Optimal Decision Making in Cognitive Radio Networks

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Senthuran Arunthavanathan

Bachelor of Electrical and Computer Systems Engineering (Hons.) – Monash University

School of Electrical and Computer Engineering

College of Science Engineering and Health

RMIT University

15/07/2016

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Senthuran Arunthavanathan

15/07/2016

RMIT UNIVERSITY

Extended Abstract

School of Electrical & Electronic Engineering

Electrical & Electronic Engineering

Doctor of Philosophy

Optimal Decision Making in Cognitive Radio Networks

by Senthuran ARUNTHAVANATHAN

Cognitive Radio Networks are being researched upon heavily in the various layers of the communication. The task of bringing software in the physical layer of communication system was a great idea by S. Haykin. He proposed the concept of a smart radio being able to learn, adapt and make intelligent decisions in an autonomous manner by use of a Software Defined Radio. This thesis reveals the content of research under Cognitive Radios in four paths.

Firstly, we look at radio sensing schemes in the Physical Layer for various radio propagation characteristics. A theoretical analysis is performed to provide a basis of comparison for the simulation results. The simulations are preformed as Monte Carlo simulations which provides an extended analysis in cases where theoretical knowledge is lacking. Theoretical Analysis was done on AWGN, Rayleigh fading and Rician fading channels, and lognormal shadowing model under existing literature. Existing literature covers receiver operating characteristics (ROC-from a receiver′s point of view) for the mentioned radio propagation characteristics and provides a closed form expression for ROCs (probability of detection and probability of false alarm) under smaller sampling sizes of the received signal. This is true under the Rician channel where theoretical expressions for ROCs are constrained to small sampling sizes. The novelty in our research is to understand the ROC behaviour for larger sampling values in Rice channels and apply it to LAP models in the LTE platform.

Secondly, we look upon the use of Reinforcement Learning and Markov models to decide upon transmitting or not based on the acquired sensing information from a Cognitive Radio Network. The Markov models are developed for the ongoing dynamic PU transmissions and the SU transmissions. These models are characterized by their arrival and death rates, and probability of detection and false alarm, respectively. Hence, this proposes a method in which the SU can make a decision through reward-based learning, without any prior knowledge initially. Then, the model is analysed in terms of transmitter operating characteristics (TOC - probability of detection and false alarm) and extended to a performance analysis in terms of interference/ collision and wastage probability with respect to system model parameters.

Thirdly, we look upon the game theoretic approach coupled with Reinforcement Learning to determine an optimal decision out of a set of decisions for the legitimate secondary user in order to minimize the effect of jamming/ collision in the selected channels. A non-cooperative game model termed as the zero sum game is applied to understand the behaviour between the two users/ players. A utility function is proposed with the idea of successful transmission in collided channels based on the two users' power constraints, distance from the receiver and number of channels accessed. The theoretical analysis is used to determine an optimal number of channels for the legitimate user to select for transmission given respective jamming strategies from a malicious user. Later, a Nash Equilibrium is proven to exist given the system model constraints under a static game play. There may exist a unique equilibrium or more than one, depending on the system model and operational parameters that are used. The work is extended to provide a means of dynamic game play by incorporating the Fictitious Play model, where the players possess past behavioural knowledge of the opponent and then makes a calculated move. It is accepted that players converge to pure (single Nash Equilibrium) or mixed (more than one Nash Equilibrium) strategies. Additionally, provided the cognitive capability of the legitimate user and a given strategy of the malicious user, the legitimate user uses Reinforcement Learning to determine the local optimum for the given jamming strategy.

Finally, we look upon the application of a complete Markov Decision Process to determine whether it is safe for the user of interest to transmit or not on the channel based on a system model performing dynamic access. In this case, the user of interest has full knowledge of the system after learning, and is able to make instantaneous decisions to transmit or not. Additionally, the work addresses the performance of the model in terms of interference/ collision and wastage with varying channel occupancy rates. The work also compares the performance of the system model with existing relevant Partially Observable Markov Decision Process and Hidden Markov models in order to understand the degree of impact on utilization. However, one must understand that the Markov Decision Process performs better but is subjected to more constraints than the other models, thus our work aims to show the need of other models and how they could be used in conjunction with our current work.

The four approaches are catered to the application of Cognitive Radio Networks that require immediate deployment and have a dynamic and robust behaviour. This is mainly applicable in the emergency communications required during natural disasters, large scale events and in mobile wireless communications. Such applications come under the "Internet of Things".

Acknowledgements

I give my sincere and deepest gratitude to my primary supervisor Associate Prof. Kandeepan Sithamparanathan who gave me the opportunity to pursue an academic career, and has been a pivotal role in the development of my knowledge and professional growth in the academic career. Next, I express my appreciation to Prof. Robin J. Evans for imparting his pearls of knowledge and wisdom with me during the course of this research. I, also thank the 2 "anonymous" reviewers for their insights and interest in this work.

I thank my colleagues and friends [Sathyanarayanan Chandrasekharan, Akram Al-Hourani, Kagiso Magowe] from RMIT University, [Dr. Leonardo Goratti, Dr. Francesco de Pellegrini, and Dr. Lorenzo Maggi] Create-Net, and [Prof. Sam Reisenfield] Macquarie University who provided me with a plethora of knwoledge, insight and expertise that greatly assisted the path of this research, although they may not agree with all of the interpretations/conclusions of this thesis. A special thanks goes to Dr. Tinku Rasheed for providing me with the opportunity to work at Create-Net, Italy which was a very momentous experience during my PhD candidature and life.

I give my deepest appreciation to my family [Mr. and Mrs. Arunthavanathan, Aarthika Arunthavanathan, Sheiyohn Arunthavanathan, Nidhya Gopinath,...] and friends for supporting and motivating me throughout this venture. I give my sincere appreciation to my friend [Chen Yang], a fellow PhD from Monash University for her support and constant affirmation of the completion of this work on due time. I also give my thanks to my fellow PhD friends [Ivan Chew, Tridib Saha, Dilukshan Karunatilake] from Monash University Malaysia for their support.

This research has received partial funding and support from the projects: EC 7 Framework Programme (FP7-2011-8) under the Grant Agreement FP7-ICT-318632 and under Australian Research Council's Discovery Projects funding scheme (Cognitive Radars for Automobiles, DP150104473), and organizations: RMIT University, NICTA Victoria, Australia, and Create-Net, Italy.

Contents

List of Figures

List of Tables

List of Abbreviations

CR	Cognitive Radio
CRN	Cognitive Radio Network
PHY	Physical Layer
PU	Primary User
SU	Secondary User
FCC	Federal State Communications Commission
SDR	Software Defined Radio
RF	Radio Frequency
DSA	Dynamic Spectrum Access
CSI	Channel State Information
CSMA	Carrier Sense Multiple Access
LTE	Long Term Evolution
LAP	Low Altitude Platform
FDD	Frequency Division Duplexing
TDD	Time Division Duplexing
TDD	Sounding Reference Signals
AWGN	Additive White Gaussian Noise
ROC	Receiver Operating Characteristics
TOC	Transmitter Operating Characteristics
C-ROC	Complementary Receiver Operating Characteristics
I-UE	Incumbent User Equipment
I-LTE	Incumbent Long Term Evolution
S-LTE	Secondary Long Term Evolution
SC-FDMA	Single carrier Frequency Division Multiple Access
RL	Reinforcement Learning

MM	**Markov Model**
HMM	**Hidden Markov Model**
POMDP	**Partially Observable Markov Decision Process**
MDP	**Markov Decision Process**
NE	**Nash Equilibrium**
FP	**Fictitious Play**
PRB	**Physical Resource Block**
LU	**Legitimate User**
MU	**Malicious User**
D2D	**Device-to-Device**

Chapter 1

Introduction

Over the decades, wireless communications have strongly moved forward in terms of infrastructure, framework, and policies. The first digital mobile wireless technological standard brought upon to the world was termed as Second Generation (2G), which denotes the beginning of wireless cellular mobile communication, that provided data transmission capacity of 50 Kbits/s to 1 Mbits/s. It was intended mainly for large volumes of voice services and provided slow data transmissions. The major advantage was that digital calls were free of static and background noise, albeit its poor signal strength on higher frequencies for larger distances.

The next technological standard developed was termed as Third Generation (3G), which brought upon greater network capacity and security, faster data transmission that allowed for data applications, though restrictive to applications that required high bandwidth. Though debatable, the main widespread advantage of 3G over 2G communications is its greater network coverage capacity, and faster data transmission rates.

After 3G, there was a growing realization that the next wireless communication technological standard required significant focus on infrastructure that supported data applications which require large amount of data bandwidth.

Later, as 3G standards were not sufficient enough due to the increasing population of mobile wireless devices and the exponential rising demand for high bandwidth multimedia applications. This brought upon the revolutionary Fourth Generation (4G) standard. The 4G technological standard revolutionized mobile communications by bringing focus on greater data transmissions, thus allowing mobile web access, IP telephony, gaming services, video conferencing, and high-definition mobile TV services.

Such services are possible in 4G as the infrastructure is based on an all Internet Protocol (IP) packet switched network. It is able to provide peak data rates that vary from 100 Mbits/s to 1 Gbits/s depending on the mobility of the wireless devices, and are able to dynamically share and use the network resources, hence support more simultaneous users, thus increasing the overall network coverage density.

Long Term Evolution (LTE, also termed as 4G LTE) technology was brought forward, which then further developed to LTE-Advanced (LTE-A). LTE enables high-speed wireless business-communication for mobile phones and data terminals. There are two major types of LTE: LTE-TDD, and LTE-FDD. They are dependant on how data is uploaded and downloaded, and what frequency spectrum the networks are deployed in.

LTE-A provides significant improvement to the LTE radio technology and architecture, higher performance requirements for working with legacy radio technologies, greater backward compatibility of LTE-Advanced with LTE, and consideration of decisions on access of frequency bands in order to ensure that LTE-A accommodates the geographically available spectrum for channels ranging above 20 MHz.

The next step in wireless communication infrastructure is the proposal, design, and development of Fifth Generation (5G) standard that comes under the Internet of Things (IoT). The requirements of 5G are: data rates up to 10 Gbits/s, lower latency times compared to LTE (1ms), signalling efficiency, spectral efficiency and network coverage improvement, and greater network coverage density which includes the massive number of wireless sensor network devices. D2D is an exciting innovative feature that is to be introduced in 5G, that allows nearby wireless mobile devices to communicate without the need of a base station. It should facilitate the inter-operability between critical public safety networks and ubiquitous commercial networks based on the LTE-A technology.

1.1 Cognitive Radio Fundamentals

The use of electromagnetic spectrum is licensed and regulated by governments through the form of transmitters and receivers catering to specific parts of the spectrum. The Federal State Communications Commission (FCC) published a report prepared by the

Spectrum-Policy Task Force that aimed at the improvement of the method in which this precious resource is managed in the United States. This report led to one conclusion [29]:
“In many bands, spectrum access is a more significant problem than the physical scarcity of spectrum, in large part due to legacy command-and-control regulation that limits the ability of potential spectrum users to obtain such access.”
The underlying problem is the exponentially growing trend in the need for higher data rates which is the result from the transition from voice-only communications to multimedia-type communications. The technological changes that have occurred as the 3G framework moved towards 4G and as of current to 4G LTE, which clearly depicts the issue of the stated transition.

The demand for multimedia-type communications is ever increasing due to two main reasons:

- The mobile population is dramatically increasing, especially in the developing countries.
- The required data for multimedia-type communications is increasing as well.

