, H is the hopping cost, $g(C^T)$ is a function that expresses the transmitter's cost of transmission when power C^T is used, f^T is the channel currently used by the transmitter-receiver pair, $\alpha = 1$ if transmission is successful and $\alpha = 0$ if not, and $\beta = 1$ if the transmitter decides to hop and $\beta = 0$ otherwise. In this notation, subindices are used to denote steps, and superindices to denote the players.

$$P_s^T \left(C_s^T, f_s^T, C_s^J, f_s^J \right) = R^T \alpha - X^T (1 - \alpha) - H \cdot \beta - g \left(C_s^T \right) \quad (1)$$

Similarly, jammer J 's generalized payoff for the step s is given as (2). Here, R^J is the jammer's reward for successful jamming, X^J is the sustained fixed penalty for the unsuccessful jamming, $g(C^J)$ is the jammer's cost of

transmission when power C^J is used. Finally, $\gamma = 1$ if the jammer decides to hop and 0 if it does not.

$$P_s^J \left(C_s^T, f_s^T, C_s^J, f_s^J \right) = R^J(1-\alpha) - X^J\alpha - H\gamma - g(C_s^J) \quad (2)$$

3.1 Equilibrium analysis of the game

Nash equilibrium is inarguably the central concept in game theory, representing the most common notion of rationality between the players involved in the game. It is defined as the set of distributions of players' strategies designed in a way that no player has an incentive to unilaterally deviate from its strategy distribution.

Let n_f be a discrete number of channels available to both players for channel hopping, and let n_{CT} and n_{CJ} be the discrete number of transmission powers for the transmitter and the jammer, respectively. For the game with $n_f \cdot n_{CT}$ ($n_{CT} = n_{CJ}$) pure strategies available to each player, we define S^T as the set of pure strategies of the transmitter and S^J as the set of pure strategies of the jammer. Then, $x \in \mathbb{R}^{S^T}$ and $y \in \mathbb{R}^{S^J}$ represent the mixed strategies of the transmitter and jammer, respectively. By denoting the payoff matrices of the transmitter and jammer as A and B , respectively, a best response to the mixed strategy y of the jammer is mixed strategy x^* of the transmitter that maximizes its expected payoff $x^{*\top}Ay$. Similarly, the jammer's best response y^* to the transmitter's mixed strategy x is the one that maximizes $x^\top By^*$. A pair (x^*, y^*) that are best responses to each other is a Nash equilibrium of the bimatrix game, i.e., for any other combination of mixed strategies (x, y) the following equations hold true:

$$xAy^{*\top} \leq x^*Ay^{*\top}, \quad (3)$$

$$x^*By^\top \leq x^*By^{*\top}. \quad (4)$$

In 1951, Nash proved that all finite non-cooperative games have at least one mixed Nash equilibrium [23]. Particularization of this proof for bimatrix games may be given as follows [24]:

Let x and y be arbitrary pairs of mixed strategies for the bimatrix game (A, B) , and A_i and B_j represent the i th column and the j th row of the matrices A and B , respectively. Then,

$$c_i = \max \{ A_i y^\top - xAy^\top, 0 \}, \quad (5)$$

$$d_j = \max \{ xB_j - xBy^\top, 0 \}, \quad (6)$$

$$x'_i = \frac{x_i + c_i}{1 + \sum_k c_k}, \quad (7)$$

$$y'_j = \frac{y_j + d_j}{1 + \sum_k d_k}. \quad (8)$$

Since $T(x, y) = (x', y')$ is continuous and x' and y' are mixed strategies, it can be shown that $(x', y') = (x, y)$ if and only if (x, y) is an equilibrium pair. Furthermore, if (x, y) is an equilibrium pair, then for all i :

$$A_i y^\top \leq xAy^\top, \quad (9)$$

hence $c_i = 0$ (and similarly $d_j = 0$ for all j), meaning that $x' = x$ and $y' = y$. Assume now that (x, y) is not an equilibrium pair, i.e., there either exists \bar{x} such that $\bar{x}Ay^\top > xAy^\top$, or there exists \bar{y} such that $xB\bar{y}^\top > xBy^\top$. Assuming the first case, as $\bar{x}Ay^\top$ is a weighted average of $A_i y^\top$, there must exist i for which $A_i y^\top > xAy^\top$, and hence some $c_i > 0$, with $\sum_k c_k > 0$. As $\bar{x}Ay^\top$ as a weighted average of $A_i y^\top$, there must exist $A_i y^\top \leq xAy^\top$ for some i such that $x_i > 0$. For this i , $c_i = 0$, hence:

$$x'_i = \frac{x_i + c_i}{1 + \sum_k c_k} < x_i, \quad (10)$$

and so $x' \neq x$. In the same way, it can be shown that $y' \neq y$, leading to the conclusion that $(x', y') = (x, y)$ if and only if (x, y) is an equilibrium. As the transformation $T(x, y) = (x', y')$ is continuous, it must have a fixed point, and so by applying Brouwer's fixed point theorem [25], it follows that this fixed point indeed represents an equilibrium point. This concludes the proof of the existence of mixed-strategy equilibrium points in a bimatrix game.

However, efficient computation of equilibria points, as well as proving uniqueness of an equilibrium, remains an open question for many classes of games. Lemke-Howson (LH) [26] is the most well-known algorithm for the computation of Nash equilibria for bimatrix games and is our algorithm of choice for finding the Nash equilibrium strategies. A bimatrix game requires the game to be fully defined by two payoff matrices (one for each player). Since in our case the immediate payoff of every player in each step depends not only on his own action and the action of the opponent but also on the previous state of the player (influence of the hopping cost), our game as a whole cannot be represented by two deterministic payoff matrices. For this reason, we divide the game into $n_f \cdot n_{CT}$ subgames, where each subgame corresponds to a unique combination of possible states of the transmitter and the jammer. Since each subgame can be treated as the separate game in a bimatrix form, we proceed to apply the LH method to find mixed strategy Nash equilibria (one per subgame). Hence, in each step, every player plays an equilibrium strategy corresponding to that step. A union of equilibria strategies of all the $n_f \cdot n_{CT}$ combinations of the states within the game may be considered as the Nash equilibrium of the game.

Gambit [27], an open-source collection of tools for solving computational problems in game theory, was used for finding equilibrium points using the LH method. For details on the implementation of the LH algorithm, an interested reader is referred to [28].

Each of the subgames (A_{ij}, B_{ij}) where $i = 1 \dots n_f$ and $j = 1 \dots n_{CT}$ is a nondegenerate bimatrix game. Then, following Shapley's proof from [26], we may conclude that there exists an odd number of equilibria for each subgame. In [29], the upper bound on the number of equilibria in

$d \times d$ bimatrix games was shown to be equal to $\frac{2.41^d}{d^{1/2}}$; however, the uniqueness of Nash equilibrium may still be proven only for several special classes of bimatrix games. Here, we provide conditions that the bimatrix game has to satisfy in order to have a unique completely mixed Nash equilibrium. Completely mixed Nash equilibrium is an equilibrium in which the supports of each of the mixed equilibrium strategies are equal to the number of available pure strategies (i.e., each strategy from a mixed strategy set is played with a non-zero probability). As shown by [30], whose proof we re-state, a bimatrix game (A, B) whose matrices A and B are a square, has a unique completely mixed Nash equilibrium if $\det(A, \mathbf{e}) \neq 0$ and $\det(B, \mathbf{e}) \neq 0$, i.e.:

$$\det(A, \mathbf{e}) \cdot \det(B, \mathbf{e}) \neq 0, \quad (11)$$

where \mathbf{e} is a column vector with all entries 1.

The saddle point matrix (A, \mathbf{e}) is given by:

$$(A, \mathbf{e}) = \begin{bmatrix} A & \mathbf{e} \\ \mathbf{e}^T & 0 \end{bmatrix}. \quad (12)$$

Then, the equilibrium strategies of the players are given as:

$$x^*_i = -\frac{\det B^i}{\det(B, \mathbf{e})}, \quad (13)$$

$$y^*_i = -\frac{\det A_i}{\det(A, \mathbf{e})}, \quad (14)$$

where B^i (A_i) is the matrix of B (A) with all entries of the i th column (row) replaced by 1.

Let us now suppose that (x^*, y^*) is an equilibrium point of the bimatrix game (A, B) , where x^* is completely mixed. Then, every pure strategy would give that player the same payoff P against the opponent's strategy y^* , i.e.:

$$Ay^* = Pe. \quad (15)$$

Since y^* is a vector of probabilities,

$$\mathbf{e}^T y^* = 1. \quad (16)$$

Or, in matrix form:

$$\begin{bmatrix} A & \mathbf{e} \\ \mathbf{e}^T & 0 \end{bmatrix} \begin{bmatrix} y^* \\ -P \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (17)$$

Following the assumption $\det(A, \mathbf{e}) \neq 0$ and by applying Cramer's rule, it follows from (17) that (14) is true for $(i = 1, 2, \dots, n)$ (in our case, $n = n_{CT} \cdot n_f$). Similarly, the same holds for x^*_i . As shown in [30]:

$$\det(A_i, \mathbf{e}) = \det(A_i - \mathbf{e}\mathbf{e}^T) - \det(A_i) = -\det A_i, \quad (18)$$

hence (13) and (14) are shown to be true. This concludes the proof of the uniqueness of the completely mixed equilibrium.

It may be computationally shown that all of the $n_f \cdot n_{CT}$ subgames constructed within the considered game satisfy (11). Furthermore, by observing the Markov state chains

corresponding to the equilibrium points found by the LH method, it may indeed be observed that $\text{supp}(x^*) = \text{supp}(y^*) = n_f \cdot n_{CT}$, i.e., the equilibria are completely mixed. Trying to find multiple equilibria for each subgame using other computational methods available within [27] has also resulted in a single (completely mixed) equilibrium for each subgame: empirical evaluation of these results, based on the algorithms to find all possible equilibrium points of the bimatrix game, further points to the existence of a unique Nash equilibrium for each subgame.

One of the common criticisms of using computational algorithms such as LH for finding Nash equilibria is that they fail to realistically capture the way that the players involved in the game may reach the equilibrium point. For this reason, it is useful to discuss the payoff performance and the convergence properties to Nash equilibrium of the algorithms realistically used for learning in games. This discussion is done for two multi-agent learning algorithms considered within this work: fictitious play (Section 3.2.1) and payoff-based adaptive play (Section 3.2.2).

