EPT University of Carthage

APPLIED SCIENCES

Laboratory of electronic systems and communication networks

Royal Military Academy

POLYTECHNIC FACULTY

Department of Communication, Information, Systems and Sensors

Study of solutions against security threats in the context of cognitive radio

Thesis submitted in fulfillment of the requirements for the award of Doctor in Telecommunications by

Feten Slimeni

Mai 2017

Promotors: Prof. Bart Scheers
Prof. Rabah Attia

Copromotors: Prof. Zied Chtourou

Acknowledgements

Firstly, I gratefully acknowledge the Tunisian Ministry of Defense that provides me the Ph.D opportunity and funds my work.

I would like to thank my advisors Prof. Bart Scheers, Prof Zied Chtourou and Prof. Rabah Attia for the continuous support of my Ph.D study and related research, for all their contributions and motivation. My sincere gratitude also to Dr. Ir. Vincent Le Nir for his guidance and help during research and writing of this thesis. Special thanks to Dr. Ir. Hafeez M. Chaudhary who has always been available for advice and sharing knowledge.

For the non-scientific side of my thesis, I particularly want to thank my husband Mohamed Hedi and my son Ghassen for enduring my absence, for patience and spiritual support. Last but not the least, I would like to thank my parents, my brother and my sisters for all the love and encouragement.

Acronyms

API application programming interface.

ASIC application specific integrated circuit.

CCC common control channel.

CE cognitive engine.

CLARION connectionist learning with adaptive rule induction on-line.

CR cognitive radio.

CRC cyclic redundancy check.

CRN cognitive radio network.

DoS denial of service.

DP dynamic programming.

DS-SS direct-sequence spread spectrum.

DSA dynamic spectrum access.

DSC diagonally strictly concavity.

DSM dynamic spectrum management.

FC fusion center.

FEC forward error correction.

FH frequency hopping.

IE incumbent emulation.

JENNA jamming evasive network coding neighbor discovery algorithm.

JRCC jamming-resilient control channel.

KKT Karush Kuhn-Tucker.

KLD Kullback-Leibler divergence.

LocDef localization defense.

MARL multiagent reinforcement learning.

MDP Markov decision process.

MLE maximum likelihood estimation.

MME maximum-minimum eigenvalue.

NE Nash equilibrium.

OPSQ-learning on-policy synchronous Q-learning.

OSA opportunistic spectrum access.

POMDP partially observable Markov decision process.

PSR packet success rate.

PU primary user.

PUE primary user emulation.

RL reinforcement learning.

RRC root raised Cosine.

RSS received signal strength.

SDR software defined radio.

SINR signal-to-interference-plus-noise ratio.

SR software radio.

SSDF spectrum sensing data falsification.

SU secondary user.

SVDD support vector data description.

TDM time division multiplexing.

ToA time of arrival.

UHD USRP hardware driver.

USRP universal software radio peripheral.

UWB ultra wideband.

WBSS wideband spectrum sensing.

Nomenclature

C The value of the transmission capacity.

F The expression of the transmission capacity.

J The jammer's total power.

L The Lagrangian expression.

M The number of available channels.

N The number of states.

N_k The fictive noise over channel k.

P The CR's total power.

Q The Q-matrix of the learning algorithm.

R The reward function.

S The state.

α The learning factor.

β Coefficient of proportionality.

ϵ Convergence tolerance.

γ The discount factor.

λ KKT multiplier.

μ KKT multiplier.

∇ The gradient vector.

$\mathbf{G}$ The Jacobian.

$\mathbf{G}^T$ The transpose of matrix $\mathbf{G}$.

gr The pseudo-gradient vector.

j The jammer's vector of powers over the M channels.

p The CR's vector of powers over the M channels.

a The action.

c The column index.

f The central frequency of the considered channel.

g_k The gain of channel k for the jammer.

h_k The gain of channel k for the CR.

j_k The jammer's power over channel k.

k The channel index.

l The row index.

n_k The noise variance over channel k.

p_k The CR's power over channel k.

t Time slot.

Contents

1 Introduction **3**
- 1.1 Problem statement and motivations . . . 3
- 1.2 Research contributions . . . 4
- 1.3 Thesis document organization . . . 5

2 Dynamic spectrum management and cognitive radio technology **7**
- 2.1 Introduction . . . 7
- 2.2 Dynamic spectrum management . . . 7
  - 2.2.1 Exclusive-Use . . . 8
  - 2.2.2 Shared-Use of Licensed Spectrum . . . 8
  - 2.2.3 Commons . . . 8
- 2.3 Cognitive radio technology . . . 9
  - 2.3.1 From hardware to software defined radio . . . 9
  - 2.3.2 Cognitive engine . . . 10
  - 2.3.3 Cognitive Radio Functions . . . 12
- 2.4 Conclusion . . . 14

3 Threats in Cognitive Radio context **15**
- 3.1 Introduction . . . 15
- 3.2 Related works . . . 15
- 3.3 Spectrum Sensing Attacks and countermeasures . . . 16
  - 3.3.1 Detection of incumbent emulation attack . . . 17
  - 3.3.2 Defense against incumbent emulation attack . . . 18
- 3.4 Spectrum Decision Attacks and countermeasures . . . 18
  - 3.4.1 Spectrum sensing data falsification attack . . . 18
  - 3.4.2 Biased utility function attack . . . 20
- 3.5 Spectrum Sharing Attacks and countermeasures . . . 21
  - 3.5.1 Common Control Channel jamming attack . . . 21
  - 3.5.2 Intentional Jamming attack . . . 23
- 3.6 Spectrum Mobility Attacks and countermeasures . . . 27
  - 3.6.1 Routing information Attack . . . 27
  - 3.6.2 Key Depletion Attack . . . 27
- 3.7 Discussion . . . 28

3.8 Conclusion 29

4 Optimal power allocation in cognitive radio and jammer games 31
4.1 Introduction 31
4.2 Key terms 32
4.3 Related works 33
4.4 System model 34
4.5 Unilateral games 36
4.5.1 CR user Unilateral Game 36
4.5.2 Jammer Unilateral Game 37
4.6 Nash game 38
4.7 Stackelberg game 38
4.7.1 The jammer as the leader 38
4.7.2 The CR as the leader 39
4.8 Optimal solution: minmax/maxmin strategies 40
4.8.1 The CR user's maxmin strategy 41
4.8.2 The jammer's minmax strategy 41
4.9 Proof of the existence and uniqueness of the equilibrium in pure strategies 42
4.9.1 Existence of Nash equilibrium in pure strategies 42
4.9.2 Uniqueness of the Nash equilibrium in pure strategies . . 43
4.10 Closed form Expression of the Saddle Point 44
4.10.1 General case 45
4.10.2 Case all channels are used by both the CR and the jammer 46
4.10.3 Case of proportional fading channels 47
4.11 Simulation results and discussion 48
4.11.1 CR user unilateral game 48
4.11.2 Jammer unilateral game 49
4.11.3 Nash game 49
4.11.4 Stackelberg game: Jammer as the leader 50
4.11.5 Stackelberg game: CR user as the leader 50
4.11.6 Minmax/maxmin optimal solutions 51
4.11.7 Comparing analytical saddle point to the NE 51
4.11.8 Saddle point example 51
4.11.9 Nash equilibrium in the general case 52
4.12 Conclusion 52

5 Learning based anti-jamming technique 54
5.1 Introduction 54
5.2 Key terms 55
5.2.1 The Markov decision process 55
5.2.2 The Q-learning algorithm 56
5.3 ON-Policy Synchronous Q-learning (OPSQ-learning) 59
5.4 Learning based channel selection 59

5.4.1 Simulation results 60
5.4.2 Discussion 64
5.5 Ameliorated reward function for channel selection 66
5.5.1 Simulation results 66
5.5.2 Comparison in terms of reward per trial 67
5.5.3 Discussion 70
5.6 Learning based multi-channel power allocation game 70
5.6.1 CR using OPSQ-learning against fixed jamming strategy . 70
5.6.2 Both the CR and the jammer using Q-learning 71
5.6.3 Simulation and discussion 73
5.7 Conclusion 80

6 Cooperative learning based anti-jamming technique 81
6.1 Introduction 81
6.2 Cooperative learning algorithm 81
6.2.1 State definition 82
6.2.2 Reward function 82
6.2.3 Learning through cooperation 82
6.3 Simulation setup and results 85
6.3.1 Simulation setup 85
6.3.2 Simulation results 86
6.4 Experimental setup and measurements 90
6.4.1 Software development and hardware environment 90
6.4.2 Software defined radio measurements 93
6.5 Conclusion 97

7 Conclusions 99

List of Figures

2.1 Dynamic Spectrum Management 7
2.2 SDR and cognitive engine components within a CR 10
2.3 A SDR transceiver [1] 11
2.4 The cognition cycle [2] 12

3.1 The PUE attack 16
3.2 The SSDF attack 19

4.1 Waterfilling definition 33
4.2 Scenario of CR jamming attack 35
4.3 CR user unilateral game 48
4.4 Jammer unilateral game 49
4.5 The strategies at the NE 50
4.6 The saddle point for two channels 52
4.7 The strategies at the NE in general case 53

5.1 Exploitation of the learned policy against a sweeping jammer . . . 63
5.2 Exploitation of the learned policy against a reactive jammer 64
5.3 Exploitation of the learned policy against a pseudo random jammer 65
5.4 Exploitation of the optimal policy against a reactive jammer with the ameliorated reward function 68
5.5 Exploitation of the optimal policy against pseudo random jammers with the ameliorated reward function 69
5.6 Comparison between the reward functions 70
5.7 The transmission capacity over flat fading channels in the presence of sweeping jammer 74
5.8 The learned anti-jamming strategy against sweeping jammer over flat fading channels 75
5.9 The learned anti-jamming strategy against sweeping jammer over selective channels 77
5.10 The learned anti-jamming strategies against sweeping jammer attacking the same channel for 2 TSs and 3 TSs 77
5.11 The transmission capacity over selective channels against a jammer using Q-learning 78

5.12 The jamming and anti-jamming strategies at the NE 79

6.1 Cooperation diagrams 84
6.2 Simulation scenario 86
6.3 Simulation of best channel selection based on sensing (a) versus channel selection based on learning (b) against a sweeping jammer 87
6.4 Simulation of channel selection based on learning (a) versus channel selection based on cooperative learning (b) against a pseudo random jammer 90
6.5 Simulation of channel selection based on learning (a) versus channel selection based on cooperative learning (b) against a hidden reactive jammer 91
6.6 Experimental environment 91
6.7 Cooperation based on stop and wait protocol 92
6.8 Cooperation spectrum 92
6.9 Packet structure 93
6.10 Acknowledgment structure 93
6.11 Scenario1 94
6.12 Scenario2 94
6.13 USRP implementation of best channel selection based on sensing (a) versus channel selection based on learning (b) against a sweeping jammer 95
6.14 USRP implementation of channel selection based on learning against a pseudo random jammer 97
6.15 USRP implementation of channel selection based on learning against the first reactive jammer 97
6.16 USRP implementation of channel selection based on learning against the second reactive jammer 98

List of Tables

3.1 Classifications of CR attacks 16
3.2 Detection techniques of IE attacker 17
3.3 SSDF reputation based countermeasures 30

5.1 The Q^* matrix in a sweeping jammer scenario 62
5.2 The $Q*$ matrix in a reactive jammer scenario 63
5.3 The $Q*$ matrix in a pseudo random jammer with a period of 5 time slots 64
5.4 The $Q*$ matrix against a reactive jammer with the ameliorated reward function 68
5.5 The $Q*$ matrix in a pseudo random jammer (5TS) with the ameliorated reward function 69
5.6 CR using OPSQ-learning/waterfilling against sweeping jammer over flat fading channels 75
5.7 CR using OPSQ-learning/waterfilling against sweeping jammer over selective channels 76
5.8 Knowledge effect on the NE 79

6.1 Simulation setup 86
6.2 Simulation results: Packet Success Rate against slow sweep jammer 87
6.3 Simulation results: Packet Success Rate against fast sweep jammer 88
6.4 Simulation results: Packet Success Rate against pseudo random jammer 89
6.5 Simulation results: Packet Success Rate against reactive jammer . 90
6.6 USRP implementation results: Packet Success Rate against a sweeping jammer 95
6.7 USRP implementation results: Packet Success Rate against a pseudo random jammer 96
6.8 USRP implementation results: Packet Success Rate against a reactive jammer 96

Résumé

La croissance importante de services de communication avec des besoins spectrales plus élevés a révélé le problème de pénurie de spectre. Ce shortage est due à la technique d'allocation statique des bandes de fréquences. Cependant, des mesures de l'occupation spectrale ont prouvé qu'une large bande de fréquences est sous-utilisée en fonction du temps et de l'espace. Les techniques de gestion dynamique du spectre et les terminaux radio cognitives constituent une solution prometteuse pour améliorer l'utilisation du spectre radio. En revanche, due aux caractéristiques spécifiques de la radio cognitive, des mesures de sécurité doivent être developpées pour se protéger des attaques.

Dans cette dissertation, nous exploitons la théorie des jeux et la capacité d'apprentissage de la radio cognitive afin de se défendre du brouillage et d'accéder aux opportunités spectrales efficacement. Une plateforme cognitive a été utilisée pour développer, tester et valider les solutions proposées.

Dans une phase préliminaire, nous discutons les différentes attaques et leurs contre-mesures proposées dans la littérature, avec projection sur le contexte militaire. Nous suggérons d'étudier le brouillage considéré parmi les menaces graves dans les réseaux cognitives, en particulier lorsque le brouilleur est équipé d'un terminal intelligent. En première partie, nous analysons le brouillage en terme d'allocation de puissance entre plusieurs canaux. Nous utilisons le modèle de jeu à somme nulle et nous montrons sa convergence. Les allocations optimales de puissance, pour la radio cognitive et le brouilleur, sont déterminées sous l'hypothèse d'information complète. Ensuite, nous proposons une version modifiée de l'algorithme d'apprentissage Q-learning pour résoudre le problème sans avoir accès aux informations nécessaires concernant l'environnement et le brouilleur. Nous commençons par appliquer l'algorithme proposé en terme de sélection de canal de transmission avant de le généraliser pour résoudre le jeu d'allocation de puissance. Dans la première application, il permet à l'équipement radio d'apprendre à éviter les canaux brouillés. Dans le contexte multi-canaux, la stratégie apprise est comparée à la solution obtenue avec information complète. Enfin, nous nous abordons le problème de brouilleur caché de l'émetteur ainsi que le scénario réel de brouilleur asynchrone. Nous présentons un modèle approprié pour le problème de sélection de canal et nous améliorons l'algorithme proposé via la coopération entre deux radios cognitives. Nous fournissons les résultats de simulation et les mesures réelles à l'aide de plateforme radio programmable.

La solution proposée peut s'appliquer non seulement pour éviter les brouilleurs, mais aussi pour la coexistence de radio cognitive avec les titulaires des bandes spectrales.

Abstract

During the last century, the growing number of communication services with higher spectral requirements revealed the problem of spectrum scarcity. The shortage is due to the frequency allocation based on exclusive licensing. However, measurements of the spectral occupancy showed that a large portion of the assigned spectrum are under-utilized. Dynamic spectrum management (DSM) models and cognitive radio (CR) technology are presented as promising solutions to this imbalance between scarcity and under-utilization. However, the CR security threats emerge as a challenging issue for its military and commercial deployment.

In this dissertation, we aim at enabling the CR efficient exploitation of the spectrum in the presence of jamming attack. For that purpose, we use game theory models and the learning capacity of the CR. A Software Defined Radio/Cognitive Radio test bed architecture is used to develop, test and validate the proposed solutions.

In a preliminary part, we discuss the CR attacks and their potential countermeasures with the focus on military context. We chose to focus on the jamming attack, one of the most severe threats in cognitive radio networks especially when the jammer has the cognitive features. Firstly, we analyze the jamming attack in terms of multi-channel power allocation. We model the interaction between the CR and the jammer using zero-sum game. We prove its convergence to the pure strategy Nash equilibrium. We provide the optimal jamming and anti-jamming power allocations under the assumption of complete information. Secondly, we propose a modified version of the Q-learning algorithm to overcome the lack of information about the environment and the jamming tactics. The proposed algorithm is applied in terms of one channel selection, then it is adapted to solve the multi-channel power allocation game. In the first application, it enables the CR to pro-actively avoid the jammed channels. In the multi-channel application, we compare the learned strategy to the case of complete information. Finally, we address hidden jammer problem and real scenario of asynchronous jammer. We present a suitable model to the channel selection problem and we enhance the proposed algorithm through the cooperation between two cognitive radio nodes. We provide both simulation results and real measurements using software defined radio platforms.

The proposed solution may be applicable not only to avoid jammers, but also for the CR coexistence with incumbents.

Chapter 1

Introduction

Recently, there have been an excessive devise of communication technologies and new wireless services with significant growth of bandwidth requirements. This development is threatened by spectrum scarcity since, unlike wired communications using dedicated connections, wireless communications share a common connection medium. The spectral shortage is due to the traditional spectrum management based on exclusive licensing. Historically, regulators have assigned frequencies by issuing licenses to specific users specifying what equipment to use, where and at what power level it can be used. The goal of this rigid spectrum allocation was to ensure secure and reliable wireless communication, when the wireless technology was limited. But, measurements indicate that a large portion of the assigned spectrum is used only sporadically by the licensees, therefore leading to temporal and spatial spectrum under-utilization. This imbalance between scarcity and under-utilization requires striving for efficient spectrum allocation techniques ensuring that sufficient amounts of spectrum are available and accessible for current and future needs. The former technique of fixed spectrum allocation is giving way in many countries to DSM models improving the efficiency of spectrum use. The key technology enabling the DSM techniques is the CR technology build on software defined radio platform [1]. This concept was introduced by Joseph Mitola III in [2] to describe a conscious, intelligent and flexible radio. The CR is able to make autonomous decisions and adapt its characteristics according to the environment variations.

The following sections discuss CR security problem, subject of the thesis, and highlight the main contributions before detailing the roadmap of this dissertation.

1.1 Problem statement and motivations

One of the major issues in military and commercial deployment of cognitive radio technology is security, since its flexibility and dynamic access capacity lead to potential communication resource misuses and security threats [3]. In addition to common wireless security problems like eavesdropping and information tamper-

ing, cognitive radio networks (CRNs) introduce new classes of security threats due to the following specific characteristics [4]:

- High sensing sensitivity
  It is an intrinsic characteristic of CR technology to detect even low signals to avoid interference with legacy users. However, it may lead to false detections resulting in inefficient use of the spectrum.

- Hidden terminal problem:
  This problem occurs when the CRN can't detect the communication of a non cognitive radio by sensing the medium. Therefore, a CR can start transmission and interfere with the legacy user.

- Synchronization requirement:
  The CRs involved in cooperative spectrum sensing have to sense the spectrum periodically and transfer sensing reports to a fusion center (FC), which takes the decision about the spectrum exploitation. This process requires the synchronization between the cooperative CRs.

- Lack of common control channel (CCC)
  In military communication system, a common control channel is often avoided because it is a signal point of failure. This means that the network needs to search for control signals across the entire spectral band.

An attack in CRNs can be defined as a use of these reliability issues and may result in denial of service, unacceptable interference to the non cognitive user and missed opportunities for cognitive users, which threats the radio spectrum sharing policy used to manage the spectrum access. Even more, the cognitive radio technology may constitute a double-edged sword in the communication electronic warfare. Its cognitive capabilities along with its self-reconfiguration abilities aid in the development of advanced security measures and may also be exploited by malicious users to deploy advanced attacks. Recent tactical military scenarios require interoperability among different military services and allied forces, as well as among military and civil authorities.

Therefore, the CR threats emerge as a challenging issue for advanced defense applications, and have to be resolved to enhance the deployment of CRNs. The work presented in this thesis studies jamming and anti-jamming tactics using cognitive radio technology with the focus on military context. This research is mainly based on game theory models and the Q-learning artificial intelligence algorithm to deal with diverse jamming tactics. A Software Defined Radio/Cognitive Radio test bed architecture is used to develop, test and validate the proposed solutions.

1.2 Research contributions

The main contributions of this thesis are summarized below.

- A comprehensive review and classification of the main security issues related to cognitive radios and the potential countermeasures with projection on military context.
- Diverse game-theoretical scenarios are presented to analyze cognitive radio jamming/anti-jamming strategies in terms of multi-channel power allocation over parallel Gaussian channels. We give the theoretical proof of existence and uniqueness of the equilibrium in pure strategies and we develop the analytical expression of the saddle-point.
- Proposition of a real-time reinforcement learning algorithm denoted on-policy synchronous Q-learning (OPSQ-learning). Its application to solve the jamming attack in terms of one channel selection and its generalization to multi-channel power allocation.
- Enhancing the developed OPSQ-learning solution through the cooperation between two cognitive radio nodes, and detailing the corresponding implementation on the test bed architecture. We provide high fidelity simulation results and real measurements using software defined radio platforms.

1.3 Thesis document organization

This dissertation is composed of six chapters that are based on a number of journals and conference papers. The next chapter is dedicated to outline the dynamic spectrum management techniques and present the cognitive radio technology components and its main functions.

Chapter 3 provides a comprehensive review of the literature related to common CR attacks and their potential countermeasures with the focus on military context. The classification of attacks is based on the four main functions of the cognitive radio, not according to the layers of the OSI model as usually done. Through this classification, we try to provide directions for related researches to discern which cognitive function has to be insured against each threat. We compare the CR attacks in terms of harmfulness and required knowledge. We opt for the study of the jamming attack since it is easy to happen and hard to mitigate. It is one of the most severe threats in cognitive radio networks especially when the jammer has the same cognitive features as the CR nodes. CRNs are characterized by dynamic spectrum access (DSA) and by mainly distributed architectures which make it difficult to implement effective jamming countermeasures.

The developed work in the remaining chapters is based on game theory models and the CR learning capacity to propose solutions for CR jamming attacks. Cognitive jammers are able to deploy advanced strategies and adapt their tactics to degrade the performance of cognitive radio communication, through injecting the total power to one channel or sharing it between multiple channels. We start by analyzing the problem of power allocation in cognitive radio user and jammer games,

over parallel Gaussian channels, in chapter 4. We model the interaction between a transmitter-receiver pair and a jammer using zero-sum games with continuous action sets; we describe unilateral, Nash and Stackelberg games. We compare the Nash equilibrium, the Stackelberg equilibrium and the minmax/maxmin optimal power allocations through the simulation of the diverse game scenarios. Furthermore, we give the theoretical proof of existence and uniqueness of the equilibrium in pure strategies and we develop the analytical expression of the saddle-point.

Chapter 5 presents a real-time learning algorithm to overcome the lack of information about the environment and the jamming tactics. The proposal is applied in terms of one channel selection, then it is adapted to solve the multi-channel power allocation game studied in chapter 4. In the first application, we model the jamming scenario as a Markov decision process (MDP) and discusses how Q-learning can be used to pro-actively avoid the jammed channels. Since, Q-learning needs a long training period to learn the behavior of the jammer, wideband spectrum sensing is considered to speed up the learning process and the already learned information are exploited to minimize the number of collisions with the jammer. The learned anti-jamming strategy depends on the chosen reward function which reflects the preferences of the cognitive radio. We define reward values based on the avoidance of the jammed channels and we compare the result to the original Q-learning algorithm. The effectiveness of our proposal is evaluated in the presence of different jamming strategies which reveals some limits. In the multi-channel application, we start by comparing the learned anti-jamming power allocation strategy to the common waterfilling technique. Then, we consider the power allocation game using Q-learning for both the cognitive radio and the jammer. The learned strategies will be compared to the Nash equilibrium found under the assumption of complete information.

Chapter 6 generalizes the proposed OPSQ learning algorithm for asynchronous jammer scenario. We define a suitable state space modeling and a more significant reward function related to the detected energy during the sensing period. Moreover, we enhance the proposed solution through the cooperation between two cognitive radio nodes to overcome the hidden jammer problem and to defeat various jamming strategies. We describe the simulation setup and detail the test-bed implementation of the proposed anti-jamming channel selection algorithm in the universal software radio peripheral (USRP) platform. Both high fidelity simulations and real USRP measurements reveal that the presented solution achieves a higher packet success rate compared to both the classical fixed channel selection and the best channel selection based only on sensing. Results are given for various scenarios and diverse jamming strategies.

Finally, we conclude the presented work, highlight the contributions and outcomes, and discuss possible future extensions.

Chapter 2

Dynamic spectrum management and cognitive radio technology

2.1 Introduction

This chapter starts by the definition of the dynamic spectrum management concept recognized as a promising solution to the imbalance between spectrum scarcity and under-utilization. Section 2.3 defines the cognitive radio technology enabling the implementation of the dynamic spectrum management techniques. We start by presenting the technology evolution from hardware to software until cognitive radio before defining its main functions: spectrum sensing, decision, sharing and mobility.

2.2 Dynamic spectrum management

DSM is a set of theoretical techniques developed to manage spectrum sharing to improve the network performance. As represented in figure 2.1, the DSM concept defines three main techniques: exclusive use, shared use and commons [5].

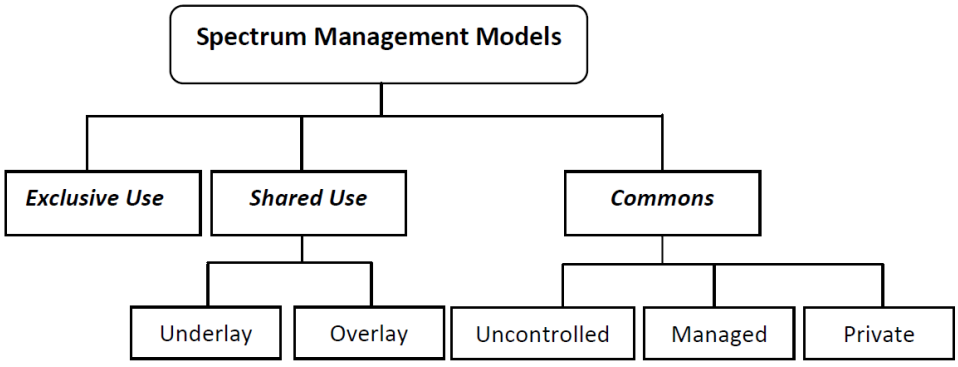


Figure 2.1: Dynamic Spectrum Management

2.2.1 Exclusive-Use

In this model, spectrum bands are licensed to serve for exclusive use and the rules of spectrum users are clearly governed by a central management body. This method solves the problem of interference between legitimate users. However, it can no longer respond to the increased demand of radio spectrum.

2.2.2 Shared-Use of Licensed Spectrum

A frequency band is simultaneously shared between two categories of users; a licensed user known as primary user (PU) and multiple non licensed users denoted as secondary users (SUs). The SUs utilize underlay or overlay approaches to exploit the spectrum without interfering with the PUs.

- Spectrum Underlay:
  The SUs use ultra wideband (UWB) techniques to transmit simultaneously with the PUs over the same channels. A spectral mask is applied to secondary signal so that the interference is below the acceptable level of the PU's signal.

- Spectrum Overlay:
  The SU is allowed to utilize licensed bands in opportunistic way by identifying and exploiting spatial and temporal unused radio spectrum called white space. The spectrum overlay technique was first denoted by Mitola as spectrum pooling technique and later called opportunistic spectrum access (OSA) in DARPA XG program. To apply spectrum overlay access, the SU should be able to detect continuously the state of the spectrum in order to exploit the idle frequency bands and to vacate the bands whenever requested by the PU.

2.2.3 Commons

According to this scheme, the spectrum bands are equitably and fairly accessible to every user and no body can claim exclusive use. The commons model has three variants:

- Uncontrolled Commons:
  Each user has open spectrum access to a common band and can have many devices operating in it. But, the participating devices have to conform to a maximum transmit power.

- Managed Commons:
  These commons are controlled jointly by a group of users with restrictions on who, when and how the resource is used, defined by the controller of the commons. The use of managed commons requires a good management protocol that encapsulates technology agnostic rules together with reliable

and scalable mechanisms that quantify rule on performance of participating devices.

- Private Commons:
  This model consists in licensed bands access at the discretion of license holders. It is like managed commons but the ultimate ownership of the licensed spectrum is still centralized with the license holder who offers either private commons service or spectrum access.

Since countries have diverse spectrum access regulations and recently military operations are often conducted in coalitions, the application of DSM in future military networks will most probably be based on the shared use of licensed spectrum and commons models [6]. Therefore, we will not use the classical PU/SU denomination, but in a more general way we classify the users as: non-cognitive legacy users having priority to access the spectrum; and cognitive users having to periodically scan and identify the vacant channels in the spectrum to communicate without interfering the non-cognitive user.

The DSM models need to be linked to the new CR technological capabilities to overcome the spectral scarcity.

2.3 Cognitive radio technology

The CR concept, introduced by Mitola, defines a radio that can be programmed and configured dynamically according to the environment changes [2]. It enables DSM techniques through its capacities of idle channels detection, learning and reasoning. The CR is built on the software defined radio (SDR) technology which is the result of an evolutionary process from purely hardware-based equipment to software-based equipment [1]. The CR can be divided into two units as given by figure 2.2; the SDR unit and an intelligence unit, which is commonly denoted cognitive engine, that adds the cognition capacity to the radio and manipulates the SDR's operating state. The cognitive engine includes a knowledge base that stores the predicates collected by the SDR unit and evolved through learning. It also stores the available actions and rules based on reasoning [7]. The operating parameters of the SDR unit (such as the frequency range, the modulation type and the maximum transmission power) can be dynamically adjusted by software [8, 9]. We start by presenting the evolution from hardware to SDR technology, then we define the cognitive engine and detail the main functions of a CR.

2.3.1 From hardware to software defined radio

Unlike, conventional fully hardware radios which are commonly based on application specific integrated circuit (ASIC) devices, a software radio (SR) is a transceiver which communication functions (e.g. amplifiers, mixers, filters, modulators/demodulators) are realized as programs running on a suitable digital pro-

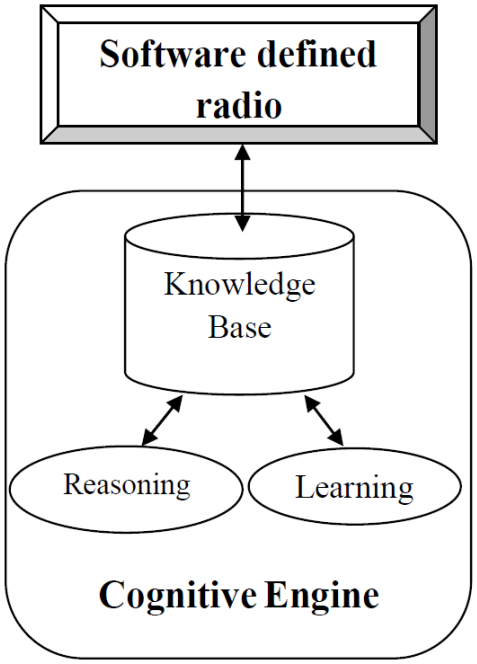


Figure 2.2: SDR and cognitive engine components within a CR

cessor. Different software transmitter/receiver algorithms, usually known as transmission standards, can be implemented on the same hardware. This evolution from hardware to software radios results from the need of serving wide variety of changing radio protocols in real time, such as for the military and cell phone services. A SR is said to be ideal if it directly samples the antenna output. A practical version of a SR, in which the received signals are sampled after a suitable band selection filter, is known as SDR [1].

Several SDR platforms [10] (e.g. BladeRF, HackRF, USRP) are available in the market, they are reconfigurable and reprogrammable with adjustable front-end which operates with different carrier frequency, signal bandwidth, variable transmit powers and different modulation types with different symbol rates. These platforms are also characterized by flexibility for different layers (forward error correction, data framing, multiplexing and scheduling), making them suitable for CR implementation. A SDR transceiver is given in figure 2.3.

The transmission parameters of the SDR can be reconfigured according to the cognitive engine decision depending on the surrounding spectrum conditions.

2.3.2 Cognitive engine

The cognitive engine (CE) analyzes the surrounding spectral information and takes the decision about its spectrum access. A cognitive engine may include a Database, a reasoning unit and a learning unit. The Database can store the historical spec-

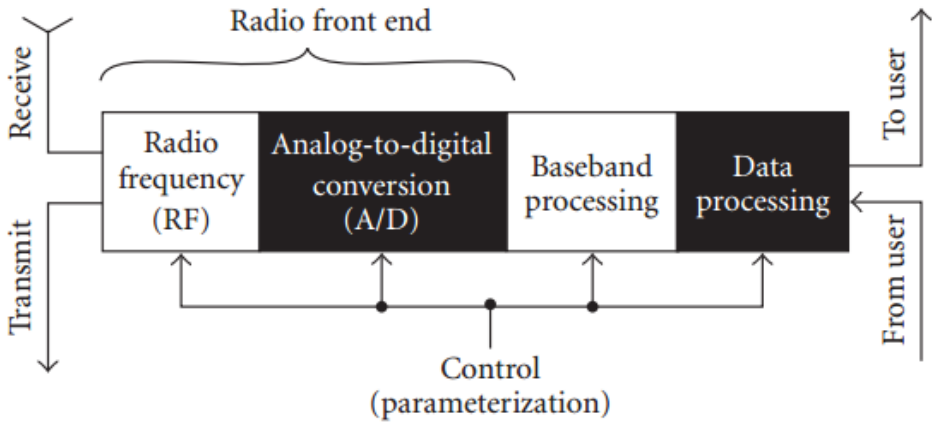


Figure 2.3: A SDR transceiver [1]

tral information and taken decisions. The reasoning unit is a set of logical rules mapping known spectral states to radio configurations (decisions). The learning unit allows creation of new rules and update of existing ones through trying radio configurations, observing and analyzing the performance in order to optimize the decisions. It makes the CR able to handle new situations using learning algorithms [11–14].

