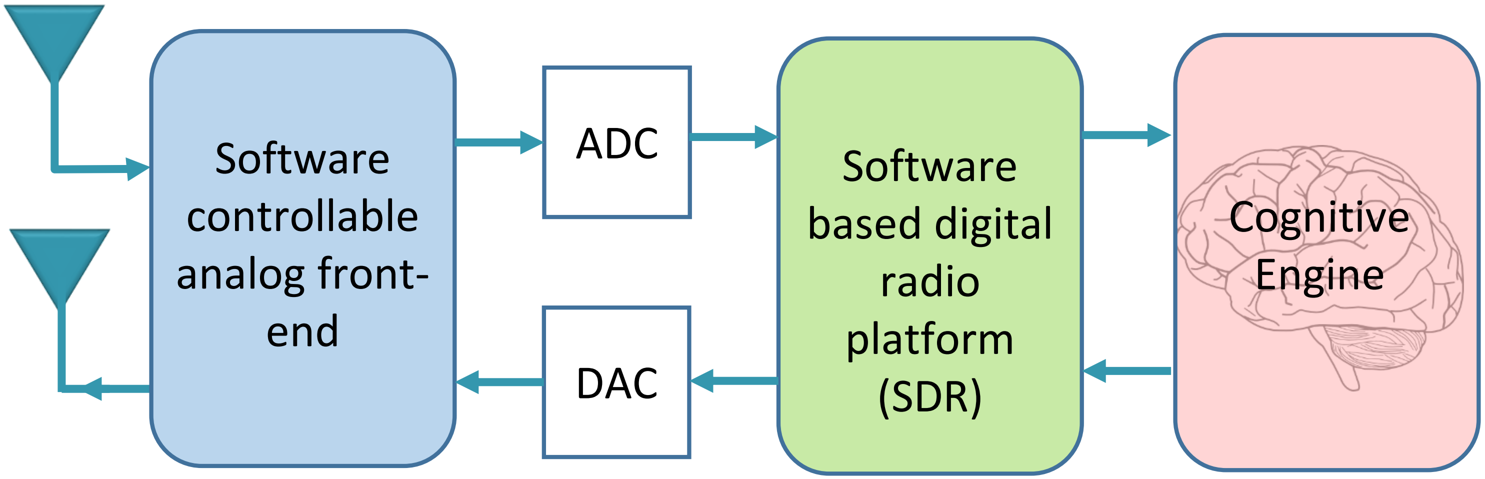
A Novel Cognitive Anti-jamming Stochastic Game

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

Introduction

Cognitive radio as an evolution of software-defined radio (SDR)

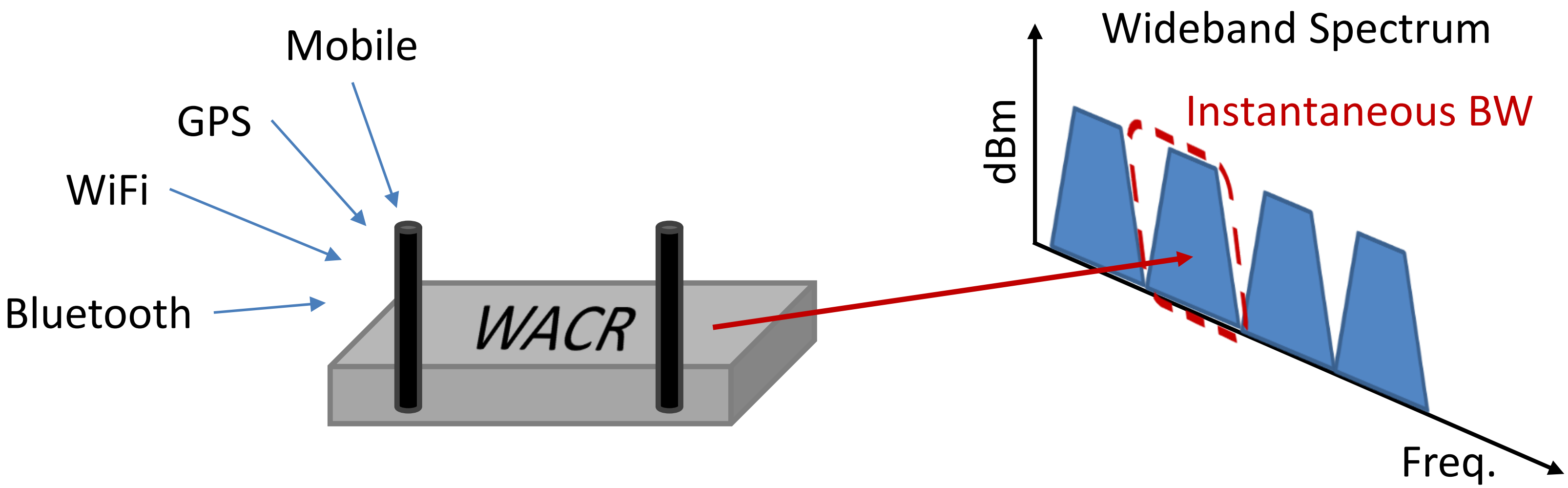


- A cognitive radio is a multiband, multimode, wideband software-defined radio (SDR) with autonomous decision-making and learning abilities that can optimally reconfigure its operation mode in response to its surrounding RF environment and user needs.