In order to provide such functionality, and hence improve overall spectrum utilization and meet the ongoing demand for higher data rates; a revolutionary concept called Cognitive Radio (CR) was put forward by J. Mitola [34].
"Cognitive Radio is an intelligent wireless communication system that is aware of its surrounding environment (i.e., outside world), and uses the methodology of understanding by building to learn from the environment and adapt its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g., transmit-power, carrier-frequency, and modulation strategy) in real-time with two primary objectives in mind: (1) Highly reliable communications whenever and wherever needed; (2) Efficient utilization of the radio spectrum. " Furthermore, an additional and essential feature such as reconfigurability is endowed into the CR, and this is provided by a platform known as Software-defined Radio (SDR). As of current, the SDR is a practical reality due to the convergence of two key underlying technologies: digital radio and computer software. The IEEE802.22

standard for Wireless Regional Area Network (WRAN) is the first effort to make commercial applications based on CR technology feasible.

1.1.1 Spectrum Utilization

To account for a simultaneous rapid rise, the concept of efficient spectrum utilization was considered. In general, parts of the spectrum are heavily used by the terrestrial communication network termed as the primary user, while some remain under-utilised or not used at all. The orthodox definition of spectrum opportunity is "a band of frequencies that are not being taken access by the leased user of that band at a particular time in a particular geographic area", and only exploits five dimensions of the spectrum space: frequency, time, space, code and angle. This is the spectral opportunity in the frequency and time domain as illustrated in Figure 1.1. Considering solely the frequency domain, the available frequencies are divided into narrower chunks which mean that not all bands are used simultaneously. This allows some bands to be utilized and some not to be used at a given time.

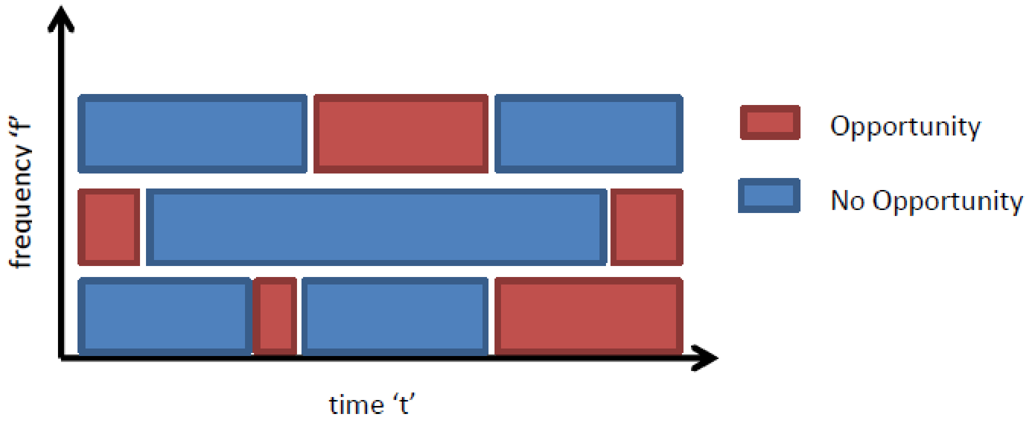


FIGURE 1.1: Opportunistic Access.

Such opportunities in the electromagnetic spectrum are termed as spectrum holes. "A spectrum hole is a band of frequencies assigned to a primary user, but at a particular time and specific geographical location, the band is not being utilized by that user."

They fall into two categories:

- White bands – Frequencies that are not used at all most of the time.

- Grey bands – Frequencies that are under-utilized most of the time.

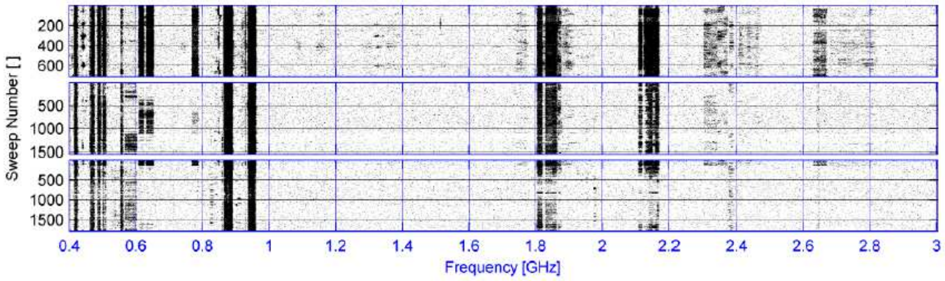


FIGURE 1.2: Spectrum Utilization[3].

These bands could be employed by a SU who can access the spectrum holes and thus, significantly improve the overall spectrum utilization as illustrated in Figure 1.2.

1.1.2 Primary & Secondary Users

The licensed users who have their spectrum assigned by the FCC [21] are called as Primary Users (PUs). This results in such users having higher priority or legacy rights over a part of the spectrum within a certain geographical area. Existing terrestrial communication systems fall under this category.

On the other hand, there are other users who have a lower priority and no legacy rights over a part of the spectrum at that geographical area, hence do not have the license to operate at the specified spectrum within the geographical area. Such users are termed as Secondary Users (SUs). The SUs can exploit the unused spectrum or share the spectrum with the PU, while ensuring there is minimal interference between the PU and SU transmissions. Hence, the SUs should be capable of CR functionality such as spectrum sensing in order to be aware of the spectrum usage and PUs' existence before commencing access.

1.1.3 Dynamic Spectrum Access

Dynamic Spectrum Access (DSA) is a spectrum sharing paradigm that allows SUs to access the spectrum holes in the licensed spectrum bands. There are two types of DSA that CRs are involved in the operational level of spectral access, which is illustrated in Figure 1.3. They are:

- Underlay - Simultaneous PU and SU transmissions may occur only if the interference generated by the SU Tx at the PU Rx is below an acceptable threshold.
- Overlay - SU transmits on PU channels in an opportunistic manner whenever PUs are not transmitting.

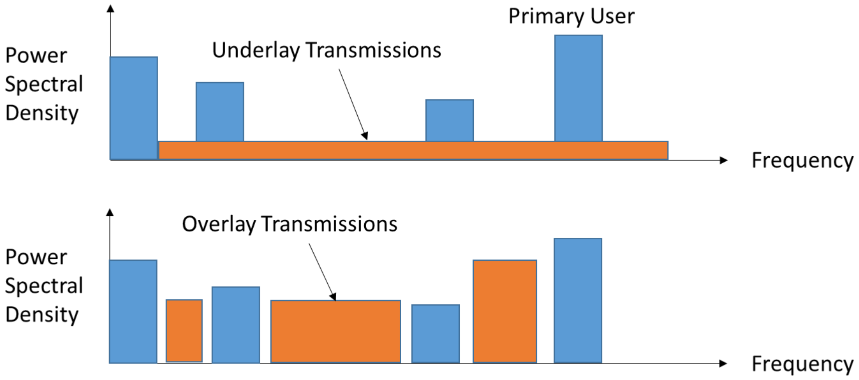


FIGURE 1.3: Types of DSA.

Furthermore, as one looks upon Cognitive Radio Networks (CRNs), issues such as resilience against jamming attacks from other malicious users are taken into consideration. A game theoretic and unsupervised learning approach are considered for the user to make optimal decisions to form a mode of best response against such events.

1.1.4 Cognitive Cycle (An Overall Background)

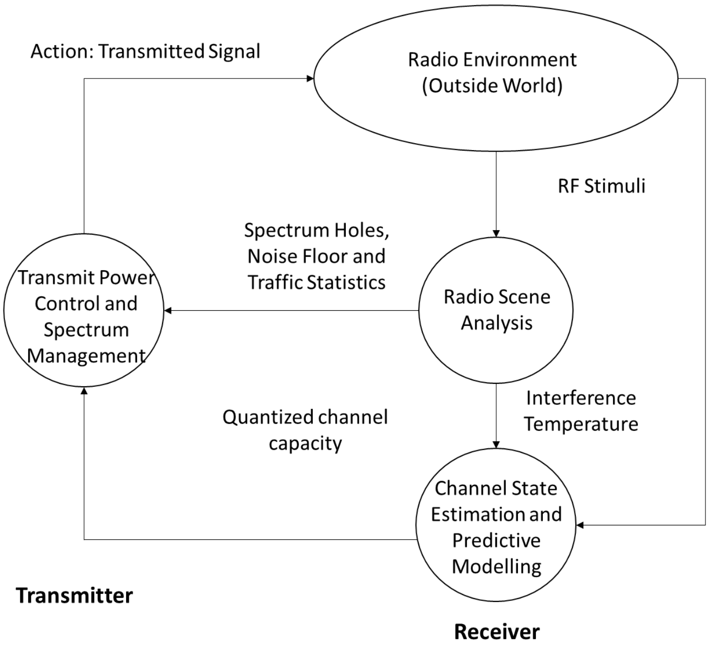

Action: Transmitted Signal
Radio Environment (Outside World)
RF Stimuli
Spectrum Holes, Noise Floor and Traffic Statistics
Transmit Power Control and Spectrum Management
Radio Scene Analysis
Interference Temperature
Quantized channel capacity
Channel State Estimation and Predictive Modelling
Transmitter
Receiver

FIGURE 1.4: Basic Cognitive Cycle by S. Haykin [29].

Besides the inclusion of a SDR for reconfigurability purposes, there are other tasks of a cognitive nature that takes into account signal-processing and machine-learning procedures for implementation. The cognitive process begins with the passive sensing of RF stimuli and culminates with action. The three cognitive tasks are:

- **Radio Scene Analysis**, which takes into account the estimation of interference of the radio environment and detection of spectrum holes. This determines the spectral usage and provides awareness of PU transmissions,
- **Channel-state Identification** that encompasses the estimation of channel-state information (CSI) and prediction of channel capacity for use by the transmitter,
- **Transmit-power control and dynamic spectrum management**, which determines the power allocation per channel for transmission purposes.

It is observed from Figure 1.4 Basic Cognitive Cycle by S. Haykin that tasks (1) and (2) are carried out at the transmitting end and task (3) is executed at the receiving end. Therefore, it is apparent that the cognitive module in the transmitter must work harmoniously with the ones in the receiver at all times. This is done by means of a feedback channel connecting the transmitter and the receiver. Hence, the CR is a fine example of a feedback communication system.

1.1.5 Sniffing & Learning

One of the major functions of the CR in SUs is the detection of spectral bands that are being utilized and non-utilized over time. The CR should be able to learn which spectral bands are being used and not used over time, determine the level of interference on each spectral band over time, and determine the frequency of PU transmissions or channel usage. The CR performs sensing to determine the characteristics of the spectral band/ channel and hence acquire information of the channel state behaviour. This behaviour is learnt through a cognitive process.

1.1.6 Decision Making

Decision making is vital for SUs in terms of determining the transmission time, number of channels and channel locations to transmit, and when to transmit in different wireless environments. This is the next step to be taken after learning in order for the SUs to perform a suitable action with minimal interference to the wireless environment. Decision making is important as CRs are defined to be autonomous and hence, perform actions without any supervision. Such actions should help them learn the outcome by means of a feedback loop.

1.2 Research Objectives & Contribution

Although CRs are able to accomplish the task of spectrum sensing, there are many challenges to be faced in this rising technology. There is a significant amount of research that is done in this area. The need for CR technology to be developed is mainly because of its vast applications in the real world such as in many forms of transportation and communication. CRs had been under extensive research for some time, even though it is still in the conceptual stage. This is because the required hardware technology and digital applications were not available in the past and not sophisticated in comparison to the current ones.