A great number of cognitive architecture frameworks (e.g. ACT-R, Soar and Clarion) have been presented in [15] as basis for implementation of cognitive engines for cognitive radio. [16] discussed desired properties for a cognitive engine in a complex unpredictable radio environment such as in a military ad hoc deployment. Such scenarios may include malicious activity like spoofing and jamming, which make simple spectrum access algorithms vulnerable and transparent for attackers. The requirements for these scenarios are as follows:

- Fast reactive response
- Compile new fast reactive responses through learning and reasoning: to complement initial fast reactive responses with new patterns of learnt settings
- Reactive response through reasoning and problem solving
- Continuous learning due to the environment and the radio sets changing over time
- Proactivity and planning of action strategies
- Interaction and communication to benefit from knowledge learned by other cognitive radios

- Metacognition for reasoning about which type of cognitive processes to use (e.g. when to use reactive reasoning and when to use a plan)
- Implementable on limited-resource embedded platforms for handheld and portable devices, the CE must be able to run on low-power processors.

Based on the two previous references, we can conclude that the framework that constituted the best match to military scenarios is connectionist learning with adaptive rule induction on-line (CLARION). A prototype CE made using CLARION framework is tested in [16] towards a simplified, simulated environment with spectrum occupation and reactive jamming. It is seen to learn how to behave in both non-jammed and jammed environments.

2.3.3 Cognitive Radio Functions

Figure 2.4 gives the cognition cycle as introduced by Mitola. This cycle describes how a cognitive radio interacts with the environment. The four major functions of this cycle are: spectrum sensing, spectrum decision, spectrum sharing and spectrum mobility.

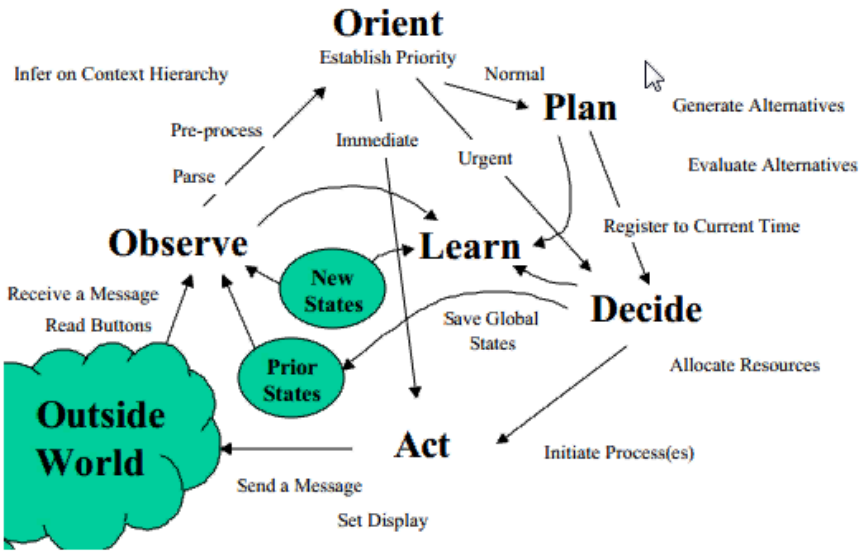


Figure 2.4: The cognition cycle [2]

- **Spectrum sensing**

Spectrum sensing is the first function in the cognition cycle, it consists in detecting available portions of the spectrum. Then, CRs can temporarily transmit over these spectrum holes without creating interference to the legitimate users.

Moreover, it has to periodically sense the spectrum to detect the presence of incumbents and quit the band once detected. The CR first gathers (observe) information about its external electromagnetic environment to detect the unused spectrum bands called white spaces. Then, this information is evaluated (orient) to know its significance and to determine the features of each band. To improve the detection performance, cognitive nodes may collaborate to combat sensing issues such as the problem of hidden transmitting nodes. Cooperative uses various data fusion schemes which can be classified mainly into hard fusion when the fusion center (FC) collects the local decisions of cognitive nodes, and soft fusion if it collects the local detected signal of each node. [17] gives a comparison between the three main spectrum sensing techniques: energy detection, matched filtering and cyclostationarity. Energy detection method consists in detecting the incumbent presence based on the sensed energy. It is the most popular spectrum sensing technique since it is simple and does not need a priori information. After filtering the detected signal, its integral over time is compared to a predefined threshold to decide if an incumbent signal exists. Matched filtering is based on a designed filter that depends on prior knowledge of the incumbent signal. Cyclostationary spectrum sensing looks for periodic statistics, characterizing cyclostationary signals, based also on prior information about the incumbent signal.

- **Spectrum management known as spectrum decision**

Based on the evaluation of SS report taken into account the information from policy Database (knowledge base), the CR determines (plan) its alternatives to meet user communication requirements. Then, it chooses (decide) the most appropriate frequency band. In cooperative CRN, the spectrum decision is taken by the FC through combining either local decisions or local detected signals received from cognitive nodes. In [18], the authors present a simulation comparison of six fusion rules: likelihood ratio combining, soft optimal linear combining, soft equal weight combining, hard decision combining using the OR/AND/Majority counting rules.

- **Spectrum sharing**

Spectrum sharing techniques manage the allocation of available frequency bands to provide a fair spectrum scheduling among the users and to avoid the interference. The spectrum sharing function includes channel and power allocation to prevent both interference with incumbent users and CRs collision in overlapping portions of the spectrum. A CCC facilitates the spectrum sharing challenge, but fixed CCC implementation is infeasible because every channel must be vacated when it is solicited by the incumbent user. The spectrum sharing techniques can be applied inside a CRN (intra-network spectrum sharing) or among multiple coexisting CRNs (inter-network spectrum sharing) [19].

- **Spectrum mobility**

It is known as spectrum handoff and defined as the process where the cognitive user changes its frequency of operation or vacate it to the incumbent user. During this operation, the transmission of the CR is interrupted and its packets wait

in the transmission buffer. The communication can be resumed only when the connection between the transmitter and the receiver is successfully restored in a new channel. [20] has proposed spectrum handoff approaches for CRNs in order to support delay sensitive applications.

2.4 Conclusion

This chapter defines the DSM techniques that manage spectrum coexistence of non-cognitive legacy users having priority to access the spectrum; and cognitive users having to periodically scan and identify the vacant channels in the spectrum to communicate without interfering the non-cognitive user. We have defined the CR technology composed of software defined unit and cognitive unit. We have detailed the CR main functions: spectrum sensing, decision, sharing and mobility. Based on sensing, the CR detects the available portions of the spectrum. It evaluates the sensing report taken into account the information from knowledge base to decide its alternatives to meet user communication requirements. Spectrum sharing manages the allocation of available frequency bands to provide a fair spectrum scheduling among the users and to avoid the interference. Through spectrum mobility, the CR changes its frequency of operation to vacate it for the incumbent user. In the next chapter, we will provide a review of the CR threats and solutions through a classification based on the four functions.

Chapter 3

Threats in Cognitive Radio Context

3.1 Introduction

The decisions made by each CR or the decision made by the fusion center of a centralized CRN depend on; (i)the incumbent users' activities, (ii) malicious users' behaviors and (iii) wireless channels' characteristics. These three factors constitute sources of uncertainty and unreliability to the CR about its surrounding environment. This chapter reveals how malicious users threaten the efficient exploitation of cognitive radio technology. It provides a comprehensive review of common CR attacks and their potential countermeasures. We choose to rank specific cognitive attacks according to the four main functions of the CR to better evaluate its impacts and to discern which cognitive function has to be insured against each threat which may provide directions for related researches. Section 3.2 provides related works. It is followed by a review of studies and existing solutions to the attacks of each CR function, separately. In section 3.7, we compare the harmfulness of the presented attacks and reveal the attack chosen to be studied in this thesis.

3.2 Related works

Table 3.1 presents diverse classifications used in related surveys to describe the CRN attacks. Most of the papers opt the OSI model layer classification (physical, link, network, transport and cross-layer attacks). Some works use a classification according to the steps of the cognition cycle. However these papers sometimes misclassify certain attacks, such as the spectrum sensing data falsification (SSDF) attack which happens during the decision function but usually defined as a spectrum sensing attack. Furthermore, these papers do not detail the consequences of given attacks on the cognition cycle. We provide in the following sections the CR attacks classified according to its main functions (spectrum sensing, spectrum decision, spectrum sharing and spectrum mobility), to discern which function should

be enhanced against which threat.

CRN attack classification	Related paper
Cognition cycle steps	[21, 22]
Inside/outside the CRN	[23]
OSI layer classification	[24–26]
Nature of manipulation (sensor, belief, control...)	[27]
Attacker type: malicious/greedy	[28]
CR components	[29]

Table 3.1: Classifications of CR attacks

3.3 Spectrum Sensing Attacks and countermeasures

The sensing information can be falsified by malicious users. The most common attack threatening the function of SS is the primary user emulation (PUE) attack described in figure 3.1.

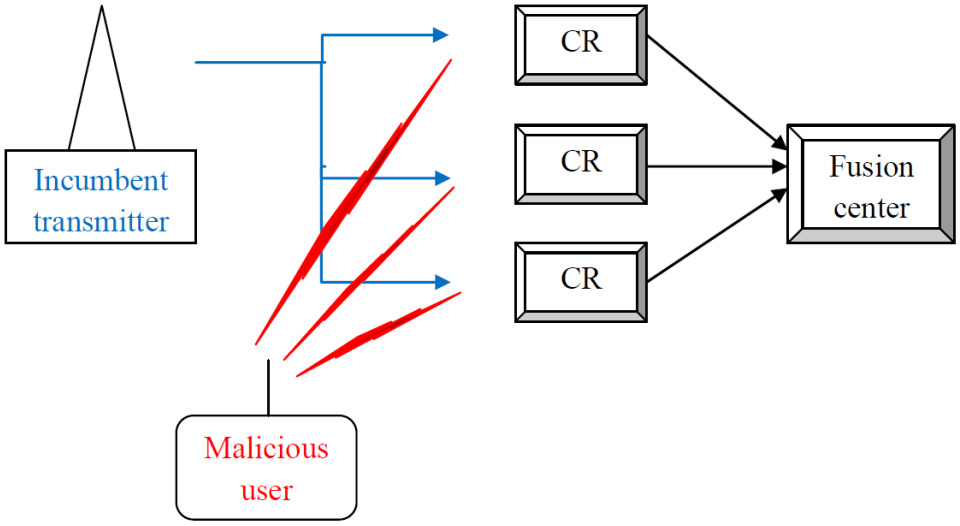


Figure 3.1: The PUE attack

In this attack, a malicious user emulates the signal of incumbent user during the spectrum sensing period to get the priority, since other cognitive users will falsely sense that the frequency is in use by a legitimate user and vacate it. In commons DSM model, we don't have the concept of primary/secondary user, so it is better to call this attack as incumbent emulation (IE) attack to cover different DSM models mainly in military scenarios. A variety of IE attack solutions are presented in the review [30]. The author enlisted various methods but without details. The

paper doesn't discern between techniques used just to detect an IE attacker and techniques used as countermeasure.

In the next paragraph, we present the common detection methods followed by some possible defensive techniques along with related references.

3.3.1 Detection of incumbent emulation attack

The CR has the challenge of distinction between the legacy incumbent node and the imitating attacker. To overcome this issue, the CR can localize the transmitting node and compare its position with known legacy users positions. It can also determine the characteristics of the detected signal and compare it with incumbent signals. We explain in table 3.2 common detection techniques and we refer to some papers.

Detection method	**Technique description**	**Related paper**
Transmitter verification	-Distance difference and distance ratio to verify the transmitter location (DDT & DRT) -RSS, or ToA to determine the characterizations of the detected signal -LocDef combining location and characterization previous techniques	[31]
Analytical detection model	Fenton Approximation Method to determine the mean and the variance of the detected signal + Markov inequality or Weighted Sequential Probability Ratio Test (WSPRT) to determine a threshold on the probability of a successful IE attack	[32], [33]
Signal activity Pattern Acquisition and Reconstruction System	Reconstruction of the observed signal activity pattern (such as the ON/OFF periods) through spectrum sensing. + Examination of the reconstruction error to distinguish incumbent user's signal activity pattern from attacker's one.	[34]

Table 3.2: Detection techniques of IE attacker

3.3.2 Defense against incumbent emulation attack

After identifying the IE attacker, the CR can either avoid or counteract the attacker. The first approach consists in avoiding the channels used by the IE attacker. The CR can use game theory and learning algorithms, such as the Q-learning algorithm, in order to learn how to avoid jammed channels. For example, a defense scheme is proposed in [35] and called dogfight in spectrum. The scenario is modeled as a zero sum game between the IE attacker and defending CRs. It is based on randomly choosing a channel to sense and transmit at each time so as to avoid the IE attack statistically.

A countermeasure proposed in [36], assigns weights to the local detected energies to eliminate the malicious signal sent by the IE attacker. The problem is solved in [37] by using spatial correlation based user selection to choose the members taking part in cooperative spectrum sensing, and the maximum-minimum eigenvalue (MME) based detection mechanism to perform the cooperation. A technique of belief propagation of location information, presented in [38], enables not only detecting but also mitigating the IE attack. In this approach, each cognitive sensor calculates the location information based on RSS measurements and exchanges messages with the neighbors to detect the IE attacker according to the mean of the final beliefs based on a belief threshold. Then, a broadcast message informs all cognitive users about the characteristics of the malicious signal to avoid it in the future sensing period, thus the IE attacker will no longer falsify the sensing results.

3.4 Spectrum Decision Attacks and countermeasures

In cooperative cognitive radio network, each CR senses the spectrum periodically and reports the measurement results to the FC node, which combines the data and makes the final decision of whether the incumbent user is present or not. However, this result is based on the assumption that all users sending the sensing reports are honest and there is no malicious entity that can manipulate the spectrum decision process. This defect leads to several attacks by malicious nodes inside the CRN. The common threats to the spectrum decision function are the spectrum sensing data falsification (SSDF) and the objective function attacks.

3.4.1 Spectrum sensing data falsification attack

In cooperative spectrum sensing process, malicious user inside the CRN can mislead the final result by sending false information such as reporting the presence of legacy user to occupy the spectrum himself, or hiding the existence of legacy user to cause collision. This attack is known as SSDF attack or Byzantine attack, it is described by figure 3.2.

Almost all related papers classify this issue as a sensing attack, but we consider it as a decision attack because it threatens the decision process after receiving the

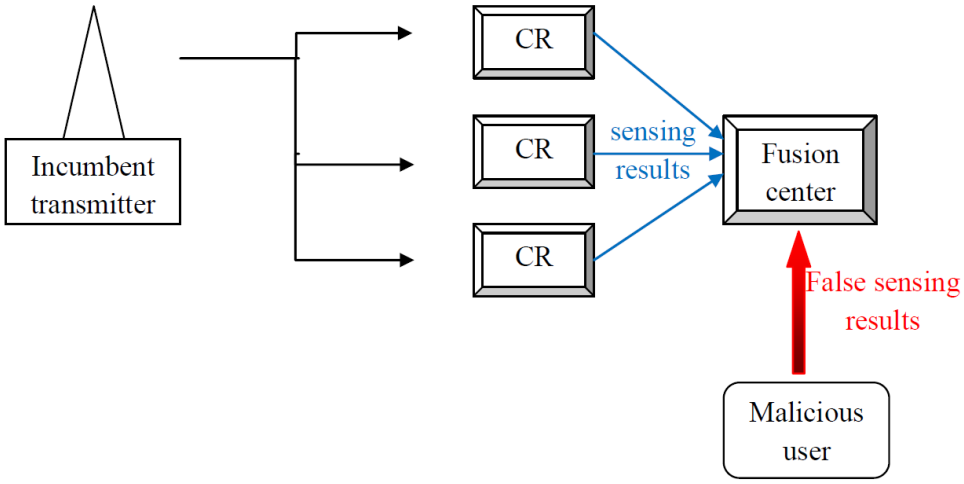


Figure 3.2: The SSDF attack

sensing reports. We present in the following paragraph the proposed solutions which can be classified into three categories, along with some related works:

3.4.1.1 Reputation based Approaches

These approaches consist in assigning suspicion levels to the cognitive nodes. If the suspicion level of any node exceeds a certain threshold, it is marked as malicious and removed from decision process. However, this method assumes that the base station has prior knowledge about the activities of attackers which is not very common. Without such information, the thresholds are approximated, resulting in significant false detections of attackers. In table 3.3, we present and discuss diverse proposed approaches to solve this problem.

3.4.1.2 Data Mining Approaches

Using these approaches, the fusion center have to intuitively interpret the received sensing report to decide to discard it if it is from a stealthy attacker. An abnormality detection in the reported data mining is used in [46] as approach to detect independent attackers. This approach starts by representing the history of reports of each CR by a point in the space. Then, it calculates the Hamming distance between each pair of two CRs and declares the presence of attackers when the distance deviates from a normal level. However, when attackers collaborate, they can successfully evade this detection approach.

More robust approaches can analyze particular pieces of sensing reports using a biweight estimate and median absolute deviation to calculate magnitudes, which are then compared against thresholds to identify the attackers [47]. The proposed method increases missed detections when using incorrect static thresholds because

inaccurately identified CRs could be excluded from the decision process. The correct setting of the detection thresholds can only be achieved with prior knowledge of attacker distribution which is unlikely to be available.

The detection of abnormal sensing reports can be combined with the verification of CRs locations, as proposed in [48]. In this reference, spatial correlation of the received signal strength among CRs is exploited to get the evidence whether the received signal strength is consistent with the location from where it is generated. Then, Dempster-Shafer theory is used to filter out abnormal reports by combining the evidence collected from the spatial correlation algorithm in each sensing period.

3.4.1.3 Artificial Intelligence Approaches

The decision process is susceptible to long-term manipulations caused by the extension of malicious inaccurate information which become a historical fact. To avoid the propagation of corrupted reports, learned values should be updated automatically by certain level of common sense. The authors in [7] proposed the use of swarm behavior in determining a global decision on whether a sensed signal was actually generated by a malicious user, along with a trust-based scheme.
The SSDF attack requires sending a falsified sensing report to the FC leading to wrong decision, but an attacker can also maximize his own gain (in the transmission power or in the spectrum access) by a simple manipulation of his utility function, as described in the following attack.

3.4.2 Biased utility function attack

The CR should adjust its transmission parameters according to the environment, such as its center frequency, bandwidth, power, modulation type, coding rate, channel access protocol, encryption type and frame size. According to [49], the radio might have three goals: low energy consumption, high data rate, and high security. Each of these goals has a different weight, which leads to a different objective function for each application.

The strategy of the biased utility function attack is the following: a malicious user can manipulate the transmission parameters to make the FC decision biased towards its benefit. For example, if a malicious user tweaks its utility function to transmit at higher power, it will result in other users getting less bandwidth. Some CRs may not even get to transmit. A scenario presented in [49] consists in an objective function composed of transmission rate (R) and security (S). An attacker may reduce the transmission rate by launching a jamming attack and hence reducing the overall objective function. Then, the CR will be forced to use a low security level and therefore it will be easily hacked. This attack is also known as objective function attack and belief-manipulation attack.

Game theory can be exploited to model utility function problems. For example, in [50], the authors propose an objective function to adjust CRs transmission power

with the constraint that the interference temperature due to the CRN transmissions on the incumbent receivers is below a given threshold. The problem is formulated as a public game and nash equilibrium solutions for a global optimum determines the transmit powers of the CRs.

There have been only few effective methods of mitigating objective function attacks. The paper [51] suggested defining thresholds for each of the adjustable parameters, so communications would be prevented when one or more of the parameters does not respect its predefined threshold. A secure spectrum decision protocol for a clustered infrastructure based network is proposed in [52]. In this solution, the spectrum decisions are made periodically and independently in each cluster. The suggested protocol consists of three steps: (1) Each node should communicate with the cluster head (CH) to join the decision process. It generates a sequence of symmetric keys using iterative application of a hash function to some initial value. The CH checks the authenticity of the message using the public key related to each node, then it stores the identity of the node and the related information if the verification is successful. The CH sends back a signed message including information about communication parameters such as the time/frequency schedule for submitting sensing data and available channels for ordinary communication. (2) The accepted nodes in each cluster send their spectrum sensing data to the CH using the key chain generated in the joining operation to protect the authenticity of the sensing information. The CH verifies the authenticity of the received messages to use it in the final decision. (3) The CH makes the decision and sends back the final channel assignment to the nodes of its cluster.

In this section, we have reviewed the common attacks to the spectrum decision function in cooperative CRNs. These attacks are harmful to cooperative military CRNs, such as in a scenario with coalition forces. It can lead to interferences with incumbent transmitters and prevent the efficient and secure spectrum access because of wrong decisions.

3.5 Spectrum Sharing Attacks and countermeasures

Generally, the management of the spectrum needs CCCs to coordinate the DSA and to exchange control messages such as local sensing reports. However, military CRNs may operate in dynamic spectrum-scarce and hostile environment. Therefore, CCCs could not be constantly availables to CRs for control message exchange and could be susceptible to malicious behavior such the jamming and saturation attacks.

3.5.1 Common Control Channel jamming attack

Cooperative CRNs use common control channels to achieve spectrum sharing. However, this channel is susceptible to jamming and saturation attacks. Jamming the control channel can disrupt the communication among CRs, resulting in packet

losses and sensing delays which may degrade the system performance. It can even lead to DoS, once the CCC is saturated by attackers. Among common approaches to mitigate CCC jamming attack, we find cross-channel control messages, random key distribution, channel hopping, intrusion defense strategies and game theory based solutions.

3.5.1.1 Cross-channel control messages

Using this defense approach, CRs continue to transmit on the jammed channel under interference to deceive the attacker and notify others about the new CCC for receiving control messages. As a result, the channels for transmitting and receiving control messages can be different to maintain the control message exchange with neighbors under jamming [53].

3.5.1.2 Random key distribution to hide CCC locations

Each CR has a valid key to be able to locate the allocated CCCs by using keyed hash functions. Any compromised node having only partial keys in the key space will not be able to jam all the CCCs. The random CCC key assignment reduces the risks of learning the key assignment structure from the attackers [53].

3.5.1.3 Channel hopping

The cluster heads are responsible of predetermining the hopping sequences for common control within the clusters. During the jamming attack, CRs hop on different sequences and communicate throw the predetermined CCC in the designated time slots without knowing the hopping sequences of others [54].

3.5.1.4 Intrusion defense strategies

Diverse intrusion detection techniques are exploited to mitigate the CCC jamming attack [55]. Here some examples:
(a) **Action Strategy Coordination (ASC)**
This technique is used to coordinate the action strategies among the CRs to establish a CCC by the exchange of a short control message including coordination parameters (current state, selected action and learning rate). The CRs update their action selection strategy with their own coordination parameters and those received from their neighbors. Such strategy increases the probability of selecting commonly available channels as control channels.
(b) **Best-Effort Cooperative Sensing (BCS)**
This technique can combat jamming attacks and enhance jamming resilience by reducing sensing errors. BCS is a distributed cooperative sensing scheme that CRs make the best efforts to share local sensing data with neighbors by using control links established in the previous stage yet still valid in the current stage, and individually make sensing decisions based on any collected sensing data. Unlike

conventional distributed sensing schemes, BCS does not require the participation of all neighbors or multiple iterations of message exchanges.
(c) **Deployment Density and Scalability**
It means the deployment of CRs in a given area. Increasing the deployment density in the jamming region leads to decrease of the average distance between the CRs while the average distance between the attacker and victims remains the same. This results in better SINR and less effective jamming perceived at the CRs.

3.5.1.5 Game theory exploitation

The interaction between the CCC jammer and the CRs can be modeled as a game and an optimal anti-jamming strategy may be found when the game reaches the Nash equilibrium. For example, the authors in [55] model the interactions of intelligent jammers and CRs as a stochastic general-sum game, called jamming-resilient control channel (JRCC) game. In this scenario, the CRN selects the optimal control channel allocation strategy by using an enhanced multiagent reinforcement learning (MARL) algorithm along with cooperative intrusion defense strategies.

Even, during a communication session, a jammer can intentionally and continuously transmit packets to prevent the CRs from exploiting the shared spectrum.This attack is known as intentional jamming attack.

3.5.2 Intentional Jamming attack

Malicious attackers may jam the current CR channel to make its signal-to-interference-plus-noise ratio (SINR) below the required threshold and to prevent it from an efficient exploitation of the spectrum. This attack can be amplified by jamming with high power in several spectral bands. Intentional jamming is one of the most easy attacks to happen in CRNs and can hinder both cognitive and non cognitive network communications. It can be a dangerous attack to CR military networks because it presents many challenges, such as the time taken to detect the malicious user and to mitigate the attack which effects severely the network performance and reliability. This attack can be carried out in several ways, and jammers can be classified according to the following criteria:

- Spot/Sweep/Barrage jamming:Spot jamming consists in attacking a specific frequency, while a sweeping jammer sweeps across an available frequency band and a barrage jammer will jam a range of frequencies at once.
- Single/Collaborative jamming:The jamming attack can be done by a single jammer or in a coordinated way between several jammers to gain more knowledge about the network and to efficiently reduce the throughput of the cognitive users.
- Constant/Random jamming:The jammer can either send jamming signals continuously on a specific channel or alternate between jamming and sleeping.

- Deceptive/Reactive jamming:A deceptive jammer continuously transmits signals in order to imitate a legitimate user. A reactive jammer transmits only when it detects busy channel to cause collisions.
- Power-Fixed/Adaptive jamming: A power-fixed attack has an invariable predefined power level regardless of the actual cognitive signal power and the surrounding radio environment. An adaptive jammer can employ estimation techniques and learning methods to adapt its transmitting power to the CR and channel parameters.
- Static/Mobile jamming: Static jammer has a fixed location, which can be revealed by using positioning techniques such as and ToA. A mobile jammer constantly changes its location to escape from localization.

More details about the classification of CRN jamming strategies are given in [56]. This reference deals with the problem of spectrum coordination between CRs in the presence of jammers.

The traditional anti-jamming solutions used in wireless networks consist in spread spectrum techniques by the use of either frequency hopping (FH) or direct-sequence spread spectrum (DS-SS) methods [57]. These solutions may be enhanced to mitigate the jamming attack in CRNs.

3.5.2.1 Frequency hopping

The CR is characterized by its ability of dynamic spectrum access which allows opportunistic use of the spectrum. This ability can be exploited to overcome jamming attacks since the CR can change its operating frequency to avoid the jammers. However, the exploitation of frequency hopping in CRN anti-jamming approaches present a trade-off between the resource consumption every time the CR hops from a jammed frequency and the jamming impact if the CR keeps using the same frequency even jammed.

Recently, diverse CRN frequency hopping defense strategies, were analyzed in [58]. It presented proactive or impetuous hopping (selecting a new set of frequencies at every slot, irrespective of the jamming) and reactive or conservative hopping (unjammed users keep the same frequencies for the next slot, while the jammed users choose a set of new unused frequencies that exclude the jammed ones). The authors proposed a multi-tier proxy based cooperative defense strategy, in which users form tiers to exploit the temporal and spatial diversity to avoid jamming. The jammer's success was based on selecting a channel to jam, that is in use by a regular node. To increase its chances of success, it might use the approach of equal power partial band spoofing, by distributing its transmit power budget among multiple randomly selected channels. The authors started by computing the effect of a single jammer on a single receiver, then summed the jamming signal strengths to compute the total interference as the collaborative jammers distribute their transmit power budget over multiple channels.

The behaviors of the CR, doing transitions between available frequencies, and the jammer, trying to prevent it from efficiently utilizing the spectrum, can be modeled using game theory. In this context, several works have been using game models and learning algorithms to find optimal anti-jamming strategy for the CR. For example in [59], the authors model the CRN jamming scenario as zero-sum game because of the opposite CR and jammer objectives. Furthermore, they implement the minimax-Q learning algorithm to find the optimal defense policy. Recently, the problem is formulated as a non-zero-sum game in [60], by taking into account different hopping and transmission costs, as well as diverse reward factors for both the transmitter and the jammer side. Authors make use of fictitious play learning algorithm to learn optimal defense strategy.

3.5.2.2 Direct-sequence spread spectrum (DSSS)

This spread spectrum technique consists in spreading the signal over several pieces of non-overlapping channels. It can be exploited as an anti-jamming technique because the jammer will have to choose either to jam a large number of channels with negligible jamming effect in each one or to jam only few channels with important effect. The authors, in [61], proposed an uncoordinated spread spectrum technique that enables anti-jamming broadcast communication without predefined shared secrets. They aimed to improve the common spread spectrum which depends on secret pairwise or group keys shared between the sender and the receivers before the communication, to adapt it for critical applications such as emergency alert broadcasts and military communications.

A random channel sharing was proposed, in [62], for broadcast CR communication to mitigate the insider jamming attack which resist to spread spectrum techniques. Spread spectrum has long been an effective technique to mitigate jamming attacks. However, in broadcast communication characterized by many receivers, once the attacker compromises a single receiver, he can discover which channels are in use and directly block those channels. The proposed idea is to organize receivers into multiple broadcast classified trusted/suspicious groups and use different channels for different groups. This ensures that a compromised receiver can only affect the members of the group it has been assigned to. A *'divide and conquer'* strategy is then used to isolate malicious receivers. The receivers are adaptively regrouped if the attacker launches a jamming attack so that the benign nodes are more likely to be merged into the trusted group, and the traitors are more likely to be included in a number of smaller suspicious groups. The random channel sharing improve the group-based approach by dynamically assign channels to groups such that different groups will randomly share their assigned channels. The data sent over the shared channel can reach more than one group, saving substantial communication cost. The receivers themselves do not know if their channels are shared with other receivers. Therefore, if a given channel is jammed and this channel is only assigned to one receiver, that receiver will be considered as one of the insiders. Thus, no matter when the insider chooses to jam the channel, there

is a chance that he will be detected and removed from future channel assignments. This scheme requires each receiver to listen to one channel at a time, instead of multiple channels.

3.5.2.3 Other anti-jamming techniques

In addition to approaches trying to evade the jammers, the CR can use coding techniques to mitigate the effect of the jamming attack on the transmitted signal. For example in [63, 64], a hybrid jamming mitigating approach is proposed to better handle the effect of malicious jamming nodes in the context of fault model classifications (including transmissive and omissive value faults due to the jamming attack) and their respective fault handling. A transmissive fault results from delivery of erroneous value to one or more receivers, and omissive fault results from failure to deliver any value to one or more receiver. The presented approach is based on a hybrid forward error correction (FEC) code defined by the concatenation of Raptor codes (used to regain lost data due to Omissive faults through data redundancy) and SHA-2 hash function (used to handle transmissive faults). Another coding approach is presented in [65], the authors combine random linear network coding with random channel hopping sequences to overcome the jamming effect on the transmitted control packets. Their proposed algorithm is called jamming evasive network coding neighbor discovery algorithm (JENNA).

Instead of using coding technique to repair the already jammed data, the concept of honeynode has been shown in [66] to be effective in deceiving jammers about the transmitting nodes. In this reference, a single honeynode is dynamically selected for each transmitting period, to act as a normal transmitting CR in order to attract the jammer to a specific channel.

Closed-form expressions to the jamming probabilities and the throughput of the CRN under various jamming attack models, was determined in [67] using the concept of Markov chain. Furthermore, the authors calculated the minimum and the maximum CRN throughput expressions under jamming, along with optimization of important anti-jamming parameters.

The authors in [59, 68, 69] use game theory to model the CRN jamming attack and apply reinforcement learning algorithms to learn how to avoid jammed channels. Other than learning anti-jamming channel selection, the authors in [70, 71] propose learning algorithms joining one channel selection and power control as anti-jamming strategy. Anti-jamming strategy based on multi-channel power allocation is studied in [72] using Colonel Blotto game and in [73] using a Bayesian approach.

The jamming attack has been widely exploited as strategic maneuver in military wireless communications. This problem has been intensively researched for traditional wireless networks but it is still a challenging issue in CRNs.

3.6 Spectrum Mobility Attacks and countermeasures

The CR have to vacate the current channel whenever it detects an activity of the incumbent user in the same channel. In order to establish smooth communication as soon as possible, the CR needs to select a new appropriate channel, and move immediately. This process is called spectrum mobility or hand-off.

An attack during spectrum mobility consists in forcing the CR to do handoff in wrong moment disturbing higher layers functions like routing protocols or security mechanisms. This problem is still less researched and most works are based on the assumption of successful handoffs. In this section, we enlist the common spectrum mobility attacks and we propose directions to future countermeasures.

3.6.1 Routing information Attack

A routing information attack is initiated when a malicious node causes spectrum handoff in the victim node just before it exchanges the routing information. During spectrum mobility, the victim node stops all ongoing communication, vacates the spectral band, opportunistically selects a new spectrum for transmission, scans the entire spectrum band to identify the neighboring nodes and informs them of the new frequency. Only, after all these operations, the CR can exchange the updated routing information with its neighbors. Until this period any path that goes through the victim node and its neighbors uses stale routing information.

One proposed solution to this attack is collision-free resident channel selection based solution. It consists in selecting a resident channel by each node from the available channel set during network initialization. It then broadcasts this selection with its neighbors. Nodes are expected to receive any updates on the resident channel. However, this protocol requires that each cognitive node is equipped with two half duplex transceivers with one waiting on the resident channel for a request of control message exchange, and the other using the data transmission channel [74].