3.2 Learning algorithms

Learning algorithms for MDPs have been extensively studied in the past [31,32]. Based on their spectrum occupancy inference capabilities, an illustrating example of the corresponding learning algorithms for the considered game and the dimensionality of the action space is given in Figure 1.

For CRs not equipped with spectrum sensing capabilities (geolocation/database-driven CRs and CRs utilizing beacon rays), payoff-based reinforcement algorithms impose themselves as the optimal viable learning algorithms. In these cases, each player is able to evaluate the payoff received in every step and modify its strategy accordingly.

CRs able to perform energy detection spectrum sensing, in addition, also have the possibility of observing their opponents' actions in each step (influenced possibly by the accuracy of the deployed spectrum sensing mechanism). By incorporating these observations into their future decision-making process, the players may build and update a belief regarding the opponents' strategy distribution. This learning mechanism is called fictitious play.

Finally, CRs able to perform feature detection spectrum sensing may recognize important parameters of the opponent's signal and use these observations to their advantage. Since various waveforms exhibit different jamming and anti-jamming properties, depending mainly on their modulation and employed coding (see, for example, [33]), increased action space could consist of switching between multiple modulation types or coding techniques.

In this paper, we focus our analysis on the first two cases. Algorithm 1 illustrates the general formulation of the game. It can be seen how, in every step, each player takes a decision d_s for his next action based on their

expected utility $\bar{P}_s = E[P_s|P_{1:s-1}]$ under PBAP or $\bar{P}_s = E[P_s|P_{1:s-1}, ss_{1:s-1}]$ under fictitious play. Received payoffs P_s are calculated for each player using (1) and (2). Thereafter, spectrum sensing is performed and the expected payoff is updated with the new information available. To simplify explanation of the learning strategies and Algorithm 1, it is assumed that both players perform the spectrum sensing step; however, the result of this step is used only under fictitious play framework. For the players with perfect spectrum sensing capabilities, $ss_s^T = d_s^T$ and $ss_s^J = d_s^J$.

Algorithm 1 Game pseudocode

```

1: function TRANSMITTER JAMMER GAME
2:    $nSteps \leftarrow$  Number of steps
3:    $R^T, R^J \leftarrow$  Rewards
4:    $C^T, C^J \leftarrow$  Cost of hopping
5:
6:   Initialize the expected utilities
7:    $s \leftarrow 0$ 
8:   while  $s < nSteps$  do
9:      $d_s^T \leftarrow$  Decide Transmitter ▷ Section 3.3
10:     $d_s^J \leftarrow$  Decide Jammer ▷ Section 3.3
11:     $P_s^T \leftarrow$  Transmitter utility ▷ Equation 1
12:     $P_s^J \leftarrow$  Jammer utility ▷ Equation 2
13:     $ssd_s^T \leftarrow$  Transmitter Spectrum Sensing
14:     $ssd_s^J \leftarrow$  Jammer Spectrum Sensing
15:     $Transmitter.learn(P_s^T, ssd_s^T)$  ▷ Section 3.2
16:     $Jammer.learn(P_s^J, ssd_s^J)$  ▷ Section 3.2
17:     $s \leftarrow s + 1$ 
18:  end while
19: end function

```

Note from the pseudocode that the game consists of two main parts: the learning algorithm, in charge of updating the expected payoffs, and the decisioning policy, which uses the available observations to decide upon the future actions.

Let us assume that in step s the transmitter was transmitting with power C_s^T on the frequency f_s^T . Using one of the decisioning policies described in Section 3.3, its action in the next step constitutes of transmitting with power C_{s+1}^T on frequency f_{s+1}^T . We denote this action as a list of four elements $d_s^T = [C_s^T, f_s^T, C_{s+1}^T, f_{s+1}^T]$ for the transmitter and the equivalent values $d_s^J = [C_s^J, f_s^J, C_{s+1}^J, f_{s+1}^J]$ for the jammer.

3.2.1 Fictitious play

Fictitious play [34] is an iterative learning algorithm where, at every step, each player updates his belief about the stochastic distributions of the strategies of the other

players in the game. The application of a learning mechanism based on fictitious play to the modelled game is constructed under the assumption that the player is necessarily endowed with the spectrum sensing capabilities, allowing him to infer the actions of the other player. A payoff of a particular action given the player's current state and the opponent's action is deterministic and may be calculated using (1) and (2) for transmitter and jammer, respectively. If the player has the information regarding the opponents' action in each step, then it is possible to calculate the expected utility more precisely, by accessing the history of the opponents' actions. This is particularly true for the jammer because of the higher number of non-jammed states compared to the states of successful jamming. Hence, learning the transmitter's pattern as soon and with as much precision as possible makes a significant difference to the overall payoff. This updating process is denoted in Algorithm 2.

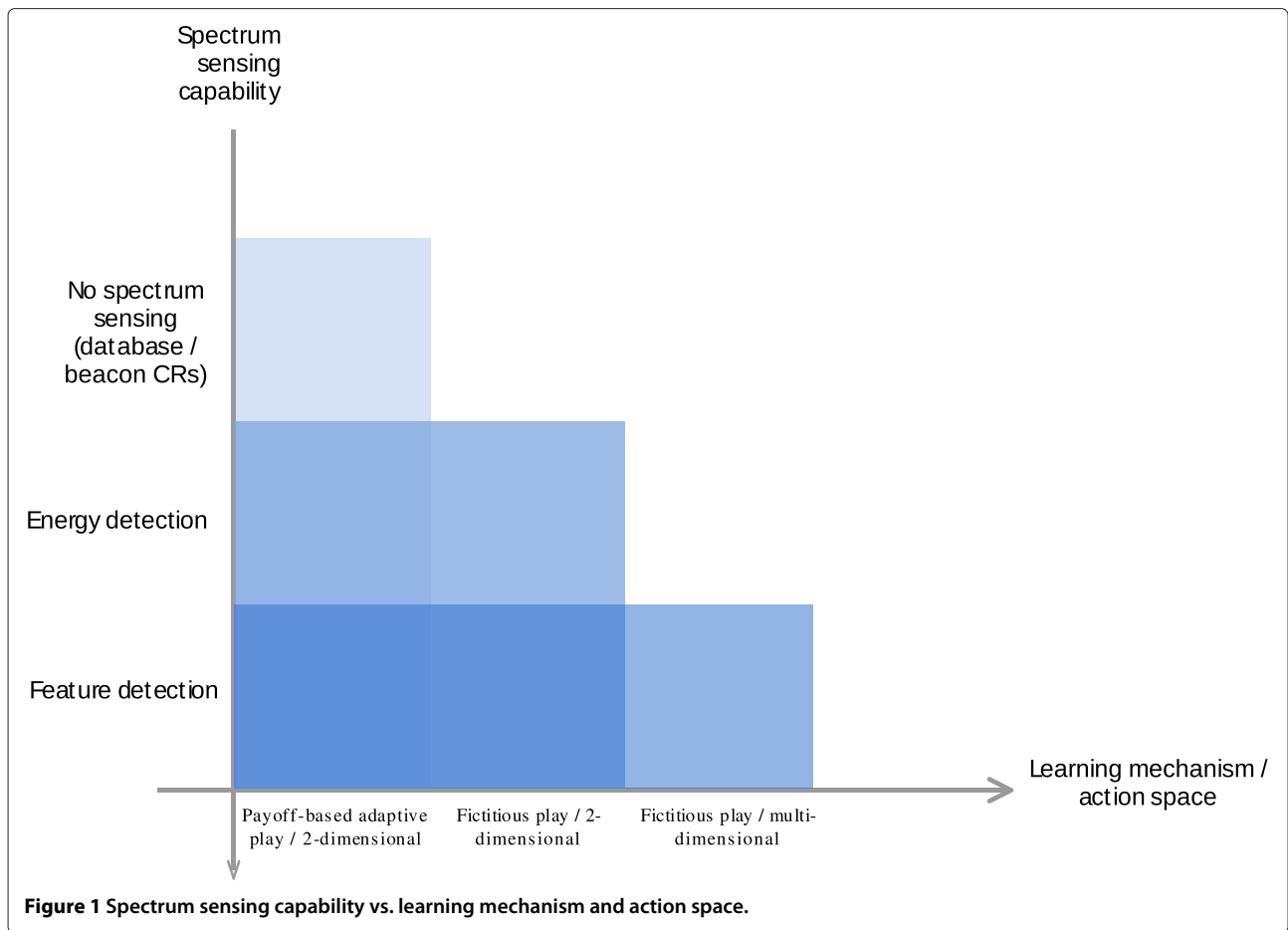
Algorithm 2 Expected utility update under fictitious play

```

1: function FICTIOUS EXPECTED UTILITY UPDATE
2:    $powers \leftarrow$  Available powers
3:    $freqs \leftarrow$  Available frequencies
4:    $SS \leftarrow$  Opponent's state [Spectrum Sensing]
5:    $flist \leftarrow$  Opponent's previous states
6:
7:    $flist.append(SS)$ 
8:   for  $d \in$  possible actions do
9:      $sum \leftarrow 0$ 
10:    for  $C, f \in powers, freqs$  do
11:       $N \leftarrow$  count( $C, f$ ) in flist
12:       $sum \leftarrow sum + N \cdot P(d[3], d[4], C, f)$ 
13:    end for
14:     $P_{s+1}^T(d) = \frac{sum}{s}$ 
15:  end for
16: end function

```

It is known that the convergence of the fictitious play to Nash equilibrium is guaranteed only for several special cases, such as zero-sum games, non-degenerate $2 \times n$ games with generic payoffs, games solvable by iterated strict dominance and weighted potential games. For other types of games, including the game considered within this work, convergence to Nash equilibrium is not guaranteed, and even when it converges, the time needed to run the algorithm to convergence may be very long due to the problem being polynomial parity arguments on directed graphs (PPAD)-complete [35]. This has led to the introduction of the concept of approximate Nash equilibrium (ϵ -equilibrium). Here, ϵ is a small positive quantity representing the maximum increase in payoff that a player could gain by choosing to follow a different strategy.



Author in [36] has shown that fictitious play achieves the worst-case guarantee of $\epsilon = (r + 1)/(2r)$ (where r is the number of FP iterations) and in reality provides even better approximation results. Furthermore, as recently shown in [37], fictitious play may in some cases outperform any actual Nash equilibrium - for this reason, it is useful to study the performance of the FP algorithm in terms of the average and final payoff compared to the Nash equilibrium.