1.2.1 Research Objectives

After reviewing the current literature in the area of CR communications based on sensing, learning and decision making, we observed a gap in the literature. The research objectives in this thesis addressed are applicable for the scenarios: Aerial-to-Terrestrial Communications, Vehicular Communication. The research objectives are the following:

- **RO1-Sensing and Learning in robust environments** - CRs should sense other users in dynamic radio frequency (RF) environments that require devices to be immediately deployed. The environment should take into consideration of various radio propagation characteristics in an aerial-terrestrial model, in particular where performance is not quantifiable by theoretical studies. Furthermore, CRs should learn the RF environment which comprises of spectral band usage, and be aware of PUs' existence. The CRs should learn the temporal behaviour of the PU transmissions and determine the appropriate action to perform over time.

- **RO2-Resilience in CRNs against jamming attacks from other SUs** - CRs should have a means of resilience in the PHY layer against other unauthorized or malicious SUs that seek to destroy the transmission channels of the authorized SUs. One should take into account that all SUs are capable of CR functionality. Resilience should allow authorized CRs to have a means of successful communication despite the jamming attacks.

- **RO3-Optimal decision making for improving spectral efficiency** CRs should learn the dynamic RF environment and make decisions based upon predictions to minimize interference between other users including both PUs and SUs. Wastage of the channel should also not be tolerated, hence a decision is made based on the trade-off between interference and wastage of the channel.

1.2.2 Contributions

Each contribution addresses the research questions put forward respectively.

- **RO1**: Spectrum Sensing using Energy Detection in various radio propagation channels was studied. The model is representative to the ABSOLUTE scenario

(www.absolute-project.eu) and the concept is for the aerial eNodeB of the SU network to detect the incumbent users during the roll-in phase and roll-out phase. This is where the SU network is deployed till the primary terrestrial infrastructure recovers. The outcome of this research are published in the following papers:
a. Arunthavanathan. S, Kandeepan .S, and Evans, R.J., "Spectrum Sensing and Detection of Incumbent-UEs in Secondary-LTE based Aerial-Terrestrial Networks for Disaster Recovery", IEEE CAMAD, pp. 201-206, 2013.
b. Book Chapter: Temporary Cognitive Small Cell Network for Rapid and Emergency Deployments, Cambridge Press, 2015.

- **RO1**: A Reinforcement Learning approach was considered using the concept of Markov models and Markov Decision Processes to present analytical methods of the performance of the Secondary network determining whether it is safe to transmit or not. The characteristics such as interference between existing users and wastage of bands that are not occupied are taken as the performance metric. The outcome of this research are published in the following papers:
a. Arunthavanathan. S, Kandeepan.S, and Evans, R.J., "Reinforcement Learning based Secondary User Transmissions in Cognitive Radio Networks" IEEE Globecom Workshops (GC Wkshps), pp. 374-379, 2013.

- **RO2**: Direct communication between nearby mobile radios is an intriguing and innovative feature of next-generation cellular networks. A real threat is an unauthorized hand-held device shall be deployed to intentionally jam the ongoing transmissions of other legitimate users. A non-cooperative game approach is taken to provide the authorized users a mode of resilience by making optimal decisions. The outcome of this research are published in the following papers:
a. Arunthavanathan. S, Goratti. L, Maggi. L, De Pellegrini. F, Kandeepan. S,"On the Achievable Rate in a D2D Cognitive Secondary Network Under Jamming Attacks", IEEE Crowncomm, pp. 39-44, 2014.
b. Arunthavanathan. S, Goratti. L, Maggi. L, De Pellegrini. F, Reisenfield. S, and Kandeepan. S, "An Optimal Transmission Strategy Under Jamming Attacks with Transmit Power Constraints", IEEE TCCN 2016 (submitted).

- **RO3**: The ARC project deals with Cognitive communications via Radar for use in automobiles that allow efficient and cognitive communication between them. A Spectral Access with Markov Decision Process was devised in order to determine an approach for each automobile user in a robust environment with pre acquired or determined knowledge. The outcome of this research are to be published in the following papers: a. Arunthavanathan. S, Kandeepan.S, and Evans, R.J., "Spectral Access with Markov Decision Process", TWL (Accepted at IEEE TWL).

1.3 List of Publications

1.3.1 Book Chapters

- Akram Al-Hourani, S. Kandeepan, and **S. Arunthavanathan**, "Temporary Infrastructure with Cognitive Radios for Small Cell Networks," in the book "Design and Deployment of Small Cell Networks", by Cambridge University Press, 2015, ISBN-13: 9781107056718.

1.3.2 Journals

- **S. Arunthavanathan**, L. Goratti, L. Maggi, F. de. Pellegrini, S. Reisenfield, and S. Kandeepan, "An Optimal Transmission Strategy Under Jamming Attacks with Transmit Power Constraints", (Submitted to IEEE Transactions on Cognitive Communications and Networking).

- **S. Arunthavanathan**, S. Kandeepan, and R.J. Evans, "A Markov Decision Process based Opportunistic Spectral Access", (Accepted in IEEE Wireless Communication Letters).

1.3.3 Conference Proceedings

- **S. Arunthavanathan**, S. Kandeepan, and R.J. Evans, "Spectrum sensing and detection of incumbent-UEs in secondary-LTE based aerial-terrestrial networks for disaster recovery," IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), pp. 201-206, Sept. 2013, DOI: 10.1109/CAMAD.2013.6708117.

- **S. Arunthavanathan**, S. Kandeepan, and R.J. Evans, "Reinforcement learning based secondary user transmissions in cognitive radio networks," IEEE Globecom Workshops, pp. 374-379,Dec. 2013, DOI: 10.1109/GLOCOMW.2013.6825016.
- **S. Arunthavanathan**, S. Kandeepan, and R.J. Evans, "On the achievable rate in a D2D cognitive secondary network under jamming attacks," International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), pp. 39-44, June 2014, DOI:10.4108/icst.crowncom.2014.255665.

1.3.4 Project Deliverables

- **S. Arunthavanathan** et. al,"FP7-ICT-2011-8-318632-ABSOLUTE/D4.5.1 Security Solutions for Public Safety: First Issue", FP-7 Project ABSOLUTE-Aerial Base Stations with Opportunistic Links for Unexpected and Temporary Events.
- **S. Arunthavanathan** et. al,"FP7-ICT-2011-8-318632-ABSOLUTE/D4.5.2 Security Solutions for Public Safety: Final Issue", FP-7 Project ABSOLUTE-Aerial Base Stations with Opportunistic Links for Unexpected and Temporary Events.
- Hanwen Cao, et.al, .., **S. Arunthavanathan**"FP7-ICT-2011-8-318632-ABSOLUTE/D3.1.1 Requirements and Specification for Spectrum Awareness", FP-7 Project ABSOLUTE-Aerial Base Stations with Opportunistic Links for Unexpected and Temporary Events.

1.4 Projects

There are many possible uses for the CR in the current market. The major needs are in public safety, provision of communications in rural areas, military purposes, cellular networks (Femto-cell) and vehicular communications. In my research, we shall focus on the following two projects that were conducted.

1.4.1 EU FP7 - ABSOLUTE: Public Safety in disaster scenarios and Large-scale events

After a large scale disaster, the normal terrestrial network infrastructure is compromised and cannot guarantee reliable and large scale coverage for rescue teams and citizens. The ABSOLUTE project architecture (www.absolute-project.eu) performs its functionality until the primary infrastructure of the particular geographical area is able to handle the traffic load or returns back from shutdown in times of major temporal events and disaster scenarios, respectively. Current critical communication systems cannot provide dependable and resilient network connectivity at higher data rates over large coverage areas. There exists an urgent requirement for a rapidly deployable multi-purpose, multi service and multi-band interoperable and integrated network infrastructure capable of supporting reliable high data rate applications to serve large scale disaster emergency situations and the temporary event scenarios, hence CRNs are considered as they are capable of DSA and the aforementioned factors.

1.4.2 ARC Discovery Project Cognitive Radar for Automobiles

Both vehicular radar and wireless devices require access to spectrum and hence optimal access strategies need to be considered. The devices perform in a distributed and time varying approach. Therefore, interference is a highly significant issue of concern as the number of wireless devices increase with increasing number of vehicles. CRNs are capable of autonomous intelligent decision making in order to minimize the interference from other vehicles. Vehicles are mobile, thus geographical areas may vary over time, hence spectral resources may vary. This leads to the context of DSA, that is performed by CRNs.

1.5 Thesis Structure

The thesis is divided into 7 chapters: chapter 1 provides an introduction to the Cognitive Radio process, opportunities, and possible applications (projects); followed with the novel contributions to the addressed research objectives, chapter 2 provides a brief theoretical introduction on the mathematical tools that are of relevant importance in the work of this thesis, including a summary of the literature review, chapter 3 discusses the novelty of this work addressed in the research objective 1, chapter 4 discusses the novelty of this work addressed in the research objective 2, chapter 5 discusses the novelty of this work addressed in the research objectives 2 and 3, chapter 6 discusses the novelty of this work addressed in the research objective 2, and chapter 7 provides a brief conclusion on the outcome of the work done in the prior chapters; hence, concluding the overall novelty and potential of this thesis.

Chapter 2

Theoretical Background and Literature Review

We shall look upon some of the relevant mathematical tools used for research purposes in this thesis, briefly. The mathematical tools described are: Spectrum sensing which focuses on Energy Based Sensing, Markov Models and Markov Decision Processes followed with Reinforcement Learning, and Game theoretic models focusing on non-cooperative games like zero-sum games. Such mathematical tools are applied in this thesis to address the prior mentioned research objectives regarding the gaps of research in CRNs.

2.1 Fundamentals for Spectrum Sensing in Cognitive Radio Networks

Before getting into the context of the various sensing techniques,there are many challenges that must be faced to bring CR to be a feasible concept: hardware requirements for complex processing, hidden PU problem, detecting spread spectrum PUs, operating constraints such as sensing duration and frequency, decision making in cooperative sensing, and resilience and security. As hardware requirements cater to the individual needs and design of the system, it shall not be looked upon in detail. One shall not dwell into further detail at cooperative sensing as it has not been looked upon. The subject of resilience and security shall be looked upon later in the upcoming chapters.

2.1.1 Hidden PU Problem

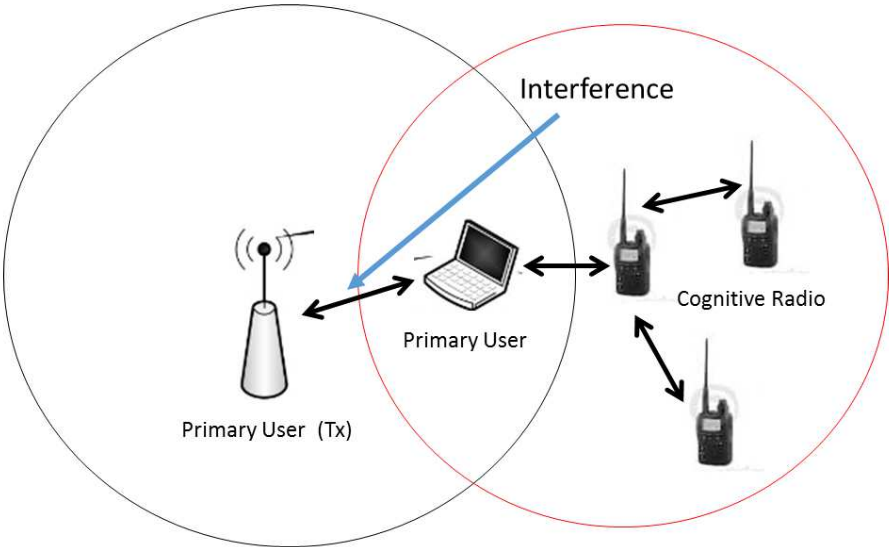


FIGURE 2.1: Hidden PU Problem.