3.6.2 Key Depletion Attack

Despite cryptographic measurements, the security of military CRNs could be degraded significantly with the important number of handoffs due to malicious issues. Frequent spectrum handoffs result in multiple sessions needed for any given application, and hence large number of cryptographic keys is used at the beginning of every transport layer session. Therefore, the probability of using the same key twice will increase. Key repetitions can be exploited to break the cipher system.

For example, the wired equivalent privacy and the temporal key integrity protocols used in IEEE 802.11 link layer are vulnerable to key repetition attacks [75]. The counter cipher mode with block chaining message authentication code protocol is designed to exponentially delay key repetitions. It offers enhanced security compared to TKIP by using 128-bit keys with a 48-bit initialization vector. It takes

128 bit key blocks of data through the AES encryption standard and uses WPA1 and WPA2 to allow for a quick handoff cipher block [76].

The security of the spectrum mobility function is still a challenging issue and an interesting road of research. Other security measurements should be added to the cryptographic algorithms to enhance the resistance against the handoff process attacks.

3.7 Discussion

The exploitation of CRs in complex scenarios such as in a military ad hoc deployment requires a high focus on the cognitive engine, since simple spectrum access algorithms may be vulnerable to malicious activities such as IE attack during SS, FC attacks for centralized spectrum decision, spectrum handoff attacks and the jamming attack. According to the bibliographic study presented in this chapter, game theory has been exploited to model the interaction between the attacker and the CR and helps to find a solution to each attack.

Let us compare the presented attacks in terms of both harmfulness of each attack and knowledge requirements to launch the malicious activity:
1-The IE attack requires knowledge about the incumbent user to succeed in its emulation, without such knowledge the attacker is said to be a jammer since it will transmit its own jamming signal. It is said to be a selfish user if it aims to increase its share of the spectrum resource, and a malicious attacker if on top of being selfish it aims also to interrupt the CRN current service.
2-The FC attacks such as SSDF and objective function attacks require that the attacker is a member of the CRN (inside the network) to be able to send an acceptable SS report to the FC or to alter the objective function.
3-The Spectrum handoff attacks are results of forcing the CR to do frequency hopping, which may also results from the jamming attack, so solving the jamming attack may reduce handoff decisions.
4-The jamming attack is easy to implement because a jammer doesn't need to be member of the existing network or to gather information neither about the incumbent network nor the CRN. The jamming attack may result in many troubles in military CRNs, not only for the spectrum sharing function, but also for the other steps of the cognition cycle. The potential consequences and impacts on each step are:

- Sensing step
  Through periodic spectrum sensing, the CRs are able to exploit frequency white spaces. However, jammers may steal these resources.

- Decision step
  It is not always possible to identify correctly the true legitimate user from the jammer which leads to wrong decisions and interferences with other users.

- Sharing step
  Jammers may inject interference signal to disturb on-going communications and may saturate the common control channel which can lead to denial of service (DoS).

- Mobility step
  Jammers may force CRs to change the operating channels, which can lead to frequent spectrum handoff inducing QoS degradation and increasing connection unreliability.

Generally, the attacker must have some information (such as the protocol or architecture) about the network or even be part of it to carry out its attack. Since this is difficult for the case of military communications, the choice of attack may be reduced to the jamming attack. For that and for its harmful impacts, we choose to work on it. The jamming attack can be also exploited in order to disturb enemy communications or to deceive eavesdroppers.

3.8 Conclusion

In this chapter, we tried to give a broad comprehensive review of CRN attacks along with related works with the focus on tactical military applications. We have differently classified intrinsic CR threats according to the main functions of the CR (sensing, decision, sharing and mobility) to better understand the effect of each attack. Finally, we have compared the threats in terms of harmfulness and required information, to explain our choice of studying the jamming attack.

In the reminder of this thesis, we model the interaction between the CR and a jammer as a game. In the next chapter, we develop the optimal power allocation for both of the players under complete information then we propose a learning algorithm to solve real scenarios.

Approach	Description	Paper
Onion peeling	Reputation values based on estimations or Bayesian statistics	[39], [40]
Weighted sequential probability ratio test	-Combines reputation and SPRT to identify malicious nodes -Outperforms standard FC decision making strategies (e.g. OR, AND) in both minimizing missed detections and maximizing the correct sensing ratio.	[41]
Game theory	-Zero-sum game between attackers and the FC -Use of minimax approach to find optimal defense strategy -KLD and error probability at the FC considered as performance metrics to characterize detection performance. -Performance limits boundaries are established for independent and cooperative attackers	[42]
Dempster-Shafer	The trust value is based on direct and indirect users observations -Degrade the impact of malicious entities during distributed cooperative spectrum sensing	[43]
Adaptative reputation based clustering algorithm	-Clustering the nodes according to the sensing history and initial reputations. -Then each node is assigned a positive or negative share based on its participation in the final decision to adjust its reputation. -The adjusted reputations are used to adjust the number of clusters for the next step.	[44]
SVDD algorithm	-SVDD is a kind of one-class classification method based on Support Vector Machine and described by a few target objects, known as support vectors. -It tries to construct the boundary around the target data enclosed within a minimum hyper-sphere. -Then the algorithm votes between trusted nodes to decide whether the spectrum is empty.	[45]

Table 3.3: SSDF reputation based countermeasures

Chapter 4

Optimal power allocation in cognitive radio and jammer games

4.1 Introduction

A cognitive radio device can adjust its power level intelligently in real time according to the wireless environment changes. Until recently, the problem of optimal power allocation in CRNs has been studied to solve either the primary and secondary users (PUs/SUs) coexistence, or the spectrum sharing between the CR users. In both scenarios, the nodes have the same goal which consists in maximizing the same utility function (capacity, SINR, spectrum exploitation, etc) and avoiding unintentional interference. However, the CR technology can be exploited by malicious users to prevent the efficient management of the available frequency bands. A jammer may be able to adjust its power allocation across tones in order to cause maximal intentional interference and harm the communication in the most efficient way. Game theory is a suitable model to this scenario since it studies decision making where the best course of a player's action depends upon the decisions made by others. Power allocation games against a jammer have been studied in some recent works for both wireless communication networks and cognitive networks. Most of related papers proved the existence and uniqueness of the pure strategy Nash equilibrium (NE) considering finite action sets without dealing with analytic computation of the optimal strategies.

In this chapter, we start by the key terms definition and the related works. In section 4.4, we model the interaction between a CR pair (a transmitter-receiver pair) and a jammer as a two players zero-sum game with the transmission capacity as the utility function. We consider that the actions of both players can be selected from continuous sets. In section 4.5, we consider no interaction between jammer and CR and we model their actions as two unilateral games. In each unilateral game, we consider only one player as the unique decision maker. The other player

has a fixed strategy. In section 4.6, we consider the Nash game in which some player moves first, the other player observes the choice made and then adapts his power allocation. The game consists in playing the two unilateral games alternately until reaching the Nash equilibrium. We determine this equilibrium in terms of pure strategy. Then, we study Stackelberg game in section 4.7 where the first player (the leader) has knowledge of the follower's reaction function and makes the optimal decision reaching the Stackelberg equilibrium. Furthermore, we determine the maxmin and minmax optimal power allocations for the CR user and the jammer under complete knowledge in finite action subsets. The simulation results give equality of the Nash equilibrium, the Stackelberg equilibrium and the minmax/maxmin optimal power allocations. In section 4.9, we theoretically prove the existence and uniqueness of this equilibrium which gives the saddle point of the considered two-person zero-sum game. In section 4.10, we derive the analytical expressions for the optimal strategies characterizing the saddle point, under the assumption that all the channels are exploited by both players. Finally, we provide and discuss all the simulation results in section 4.11.

4.2 Key terms

This chapter is based on game theory and uses key terms such as waterfilling that will be defined as follows.

- **Game theory**

It provides mathematical models of both conflict and cooperation between rational decision makers where the outcome of one player depends on the others' actions. The model includes: the players, the actions (or strategies) set and the utility function of each player. Cooperative games deal with logical decision making between allied players to gain collective payoffs. Non-cooperative games model individual participants' payoffs analyzing Nash equilibrium which corresponds to the solution of the game and specifies the equilibrium strategies of the players. The NE corresponds to a stable state in which no player can gain from unilaterally deviating from its strategy. A pure strategy NE is a deterministic NE that specifies the optimal action for each player among its set of actions, but a mixed strategy NE assigns a probability to the available actions of each player. When the gain of one player results in loss for the other one, the model is said to be zero-sum game. When players may choose a strategy from continuous strategy sets, the game is said to be continuous. More details about game categories are given in [77]. We adopt a non cooperative game-theoretic model to study the power allocation problem since the CR users and the jammer are rational and selfish, each player is interested in maximizing its own utility. A pure strategy NE to the power allocation CR and jammer game should specify the optimal power allocation for each player.

- **Waterfilling**

It is an optimal multi-channel power allocation technique providing closed form solution for capacity maximization [78]. The term waterfilling is used since the communication medium looks like a water container with an uneven bottom and each channel is like a container section having its own depth. Power allocation over the available channels is similar to water pouring into the container, as given in figure 4.1. The allocated power to each channel depends on its noise level with respect to the total power defining the water level.

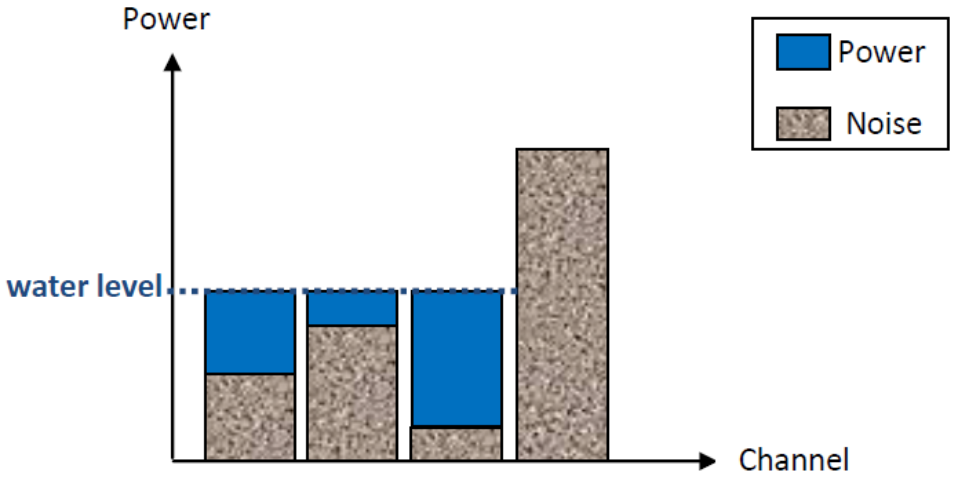


Figure 4.1: Waterfilling definition

4.3 Related works

Diverse works deal with the power allocation problem in cognitive radio networks without considering malicious users; In [79], the goal is to maximize the weighted sum effective capacities of the SUs in the presence of PUs. The authors determined the optimal power allocation through a convex optimization method using Lagrangian functions with respect to Karush Kuhn-Tucker (KKT) conditions. The authors in [80] studied the impact of channel correlation on the optimal power allocation strategy. Multiple input single output antenna techniques and antenna selection techniques are studied in [81] to combat the interference constraint and improve the capacity of the SU. The problem in [82], is modeled as a partially observable Markov decision process (POMDP) and the optimal policy is derived for relay selection, channel access, and power allocation through a dynamic programming approach. In order to maximize its transmission rate, each SU in [83] applies a waterfilling scheme and uses the greedy asynchronous distributed interference avoidance algorithm to solve the mutual interference problem. The approach is based on the dynamic adjustment of the number of used frequencies by

each user. The problem of power and chunk-based resource allocation is investigated in [84] to maximize the energy efficiency of a multi-carrier CRN. Using Dinkelbach method from non-linear fractional programming and dual optimization method, the authors developed an iterative algorithm to optimize both of the power allocation and chunk-based resource allocation.

Considering non-cooperative game against a jammer, the power allocation problem has been studied in [85] for MIMO radar system and in [86–90] for wireless communication networks with diverse utility functions such as the SINR, transmission capacity and number of successful channel access. In [85], the interaction between a smart target and a smart MIMO radar is modeled as a two-person zero-sum game. The unilateral, hierarchical, and symmetric power allocation games are studied based on the information set available for each player, and the equilibrium solutions are derived. In [86] and [87], Altman proved the existence and uniqueness of NE considering the transmission capacity as the utility function. To develop the closed form analytic expressions of the optimal power allocations, in the first paper [86] he proposed an algorithm based on the bisection method. In the second paper [87], he converted the problem to a minimax problem since the NE strategy of a zero-sum game is equal to the optimal minimax strategy [91], and he considered the particular case of proportional channel fading coefficients faced by both the jammer and the transmitter. [89] paper can be considered as a generalization of Altman's work to a game scenario between K users and a jammer. The authors develop a generalized version of the iterative waterfilling algorithm whereby all of the users and also the jammer update their power allocations in a greedy manner in order to maximize their respective utilities. Considering finite strategy sets for both the transmitter and the jammer, the authors in [90] prove the existence of NE in pure (deterministic) strategies and characterize the optimal power allocations in asymptotic regimes over independent parallel Gaussian wiretap channels where a legitimate transmitter and a legitimate receiver communicate in the presence of an eavesdropper and a jammer.

The power allocation interaction between a jammer and a CR user were studied in [72, 73, 92]. The problem is presented in [72] as Colonel Blotto game where the two opponents distribute limited resources over a number of battlefields with the payoff equal to SINR, and the equilibrium is derived in terms of mixed (probabilistic) strategy via power randomization. Likewise, the authors in [73] adopt a Bayesian approach in studying the power allocation game between the CR user and the jammer, and provide the Cumulative Distribution Functions (CDFs) of the transmission powers that should be adopted by the CR user and the jammer at NE to optimize the utility function equal to the number of successful transmissions.

4.4 System model

We consider that the CR user has the capacity of accessing multiple frequency bands at the same time with a limited power budget. This scenario is possible

for example by using the OFDM modulation. Each jammer is also assumed to be able to inject interference to all channels which is known as barrage jamming. The scenario is given in figure 4.2. The CR user adopts the 'listen-before-talk'

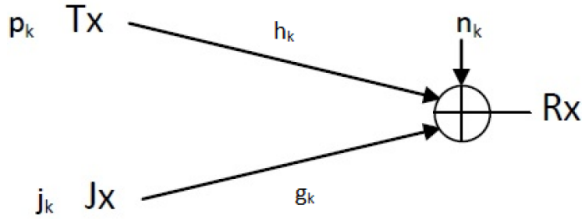


Figure 4.2: Scenario of CR jamming attack

rule, that is, sensing for spectrum opportunities at the beginning of each time slot. We consider M available channels, it allocates a power $p_k \geq 0$ to each channel k$\in [1, M]$ such that:

$$\sum_{k=1}^{M} p_k \leq P \tag{4.1}$$

An action of the CR user is designed by the vector **p**$= (p_1, \cdots, p_k, \cdots, p_M)$ in order to maximize its transmission capacity subject to (4.1) with P as the total power. At the same time, the jammer injects power $j_k \geq 0$ to the channel k such that:

$$\sum_{k=1}^{M} j_k \leq J \tag{4.2}$$

An action of the jammer is designed by the vector **j**$= (j_1, \cdots, j_k, \cdots, j_M)$ in order to minimize the transmission capacity of the CR user, subject to (4.2) with J as the total power. We use n_k to denote the noise variance of channel k, h_k and g_k to denote the complex gains of channel k for the CR user and the jammer respectively. We assume that all channel gains are common knowledge to both players, and we consider that the M channels are parallel Gaussian channels.
The Shannon capacity is proportional to

$$F(\mathbf{p}, \mathbf{j}) = \sum_{k=1}^{M} log_2(1 + \frac{|h_k|^2 p_k}{|g_k|^2 j_k + n_k}). \tag{4.3}$$

We consider F$(\mathbf{p}, \mathbf{j})$ $(-F(\mathbf{p}, \mathbf{j}))$ the utility function of the CR user (the jammer). The CR user is trying to maximize its total transmission capacity over the available channels and the jammer is trying to minimize this capacity, so their interaction can be seen as a two person zero-sum game. Provided that each element of the vectors **p** and **j** can take any value in $[0, P]$ and $[0, J]$, we have continuous set of actions for the two players.

In the remainder of this chapter, we will study diverse scenarios of the game between the two players to find the optimal power allocations. All simulations are grouped in section 4.11.

4.5 Unilateral games

We start by considering the two following extreme cases where only a player (the CR or the jammer) has to decide how to allocate his total power against an opponent having a fixed strategy.

4.5.1 CR user Unilateral Game

If the jammer's strategy is fixed, the game degenerates to a classical power allocation problem where the CR user chooses its power according to the noise plus jamming level in order to maximize the capacity. Mathematically, it can be formulated as the following nonlinear optimization problem:

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \sum_{k=1}^{M} log_2(1 + \frac{|h_k|^2 p_k}{|g_k|^2 j_k + n_k}) \\ \text{subject to} \quad & \sum_{k=1}^{M} p_k \leq P \end{aligned} \tag{4.4}$$

Allowing inequality constraints, the KKT approach generalizes the method of Lagrange multipliers to nonlinear programming. The Lagrangian is then,

$$L(\mathbf{p}, \mathbf{j}, \lambda) = \sum_{k=1}^{M} log_2(1 + \frac{|h_k|^2 p_k}{|g_k|^2 j_k + n_k}) - \lambda(\sum_{k=1}^{M} p_k - P) \tag{4.5}$$

Since L is separable in p_k, we can separately optimize each term.

$$\frac{\partial L}{\partial p_k} = \frac{|h_k|^2}{|h_k|^2 p_k + |g_k|^2 j_k + n_k} - \lambda \tag{4.6}$$

The optimal solution of this optimization problem yields the following strategy

$$p_k^* = (\frac{1}{\lambda} - N_k)^+ \tag{4.7}$$

known as waterfilling strategy, where $\frac{1}{\lambda}$ is the waterlevel. The KKT multiplier $\lambda > 0$ can be found by bisection and should satisfy

$$\sum_k (\frac{1}{\lambda} - N_k)^+ = P, \tag{4.8}$$

where $(x)^+ = \max(0, x)$ and N_k is the fictive noise power on each channel given as

$$N_k = \frac{|g_k|^2 j_k^* + n_k}{|h_k|^2} \tag{4.9}$$

Note that to find the optimal solution, the CR needs to know the channel gain g_k between the jammer and the receiver CR. In practical cases, this information is not known. This remark is valid for all the following sections.

4.5.2 Jammer Unilateral Game

On the other hand, suppose that the CR user has a fixed power allocation strategy. The game degenerates to a jamming unilateral optimization, as the CR user is not aware of this. In such a circumstance, the jammer will allocate its jamming power to minimize the total capacity. Mathematically, this is expressed as the following minimizing problem

$$\begin{aligned} &\underset{\mathbf{j}}{\text{minimize}} \quad F(\mathbf{p}, \mathbf{j}) \\ &\text{subject to} \quad \sum_{k=1}^{M} j_k \leq J \end{aligned} \tag{4.10}$$

We can write the Lagrangian as

$$\mathrm{L}(\mathbf{j}, \mu) = -F(\mathbf{p}, \mathbf{j}) - \mu(\textstyle\sum_{k=1}^{M} j_k - J) \tag{4.11}$$

Since L is separable in j_k, we can separately minimize each term as shown below

$$\frac{\partial L}{\partial j_k} = \frac{|g_k|^2 |h_k|^2 p_k}{(|h_k|^2 p_k + |g_k|^2 j_k + n_k)(|g_k|^2 j_k + n_k)} - \mu \tag{4.12}$$

After solving the resulting second order equation in j_k, we get the optimal solution

$$j_k^{*=} \left(\frac{1}{2} \sqrt{(\frac{|h_k|^2 p_k}{|g_k|^2})^2 + 4\frac{|h_k|^2 p_k}{|g_k|^2 \mu}} - \frac{|h_k|^2 p_k}{2|g_k|^2} - \frac{n_k}{|g_k|^2} \right)^+ \tag{4.13}$$

where the KKT multiplier μ is the solution of

$$\sum_{k=1}^{M} j_k^* \leq J \tag{4.14}$$

and can be found by bisection.

Unlike the CR user who uses the waterfilling strategy, the jammer applies a different strategy to dynamically allocate its power (as given in equation (4.13)).

4.6 Nash game

After solving the optimization problems independently for the CR user and the jammer, we consider here a sequential-moves game in which both the CR user and the jammer make decisions but sequentially. In game theory, a game is said to be sequential if the players choose their actions in a consecutive way and the latter player requires information about the former. The main issue is the convergence of this continuous game to a Nash equilibrium at which no player has interest in changing the power allocation. The theoretical proof of existence and uniqueness of the NE is shown in section 4.9, according to references [93] and [94].

Since $\mathbf{p}^*$ maximizes $F(\mathbf{p}, \mathbf{j}^*)$ and $\mathbf{j}^*$ minimizes $F(\mathbf{p}^*, \mathbf{j})$, we alternatively determine the CR user's power for a given jamming action, then compute the minimizing jamming power for the CR user's action. That is starting with an initial value $\mathbf{j}_0$, we perform a bisection to determine $\mathbf{p}$. Then for this $\mathbf{p}$, compute the vector $\mathbf{j}$ which minimizes $F(\mathbf{p}, \mathbf{j})$ and these two steps will be repeated.

This game implements the two unilateral games presented in section 4.5 in an iterative way until convergence to a fixed power allocation per channel within a specific tolerance ϵ. The CR user applies the waterfilling strategy, we proceed by bisection until reaching the value of λ corresponding to the allocation of the total CR user's power (equation (4.7)). For the jammer, we exploit another strategy and we proceed by bisection until reaching the value of μ corresponding to the allocation of the total jamming power (equation (4.13)).

4.7 Stackelberg game

In the previous section, we have considered a sequential-moves game played over time which is usually applied either when the rules of the game are unknown or when directly solving is difficult [95]. In this section we will consider a sequential one-shot game known as a Stackelberg game, in which the leader should anticipate the follower's reaction function in order to alleviate the worst case. Subsequently, the follower observes the action taken by the leader and plays an unilateral game. The solution for this scenario is known as Stackelberg equilibrium and can be found by backward induction. We start by determining the action of the follower, then we derive that of the leader. We will start by a scenario in which the jammer is the leader, then we will solve the game in which the CR user is the leader.

4.7.1 The jammer as the leader

In this scenario, the jammer knows the reaction function of the CR user which is given in equation (4.7), and he should use it to substitute p_k in his minimizing problem (equation (4.10)) to find his optimal power allocation.

Replacing p_k with the expression (4.7), we get:

$$\begin{aligned} log_2(1+\frac{|h_k|^2 p_k}{|g_k|^2 j_k+n_k}) &= log_2(1+\frac{\frac{|h_k|^2}{\lambda}-|g_k|^2 j_k-n_k}{|g_k|^2 j_k+n_k}) \\ &= log_2(\frac{|h_k|^2}{\lambda(|g_k|^2 j_k+n_k)}) \end{aligned} \tag{4.15}$$

We should get the expression of λ as a function of j_k. For that, the jammer have to consider the following constraint:

$$\sum_{k=1}^{M} p_k = P, \tag{4.16}$$

which results in,

$$\sum_{k=1}^{M}(\frac{1}{\lambda}-\frac{|g_k|^2 j_k+n_k}{|h_k|^2}) = P. \tag{4.17}$$

So, we obtain

$$\lambda = \frac{M}{P+\sum_k \frac{g_k^2 j_k+n_k}{h_k^2}}, \tag{4.18}$$

and the minimizing problem of the jammer becomes

$$\begin{aligned} \underset{\mathbf{j}}{\text{minimize}} \quad & \sum_{k=1}^{M} log_2(\frac{|h_k|^2(P+\sum_k \frac{g_k^2 j_k+n_k}{h_k^2})}{M(|g_k|^2 j_k+n_k)}) \\ \text{subject to} \quad & \sum_{k=1}^{M} j_k \leq J \end{aligned} \tag{4.19}$$

Since the utility function is no longer separable in j_k, we can no longer derive the Lagrangian expression independently for each j_k. To solve this optimization problem, the jammer can apply a one dimensional exhaustive search over his possible power allocations to find the optimal vector $\mathbf{j}$ minimizing this new capacity expression, since it is no longer function of $(\mathbf{p}, \mathbf{j})$, it is only a function of $\mathbf{j}$. Then, the CR user (as follower) has to exploit the available information about the jammer's power allocation in order to maximize his total capacity. Hence, he plays the unilateral game described in subsection 4.5.1.

4.7.2 The CR as the leader

In this scenario, the CR user knows the reaction function of the jammer which is given in equation (4.13), and he should use it to replace j_k in his maximizing

problem (4.4) to find the optimal power allocation.
To simplify, we replace j_k with $U(p_k)$ which is equal to the expression (4.13):

$$U(p_k) = \frac{1}{2}\sqrt{(\frac{|h_k|^2 p_k}{|g_k|^2})^2 + 4\frac{|h_k|^2 p_k}{|g_k|^2 \mu}} - \frac{|h_k|^2 p_k}{2|g_k|^2} - \frac{n_k}{|g_k|^2}. \tag{4.20}$$

Subsequently, we get as new utility function:

$$log_2(1 + \frac{|h_k|^2 p_k}{|g_k|^2 j_k + n_k}) = log_2(1 + \frac{|h_k|^2 p_k}{|g_k|^2 U(p_k) + n_k}). \tag{4.21}$$

In $U(p_k)$ we have a Lagrangian parameter which is μ that depends in p_k and we should determine its closed form expression. For that, we have to solve the equation

$$\sum_{k=1}^{M} U(p_k) = J \tag{4.22}$$

Even it is a complicated equation especially because μ is inside the square root, we can remark that μ is function of all the p_k, $\forall\ k \in [1, M]$ so the utility function of the CR user is not separable in p_k and we can't derive it with respect to each p_k independently.
The maximizing problem (4.4) of the CR user becomes

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \sum_{k=1}^{M} log_2(1 + \frac{|h_k|^2 p_k}{|g_k|^2 U(p_k) + n_k}) \\ \text{subject to} \quad & \sum_{k=1}^{M} U(p_k) = J \\ \text{subject to} \quad & \sum_{k=1}^{M} p_k \leq P \end{aligned} \tag{4.23}$$

We solve this maximizing problem through exhaustive search over all possible power allocations **p** that respect the second constraint. We start by determining μ for each possible **p** through bisection with respect to the first constraint, then using this μ value we calculate the corresponding utility function. The optimal power allocation $\mathbf{p}^*$ corresponds to the maximizer of this function. For the jammer (as follower), we implement the expression (4.13) found in the jammer unilateral game since he can observe the CR user's power allocation.

4.8 Optimal solution: minmax/maxmin strategies

Here we consider the perfect scenario of complete knowledge and we define finite action sets for the two players. The minmax search is especially known for its usefulness in calculating the best move in two-player games where all the information

is available. Each player in this game, knows that its strategy will be intercepted by its opponent. By considering conservativeness and rationality assumptions of the Minimax theorem [96], each player may adopt the strategy which can alleviate the worst case. This means that the strategy of the jammer is the minimizer to the maximum payoff of the CR user, it is also the minimizer to his own maximum loss (since the game is zero-sum). likewise, the CR user's strategy is the maximizer to the worst case (i.e. maximize the minimum payoff of the CR user). So, neither the CR user nor the jammer will profit when changing its strategy and moving from the equilibrium.

4.8.1 The CR user's maxmin strategy

Consider that the jammer is able to sense the CR user's power allocation and that the CR user is aware of it. Then, a conservative CR user may select its strategy based on the following optimization problem:

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}}\ \underset{\mathbf{j}}{\text{minimize}} \quad & F(\mathbf{p},\mathbf{j}) \\ \text{subject to} \quad & \sum_{k=1}^{M} p_k \leq P, \\ & \sum_{k=1}^{M} j_k \leq J, \\ & \mathbf{p} > 0, \mathbf{j} > 0, \end{aligned} \tag{4.24}$$

to maximize the capacity in the worst case (i.e. in the situation where the jammer plays the strategy which cause the greatest harm to the CR user).

4.8.2 The jammer's minmax strategy

The CR device possesses sufficient interception capacity that it can sense interference, due to its wideband spectrum sensing capacity. If the jammer behaves in a conservative way, he will distribute his power as to minimize the possible maximum capacity, which corresponds to solving the two-stage optimization problem [97],

$$\begin{aligned} \underset{\mathbf{j}}{\text{minimize}}\ \underset{\mathbf{p}}{\text{maximize}} \quad & F(\mathbf{p},\mathbf{j}) \\ \text{subject to} \quad & \sum_{k=1}^{M} p_k \leq P, \\ & \sum_{k=1}^{M} j_k \leq J, \\ & \mathbf{p} > 0, \mathbf{j} > 0 \end{aligned} \tag{4.25}$$

We implemented this scenario using exhaustive search over a finite set of possible power allocations. We calculate a matrix of capacity values, its rows are the possible jammer's power allocations and its columns are the CR user's power allocations. For the CR user's maxmin, we determine a row of minimum capacity values over all the rows, and finally we determine the column corresponding to the maximum value in this row of minimums. And for the jammer's minmax, we determine a column of maximum capacity values over all the columns, and finally we determine the row corresponding to the minimum value in this column of maximums.

We have presented the jamming game in diverse scenarios (unilateral, Nash, Stackelberg, minmax/maxmin) and studied the equilibrium of each one. Before providing the simulation results (section 4.11), we will prove theoretically the existence and uniqueness of the pure strategy equilibrium for this game (section 4.9) and give its closed form expression (section 4.10).

4.9 Proof of the existence and uniqueness of the equilibrium in pure strategies

In this section, the jamming scenario is described as a two-player zero-sum game with continuous action sets. The existence of Nash Equilibrium can be proved from the properties of the action sets and the utility functions:

- The action sets, $[0, P]^M$ and $[0, J]^M$ are non-empty convex and compact.
- The utility functions are continuous in $(\mathbf{p}, \mathbf{j})$.

So, this game is said to be a continuous game for which the NE is guaranteed [98], but we have to determine if the NE exists in pure strategies or mixed strategies.

4.9.1 Existence of Nash equilibrium in pure strategies

According to the definition of quasi-convex and quasi-concave utility functions given in [99], the utility function $F(\mathbf{p}, \mathbf{j})$ is quasi-concave in $\mathbf{p}$ and quasi-convex in $\mathbf{j}$.

We can conclude that we have a non-empty compact convex action sets and the utility function is continuous, quasi-concave in $\mathbf{p}$ and quasi-convex in $\mathbf{j}$. Then according to [98, 100],

$$sup_{\mathbf{p}\in A} inf_{\mathbf{j}\in B} F(\mathbf{p}, \mathbf{j}) = inf_{\mathbf{j}\in B} sup_{\mathbf{p}\in A} F(\mathbf{p}, \mathbf{j}), \tag{4.26}$$

which is equal to the optimal value of the game. So, this game has a Nash equilibrium in pure strategies.

The uniqueness of the NE issue with continuous action sets will be proved by verifying the sufficient condition of diagonally strictly concavity, given in [93].

4.9.2 Uniqueness of the Nash equilibrium in pure strategies

Let's define the pseudo-gradient vector [101]

$$\mathbf{gr}(\mathbf{p},\mathbf{j}) = \left[\nabla_{\mathbf{p}} u_{CR}(\mathbf{p},\mathbf{j}), \nabla_{\mathbf{j}} u_{Jx}(\mathbf{p},\mathbf{j})\right]^T \tag{4.27}$$

Where, u_{CR} and u_{Jx} are respectively the utility functions of the CR user and the jammer verifying: $u_{CR} = -u_{Jx} = F(\mathbf{p},\mathbf{j})$ and their gradient vectors are

$$\nabla_{\mathbf{p}} u_{CR}(\mathbf{p},\mathbf{j}) = \nabla_{\mathbf{p}} F(\mathbf{p},\mathbf{j}) \quad = \left[\frac{\partial F}{\partial p_1}, \cdots, \frac{\partial F}{\partial p_k}, \cdots, \frac{\partial F}{\partial p_M}\right]^T \tag{4.28}$$

and

$$\nabla_{\mathbf{j}} u_{Jx}(\mathbf{p},\mathbf{j}) = -\nabla_{\mathbf{j}} F(\mathbf{p},\mathbf{j}) \quad = -\left[\frac{\partial F}{\partial j_1}, \cdots, \frac{\partial F}{\partial j_k}, \cdots, \frac{\partial F}{\partial j_M}\right]^T \tag{4.29}$$

Let $\mathbf{G}(\mathbf{p},\mathbf{j})$ denote the Jacobian of the pseudo-gradient $\mathbf{gr}(\mathbf{p},\mathbf{j})$. To justify the diagonally strictly concavity (DSC) condition, we have to prove that the symmetric matrix $(\mathbf{G}(\mathbf{p},\mathbf{j}) + \mathbf{G}^T(\mathbf{p},\mathbf{j}))$ is negative definite for all possible $(\mathbf{p},\mathbf{j})$, which is a sufficient condition [93], $\mathbf{G}^T$ is the transpose of the matrix $\mathbf{G}$.