Introduction

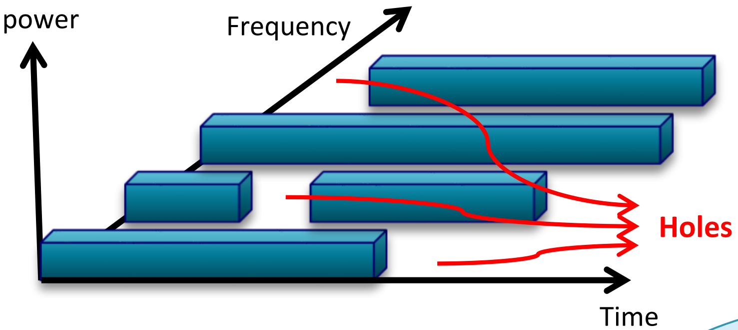
Wideband Autonomous Cognitive Radios (WACR)

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

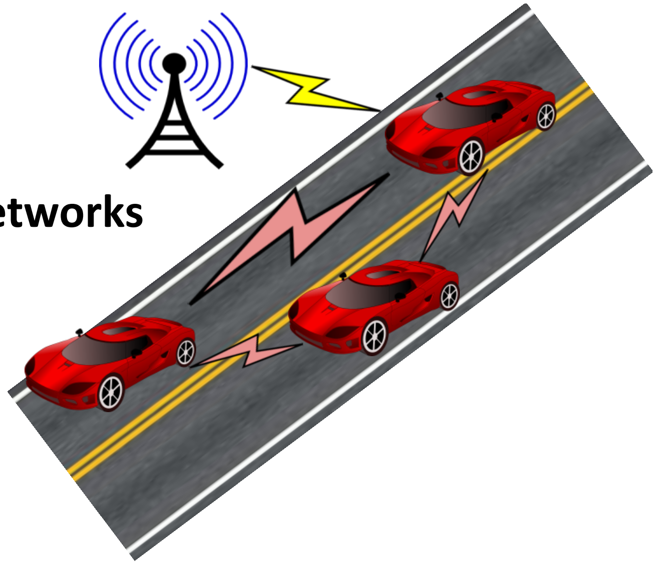


Introduction

Dynamic spectrum sharing (DSS)

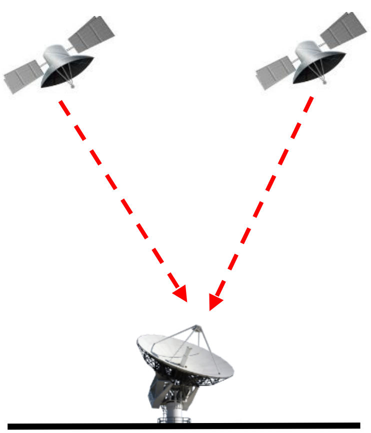


Vehicular networks



Cognitive Radio Applications

Space



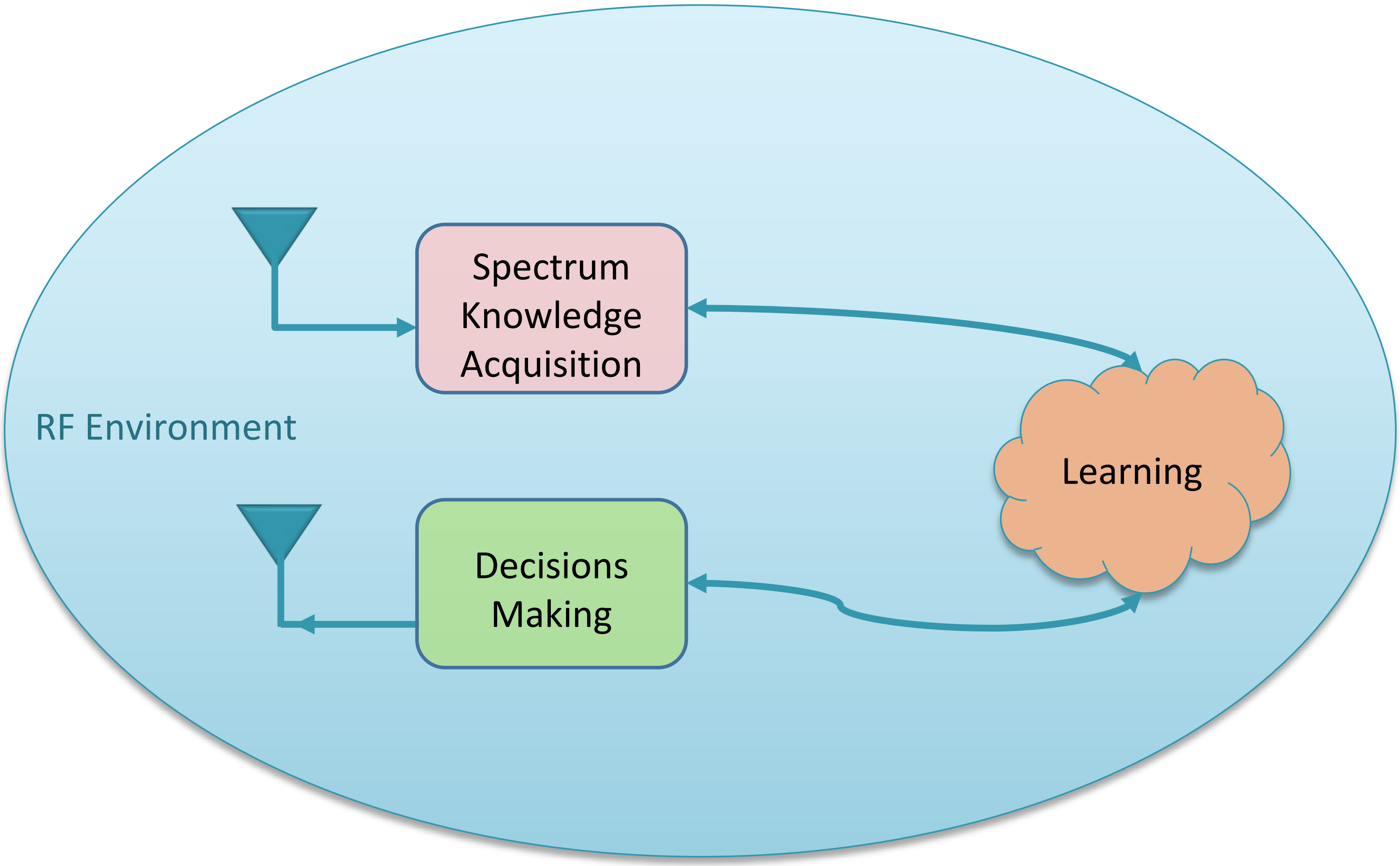
Health care

Smart grid

Military

Source: http://mil-embedded.com/articles/evolving-technology-sdr-cognitive-radio/

Basic Cognitive Radio Functions



Source: S. K. Jayaweera,"Signal Processing for Cognitive Radio," John Wiley & Sons, Hoboken, NJ, USA.