The hidden PU problem characterizes similarly to the hidden node problem in Carrier Sense Multiple Access (CSMA). It is caused by various factors including multi-path fading, shadowing observed by SUs while scanning for PU transmissions. Figure 2.1 shows an illustration of a hidden node problem where the black and red circles show the operating domain of the PU and the CR device. Here, CR device causes unwanted interference to the PU receiver as the PU transmitter's signal is not detected, because of the locations of devices. Cooperative sensing is an ideal proposition as pointed out in the literature [27][22].

2.1.2 Types of Sensing

A CR should be aware of its Radio Frequency (RF) environment, which means that it should sense the surrounding environment and localize [63, 6] the related RF activities. This is done by means of Spectrum Sensing. Different spectrum sensing methods have been proposed in literature such as the cyclostationary based sensing, matched-filter based sensing, waveform based sensing, cooperative sensing, distributed sensing and energy based sensing [25][74], which are summarized in Figure 2.2.

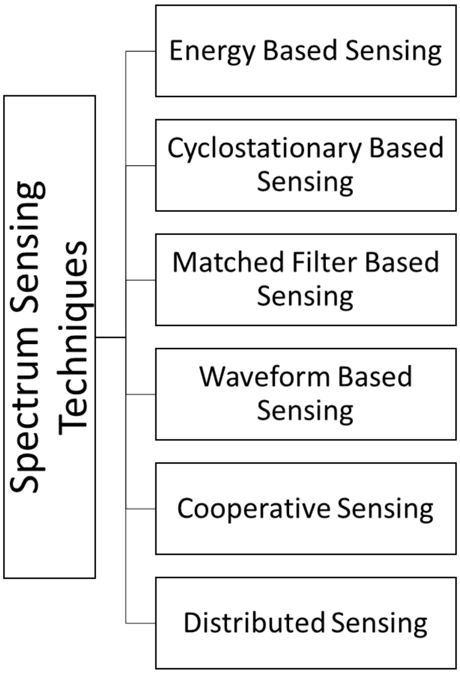

Spectrum Sensing Techniques
Energy Based Sensing
Cyclostationary Based Sensing
Matched Filter Based Sensing
Waveform Based Sensing
Cooperative Sensing
Distributed Sensing

FIGURE 2.2: Various Sensing Methods.

The above stated sensing algorithms are summarized accordingly [89] for additional information. Our work focuses solely on energy based sensing as it is a less complex structure which is widely used and accepted in literature. Hence, we look upon energy based sensing in our literature review.

2.1.3 Energy Based Sensing

Energy detector based scheme, also termed as radiometry or periodogram, is the most familiar way of spectrum sensing due to its low computational and implementation complexities. It is heavily used as it does not require any knowledge on the PU's signal. The signal is detected by comparing the output of the energy detector with a threshold that depends on the noise floor. The major challenge is the selection of the threshold for detecting PUs, as it depends accordingly to the system used and its operating constraints. Energy based sensing is looked upon in greater detail in the later sections, as it is a major part of the contribution to this chapter.

2.1.4 Summary of Literature Review

A CR should be aware of its Radio Frequency (RF) environment, which means that it should sense the surrounding environment and localize the related RF activities. This is done by means of Spectrum Sensing. Spectrum sensing has been researched heavily, and shall have a wide context to look up. There are many published works in the various aspects of sensing. Energy Based Sensing involves the acquirement of the spectral content or RF energy over the spectrum. Our novelty lies in the comparison of Rice, Rayleigh and AWGN for an immediate deployment scenario. Current literature provides a detailed view of the various sensing techniques involved and the implications behind them. A survey of spectrum sensing techniques was summarized in the following paper [89]. It describes the challenges and opportunities that sensing techniques face in CRNs and give an overall view of the Spectrum Sensing literature. Research on unknown deterministic channels and known fading channels was worked upon in the area of energy based sensing [25][77]. Fading channels such as Rayleigh and Rice were discussed and the implication on each channel arising from energy based sensing was explained. The paper [59] proposes the energy detection on channels subjected to not only multi-path fading but also shadowing. It describes the performance of the energy detector under such constraints. The paper [90] studies the spectrum characterization for opportunistic access as a whole. Additionally, the paper [9] provides experimental results on the cyclostationary properties of the IEEE802.11n Wi-Fi transmissions, and the usage of the cyclostationary features to detect IEEE802.11n radios in the context of ultra wide band (UWB) based CR. The IEEE802.11n lies in the UWB frequency range of 5.2GHz, and the CR needs to successfully detect the legacy user to avoid interference. In addition, the paper [10] considers the spectrum sensing performance and requirements for detecting legacy users in CR with periodic scanning. The performance and requirements are studied based on the temporal spectral occupancy statistics of the legacy user and the sensing signal to noise ratio levels in order to achieve a certain level of detection probability. It is modelled as the temporal statistics of the user in a Poisson Pareto burst process (PPBP) describing a typical WEB service application and the noise as an additive white Gaussian noise (AWGN) process. The paper [72] extends the simple cooperative spectrum sensing communication model to admit transmission imperfections and considers the case where the local hard CR decisions that are based on any local detection scheme are corrupted by additive noise during transmission from CR to base station. Significant research has been conducted in the areas of spectrum sensing as seen in the above short summary of the literature done. Spectrum sensing techniques still require further improvement especially in the concept of dynamic and robust communications environment, as the primary objective of the CRN is to ensure successful transmission, with minimal interference to the Primary system.

2.2 Fundamentals in Markov Decision Processes

Probability theory states that a Markov Model is defined as a stochastic model that is used to model randomly changing systems where the assumption of future states depend only on the current state and not on the events that occurred before it is held valid. There are different types of MM: Markov Chains and Hidden Markov Models (HMMs). HMMs are models where the states are only partially observable. Observations are related to the state of the system, but they are typically insufficient to accurately determine the state. This means that the user enters the current state which is not completely tangible with the expected state. An extension to such HMMs is the inclusion of actions performed at each state that leads to the next possible state with a reward. Such cases are known as Markov Decision Processes (MDPs). However, the extension of an HMM leads to a Partially Observable Markov Decision Process (POMDP) due to the states consisting of partial information. At each time interval, the agent gets to make some observations that depend on the state. The agent only has access to the history of observations and previous actions when making a decision. It cannot directly observe the current state, hence unable to acquire complete information regarding the current state.

2.2.1 Markov Decision Processes

If an agent must reason about an ongoing process or has no knowledge on how many actions it will be required to perform, such problems are called infinite horizon problems as the process may continue forever. They are also termed as indefinite horizon problems which states that the agent will eventually stop, but does not know when it will halt.

To model such situations, one should augment the MM with actions. At each stage, the agent decides upon an action to perform; the resulting state depends on both the previous state and the current action performed. An agent can receive a sequence of rewards. These rewards incorporate the action costs in addition to any prizes or penalties that may be awarded. Negative rewards are called punishments. A MDP consists of:

- S, a set of states of the domain.
- A, a set of all possible actions.
- $P : S \times S \times A \rightarrow [0, 1]$, specifies the dynamics of the system. This is written P(s′ | s,a), where $\forall s \in S, \forall a \in A, \sum_{s' \in S} P(s'|s, a) = 1$.

$P(s'|s, a)$ specifies the transition probability to state s' provided that the agent is in state s and performs action a. The reward $R(s, a, s')$ gives the expected immediate reward from doing action a and transition to state s' from state s. Furthermore, the rewards can be stochastic and hence vary over time by following a stochastic distribution at the resulting state. In a fully observable MDP, the agents observes the current state and acquires full knowledge in order to perform an action to move to the next state.

2.2.2 Partially Observable Markov Decision Process

A POMDP consists of the following:

- S, a set of states of the world;
- A, a set of actions;
- O, a set of possible observations;

- $P(S_0)$, which gives the probability distribution of the starting state;
- $P(S'|S, A)$, which specifies the dynamics - the probability of getting to state S' by doing action A from state S;
- $R(S, A, S')$, which gives the expected reward of starting in state S, doing action A, and transition to state S'
- $P(O|S)$, which gives the probability of observing O given the state is S.

2.2.3 Stationary Policy

A stationary policy is a function $\pi : S \rightarrow A$, where an action is assigned to each state. Provided a reward criterion, a policy has an expected value for every state. Let $V_\pi(s)$ be the expected value of following π in state s. This specifies how much value the agent expects to receive from following the policy in that state. Policy π is an optimal policy if there is no policy $\pi^{'}$ and no state s such that $V_\pi(s)^{'} > V_\pi(s)$. This means that the policy that has a greater or equal expected value at every state than any other policy.

The expected value of a policy π for the discounted reward, with discount γ, is defined in terms of two interrelated functions, V_π and Q_π. Let $Q_\pi(s, a)$, where s is a state and a is an action, be the expected value of performing a in state s and then following policy π.

Q_π and V_π are defined recursively in terms of each other. If the agent is in state s and performs action a, and arrives in state s', it gets the immediate reward of $R(s, a, s')$ plus the discounted future reward, $\gamma V_\pi(s')$. When the agent is planning and it does not know the actual resulting state, it uses the expected value, averaged over the possible resulting states:

$$Q_\pi(s, a) = \sum_{s'} P(s'|s, a)(R(s, a, s') + \gamma V_\pi(s')). \tag{2.1}$$

$V_\pi(s)$ is acquired by performing the action specified by π:

$$V_\pi(s) = Q_\pi(s, \pi(s)). \tag{2.2}$$

Let $Q^*(s, a)$, where s is a state and a is an action, be the expected value of performing a in state s and following the optimal policy. Let $V^*(s)$, where s is a state, be the expected value of following an optimal policy from state s. Q^* can be defined analogously to Q_π:

$$Q^*(s, a) = \sum_{s'} P(s'|s, a)(R(s, a, s') + \gamma V^*(s')). \tag{2.3}$$

$V^*(s)$ is obtained by performing the action that gives the best value in each state:

$$V^*(s) = \max_a Q^*(s, a). \tag{2.4}$$

An optimal policy π^* is one of the policies that gives the best value for each state:

$$\pi^*(s) = \arg\max_a Q^*(s, a). \tag{2.5}$$

Note that $\arg\max_a Q^*(s, a)$ is a function of state s, and its value is one of the a's that results in the maximum value of $Q^*(s, a)$.

2.3 Fundamentals in Reinforcement Learning

RL strategies dates back to the early days of cybernetics security and also works in statistics, neuroscience and computer science. This strategy has attracted a rapid growth in machine learning and artificial intelligence for the last decade. Its enticing factor is the capacity to program agents by introducing a reward-punishment scheme. However, there are some difficult computational complexities in terms of processing the knowledge. RL is defined as a method of machine learning, where the learner is the decision-making agent which takes actions in an environment and gets a reward or penalty based on its actions to solve the problem. In our case, the learner is the SU, the environment is the PU transmissions or the spectrum available. The problem is defined as the ability of the SU to learn the occurrence of PU transmissions and be capable of predicting such transmissions in the near future, thus reducing the possibility of interference. Figure 2.3 presents the RL application in a CR functioning in a dynamic RF environment.