$\mathbf{G}(\mathbf{p},\mathbf{j})$ is $2M * 2M$ matrix, in which the first M columns are the partial derivatives of $\mathbf{gr}(\mathbf{p},\mathbf{j})$ with respect to the M elements of the vector $\mathbf{p}$ and the second M columns are its partial derivatives with respect to the vector $\mathbf{j}$, so we can represent the matrix $\mathbf{G} = (g_{lc})_{1\leq l,c\leq 2M}$ using four $M * M$ sub-matrices

$$\mathbf{G}(\mathbf{p},\mathbf{j}) = \begin{bmatrix} [\mathbf{A}] & [\mathbf{B}] \\ [\mathbf{C}] & [\mathbf{D}] \end{bmatrix} \tag{4.30}$$

let's give the expressions of these sub-matrices, using l to denote the row index and c for the column index

- $\forall 1 \leq l, c \leq M$, (the submatrix A)

$$\begin{aligned} g_{lc} = a_{lc} &= \frac{\partial^2 F(\mathbf{p},\mathbf{j})}{\partial p_l \partial p_c} \\ &= \begin{cases} -\left(\frac{h_l^2}{h_l^2 p_l + g_l^2 j_l + n_l}\right)^2 & if \quad c = l \\ 0 & else \end{cases} \end{aligned} \tag{4.31}$$

- $\forall M+1 \leq l, c \leq 2M$, let's $x = l - M$ and $y = c - M$, (the submatrix D)

$$\begin{aligned} g_{lc} = d_{xy} &= -\frac{\partial^2 F(\mathbf{p},\mathbf{j})}{\partial j_x \partial j_y} \\ &= \begin{cases} -\frac{h_x^2 g_x^2 p_x (2 g_x^2 (g_x^2 j_x + n_x) + h_x^2 g_x^2 p_x)}{(h_x^2 p_x + g_x^2 j_x + n_x)^2 (g_x^2 j_x + n_x)^2} & if \quad x = y \\ 0 & else \end{cases} \end{aligned} \tag{4.32}$$

- $\forall M+1 \leq l \leq 2M$ and $1 \leq c \leq M$, let's $x = l - M$, (the submatrix C)

$$\begin{aligned} g_{lc} = c_{xc} &= -\frac{\partial^2 F(\mathbf{p},\mathbf{j})}{\partial j_x \partial p_c} \\ &= \begin{cases} \frac{h_x^2 g_x^2}{(h_x^2 p_x + g_x^2 j_x + n_x)^2} & if \quad c = x \\ 0 & else \end{cases} \end{aligned} \tag{4.33}$$

- $\forall 1 \leq l \leq M$ and $M+1 \leq c \leq 2M$, let's $y = c - M$, (the submatrix B)

$$\begin{aligned} g_{lc} = b_{ly} &= \frac{\partial^2 F(\mathbf{p},\mathbf{j})}{\partial p_l \partial j_y} \\ &= \begin{cases} -\frac{h_l^2 g_l^2}{(h_l^2 p_l + g_l^2 j_l + n_l)^2} & if \quad y = l \\ 0 & else \end{cases} \end{aligned} \tag{4.34}$$

As we can see from these expressions, all the four sub-matrices are diagonal matrices, we also have $\mathbf{B} = -\mathbf{C} < 0$, $\mathbf{A} < 0$ and $\mathbf{D} < 0$. Now we can calculate the symmetric matrix $(\mathbf{G}(\mathbf{p},\mathbf{j}) + \mathbf{G}^T(\mathbf{p},\mathbf{j}))$ and determine if it is a negative definite matrix.

$$\mathbf{G}(\mathbf{p},\mathbf{j}) + \mathbf{G}^T(\mathbf{p},\mathbf{j}) = \begin{bmatrix} [2\mathbf{A}] & [\mathbf{B}+\mathbf{C}] \\ [\mathbf{B}+\mathbf{C}] & [2\mathbf{D}] \end{bmatrix} = \begin{bmatrix} [2\mathbf{A}] & [\mathbf{0}] \\ [\mathbf{0}] & [2\mathbf{D}] \end{bmatrix} \tag{4.35}$$

Since the diagonal sub-matrices $\mathbf{A}$ and $\mathbf{D}$ are negative definite, we can conclude that $(\mathbf{G}(\mathbf{p},\mathbf{j}) + \mathbf{G}^T(\mathbf{p},\mathbf{j}))$ is a negative definite matrix, which is sufficient to prove the condition of diagonally strictly concavity. So, this game has a unique NE.

In two-person zero-sum game, this unique equilibrium equals the Saddle Point of the game [102], which is the subject of the following section.

4.10 Closed form Expression of the Saddle Point

The saddle point is so called because if we represent the payoff values as a matrix, the equilibrium value is the minimum in its row and the maximum in its column, this value is the value of the game, and the players' actions are the row and column that intersect at that point. This description of the saddle-point refers to a saddle sit on a horse's back at the lowest point on its head-to-tail axis and highest point on its flank-to flank axis [103]. As example, we determine in figure 4.6 the saddle point of this game over two flat fading channels.

The saddle point of this game corresponds to the optimal power allocations for both the jammer and the CR. The explicit solution to this game allows the CR to study the jamming strategy and to proactively use the corresponding optimal anti-jamming power allocation. Computing the closed form of the saddle point through exhaustive search over all the possible power allocations of the two players turns out to be hard to do in terms of resource and time consumption. It turns out that it is possible to develop its analytical expression under certain condition, we will start by the general case before presenting the conditions.

4.10.1 General case

Based on the equations (4.6) and (4.12) given to solve each player's decision problem, a vector of powers $(\mathbf{p}, \mathbf{j})$ constitutes the saddle point if and only if there are KKT multipliers λ and μ such that [104]:

$$\frac{\partial F}{\partial p_k} = \frac{|h_k|^2}{|h_k|^2 p_k + |g_k|^2 j_k + n_k} = \lambda \tag{4.36}$$

and

$$\frac{\partial(-F)}{\partial j_k} = \frac{|g_k|^2 |h_k|^2 p_k}{(|h_k|^2 p_k + |g_k|^2 j_k + n_k)(|g_k|^2 j_k + n_k)} = \mu \tag{4.37}$$

Equation (4.36) gives the expression of p_k as

$$p_k = \frac{1}{\lambda} - \frac{n_k + |g_k|^2 j_k}{|h_k|^2} \tag{4.38}$$

Now we replace p_k in (4.37) by the expression (4.38) to find j_k

$$j_k = \frac{|h_k|^2}{\lambda |g_k|^2 + \mu |h_k|^2} - \frac{n_k}{|g_k|^2} \tag{4.39}$$

If $j_k \geq 0$, we can replace j_k in (4.38) to get the expression of p_k

$$p_k = \frac{\mu}{\lambda} \frac{|h_k|^2}{\lambda |g_k|^2 + \mu |h_k|^2} \tag{4.40}$$

So, we can give the equilibrium strategies closed forms for $k \in [1, M]$

$$p_k = \begin{cases} \frac{\mu}{\lambda} \frac{|h_k|^2}{\lambda |g_k|^2 + \mu |h_k|^2} & \text{if} \quad \frac{n_k}{|h_k|^2} < \frac{|g_k|^2}{\lambda |g_k|^2 + \mu |h_k|^2} \\ \frac{1}{\lambda} - \frac{n_k}{|h_k|^2} & \text{if} \quad \frac{|g_k|^2}{\lambda |g_k|^2 + \mu |h_k|^2} \leq \frac{n_k}{|h_k|^2} < \frac{1}{\lambda} \\ 0 & \text{if} \quad \frac{n_k}{|h_k|^2} > \frac{1}{\lambda} \end{cases} \tag{4.41}$$

and

$$j_k = \begin{cases} \frac{|h_k|^2}{\lambda |g_k|^2 + \mu |h_k|^2} - \frac{n_k}{|g_k|^2} & \text{if} \quad \frac{n_k}{|h|^2} < \frac{|g_k|^2}{\lambda |g_k|^2 + \mu |h_k|^2} \\ 0 & \text{if} \quad \frac{n_k}{|h_k|^2} \geq \frac{|g_k|^2}{\lambda |g_k|^2 + \mu |h_k|^2} \end{cases} \tag{4.42}$$

To simplify and explain these power allocation expressions, we define a new parameter $\tau_k = \lambda + \mu \frac{|h_k|^2}{|g_k|^2}$. We get $\forall k \in [1, M]$

$$p_k = \begin{cases} \frac{\mu}{\lambda} \frac{|h_k|^2}{\lambda |g_k|^2 + \mu |h_k|^2} & \text{if} \quad \frac{n_k}{|h_k|^2} < \frac{1}{\tau_k} \\ \frac{1}{\lambda} - \frac{n_k}{|h_k|^2} & \text{if} \quad \frac{1}{\tau_k} \leq \frac{n_k}{|h_k|^2} < \frac{1}{\lambda} \\ 0 & \text{if} \quad \frac{n_k}{|h_k|^2} > \frac{1}{\lambda} \end{cases} \tag{4.43}$$

and

$$j_k = \begin{cases} \frac{|h_k|^2}{|g_k|^2}\left(\frac{1}{\tau_k} - \frac{n_k}{|h_k|^2}\right) & \text{if} \quad \frac{n_k}{|h|^2} < \frac{1}{\tau_k} \\ 0 & \text{if} \quad \frac{n_k}{|h_k|^2} \geq \frac{1}{\tau_k} \end{cases} \tag{4.44}$$

We can draw the following three cases controlled by the three power levels: $\frac{1}{\lambda}$ related to the CR, $\frac{1}{\tau_k}$ related to the jammer and $\frac{n_k}{|h_k|^2}$ related to the noise:

- (a) Since $\frac{1}{\tau_k} < \frac{1}{\lambda}, \forall k \in [1, M]$, a bad channel for the CR ($\frac{n_k}{|h_k|^2} > \frac{1}{\lambda}$) is also a bad channel for the jammer ($\frac{n_k}{|h_k|^2} > \frac{1}{\tau_k}$). The jammer does not attack a channel which is not occupied by the CR, i.e. if $p_k = 0$ then $j_k = 0$
- (b) In channels verifying $\frac{1}{\tau_k} \leq \frac{n_k}{|h_k|^2} < \frac{1}{\lambda}$, the CR succeeds to transmit without being jammed; i.e. $p_k > 0$ and $j_k = 0$, these channels are considered unfavorable for the jammer. It avoids these channels may be because of low g_k values which may force it to send with very high power to achieve the CR attack. A solution for the jammer to minimize the number of channels verifying this condition (since it can be considered as favorable opportunity for the CR), is to be close to the receiver node in order to get high g_k values and so $\frac{1}{\tau_k} \approx \frac{1}{\lambda}$.
- (c) If $\frac{n_k}{|h_k|^2} < \frac{1}{\tau_k}$, the channel is considered good for the two players and so occupied by both the CR and the jammer.

We provide in figure 4.7 an example covering these three situations.

4.10.2 Case all channels are used by both the CR and the jammer

Under the assumption that the jammer and the CR use all the channels ($p_k, j_k > 0, \forall\ k \in [1, M]$), which means $\frac{|g_k|^2}{\lambda|g_k|^2+\mu|h_k|^2} \geq \frac{n_k}{|h_k|^2}$, then we can give the power allocation closed forms at the NE for $k \in [1, M]$

$$\begin{cases} p_k &= \frac{\mu}{\lambda}\frac{|h_k|^2}{\lambda|g_k|^2+\mu|h_k|^2} \\ j_k &= \frac{|h_k|^2}{\lambda|g_k|^2+\mu|h_k|^2} - \frac{n_k}{|g_k|^2} \end{cases} \tag{4.45}$$

The power allocations should respect the conditions (4.1) and (4.2) which give

$$\begin{cases} \frac{\mu}{\lambda}\sum_{k=1}^{M}\frac{|h_k|^2}{\lambda|g_k|^2+\mu|h_k|^2} &= P \\ \sum_{k=1}^{M}\frac{|h_k|^2}{\lambda|g_k|^2+\mu|h_k|^2} - \sum_{k=1}^{M}\frac{n_k}{|g_k|^2} &= J \end{cases} \tag{4.46}$$

it gives the following relation between λ and μ

$$\frac{\lambda}{\mu} = \frac{J + \sum_{k=1}^{M}\frac{n_k}{|g_k|^2}}{P} \tag{4.47}$$

so, we can replace μ in p_k, and λ in j_k to get

$$\begin{cases} p_k & = 1/\lambda \left(1 + \frac{|g_k|^2}{|h_k|^2} \frac{J + \sum \frac{n_k}{|g_k|^2}}{P}\right) \\ j_k & = 1/\mu \left(1 + \frac{|g_k|^2}{|h_k|^2} \frac{\left(J + \sum \frac{n_k}{|g_k|^2}\right)}{P}\right) - \frac{n_k}{|g_k|^2} \end{cases} \tag{4.48}$$

Using the conditions (4.1) and (4.2), we get the closed form expressions of λ and μ

$$\begin{cases} \lambda & = \sum_{k=1}^{M} 1/\left(P + \frac{|g_k|^2}{|h_k|^2}\left(J + \sum \frac{n_k}{|g_k|^2}\right)\right) \\ \mu & = \frac{1}{J + \sum \frac{n_k}{|g_k|^2}} \sum_{k=1}^{M} 1/\left(1 + \frac{|g_k|^2}{|h_k|^2} \frac{\left(J + \sum \frac{n_k}{|g_k|^2}\right)}{P}\right) \end{cases} \tag{4.49}$$

Finally, replacing λ and μ in (4.48) gives the closed form expressions of the power allocations at the NE, and the following relation

$$j_k = \frac{J + \sum \frac{n_k}{|g_k|^2}}{P} p_k - \frac{n_k}{|g_k|^2} \tag{4.50}$$

This analytical result will be compared in section 4.11 with the NE found by simulation through playing iteratively the unilateral games.

4.10.3 Case of proportional fading channels

Now, let's consider the particular case studied in [87] of proportional fading coefficients,

$$g_k = \beta h_k, \forall k \in [1, M] \tag{4.51}$$

we define

$$\tau = \lambda\beta + \mu \tag{4.52}$$

So, the expression of λ in (4.49) becomes

$$\lambda = \frac{M}{P + \beta(J + \sum_{k=1}^{M} \frac{n_k}{|g_k|^2})} \tag{4.53}$$

Replacing λ in (4.48) results in

$$\begin{cases} p_k & = \frac{P}{M} \\ j_k & = \frac{J + \sum_{k=1}^{M} \frac{n_k}{|g_k|^2}}{M} - \frac{n_k}{|g_k|^2} \end{cases} \tag{4.54}$$

which brings us to the same conclusion as [87] about uniform power allocation; i.e. if the jammer tries to jam all the channels, then the optimal anti-jamming strategy for the CR is to allocate its power equally over the channels, under the assumption of proportional fading coefficients.

4.11 Simulation results and discussion

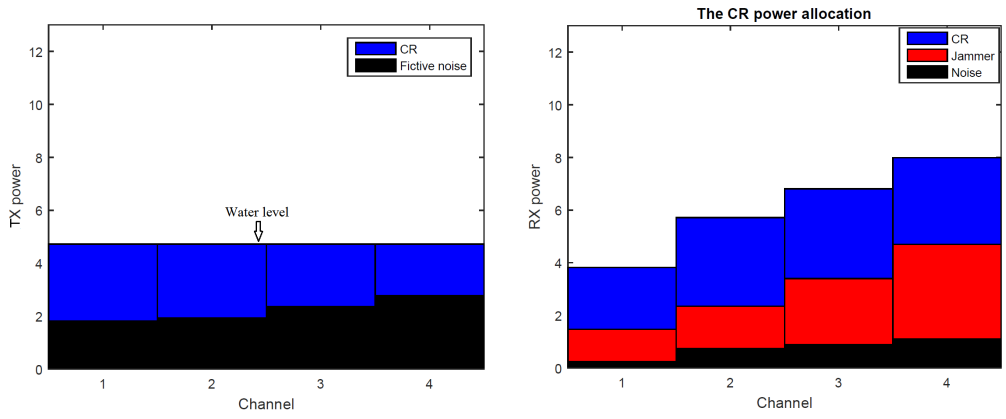
In the following simulations, we consider the system model of section 4.4 and the game scenarios respectively described in the previous sections (unilateral, Nash, Stackelberg, minmax/maxmin). Furthermore, we compare the simulation results to the analytical expression and we end the section by the saddle point illustration. We suppose that there are $M = 4$ available channels, the noise level vector equals $\mathbf{n} = (0.25, 0.75, 0.9, 1.1)$, $P = 10$ and $J = 10$ are the total power respectively for the CR user and the jammer, the channel coefficients are given by $\mathbf{h} = (0.9, 1.1, 1.2, 1.3)$ and $\mathbf{g} = (0.7, 0.8, 1, 1.2)$.

4.11.1 CR user unilateral game

To implement the solution of the CR user unilateral game described in subsection 4.5.1, we consider the fictive noise level in every channel as given by the expression (4.9). We proceed by bisection until reaching the maximum water level corresponding to the allocation of the total power of the CR, as illustrated by figure 4.3-(a).

As a fixed jamming action, we consider $\mathbf{j} = (2.5, 2.5, 2.5, 2.5)$. The waterfilling strategy of the CR user results in $\mathbf{p}^* = (2.9053, 2.7842, 2.3652, 1.9453)$ and a capacity $C = 4.4254$. Figure 4.3-(b) gives the total received power per channel ($k \in [1, 4]$) in terms of noise (n_k), jamming signal ($|g_k|^2 j_k$) and CR user's signal ($|h_k|^2 p_k$).

Let us compare the total transmission capacity resulting from the application of the waterfilling strategy to the result of using flat power allocation. If the CR user assigns a power level $p_k = \frac{P}{M}$ to each channel $k \in [1, M]$, the total capacity will be equal to $C = 4.4073$ which results in a payoff loss compared to the optimal waterfilling strategy.



(a) Waterfilling (b) The received power

Figure 4.3: CR user unilateral game

4.11.2 Jammer unilateral game

To implement the jammer unilateral game described in subsection 4.5.2, we proceed by bisection and we calculate the sum of the allocated powers to all the channels (using equation (4.13)) until reaching the value of μ corresponding to the allocation of the total jamming power J.

Under the same conditions as the above simulation, we consider that the CR user's power allocation is fixed to $\mathbf{p} = (1, 2, 3, 4)$. The reaction of the jammer is given by $\mathbf{j}^* = (2.0723, 2.2849, 2.7064, 2.9364)$ and the resulting capacity is equal to $C = 4.0979$. From this result, note that the jammer pursues the CR user in terms of power allocation. It assigns a higher power to the channels having higher CR user's power. The received power per channel is given in figure 4.4.

Under the scenario of imperfect knowledge of the opponent's strategy and the channels gain coefficients, the trivial solution for the jammer would be a flat power allocation. The resulting capacity for the CR user will be $C = 4.1217$ which is higher than the result of applying the described technique based on bisection. Hence, the jammer using flat power allocation loses in terms of payoff since his goal is to minimize the CR user's total transmission capacity.

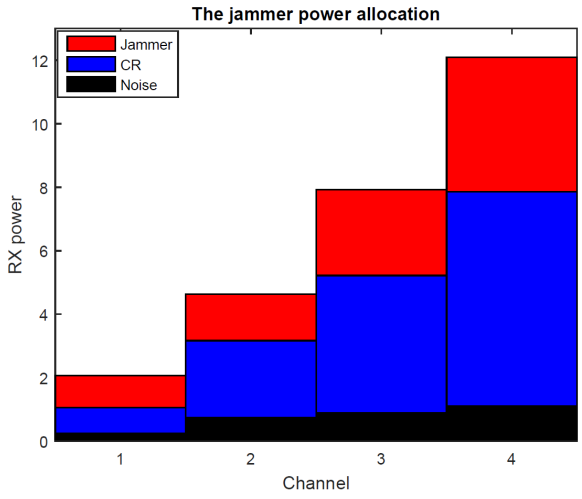


Figure 4.4: Jammer unilateral game

4.11.3 Nash game

The Nash game scenario between the CR user and the jammer, described in section 4.6, consists in playing iteratively the two unilateral games presented in section 4.5 until convergence to almost fixed power allocation per channel within a specific tolerance $\epsilon = 1e - 15$.

Considering the same conditions as the previous games, we find at the convergence to the NE $\mathbf{j}^* = (2.9625, 2.5073, 2.3574, 2.1729)$, $\mathbf{p}^* =$

$(2.602, 2.7568, 2.4407, 2.2005)$ and $C = 4.4017$. Figure 4.5 gives the received power per channel, at the NE. Contrary to the CR user who allocates higher power to the less occupied channels, the jammer allocates higher power to the more occupied channels since he tries to minimize the CR user's payoff.

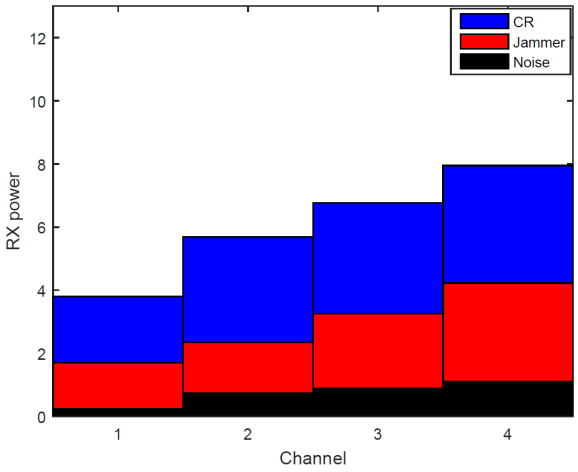


Figure 4.5: The strategies at the NE

4.11.4 Stackelberg game: Jammer as the leader

As described in subsection 4.7.1, we consider that the jammer is the leader and knows the explicit expression of the CR user's reaction function. To implement this game, the jammer does one dimensional exhaustive search over its possible power allocations to find the optimal power allocation $\mathbf{j}^*$ which minimizes the CR user's transmission capacity. The CR user, playing as follower, determines its optimal power allocation $\mathbf{p}^*$ by using equation (4.7) found in the CR user unilateral game since he can observe the jammer's startegy. We found the same power allocations and the same capacity value as for the NE. Hence, the jammer playing as a leader with knowledge about the reaction function of the opponent finds the same optimal jamming strategy compared to the scenario of playing in iterative way by only observing the instantaneous action of the opponent.

4.11.5 Stackelberg game: CR user as the leader

We consider the Stackelberg game described in subsection 4.7.2. The CR user is the leader and knows the reaction function of the jammer. Hence, the CR user does exhaustive search over all its possible power allocations $\mathbf{p}$ to find the optimal power allocation $\mathbf{p}^*$ maximizing its transmission capacity. The jammer, as follower, uses the expression (4.13) found in the jammer unilateral game since it can observe the CR user's power allocation. Also for this scenario, we find the same power allocations and the same capacity value as the result found at the NE.

According to the simulation results of both Stackelberg games (with the jammer or the CR user as the leader), neither the leader wins due to the knowledge of the opponent's reaction nor the follower loses compared to the Nash game.

4.11.6 Minmax/maxmin optimal solutions

We consider the same parameters considered in the previous simulations to find the NE and the SE. Here, we determine the optimal solutions with the help of the characteristics of the solution found at the NE, otherwise the exhaustive search will be difficult to launch in continuous action sets. We limit the research to the interval $[2, 3]$ where we have found the NE and we consider a step of 0.01. We found the maxmin CR user's power allocation: $\mathbf{p}_{maxmin} = (2.6, 2.76, 2.44, 2.2)$ with $C_{maxmin} = 4.4017$, and the minmax jammer's power allocation $\mathbf{j}_{minmax} = (2.96, 2.51, 2.36, 2.17)$ with the capacity $C_{minmax} = 4.4017$.

Comparing the simulation results, we can note that the optimal values found by exhaustive search under the assumption of finite action subsets and a step of 0.01, give a very near approximation to the power allocations at the NE found in the continuous action sets. Accordingly, the power allocations at the Nash equilibrium and at the Stackelberg equilibrium are equal to the optimal minmax/maxmin power allocations.

4.11.7 Comparing analytical saddle point to the NE

Let's start by replacing the parameters $(|\mathbf{h}|^2, |\mathbf{g}|^2, \mathbf{n}, P, J, M)$ in the analytical expressions of subsection 4.10.2. According to the optimal power allocations given by the expressions (4.48) and (4.49), we get $\mathbf{j} = (2.9625, 2.5073, 2.3574, 2.1729)$ and $\mathbf{p} = (2.602, 2.7568, 2.4407, 2.2005)$. Which results in a payoff value of $C = 4.4017$. This analytical result equals the simulation result of the Nash game.

4.11.8 Saddle point example

Just to illustrate the concept of saddle-point, we consider $M = 2$ flat fading channels with gain coefficients $h_k = g_k = 1$, $\forall k \in [1, M]$. We choose $P = 30$ and $J = 20$ as the total power for the CR and the jammer respectively. We consider only finite sets of power allocations with steps of 1, so $p_k \in \{0, 1, 2, \cdots, P\}$ and $j_k \in \{0, 1, 2, \cdots, J\}$. We have implemented this scenario in Matlab using the exhaustive search over the finite set of possible power allocations. We calculated a matrix of capacity values, its rows are the possible jammer's power allocations and its columns are the CR's power allocations. We found the optimal maxmin CR's power allocation: $\mathbf{p}^* = (15, 15)$ corresponding to the column number 16, and optimal minimax jammer's power allocation $\mathbf{j}^* = (10, 10)$ corresponding to the row 11. Figure 4.6 illustrates this saddle point given by the indexes of $\mathbf{p}^*$ and $\mathbf{j}^*$.

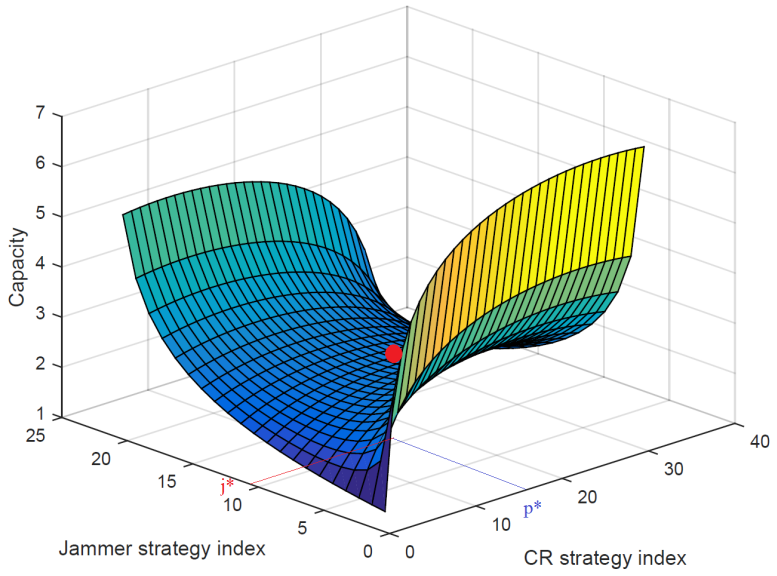


Figure 4.6: The saddle point for two channels

4.11.9 Nash equilibrium in the general case

To cover the general case detailed in subsection 4.10.1, we consider the system model described in figure. 4.2 with $M = 4$ parallel Gaussian channels, $P = 10$ and $J = 10$ as the total CR's and jammer's powers in watts, the background noise over the four channels $\mathbf{n} = (2, 0.75, 0.9, 1.1)$ and the channel gain coefficients $\mathbf{h} = (0.1, 1.1, 1.2, 1.3)$, $\mathbf{g} = (0.7, 0.8, 0.1, 1.2)$.

After convergence of the iterative game to almost fixed power allocations with tolerance $\epsilon = 1e - 10$, we get $\mathbf{j} = (0, 5.704, 0, 4.296)$ and $\mathbf{p} = (0, 2.5543, 5.5661, 1.8797)$. Which results in a payoff value of $C = 4.5978$ with $\frac{1}{\lambda} = 6.1911$ and $\mu = 0.06$. Figure 4.7 gives the received power due to the noise, jammer and CR's powers in each channel at the NE.

We can see that in channel 1, $p_k = j_k = 0$ since $\frac{n_1}{|h_1|^2} > \frac{1}{\lambda}$ which corresponds to the case (a) in paragraph 4.10.1. Channel 3 receives $p_k > 0$ but $j_k = 0$, since $\frac{1}{\tau_3} < \frac{n_3}{|h_3|^2} < \frac{1}{\lambda}$ which corresponds to case (b). Channels 2 and 4 corresponds to case (c) since $\frac{n_k}{|h_k|^2} < \frac{1}{\tau_k}$ which results in $p_k > 0$ and $j_k > 0$.

4.12 Conclusion

In this chapter, we have exploited the CR capacities of simultaneous multi-frequency access and dynamic power allocation as the anti-jamming strategy. We have modeled the interaction between the two players, using different strategies to dynamically update their power allocations, as a zero-sum game with continuous action sets. Then, we have considered different game scenarios, for which we have

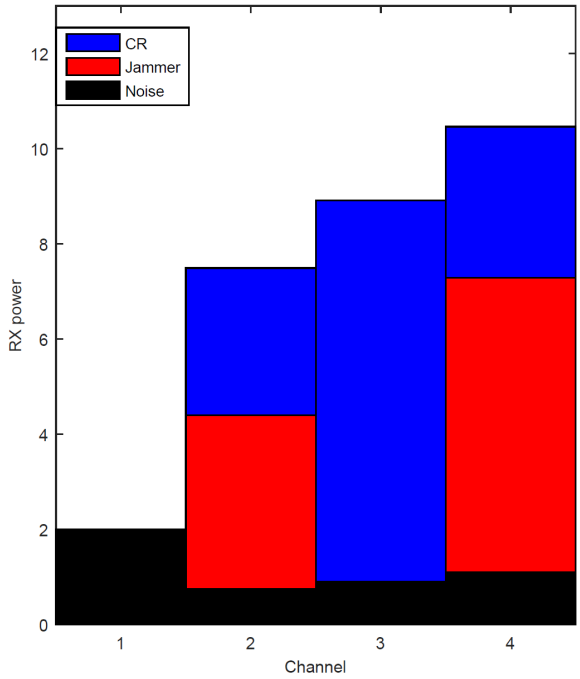


Figure 4.7: The strategies at the NE in general case

determined the NE, SE and the optimal minmax/maxmin power allocations. The simulation results have given equality between the solutions of all the considered game scenarios. We have proved theoretically that this game has a unique equilibrium which is equal to the saddle point given in closed form, under the assumption that both the CR and the jammer are using all the channels (i.e. p_k, $j_k > 0$, $\forall$ $k \in [1, M]$).

To solve the presented game and find the optimal power allocation strategies, we have considered complete information for both the CR and the jammer. Under this assumption, each player has all relevant information with which to make a decision in each step of the game. But in real scenario the players has no information about the required parameters to calculate their optimal strategies. To overcome this problem, in the next chapter we exploit the CR capacity of learning and reasoning to develop an anti-jamming technique under incomplete information. The proposed learning algorithm will be applied as a jamming solution in terms of channel selection before adapting it to the multi-channel power allocation game.

Chapter 5

Learning based anti-jamming technique

5.1 Introduction

The cognitive radio capacities of sensing and learning can be exploited to deal with the lack of information required to make suitable decisions. The Q-learning is a common model-free reinforcement learning (RL) algorithm applied in CRN jamming study to deal with incomplete knowledge about the environment [105–107]. Differently from work available in literature, we aim to provide a modified version of the Q-learning algorithm to speed up the training period and to make it appropriate for on-line learning. The proposed anti-jamming algorithm will be applied for one channel selection, then generalized for the the multi-channel game presented in the previous chapter.

We start by defining the key terms: Markov Decision Process (MDP) and Q-learning algorithm, before providing the proposed learning algorithm in section 5.3. The channel selection application is presented in section 5.4. We present each component of the MDP modeling the problem in the presence of fixed jammer and we adapt the proposed algorithm to solve it. During learning, the CR tries to maximize its long term return which combines, into the Q-values, the sequence of rewards related to the visited states and taken actions. In section 5.5, we propose an ameliorated reward function to stay as long as possible in the same channel without being jammed.We evaluate the effectiveness of the modified Q-learning algorithm in the presence of different jamming strategies. We present and compare also the learned anti-jamming strategies related to the two proposed reward functions. In section 5.6, the proposed learning algorithm is adapted to the multi-channel power allocation game. We start by considering fixed jamming strategy to compare the learned anti-jamming strategy to the solution of chapter 4, based on waterfilling. Then, we will consider the scenario of a jammer using Q-learning algorithm to minimize the total transmission capacity of the CR. We compare the result with the Nash game of the previous chapter obtained under the assumption of complete

channel information for both of the CR transmitter and the jammer.

5.2 Key terms

5.2.1 The Markov decision process

The MDP is a discrete time stochastic control process. It provides a mathematical framework to model the decision problem faced by an agent to optimize his outcome. The goal of solving the MDP is to find the optimal strategy for the considered agent. In CRN jamming scenario, it means finding the best actions (to hop to a different channel or to stay) for the CR to avoid the jammed frequency.
A MDP is defined by four essential components:

- A finite set of states $\{S_0, \cdots, S_N\}$.
- A finite set of actions $\{a_1, \cdots, a_M\}$.
- $P_a(S, S') = Pr(S_{t+1} = S' | S_t = S, a_t = a)$ the transition probability from an old state S to a new state S' when taking action a.
- $\mathrm{R}_a(S, S')$ the immediate reward after transition to state S' from state S when taking action a.

The process is played in a sequence of time slots $t = 0, 1, 2, \cdots$. At every slot, the agent being in one state selects an action to move to a new state with the corresponding transition probability. The agent receives a payoff, also called reward, which depends on the current state and the taken action. He continues to play until finding the optimal policy, which is the mapping from states to actions that maximizes the state values. The standard family of algorithms used to calculate this optimal policy requires storage of two arrays indexed by state:

- State value $V(S)$, which contains a real value corresponding to the discounted sum of the rewards received when starting from each state.
- Policy $\pi(S)$ which gives the action taken in every state.