Wideband Spectrum Knowledge Acquisition

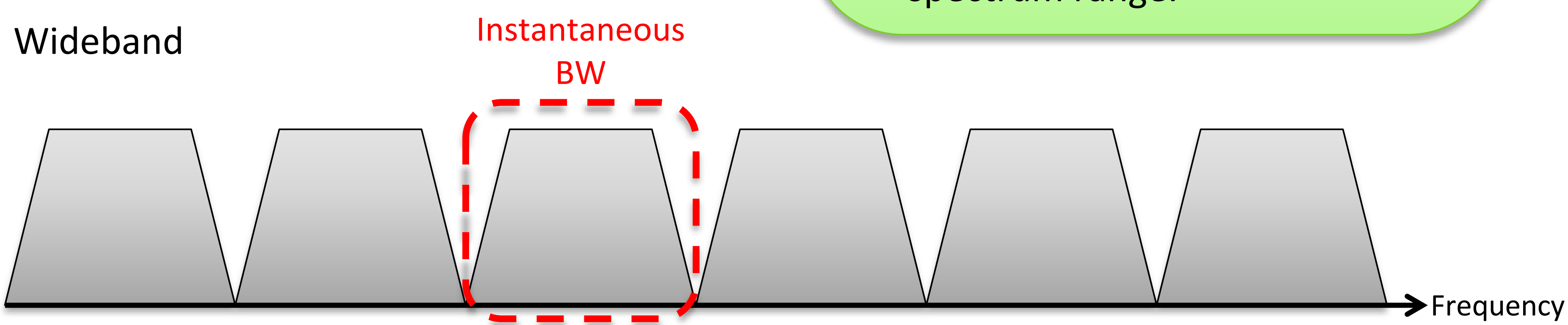
Spectrum Knowledge Acquisition

Wideband spectrum scanning

Spectral activity detection

Signal classification

- 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

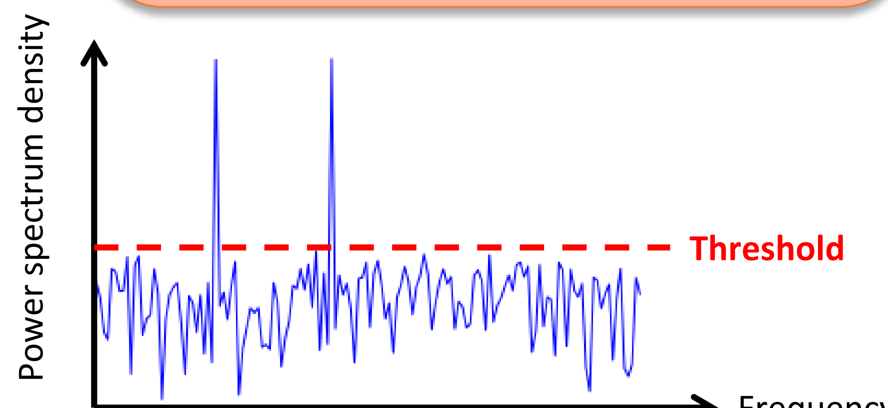
Spectrum Knowledge Acquisition

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.



Wideband Spectrum Knowledge Acquisition

Spectrum Knowledge Acquisition

Wideband spectrum scanning

↓

Spectral activity detection

↓

Signal classification →

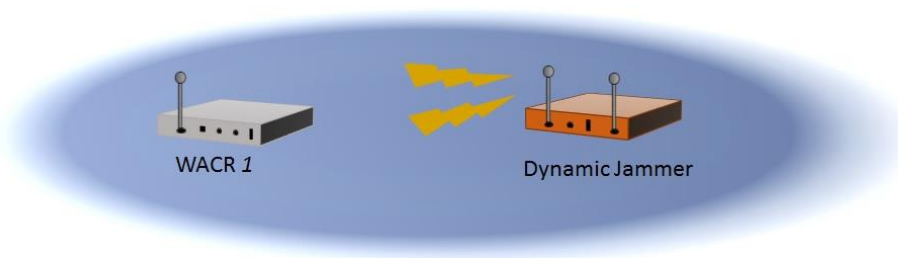
- Detected signals may belong to different radio systems.