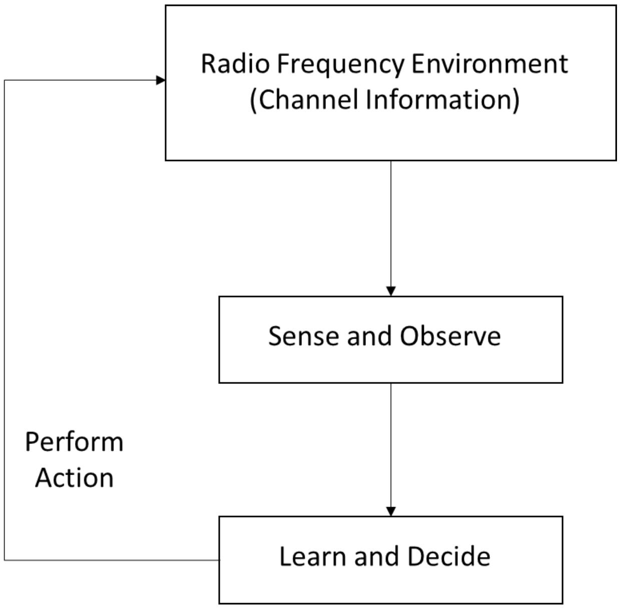


FIGURE 2.3: A simple Reinforcement Learning Model applied to a CR in a RF environment [44].

The RL strategy falls under the category of unsupervised learning such as artificial neural networks (ANN). There lies a significant difference where there is no guidance for the learner in RL strategies compared to unsupervised learning, but solely by critic-learning. Critic learning is stated as learning from experience in which the user learns how well its past actions have helped achieve the outcome based on a reward-punishment scheme. RL techniques have been developed through three major pathways:

- Trial and Error
- Dynamic Programming
- Temporal Difference.

The idea of RL strategy is to deal with problems that require a correct sequential set of actions to be taken by the agent or learner, unlike in unsupervised learning where the main objectives are one or two actions and not a set of actions. The correct set of actions are determined by the maximum cumulative reward the agent (SU) acquires. The agent is connected to the environment via perception and action. On each iteration, the agent receives an input that indicates the current state of the environment. The agent then chooses an action to generate as output to the environment. This action induces the change in the state of the environment. The transition induces a value which is received by the agent as the reward or punishment (negative reward). The behaviour of the actions of the agent should increase the cumulative reward.

Figure 2.3 portrays the interactions between the CR and the RF environment. The learning agent (CR) receives an observation o_t at the state s_t at a given time instant t. The observation is accompanied with a given reward r_t which indicates the reward received for the action a_{t-1} in state s_{t-1} taken at time t. The observation o_t and reward r_t are used by the CR to determine the action a_t. This results in a state transition from s_t to s_{t+1} and delayed reward of r_{t+1}. It is to be noted that the CR is not passive and does not only observe the outcomes from the environment, but also affects the system state via its actions such that it drives the environment to a desired state that brings the highest reward to the agent. MDP is considered to be one of the elements in RL strategy to model the learning process of the agent (SU) and its environmental states. However, MDP requires complete knowledge on the environment. This leads to the development of partial observable MDP where the environment is known only partially and the process is also based on the sequential set of actions of the agent (SU). The other key element to be taken into account is the converging factor or the discount factor. This is used to keep the maximum cumulative rewards to a finite value and for convergence purposes. In order to use RL techniques, one must formulate the current problem. Generally, this is done through a Markov model or MDP. It is assumed that the PU channel activity follows a Markov chain and the SU's attempt to access those channel during idle periods. The key advantage of using temporal difference RL techniques on a MDP based problem is that it yields the optimal result at a lower complexity of processing compared to dynamic programming.

The CR observes the current state s and chooses the action a for the next stage. This is done through the stochastic policy $\pi : A \times S \rightarrow [0,1]$, where $\pi(a,s)$ defines the probability of taking action a at state s. Such a policy is said to be an optimum policy if it maximizes the expected reward r_t, which is usually discounted by a discount factor

$\gamma \in [0, 1]$ in the case of an infinite time horizon.

$$r_t = \mathbb{E}\sum_{k=0}^{\infty} \gamma^k r_{t+k}(s_{t+k}, a_{t+k}) \tag{2.6}$$

RL strategies are usually introduced in economic games due to their capability of learning from experience and the concept was applied from real game models such as chess. The aim of the RL strategy is for the agent to maximize the reward received from the environment and minimize the possibility of punishment.

There are many existing algorithms that are considered for the RL strategies, especially ones based on dynamic programming. One of the most popular ones are the adhoc strategies: Greedy strategies, Randomized strategies, interval-based techniques. Greedy strategies are heuristics where it states the agent should pick the highest estimated pay-off. The problem in such a heuristic is that unlucky sampling could indicate the best action reward is less than the reward from a sub-optimal action. Randomized strategies are somewhat similar to greedy heuristics though with a minor difference. Randomized strategies allow the agent to pick the action with the highest estimated reward by default, but chooses an action randomly with a probability p. This is to prevent the agent's learning curve from remaining at the known best reward (local optimum) and not trying to find alternate set of actions. Interval-based heuristics or softmax action selection heuristics are ones where the greedy action is given the highest selection probability but the other actions are ranked and weighted according to their respective reward values. This leads to the agent choosing an action a on the n_{th} iteration that is calculated by the upper bound confidence interval $(100(1 - \alpha))$ on the success probability of each action. The above techniques mentioned are mainly used when the agent does not need prior knowledge of the environment and thus does not need to perform exploration of the environment. However, our case does require the agent (SU) to perform exploration of the environment (spectrum availability). There are two main features that characterize the RL strategy: trial-and-error and delayed reward. The trial-and-error is assumed that the agent does not have any prior knowledge about the environment and performs actions in a blind manner to explore the environment. The delayed reward feature is the feedback signal that the agent (SU) receives from the environment after performing the action in each iteration. The reward is either positive or negative (punishment) determining how well the performed action was. This leads to the evolution of RL through temporal difference. This consists of updating an evaluation function based on the environment, thereby, improving the total reward. There are a few classic models of RL in temporal difference learning: Value Iteration algorithm and Q-Learning.

2.3.1 Value Iteration

This iterative algorithm is based on the Bellman's principle of optimality [88], where the value function V_t at time t is estimated with respect to V_{t-1} at time $t-1$.

$$V^t(s) = \max_{a \in A} r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V^{t-1}(s') \tag{2.7}$$

It is shown that the value-iteration algorithm guarantees that the estimated value function is Q-optimal over an infinite horizon. Value iteration begins at $t = 0$ and $V_{t=0}$ as a random estimate of the value function. It iterates repeatedly, computing V_{t+1} for all states $s \in S$, until it converges with the left-hand side equal to the right-hand side, which is termed as the Bellman equation.

Incorporating the RL method, one finds an optimal solution to the MDP, without any knowledge of the transition probabilities, thus resulting in the RL scheme being a desired approach for problems with incomplete knowledge. The RL algorithm is based on the temporal difference learning approach that updates the value of each state $V(s)$, after each interaction, is as expressed:

$$V(s_t) \leftarrow V(s_t) + \beta_V [r_{t+1} + \gamma V(s_{t+1}) - V(s_t)]. \tag{2.8}$$

where β_V is the step-size parameter or also termed as the value learning rate.

2.3.2 Q-learning

This temporal difference algorithm that was developed by Watkins [81] provides an estimation of Q-values, $Q(s,a)$ of the joint-action pairs (s,a). This represents the return function of action a when system is in state s and hence defined as:

$$Q(s,a) = \mathbb{E}\sum_{k=0}^{\infty}\gamma^{k}r_{t+k}|s_t = s, a_t = a \tag{2.9}$$

The most commonly used one-step Q-learning is defined as follows:

$$Q_{t+1}(s_t,a_t) \leftarrow Q_t(s_t,a_t) + \alpha_{RL}.(r_{t+1} + \gamma \max_{a'}(Q_t(s_{t+1},a)) - Q(s_t,a_t)) \tag{2.10}$$

where $alpha_{RL}$ is the learning rate and γ is the discount factor.

The update function approximates the optimal Q-value noted as $Q*$. However, it is required that all state-action pairs are continuously updated in order to guarantee correct convergence. This is achieved by applying an ϵ-greedy policy that ensures that all state-action pairs are updated with a non-zero probability, thus leading to an optimal policy. The advantage of Q-learning is that it does not depend on the problem model used as it can be applied on any MDP problem formulation to seek an optimal $Q*$.

2.3.3 Summary of Literature Review

Research in the area of RL incorporated CRNs has not been heavily focused. As a mathematical tool, it is used to learn the environment by means of action rewards for a given environment. There are various methods in the concept of RL, and it has been summarized as a survey in the following paper [44]. It studies the concept of decision making using various learning techniques and identifies the policies used for each methodology. It explains the advantages and the challenges faced in each methodology when dealing with CRNs, especially in decentralized and non-markov environments. Existing similarities and the differences between the respective techniques are discussed as well. One paper that strongly focuses on this area is the paper referenced at [76]. This paper focuses on the detection of the PU system's allocation and detection of spectral resources using RL. The detection is performed frequently on all sub channels of the PU transmission spectrum by using Fast Fourier Transform (FFT). It focuses on the idea of overlay systems as to identify the bands that are not heavily or not used at all. The paper [41] is based on the idea of inclusion of RL in cooperative cognitive Ad Hoc networks for solving the overhead problems such as sensing delay for reporting local decisions and the increase in the control of traffic in the network. The major focus is the ability of the SU to minimize sensing delay and find an optimal set of cooperating neighbours. Meanwhile, the work published in [33] describes a sensing policy created by use of RL to improve energy efficiency in Spectrum Sensing in CRNs. Additionally, the paper [32] studies the model of a distributed multi-agent, and multi-band reinforcement learning based sensing policy for CR Ad-Hoc network. The goal is to maximize the available spectral bands for SUs provided a desired diversity order. The paper [46] discusses the collaboration between the SUs to acquire spectrum leasing, which has been adopted in various schemes such distributed channel sensing and channel access, is an intrinsic characteristic of CR to improve network performance. Though, the requirement to collaborate has inevitably opened doors to various forms of attacks by malicious SUs, and this is addressed using Trust and Reputation Management (TRM) scheme. TRM detects malicious SUs including honest SUs that turn malicious. To achieve a more efficient detection, the use of Reinforcement Learning (RL) is advocated, since it is flexible and adaptable to the changes in operating environment in order to achieve optimal network performance. On the other hand, the paper [47] studies the optimality in selecting cluster-heads in CRNs. It investigates the effectiveness of trust and reputation model (TRM) in clustering as an approach to achieve higher network performance in CRNs. The performance of both the traditional TRM and RL-based TRM schemes was analysed using performance metrics such as packet transmission and dropping in the network The RL-based TRM scheme demonstrated faster detection of malicious SUs, thus indicating the advantageous part of learning trust over time. Next, the paper [85] proposes a novel Q-learning based auction (QL-BA) algorithm for dynamic spectrum access in a one PU multiple SUs scenario. In the auction market, the SU provides a bidding price dynamically and intelligently using a Q-learning based bidding strategy to compete for current access opportunity; meanwhile PU decides to whom to release the unused spectrum according to the maximum bidding principle. From the above literature in RL work, one can conclude that it has been applied in the different aspects of CRs ranging from spectrum sensing, leasing, and cluster head selection to auction of spectrum and trust issues between the adjacent SUs.