Every MDP has at least one optimal policy π^* that is stationary and deterministic. π^* is called stationary since it does not change as a function of time and it is called deterministic since the same action is always chosen whenever the agent is in one state S. At the end of the algorithm, π^* will contain the optimal solution and $V(S)$ will contain the discounted sum of the rewards to be earned by following that policy from state S.

Markov decision processes can be solved via dynamic programming (DP) when we have complete knowledge about transition probabilities and the reward of every action. However in real situations of dynamic environment and incomplete knowledge about transition probabilities and rewards, MDP is solved using RL algorithms [105].

DP techniques require an explicit, complete model of the process to be controlled. It is known as model based techniques, since we have to reconstruct an approximate model of the MDP and then solve it to find the optimal policy. The most popular DP techniques is the value iteration algorithm which consists in solving the following Bellman equation until convergence to the optimal values $V^*(S)$, from which we can derive the corresponding optimal policy:

$$Q(S,a) = R_a(S,S') + \gamma \sum_{S'} P_a(S,S')V^*(S') \tag{5.1}$$

$$V^*(S) = max_a Q(S,a) \tag{5.2}$$

where γ is the discount factor that controls how much effect future rewards have on the optimal decisions. Small values of γ emphasizing near-term gain and larger values giving significant weight to later rewards. Equation (5.1) is repeated for all possible actions in each state S. It calculates the sum of the immediate reward $R_a(S,S')$ of the taken action and the expected sum of rewards over all future steps. Then, equation (5.2) gives the optimal action which corresponds to the maximum $V(S)$ value. The value iteration algorithm reaches convergence when $|V_{t+1}(S) - V_t(S)| < \epsilon$ is met for all states S, where $V_t(S)$ corresponds to the calculated $V(S)$ value at time slot t.

However, in real scenarios the CR is acting in hostile and dynamic environment without complete information. It doesn't know either the resulting new state after taking an action or the reward/cost of its action. For example, hopping to another frequency may lead to jamming situation or successful transmission. This situation can be defined as a RL problem, in which an agent wanders in an unknown environment and tries to maximize its long term return by performing actions and receiving rewards [106]. Therefore, the CR should use learning algorithms to learn incumbent user's and jammer's activities. After learning the jammers' policy, it can predict the next action of the jammer and plan its own next course of action to avoid jammed channels.

5.2.2 The Q-learning algorithm

Learning algorithms can be used as a model-free simulation tool for determining the optimal policy π^* without prior information about the action rewards and the transition probabilities. Autonomous RL is completely based on interactive experience to update the information step by step and derive an estimate of the optimal policy. The most popular RL method is the Q-learning algorithm, which is an extension to the value iteration algorithm to be applied in non deterministic Markov decision processes.

As first introduced in [107], the Q-learning algorithm is a simple way for agents to learn how to act optimally by successively improving its evaluations of the quality of each action. It consists in approximating the unknown transition probabilities

by the empirical distribution of states that have been reached as the process unfolds. The goal is finding a mapping from state/action pairs to Q-values. This result can be represented by a Q-matrix of N rows, where N is the number of states S, and M columns corresponding to possible actions a. The Bellman equation (5.1) is replaced in this algorithm by an iterative process; at every time slot the algorithm measures the feedback rewards of taking an action a in a state S, and updates the corresponding $Q(S, a)$:

$$Q(S,a) \leftarrow Q(S,a) + \alpha \left[R_a(S,S') + \gamma \, max_x Q(S',x) - Q(S,a)\right] \quad (5.3)$$

which gives:

$$Q(S,a) \leftarrow (1-\alpha)Q(S,a) + \alpha \left[R_a(S,S') + \gamma \, max_x Q(S',x)\right] \quad (5.4)$$

where $0 < \alpha \leq 1$ is a learning rate that controls how quickly new estimates are blended into old estimates. The Q-value is a prediction of the sum of the discounted reinforcements (rewards) received when performing the taken action and then following the given policy thereafter. It can be considered as a measure of the goodness of that action choice.

The Q-learning algorithm updates the values of $Q(S, a)$ through many episodes (trials) until convergence to stationary Q^* values ($|Q_{t+1}(S, a) - Q_t(S, a)| < \epsilon$, for all state S and action a); this is known as the training/learning stage of the algorithm. Each episode starts from a random initial state S_0 and consists of a sequence of time slots during which the agent goes from state to another and updates the corresponding Q value. Each time the agent reaches the goal state, which have to be defined depending on the scenario, the episode ends and he starts a new trial. The convergence to the optimal Q^* matrix requires visiting every state-action pair as many times as needed. In simulation, this problem is known as the exploration issue. Random exploration takes too long time to focus on the best actions which leads to a long training period of many episodes. Furthermore, it does not guarantee that all states will be visited enough, as a result the learner would not expect the trained Q function to exactly match the ideal optimal Q^* matrix for the MDP [108]. The training phase of the Q-learning process is described in algorithm 1 [109].

Two main characteristics of the standard Q-learning algorithm are: (i) it is said to be an asynchronous process since at each time slot the agent updates a single $Q(S, a)$ value (one matrix cell), corresponding to his current state S (row S) and his action a (column a) taken at this time slot [110]. (ii) The Q-learning method does not specify what action a the agent should take at each time slot during the learning period, therefore it is called OFF-policy algorithm allowing arbitrary experimentation until convergence to stationary Q values [111]. The optimal Q^* matrix resulting from the learning period will be exploited by the agent as the best policy. During the exploitation phase, when he is in a state S, he has to take the action corresponding to the maximum value in the matrix line $Q^*(S, :)$.

An off-line application of this technique seems to be inefficient for the CR, because until the convergence of the Q-learning algorithm other jammers may emerge

Algorithm 1 Pseudocode of the Q-learning algorithm

```
Set the γ parameter and the matrix R of environment rewards.
Initialize the Q matrix to zero matrix.
while Not convergence do
    Select a random initial state S = S_0.
    while The goal state hasn't been reached do
        Select one action a among all possible actions for the current state.
        Using this possible action, consider going to the next state S'.
        Get maximum Q value for this next state based on all possible actions
        max_a(Q(S', a)).
        Update the Q(S, a) value using equation (5.4)
        Set the next state as the current state S = S'.
    end while
end while
```

and legacy spectrum holders (primary users) activity may change. During the training phase of the Q-learning algorithm, the CR can already exploit the communication link, denoted as on-line learning, but it may lose many data packets because of the random learning trials.

In [112], a decentralized Q-learning algorithm is proposed to deal with the problem of aggregated interference generated by multiple CRs at passive primary receivers. Furthermore, it is implemented in the cognitive architecture Clarion to represent implicit learning and it is also incorporated into the cognitive architecture Soar [15, 16]. Moreover, [68] and [69] study the Q-learning algorithm to solve the CR jamming problem. In the first paper, the authors start by deriving a defense strategy for the CR using an MDP model under the assumption of complete knowledge, in terms of transition probabilities and rewards. Further, they propose two learning schemes for CRs to gain knowledge of adversaries to handle cases of incomplete knowledge: maximum likelihood estimation (MLE), and an adapted version of the Q-learning algorithm. However the modified Q-learning algorithm is given without discussion or simulation results. The second paper gives a MDP model of the CRN jamming scenario and proposes a modified Q-learning algorithm to solve it. Again, as in the previous reference, no details are given on how to implement the described theoretical anti-jamming scheme.

We have explained the MDP and the Q-learning algorithm tools commonly used to model and solve the CRN scenario under static jamming strategy. The CR can apply the Q-learning algorithm to learn the jammer's behavior, but it has to wait for a long training period before getting the optimal anti-jamming strategy. Moreover, as the CR has to try random actions before the convergence of the Q-learning algorithm, it is not suitable to do learning in an operational communication link because the CR may lose many transmitted packets. As a solution to these challenges, we propose in the next section a modified version of the Q-learning algorithm denoted ON-Policy Synchronous Q-learning (OPSQ-learning)

algorithm.

5.3 ON-Policy Synchronous Q-learning (OPSQ-learning)

We present in algorithm 2, a modified version of the Q-learning process denoted as the OPSQ-learning, because of the two following modifications: (i) We have replaced the OFF-policy characterizing the standard Q-learning algorithm by an ON-policy, i.e. at each time slot, the CR follows a greedy strategy by selecting the best action corresponding to $max_x Q(S, x)$ instead of trying random action. (ii) We have exploited the CR ability of doing wideband spectrum sensing [113–115], to do synchronous update of M Q-values instead of the asynchronous update of only one cell in the Q matrix, i.e. after moving to a next state, the CR can detect the frequency of the jammer at that moment, using its wideband sensing capability, and hence do an update of all state-action pairs, corresponding to the possible actions which can be taken from its previous state S (update all columns of the Q matrix line $Q(S, :)$).

We have to mention that we are assuming perfect spectrum sensing and full observations for simplicity. But we cite some interesting references dealing with the influence of the radio channel in the estimation of the detected signal. For example, [116] develops and analyzes an adaptive spectrum sensing scheme according to the variation of time-varying channels, [18] studies the cooperative spectrum sensing for a cognitive radio system operating in AWGN, correlated/uncorrelated shadowing, and in channels featuring composite large-scale and small scale fading. Also, [117] provides a comprehensive overview of the propagation channel models that will be used for the design of cognitive radio systems and deals with the time variations of the channel response which determine how often potential interference levels have to be estimated and, thus, how often transmission strategies may have to be adapted.

In the rest of this chapter, the proposed algorithm will be applied to solve the jamming attack in terms of channel selection before being generalized to the multi-channel power allocation game.

5.4 Learning based channel selection

We consider a fixed jamming strategy to solve the decision making problem from the side of the CR trying to find an anti-jamming strategy. Assume there are M available channels for the CR and there is a jammer trying to prevent it from an efficient exploitation of these channels. As a defense strategy, the CR have to choose at every time slot either to keep transmitting over the same channel or to hop to another one. The challenge is to learn how to escape from jammed channels without sacrifying a long training period to learn the jammer's strategy. We will start by defining a Markov decision process to model the CR's available states and actions, with the consideration of unknown transition probabilities and unknown

Algorithm 2 Pseudocode of ON-policy synchronous Q-learning (OPSQ-learning)

Set the γ parameter and the matrix R of environment rewards.
Initialize the Q matrix to zero matrix.
while Not convergence **do**
  Select a random initial state $S = S_0$
  while The goal state hasn't been reached **do**
    Select the best action a verifying $max_a Q(S, a)$
    Using this action, consider going to the next state S'.
    Get maximum Q value for this next state based on all possible actions $max_a(Q(S', a))$.
    Update all the $Q(S, :)$ values using equation (5.4)
    Set the next state as the current state $S = S'$.
  end while
end while

immediate rewards of the taken actions. Then, we will adapt the OPSQ-learning algorithm to solve the defined MDP model.

The state of the CR is defined by a pair of parameters: $S = (f_{TX}, n)$, where f_{TX} is its operating frequency associated to that state and n is the number of successive time slots using this frequency. We have opt for mixing spatial and temporal properties in the state space definition to consider the CR staying in the same channel more than one time. At every state, the CR should choose an action to move to another state, which means that it has to choose its future frequency. Therefore, we define its possible actions as a set of M actions, which are the M available channels: $\{a_1, a_2, ..., a_M\}$=$\{f_1, f_2, ..., f_M\}$. An example of the Q matrix composed by these states and actions is given in Table 5.1. Under the assumption of synchronized jammer and CR (meaning that they have the same notion of time slot), we give a reward equal to zero $R_a(S, S') = 0$ whenever the new frequency after choosing the action a is not jammed, and $R_a(S, S') = -1$ when the CR takes an action a resulting in a jammed frequency. We consider the jammed state as a failure and a situation that should be avoided.

We present in algorithm 3, an adaptation of the OPSQ-algorithm to solve the described MDP model under incomplete information about the jammer. In this scenario, our episode starts from a random frequency, going from one state to another by taking the best action available at every time slot, and ends whenever the CR goes to a jammed frequency.

The next subsection presents the simulation results in the presence of various jamming strategies.

5.4.1 Simulation results

We have considered in the simulations four available frequencies ($M = 4$) for the CR. We have implemented both the standard (algorithm 1) and the modified

Algorithm 3 Channel selection using OPSQ-learning

Set the γ parameter and initialize the rewards to zero values.
Initialize the Q matrix to zero matrix.
while Not convergence **do**
  Select a random initial state $S = S_0$ and set $timeslot = 1$
  while The goal state hasn't been reached **do**
    Calculate the learning coefficient $\alpha = 1/timeslot$
    Select the best action a verifying $max_a Q(S, a)$ and corresponding to frequency f'.
    Using this action, consider going to the next state S'.
    Detect the jammed frequency f_{JX} %(due to wideband spectrum sensing)
    Update all the $Q(S, :)$ values, associated to the state S, by doing:
    for $i = 1 : M$ **do**
      Observe the fictive state S_{tmp} of taking fictive action f_i
      if $f_i = f_{JX}$ **then**
        $R_{f_i}(S, S_{tmp}) = -1$
      else
        $R_{f_i}(S, S_{tmp}) = 0$
      end if
      Update $Q(S, f_i) = (1 - \alpha)Q(S, f_i) + \alpha[R_{f_i}(S, S_{tmp}) + \gamma\, max_a Q(S_{tmp}, a)]$
    end for
    Set the next state as the current state $S = S'$ and increment $timeslot = timeslot + 1$.
  end while
end while

(algorithm 3) versions of the Q-learning algorithm, under sweeping, reactive and pseudo random jamming strategies.

We started by the implementation of the standard version of Q-learning algorithm. We found, by averaging over many simulations, that it takes about one hundred episodes to converge to the matrix Q^*. Then, we have implemented the modified Q-learning version (OPSQ-learning) and we give the results in the following paragraphs. The following figures display the anti-jamming strategy in the exploitation phase, after running the learning algorithm. We are using the red color to indicate the jammed frequencies and the blue color to indicate the CR frequencies.

5.4.1.1 Scenario with a sweeping jammer

As a first scenario, we consider a jammer sweeping over the available spectrum frequencies by attacking at each time slot one frequency. The OPSQ-learning algo-

rithm converges after only one or two episodes. The Q^* matrix is given in table 5.1. The strategy given by this resulting Q^* matrix is shown in figure 5.1 starting from the frequencies f_2 and f_3 respectively. Being in f_1 for the first time (the first row of the Q matrix), the CR takes the first action (selects the column) having the maximum Q value which results in staying in frequency f_1. Due to this decision, the CR moves to the state $(f_1, 2)$ (the second row of the Q matrix). Being in this state, the best action corresponds to staying in the same frequency and results in the state $(f_1, 3)$. The next best action corresponds to moving to frequency f_2, so going to the state $(f_2, 1)$. From this new state, the first maximum Q value corresponds to frequency f_1. So, the CR stays three time slots in frequency f_1 before moving to f_2 from which he returns to f_1 and so on, as illustrated by figure 5.1.

State \ Action	f_1	f_2	f_3	f_4
(f_1,1)	0	0	-0.8356	0
(f_1,2)	0	0	0	-0.6768
(f_1,3)	-0.5770	0	0	0
⋮	⋮	⋮	⋮	⋮
(f_2,1)	0	-0.3822	0	0
⋮	⋮	⋮	⋮	⋮
(f_3,1)	0	-1	0	0
⋮	⋮	⋮	⋮	⋮
(f_4,1)	0	0	0	0
⋮	⋮	⋮	⋮	⋮

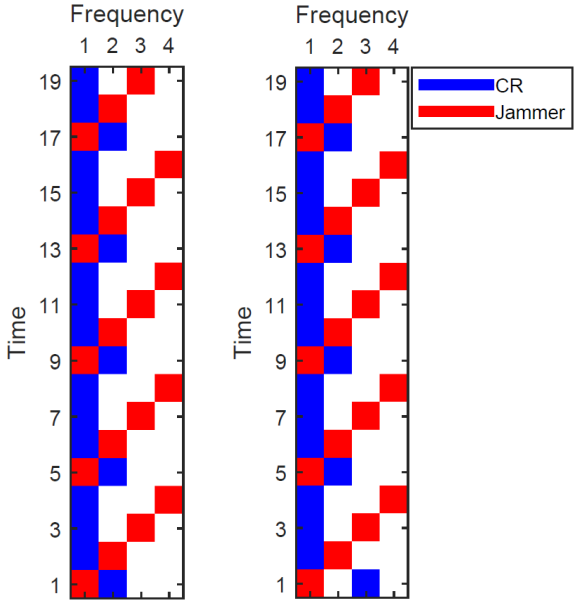
Table 5.1: The Q^* matrix in a sweeping jammer scenario

5.4.1.2 Scenario with a reactive jammer

In this scenario, we consider a reactive jammer. We assume that this jammer needs a duration of two time slots before jamming the detected frequency, because it has to do the spectrum sensing, then make the decision and finally hop to the detected frequency. The OPSQ-learning algorithm converges in this scenario after four episodes. The Q^* matrix is given in table 5.2. According to the resulting Q^* matrix, the CR succeeds to learn that it has to change its operating frequency every two time slots to escape from the reactive jammer. The learned strategy is given in figure 5.2 when the CR starts respectively from the frequencies f_2 and f_3 as initial state S_0.

5.4.1.3 Scenario with a pseudo random jammer

In this scenario, we consider a jammer with a pseudo random strategy. We assume that at every time slot, this jammer attacks randomly one of the four frequencies, and after a period T it repeats the same sequence of the attacked frequencies. We started with a period $T = 5$ during which the random sequence



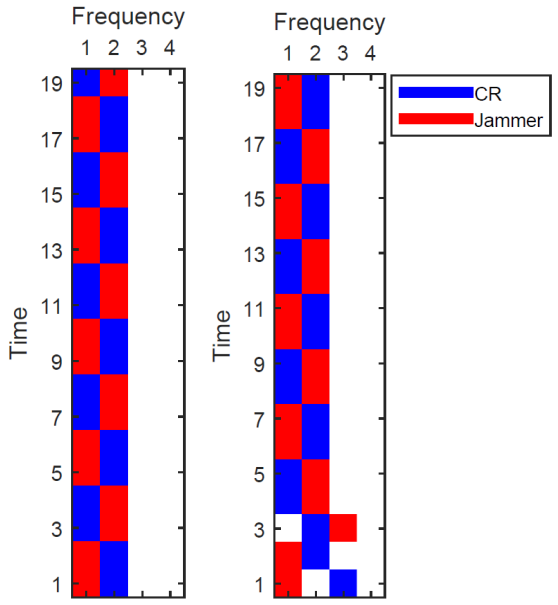
(a) $S_0 = (f_2, 1)$ (b) $S_0 = (f_3, 1)$

Figure 5.1: Exploitation of the learned policy against a sweeping jammer

State \ Action	f_1	f_2	f_3	f_4
(f_1,1)	0	-0.8047	0	0
(f_1,2)	-0.6986	0	0	0
⋮	⋮	⋮	⋮	⋮
(f_2,1)	-1	0	0	0
(f_2,2)	0	-0.6861	0	0
⋮	⋮	⋮	⋮	⋮
(f_3,1)	-1	0	0	0
⋮	⋮	⋮	⋮	⋮
(f_4,1)	-1	0	0	0
⋮	⋮	⋮	⋮	⋮

Table 5.2: The $Q*$ matrix in a reactive jammer scenario

is $(f_1, f_3, f_2, f_4, f_2)$, we found that the OPSQ-learning algorithm converges in this scenario after four episodes. Then, we considered a period $T = 10$ during which the random sequence is $(f_1, f_1, f_4, f_3, f_2, f_1, f_3, f_3, f_4, f_2)$, we found that the OPSQ-learning algorithm converges in this scenario after five episodes. The Q^* matrix is given in table 5.3. The CR succeeds to learn the pseudo random strat-



(a) $S_0 = (f_2, 1)$ (b) $S_0 = (f_3, 1)$

Figure 5.2: Exploitation of the learned policy against a reactive jammer

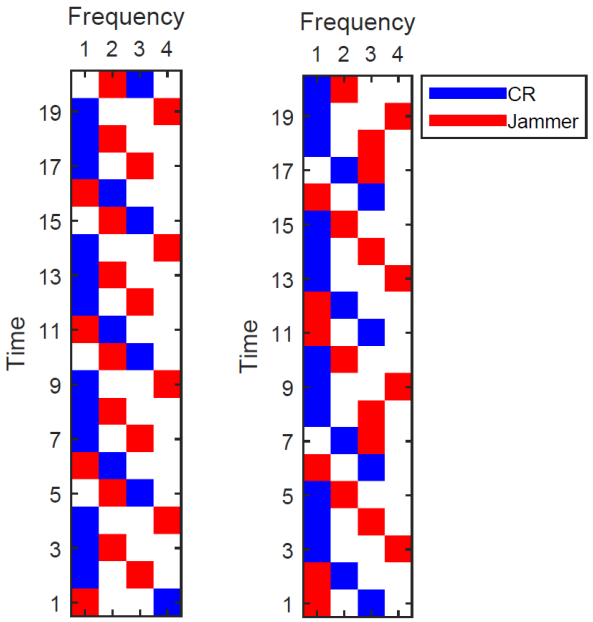
egy of the jammer, and the learned anti-jamming strategies are given in figure 5.3 when the periods of the pseudo random jamming sequences are respectively $T = 5$ and $T = 10$ time slots.

State \ Action	f_1	f_2	f_3	f_4
(f_1,1)	0	-0.8235	0	-0.0882
(f_1,2)	0	-0.1130	0	-0.6610
(f_1,3)	-0.1100	-0.5602	0	0
⋮	⋮	⋮	⋮	⋮
(f_2,1)	0	0	-0.3236	0
⋮	⋮	⋮	⋮	⋮
(f_3,1)	-1	0	0	0
⋮	⋮	⋮	⋮	⋮
(f_4,1)	0	0	-1	0
⋮	⋮	⋮	⋮	⋮

Table 5.3: The $Q*$ matrix in a pseudo random jammer with a period of 5 time slots

5.4.2 Discussion

The standard Q-learning algorithm converges after about one hundred episodes; each episode starts from a random frequency, going randomly from one frequency



(a) $T = 5$ (b) $T = 10$

Figure 5.3: Exploitation of the learned policy against a pseudo random jammer

to another taking random decisions until collision with the jammer. The CR applying this technique have to either wait for all this training period to get an anti-jamming strategy or to use it during real time communication and sacrifice about hundred lost packets. The ON-policy synchronous Q-learning algorithm converges faster than the standard Q-learning algorithm; in the simulated scenarios, it gives a suitable defense strategy after about four training episodes against sweeping and reactive jammers. This is due to the synchronous update of all Q-values of possible actions from a current state, which helps the CR to faster improve its beliefs about all decisions without trying all of the actions. Furthermore, the choice of taking at every time slot the best action promotes the real time exploitation of the OPSQ-learning algorithm during the CR communication. But, We should mention that the proposed OPSQ-learning algorithm doesn't optimize the entire matrix Q, it just optimizes the Q-values of state/action pairs that the CR goes through until finding an anti-jamming strategy. The CR using the proposed algorithm succeeds

to learn how to avoid the jammed channels, but as we can see in figures 5.2(b) and 5.3, the CR does unneeded frequency switching. It means that he learned to jump from frequency to another even if the first one will not be jammed in the next time slot, which costs in terms of time, frequency and power consumption. This disadvantage is due to the elected reward strategy, in which we accord a penalty of -1 just to the choice of a jammed frequency, otherwise the CR receives zero as reward. In the next section, we propose an ameliorated reward strategy trying to find the optimal anti-jamming strategy.

5.5 Ameliorated reward function for channel selection

In this section, we use $R1$ to denote the reward function defined in the previous section and propose another reward function denoted $R2$ associating a penalty of -1 not only for the choice of a jammed frequency but also for the frequency switching without having the previous frequency attacked by the jammer. We integrate this new reward strategy in the OPSQ algorithm given by the pseudocode 3 to get the new learning algorithm is given by pseudocode 4.

5.5.1 Simulation results

We have implemented this new algorithm considering the same simulation parameters as given in subsection 5.4.1. Against a sweeping jammer we get the same simulation results as the results given in the previous section. But against reactive and pseudo random jammers, the CR succeeds to avoid the jammed channels with the minimum number of frequency switching.

5.5.1.1 Scenario with a reactive jammer

The CR succeeds to learn not only that it has to change its operating frequency every two time slots to escape from the reactive jammer, but also that starting from frequency f_3 he doesn't need to hop to the frequency f_2 as he does in figure 5.2(b). The Q^* matrix is given in table 5.4 and the ameliorated learned strategy is given in figure 5.4 when the CR starts from f_3.

5.5.1.2 Scenario with a pseudo random jammer

In this scenario, we consider jammers with the same pseudo random strategies as the previous section: the same random sequence $(f_1, f_3, f_2, f_4, f_2)$ of period $T = 5$ and the same random sequence $(f_1, f_1, f_4, f_3, f_2, f_1, f_3, f_3, f_4, f_2)$ of period $T = 10$. The CR succeeds not only to learn the pseudo random strategies of the jammers, but with the minimum number of frequency switching actions compared to figure 5.3. The Q^* matrix is given in table 5.5. The learned anti-jamming strategies are given in figure 5.5 when the periods of the pseudo random jamming sequences are respectively $T = 5$ and $T = 10$ time slots.

Algorithm 4 Channel selection using OPSQ-learning with modified reward function

Set the γ parameter and initialize the rewards to zero values.
Initialize the Q matrix to zero matrix.
while Not convergence **do**
  Select a random initial state $S = S_0$ and set $timeslot = 1$
  while The goal state hasn't been reached **do**
    Calculate the learning coefficient $\alpha = 1/timeslot$
    Select the best action a verifying $max_a Q(S, a)$ and corresponding to frequency f'.
    Using this action, consider going to the next state S'.
    Detect the jammed frequency f_{JX} %(due to wideband spectrum sensing)

    Update all the $Q(S, :)$ values, associated to the state S, by doing:
    for $i = 1 : M$ **do**
      observe the fictive state S_{tmp} of taking fictive action f_i
      if $f_i = f_{JX}$ **then**
        $R_{f_i}(S, S_{tmp}) = -1$ % (jammed)
      else
        if $f_i = f$ **then**
          $R_{f_i}(S, S_{tmp}) = 0$
        else
          if $f = f_{JX}$ **then**
            $R_{f_i}(S, S_{tmp}) = 0$
          else
            $R_{f_i}(S, S_{tmp}) = -1$ % (unneeded hop)
          end if
        end if
      end if
      Compute $Q(S, f_i) = (1 - \alpha)Q(S, f_i) + \alpha[R_{f_i}(S, S_{tmp}) + \gamma\, max_x Q(S_{tmp}, x)]$
    end for
    Set the next state as the current state $S = S'$ and increment $timeslot = timeslot + 1$.
  end while
end while

5.5.2 Comparison in terms of reward per trial

Until the convergence of the algorithm, we calculate for each trial (sequence of time slots that ends if there is collision with the jammer) the sum of reward values associated to each time slot. We calculate the average reward per trial for one hundred executions. The OPSQ algorithm learning a safe strategy (it takes the action

State \ Action	f_1	f_2	f_3	f_4
(f_1,1)	0	-0.727	-0.727	-0.727
(f_1,2)	-0.6287	0	0	0
⋮	⋮	⋮	⋮	⋮
(f_2,1)	-1	0	-1	-1
(f_2,2)	0	-0.4164	0	0
⋮	⋮	⋮	⋮	⋮
(f_3,1)	-1	-1	0	-1
(f_3,1)	-1	-1	0	-1
(f_3,2)	0	0	-0.75	0
⋮	⋮	⋮	⋮	⋮
(f_4,1)	-1	-1	-1	0
⋮	⋮	⋮	⋮	⋮

Table 5.4: The $Q*$ matrix against a reactive jammer with the ameliorated reward function

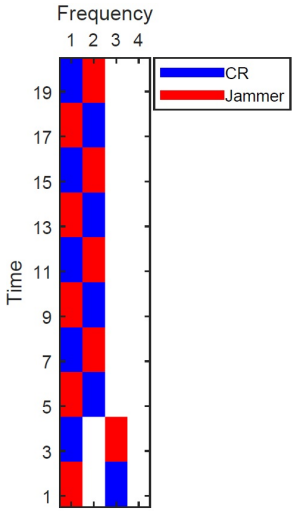
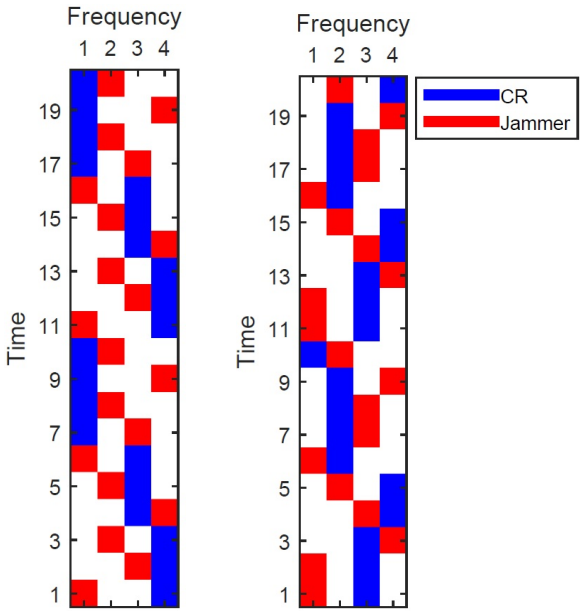


Figure 5.4: Exploitation of the optimal policy against a reactive jammer with the ameliorated reward function

selection method into account when learning) receives a higher average reward per trial than Q-learning as given in figure 5.6. Concerning the new reward strategy (reward R2) with two penalties (one for being jammed and the other for extra frequency switching), it results in reaching later the maximum level $reward = 0$ than the algorithm using the simple reward strategy (R1: according a penalty just for being jammed), but it results in gain in terms of frequency hopping as explained in the simulation results.

State \ Action	f_1	f_2	f_3	f_4
(f_1,1)	-0.4334	-0.996	-1.2795	-1.2748
(f_1,2)	-0.3871	-1.052	-0.9825	-0.7728
(f_1,3)	-0.1482	-0.7613	-0.9579	-0.9555
(f_1,4)	-0.6294	-0.2356	-0.1793	-0.1727
⋮	⋮	⋮	⋮	⋮
(f_2,1)	-0.643	-0.5728	-0.4717	-0.6321
⋮	⋮	⋮	⋮	⋮
(f_3,1)	-1.4148	-1.0277	-0.3695	-1.3616
(f_3,2)	-1.028	-1.397	-0.4091	-1.3918
(f_3,3)	-0.4521	-0.4877	-0.9343	-0.4569
⋮	⋮	⋮	⋮	⋮
(f_4,1)	-1.4152	-1.4481	-1	-0.3657
(f_4,2)	-1.3753	-0.9765	-1.3048	-0.3504
(f_4,3)	-0.415	-0.4296	-0.3436	-1
⋮	⋮	⋮	⋮	⋮

Table 5.5: The $Q*$ matrix in a pseudo random jammer (5TS) with the ameliorated reward function



(a) $T = 5$ (b) $T = 10$

Figure 5.5: Exploitation of the optimal policy against pseudo random jammers with the ameliorated reward function

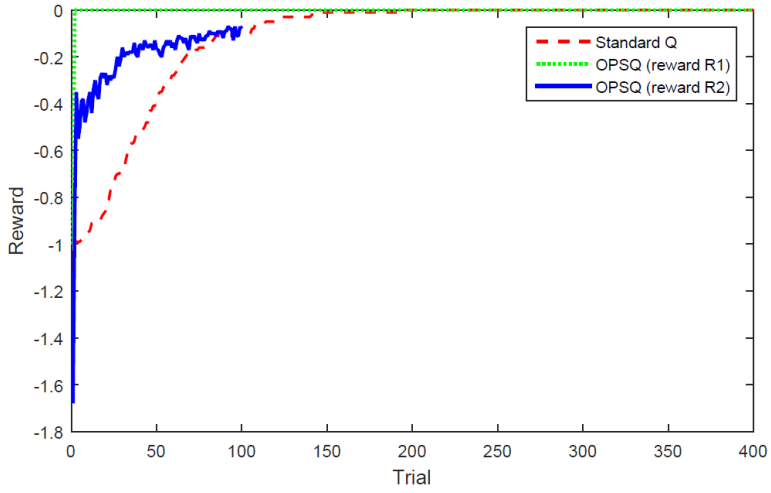


Figure 5.6: Comparison between the reward functions

5.5.3 Discussion

The MDP model presented in this chapter limits the convergence of the proposed OPSQ-learning algorithm to the scenario of a synchronized jammer in terms of time slot duration and starting time, as illustrated in all the previous figures. This is mainly due to the state definition which is related just to the CR node and is independent of the jammer's behavior. The spatio-temporal state definition allows the learning CR to stay the longest time in the same channel, but assumes that each time the CR revisit a state, the jammer will be in the same channel for the same period as during previous visits. Furthermore, the learning node may not detect the jammer that may be hidden and succeeds in jamming the packets. These limits will be addressed in the next chapter. In the next section, the proposed algorithm will be ameliorated and adapted to the multi-channel power allocation game studied in the previous chapter.

