A Novel Cognitive Anti-jamming Stochastic Game

Mohamed Aref and Sudharman K. Jayaweera
Communication and Information Sciences Laboratory (CISL)
ECE, University of New Mexico, Albuquerque, NM
and
Bluecom Systems & Consulting LLC, Albuquerque, NM
Outline

1. Introduction
2. Problem formulation
3. System model
4. Q-learning-aided cognitive anti-jamming algorithm
5. Proposed anti-jamming stochastic game
6. Simulation results
Cognitive radio as an evolution of software-defined radio (SDR)

- A cognitive radio is a multiband, multimode, wideband software-defined radio (SDR) with autonomous decision-making and learning abilities that can optimally reconfigure its operation mode in response to its surrounding RF environment and user needs.
Wideband Autonomous Cognitive Radios (WACR)

- Senses a wide frequency range.
- Comprehend its operating RF environment.
- Autonomous operation.
- Learn communication protocols and policies.

Introduction
Introduction

Dynamic spectrum sharing (DSS)

Space

Frequency

Time

Holes

Vehicular networks

Cognitive Radio Applications

Health care

Military

Smart grid

Source: http://mil-embedded.com/articles/evolving-technology-sdr-cognitive-radio/
Basic Cognitive Radio Functions

Hardware constraints limit the instantaneous sensing bandwidth of most state-of-the-art software-defined radio (SDR) platforms to about 100MHz.

There is a need to design an efficient scheme to achieve real-time sensing over a wide spectrum range.
Wideband Spectrum Knowledge Acquisition

Spectrum Knowledge Acquisition

- Wideband spectrum scanning
- Spectral activity detection
- Signal classification

- Detection can be done by defining a threshold (simple).
- Any power spectral activity above this threshold is considered as an active signal.

**Chart:**
- Power spectrum density vs. Frequency
- Dashed line represents the threshold.
Signal classification

• Detected signals may belong to different radio systems.

Classification

Wi-Fi  Bluetooth  Mobile  ...  Others
Problem formulation

• Deliberate radio jammers and unintentional interference can disrupt communication systems.
  – In both commercial and military systems
Problem formulation

- In practice, this will result in a complicated multi-agent environment.

**Goal:** find optimal anti-jamming and interference avoidance policies for the WACRs that switches transmission **before** getting jammed.
System model

- Spectrum is divided into $N_b$ sub-bands.

Sub-band dynamics:

- Single sub-band has 2 Markov states: available/not-available.
  - If the sub-band is jammed or faces interference, it is considered to be in state “0” (not-available).
  - Otherwise, it is considered to be in state “1” (available).
- The set of sub-band states can be denoted by $\mathcal{V} = \{0, 1\}$. 
• Each operation will have its own learning algorithm with different targets, but they both will experience the same RF environment.

• Essentially, if the sensing operation were to learn an optimal policy, the WACR would be able to accurately predict the jammed/interfered sub-bands.

• This will help the transmission operation as follows:
  • if the current operating sub-band is predicted to be jammed during the next time instant by the sensing policy, the WACR will switch to another sub-band thereby avoiding the possibility of getting jammed.
System model

- For the game state, we choose a simple definition for both sensing and transmission operations, where $s_s[n] \in S$ and $s_t[n] \in S$ represent the index of selected sub-bands for sensing and transmission, respectively, at time $n$. Thus, the state space is given by $S = \{1, \cdots, N_b\}$.

- At any time instant, the state of operating sub-bands for both sensing and transmission (the value of $v \in V$ for sub-band index $s \in S$) has to be identified.
  - During sensing operation: the WARC will perform spectral activity detection (spectrum sensing) to detect any active signals in the sensed sub-band and hence identify whether the sub-band is available or not.
  - During transmission operation: the communications link quality will determine if transmission over the current operating sub-band is acceptable.

- After determining the states of both operating sub-bands, the WACR will select and execute actions for both operations.
  - We define actions $a_s[n]$ and $a_t[n]$ as the indices of the selected new operating sub-bands for sensing and transmission, respectively, at time $n$.

- The action space can thus be defined as $A = \{1, \cdots, N_b\}$.
Q-learning-aided Cognitive Anti-jamming

Algorithm 1 Q-learning-aided cognitive anti-jamming communications algorithm

1: **Initialize:**
   \[ \alpha, \gamma, \epsilon \in [0, 1] \]
   \[ Q(s, a) \leftarrow 0 \quad \forall s \in S, \forall a \in A \]
2: **for** each stage \( n \) **do**
3:   Identify the state \( (v \in \mathcal{V}) \) of operating sub-band \( s \)
4:   **if** sub-band state \( v = 0 \) **then**
5:     Compute reward \( r \) for current state \( s \) and action \( a \)
6:     Update Q-value \( Q(s, a) \) as follow:
7:     \[ Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha \left[ r + \gamma \max_{a'} Q(s', a') \right] \]
8:   **Select** new action \( a' \in \mathcal{A} \) for the new state \( s' \)
   **according** to the following:
9:   \[ a' = \begin{cases} \arg \max_{a \in \mathcal{A}} Q(s', a) & \text{with probability } 1 - \epsilon, \\ \sim \mathcal{U}(\mathcal{A}) & \text{with probability } \epsilon, \end{cases} \]