3.2.2 Payoff-based adaptive play

Payoff-based adaptive play [38] is a form of reinforcement learning algorithm, where it is assumed that the player does not have access to the information about the state of the other player and relies on the history of his own previous payoffs. The expected utility of d_s given previous payoffs is given by Equation 19.

$$\overline{P_{s+1}(d_s)} = E[P_s(d_s)|P_{1:s-1}(d_s)] = \frac{\overline{P_s(d_s)} \cdot s + P_s(d_s)}{s + 1} \quad (19)$$

PBAP has been shown to converge to Nash equilibrium for zero-sum games [39]. For general finite two-player games, it was shown to converge to close-to-optimal solutions in polynomial time [40].

In addition to comparing the performance of the PBAP to the computed Nash equilibrium strategy from Section 3.1, of particular interest to this work is the comparison to the performance of the FP. This comparison should reflect the benefit that each player gains by being equipped with the spectrum sensing algorithm (FP) over not being equipped with it (PBAP).

3.3 Decisioning policies

A decisioning policy of the learning algorithm corresponds to the set of rules that the player uses to select his future actions.

3.3.1 Greedy decisioning policy

The most intuitive decisioning policy consists of always choosing the action that is expected to yield the highest possible value based on the current estimates - the so-called greedy decisioning policy [41]. However, a greedy method is overly biased and may easily lead the learning algorithm to 'get stuck' in local optimal solutions. An example of this is given in Figure 2, where both players are employing the greedy decisioning policy. Here, each player fairly quickly learns the 'best response' to an opponent's action and starts relying on using it. Then, a

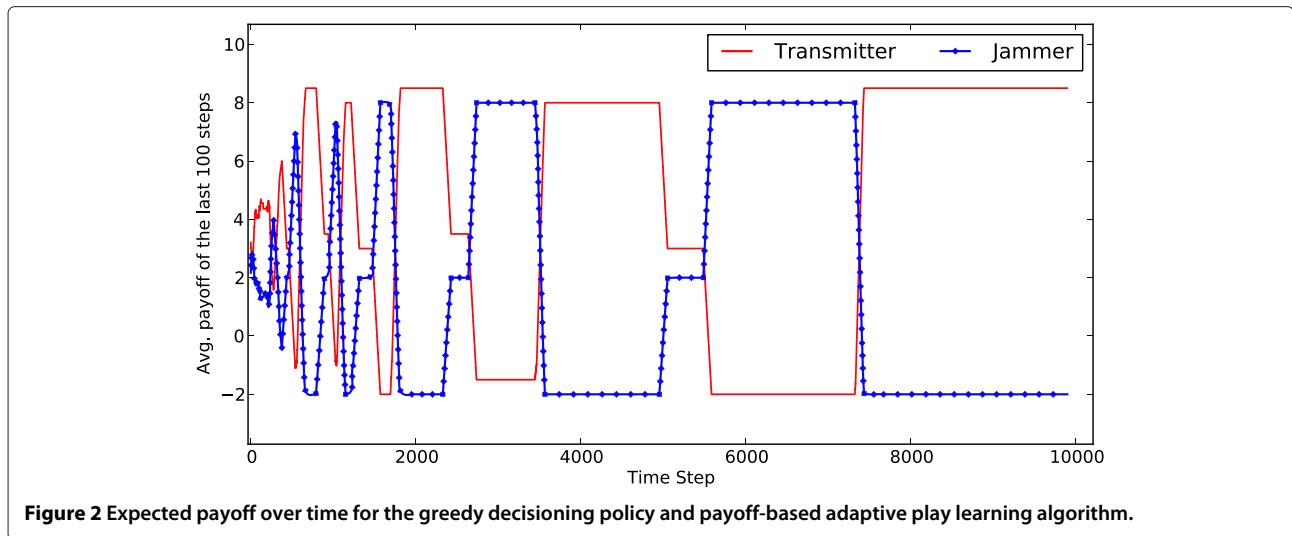


Figure 2 Expected payoff over time for the greedy decisioning policy and payoff-based adaptive play learning algorithm.

significant amount of time has to pass before his expected payoff for the given action drops enough that another action starts being considered as ‘best response’, where in the meantime significant payoff losses are sustained. This could partially be mitigated by introducing temporal forgiveness into the learning algorithm.

3.3.2 Stochastically sampled decisioning policy

Another common approach to this issue is choosing a stochastically sampled policy (also known as ϵ -greedy policy, [42]) where, at each step, a randomly sampled action is taken with a probability p . We propose a variation of the stochastically sampled policy where sampling is performed by scaling the expected payoff value of each action to the minimum possible payoff for the game. For a minimum payoff $PMIN$ and n actions with expected payoffs, $\overline{P(1)} \dots \overline{P(n)}$ the probability of choosing an action d is given by (20):

$$p(d) = \frac{\overline{P(d)} - PMIN}{\sum_{k=1}^n \overline{P(k)} - PMIN} \quad (20)$$

4 Experimental setup

In order to infer the parameters related to the occurrence of jamming and to be able to extract the physical parameters relevant for the game, a set of experiments using the real-life SDR test bed [43] is performed.

4.1 Test bed description

A coaxial test bed is implemented for the frequency range of interest. The coaxial test bed eliminates the typical uncertainties characteristic to wireless transmission and allows for repeatability of the experiments. An implemented test bed, shown in Figure 3, consists of two interconnected SWAVE handheld (HH) SDRs [44], and the dual directional coupler with 50-dB attenuation placed

in between, emulating the channel. SWAVE HH is an SDR terminal designed to operate in very high frequency (VHF) (30 to 88 MHz) and ultra high frequency (UHF) (225 to 512 MHz) bands. It is compliant with the Software Communications Architecture (SCA) 2.2.2 standard and supports a multitude of legacy as well as new waveforms. Each HH is connected to the personal computer (PC) via Ethernet as well as the RS-232 serial connection. Interfaces between the HH and the PC are illustrated in Figure 4.

Ethernet connection is used for the external control of HH’s transmission-related parameters. Using the Simple Network Management Protocol (SNMP) v3, values of all of the relevant parameters - for our purposes transmission frequency and transmission power - may be read out and altered on-the-fly.

Serial connection is used for transferring unprocessed spectrum data from the HH to the PC. There, this data is analyzed by the developed energy detection spectrum sensing mechanism and outputted to the spectrum intelligence mechanism (currently under development, [45]).

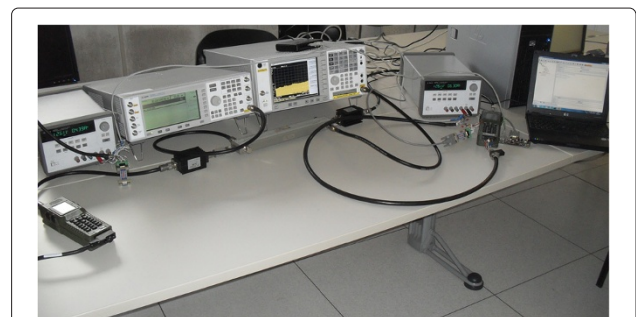


Figure 3 Implemented coaxial test bed.

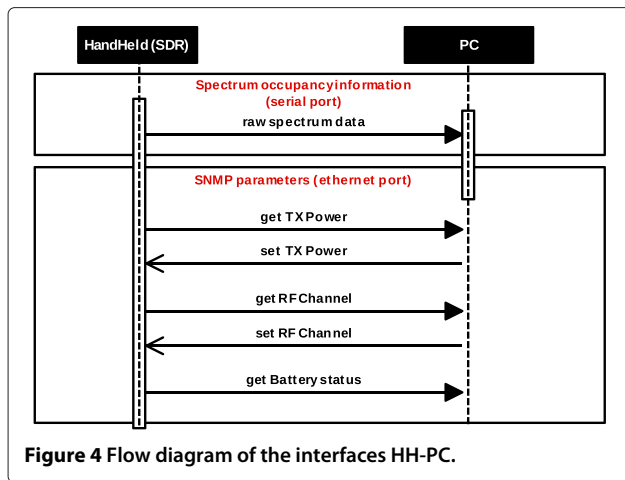


Figure 4 Flow diagram of the interfaces HH-PC.

4.2 Jammer implementation

In order to infer influence of intentional interference on the communication, a vector signal generator is used as a jammer emulator. Interference of various types (pulse tones, multitones, Global System for Mobile Communications (GSM) signal, additive white Gaussian noise (AWGN)), occupied bandwidth, and power may be created and injected in the channel.

5 Experimental methodology and results

The measurements and parameters relevant for the constructed game are:

- Impact of interference on the quality of communication link,
- Transmission levels,
- Battery life of the HHs for varying transmission levels
- Number of considered channels,
- Time needed to perform frequency hopping,
- Spectrum sensing time,
- Spectrum sensing detection accuracy.

The connection between the HHs is established using the soldier broadband waveform (SBW). SBW is a wideband multi-hop mobile *ad hoc* network (MANET) waveform, encompassed with self-establishment and self-awareness of the network structure and topology. The waveform's bandwidth is 1.3 MHz, and channel spacing is 2 MHz - large enough to disregard the influence of potential energy spillover between adjacent channels. Experiments are done at 300 MHz central carrier frequency.

Interference is created by injecting a pulse-shaped signal onto the central carrier frequency of the HHs. To measure the impact of interference, a set of BER tests was performed for varying levels of transmission power and varying levels of interference. Results for three discrete values of transmission power: -12, 4, and 7 dBW, respectively, are presented in Figure 5. By setting the threshold for the communication failure at $BER = 10^{-1}$, corresponding interference powers for the observed values of transmission powers are found, equaling to: 1, 6, and 9 dBW, respectively.