2.4 Fundamentals in Game Theoretic Models

The approach of game theory into the PHY layer security in CRNs started in recent years. It is quite a new area of research that has commenced recently and is also of heavy interest. Game theory in lamyman's terms, is the mathematical tool that allows the system/ user/ player of interest to make an informed decision that optimizes the overall objective and reaches an equilibrium based on the presence of other players. In order to provide reliable transmission to the authorized user with minimal interference from attackers on the various channels available for transmission, game theoretic models were proposed as a means of PHY layer security and resilience. The concept of physical security is mainly divided into two categories:

- The first type is to provide a reliable form of transmission to the receiving end, even when the authorized users are subject to interference. This focuses on the secrecy capacity and Shannon's theorem of secure transmissions.
- The second type is to provide a method of counterattacking the attackers and preventing the authorized users being subject to interference. In this category, we shall focus on the application of game theoretical models such as Colonel Blotto, zero sum games with Nash equilibria, evolutionary games, Stackelberg games, etc.

A through literature study was done on the above categories and the relevant approaches in relation to the scenarios presented in the ABSOLUTE architecture.

In general, game theory is defined as a mathematical tool that analyses the strategic interactions among multiple decision makers or players. Game theory history dates back to the original work of J. Von Neumann and O. Morgenstern [49] that included the study of zero-sum games and laid the foundation of game theory. Later on, cooperative game theory was introduced which is the analysis of optimal strategies for groups of players based on the fact that collaboration can be enforced to jointly improve their status in the game model. Furthermore, J. Nash established an innovative criterion known as the Nash Equilibrium to characterize the mutual consistent strategies of players, which was mainly brought in for non-cooperative game models. The users in a CRN make intelligent decisions on their spectrum usage and operating parameters based on sensed spectrum dynamics and actions adopted by other users. Figure 2.4 depicts the four types of game theoretical approaches [17] in spectral sharing approaches.

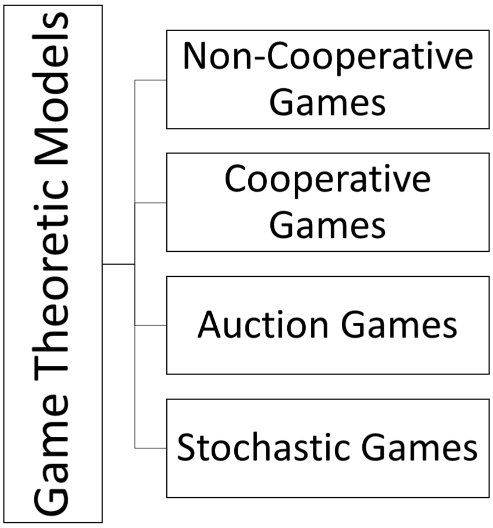


FIGURE 2.4: Types of Game Theoretic Models.

The general reason for a game theoretic study on CRNs, is that users who compete for spectral resources may compete with each other and behave selfishly. Thus, it is feasible to study the intelligent behaviours and interactions of the players from a game theoretical perspective. To isolate the general scenario in a manner similar to the ABSOLUTE scenarios, we commence a study on such intelligent behavioural patterns to provide some means of reliable communication and minimize the overall interference to the architecture from intentional or unintentional jamming.

The interactions between the authentic user and attacker could be modelled as a zero-sum non-cooperative game. Stochastic games would be useful in the derivation of optimal defence strategies that accommodate the environment dynamics and counter-attacking strategies. Coalitional games allow the authentic users to form parties and defend in a more optimized manner against attackers. Evolutionary game theory allows the authentic users to find stable equilibria strategies that will prevent or minimize the effect of possible future attacks. Thus, one could consider game theory as an effective means to provide security to the CRNs from existing literature.

2.4.1 Non-cooperative games

Non-cooperative games are games consisting of two or more players where each players compete for the spectral resources, make decisions that are not based on the strategies of the other players as in our case, thus resulting in non-optimal plays [17]. The strategic game is often comprised of three factors:

- N is the number of players in the game
- A set of actions 'A_i'
- Payoff/ Utility function 'U_i'

The game is denoted as $< N, (A_i), (U_i) >$. The interactions between the players are modelled as a game. Therefore, NE was introduced as a criterion or solution concept in non-cooperative games where each player selects the strategy that maximizes its possible reward. NE is defined as the solution for a non-cooperative game consisting of two or more players, where each player knows the equilibrium or optimal strategy of the other players, although one cannot gain any form of reward by unilaterally changing its strategy. NE is the solution concept where it provides the best strategy given that all the other players stick to their respective equilibrium strategies too. Most non-cooperative games are solved by zero-sum games as a means of defence against jamming attacks. In the ABSOLUTE scenarios, the game models are considered to be non-cooperative where the two different group of players compete for spectral characteristics and resources. Under non-cooperative game theory, comes the idea of zero-sum games that involves two different players, where the game is strictly competitive such that the one player's gain is balanced by the loss in the utility of the other player. The total sum of the utility of both players results in a zero value. Zero-sum games are non-cooperative and hence usually solved with the use of the NE concept. Colonel Blotto games fall under zero-sum games which consists of two players where the players are assigned to distribute limited resources over several entities [18].

Zero-Sum Games

A zero-sum game is a mathematical tool that is used to represent a game comprising of two or more non-cooperative players where the sum of all the players' payoffs/utilities sum to zero, albeit their respective strategies. In mathematical terms, zero-sum game is defined as the following:

$$G = (N, (A_i), U_i) \text{ if } \forall (a_1, ..., a_{\mathrm{N}}) \in A_1 \times .....A_N \sum_{i \in N} U_i(a_1, ..., a_N) = 0$$

The zero-sum property (one player gains, another loses) means that the result is a Pareto optimal due to its conflicting nature. A Pareto optimum is defined as a sub-optimal point where the state of the individual player/ player of interest' payoff cannot be improved without changing the other player's strategy. In this thesis, we shall focus on two-player zero-sum games, where one player tries to maximize his payoff, while simultaneously minimizing the payoff of the other player.

Nash Equilibrium

We are interested in the solution concepts for zero-sum games. A convenient way to argue about such games is to consider an interactive two stage model where one player acts first and the other player acts second; as always, each of the players tries to optimize his own payoff in a sequential manner. A powerful concept in games is represented by the celebrated Nash Equilibrium (NE), that describes the situations in which none of the two players can achieve a better payoff/ utility by unilaterally changing its transmission strategy π (i.e., when the other agent does not change its own strategy).

Let (A, U) be a game with N players, where A_i is the strategy action set for player i, $A = A_1 \times A_2 \times \cdots \times A_N$ is the set of strategy action profiles and $U(x) = (U_1, \ldots, U_x$, is its payoff/ utility function at $x \in A$. Let x_i be an action profile of player i and x_{-i} be a strategy profile of all players except player i. When each player $i \in \{1, \ldots, n\}$ chooses strategy x_i resulting in strategy profile $x = (x_1, \ldots, x_n)$ then player i obtains payoff $U_i(x)$. A strategy profile $x^* \in A$ is a NE if there is no unilateral change in strategy by any single player which is profitable for that player;

$$\forall i, x_i \in A_i : U_i({x_i}^*, {x_{-i}}^*) \geq U_i(x_i, {x_{-i}}^*). \tag{2.11}$$

Fictitious Play

Fictitious Play (FP) is a learning mechanism for game theoretic approaches. As mentioned prior, zero-sum games are played for one interval, with the given knowledge of the opponent's strategy. This is defined as a static case. FP is defined as a belief-based learning rule, where the players form beliefs about opponents' play (from the entire history of past play) and behave rationally with respect to such beliefs. FP is used to approximate the value and optimal strategies of a finite game. It is a sequential procedure that approximates the value, giving upper and lower bounds that converge to the value and strategies that achieve these bounds.

It is also myopic, since players try to maximize current payoff without considering future payoffs. They are not learning the "true model" generating the empirical frequencies (how their opponent is actually playing the game). Thus, every player plays a best response to opponents' empirical distributions. If a player plays a strategy with an empirical frequency of 1, that strategy is stated to be *pure*. If not, the player has a few strategies whose empirical frequencies sum up to 1. This strategy set is stated to be *mixed*.

Let $U(i,j)$ be a $m \times n$ payoff matrix. The method initially starts with an arbitrary value $m > i_1 > 1$ for player I. Each player then chooses his next strategy as means of a best response assuming the other player chooses among his previous choices at random equally. In other words, player II chooses strategy j_k based on player I previous chosen strategies $i_1, \ldots, i_k$, where k is the number of intervals, such that j minimizes the expectation $\frac{1}{k}\sum_{l=1}^{k} U(i_l, j)$. On the other hand, player I would choose strategy i_{k+1} at the next step so that i maximizes the expectation $\frac{1}{k}\sum_{l=1}^{k} U(i, j_l)$.

$$j_k = \arg\min \frac{1}{k} \sum_{l=1}^{k} U(i_l, j) \qquad i_{k+1} = \arg\max \frac{1}{k} \sum_{l=1}^{k} U(i, j_l) \tag{2.12}$$

As k approached infinity, the lower and upper bounds j_k and i_{k+1} converges to the equilibrium of the game.

2.4.2 Summary of Literature Review

The concept of security or channel resilience in CRNs derives from the fact that there exist unauthorized users who take advantages of the network in an unfair manner thus misusing trust, or intentionally interfering with other users to disrupt the entire network. The concept of resilience in the channels of the CRN is a novel approach that provides the authorized system the capability of successfully detecting interfered channels and providing a probabilistic means of successful transmission for channels under jamming effects.

There has been significant literature in learning methods and game theory. However, the concept of using the mathematical tool known as game theory to be applied in CRNs as means of physical reliance or defence is something that has been put forward recently. Game theory acts as an optimization tool in a set of decisions for varying dependant environments. Current literature has explained the rising opportunities and disadvantages faced, when dealing game theory with dynamic spectrum access in CRNs.

The paper [86] investigates the situation where a SU can access only one channel at a time and hop among different channels, and is modelled as an anti-jamming game. Analysing the interaction between the SU and attackers, a channel hopping defence strategy is derived using the Markov decision process approach with the assumption of perfect knowledge, leading to the proposal of two learning schemes for SUs to gain knowledge of adversaries to handle cases without perfect knowledge. The paper [4] summarizes the possible types of jamming attacks from a malicious user in CRNs based on each layer of the OSI model and describes the possible prevention mechanisms for each type. A similar paper [54] was published describing the problems involved in sensing and management, attacks on CRNs, attacks on various network layers, threats on CRNs, and the current security and privacy solutions available in CRNs. As the above papers focus on security issues, the paper [17] introduces the most fundamental concepts of game theory, and explains how these concepts are leveraged in designing spectrum sharing protocols, with an emphasis on state-of-the-art research contributions in CR networking. Research challenges and future directions in game theoretic modelling methods are also outlined. The application of game theory is well described in paper [84] which analyses the spectrum allocation problem of CRNs. To deal with the multi Nash equilibrium problem of non-cooperative game based spectrum sharing in CRNs, varying utility of CR users was considered to judge the stability after several iterations, and hence design an improved non-cooperative spectrum allocation algorithm. The paper [45] describes a stochastic auction game termed as the Stackelberg game applied to CRNs where CRs act as master-slave pairs to update their transmission powers and frequencies simultaneously. Meanwhile, the paper [13] takes a non-cooperative game approach where the SU is aware of the absence of several PUs and the presence of a malicious user, thus allocate power to the bands with a randomized strategy, in hope of alleviating the damage caused by the malicious user. The scenario is a two-player zero-sum game, and the unique Nash Equilibrium is shown to exist under certain conditions using the Colonel Blotto game approach, which provides a minimax strategy that the SU should adopt in order to mitigate the worst-case damage caused by the malicious user. Additionally, the paper [12] proposes a stochastic game framework for anti-jamming methods in CRNs. SUs observe the spectrum availability, the channel quality, and the jamming strategy from the status of jammed channels. This leads to the decision on how many channels the SU should reserve for transmitting control and data messages and to switch between the different channels based on priority of overall efficiency. Using the minimax-Q learning, SUs gradually learn the optimal policy, which maximizes the expected sum of discounted payoffs defined as the spectrum-efficient throughput. Overall, we can conclude some significant work has been done in game theory and it is mainly based on frequency hopping to the available channel, allocating and accessing spectral bands that are least used, auctioning for more power and competing between multiple CRs, and switching between the types of channels such as control and data based on priority. Aforementioned papers do not focus on the total number of channels that are successful subjected to jamming scenarios. This is where our work has been directed in.