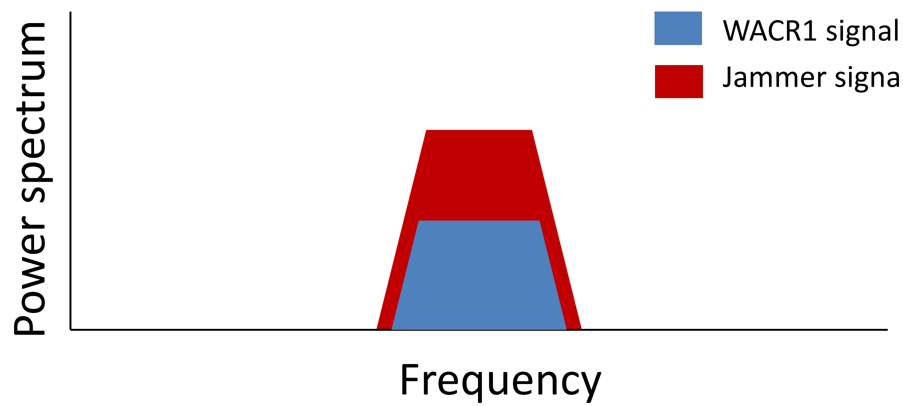
Classification

Wi-Fi | Bluetooth | Mobile | … | Others

Problem formulation

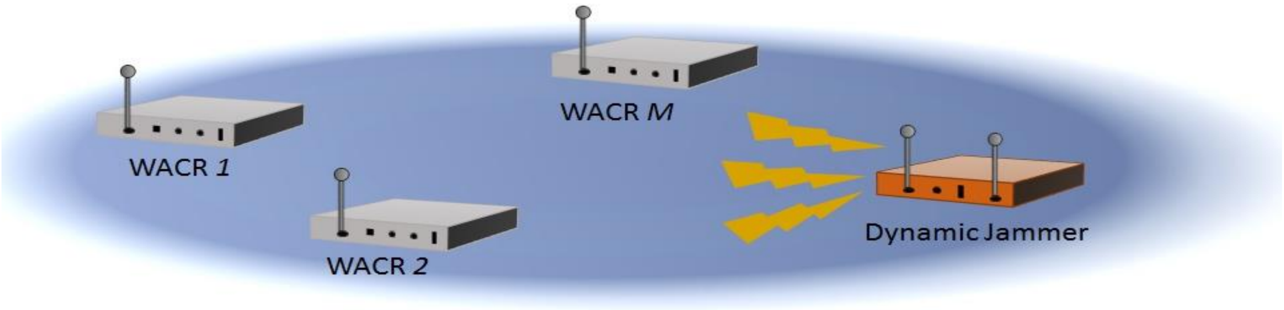
- Deliberate radio jammers and unintentional interference can disrupt communication systems.
  - In both commercial and military systems

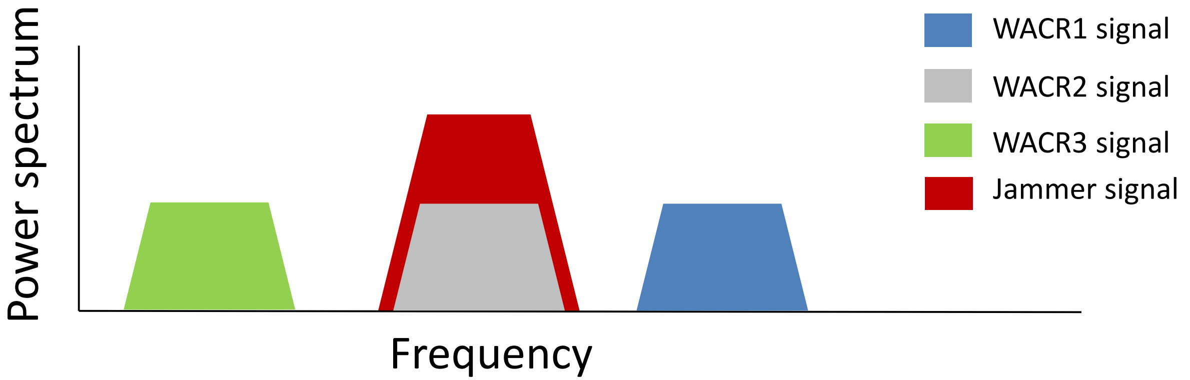




Problem formulation

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

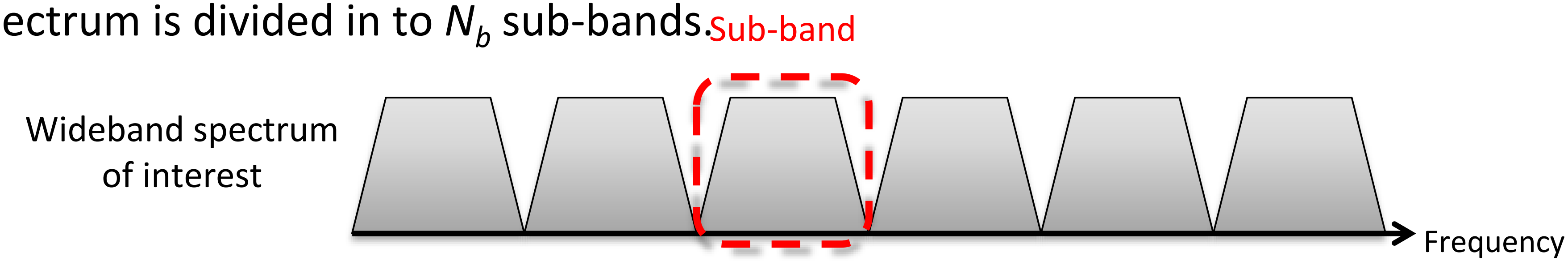




- Goal: find optimal anti-jamming and interference avoidance policies for the WACRs that switches transmission before getting jammed.

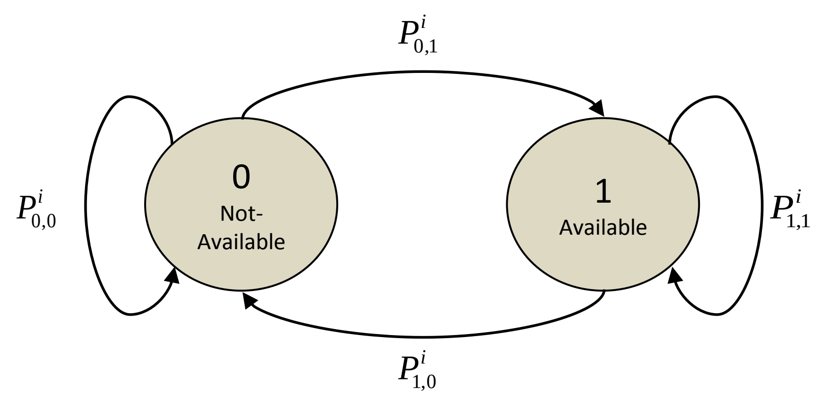
System model

- Spectrum is divided in to N_b sub-bands.