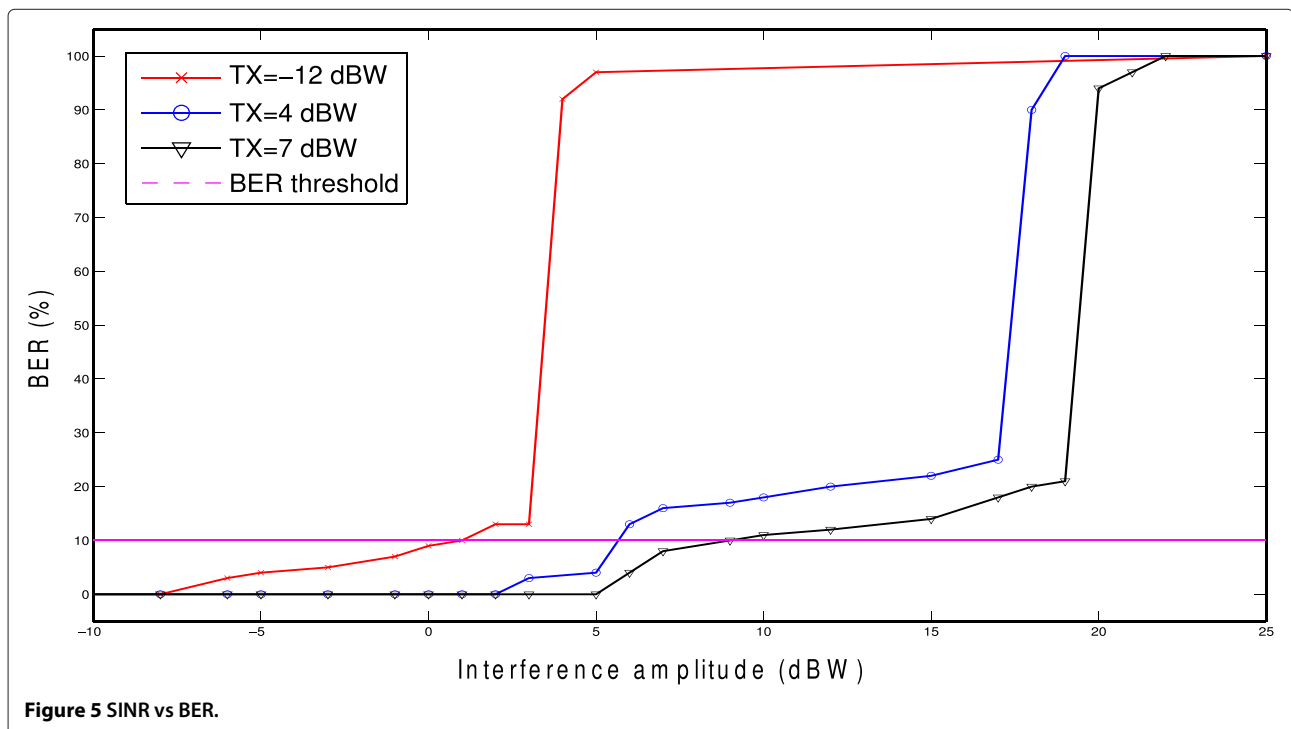


Figure 5 SINR vs BER.

Energy detection spectrum sensing is done in the following way: every 1.1 s, a burst of 8,192 samples from the HH's ADC is sent over the serial port to the PC. These samples, corresponding to 120 MHz of the bandwidth scanned around the HH's center frequency, are then converted into the frequency domain. The data may then be analyzed by the spectrum intelligence algorithm. This data processing currently lasts around 0.2 s, making the whole spectrum sensing cycle last approximately 1.3 s.

Currently, the test bed does not have the spectrum intelligence algorithm developed, whose task is processing spectrum sensing data related to the scanned wideband signal and concluding on the presence of narrowband waveforms in it. Hence, presently it is impossible to infer the detection accuracy of the spectrum sensing. For these purposes, various levels of detection accuracy manifesting in varying percentages of misdetection will be considered for the simulation results. An interested reader is referred to [46] for experimental results on the energy detection accuracy.

Frequency hopping is performed by issuing the appropriate SNMP SET command over the Ethernet port to the HH. The action of processing the SNMP request and changing the frequency takes 0.3 s, during which the radio is not transmitting. The frequency settling time of the radios is in this case negligible^b.

HH's battery time for states of continuous packet data stream (packets are generated by the BER test function) are measured for the identified relevant values of the transmission power of -12, 4, and 7 dBW, equaling to 120, 94, and 90 min, respectively. The results for the relevant transmission powers of the supposed jammer were then linearly interpolated from the aforementioned, equaling to 99, 92, and 87 min, respectively.

The relevant parameters are summarized in Table 1.

6 Application to the proposed game and simulation results

Starting from the general expressions for the payoffs of the transmitter and the jammer given in Equations 1 and 2, a short discussion is offered on the interpretation of the parameters measured in the previous section and the

feasibility of their application to the proposed game. The discussion is followed by the simulation results.

6.1 Adaptation of the measured parameters to the proposed game

One of the principal problems with introducing the experimental parameters in the theoretical model is the process of aligning the parameters with different units (namely, Watts and seconds), used in Equations 1 and 2. The first and second terms represent the transmission (jamming) reward and penalty, which may be defined arbitrarily. For the simulation purposes, we define them as $R = 1$ and $X = -R$, respectively.

Hopping cost, the third term of the equation, can be expressed as a function of the reward. If the hopping is performed and the transmission is successful, the final utility is decreased by the hopping cost, denoted as $Rh\alpha\beta$. Here, $h = \frac{0.3}{1.3}$ is the proportion of the time step where the transmission is not taking place due to the hopping process. An increase of the transmission power, on the other hand, directly influences battery life. For this purpose, transmission cost may be described as a function of battery life of the radio, as denoted in (21). Maximum battery life corresponds to the minimum transmission power of -12 dBW and equals to $B_{\max} = 120$ min. Transmission costs of higher transmission powers are then scaled with respect to this value.

$$g(C) = R \left(1 - \frac{B(C)}{B_{\max}} \right) \quad (21)$$

Finally, for each step s , expression (1) may be re-written as (22) and expression (2) as (23) for the transmitter and jammer, respectively.

$$\begin{aligned} P^T &= R^T \alpha - R^T (1 - \alpha) - R^T h \alpha \beta - R \left(1 - \frac{B(C_s^T)}{B_{\max}} \right) \\ &= R^T \left(\alpha(2 - H\beta) + \frac{B(C_s^T)}{B_{\max}} - 2 \right) \end{aligned} \quad (22)$$

$$\begin{aligned} P^J &= R^J (1 - \alpha) + R^J \alpha - R^J h \gamma (1 - \alpha) - R^J \left(1 - \frac{B(C_s^J)}{B_{\max}} \right) \\ &= R^J \left(H\gamma(\alpha - 1) + \frac{B(C_s^J)}{B_{\max}} \right) \end{aligned} \quad (23)$$

Table 1 Overview of the inferred parameters

	Transmitter	Jammer
Considered frequencies (MHz)	(300, 302.65, 305.3)	(300, 302.65, 305.3)
Transmitting powers (dBW)	(-12, 4, 7)	(1, 6, 9)
Battery life for TX powers (min)	(120, 94, 90)	(99, 92, 87)
Spectrum sensing time (s)	1.3	1.3
Frequency hopping time (s)	0.3	0.3
Signal detection accuracy (%)	(50, 70, 90, 100)	(50, 70, 90, 100)

Table 2 Overview of the parameters adapted to the game

	Transmitter	Jammer
R	1	1
h	$\frac{0.3}{1.3}$	$\frac{0.3}{1.3}$
$g(C)$	$(0, \frac{94}{120}, \frac{90}{120})$	$(\frac{99}{120}, \frac{92}{120}, \frac{87}{120})$
Default probability of misdetection [%]	0	0

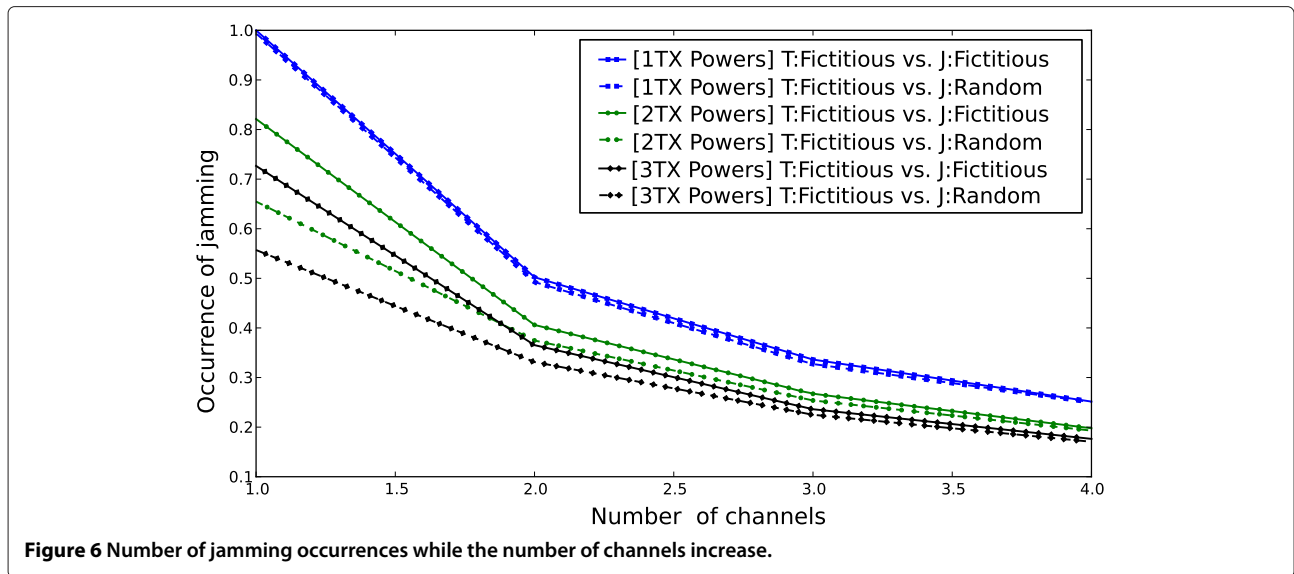


Figure 6 Number of jamming occurrences while the number of channels increase.

Following the experiments denoted in Figure 5, the occurrence of jamming in step s for the three couplets of transmission powers $C^T = (-12, 4, 7)$ and $C^J = (1, 6, 9)$ can be defined as (24). An overview of the adapted parameters is given in Table 2.

$$\alpha = \begin{cases} 1 & \text{if } T_s > J_s \text{ or } f_s^T \neq f_s^J \\ 0 & \text{if } T_s \leq J_s \text{ and } f_s^T = f_s^J. \end{cases} \quad (24)$$

6.2 Simulation results

In this subsection, we analyze the performance of the considered learning algorithms under the proposed game and compare it to the computed Nash equilibrium. All the games are constructed using the parameters denoted in Table 2, unless indicated otherwise. Default number of simulation steps is 10,000. Each simulation is repeated 100 times, and the points are averaged. It was verified that

each pair of the constructed payoff matrices satisfy condition (13), guaranteeing uniqueness of a completely mixed Nash equilibrium. In several games, a comparison with the player whose strategy is fully randomized, i.e., taken actions are irrespective of the observations, is performed. Figure 6 shows the percentage of occurrences of successful jamming for different dimensions of the players' action sets, from games with one channel and one transmission power, to four channels and three transmission powers. In all games, transmitter is playing FP, whereas jammer is alternating between FP (full lines) and random strategy (dashed lines). Benefit of having the learning algorithm for the jammer is particularly prominent for the low-dimensional games, where the transmitter is able to adapt to any static strategy of the jammer (including fully randomized) and start exploiting it significantly.