Chapter 3

Spectrum Sensing in Aerial Communications

Spectrum Sensing is of dire importance in CRNs as they comprise of SUs with no license to access the available spectrum in a geographical area at a given time. The main advantage of CRNs is the capability of spectrum sharing, where the SUs are able to access spectral holes (white, grey). This means that the SUs are sharing the spectrum with the consent of the PU system. The major objective of the SU is to minimize the amount of collision/ interference caused by its own transmissions to the PU system. Therefore, an effective spectrum sensing technique is required, taking into account of the radio propagation characteristics in aerial-terrestrial communication systems.

3.1 Background

This chapter shall focus on the spectrum sensing method to be applied by CRNs for various radio propagation characteristics such as small-scale fading, path loss, log-normal shadowing, and multi-path propagation. The characteristics are defined below for future reference.

- Fading - We shall focus on small scale fading as we shall deal with the received signal being an entirety of multiple versions of the transmitted signal over time, due to the positions of various reflectors and scatterers. The fading models considered with regards to the presence and absence of LOS are: AWGN, Rayleigh and Rician.

- Path Loss - The large-scale effects cause the received power to vary gradually due to attenuation of the signal, which is determined by the geometry of the signal path profile.
- Log-normal Shadowing - This is also a fading model, but comprises of large scale variations caused by the shadowing of obstacles along the signal path.

The sensing scheme used to detect the presence of a signal is looked upon with considerations taken for the above mentioned characteristics. This falls under spectrum detection which is one of the pivotal areas of research in CRNs due to the capability of DSA. We focus on the energy based sensing scheme for the various fading models and perform an extension in the performance of the scheme under Rician channels by simulation, where theoretical results are limited.

3.2 Need for Spectrum Sensing and my Contribution

Aerial-Terrestrial communications require spectral resources at various geographical areas due to unexpected deployment scenarios [66, 64]. Hence, there is an unexpected increasing demand for spectrum as the need for higher data rates is increasing, mainly as a result of the transition from voice-only communications to multimedia-type applications. This brings forth the need of aerial to terrestrial communications where the down link and uplink channels may vary in terms of spectral occupancy at a given geographical area. The aerial infrastructure should sense the spectrum to determine which parts of the spectrum are available before commencing utilization.

The electromagnetic spectrum being of an infinite amount is an illusion. The electromagnetic spectrum is a natural finite resource in terms of communication technology, where the number of transmitters and receivers are limited and are licensed by major authorities. The problematic situation is not the physical shortage of spectrum rather the access of the spectrum which is controlled by the federal communication authorities[34] which lead to the proposition of DSA.

Haykin [29] states that CRs are projected to be brain-empowered wireless devices that are aimed at improving the overall utilization of the electromagnetic spectrum.

This vision is to be accomplished by use of the methodology of understanding-by-building in CRs which has two major objectives: Permanent reliable communications and efficient utilization of the available spectrum resources. As stated in terms of CR terminology, the SUs have the lower priority to exploit the specified area of the spectrum in a way that does not cause interference to the PU [89]. Therefore, the SU requires the acquisition of CR capabilities in order to exploit the unused part of the spectrum at temporary events.

Energy-Based Sensing involves the acquirement of the spectral content or RF energy over the spectrum and is less complex in terms of computational processing[68, 87, 27, 22, 8, 52, 51]. The latter involves the obtaining of the spectrum usage characteristics across multiple dimensions such as time, space, frequency and code and thus, leading to powerful signal analysis techniques and increased computational complexity and processing.

We focus on the energy detection based sensing scheme for detecting the PU in the environment of an aerial-terrestrial system. The incumbent PU and SU scenario considered for the energy based spectrum sensing is described in detail in Section 3.3. We perform Monte-Carlo simulations for the proposed spectrum sensing scenario by considering different fading channels. We especially consider the Rician fading channel that models the terrestrial-aerial uplink channel where there is a LoS component. We note here that closed form theoretical expressions for the detection probability in a Rician fading channel only exists for a very limited case and therefore we present an extensive set of simulation results that is necessary to study the sensing performance for our proposed system. Moreover, we also verify our simulation model for the Rayleigh fading channel by comparing with the theoretical expressions for the detection probability which is found in the current literature.

3.3 Aerial-Terrestrial LTE Network Model

The scenario considered is a rapidly deployable temporary aerial-terrestrial network architecture with an aerial base station using the LTE technology [2], for a disaster recovery operation [7]. Note that the regular LTE network providing services in the disaster incident area will be damaged in such situations and therefore the temporary aerial-terrestrial LTE network provides supplementary services to that area for the first responders and other service men. At this point, we refer to the network that provides the regular coverage as the Incumbent-LTE (I-LTE) which is characterized as the PU and the temporary LTE network as the Secondary-LTE (S-LTE) which is characterized as the SU. Therefore, the S-LTE network uses the spectral band of the I-LTE network until the I-LTE network is reconstituted. The terrestrial nodes in this area consist of the user equipment (UE) from the S-LTE network known as secondary UE (S-UE) and UEs from the I-LTE network known as the incumbent UE (I-UE) which are temporarily disconnected from services.

Figure 3.1 presents the S-LTE aerial terrestrial communications architecture (derived from the ABSOLUTE project [7]) where the aerial LAP terminal is considered to be the eNodeB for the S-LTE network. The figure illustrates the structure and behaviour of the model consisting of the S-UEs and the (damaged) I-LTE network with the terrestrial I-UEs.

When the S-LTE network is in operation in such a disaster recovery situation, the I-LTE network will slowly commence to recover its operation and begin to provide coverage to the I-UEs. Therefore, the eNodeB of the S-LTE (i.e. the LAP) needs to perform intelligent spectrum sensing to detect the spectrum usage by the terrestrial eNodeB in the vicinity, as a Frequency Division Duplexing (FDD) LAP is considered in this scenario. The work conducted proposes the method in which the S-LTE eNodeB can perform spectrum sensing to detect the incumbent users and avoid interference to them.

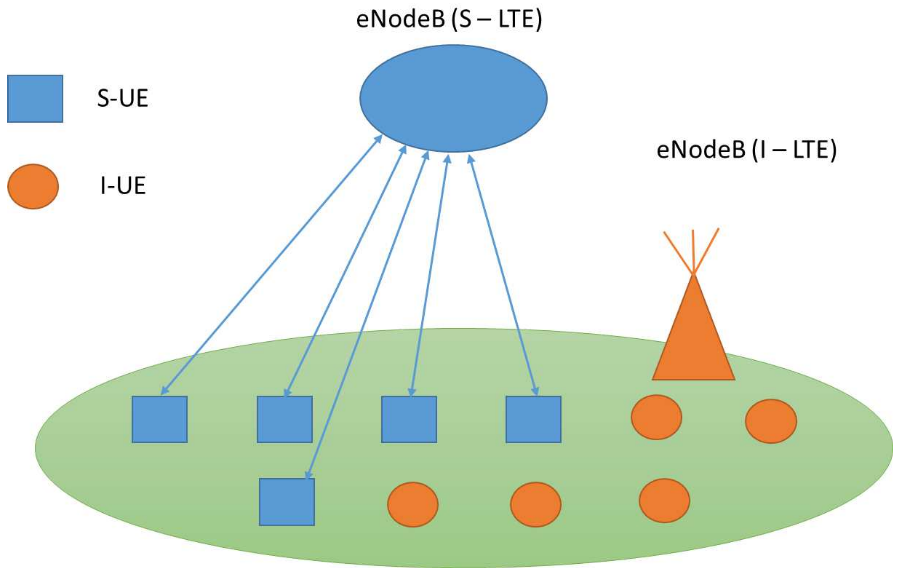


FIGURE 3.1: Temporary aerial-terrestrial LTE network supplementing a regular (damaged) LTE Infrastructure.

3.4 Spectrum Sensing by the LAP eNodeB

The main interest lies in the uplink of the LTE that uses SC-FDMA for the transmissions from the UEs to the eNodeB. The underlying problem is to detect the uplink transmissions of the I-UE (i.e. when the I-LTE starts to recover and provide services to the I-UEs) by the LAP eNodeB in the S-LTE network. Here we propose how the LAP (S-LTE) eNodeB can perform energy based detection to identify the presence of the I-UE's using the uplink signals received by the S-LTE eNodeB from the terrestrial I-UEs. In order to propose this mechanism, we consider the following two cases:

- Case 1: S-LTE eNodeB sensing for the presence of I-UEs, when a S-UE is connected to the S-LTE eNodeB in that particular frequency band.
- Case 2: S-LTE eNodeB sensing for the presence of I-UEs, when a S-UE is NOT connected to the S-LTE eNodeB in that particular frequency band.

In Case 1, in order to detect the I-UEs the S-LTE eNodeB utilizes the resource slots or channels where the S-UE transmits the Sounding Reference Signal (SRS) to perform spectrum sensing. Such reference signals are used in LTE for the UEs to have good links with the eNodeB by selecting the best resource slot by means of channel sounding [1].

TABLE 3.1: Advantages and Disadvantages of Energy Based Sensing

Advantages
Lower Computational Complexity
Low Computational Cost and Processing
Does not require any prior knowledge of the PU
Disadvantages
Inability to differentiate interference between PU and noise
Poor performance under low Signal-to-Noise Ratio(SNR)
Does not work well in spread spectrum signals

Therefore in Case 1, the S-UEs will mute their transmissions during the SRS transmit period (not always but in a periodic manner) such that the S-LTE eNodeB can use this time slots to sense for any I-UEs in the environment in that frequency band.

However in Case 2, since there are no S-UEs connected to the S-LTE eNodeB, the S-LTE eNodeB can perform sensing to learn the radio environment for the presence of any I-UEs in that particular frequency band. The sensing knowledge can be then used to perform intelligent resource allocation to the S-UEs in the S-LTE network.

3.5 The Energy Detector Model

The energy based sensing method is chosen for the detection of the I-UE in the environment by the LAP eNodeB and the sensing is performed at the LAP eNodeB. It is the most common and least complex method of spectrum sensing. This is because the receivers do not need any prior information on the PU signal so they detect the PU in a blind manner[89][77]. The signal is detected by comparing the output of the energy detector with a detection threshold that depends on the noise level[89]. Table 3.1 compares the advantages and disadvantages of Energy Based Sensing [89][90].