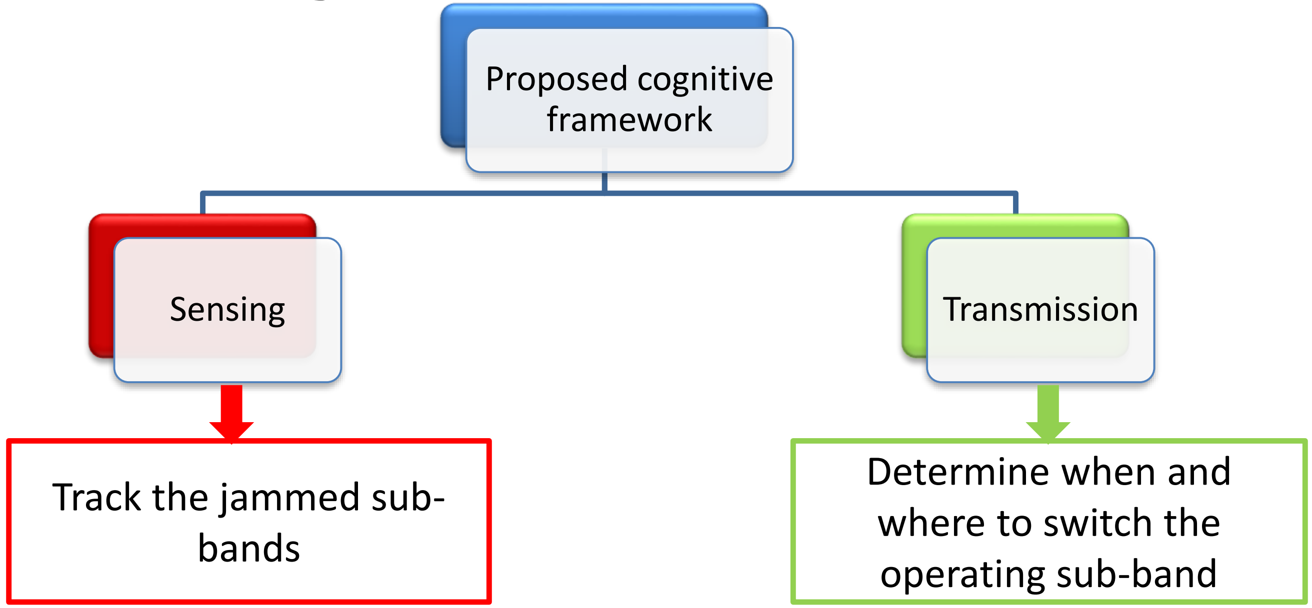


Sub-band dynamics:

- 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 $\mathcal{V} = \{0, 1\}$.



System model



- 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 $s_s[n] \in \mathcal{S}$ and $s_t[n] \in \mathcal{S}$ represent the index of selected sub-bands for sensing and transmission, respectively, at time *n*. Thus, the state space is given by $\mathcal{S} = \{1, \cdots, N_b\}$.

- At any time instant, the state of operating sub-bands for both sensing and transmission (the value of $v \in \mathcal{V}$ for sub-band index $s \in \mathcal{S}$) 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 $\mathcal{A} = \{1, \cdots, N_b\}$.

Q-learning-aided Cognitive Anti-jamming

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

1: **Initialize:**
$\alpha, \gamma, \epsilon \in [0,1]$
$Q(s,a) \leftarrow 0 \; \forall s \in \mathcal{S}, \forall a \in \mathcal{A}$
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: $$a' = \begin{cases} \arg\max_{a \in \mathcal{A}} Q(s',a) & \text{with probability } 1-\epsilon, \\ \sim U(\mathcal{A}) & \text{with probability } \epsilon, \end{cases}$$

- Learning parameters and Q-table initialization.

Q-learning-aided Cognitive Anti-jamming Algorithm

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

Initialize:
  $\alpha,\ \gamma,\ \epsilon \in [0, 1]$
  $Q(s, a) \leftarrow 0\ \forall s \in \mathcal{S},\ \forall a \in \mathcal{A}$
for each stage n **do**
  Identify the state $(v \in \mathcal{V})$ of operating sub-band s
  if sub-band state $v = 0$ **then**
    Compute reward r for current state s and action a
    Update Q-value $Q(s, a)$ as follow:
    $Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha[r + \gamma \max_{a} Q(s', a)]$
    Select new action $a' \in \mathcal{A}$ for the new state s' according to the following:

$$a' = \begin{cases} \arg\max_{a \in \mathcal{A}} Q(s', a) & \text{with probability } 1 - \epsilon, \\ \sim U(\mathcal{A}) & \text{with probability } \epsilon, \end{cases}$$

- 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:**

α, γ, $\epsilon \in [0, 1]$

$Q(s, a) \leftarrow 0\ \forall s \in \mathcal{S}, \forall a \in \mathcal{A}$

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: $$a' = \begin{cases} \arg\max_{a \in \mathcal{A}} Q(s', a) & \text{with probability } 1 - \epsilon, \\ \sim U(\mathcal{A}) & \text{with probability } \epsilon, \end{cases}$$

- If the sub-band state is “0” (not-available), the WACR updates the Q-table based on a certain observed reward (r).

Q-learning-aided Cognitive Anti-jamming Algorithm

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

1: **Initialize:**

$\alpha, \gamma, \epsilon \in [0, 1]$

$Q(s, a) \leftarrow 0 \ \forall s \in \mathcal{S}, \forall a \in \mathcal{A}$

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:

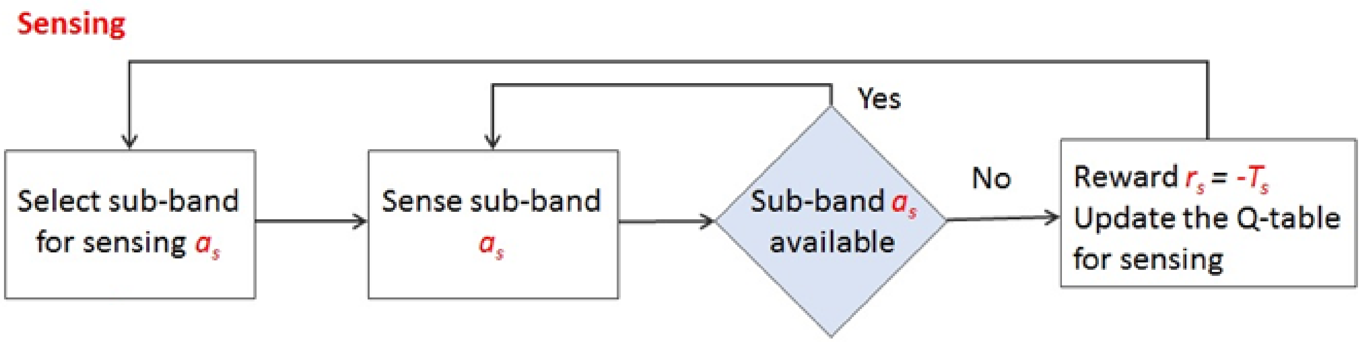
7: $Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha[r + \gamma \max_a Q(s', a)]$

8: Select new action $a' \in \mathcal{A}$ for the new state s' according to the following:

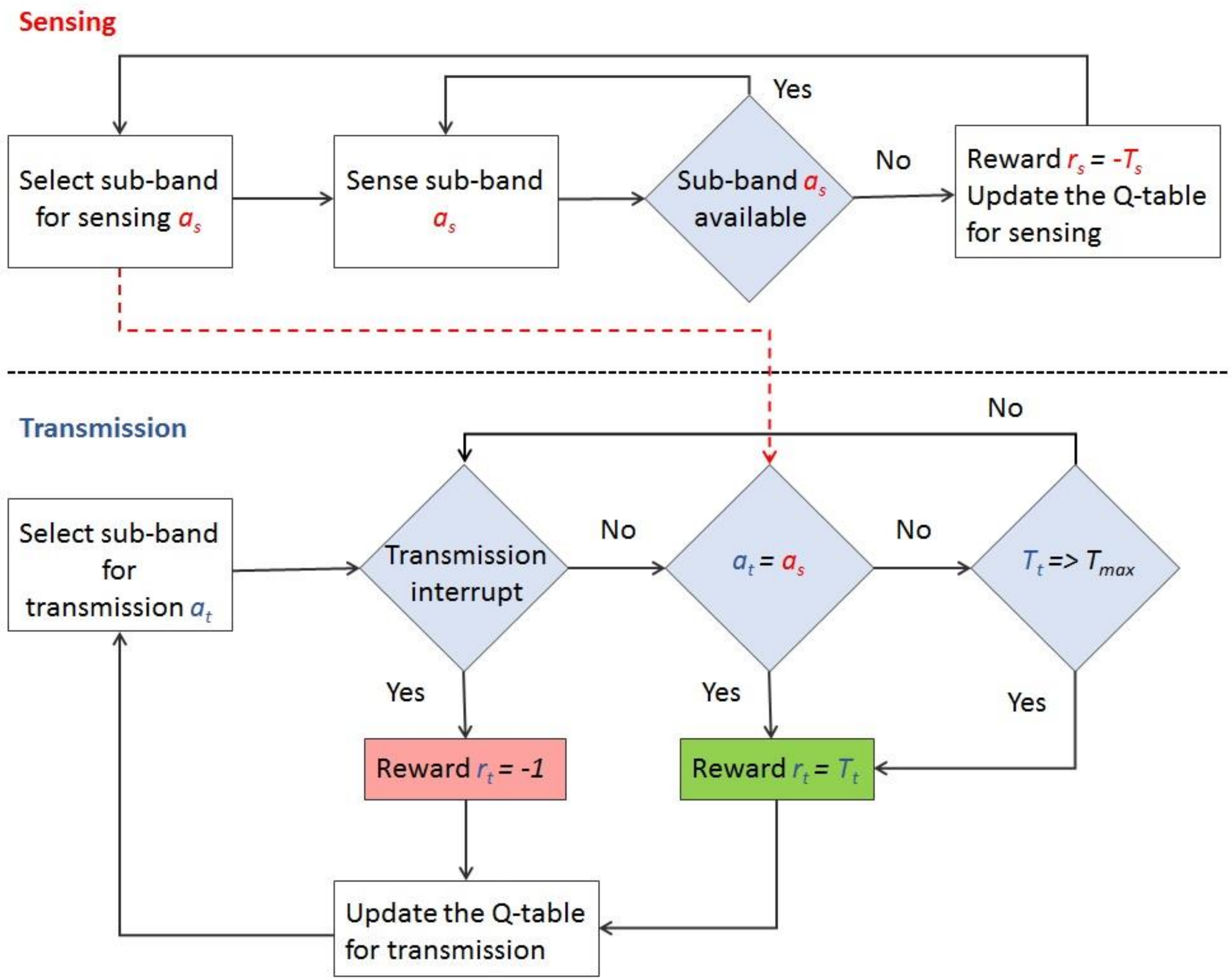
9: $$a' = \begin{cases} \arg\max_{a \in \mathcal{A}} Q(s', a) & \text{with probability } 1 - \epsilon, \\ \sim U(\mathcal{A}) & \text{with probability } \epsilon, \end{cases}$$

- Once the Q-table is updated, the WACR selects a new action a' representing the new operating sub-band.