5.6 Learning based multi-channel power allocation game

We generalize the proposed OPSQ-learning algorithm to the multi-channel power allocation scenario, as given by algorithm 5. We will start by considering fixed jamming strategy (e.g. sweeping, pseudo random) before dealing with a smart jammer using the standard version of the Q-learning algorithm.

5.6.1 CR using OPSQ-learning against fixed jamming strategy

We consider a fixed jamming strategy, which means that the jammer doesn't change its jamming policy during the game. Precisely, we are considering a sweeping jammer which allocates its total power to one channel each time. We consider

that the CR's power p_k over each channel k can be selected from K levels. The multi-channel power allocation problem can be then modeled as a MDP with incomplete information about the jammer and the environment. The CR is using OPSQ-learning to learn the optimal strategy which gives the optimal power allocation that it should choose in each state. We define the state by the pair (f_{JX}, nb) with f_{JX} the jammed frequency detected through wideband spectrum sensing (WBSS) and nb the parameter indicating the number of successive occurrence of this jammed frequency. We have opted for mixing spatial and temporal properties in the state space definition to take into consideration a jammer staying more than one time in the same channel.

In each time slot, the CR chooses the action a (its power allocation vector $\mathbf{p}$) which corresponds to the maximum Q value in the current state. This action is given by the column having the maximum Q value in the row corresponding to the current state. The CR transmits with the power levels given by the chosen action $a = \mathbf{p}$, observes the new state S', cooperates with the receiver node to measure the fictive noise N_k corresponding to the normalized interference and noise in each channel

$$N_k = \frac{n_k + |g_k|^2 j_k}{|h_k|^2}, \tag{5.5}$$

and calculates the reward defined by the expression (5.6) inspired from the maximization problem of the previous chapter (section 4.5) to allow the comparison of the results. The reward corresponds to the total transmission capacity

$$R_a(S, S') = \sum_k F_k(\mathbf{p}, \mathbf{j}) \tag{5.6}$$

where

$$F_k(\mathbf{p}, \mathbf{j}) = log_2(1 + \frac{p_k}{N_k}). \tag{5.7}$$

Having the value of N_k, $\forall k$, the CR is able to do synchronous update of all the Q values in the row corresponding to the current state S, as follows

$$Q(S, a) = (1 - \alpha)Q(S, a) + \alpha[R_a(S, S') + \gamma \, max_x Q(S', x)] \tag{5.8}$$

5.6.2 Both the CR and the jammer using Q-learning

We consider that the powers p_k, j_k , for both of the CR and the jammer, in each channel k can be selected from K levels. Their interaction in terms of multi-channel power allocation can be modeled as a stochastic game which is a generalization of MDP [118]. The CR is using the OPSQ-learning as given by algorithm 5 against a jammer who is using an adapted version of the Q-learning as given by algorithm 6. Since the jammer interacts with the CR and changes its jamming

Algorithm 5 Multi-channel anti-jamming power allocation using OPSQ-learning

Set γ and α parameters.
Initialize the Q matrix to zero matrix.
Select a random initial action and observe the associated state S
while Not convergence **do**
  Select the best action a verifying $max_a Q(S, a)$ and corresponding to the power allocation **p**.
  Measure the fictive noise N_k in each channel by (5.5)
  Transmit using the power levels of the chosen action, measure the immediate reward as given by (5.6) and consider going to the next state S'.
  Update all the $Q(S, :)$ values, associated to the state S, by doing:
  for i $\in$ {the action set of the CR} **do**
    Observe the subsequent fictive state S_{tmp} of taking fictive action i
    Observe the fictive reward $R_i(S, S_{tmp})$ as given by (5.6)
    Update $Q(S, i) = (1 - \alpha)Q(S, i) + \alpha[R_i(S, S_{tmp}) + \gamma\, max_x Q(S_{tmp}, x)]$
  end for
  Set the next state as the current state $S = S'$
end while

strategy, we define the state of this interactive game by the pair (**p**, **j**) of the actions taken by the CR and the jammer.

In each time slot, the CR chooses the action (the power allocation vector **p**) which corresponds to the maximum Q value in the current state and the jammer chooses the action (the power allocation vector **j**) which corresponds to the minimum Q value in the current state. The two players transmit with the corresponding powers in each channel, observe the new state S' and calculate the immediate reward given by equation (5.6). Having the value of N_k, the CR updates all the Q values in the row corresponding to the current state S.

In this work, we consider that the jammer can measure the SINR value resulting from its action by observing the acknowledgment packets exchanged between the transmitter-receiver pair [119]. Then, it is able to calculate the immediate reward and the Q value related just to the current taken actions p_k and j_k; i.e. even having the capacity of WBSS, the jammer can not get the required information to update more than one Q value.

In real scenario, the jammer doesn't have the required information either to apply the Q-learning algorithm or to play the Nash game (described in chapter 4). According to its power allocation expression (4.13), the jammer needs to estimate the CR's power allocation p_k and make assumptions about the parameters n_k, h_k and g_k.

A trivial solution for the jammer would be to make the assumption of flat fading channels, otherwise he has to estimate the different channel coefficients which requires extra resources. Let $h_k = h$ and $g_k = g, \forall k$. He may consider $g = 1$, which corresponds to the scenario of being near the receiver node. Also, he may neglect

Algorithm 6 Multi-channel jamming power allocation using Q-learning

Set γ and α parameters.
Initialize the Q matrix to zero matrix.
Select a random initial action and observe the associated state S
while Not convergence **do**
  Select the action a verifying $min_a Q(S, a)$
  Observe the immediate reward
  Observe the subsequent state S'
  Update the Q value: $Q(S,a) = (1-\alpha)Q(S,a) + \alpha[R_a(S,S') + \gamma\, min_x Q(S',x)]$
  Set the next state as the current state $S = S'$
end while

the noise **n**. Furthermore, the jammer may consider that what he detects through spectrum sensing is equal to transmission power **p** multiplied by the channel gain **h**.

5.6.3 Simulation and discussion

We provide the power allocation results against a fixed jamming strategy and a smart interactive jammer with both complete and incomplete information.

5.6.3.1 OPSQ-learning against fixed jamming strategies

In this scenario, we consider fixed jamming strategy and we will compare the anti-jamming strategy of the CR applying the proposed OPSQ-learning to the waterfilling strategy.

Let's consider $M = 3$ channels. The action set of a sweeping jammer is defined by $A_{\mathbf{j}} = \{(J,0,0),(0,J,0),(0,0,J)\}$ with J as the total jamming power. To apply the OPSQ-learning algorithm, we consider four power levels for the CR: $\left\{0, P, \frac{P}{2}, \frac{P}{3}\right\}$ with P as its maximum power, so the CR may use one/two or the three available channels and its action set is,

$$\begin{aligned} A_{\mathbf{p}} = \{&(P,0,0),(0,P,0),(0,0,P),\\ &(\frac{P}{2},\frac{P}{2},0),(\frac{P}{2},0,\frac{P}{2}),(0,\frac{P}{2},\frac{P}{2}),\\ &(\frac{P}{3},\frac{P}{3},\frac{P}{3}) \end{aligned}$$

We consider $P = J = 10$ as the total CR's and jammer's power, the discount factor $gamma = 0.95$, the learning rate $\alpha = 0.1$. We will compare the learned solution to the waterfilling solution in both flat and selective fading channels. Also, we provide the learned solutions against other fixed jamming strategies.

a- Comparison between OPSQ-learning and waterfilling in flat fading channels scenario

As first case, we consider flat fading channels for both the CR and the jammer with equal channel gain coefficients $\mathbf{g} = \mathbf{h} = (1, 1, 1)$ and we consider also the same noise level in all the channels $\mathbf{n} = (1, 1, 1)$. We present in figure 5.7 the average payoff fluctuations during learning. Figure 5.8 gives the CR's actions resulting from the application of the learning algorithm against the sweeping jammer. The action indexes are varying from 1 to 7 as given in the action set $A_{\mathbf{p}}$ of the CR. According to these figures, after some collisions (in time slots 1 and 4) and some successful transmissions during about 12 time slots, the CR learns to follow the optimal strategy as given by table 5.6. For each action of the sweeping jammer, we mention the index of the CR optimal action and the corresponding power allocation found at the convergence of the OPSQ-learning algorithm. As given in this table, the power allocation resulting from the waterfilling strategy equals the power allocation learned using the proposed algorithm.

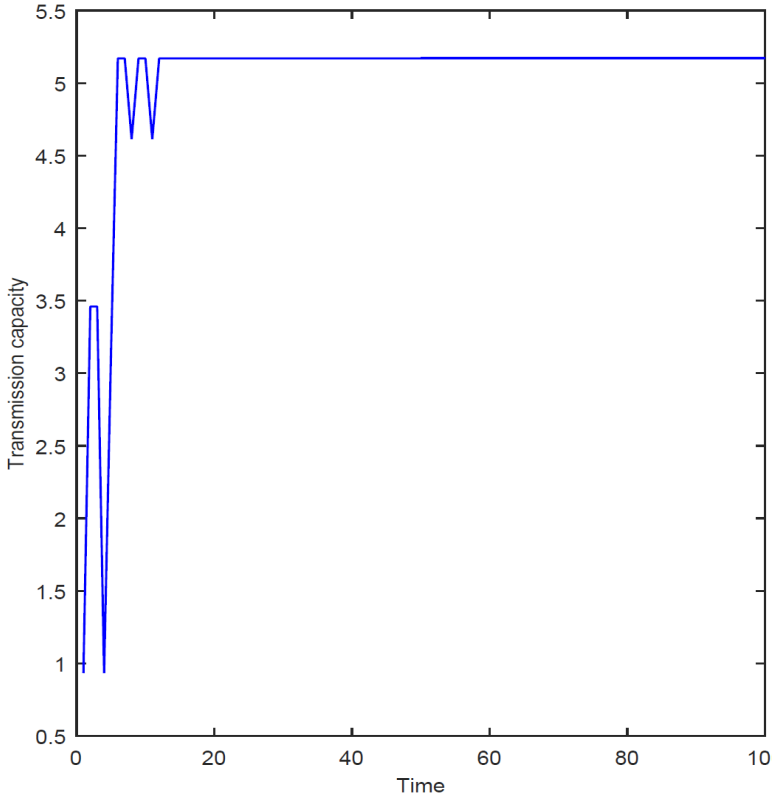


Figure 5.7: The transmission capacity over flat fading channels in the presence of sweeping jammer

Jx	index	1	2	3
	power	(10,0,0)	(0,10,0)	(0,0,10)
OPSQ-learning	index	6	5	4
	power	(0, 5, 5)	(5, 0, 5)	(5, 5, 0)
waterfilling		(0, 5, 5)	(5, 0, 5)	(5, 5, 0)
Capacity		5.1699	5.1699	5.1699

Table 5.6: CR using OPSQ-learning/waterfilling against sweeping jammer over flat fading channels

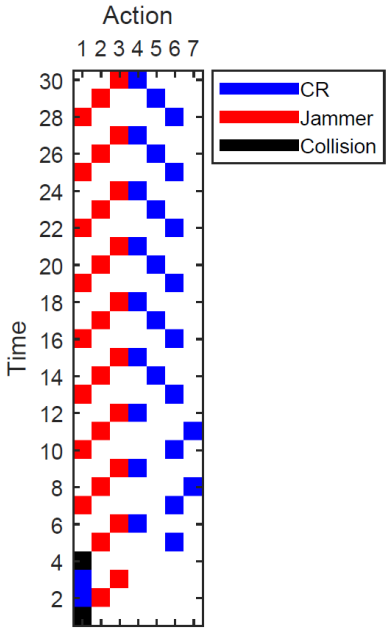


Figure 5.8: The learned anti-jamming strategy against sweeping jammer over flat fading channels

b- Comparison between OPSQ-learning and waterfilling in selective channels scenario

In this scenario, we consider selective channels for both the CR and the jammer with the channel gain coefficients $\mathbf{g} = \mathbf{h} = (2, 1, 3)$ and we consider the noise vector $\mathbf{n} = (2, 3, 1)$. We have chosen these values to make channel 3 better than channel 1 which is better than channel 2, for both the CR and the jammer.

Figure 5.9 gives the CR's actions applying the OPSQ-learning algorithm. After some collisions (in time slots 1 and 4) and some successful transmissions during

about 6 time slots, the CR learns to follow the optimal strategy as given by table 5.7. Considering the same parameters, the waterfilling solution against each action of the jammer results in a power allocation close to the solution found by using the OPSQ-learning algorithm; the two solutions avoid almost the same channels but differs slightly in the allocated power levels and the payoff values. This difference is due to the number of possibilities (i.e. the power levels) which is infinite for the waterfilling strategy and finite for the proposed algorithm.

Jx	index	1	2	3
	power	(10,0,0)	(0,10,0)	(0,0,10)
OPSQ-learning	index	6	5	4
	power	(0, 5, 5)	(5, 0, 5)	(5, 5, 0)
Capacity (OPSQ)		6.9386	8.983	4.8745
waterfilling		(0, 3.5556, 6.4444)	(4.8056, 0, 5.1944)	(6.25, 3.75, 0)
Capacity (waterfilling)		7.0104	8.9849	4.9248

Table 5.7: CR using OPSQ-learning/waterfilling against sweeping jammer over selective channels

c- OPSQ-learning against other fixed jamming strategies

Applying the proposed learning algorithm, the CR succeeds to learn anti-jamming power allocations against a sweeping jammer staying in the same channel for two and three time slots, as given by figure 5.10.

5.6.3.2 OPSQ-learning against a jammer using Q-learning

In this scenario, we consider the same previous simulation parameters. Here, the CR applies the OPSQ-learning as given by algorithm 5 and the jammer uses the Q-learning algorithm 6. We consider that the jammer has the same action set as the CR: $A_{\mathbf{j}} = A_{\mathbf{p}}$. We will compare the strategies learned by the CR and the jammer to the optimal strategies found at the convergence of the Nash game under complete information which is detailed in chapter 4.

Figure 5.11 gives the payoff fluctuations during learning. After about 13000 time slots, the CR's and the jammer's actions (power allocations) are no longer fluctuating and the transmission capacity reaches the fixed value $C = 2.4859$. The jammer's final power allocation is $\mathbf{j} = (5, 0, 5)$ and the CR's final power allocation is $\mathbf{p} = (\frac{10}{3}, \frac{10}{3}, \frac{10}{3})$.

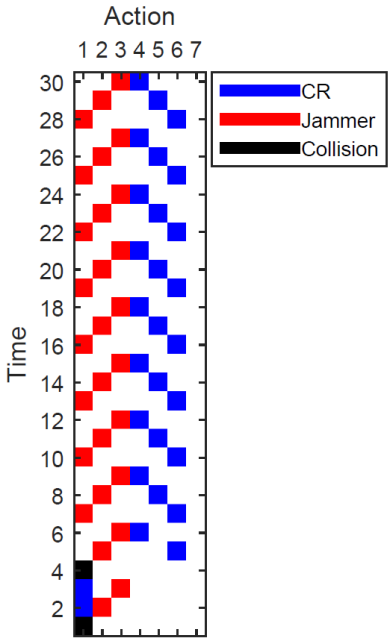


Figure 5.9: The learned anti-jamming strategy against sweeping jammer over selective channels

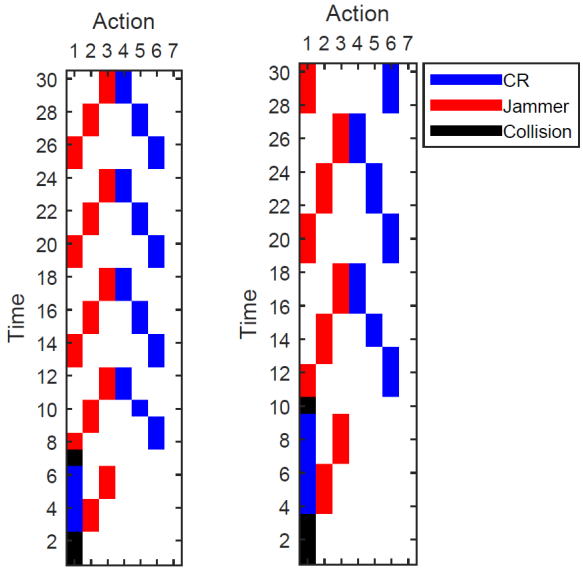


Figure 5.10: The learned anti-jamming strategies against sweeping jammer attacking the same channel for 2 TSs and 3 TSs

In the Nash game, the CR uses the waterfilling expression (4.7) and proceeds by bisection until reaching the value of λ corresponding to the alloca-

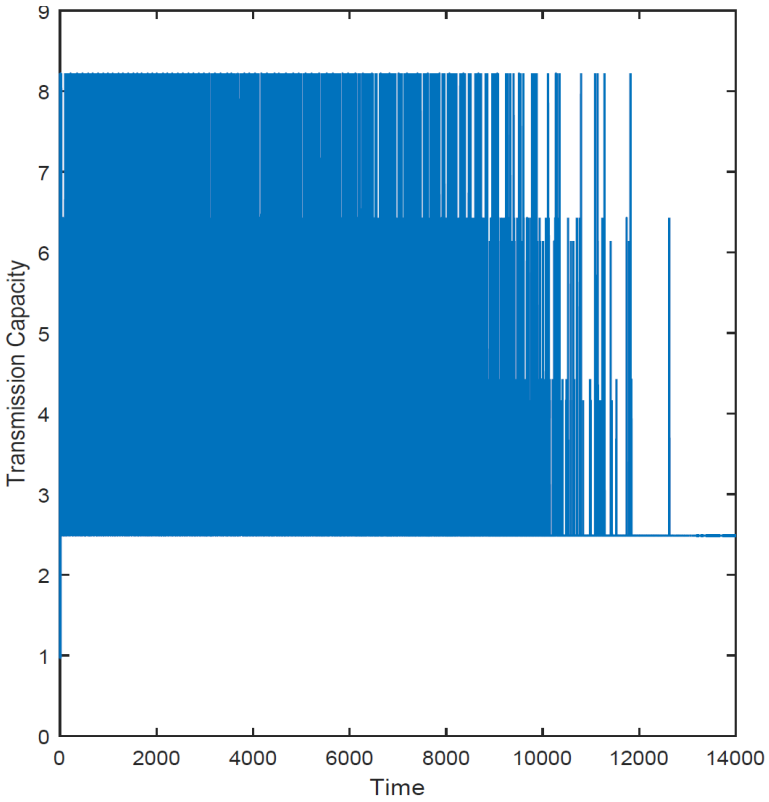


Figure 5.11: The transmission capacity over selective channels against a jammer using Q-learning

tion of the total power. The jammer proceeds by bisection and calculates the sum of the allocated powers to all the channels using the expression (4.13) until reaching the value of μ corresponding to the allocation of the total jamming power. At the NE of the described game, we get (after 37 iterations) the jammer's power allocation $\mathbf{j} = (4.0370, 1.5370, 4.4259)$ and the CR's power allocation $\mathbf{p} = (3.3333, 3.3333, 3.3333)$ with the transmission capacity $C = 2.384$, as given by figure 5.12. This result is close to the result found using OPSQ-learning for the CR and Q-learning for the jammer.

5.6.3.3 The game against a jammer with incomplete information

We consider the Nash game in two scenarios; (1) the jammer has complete information, (2) the jammer does the assumptions of flat fading channels, being near the receiver node and neglected noise level. We present in table 5.8 a comparison between the NEs of the two scenarios.

These results corresponds to the channel coefficients $|h| = (1.9821, 0.9848, 3.3178)$ and $|g| = (0.533, 0.0985, 1.1683)$. Hence, the jammer having complete information avoids bad channels (e.g. channel two since

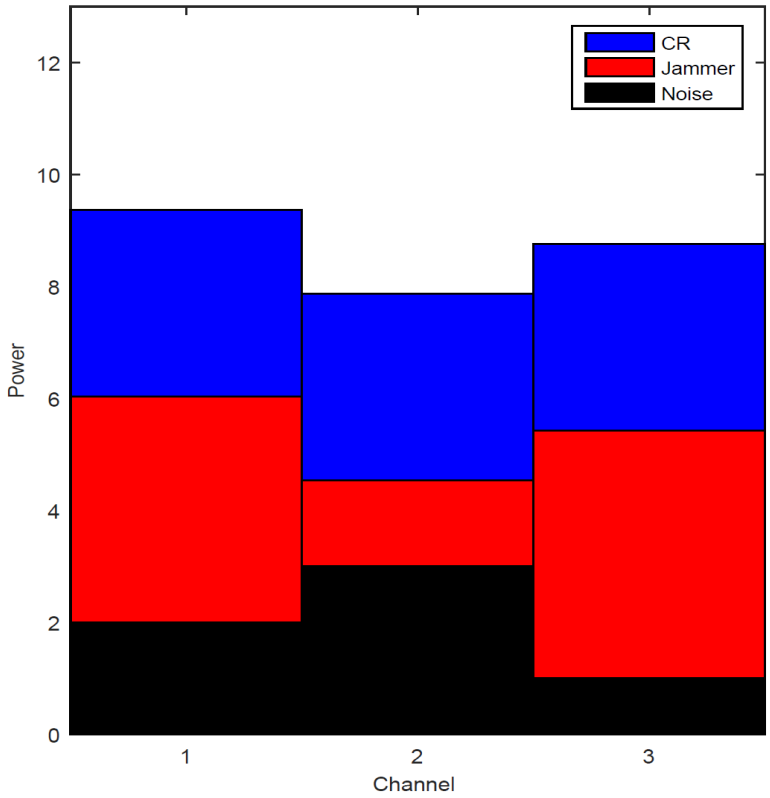


Figure 5.12: The jamming and anti-jamming strategies at the NE

it has low gain coefficient) even if it used by the CR. This allows the jammer to attack the other channels with higher powers which reduces the total channel capacity of the CR. Without complete information, the jammer occupies all the channels with almost the same power level which results in a limited payoff gain for the CR (i.e. limited loss in the effectiveness of the jamming attack).

	NE under perfect knowledge	NE under imperfect knowledge
p	(3.3184, 3.5958, 3.0858)	(3.3316, 3.4675, 3.2009)
j	(4.9801, 0, 5.0198)	(3.4679, 3.2153, 3.3167)
Capacity	10.8422	11.4391

Table 5.8: Knowledge effect on the NE

5.7 Conclusion

In this chapter, we have proposed a modified Q-learning algorithm to solve the jamming attack. The proposed algorithm is based on wideband spectrum sensing and in a greedy policy which allow its real-time application. It is denoted OPSQ-learning algorithm. In a first time, the proposed solution was applied in terms of one channel selection to learn how to pro-actively avoid the jammed channels. In a second part time, it was adapted to solve the multi-channel power allocation game. In the first application, we have modeled the scenario of fixed jamming strategy as a MDP model. To learn the optimal anti-jamming strategy, we have ameliorated the reward function in order to stay as longer as possible in the same frequency and minimize the number of frequency switching. We have presented the simulation results against sweeping, reactive and pseudo random jamming strategies. We can conclude that the OPSQ-learning version speeds up the learning period and the ameliorated reward strategy optimizes the number of channel switching which enhance its application during CRN real time communication. For the multi-channel application, we have started by a fixed jamming strategy and found that the learned solution almost equals the common explicit waterfilling solution. Furthermore, we considered a smart jammer using the Q-learning algorithm. The learned jamming and anti-jamming power allocation strategies are almost equal to the optimal Nash equilibrium strategies found under the assumption of complete information, studied in the previous chapter. Finally, we studied the real scenario when the jammer has incomplete information about the CR user and the channel gain coefficients. Under this condition, the jammer occupies all the channels with almost the same power level which results in a limited payoff gain for the CR.

The presented MDP and the proposed OPSQ-learning algorithm, for channel selection, present limits in terms of synchronization requirement with the jammer and do not solve the scenario of hidden jammers.

The next chapter presents an enhanced version of the proposed OPSQ-learning algorithm to overcome hidden jammer problem and take into consideration asynchronous jammer behavior, in terms of channel selection. Furthermore, the ameliorated OPSQ-learning algorithm will be tested in real environment using a software defined radio platform.

Chapter 6

Cooperative learning based anti-jamming technique

6.1 Introduction

the OPSQ-learning algorithms presented in previous chapter assumed both CR and jammer are synchronous and able to perform accurate WBSS. In this Chapter we propose realistic versions and implement them on state of art SDR platforms in real scenarios. We solve practical problems like asynchronous CR/JX behavior, realistic reward function based on real time WBSS results and hidden node problem. We start by performing high-fidelity simulations with fine time granularity and down to the level of IQ to allow realistic sensing. These simulations are used as reference for evaluation of SDR implementation performance. We provide both simulation results and real measurements in terms of packet success rate (PSR). The proposed DSA algorithm significantly improves the packet success rate compared to both static spectrum access and intelligent spectrum access without learning.

The rest of this chapter is organized as follows: Section 6.2 describes the new MDP model and presents the cooperative Q-learning algorithm. Section 6.3 details the programming setup and discusses the simulation results. Section 6.4 presents the hardware environment and discusses the real measurements. Finally, section 6.5 summarizes the conclusions.

6.2 Cooperative learning algorithm

This section presents the new choices in terms of state & reward definition and the enhanced version of the learning algorithm. We will explain how these modifications solve the practical problems like the synchronization between CR & jammer and hidden node problem.

6.2.1 State definition

In the previous chapter, the state of the CR was defined by a pair of parameters; (1) For channel selection against fixed synchronized jammer in 5.4, the definition $S = (f_{TX}, n)$ reflects just the CR behavior where f_{TX} is its current operating frequency and n is the number of successive time slots using this frequency, (2) to find the optimal power allocation in the presence of a fixed jammer attacking a single channel each time in 5.6.1, the definition $S = (f_{JX}, nb)$ is related to the occurrences of the jamming frequency, (3) to find the optimal power allocation in the presence of an interactive jammer in 5.6.2, the definition $S = (\mathbf{p}, \mathbf{j})$ reflects the interaction in terms of power allocation. In order to take into consideration the asynchronous jammer behavior, including its random starting time and its unknown channel occupancy, we define the state with the triplet $S = \{f_{TX}, n, f_{JX}\}$ where f_{TX} is the best idle channel, f_{JX} is the worst jammed channel and n is the number of successive occurrences of f_{TX} for the same f_{JX}. This new definition covers the real scenario of a jammer attacking any channel any time and for any period. We consider that, at each time slot, the CR is able to do wideband spectrum sensing to detect the worst (jammed f_{JX}) and the best (idle) channels among M channels defined as the possible actions: $\{a_1, \cdots, a_M\} = \{f_1, \cdots, f_M\}$.

6.2.2 Reward function

We used either basic reward functions (in algorithms 3 and 4), assuming a jammer in one channel at a time, or ideal one (equation 5.6) assuming accurate evaluation of noise, fading and jamming power. Here we consider a realistic reward based on spectrum sensing result:

$$R_f(S, S') = 1 - \frac{E(f)}{ET}, \tag{6.1}$$

where $E(f)$ is the energy measured over the channel f and ET is the total energy measured over the M channels. Such realistic reward function adapts the CR channel selection to the real time spectrum occupancy and allows a pro-active collision avoidance; i.e. an occupied or jammed channel will carry high energy which corresponds to low reward and an idle channel having low energy will be assigned a high reward.

6.2.3 Learning through cooperation

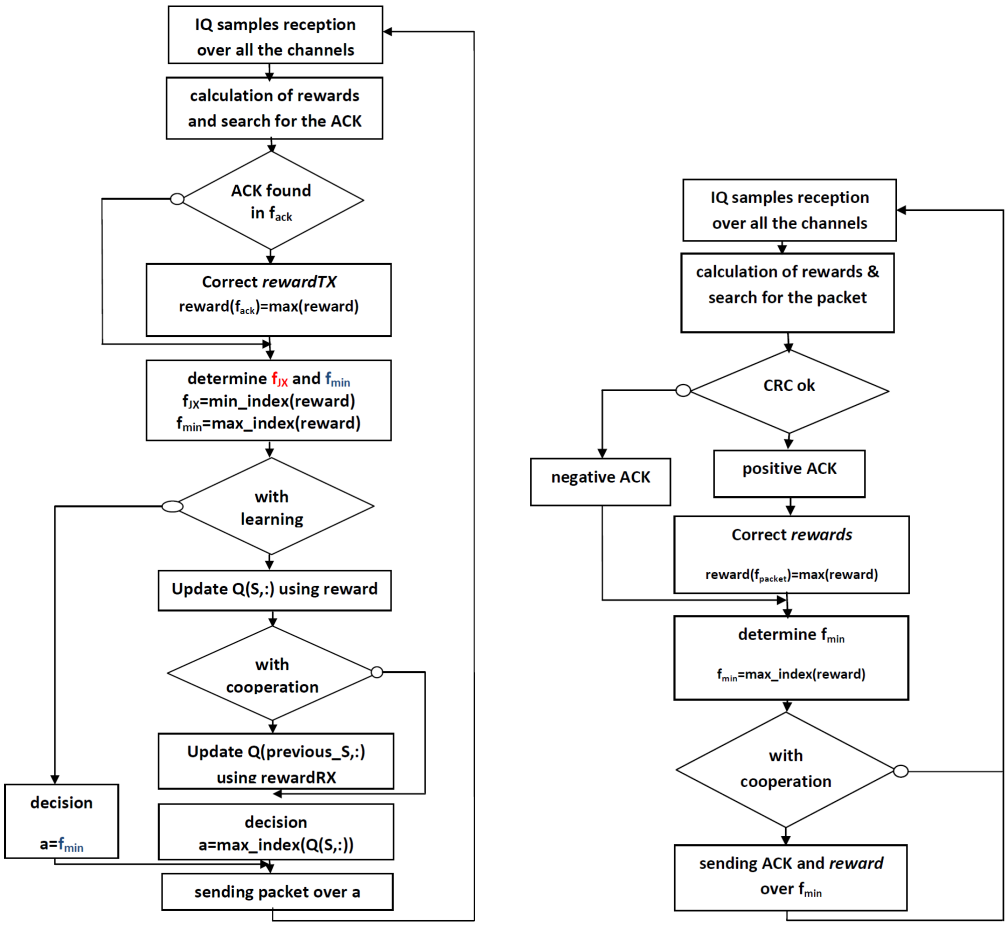
In our previous channel selection technique, the learning algorithm depends only on the sensing results of one CR node, i.e. the learning node. This reveals the problem of hidden jammers that may interfere the transmitted packets without being detected. We enhance the proposed solution via the cooperation with the node receiving the packets. This latter transmits the acknowledgment including its sensing results. This feedback about both the packet success and the spectral occupancy may correct the learned strategy since the learning node updates the Q values based

on both its sensing and the received sensing results which gives more vision about the actual and the previous channels occupancy. The proposed solution is described in algorithm 7, using $R_a^l(S, S')$ to denote the local reward $R_a(S, S')$ measured by the learning node in the current state S for each possible action a that results in a next state S'. Likewise, $R_a^r(S_p, S_p^{'})$ represents the received reward $R_a(S_p, S_p^{'})$ measured by the cooperative node during the reception of the previous packet.

Algorithm 7 pseudocode of cooperative OPSQ-learning

Select a random initial state $S = S_0$
while true **do**
The learning node does WBSS and checks for acknowledgment reception
Update all Q values at the current state S based on the local WBSS and the previous state S_p based on the received WBSS results using,$\forall a$:
$Q(S, a) = (1 - \alpha)Q(S, a) + \alpha(R_a^l(S, S') + \delta max_x Q(S^{'}, x))$
$Q(S_p, a) = (1 - \alpha)Q(S_p, a) + \alpha(R_a^r(S_p, S_p^{'}) + \delta max_x Q(S_p^{'}, x))$
Select an action a with max Q value
Take a and observe next state $S^{'}$
$S_p = S$
$S = S^{'}$
end while

Figure 6.1-(a) details the tasks performed by the learning node. The first step consists in gathering the IQ samples through wideband reception over the considered M channels. Then the rewards associated to all the possible actions are calculated using equation 6.1 based on energy detection to perform WBSS. During this processing step, the learning node looks blindly for an acknowledgment over the M considered channels without a rendez-vous or a signaling channel. If an acknowledgment is received over a channel f_{ack}, the reward calculated for that channel carrying the acknowledgment should not keep its low value (since it has high energy $E(f_{ack})$) to not falsify the decisions and be considered as a jammed channel. For that, the learning node associates to this channel the maximum reward that he has calculated. The next step consists in deciding which is the jammed channel and which is the best one (having the maximum reward). If we opt for channel selection based on sensing without learning (denoted as the best channel selection), the learning node selects the channel having the minimum energy in each time slot. If we opt for the proposed OPSQ-learning algorithm without cooperation with the receiving node, the transmitter updates just the Q values related to the actual state $Q(S, :)$. The action having the maximum Q value $a = max\_index(Q(S, :))$ is selected to transmit the packets. The last channel selection strategy consists in cooperating with the node receiving the packets to have more knowledge. So, the learning node updates not only the actual state but also the previous state based on the reward values extracted from the acknowledgment. The received rewards are related to the previous time slot when the destination node has received the transmitted packet. If the learning node does not receive the acknowledgment, he



(a) The learning node (b) The cooperative node

Figure 6.1: Cooperation diagrams

considers that the response was jammed or lost and considers null received rewards. Finally, the learning node selects the channel having maximum Q value. For each of the channel selection strategies, the packet is sent over the selected channel.

Figure 6.1-(b) describes the operations of the CR node receiving the transmitted packets. After a wideband reception of the IQ samples, the channels rewards are calculated based on the detected energies through WBSS. This node looks blindly for the packet over the considered channels and performs the cyclic redundancy check (CRC). If the packet is received correctly over a channel f_{packet}, the CR node decides to send a positive acknowledgment. He also corrects the reward that he calculated for that channel to the maximum reward since it is not a jammed one. If the CRC is false, a negative acknowledgment is sent. In both cases, he selects the best channel having maximum reward to send the ACK sign and the rewards if we have selected the cooperative strategy.