- Learning parameters and Q-table initialization.
Q-learning-aided Cognitive Anti-jamming Algorithm

Algorithm 1 Q-learning-aided cognitive anti-jamming communications algorithm

1: Initialize:
   \[ \alpha, \gamma, \epsilon \in [0, 1] \]
   \[ Q(s, a) \leftarrow 0 \ \forall s \in S, \forall a \in A \]
2: for each stage \( n \) do
3:   Identify the state \( (v \in \mathcal{V}) \) of operating sub-band \( s \)
4:   if sub-band state \( v = 0 \) then
5:     Compute reward \( r \) for current state \( s \) and action \( a \)
6:     Update Q-value \( Q(s, a) \) as follow:
7:     \[ Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha[r + \gamma \max_{a'}Q(s', a')] \]
8:     Select new action \( a' \in \mathcal{A} \) for the new state \( s' \)
9: according to the following:
   \[ a' = \begin{cases} 
   \arg \max_{a \in \mathcal{A}} Q(s', a) & \text{with probability } 1 - \epsilon, \\
   \sim \mathcal{U}(\mathcal{A}) & \text{with probability } \epsilon, 
\end{cases} \]

- Identify the state of the current operating sub-band.
- If the sub-band state is “1” (available), no further action is required.
Q-learning-aided Cognitive Anti-jamming Algorithm

Algorithm 1 Q-learning-aided cognitive anti-jamming communications algorithm

1: Initialize:
   \( \alpha, \gamma, \epsilon \in [0, 1] \)
   \( Q(s, a) \leftarrow 0 \ \forall s \in S, \ \forall a \in A \)
2: for each stage \( n \) do
3:   Identify the state \( (v \in \mathcal{V}) \) of operating sub-band \( s \)
4:   if sub-band state \( v = 0 \) then
5:     Compute reward \( r \) for current state \( s \) and action \( a \)
6:     Update Q-value \( Q(s, a) \) as follow:
7:   \( Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha \left[ r + \gamma \max_a Q(s', a) \right] \)
8:   Select new action \( a' \in A \) for the new state \( s' \)
   according to the following:
9:   \( a' = \begin{cases} 
   \arg \max_{a \in A} Q(s', a) & \text{with probability } 1 - \epsilon, \\
   \sim U(A) & \text{with probability } \epsilon,
\end{cases} \)

- If the sub-band state is “0” (not-available), the WACR updates the Q-table based on a certain observed reward \( (r) \).
Algorithm 1 Q-learning-aided cognitive anti-jamming communications algorithm

1: Initialize:
   \[ \alpha, \gamma, \epsilon \in [0, 1] \]
   \[ Q(s, a) \leftarrow 0 \quad \forall s \in S, \forall a \in A \]
2: for each stage \( n \) do
3:   Identify the state \( (v \in \mathcal{V}) \) of operating sub-band \( s \)
4:   if sub-band state \( v = 0 \) then
5:     Compute reward \( r \) for current state \( s \) and action \( a \)
6:     Update Q-value \( Q(s, a) \) as follow:
7:     \[ Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha[r + \gamma \max_{a'}Q(s', a')] \]
8:   Select new action \( a' \in \mathcal{A} \) for the new state \( s' \)
   according to the following:
9:   \[ a' = \begin{cases} \arg \max_{a \in \mathcal{A}} Q(s', a) & \text{with probability } 1 - \epsilon, \\ \sim U(\mathcal{A}) & \text{with probability } \epsilon, \end{cases} \]

- Once the Q-table is updated, the WACR selects a new action \( a' \) representing the new operating sub-band.
Proposed Anti-jamming Stochastic Game

1. Select sub-band for sensing $a_s$
2. Sense sub-band $a_s$
3. Check if sub-band $a_s$ is available
   - Yes
     - Reward $r_s = -T_s$
     - Update the Q-table for sensing
   - No

Proposed Anti-jamming Stochastic Game

### Sensing

- Select sub-band for sensing $a_s$
- Sense sub-band $a_s$
- Sub-band $a_s$ available

**Yes**
- Reward $r_s = -T_s$
- Update the Q-table for sensing

**No**

### Transmission

- Select sub-band for transmission $a_t$
- Transmission interrupt

**Yes**
- Reward $r_t = -1$
- Update the Q-table for transmission

**No**

- $a_t = a_s$

**No**

- $T_t \Rightarrow T_{max}$

**Yes**
- Reward $r_t = T_t$
Simulation results

Performance metric:

Normalized accumulated reward

\[ R_N = \frac{1}{N} \sum_{n=1}^{N} r_t(s_t[n], a_t[n]) \]

\( r_t(s_t[n], a_t[n]) \): immediate non-negative reward for transmission operation at time \( n \)

\( N \): number of iterations

Jammer model:

Sweeps the spectrum of interest from the lower to the higher frequency.