Proposed Anti-jamming Stochastic Game



Proposed Anti-jamming Stochastic Game



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:

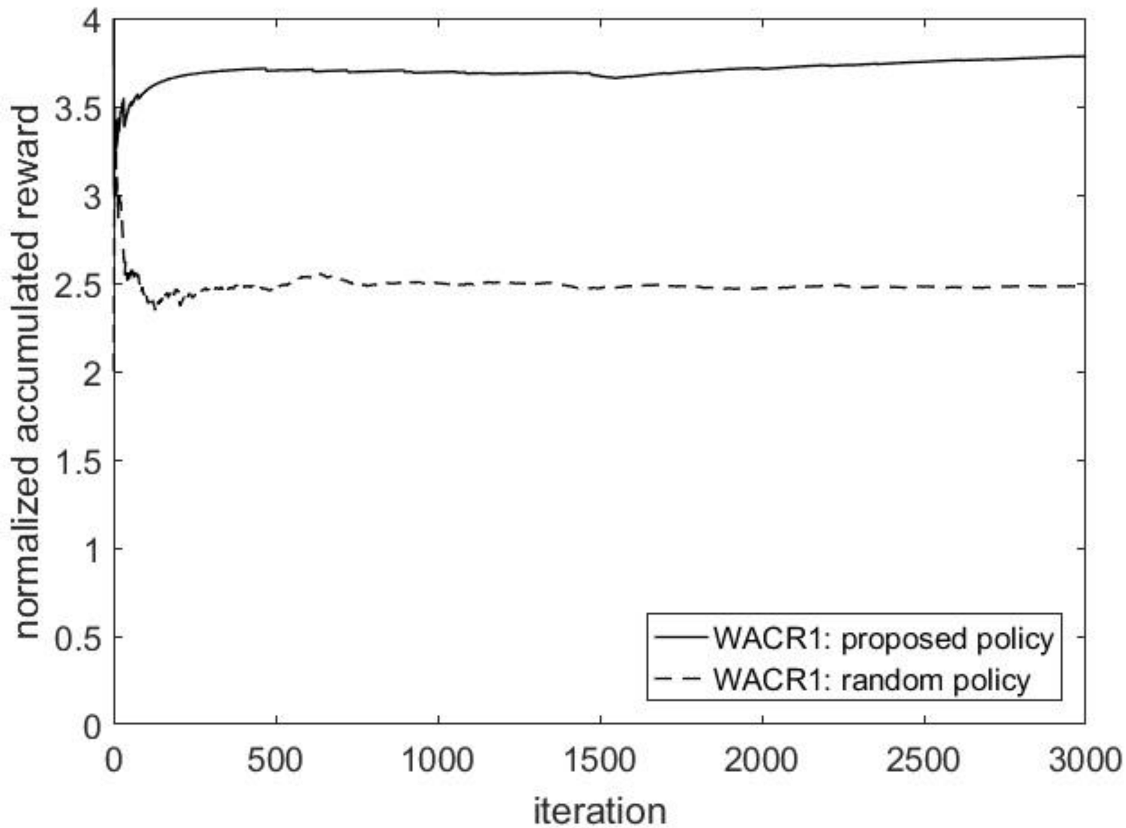
ϒ=0.8

ϵ=0.9, α =0.4 → Before Q-table convergence

ϵ=0.01, α =0.1 → After Q-table convergence

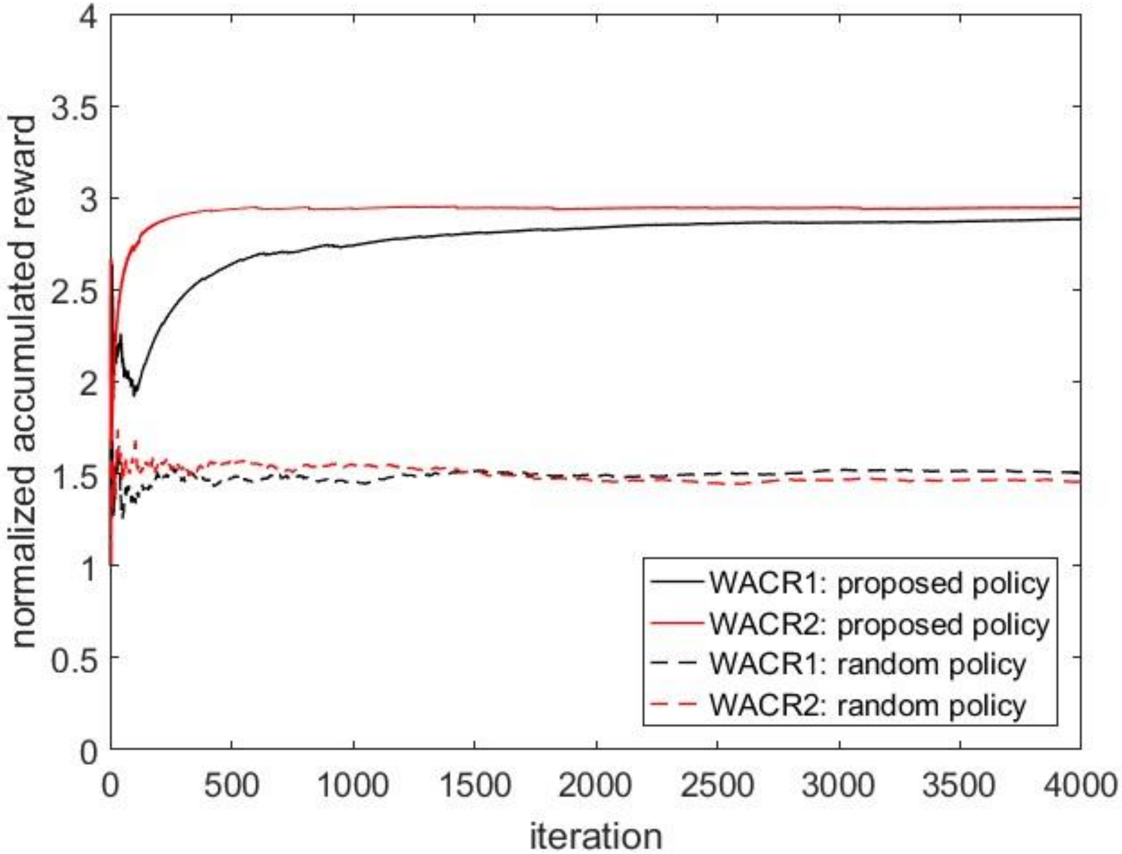
Simulation results

Experiment 1: 1 WACR and 5 Sub-bands



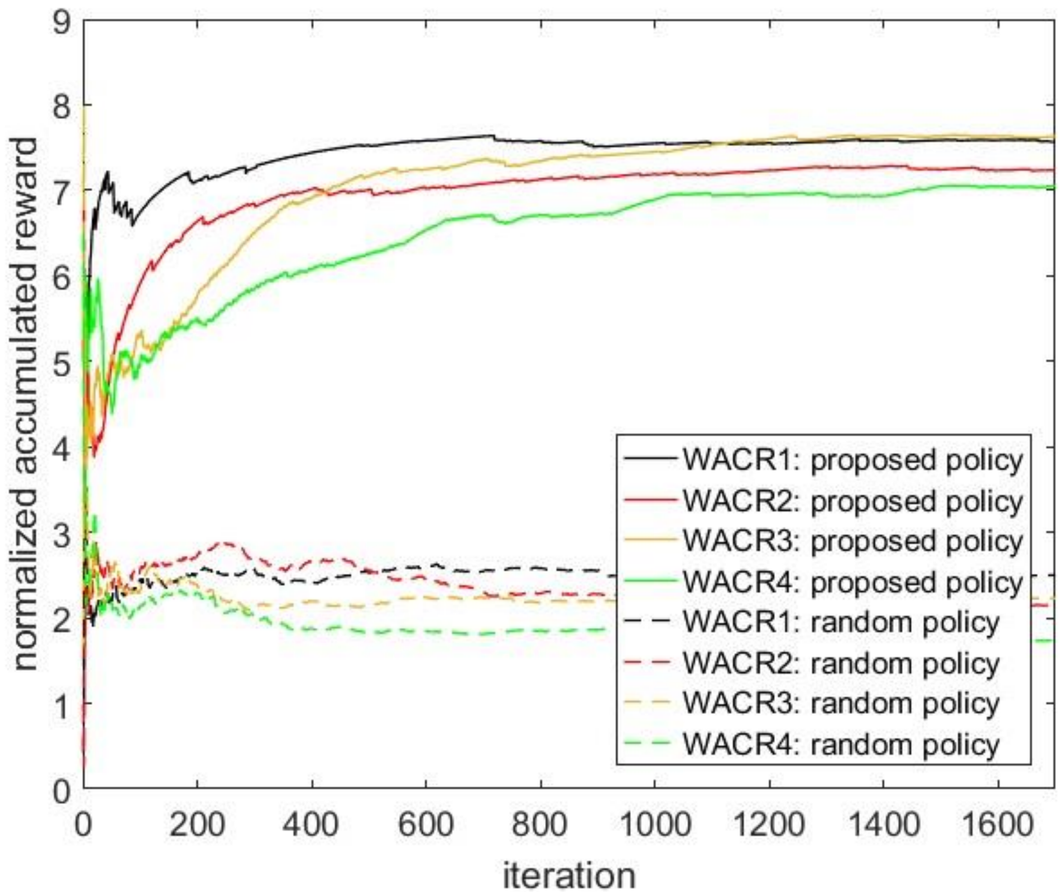
Simulation results

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



Simulation results

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



Simulation results

Table I
NORMALIZED ACCUMULATED REWARD VALUES FOR DIFFERENT SIMULATION SCENARIOS

Test case	**Scenario**	**Reward upper bound**	**WACR 1**	**WACR 2**	**WACR 3**	**WACR 4**	**Average**
1	1 WACR and 5 sub-bands	4	Proposed:3.8 Random: 2.5				Proposed:3.8 Random: 2.5
2	2 WACRs and 6 sub-bands	4	Proposed:2.8 Random: 1.5	Proposed:3 Random: 1.4			Proposed:2.9 Random: 1.45
3	4 WACR and 16 sub-bands	12	Proposed:7.5 Random: 2.5	Proposed:7.2 Random: 2.2	Proposed:7.5 Random: 2.2	Proposed:7 Random: 1.8	Proposed:7.3 Random: 2.17

Simulation results

Table II
PROBABILITIES OF GETTING JAMMED FOR DIFFERENT SIMULATION SCENARIOS

Test case	Scenario	WACR 1	WACR 2	WACR 3	WACR 4	Average
1	1 WACR and 5 sub-bands	Proposed: 0.86% Random: 1.8%				Proposed: 0.86% Random: 1.8%
2	2 WACRs and 6 sub-bands	Proposed: 2.6% Random: 47.2%	Proposed: 2.1% Random: 48%			Proposed: 2.35% Random: 47.6%