To verify the importance of spectrum sensing capabilities corresponding to the fictitious play learning algorithm, we propose the analysis of the overall utility of

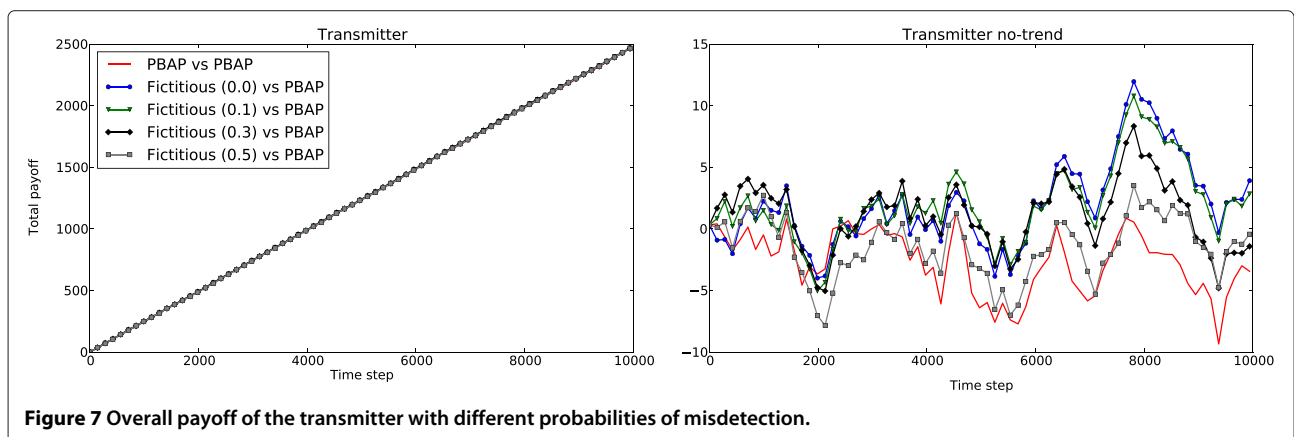


Figure 7 Overall payoff of the transmitter with different probabilities of misdetection.

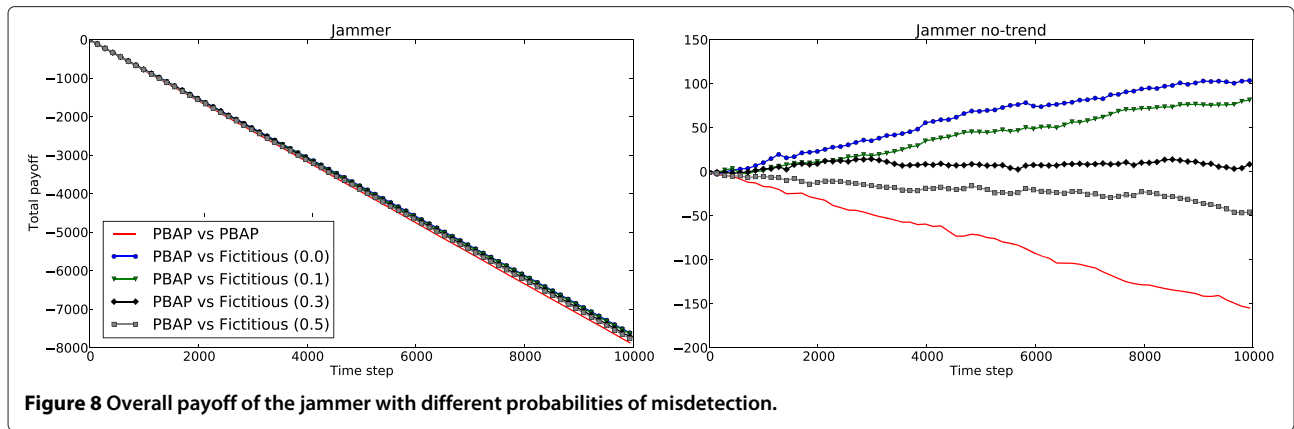


Figure 8 Overall payoff of the jammer with different probabilities of misdetection.

each player when the opponent is utilizing payoff-based adaptive play. Furthermore, in order to understand how the spectrum sensing accuracy affects the performance, we consider a spectrum sensing mechanism with a certain probability of misdetection. For the simplicity of analysis, we disregard the fact that the misdetection probability realistically depends on the instantaneous SINR. Figures 7 and 8 show the results of these simulations for transmitter and jammer, respectively. In the left side of the figures, the overall payoff obtained during the game for each player is shown. For the visualization purposes, a trend is removed in the right side of the figures.

From Figure 7, it is evident that the compared schemes perform almost equally - regardless of the misdetection probability - for the transmitter. This points to the conclusion that the optimal strategy of the transmitter under the considered game when the jammer is endowed with the learning algorithm is not too far from 'random'. On the

other hand, Figure 8 points once again to the significance of the spectrum sensing for the jammer side, as its overall payoff is significantly higher when utilizing fictitious play, compared to payoff-based adaptive play, even for sub-optimal spectrum sensing mechanisms (mechanisms with higher probabilities of misdetection).

In order to study this occurrence in more detail and in order to ease-up the comparison, we next present these results in the forms of normal distributions. Figure 9 shows the performance of the transmitter using PBAP learning algorithm in the upper part and fictitious play in the bottom part, for varying learning algorithms of the jammer. Analogously, Figure 10 shows the performance of the jammer employing PBAP learning algorithm in the upper part and fictitious play in the bottom part, for different learning algorithms of the transmitter. The title of each subplot denotes the learning algorithm followed by the observed player while colors of the lines are