Learning parameters:

\( \gamma = 0.8 \)
\( \epsilon = 0.9, \alpha = 0.4 \) \hspace{1cm} \text{Before Q-table convergence}
\( \epsilon = 0.01, \alpha = 0.1 \) \hspace{1cm} \text{After Q-table convergence}
Simulation results

Experiment 1: 1 WACR and 5 Sub-bands
Simulation results

**Experiment 2:** 2 WACRs and 6 Sub-bands

![Graph showing simulation results](graph.png)
Simulation results

**Experiment 3:** 4 WACRs and 16 Sub-bands
Simulation results

Table I
NORMALIZED ACCUMULATED REWARD VALUES FOR DIFFERENT SIMULATION SCENARIOS

<table>
<thead>
<tr>
<th>Test case</th>
<th>Scenario</th>
<th>Reward upper bound</th>
<th>WACR 1</th>
<th>WACR 2</th>
<th>WACR 3</th>
<th>WACR 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 WACR and 5 sub-bands</td>
<td>4</td>
<td>Proposed: 3.8</td>
<td>Proposed: 2.5</td>
<td></td>
<td></td>
<td>Proposed: 3.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Random: 2.5</td>
<td></td>
<td></td>
<td></td>
<td>Random: 2.5</td>
</tr>
<tr>
<td>2</td>
<td>2 WACRs and 6 sub-bands</td>
<td>4</td>
<td>Proposed: 2.8</td>
<td>Proposed: 3</td>
<td></td>
<td></td>
<td>Proposed: 2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Random: 1.5</td>
<td>Random: 1.4</td>
<td></td>
<td></td>
<td>Random: 1.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Random: 2.5</td>
<td>Random: 2.2</td>
<td>Random: 2.2</td>
<td>Random: 1.8</td>
<td>Random: 2.17</td>
</tr>
</tbody>
</table>
# Simulation results

## Table II

**Probabilities of getting jammed for different simulation scenarios**

<table>
<thead>
<tr>
<th>Test case</th>
<th>Scenario</th>
<th>WACR 1</th>
<th>WACR 2</th>
<th>WACR 3</th>
<th>WACR 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 WACR and 5 sub-bands</td>
<td>Proposed: 0.86%</td>
<td></td>
<td></td>
<td></td>
<td>Proposed: 0.86%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Random: 1.8%</td>
<td></td>
<td></td>
<td></td>
<td>Random: 1.8%</td>
</tr>
<tr>
<td>2</td>
<td>2 WACRs and 6 sub-bands</td>
<td>Proposed: 2.6%</td>
<td>Proposed: 2.1%</td>
<td></td>
<td></td>
<td>Proposed: 2.35%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Random: 47.2%</td>
<td>Random: 48%</td>
<td></td>
<td></td>
<td>Random: 47.6%</td>
</tr>
<tr>
<td>3</td>
<td>4 WACR and 16 sub-bands</td>
<td>Proposed: 6.4%</td>
<td>Proposed: 7.6%</td>
<td>Proposed: 12.4%</td>
<td>Proposed: 12.3%</td>
<td>Proposed: 9.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Random: 64.8%</td>
<td>Random: 66.3%</td>
<td>Random: 66.3%</td>
<td>Random: 72.6%</td>
<td>Random: 67.5%</td>
</tr>
</tbody>
</table>
Conclusions

• Proposed a novel cognitive anti-jamming stochastic game based on Q-learning for WACRs to avoid a dynamic jammer signal as well as unintentional interference from other WACRs.

• Developed new definitions for state, actions and rewards that enable the WACR to switch its operating sub-band before getting jammed, compared to previously proposed anti-jamming techniques in literature that switch the operating sub-band only after getting jammed.

• The cognitive framework is divided into two operations:
  – sensing and transmission.
  – Each is helped by its own learning algorithm based on Q-learning.

• The objective of the sensing operation is to track the jammed sub-bands. On the other hand, the transmission operation determines when and where to switch the operating sub-band.
  – The key difference from the previous work is that the radio will switch the sub-band before getting jammed.
  – This can be especially useful against a smart jammer since it will prevent the jammer from learning the radio’s behavior.

• Simulation results showed that the proposed cognitive protocol has a very low probability of getting jammed and acceptable value for accumulated reward.
Questions
References