6.3 Simulation setup and results

To evaluate the proposed algorithm, we have compared four channel selection strategies; The first strategy is the classical fixed channel selection that consists in transmitting over the same channel all the time with neither sensing nor learning. The second one is the best channel selection based on sensing without learning. The third strategy is based on the proposed OPSQ-learning algorithm but without cooperation. The last strategy consists in learning with the cooperation of the node receiving the packets. This section concerns MATLAB simulation of the considered four channel selection strategies, it starts by detailing the simulation setup before discussing the results.

6.3.1 Simulation setup

We have opted for a high fidelity simulation which provides the flexibility to adjust the CR configurable parameters according to the chosen strategy and to the electromagnetic environment without abstractions of the physical layer [120]. In layer 1, we manipulate the CR parameters such as the frequency, the bandwidth, the power, and we make the choices of energy detection as the spectrum sensing technique, dynamic spectrum access in overlay mode and binary phase shift keying (BPSK) as modulation scheme. In layer 2 (data link layer), our media access control (MAC) is based on time division multiple access (TDMA). We have also flexibility in terms of the jammer parameters and higher time granularity to introduce asynchronous behavior and to have a kind of reference for the comparison with real measurements on SDR platform. Furthermore, this allows going down to the level of IQ samples and includes signal processing details such as spectrum sensing, frame construction and real modulation & demodulation. The simulation setup is given in table 6.1.

For the transmission side, the modulated signal is upsampled and filtered using a root raised Cosine (RRC) pulse shaping filter. Then, the baseband pulse train is multiplied by a sinusoidal carrier considering the chosen frequency among the four available frequencies. The BPSK modulated signal is transmitted over an AWGN channel. At the reception side, the received BPSK signal is blindly downconverted to baseband considering each time one of the four channels, since the receiver does not know the chosen transmitting channel. After checking for the preamble (of the packet or the acknowledgment), the baseband signal carrying the preamble is downsampled and filtered by a RRC receiving filter equal to the RRC transmitting one. The samples are finally passed to the BPSK demodulator to take decisions about the received bits.

In the following results, the reception period is equal to the transmission period $T_{RX} = T_{packet} = T_{ACK} = 0.98ms$. We will start by considering a slow sweeping jammer with a dwell time of $T_{JX} = 2.28ms$ which corresponds to $T_{JX} \approx 2.3T_{packet}$ as represented in figure 6.2.

Parameter	Value
Modulation	BPSK
Bandwidth of the transmitted signal (packet/acknowledgment)	$12KHz$
Number of channels	4
Maximum successive occurrences of the same channel	10
Size of the Q matrix	$10 * 4 * 4 = 160$ rows and 4 columns
Learning rate α	0.1
Discount factor γ	0.1

Table 6.1: Simulation setup

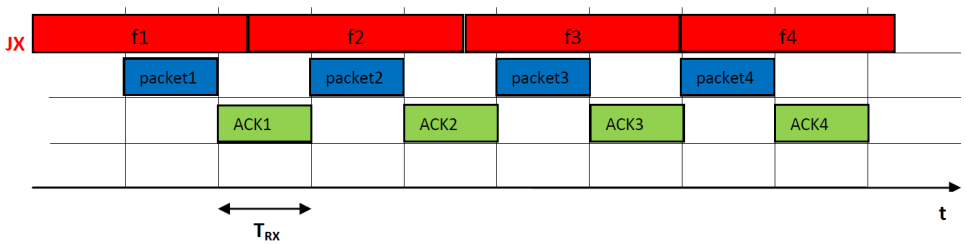


Figure 6.2: Simulation scenario

6.3.2 Simulation results

The node receiving the packets and performing CRC measures the PSR for the four channel selection strategies as given in table 6.2. The four strategies corresponds to the four rows of the table. The columns of the table corresponds to two scenarios depending on the visibility of the jammer to the learning node. The PSRs are given for 1000 transmitted packets.

In the first scenario (column 1 corresponding to a jammer detectable by both of the CR nodes), considering the slow sweep jammer, we get that the best channel selection based just on spectrum sensing (row 2) gives higher success rate than the fixed channel selection (row 1) since this latter is a blind selection staying on the same channel all the time without any information about the channels occupancy. Then, the channel selection based on OPSQ-learning (row 3) is better than selecting the best channel without learning (row 2) since the learning decision is not only based on the actual information but also on the past learned information $((1-\alpha)Q(S,a))$ and on the future expectation $(\alpha\gamma max_x Q(S',x))$ as given in the

	Jammer detectable by the learning node	Jammer hidden to the learning node
Classical fixed channel selection	66.6%	66.6%
Best channel selection without learning	80%	66.6%
Learning without cooperation	82.8%	66.6%
Learning with cooperation	96.8%	84.4%

Table 6.2: Simulation results: Packet Success Rate against slow sweep jammer

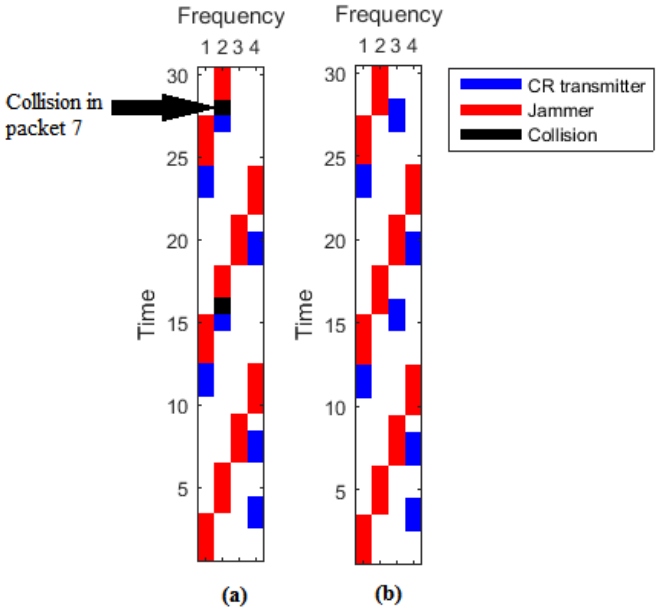


Figure 6.3: Simulation of best channel selection based on sensing (a) versus channel selection based on learning (b) against a sweeping jammer

expression of Q value update. Finally, learning with cooperation (row 4) outperforms learning without cooperation (row 3) since the cooperative node gives more information to the learning node about the jammer that may be not detected during its sensing period but appears during the transmission of the packet.

In the second scenario (column 2 corresponding to a jammer hidden from the learning node), choosing the best channel with or without learning (row 2 or row 3) are similar to staying in the same channel (row 1) since both best channel selections are based only on the sensing result of the learning node. When the destination node cooperates with the learning node (row 4), the PSR increases since the cooperative node gives an information about the channel used for the previous packet transmission: packet success implies the jammer absence and packet failure means collision with the jammer.

Figure 6.3 gives the channels occupancy for each of the learning node and the sweeping jammer over time for both the second and the third strategies. The best channel selection without learning, given in subfigure (a), results in losing more packets than the strategy based on OPSQ-learning presented in subfigure (b). For example, we consider packet number seven as indicated in the figure. The wideband spectrum sensing gives the following reward vector for both of the strategies: $reward = (0.4145; 0.9982; 0.9981; 0.5892)$, the best channel selection strategy results in the selection of the second channel resulting in collision with the jammer. However, the on-line learning algorithm calculates the $Qvalues = (0.0228; 0.1896; 0.1898; 0.1679)$. Applying the proposed learning algorithm, the third channel having the maximum Q value is selected, as presented in subfigure 6.3-(b).

According to the presented results, the cooperative OPSQ-learning (row 4) outperforms learning without cooperation. Moreover, a CR applying the proposed OPSQ-learning succeeds better than a CR just sensing the spectrum to select the best channel, if the jammer is detectable. However, these success rates especially for the cooperative learning strategy depends on the jammer's period and tactic. **In terms of the jamming period**, we have considered a faster jammer with a dwell time larger than the sensing period but lower then the sensing plus transmission periods of the learning node. The simulation results, given in table 6.3, give the same conclusions as the results against the slow sweep jammer. The noteworthy difference concerns the best channel selection without learning (row 2) which gives lower PSR than the three other strategies even the fixed channel selection. This is due to the fast sweep jammer which may be detected by the CR node in one channel during the sensing period but moves to another channel during the transmission period.

	Jammer detectable by the learning node	**Jammer hidden to the learning node**
Classical fixed channel selection	73.3%	73.3%
Best channel selection without learning	65.5%	73.3%
Learning without cooperation	77.3%	73.3%
Learning with cooperation	86%	88.7%

Table 6.3: Simulation results: Packet Success Rate against fast sweep jammer

In terms of the jamming tactic, we have applied the proposed solution against both a pseudo random jammer and a reactive one as described in figures 6.4 and 6.5. The packet success rates are given in tables 6.4 and 6.5. Concerning the pseudo random jammer, we have considered a sweep over a sequence of six channels $\{f_1, f_4, f_3, f_3, f_2, f_4\}$. We present a detailed example in figure 6.4 to compare between the OPSQ-learning strategy without (a) and with cooperation (b). Packet

number 19 is lost using learning without cooperation and received successfully when the learning node cooperates with the receiving node. Using learning without cooperation, the learning node stayed in channel 1 since he calculates the Q values $(0.1017; 0.0624; 0.0997; 0.04)$. But, when cooperating with the destination node, he calculates the Q values $(0.1844; 0.1717; 0.1925; 0.0705)$, so he takes the decision to use channel 3 for packet 19. There are other collisions like for packet 20 which are not avoided even through the cooperation, because the Q values corrections will be considered when the states (from which wrong decisions were taken) are revisited. Concerning the reactive jammer, we have considered an intelligent jammer which is capable to do spectrum sensing to jam the detected occupied channel. We assumed that this jammer needs a duration of two time slots before jamming the detected frequency, because it has to do the spectrum sensing, then make the decision and finally hop to the detected frequency. Against such jammer when hidden to the learning node and detected just by the packet receiver node, the three first strategies gives a PSR of 1% saving just the first packet as represented in figure 6.5-(a). The last strategy based on learning and cooperating gives a higher PSR as illustrated by figure 6.5-(b). Without the cooperation, the learning node updates the Q-matrix based only on the local sensing results, which results in staying in the same channel with the reactive non detected jammer. Figure 6.5-(b) illustrates the reactive behavior of the jammer who starts in channel 1 then jams two times channel 4 since he detects two transmissions over this channel, and so on.

The results of tables 6.4 and 6.5 against pseudo random and reactive jammers confirm the same conclusions as the results of table 6.2 against a sweeping jammer; The channel selection based on the cooperative OPSQ-learning algorithm outperforms the three other considered channel selection strategies for both scenarios of visible and hidden jammer. Furthermore, in the first scenario of detectable jammer, the OPSQ-learning strategy without cooperation outperforms the best channel selection strategy without learning which also outperforms the fixed channel selection. The three strategies gives the same packet success rate if the jammer is hidden to the learning node.

	Jammer detectable by the learning node	**Jammer hidden to the learning node**
Classical fixed channel selection	53.7%	53.7%
Best channel selection without learning	77.6%	53.7%
Learning without cooperation	89.5 %	53.7%
Learning with cooperation	99.4 %	74.5 %

Table 6.4: Simulation results: Packet Success Rate against pseudo random jammer

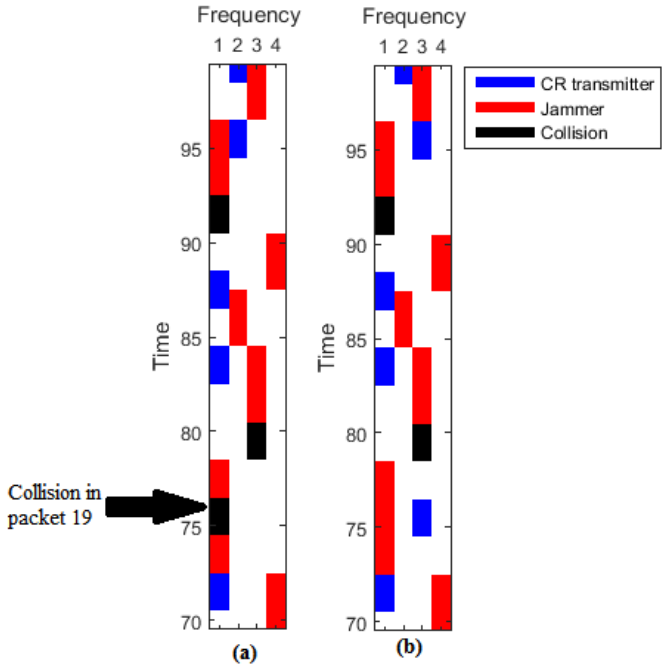


Figure 6.4: Simulation of channel selection based on learning (a) versus channel selection based on cooperative learning (b) against a pseudo random jammer

	Jammer detectable by the learning node	Jammer hidden to the learning node
Classical fixed channel selection	1%	1%
Best channel selection without learning	96%	1%
Learning without cooperation	97%	1%
Learning with cooperation	97.6%	66.9%

Table 6.5: Simulation results: Packet Success Rate against reactive jammer

6.4 Experimental setup and measurements

This section details the programming setup and describes the hardware environment. Furthermore, it discusses and compares the real measurements to the simulation results presented in the previous section.

6.4.1 Software development and hardware environment

We have implemented the physical layer signal processing steps and the four channel selection strategies described previously using Qt Creator/C++ development environment and the Universal Software Peripheral Radio platforms USRP E110

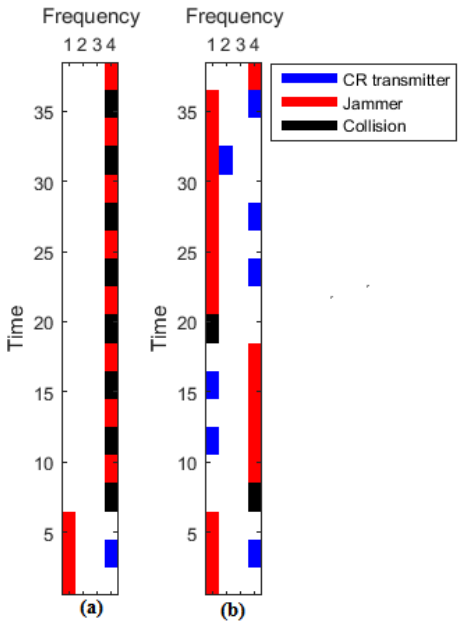


Figure 6.5: Simulation of channel selection based on learning (a) versus channel selection based on cooperative learning (b) against a hidden reactive jammer

and B205mini, see figure 6.6. The physical layer is based on BPSK modulation

Figure 6.6: Experimental environment

over the four channels: $(432.94; 432.98; 433.02; 433.06)MHz$. Without loss of generality, we have opted for the stop and wait scheme described in figure 6.7, but the presented study can be applied to any time division multiplexing (TDM) scheme. The learning node does wideband reception of the IQ samples during the reception period T_{RX} detecting acknowledgment and jamming signals. The time needed to do blind search of the ACK (for the learning node) or the packet (for the cooperative node) over the M channels is denoted $T_{process}$. After sending the packet, the learning node waits until the end of the cooperative node processing before returning to the reception step. The cooperative node respects the

same doctrine to keep synchronized with the learning node. We call radio period the sum of the reception, transmission and packet/ack processing periods: $T_{radio} = T_{RX} + 2 * T_{process} + T_{TX}$. Figure 6.8 describes the alternation between the packet and the acknowledgment transmissions by USRP nodes.

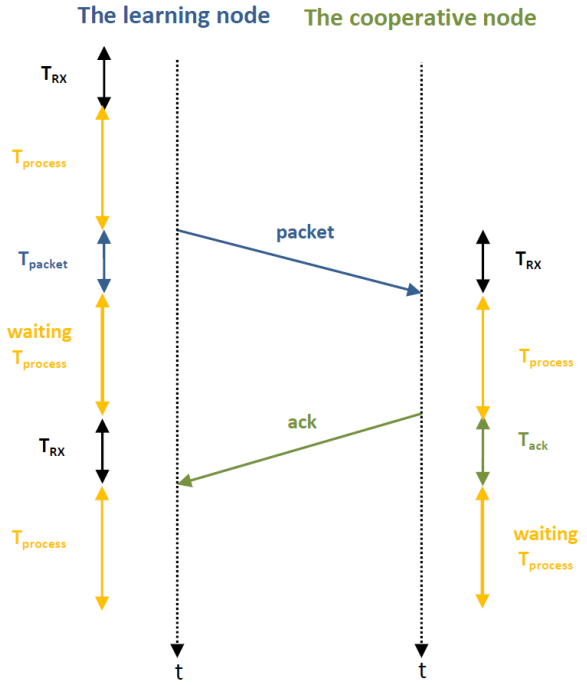


Figure 6.7: Cooperation based on stop and wait protocol

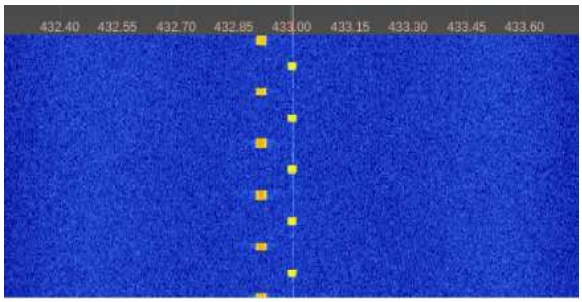


Figure 6.8: Cooperation spectrum

The packet and acknowledgment structures are given in figures 6.9 and 6.10 respectively. The transmitted data is delimited by a preamble and an end delimiter. After the addresses of the transmitter (@Source) and the receiver (@Destination), we mention the number of the transmitted packet or the acknowledgment. As a part of the payload in the transmitted acknowledgment, we find the ACK sign which is the character '+' or '-' and we find also the vector of four reward values calculated

by the cooperative node. Before the end delimiter, we have the CRC field of 32 bits. To deal with signal impairments due to the transmission in real environment and real conditions, we have used the synchronizer MPSK_Receiver which is a GNU Radio C++ signal processing block. This block performs carrier frequency and phase synchronization as well as symbol timing recovery.

Packet Preamble	@Source	@Destination	Packet Number	Delimiter	Payload	CRC	Packet End

Figure 6.9: Packet structure

Acknowledgment Preamble	@Source	@Destination	Acknowledgment Number	Delimiter	ACK	Rewards	Payload	CRC	End

Figure 6.10: Acknowledgment structure

We have considered two scenarios; scenario 1 of a jammer detectable by both the learning node & the receiving node and scenario 2 of a jammer hidden from the learning node. The tests were performed in the Royal Military Academy (RMA) in Brussels where the USRP platforms were placed in different buildings as described in figures 6.11 and 6.12. The jammer was running standalone at start up of USRP E110, the reporting node code was transferred to an Odroid-U3+ connected to one of the two used USRP B205mini, and the learning node was running on a laptop connected to the other USRP B205mini platform. To connect to the USRP platforms and adjust its parameters, we use the C++ application programming interface (API) provided by the USRP hardware driver (UHD).

6.4.2 Software defined radio measurements

The PSR measured by the CR receiving the packets is given in table 6.6 for the four considered strategies in both scenarios of a jammer detectable (scenario 1) or hidden (scenario 2) from the learning node.

The real measurements show that the cooperation ameliorates the PSR for both scenarios since the learning node receives the sensing result measured by the cooperative node which helps in learning the jammer's behavior. Without cooperation, the learning node gains in terms of PSR only if he detects the jammer since the proposed learning algorithm is based on the sensing results. Otherwise, the learned strategy has the same PSR as the fixed and the sensing based strategies. Figure 6.13 gives the best channel selection based on sensing (a) and the channel selection based on learning (b) in the presence of the sweeping jammer. Based just on sensing, the strategy presents wrong decisions due to the asynchronous jammer behavior. This latter may be detected in a channel during the sensing period, but it moves to another channel during the packet transmission period which leads to repeated collisions if this behavior is not learned to pro-actively avoid the jammed channels. Furthermore, the CR based just on sensing without learning may move

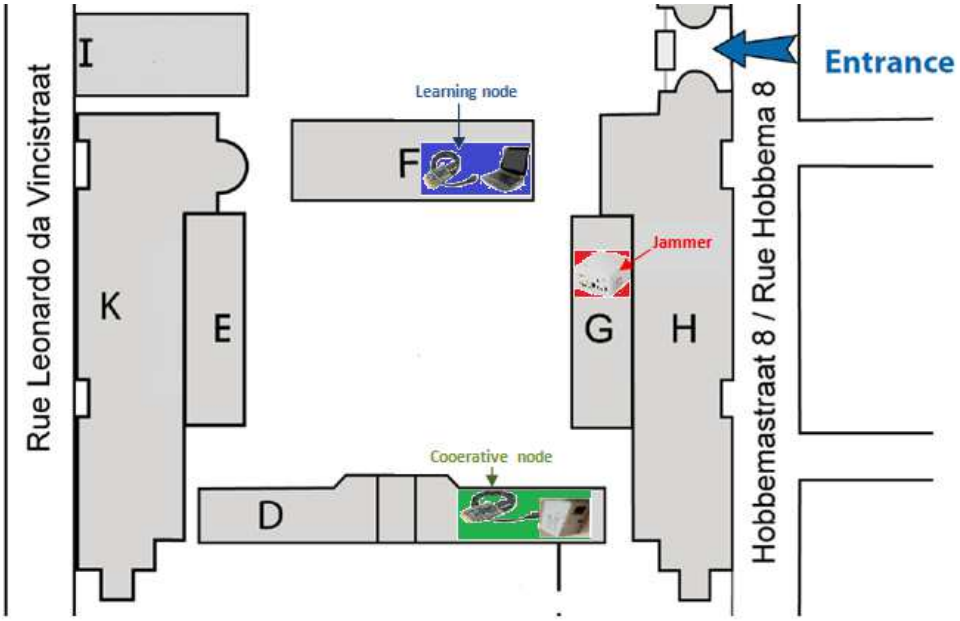


Figure 6.11: Scenario1

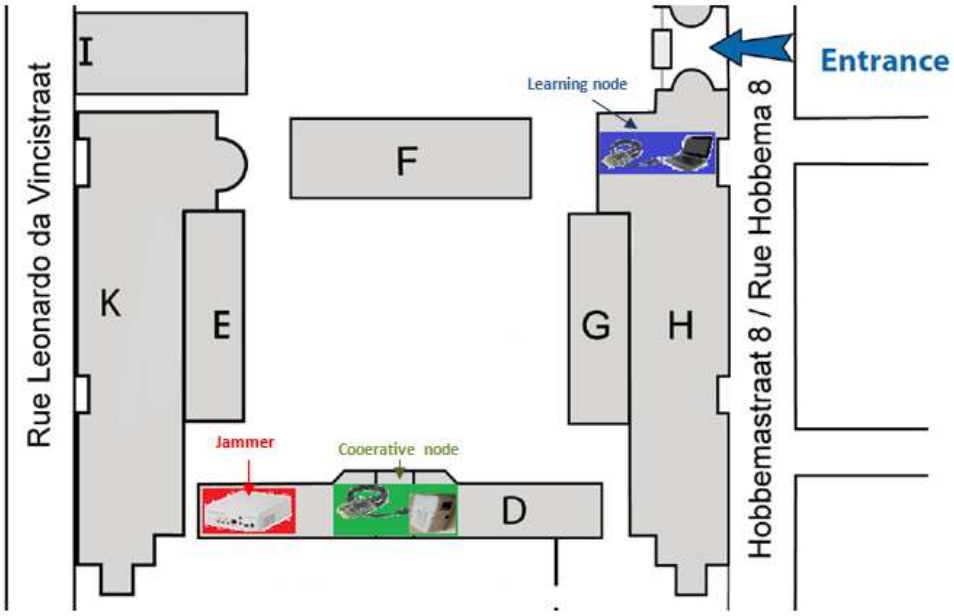


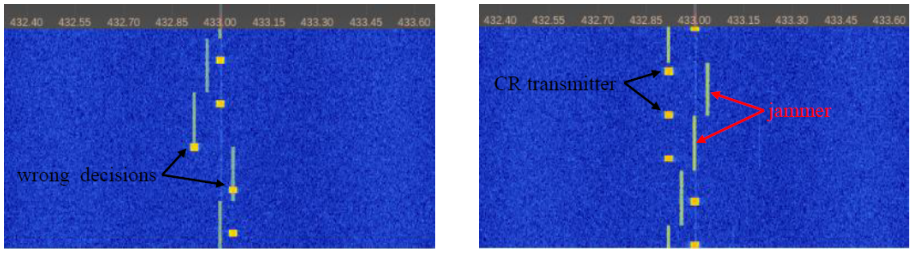
Figure 6.12: Scenario2

from channel to another without avoiding uneeded frequency alteration. However, the learning node ameliorates its behavior over time based on the goodness measures of the available decisions. The Q values are updated based not only on the

sensing results but also on the past learned information and the future expectation to take the best decision avoiding collisions.

	Jammer detectable by the learning node	**Jammer hidden from the learning node**
Classical fixed channel selection	69%	69%
Best channel selection without learning	76%	69%
Learning without cooperation	87%	69%
Learning with cooperation	94%	82%

Table 6.6: USRP implementation results: Packet Success Rate against a sweeping jammer



(a) (b)

Figure 6.13: USRP implementation of best channel selection based on sensing (a) versus channel selection based on learning (b) against a sweeping jammer

Tests using real radio equipments were also performed against pseudo random and reactive jammers. For the pseudo random jammer, we have considered the same sequence of channels used in MATLAB simulation, as given in figure 6.14. Figure 6.15 presents a first reactive jammer staying in the previous detected channel if there is no new detected transmission. Figure 6.16 describes a second reactive jammer that does not jam if there is no detected transmission. The measured PSR is given in table 6.7 against the pseudo random jammer, and in table 6.8 against the second reactive jammer. For this latter, we have considered equal receiving and jamming periods of $T_{RX} = T_{JX} = 0.512s$ which are different from the receiving and transmission periods of the CR nodes ($T_{RX} = T_{TX} = 0.409s$). Both tables confirm that the channel selection based on learning and cooperation outperforms the other strategies. Against the considered reactive jammer, the cooperation results in a PSR of 65% if the CR transmitting the packets detects the jammer and 60% if not. Furthermore, the learned channel selection without cooperation gives the same packet success rate as the best selection based only on sensing. Finally,

transmitting over the same channel gives 32% of correct packets in the presence of the described reactive jammer. This is due the asynchronous jamming behavior. The successful packets may be transmitted during the sensing period of the jammer or even during its frequency alteration process.

The measured results confirm the same conclusions as simulation results; The channel selection based on the proposed cooperative algorithm outperforms learning without cooperation which also outperforms the best channel selection without learning. However, the real USRP measurements are different from MATLAB simulation values. This is due to the implemented time division multiplexing scheme, in the USRP, that needs a processing time for the blind reception of the packets or the acknowledgments. The simulation time is different from the real time and neither the CR nodes nor the jammers need a processing time as presented in figure 6.2.

	Jammer detectable by the learning node	**Jammer hidden from the learning node**
Classical fixed channel selection	59%	59%
Best channel selection without learning	80%	59%
Learning without cooperation	89%	59%
Learning with cooperation	92%	74%

Table 6.7: USRP implementation results: Packet Success Rate against a pseudo random jammer

	Jammer detectable to the learning node	**Jammer hidden from the learning node**
Classical fixed channel selection	32%	32%
Best channel selection without learning	65%	32%
Learning without cooperation	65%	32%
Learning with cooperation	65%	60%

Table 6.8: USRP implementation results: Packet Success Rate against a reactive jammer

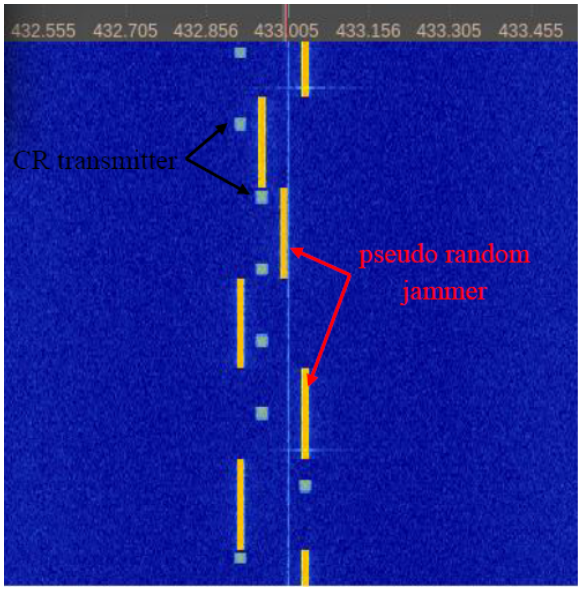


Figure 6.14: USRP implementation of channel selection based on learning against a pseudo random jammer

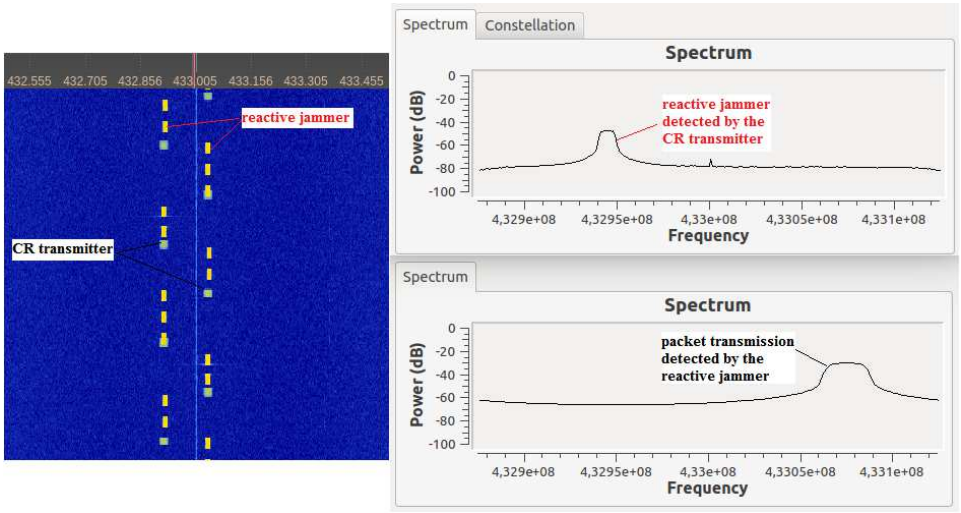


Figure 6.15: USRP implementation of channel selection based on learning against the first reactive jammer

6.5 Conclusion

In this chapter, we have proposed an enhancement of the proposed learning algorithm to go towards a realistic Q-learning algorithm solving the practical problems of synchronization requirement and hidden node problem. For that purpose, we have defined a realistic reward function based on the sensing results and we have

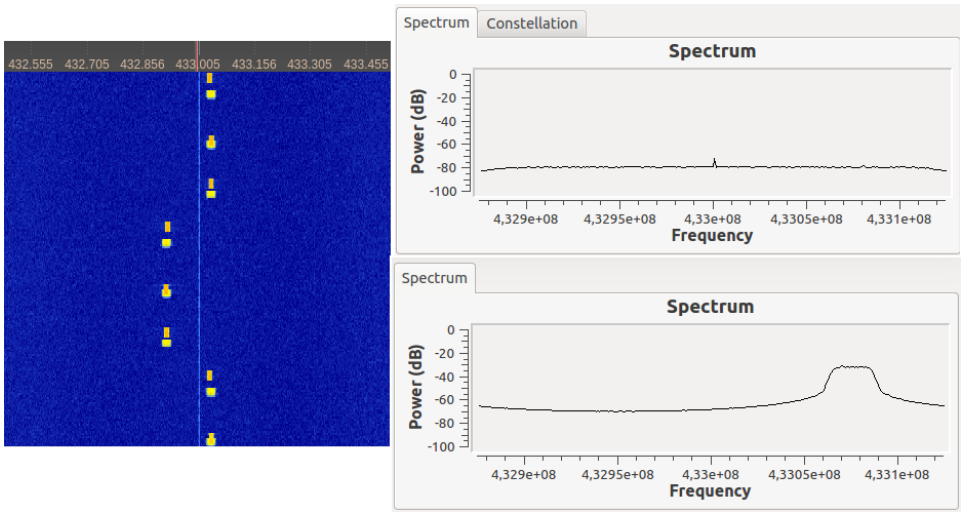


Figure 6.16: USRP implementation of channel selection based on learning against the second reactive jammer

considered the cooperation with the receiving cognitive radio. The cooperative node acknowledges each packet reception and transmits its sensing results to the CR learning node who exploits this information to update the Q values.

Simulation results and measurements using real radio equipment are given in terms of packet success rate. In MATLAB simulation, we have considered sweeping, pseudo random and reactive jammers. This latter is able to do spectrum sensing in order to detect and interfere the channel carrying the packet. For the real measurements, we have used the USRP platform and Qt Creator/C++ development environment. The results have shown that the channel selection based on the proposed learning algorithm achieves a higher packet success rate than the best channel selection based just on sensing. The results are even better when the learning CR cooperates with the CR receiving the packets to detect the jammer and update the Q values. The proposed solution is applicable in practice in real radios to avoid malicious interferes, even when hidden or not synchronized.