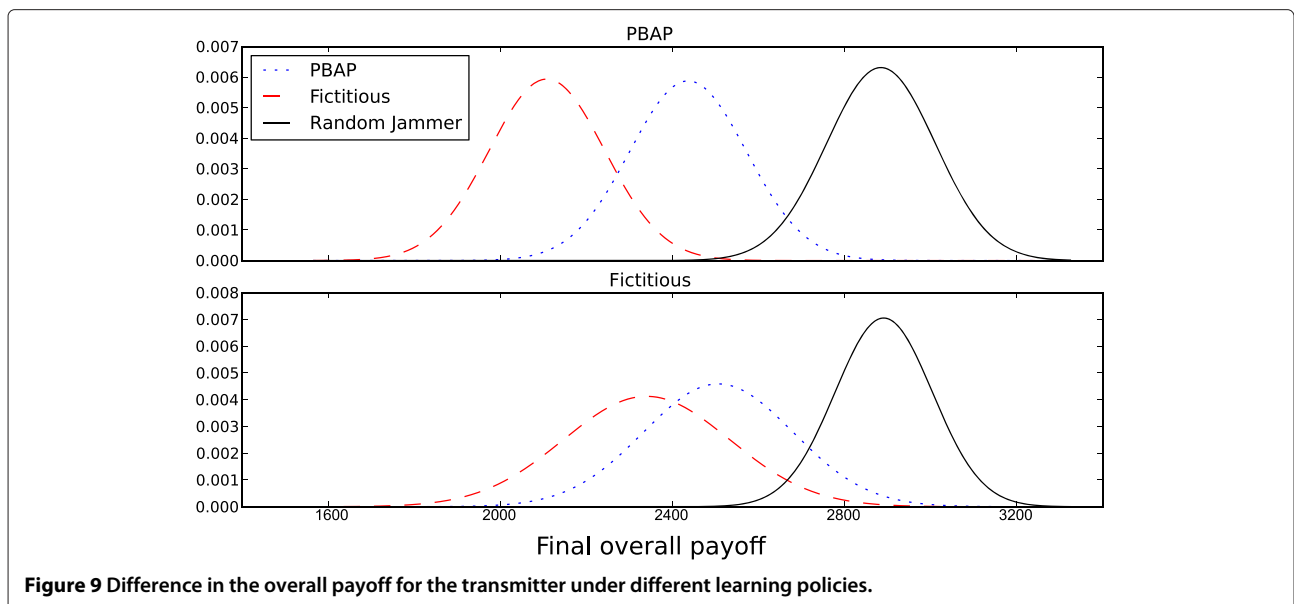
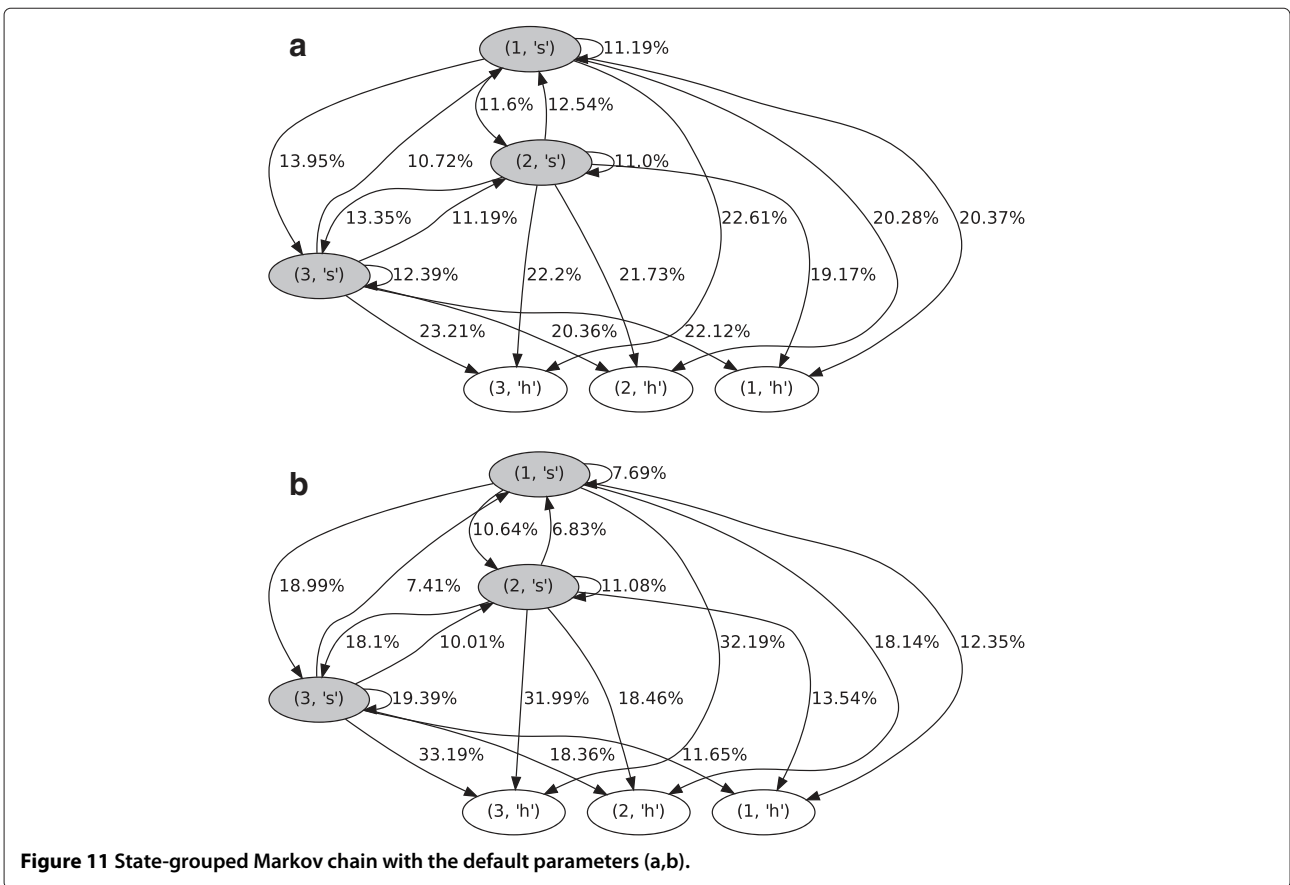
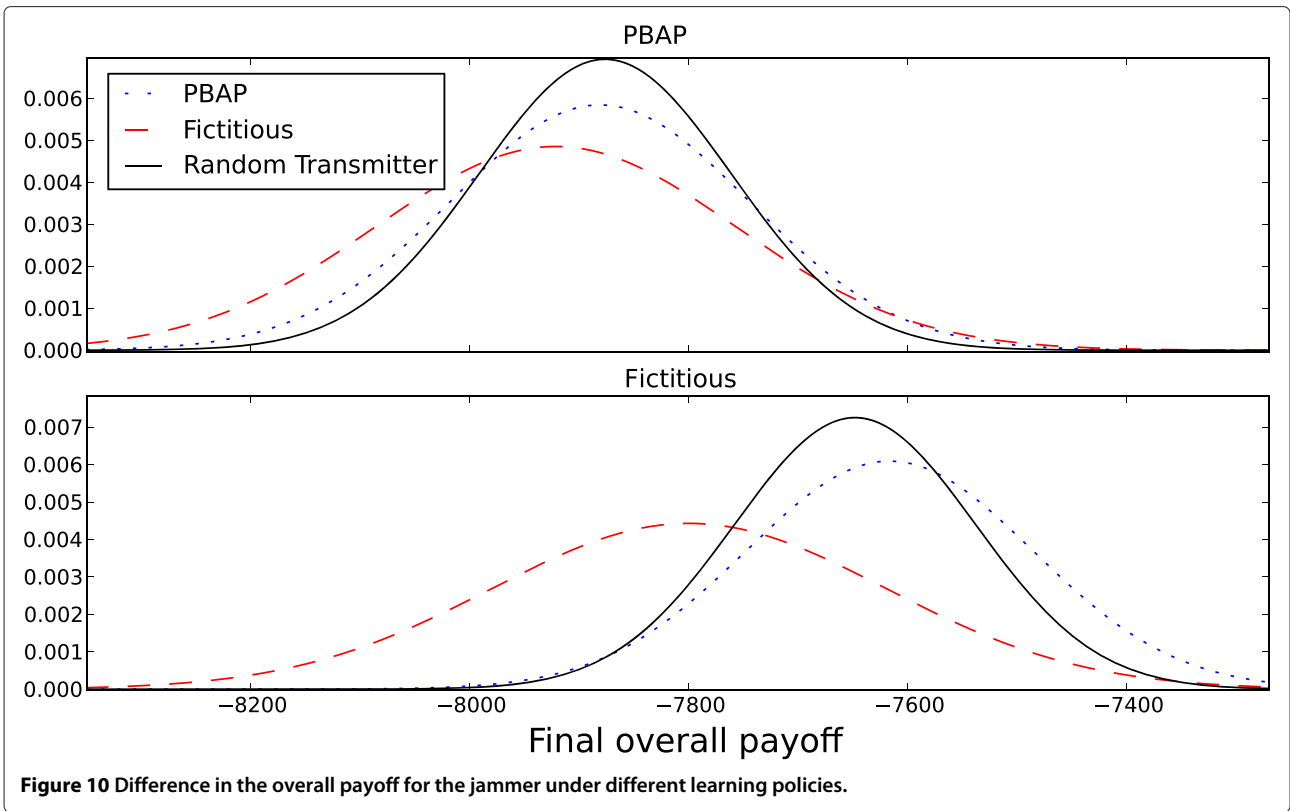


Figure 9 Difference in the overall payoff for the transmitter under different learning policies.



used to differentiate between the learning strategies of the opponent.

The results verify that the performance of the transmitter is very similar while using PBAP (top part) and fictitious play (bottom part). The exception is the case when the jammer employs fictitious learning. In this case, the transmitter will benefit slightly more by also deploying fictitious play in order to learn to infer the jammer's actions as soon as possible. The results for the jammer confirm our intuition - significantly better results for both cases are obtained using fictitious play.

Next, we aim to show how evolution of the game is influenced when the parameters of the game are modified. As explained previously, the state/action space of the players can be depicted by Markov chains, where each Markov state represents the current state of the player, and each edge the probability of taking an action leading to the new state. A graphical representation of the Markov transition probabilities is difficult to interpret for the full set of states of high-action-space games (higher than 2×2). Some examples of the full Markov chains for small action spaces may be found in [21]. This problem can partially be alleviated by creating state-grouped Markov chains, as shown in Figure 11a,b. Here, the number refers to the ordinal number of transmission power (i.e., '1' = -12 dBW for

the transmitter, '1' = 1 dBW for the jammer, etc.). Actions pertaining to frequency hopping are grouped and marked as 'h', while actions of staying on the same frequency are marked as 's'.

Then, the simulations are done for two extreme values of the hopping cost: 0.01 and 1.3, while keeping all other parameters the same. Figure 12a,b shows the differences in final stochastic distributions of the transmitter's strategies. As expected, an evident trend of the learning algorithm focuses on placing more importance on action 's' as the hopping cost increases.

Stochastic distributions of the mixed strategy Nash equilibrium for the transmitter and jammer under the default game parameters may also be shown in the form of the state-grouped Markov chains, as done in Figure 13a,b.

Finally, we perform the evaluation of the convergence to Nash equilibrium in terms of the overall payoff for the considered learning algorithms.

Figure 14 shows the convergence to Nash equilibrium in terms of payoff for fictitious play. Here, the red line shows the payoff obtained when both players are playing Nash equilibrium strategies. The blue line shows the case when the transmitter is playing the Nash strategy and jammer is deploying fictitious play. As can be seen for the jammer in the bottom part of the figure, fictitious play is

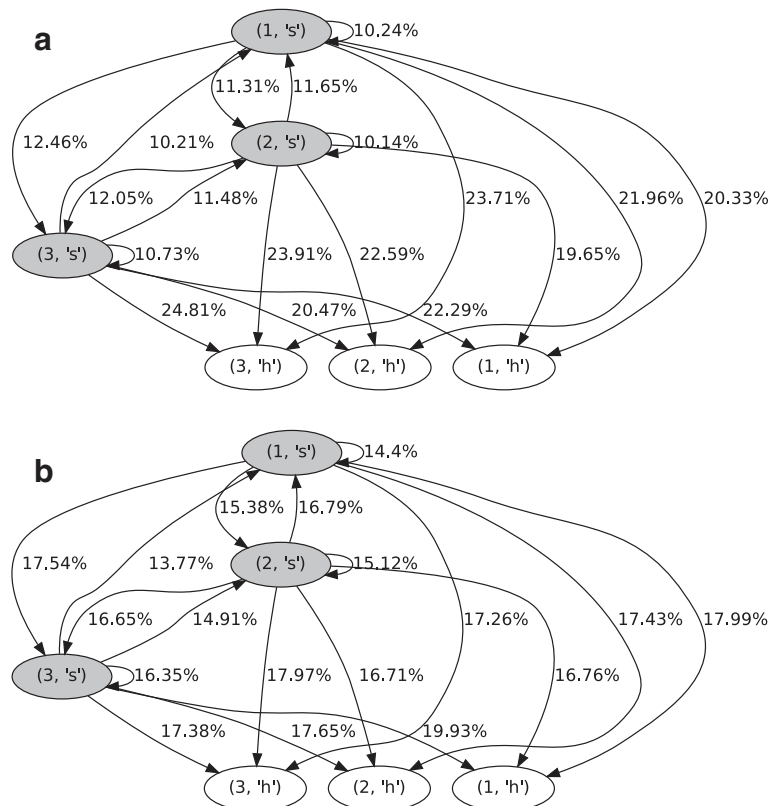


Figure 12 State-grouped Markov chain for different hopping cost (a,b).