3	4 WACR and 16 sub-bands	Proposed: 6.4% Random: 64.8%	Proposed: 7.6% Random: 66.3%	Proposed: 12.4% Random: 66.3%	Proposed: 12.3% Random: 72.6%	Proposed: 9.6% Random: 67.5%

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

1. M. A. Aref, S. K. Jayaweera and S. Machuzak, “Multi-agent Reinforcement Learning Based Cognitive Anti-jamming”, IEEE Wireless Communications and Networking Conference (WCNC’17), San Francisco, CA, Mar. 2017.

2. H. M. Schwartz, “Multi-Agent Machine Learning: A Reinforcement Approach,” John Wiley & Sons, ISBN: 978-1-118-36208-2, 2014.

3. B. Wang, Y. Wu, K. Liu, and T. Clancy, “An anti-jamming stochastic game for cognitive radio networks,” IEEE Journal on Selected Areas in Communications, vol. 29, no. 4, Apr. 2011.

4. Y. Gwon, S. Dastangoo, C. Fossa, and H. T. Kung, “Competing mobile network game: Embracing anti-jamming and jamming strategies with reinforcement learning,” IEEE Conference in Communications and Network Security (CNS’13), National Harbor, MD, Oct. 2013.

5. M. Bowling and M. Veloso, “Rational and Convergent Learning in Stochastic Games,” 17th international joint conference on Artificial intelligence (IJCAI’01), Seattle, WA, Aug. 2001.

6. S. K. Jayaweera, “Signal Processing for Cognitive Radio,” John Wiley & Sons, ISBN: 978-1-118-82493-1, 2014.

7. R. S. Sutton and A. G. Barto, “Reinforcement Learning: An Introduction,” MIT Press, 1998.