Chapter 7

Conclusions

From civilian side, the emergence of new wireless services, such as 4G LTE and 5G mobile networks that are predicted to meet consumer and business growing demands, is creating a spectrum shortage problem. Moreover, the current technique of static frequency allocation leads to inefficiency utilization of the available spectrum. From military side, recently tactical military missions are characterized by the coexistence of multiple heterogeneous wireless networks in the same geographical area, which leads to the problems of interferences and malicious users. Furthermore, growing military wireless services are continuously increasing the spectrum requirements and reveal the problem of bandwidth shortage. The investment of CR technology with the implementation of efficient DSM, may respond to the civilian requirements and may mitigate the military tactical problems. Cognitive radio and DSM concepts, aim to solve this imbalance between scarcity and under utilization of the spectrum by dynamically using the frequency bands. However, the CR technology introduces new vulnerabilities and opportunities for malicious users compared to traditional wireless networks due to its intrinsic characteristics. To enhance military and commercially CR investment, security challenges should be resolved which is the subject of this thesis report.

In the first chapter, we defined the cognitive radio technology that enables the implementation of dynamic spectrum management techniques to solve the imbalance between spectrum scarcity and under-utilization. We have detailed the CR main functions: spectrum sensing, decision, sharing and mobility. Based on sensing, the CR detects the available portions of the spectrum. It evaluates the sensing report taken into account the information from knowledge base to decide its alternatives to meet user communication requirements. Spectrum sharing manages the allocation of available frequency bands to provide a fair spectrum scheduling among the users and to avoid the interference. Through spectrum mobility, the CR changes its frequency of operation to vacate it for the incumbent user.

In chapter 3, we presented a comprehensive review of common CR attacks and their potential countermeasures with projection on military radio networks. We classified the attacks based on the four main functions of the cognitive radio, not

according to the layers of the OSI model as usually done. Through this classification, we tried to provide directions for related researches to discern which cognitive functionality has to be insured against each threat. The exploitation of CRs in complex scenarios such as in a military ad hoc deployment requires a high focus on the cognitive engine, since simple spectrum access algorithms may be vulnerable to malicious activities such as IE attack during SS, FC attacks for centralized spectrum decision, spectrum handoff attacks and the jamming attack. Moreover, we have compared the presented threats in terms of harmfulness and required information to accomplish each attack. We have explained how the jamming attack is easy to happen and can threaten the CR technology in all its cognition functionalities (spectrum sensing, decision, sharing and mobility), especially when the jammer is also equipped with the same technology. Radio jamming is a challenging attack in CRNs since (i) it may prevent CRs from detecting an available spectrum band during spectrum sensing by keeping the wireless spectrum busy, (ii) it may inject interference during an ongoing communication, so that the signal to SINR deteriorates heavily and no data can be received correctly and (iii) it may corrupt control packets by attacking a common control channel to disrupt the totality of the network. CRNs are characterized by DSA and by mainly distributed architectures which make it difficult to implement effective jamming countermeasures.

In chapter 4, we have exploited the CR capacities of simultaneous multi-frequency access and dynamic power allocation as the anti-jamming strategy. We have modeled the interaction between the two players, using different strategies to dynamically update their power allocations, as a zero-sum game with continuous action sets. Then, we have considered different game scenarios, for which we have determined the NE, SE and the optimal minmax/maxmin power allocations. The simulation results have given equality between the solutions of all the considered game scenarios. We have proved theoretically that this game has a unique equilibrium which is equal to the saddle point given in closed form, under the assumption that both the CR and the jammer are using all the channels (i.e. p_k, $j_k > 0$, $\forall$ $k \in [1, M]$). To solve the presented game and find the optimal power allocation strategies, we have considered complete knowledge for both the CR and the jammer. Under this assumption, each player has all relevant information with which to make a decision in each step of the game. But in real scenario the players has no information about the required parameters to calculate their optimal strategies. In chapter 5, we exploit the CR capacity of learning and reasoning to develop an anti-jamming technique under incomplete information. The proposed solution is a modified version of the Q-learning algorithm. We call the proposed algorithm as the OPSQ-learning algorithm, it is based on widebend spectrum sensing and on a greedy policy even during learning. The OPSQ-learning algorithm is applied in terms of channel selection before adapting it to the multi-channel power allocation game. We have modeled the channel selection, in the presence of fixed jamming strategy, as a MDP process. Then, we have adapted the proposed algorithm to learn the optimal anti-jamming strategy. We have ameliorated the reward function in order to stay as longer as possible in the same frequency and minimize

the number of frequency switching. We have presented the simulation results under sweeping, reactive and pseudo random jamming strategies. We can conclude that the OPSQ-learning version speeds up the learning period and the ameliorated reward strategy optimizes the number of channel switching which enhance its application during CRN real time communication. However, the presented MDP and the proposed OPSQ-learning algorithm, for channel selection, present limits in terms of synchronization requirement with the jammer and do not solve the scenario of hidden jammers. For the multi-channel application, we have adapted the modified version of the Q-learning algorithm (OPSQ-learning) to learn an anti-jamming power allocation strategy. Against fixed jamming strategies, the learned solution almost equals the common explicit waterfilling solution. Furthermore, we considered a smart jammer using the Q-learning algorithm. The learned jamming and anti-jamming power allocation strategies are almost equal to the optimal Nash equilibrium strategies found under the assumption of complete information, presented in chapter 4. Finally, we studied the real scenario when the jammer has incomplete information about the CR user and the channel gain coefficients. Under this condition, the jammer occupies all the channels with almost the same power level which results in a limited payoff gain for the CR.

In chapter 6 we have proposed an enhancement of the presented learning algorithm to go towards a realistic Q-learning algorithm, one that can be implemented in a real scenario with real hardware. To solve all practical problems like the synchronization between CR & jammer and hidden jammer problem, we have considered (1) a spatio-temporal state definition including CR and jammer channels, (2) a realistic reward function based on the sensing results and (3) a cooperative process between the learning and the receiving nodes. The cooperative node acknowledges each packet reception and transmits its sensing results to the CR learning node who exploits this information to update the Q values. To do this, we performed first a detailed simulation, with a smaller time granularity and down to the level of IQ to allow a realistic sensing. The high-fidelity simulation served as a reference for the implementation results on USRP platform. Simulation results and real radio measurements are given in terms of packet success rate. In MATLAB simulation, we have considered sweeping, pseudo random and reactive jammers. This latter is able to do spectrum sensing in order to detect and interfere the channel carrying the packet. For the real measurements, we have used the USRP platform and Qt Creator/C++ development environment. The results have shown that the channel selection based on the proposed learning algorithm achieves a higher packet success rate than the best channel selection based just on sensing. The results are even better when the learning CR cooperates with the CR receiving the packets to detect the jammer and update the Q values. The measured results confirm the same conclusions as simulation results, but the measurements are different from simulation values. This is due to the implemented time division multiplexing scheme that needs a processing time for the blind reception of the packets or the acknowledgments. The simulation time is different from the real time and neither the CR nodes nor the jammers need a processing time.

The proposed solution can be ameliorated if applied to an optimized time division multiplexing scheme that deal with the processing time limitation, which may give better results. Furthermore, OPSQ-learning can be used in practice with real radios and may be applied not only to avoid malicious interferes but also for the CR coexistence with incumbents, even when incumbents/jammers are hidden or not synchronized. As future work, we suggest trying other reward functions and implementing the OPSQ-learning algorithm at the fusion center of a centralized cognitive radio network to manage spectrum sharing in the presence of an incumbent network and more than one jammer.

Publications

F. Slimeni, V. Le Nir, B. Scheers, Z. Chtourou, and R. Attia. Optimal power allocation over parallel gaussian channels in cognitive radio and jammer games. *IET Communications*, 10(8):980–986, 2016.

F. Slimeni, B. Scheers, and Z. Chtourou. Security threats in military cognitive radio networks. In *2015 International Conference on Military Communications and Information Systems (ICMCIS)*, pages 1–10, May 2015.

F. Slimeni, B. Scheers, Z. Chtourou, and V. Le Nir. Cognitive radio jamming mitigation using Markov decision process and reinforcement learning. *Procedia Computer Science*, 73:199 – 208, 2015.

F. Slimeni, B. Scheers, Z. Chtourou, and V. Le Nir. Jamming mitigation in cognitive radio networks using a modified Q-learning algorithm. In *2015 International Conference on Military Communications and Information Systems (ICMCIS)*, pages 1–7, May 2015.

F. Slimeni, B. Scheers, Z. Chtourou, and V. Le Nir. Closed form expression of the saddle point in cognitive radio and jammer power allocation game. In *Cognitive Radio Oriented Wireless Networks - 11th International Conference, CROWNCOM 2016, Grenoble, France, May 30 - June 1, 2016, Proceedings*, pages 29–40, 2016.

F. Slimeni, B. Scheers, Z. Chtourou, V. Le Nir, and R. Attia. A modified Q learning algorithm to solve cognitive radio jamming attack. *Int. J. of Embedded Systems*, in press.

F. Slimeni, B. Scheers, V. Le Nir, Z. Chtourou, and R. Attia. Learning multi-channel power allocation against smart jammer in cognitive radio networks. In *2016 International Conference on Military Communications and Information Systems (ICMCIS)*, pages 1–7, May 2016.

References

[1] Friedrich K. Jondral. Software-defined radio-basic and evolution to cognitive radio. *EURASIP Journal on Wireless Communication and Networking*, 2005.

[2] Mitola Joseph. *Cognitive Radio:An Integrated Agent Architecture for Software Defined Radio*. PhD thesis, Royal Institute of Technology (KTH), May 2000.

[3] Yuan Zhang, Gaochao Xu, and Xiaozhong Geng. Security Threats in Cognitive Radio Networks. In *High Performance Computing and Communications, 2008. HPCC '08. 10th IEEE International Conference on*, pages 1036–1041, Sept 2008.

[4] Mahmoud Qusay. *Cognitive Networks: Towards Self-Aware Networks*. John Wiley and Sons, 2007.

[5] Qing Zhao. A survey of dynamic spectrum access: signal processing, networking, and regulatory policy. In *in IEEE Signal Processing Magazine*, pages 79–89, 2007.

[6] Scheers B. Introduction of Dynamic Spectrum Access Technology in NATO Europe Tactical Communications. In *IEEE Military Communications Conference (MILCOM'2013), San Diego, CA, USA. (Invited Paper)*, pages 737–742, November 2013.

[7] T. Charles Clancy and Nathan Goergen. Security in Cognitive Radio Networks: Threats and Mitigation. In *Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2008). 3rd International Conference, Singapore*, pages 1–8, 15-17 May 2008.

[8] Mitola J. The software radio architecture. *Communications Magazine, IEEE*, 33(5):26–38, May 1995.

[9] Raut R.D. and Kulat K.D. SDR design for cognitive radio. In *Modeling, Simulation and Applied Optimization (ICMSAO), 2011 4th International Conference on*, pages 1–8, April 2011.

[10] Ojonav Hazarika and Amit Kumar Mishra. *A Review of Hardware Platforms for Whitespace Communication*, pages 33–48. Springer International Publishing, 2015.

[11] Chen Sheng, Li Xiaochen, Cai Qiao, Hu Nansai, He Haibo, Yao Yu-Dong, and Mitola Joseph. Classification and Control of Cognitive Radios Using Hierarchical Neural Network, 2010.

[12] Qiao Cai, Sheng Chen, Xiaochen Li, Nansai Hu, Haibo He, Yu-Dong Yao, and Mitola J. An integrated incremental self-organizing map and hierarchical neural network approach for cognitive radio learning. In *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pages 1–6, July 2010.

[13] Waheed M. and Cai A. Evolutionary algorithms for radio resource management in cognitive radio network. In *Performance Computing and Communications Conference (IPCCC), 2009 IEEE 28th International*, pages 431–436, Dec 2009.

[14] Udgata S.K., Kumar K.P., and Sabat S.L. Swarm intelligence based Resource Allocation Algorithm for cognitive radio network. In *Parallel Distributed and Grid Computing (PDGC), 2010 1st International Conference on*, pages 324–329, Oct 2010.

[15] Langley Pat, Laird John E., and Rogers Seth. Cognitive Architectures: Research Issues and Challenges. *Cogn. Syst. Res.*, 10(2):141–160, June 2009.

[16] Ulversoy T. Cognitive engine for critical communications cognitive radio. In *Communications and Information Technologies (ISCIT), 2012 International Symposium on*, pages 550–555, Oct 2012.

[17] A.M. Fanan, N.G. Riley, M. Mehdawi, M. Ammar, , and M. Zolfaghari. Survey: A Comparison of Spectrum Sensing Techniques in Cognitive Radio. In *Int'l Conference Image Processing, Computers and Industrial Engineering (ICICIE'2014)*, Jan. 15-16 2014.

[18] Spyros Kyperountas, Neiyer Correal, and Qicai Shi. A Comparison of Fusion Rules for Cooperative Spectrum Sensing in Fading Channels.

[19] Huseyin A. *Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems*. Springer, 2007.

[20] Adisorn Lertsinsrubtavee, Naceur Malouch, and Serge Fdida. Spectrum Handoff Strategies for Multiple Channels Cognitive Radio Network. In *Proceedings of the ACM CoNEXT Student Workshop*, CoNEXT '10 Student Workshop, pages 20:1–20:2, New York, NY, USA, 2010. ACM.

[21] Wajdi Alhakami, Ali Mansour, and Ghazanfar A Safdar. Spectrum Sharing Security and Attacks in CRNs: a Review. *International Journal of Advanced Computer Science and Applications(IJACSA)*, 5(1):76–87, 2014.

[22] Saman T. Zargar, Martin B.H. Weiss, Carlos E. Caicedo, and James B.D. Joshi. Security in Dynamic Spectrum Access Systems: A Survey. Working paper, University of Pittsburgh, December 2009.

[23] Wen Hong, Li Shaoqian, Zhu Xiping, and Zhou Liang. A framework of the PHY-layer approach to defense against security threats in cognitive radio networks. *IEEE Network*, 2013.

[24] Hlavacek D. and Chang J. M. A layered approach to cognitive radio network security: A survey. *Computer Networks*, 2014.

[25] Fragkiadakis Alexandros G., Tragos Elias Z., and Askoxylakis Ioannis G. A Survey on Security Threats and Detection Techniques in Cognitive Radio Networks. *IEEE Communications Surveys and Tutorials*, 2013.

[26] Parvin Sazia, Hussain Farookh Khadeer, Hussain Omar Khadeer, Han Song, Tian Biming, and Chang Elizabeth. Cognitive radio network security: A survey. *J. Network and Computer Applications*, 2012.

[27] Bhattacharjee Shameek, Sengupta Shamik, and Chatterjee Mainak. Vulnerabilities in cognitive radio networks: A survey. *Computer Communications*, 2013.

[28] Alireza Attar, Helen Tang, Athanasios V. Vasilakos, F. Richard Yu, and Victor C. M. Leung. A Survey of Security Challenges in Cognitive Radio Networks: Solutions and Future Research Directions. *Proceedings of the IEEE*, pages 3172–3186, 2012.

[29] Brown Timothy X. and Sethi Amita. Potential Cognitive Radio Denial-of-service Vulnerabilities and Protection Countermeasures: A Multi-dimensional Analysis and Assessment. *Mob. Netw. Appl.*, 13(5), oct 2008.

[30] Abhilasha Singh and Anita Sharma. A Survey of Various Defense Techniques to Detect Primary User Emulation Attacks. *International Journal of Current Engineering and Technology*, April 2014.

[31] Chen Ruiliang, 0001 Jung-Min Park, and Reed Jeffrey H. Defense against Primary User Emulation Attacks in Cognitive Radio Networks. *IEEE Journal on Selected Areas in Communications*, (1):25–37, 2008.

[32] S. Anand, Z. Jin, , and K. P. Subbalakshmi. An analytical model for PUEA in cognitive radio networks. *IEEE Symposium of New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, 2008.

[33] Z. Jin, Santhanakrishnan Anand, and K. P. Subbalakshmi. Detecting Primary User Emulation Attacks in Dynamic Spectrum Access Networks. In *Proceedings of IEEE International Conference on Communications, ICC, Dresden, Germany*, pages 1–5, 14-18 June 2009.

[34] Chunsheng Xin and Min Song. Detection of PUE Attacks in Cognitive Radio Networks Based on Signal Activity Pattern. *IEEE Trans. Mob. Comput.*, 2014.

[35] Li Husheng and Han Zhu. Dogfight in Spectrum: Combating Primary User Emulation Attacks in Cognitive Radio Systems, Part I: Known Channel Statistics. *IEEE Transactions on Wireless Communications*, 2010.

[36] Chao Chen. *Investigation of Primary User Emulation Attack in Cognitive Radio Networks*. thesis, Stevens Institute of technology, 2011.

[37] Fayu Liu, Huifang Chen, Lei Xie, and Kuang Wang. Maximum-minimum Eigenvalue Detection-based Method to Mitigate the Effect of the PUEA in Cognitive Radio Networks. *Wireless Communications and Signal Processing (WCSP), 2011 International Conference*, pages 1–5, November 9-11 2011.

[38] Yuan Zhou, Niyato Dusit, Li Husheng, and Han Zhu. Defense against primary user emulation attacks using belief propagation of location information in cognitive radio networks. In *Proc. IEEE WCNC*, pages 599 – 604, March 2011.

[39] Kun Zeng, Qihang Peng, and Youxi Tang. Mitigating spectrum sensing data falsification attacks in hard-decision combining cooperative spectrum sensing. *Science China Information Sciences*, 2014.

[40] Wang Wenkai, Li Husheng, Sun Yan, and Han Zhu. CatchIt: Detect Malicious Nodes in Collaborative Spectrum Sensing. In *Proceedings of the 28th IEEE Conference on Global Telecommunications (GLOBECOM'09), Honolulu, Hawaii, USA*, 2009.

[41] Chen Ruiliang, Park Jung Min, and Bian Kaigui. Robust Distributed Spectrum Sensing in Cognitive Radio Networks. In *Proc. 27th Conf. Comput. Commun., Phoenix, AZ*, pages 1876–1884, Apr. 13-18 2008.

[42] Rawat Ankit Singh, Anand Priyank, Chen Hao, and Varshney Pramod K. Collaborative Spectrum Sensing in the Presence of Byzantine Attacks in Cognitive Radio Networks. *IEEE Transactions on Signal Processing*, 2011.

[43] Jinlong Wang, Shuo Feng, Qihui Wu, Xueqiang Zheng, Yuhua Xu, and Guoru Ding. A robust cooperative spectrum sensing scheme based on Dempster-Shafer theory and trustworthiness degree calculation in cognitive radio networks. *EURASIP J. Adv. Sig. Proc.*, 2014.

[44] Chowdhury Sayeed Hyder, Brendan Grebur, and Li Xiao. Defense against Spectrum Sensing Data Falsification Attacks in Cognitive Radio Networks. In *SecureComm*, 2011.

[45] Farhad Farmani, Mohssen Abbasi Jannat-Abad, and Reza Berangi. Detection of SSDF Attack Using SVDD Algorithm in Cognitive Radio Networks. In *CICSyN*, 2011.

[46] Li H. and Han Z. Catching Attacker(s) for collaborative spectrum sensing in cognitive radio systems: an abnormality detection approach. *Proceedings of IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*, 2010.

[47] Kaligineedi Praveen, Khabbazian Majid, and Bhargava Vijay K. Malicious User Detection in a Cognitive Radio Cooperative Sensing System. *IEEE Transactions on Wireless Communications*, 2010.

[48] Jun Du, Daoxing Guo, Bangning Zhang, and Ligang Shang. A Secure Cooperative Spectrum Sensing Scheme in Mobile Cognitive Radio Networks. *IJDSN*, 2014.

[49] Wassim El-Hajj, Haidar Safa, and Mohsen Guizani. Survey of Security Issues in Cognitive Radio Networks. In *Journal of Internet Technology*, volume 12, 2011.

[50] Xing Yiping, Mathur Chetan Nanjunda, Mohamed A. Haleem, Chandramouli Rajarathnam, and Subbalakshmi K. P. Priority Based Dynamic Spectrum Access with QoS and Interference Temperature Constraints. In *ICC*. IEEE, 2006.

[51] Leon Olga, Hernandez-Serrano Juan, and Soriano Miguel. Securing cognitive radio networks. *Int. J. Communication Systems*, 2010.

[52] Goce Jakimoski and K. P. Subbalakshmi. Towards Secure Spectrum Decision. In *ICC*, 2009.

[53] Lo Brandon F. A survey of common control channel design in cognitive radio networks. *Physical Communication*, 2011.

[54] Lazos Loukas, Liu Sisi, and Krunz Marwan. Mitigating control-channel jamming attacks in multi-channel ad hoc networks. In *WISEC*. ACM, 2009.

[55] Brandon Fang-Hsuan Lo. *Design and analysis of common control channels in cognitive radio ad hoc networks*. thesis, Stevens Institute of technology, 2013.

[56] Pietro Roberto Di and Oligeri Gabriele. Jamming mitigation in cognitive radio networks. *IEEE Network*, 27(3):10–15, 2013.

[57] Gowrilakshmi Ponuratinam, Bhumika Patel, and Syed S. Rizvi Khaled M. Elleithy. Improvement in the Spread Spectrum System in DSSS, FHSS, AND CDMA. 2013.

[58] Wang Wenjing, Bhattacharjee Shameek, Chatterjee Mainak, and Kwiat Kevin. Collaborative jamming and collaborative defense in cognitive radio networks. *Pervasive and Mobile Computing*, 9(4):572–587, 2013.

[59] Wang Beibei, Wu Yongle, Liu K. J. Ray, and Clancy T. Charles. An Anti-Jamming Stochastic Game for Cognitive Radio Networks. *IEEE Journal on Selected Areas in Communications*, 2011.

[60] Kresimir Dabcevic, Alejandro Betancourt, Lucio Marcenaro, and Carlo S. Regazzoni. A fictitious play-based game-theoretical approach to alleviating jamming attacks for cognitive radios. In *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference*, 2014.

[61] Pöpper Christina, Strasser Mario, and Čapkun Srdjan. Anti-jamming broadcast communication using uncoordinated spread spectrum techniques. *IEEE Journal on Selected Areas in Communications*, 2010.

[62] Dong Qi, Liu Donggang, and Wright Matthew. Mitigating jamming attacks in wireless broadcast systems. *Wireless Networks*, 2013.

[63] Balogun V. and Krings A. On the Impact of Jamming Attacks on Cooperative Spectrum Sensing in Cognitive Radio Networks. In *Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop (CSIIRW '13)*, 2013.

[64] Victor Balogun. Anti-jamming Performance of Hybrid FEC code in the Presence of CRN Random Jammers. *International Journal of Novel Research in Engineering and Applied Sciences (IJNREAS)*, 1(1), 2014.

[65] Asterjadhi Alfred and Zorzi Michele. JENNA: a jamming evasive network-coding neighbor-discovery algorithm for cognitive radio networks. *IEEE Wireless Communications*, 17(4):24–32, 2010.

[66] Suman Bhunia, Xing Su, Shamik Sengupta, and Felisa J. Vázquez-Abad. Stochastic Model for Cognitive Radio Networks under Jamming Attacks and Honeypot-Based Prevention. In *Distributed Computing and Networking - 15th International Conference (ICDCN '14), Coimbatore, India*, pages 438–452, January 4-7 2014.

[67] Wednel Cadeau, Xiaohua Li, and Chengyu Xiong. Markov Model Based Jamming and Anti-Jamming Performance Analysis for Cognitive Radio Networks. Communications and Network, 2014.

[68] Yongle Wu, Beibei Wang, and K. J. Ray Liu. Optimal Defense against Jamming Attacks in Cognitive Radio Networks Using the Markov Decision Process Approach. In *GLOBECOM'10*, pages 1–5, 2010.

[69] Chen Changlong, Song Min, Xin Chunsheng, and Backens Jonathan. A game-theoretical anti-jamming scheme for cognitive radio networks. *IEEE Network*, 27(3):22–27, 2013.

[70] Tianhua Chen, Jinliang Liu, Liang Xiao, and Lianfen Huang. Anti-jamming transmissions with learning in heterogenous cognitive radio networks. In *Wireless Communications and Networking Conference Workshops (WCNCW), 2015 IEEE*, pages 293–298, March 2015.

[71] Liang Xiao, Yan Li, Jinliang Liu, and Yifeng Zhao. Power control with reinforcement learning in cooperative cognitive radio networks against jamming. *The Journal of Supercomputing*, 71(9):3237–3257, 2015.

[72] Yongle Wu, Beibei Wang, K. J. Ray Liu, and T. Charles Clancy. Anti-Jamming Games in Multi-Channel Cognitive Radio Networks. *IEEE Journal on Selected Areas in Communications*, 30(1):4–15, 2012.

[73] R. El-Bardan, S. Brahma, and P.K. Varshney. Power control with jammer location uncertainty: A Game Theoretic perspective. In *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, pages 1–6, March 2014.

[74] Xiangquan Zheng, Ying Li, and Haicheng Zhang. A collision-free resident channel selection based solution for deafness problem in the cognitive radio networks. *IEEE International Conference, Wireless Information Technology and Systems (ICWITS)*, pages 1–4, 2010.

[75] Vernekar Deepraj S. *An Investigation of Security Challenges in Cognitive Radio Networks*. thesis, University of Nebraska - Lincoln, 2012.

[76] Faith M. Heikkila, Kristen Zarcadoolas, Ed Sale, and Pivot Group. Securing Telecommunications: Mission Impossible? *International Legal Technology Association (ILTA) White Papers and Surveys, Creating Omnipresence Through Telecommunications Technologies*, page 3, November 2005.

[77] Don Ross. Game Theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. 2014.

[78] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, NY, USA, 1991.

[79] Ding Xu and Qun Li. Effective capacity region and power allocation for two-way spectrum sharing cognitive radio networks. 58(6):1–10, 2015.

[80] Marios I. Poulakis, Stavroula Vassaki, Athanasios D. Panagopoulos, and Philip Constantinou. Effects of spatial correlation on QoS-driven power allocation over Nakagami-m fading channels in cognitive radio systems. *Transactions on Emerging Telecommunications Technologies*, 26(5):851–860, 2015.

[81] Fuhui Zhou, Zan Li, Jiangbo Si, Boyang Liu, and Haiyan Huang. Optimal power allocation for multiple input single output cognitive radios with antenna selection strategies. *Communications, IET*, 9(10):1337–1345, 2015.

[82] Changqing Luo, Geyong Min, F.R. Yu, Yan Zhang, L.T. Yang, and V.C.M. Leung. Joint Relay Scheduling, Channel Access, and Power Allocation for Green Cognitive Radio Communications. *Selected Areas in Communications, IEEE Journal on*, 33(5):922–932, May 2015.

[83] Saygin Baksi and Dimitrie C. Popescu. Distributed power allocation for rate maximization in cognitive radio networks with horizontal spectrum sharing. In *2015 IEEE Wireless Communications and Networking Conference, WCNC 2015, New Orleans, LA, USA, March 9-12, 2015*, pages 932–936, 2015.

[84] Ding Xu and Qun Li. Energy efficient joint chunk and power allocation for chunk-based multi-carrier cognitive radio networks. In *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, pages 943–948, March 2015.

[85] Song Xiufeng, Willett Peter, Zhou Shengli, and PB Luh. The MIMO Radar and Jammer Games. *IEEE Transactions on Signal Processing*, 60(2):687–699, 2012.

[86] Eitan Altman, Konstantin Avrachenkov, and Andrey Garnaev. A Jamming Game in Wireless Networks with Transmission Cost. In *NET-COOP*, volume 4465 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2007.

[87] Eitan Altman, Konstantin Avrachenkov, and Andrey Garnaev. Fair resource allocation in wireless networks in the presence of a jammer. In *3rd International ICST Conference on Performance Evaluation Methodologies and Tools, VALUETOOLS 2008, Athens, Greece, October 20-24, 2008*, page 33, 2008.

[88] Dejun Yang, Guoliang Xue, Jin Zhang, Andréa W. Richa, and Xi Fang. Coping with a Smart Jammer in Wireless Networks: A Stackelberg Game Approach. *IEEE Transactions on Wireless Communications*, 12(8):4038–4047, 2013.

[89] Ramy H. Gohary, Yao Huang, Zhi-Quan Luo, and Jong-Shi Pang. A generalized iterative water-filling algorithm for distributed power control in the presence of a jammer. In *ICASSP'09*, pages 2373–2376, 2009.

[90] M. Ara, H. Reboredo, S.A.M. Ghanem, and M.R.D. Rodrigues. A zero-sum power allocation game in the parallel Gaussian wiretap channel with an unfriendly jammer. In *Communication Systems (ICCS), 2012 IEEE International Conference on*, pages 60–64, Nov 2012.

[91] Zhengyu Yin, Dmytro Korzhyk, Christopher Kiekintveld, Vincent Conitzer, and Milind Tambe. Stackelberg vs. Nash in security games: interchangeability, equivalence, and uniqueness. In *AAMAS*, pages 1139–1146. IFAAMAS, 2010.

[92] K Dabcevic, A Betancourt, L Marcenaro, and C.S. Regazzoni. Intelligent cognitive radio jamming-a game-theoretical approach. *EURASIP Journal on Advances in Signal Processing*, 2014:1–18, 2014.

[93] J. B. Rosen. Existence and Uniqueness of Equilibrium Points for Concave N-Person Games. *Econometrica*, 33(3):520–534, 1965.

[94] Samson Lasaulce, Merouane Debbah, and Eitan Altman. Methodologies for analyzing equilibria in wireless games. *IEEE Signal Processing Magazine*, 26(5):41–52, 2009.

[95] Tuomas Sandholm. Perspectives on multiagent learning. *Artif. Intell.*, 171 (7):382–391, 2007.

[96] Tinne Hoff Kjeldsen and Communicated J. Gray. John von Neumanns Conception of the Minimax Theorem: A Journey Through Different Mathematical Contexts. *Arch. Hist. Exact Sci.*, (56):39–68, 2001.

[97] Tamer Basar and Geert J. Olsder. *Dynamic Noncooperative Game Theory*. Classics in Applied Mathematics. Soc for Industrial & Applied Math, 2. ed edition, 1999.

[98] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, Cambridge, MA, 1991. Translated into Chinesse by Renin University Press, Bejing: China.

[99] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[100] Robert Aumann and Hart S., editors. *Handbook of Game Theory with Economic Applications*, volume 2. Elsevier, 1 edition, 1994.

[101] Asuman Ozdaglar. Strategic Form Games and Nash Equilibrium, MIT, 2013.

[102] Tamer Basar. Lecture Notes on Non-Cooperative Game Theory, July 2010.

[103] Andrew M. Colman. *Game theory and its applications in the social and biological sciences*. International series in social psychology. Oxford: Butterworth-Heinemann & London: Routledge, 2. ed edition, 1995.

[104] Jean Céa. Optimisation : théorie et algorithmes, 1971.

[105] Csaba Szepesvári and Michael L. Littman. Generalized Markov Decision Processes: Dynamic-programming and Reinforcement-learning Algorithms. Technical report, 1996.

[106] Carlos Henrique Costa Ribeiro. A Tutorial on Reinforcement Learning Techniques.

[107] Watkins Christopher John Cornish Hellaby. *Learning from Delayed Rewards*. PhD thesis, King's College, Cambridge, UK, May 1989.

[108] Tesauro Gerald. Extending Q-Learning to General Adaptive Multi-Agent Systems. In *NIPS*. MIT Press, 2003.

[109] Sutton Richard S. and Barto Andrew G. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1998.

[110] Jinane Abounadi, Dimitri P. Bertsekas, and Vivek S. Borkar. Stochastic Approximation for Nonexpansive Maps: Application to Q-Learning Algorithms. *SIAM J. Control and Optimization*, 41(1):1–22, 2002.

[111] Eyal Even-Dar and Yishay Mansour. Learning Rates for Q-learning. *Journal of Machine Learning Research*, 5:1–25, 2003.

[112] A. Galindo-Serrano and L. Giupponi. Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks. *IEEE Transactions on Vehicular Technology*, 59(4):1823–1834, May 2010.

[113] Zhi Tian. A wavelet approach to wideband spectrum sensing for cognitive radios. In *in Proc. 1st Int. Conf. on Cognitive Radio Oriented Wireless Networks Coms. (CROWNCOM), Mykonos*, 2006.

[114] Zhi Quan, Shuguang Cui, A.H. Sayed, and H.V. Poor. Wideband Spectrum Sensing in Cognitive Radio Networks. In *Communications, 2008. ICC '08. IEEE International Conference on*, pages 901–906, May 2008.

[115] Hongjian Sun, A. Nallanathan, Cheng-Xiang Wang, and Yunfei Chen. Wideband spectrum sensing for cognitive radio networks: a survey. *Wireless Communications, IEEE*, 20(2):74–81, April 2013.

[116] A.Arokkiaraj and T.Jayasankar. OFDM Based Spectrum Sensing In Time Varying Channel. *International Refereed Journal of Engineering and Science (IRJES)*, 3(4):50–55, 2014.

[117] Andreas F. Molisch, Larry J. Greenstein, and Mansoor Shafi. Propagation Issues for Cognitive Radio. *Proceedings of the IEEE*, 97(5), 2009.

[118] L. S. Shapley. Stochastic Games. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10):1095–1100, 1953.

[119] Z. Serceki and M. Wilhoyte. Method for determining packet error rate of wireless LAN stations, 22 April 2004.

[120] V. Le Nir and B. Scheers. Evaluation of open-source software frameworks for high fidelity simulation of cognitive radio networks. In *2015 International Conference on Military Communications and Information Systems (ICMCIS)*, pages 1–6, May 2015.