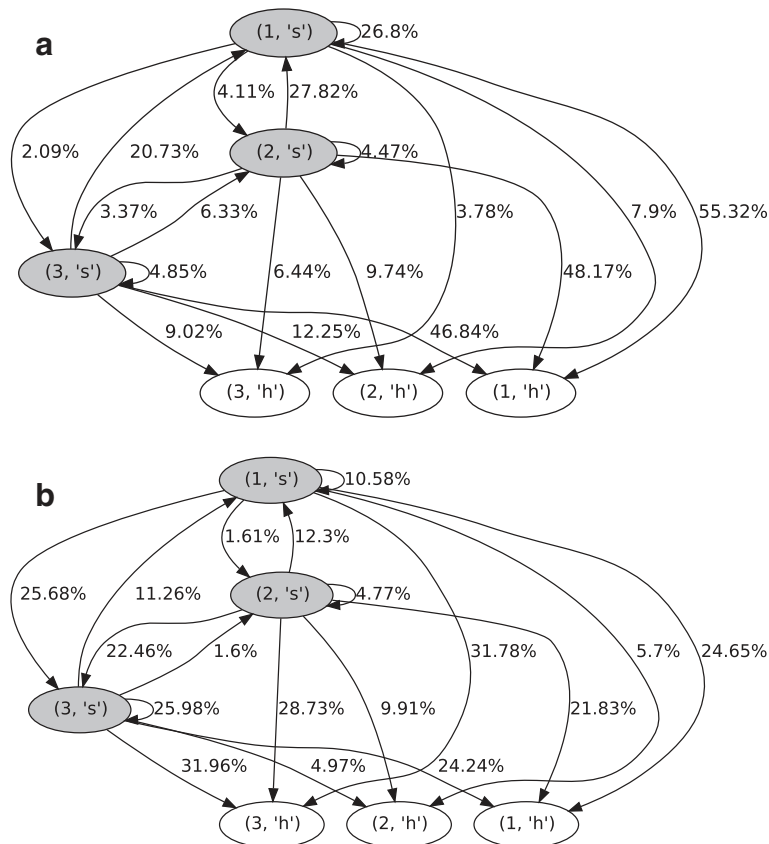


Figure 13 State-grouped Markov chain for the transmitter and jammer playing Nash equilibrium (a,b).

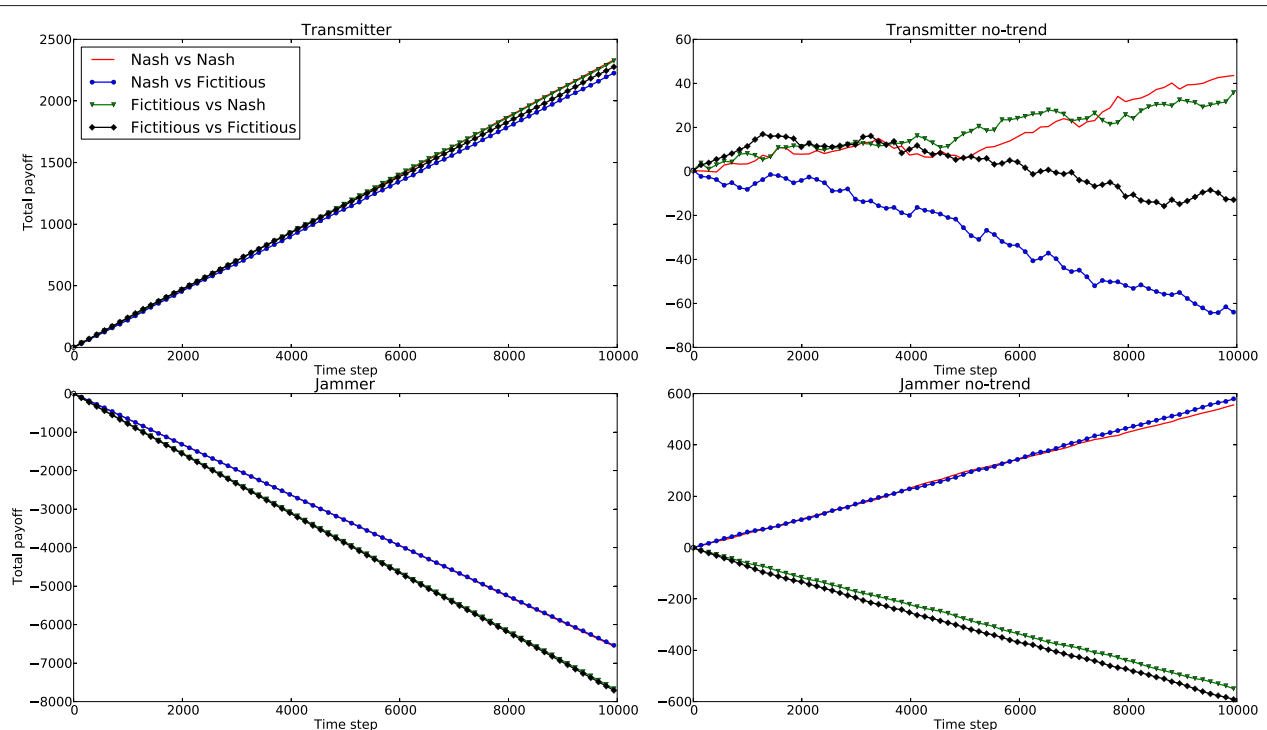


Figure 14 Comparison of fictitious play to Nash equilibrium strategy.

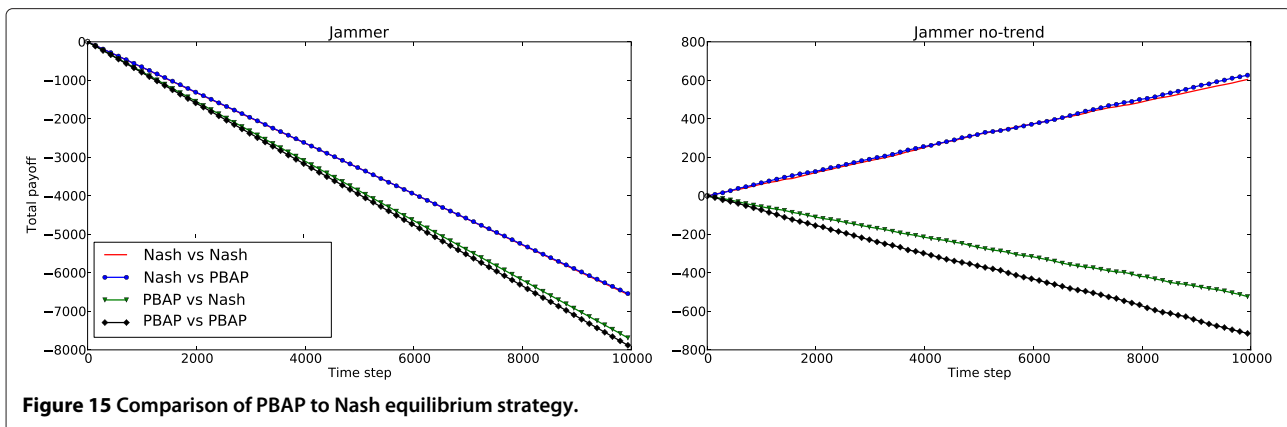


Figure 15 Comparison of PBAP to Nash equilibrium strategy.

able to obtain performance nearly as good as the strategy played in Nash equilibrium, when the opponent is playing according to Nash. Similar conclusions, although once again less prominent, may be drawn from the upper part of the figure for the transmitter playing fictitious play and jammer playing according to Nash equilibrium. The results are compared to the flow of the game when both players are playing according to fictitious play (black line). The results correspond to the results of [36] - fictitious play indeed seems to converge in the payoff to ϵ -equilibrium.

Similar results are obtained for the PBAP when faced against the Nash strategy. Figure 15 shows the convergence comparison for the jammer.

7 Conclusions

In the paper, a cognitive radio stochastic jamming/anti-jamming game between two players was modelled. Increased action space of the anti-jamming algorithm was created by combining power alteration and channel hopping. Two learning algorithms were considered: payoff-based adaptive learning corresponding to radios without spectrum sensing capabilities and fictitious play which may be utilized by the spectrum-sensing radios. In addition to their performance, their convergence properties to Nash equilibrium in terms of overall payoff and empirical distributions of the strategies were studied. In order to narrow the gap between the theoretical constraints inherent to game theory and practical aspects of the communication systems, relevant parameters for the game were inferred by performing a set of experiments using the real-life software-defined radio test bed. The major finding of the paper is the importance of the spectrum sensing endowment for the jamming side, compared to relatively insignificant benefits for the transmitting side in proactive anti-jamming games. In addition, evolution dynamics for different game parameters were presented.

Deployment of the feature detectors is a logical next step in the arms race between the narrowband jammers

and the anti-jamming systems. However, introduction of the additional parameters under the currently proposed framework would increase the action space to the point of infeasibility for analysis. For this purpose, future work will focus on finding ways for clusterizing overly complex action spaces and further optimizing their graphical representations by the means of state-grouped Markov chains.

In addition, once that the deployed SDR/CR test bed becomes endowed with the spectrum intelligence and automatic-reconfigurability capabilities, it will be used for testing and verification of the adaptation of the proposed game-theoretical schemes.

Endnotes

^aThis assumption may be built upon the existence of the threshold effect, characteristic for digital communication systems, where there is a certain SINR below which the BER significantly rises, and the communication systems perform poorly [33].

^bIn the cases of devices able to perform fast frequency hopping, and whose spectrum sensing and processing time is comparable to the frequency settling time, or in the games with smaller step sizes, this parameter would play a difference and should not be disregarded.

Abbreviations

AWGN: additive white Gaussian noise; BER: bit error rate; CR: cognitive radio; DSA: dynamic spectrum access; FP: fictitious play; GTO: game theory optimal; HH: handheld; LH: Lemke-Howson; MDP: Markov decision process; OSA: opportunistic spectrum access; PBAP: payoff-based adaptive play; PPAD: polynomial parity arguments on directed graphs; RF: radio frequency; SBW: soldier broadband waveform; SDR: software defined radio; SINR: signal to interference plus noise ratio; SNMP: Simple Network Management Protocol; UHF: ultra high frequency; VHF: very high frequency.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was partially developed within the nSHIELD project (<http://www.newshield.eu>) co-funded by the ARTEMIS Joint Undertaking (Sub-programme SP6) focused on the research of SPD (Security, Privacy, Dependability) in the

context of Embedded Systems. This work was supported in part by the Erasmus Mundus joint Doctorate in interactive and Cognitive Environments, which is funded by the EACE Agency of the European Commission under EMJD ICE. The authors would like to thank Selex ES and Sistemi Intelligenti Integrati Tecnologie (SIIT) for providing the equipment for the test bed and the laboratory premises for the test bed assembly. Particular acknowledgment goes to Virgilio Esposto of Selex ES, for providing expertise and technical assistance.

Received: 16 March 2014 Accepted: 17 November 2014
Published: 3 December 2014

References

1. J Mitola, GQ Maguire Jr, Cognitive radio: making software radios more personal. *IEEE Pers. Comm.* **6**(4), 1999. doi:10.1109/98.788210
2. Y Zeng, Y-C Liang, A Hoang, R Zhang, A review on spectrum sensing for cognitive radio: challenges and solutions. *EURASIP J. Adv Signal Process.* **2010**(1), 381465 (2010). doi:10.1155/2010/381465
3. MO Mughal, A Razi, SS Alam, L Marcenaro, CS Regazzoni, in *Next Generation Mobile Apps, Services and Technologies (NGMAST) 2013 Seventh International Conference On*. Analysis of energy detector in cooperative relay networks for cognitive radios (IEEE Prague, Czech Republic, 2013), pp. 220–225. doi:10.1109/NGMAST.2013.47
4. S Kapoor, SVRK Rao, G Singh, in *Communication Systems and Network Technologies (CSNT), 2011 International Conference on*. Opportunistic spectrum sensing by employing matched filter in cognitive radio network (IEEE Jammu, India, 580). doi:10.1109/CSNT.2011.124
5. L Yang, Z Chen, F Yin, Cyclo-energy detector for spectrum sensing in cognitive radio. *Int. J. Electron. Commun.* **66**(1), 89–92 (2012). doi:10.1016/j.aeue.2011.05.004
6. Z Khalaf, A Nafkha, J Palicot, M Ghazzi, in *Telecommunications (AICT), 2010 Sixth Advanced International Conference on*. Hybrid spectrum sensing architecture for cognitive radio equipment (IEEE Barcelona, Spain, 2010), pp. 46–51. doi:10.1109/AICT.2010.42
7. M Nekovee, in *Ultra Modern Telecommunications & Workshops, 2009. ICUMT'09. International Conference on*. A survey of cognitive radio access to TV White Spaces (Russia IEEE St. Petersburg, 2009), pp. 2070–2077. doi:10.1155/2010/236568
8. Z Lei, F Chin, A reliable and power efficient beacon structure for cognitive radio systems. *IEEE Trans. Broadcast.* **54**(2), 182–187 (2008). doi:10.1109/TBC.2008.917737
9. K Dabcevic, L Marcenaro, CS Regazzoni, in *Evolution of Cognitive Networks and Self-Adaptive Communication Systems*, ed. by TD Lagkas, P Sarigiannidis, M Louta, and P Chatzimisios. Security in cognitive radio networks, vol. 1, 1st edn. (IGI Global Hershey, 2013), pp. 301–333
10. J Blesa, E Romero, A Rozas, A Araujo, PUE attack detection in cwsns using anomaly detection techniques. *EURASIP J. Wireless Commun. Netw.* **2013**(1), 215 (2013). doi:10.1186/1687-1499-2013-215
11. P Anand, AS Rawat, C Hao, PK Varshney, in *Communication Systems and Networks (COMSNETS), 2010 Second International Conference On*. Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks (IEEE Bangalore, India, 2010), pp. 168–176. doi:10.1109/COMSNETS.2010.5432012
12. W Wang, H Li, YL Sun, Z Han, Securing collaborative spectrum sensing against untrustworthy secondary users in cognitive radio networks. *EURASIP J. Adv. Signal Process.* **2010**(1), 695750 (2010). doi:10.1155/2010/695750
13. J Zhu, Security-reliability trade-off for cognitive radio networks in the presence of eavesdropping attack. *EURASIP J. Adv. Signal Process.* **2013**(1), 169 (2013). doi:10.1186/1687-1499-2013-169
14. RD Pietro, G Oliveri, Jamming mitigation in cognitive radio networks. *IEEE Netw.* **27**(3), 10–15 (2013). doi:10.1109/MNET.2013.6523802
15. P Morerio, K Dabcevic, L Marcenaro, CS Regazzoni, in *Complexity in Engineering (COMPENG), 2012*. Distributed cognitive radio architecture with automatic frequency switching (IEEE Aachen, Germany, 2012), pp. 139–142. doi:10.1109/CompEng.2012.6242968
16. E Altman, K Avrachenkov, A Garnaev, in *Proceedings of the 1st EuroFGI International Conference on Network Control and Optimization, NET-COOP'07*. A jamming game in wireless networks with transmission cost (Springer Berlin, Heidelberg, 2007), pp. 1–12
17. B Wang, Y Wu, KJR Liu, TC Clancy, An anti-jamming stochastic game for cognitive radio networks. *IEEE J. Sel. Area Comm.* **29**(4), 877–889 (2011). doi:10.1109/JSAC.2011.110418
18. C Chen, M Song, C Xin, J Backens, A game-theoretical anti-jamming scheme for cognitive radio networks. *IEEE Netw.* **27**(3), 22–27 (2013). doi:10.1109/MNET.2013.6523804
19. N Buchbinder, L Lewin-Eytan, I Menache, J Naor, A Orda, Dynamic power allocation under arbitrary varying channels: an online approach. *IEEE/ACM Trans. Netw.* **20**(2), 477–487 (2012). doi:10.1109/TNET.2011.2170092
20. A Garnaev, Y Hayel, E Altman, in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt) 2012 10th International Symposium On*. A bayesian jamming game in an ofdm wireless network (IEEE Paderborn, Germany, 2012), pp. 41–48
21. K Dabcevic, A Betancourt, L Marcenaro, CS Regazzoni, in *Acoustics, Speech and Signal Processing (ICASSP) 2014 IEEE International Conference On*. A fictitious play-based game-theoretical approach to alleviating jamming attacks for cognitive radios (IEEE Florence, Italy, 2014), pp. 8158–8162. doi:10.1109/ICASSP.2014.6855191
22. JF Mertens, A Neyman, Stochastic games. *Int. J. Game Theor.* **10**(2), 53–66 (1981). doi:10.1007/BF01769259
23. J Nash, Non-cooperative games. *Ann. Math.* **54**(2), 286–295 (1951)
24. G Owen, *Game Theory*, 3rd edn. (Academic Press, San Diego, 1995)
25. LEJ Brouwer, Über abbildung von mannigfaltigkeiten. *Mathematische Annalen.* **71**(1), 97–115 (1911)
26. LS Shapley, in *Pivoting and Extension*. Mathematical Programming Studies, vol. 1., ed. by ML Balinski. A note on the Lemke-Howson algorithm (Springer Berlin Heidelberg, 1974), pp. 175–189. doi:10.1007/BFb0121248
27. RD McKelvey, AM McLennan, TL Turocy, Gambit: Software Tools for Game Theory, Version 13.1.2 (2014). <http://www.gambit-project.org>
28. BV Stengel. Chapter 45 Computing equilibria for two-person games. *Handbook of Game Theory with Economic Applications*, vol. 3 (Elsevier, 2002), pp. 1723–1759. doi:10.1016/S1574-0005(02)03008-4
29. B Von Stengel, New maximal numbers of equilibria in bimatrix games. *Discrete Comput. Geom.* **21**(4), 557–568 (1999)
30. I Milchtaich, T Ostrowski, On some saddle point matrices and applications to completely mixed equilibrium in bimatrix games. *Int. J. Math. Game Theory Algebra.* **18**(2), 65–72 (2008)
31. JS Shamma, G Arslan, Dynamic fictitious play, dynamic gradient play, and distributed convergence to nash equilibria. *IEEE Trans. Automat. Contr.* **50**(3), 312–327 (2005). doi:10.1109/TAC.2005.843878
32. RS Sutton, AG Barto, *Introduction to Reinforcement Learning*, 1st edn. (MIT Press, Cambridge, 1998)
33. R Poisel, *Modern Communications Jamming: Principles and Techniques*. Artech House intelligence and information operations series (Artech House, Norwood, 2011)
34. J Robinson, An iterative method of solving a game. *Ann. Math.* **54**(2), 296–301 (1951)
35. C Daskalakis, PW Goldberg, CH Papadimitriou, in *Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing STOC '06*. The complexity of computing a nash equilibrium (ACM New York, 2006), pp. 71–78. doi:10.1145/1132516.1132527
36. V Conitzer, in *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing Allerton'09*. Approximation guarantees for fictitious play (IEEE Press Piscataway, 2009), pp. 636–643
37. G Ostrovski, S van Strien, Payoff performance of fictitious play. *CoRR. abs/1308.4049* (2013)
38. R Cominetti, E Melo, S Sorin, A payoff-based learning procedure and its application to traffic games. *Game Econ. Behav.* **70**(1), 71–83 (2010). doi:10.1016/j.geb.2008.11.012. Special Issue In Honor of Ehud Kalai
39. C Daskalakis, R Frongillo, CH Papadimitriou, G Pierrakos, G Valiant, in *Proceedings of the Third International Conference on Algorithmic Game Theory SAGT'10*. On learning algorithms for nash equilibria (Springer Berlin, Heidelberg, 2010), pp. 114–125
40. B Banerjee, J Peng, Efficient no-regret multiagent learning. *Proceedings of the 20th National Conference on Artificial Intelligence - Volume 1. AAAI'05*, 41–46 (2005)
41. K Wang, Q Liu, L Chen, Optimality of greedy policy for a class of standard reward function of restless multi-armed bandit problem. *IET Signal Process.* **6**(6), 584–593 (2012). doi:10.1049/iet-spr.2011.0185

42. M Tokic, in *KI 2010: Advances in Artificial Intelligence*, ed. by R Dillmann, J Beyerer, U Hanebeck, and T Schultz. Adaptive ϵ -greedy exploration in reinforcement learning based on value differences, vol. 6359 (Springer Karlsruhe, 2010), pp. 203–210. Lecture Notes in Computer Science
43. K Dabcevic, L Marcenaro, CS Regazzoni, in *Proceedings of the 4th International Conference on Pervasive and Embedded Computing and Communication Systems*. Spd-driven smart transmission layer based on a software defined radio test bed architecture (SciTePress Lisbon, Portugal, 2014), pp. 219–230
44. SelexES, SWAVE HH - Secure Wideband Multi-role - Single-Channel Handheld Radio (2013). <http://www.selexelsag.com/internet/localization/IPC/media/docs/SWave-Handheld-Radio-v1-2012Selex.pdf>. Accessed: 2014-07-31
45. K Dabcevic, MO Mughal, L Marcenaro, CS Regazzoni, in *Proceedings of SDR'14 - WinnComm-Europe*. Spectrum intelligence for interference mitigation for cognitive radio terminals (Wireless Innovation Forum Rome, Italy, 2014)
46. D Cabric, A Tkachenko, RW Brodersen, in *Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum TAPAS '06*. Experimental study of spectrum sensing based on energy detection and network cooperation (ACM New York, 2006). doi:10.1145/1234388.1234400

doi:10.1186/1687-6180-2014-171

Cite this article as: Dabcevic et al.: Intelligent cognitive radio jamming - a game-theoretical approach. *EURASIP Journal on Advances in Signal Processing* 2014 **2014**:171.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