The received signal at the LAP eNodeB from the I-UE terrestrial terminals in a random wireless channel is given by:

$$r(t) = \frac{1}{\sqrt{L(d)}} h(t)s(t) + w(t) \tag{3.1}$$

where, $r(t)$ is the received signal, $s(t)$ is the transmitted signal from the terrestrial terminal, $h(t)$ is the random wireless channel explained in detail later, $w(t)$ is the Additive

White Gaussian Noise (AWGN), and $L(d)$ is the path-loss between the transmitting terrestrial terminal and the LAP base station at a distance of d m. We assume that the altitude of the LAP is also given by d. The path-loss is given by the log-distance path-loss model as,

$$L(d) = L(d_0) + 10\alpha \log_{10}(d/d_0), \tag{3.2}$$

where α is the path-loss exponent, $L(d_0)$ is the free space path-loss in dB at a distance of d_0 given by $L(d_0) = 20\log_{10}(4\pi d_0 f_c/c)$ and f_c is the carrier frequency of $s(t)$ and $c = 3 \times 10^8$. The test statistic used to perform the decision for detecting the presence of PU is the signal energy given by:

$$\epsilon = \int_0^t |r(t)|^2 dt. \tag{3.3}$$

In the discrete domain, the energy of the received signal ϵ is given by,

$$\epsilon = \sum_{n=0}^{N_s-1} |r(n)|^2, \tag{3.4}$$

where N_s is the number of samples per estimate. If T is the duration of the signal and F_s is the sampling frequency, it leads to $N_s = TF_s$ also know as the time bandwidth product. The mean signal-to-noise ratio of the received signal in terms of power ratio is defined as:

$$\rho = \frac{P_s}{P_n} = \frac{1}{N_s}\left(\frac{E}{N_0}\right) \tag{3.5}$$

where, P_s is the mean power of the signal term $h(t)s(t)/L(d)$, P_n is the power of the AWGN term $w(t)$ in (6.1), E is the mean energy and N_0 is the noise power spectral density. The noise power spectral density is estimated through a mean and variance based on a gaussian distribution.

3.5.1 Random Fading Channel Models

There are three types of communication channel models between the I-UE and the LAP eNodeB that are taken into account, namely:

- (i) AWGN only model without fading with $h(t) = 1$

- (ii) slow fading Rayleigh envelope model where $h(t)$ follows a Rayleigh distribution
- (iii) slow fading Rician envelope model where $h(t)$ follows a Rician distribution

Under the two fading models, Rayleigh and Rician, the instantaneous energy per noise spectral density ratio γ follows the following distributions respectively [25][38],

$$f_\gamma(\gamma) = \frac{1}{\bar{\gamma}} \exp\left(\frac{-\gamma}{\bar{\gamma}}\right) \tag{3.6}$$

$$f_\gamma(\gamma) = \frac{1+K}{\bar{\gamma}} \exp\left(-K - \frac{-(1+K)\gamma}{\bar{\gamma}}\right) \Phi \tag{3.7}$$

where $\bar{\gamma} = E/N_0$ is the mean energy per power spectral density ratio, $\Phi = I_0\left(2\sqrt{K(1+K)\gamma/\bar{\gamma}}\right)$, I_0 is the zeroth order Bessel function of the first kind and K is the Rice factor.

3.5.2 The Detection Process

Let us define the two binary hypotheses related to the received signal as the following,

$$\begin{aligned} H_0 &: r(t) = w(t) \\ H_1 &: r(t) = \frac{1}{L(d)} h(t)s(t) + w(t). \end{aligned} \tag{3.8}$$

Then, the decision D determined as channel occupancy detection is made based on the presence or absence of the PU signal by means of the following rules:

$$\begin{aligned} D &= 0 \text{ if } \epsilon < \lambda \\ D &= 1 \text{ if } \epsilon \geq \lambda. \end{aligned} \tag{3.9}$$

Finding the optimal threshold for a given detection criteria is a challenging task, especially with the inclusion of fading channels.

3.5.3 Detection Performance

The performance of the detector is quantified by the two probabilities, probability of detection and the probability of false alarm. The probability of detection P_D is defined by,

$$P_D = Pr(\epsilon > \lambda | H_1) \tag{3.10}$$

and the miss detection probability is given by $P_M = 1 - P_D$. The probability of false alarm P_F is defined as,

$$P_F = Pr(\epsilon > \lambda | H_0). \tag{3.11}$$

The performance curve used for analyzing the detection performance is given by the Receiver Operating Characteristics (ROC) curve which is a plot of P_D versus P_F for various values of λ or equivalently or alternatively the complementary ROC (C-ROC) curves which is a plot of P_M versus P_F.

3.6 Theoretical Detection Performance

In this section, the theoretical expressions are presented for P_D and P_F for the channel models presented previously, which are explored in the existing literature.

3.6.1 AWGN Channel

In an AWGN channel, the energy based statistic follows a non-central and a central chi-squared distribution under H_0 and H_1 respectively with N_s degrees of freedom [77]. This leads us to the computation of the theoretical P_D and P_F as shown respectively [25][59],

$$P_D = Q_{N_s/2}(\sqrt{N_s \rho}, \sqrt{\Lambda}) \tag{3.12}$$

$$P_F = \Gamma(N_s/2, \Lambda/2) \tag{3.13}$$

where $Q_{N_s/2}(.,.)$ is the generalized Marcum Q-function and $\Gamma(.,.)$ is the normalized incomplete upper gamma function and Λ is the detection threshold of the energy detector.

3.6.2 Rayleigh and Rice Channels

In the fading case, the energy detection performance varies when the received signal component has a time varying envelope. The mean probability of detection $\bar{P}_D$ for fading channels is obtained by averaging P_D in (3.12) which is a function of γ with respect to the distribution of γ as follows,

$$\bar{P}_D = \int_0^\infty P_D(\gamma) f_\gamma(\gamma) d\gamma \tag{3.14}$$

The above integral has been solved for the Rayleigh and the Rice channels in [59], and are presented below. For the Rayleigh channel the mean detection probability is given by,

$$\bar{P}_D = \exp\left(\frac{-\lambda}{2}\right) \sum_{n=0}^{N_s/2-2} \frac{1}{n!}\left(\frac{\lambda}{2}\right)^n + \left(\frac{1+\bar{\gamma}}{\bar{\gamma}}\right)^{N_s-1} \Lambda \tag{3.15}$$

where,

$$\Lambda = \exp\left(\frac{-\lambda}{2+2\bar{\gamma}}\right) - \exp\left(\frac{-\lambda}{2}\right) \sum_{n=0}^{N_s/2-2} \frac{1}{n!}\left(\frac{\lambda\bar{\gamma}}{2(1+\bar{\gamma})}\right)^n \tag{3.16}$$

and for the Rice channel the mean detection probability is given for the special case of $N_s = 2$ as,

$$\bar{P}_D = Q\left(\sqrt{\frac{2K\bar{\gamma}}{1+K+\bar{\gamma}}}, \sqrt{\frac{\lambda(1+K)}{1+K+\bar{\gamma}}}\right) \tag{3.17}$$

Note that since the closed form expression for the mean detection probability under the Rice channel does not exist for the generic N_s case, Monte Carlo simulations are conducted for the studying the performance of the energy detector under such fading characteristics.

3.7 Simulation Results & Performance Analysis

We perform Monte Carlo simulations to compare the theoretical results with the simulation. The results for energy based spectrum sensing under various fading channels are described in this section based on varying fading channels and their properties. Firstly, the simulation results are verified by comparing them with the theoretical results for AWGN, Rayleigh and Rice (for Rice the theory only exists for $N_s = 2$) channels, and only then, the simulations are extended to show the energy detector's performance under Rice fading channel by varying N_s, the LAP eNodeB altitude, ρ and K. The following transmission parameters were fixed in our simulations; $f_c = 2.6$GHz, transmit power $P_t = 13$dBm, $\alpha = 2.4$, and $P_n = -105$dBm. The terrestrial terminal is considered to be directly below the LAP or in other words the distance d between the transmitting terrestrial terminal and the sensing LAP base station is the same as the LAP altitude without loss of generality. Furthermore, the received signal power in dBm is given by $P_s = P_t - L(d)$.

Figure 3.2 shows the C-ROC curves for the AWGN and the Rayleigh fading channels under different values of ρ and N_s. From the figure, it is clearly seen the degradation in the detection performance for the Rayleigh fading case compared to the AWGN case, and furthermore, verification of the simulation and the theoretical results matching is conducted for the AWGN and the Rayleigh channels. Figure 3.3 depicts the C-ROC curves for the Rice fading channel (specifically for $N_s = 2$) for different values of ρ and K. From the figure, it is observed that the detection performance under the Rice channel approaches the detection performance of the AWGN channel as expected. Thus, it can be observed that the simulation and theoretical results match well with each other verifying our simulation model for the Rice channel as well.

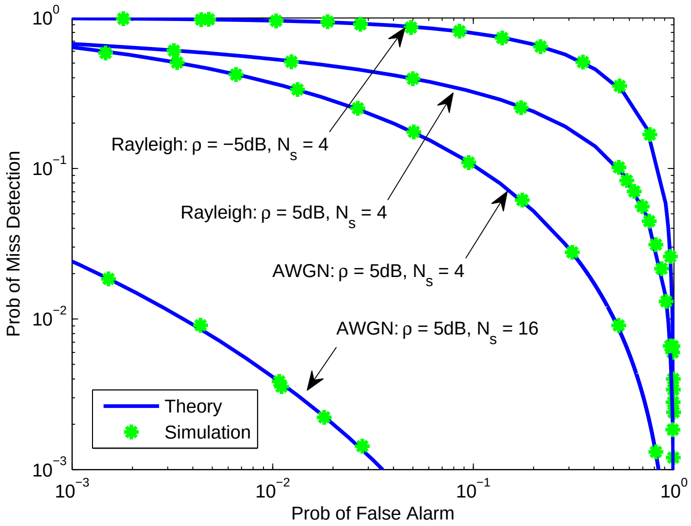


FIGURE 3.2: Complementary ROC for the AWGN and Rayleigh envelope fading channels, theory versus simulation comparisons.

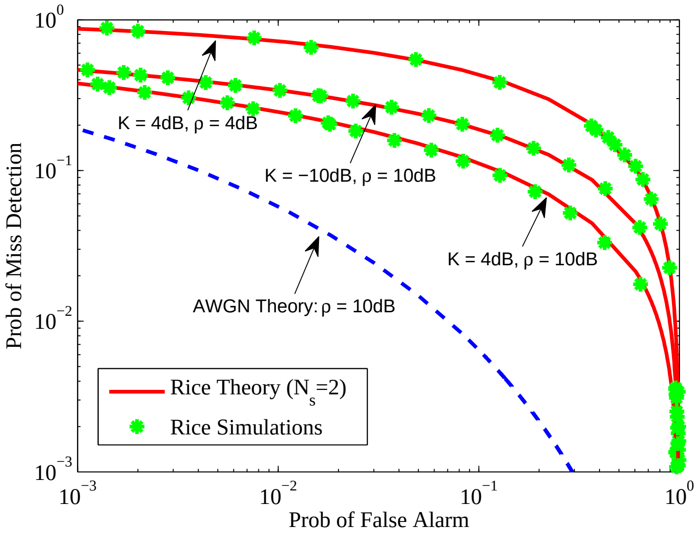


FIGURE 3.3: Complementary ROC curves for the Rice envelope fading channel, with $N_s = 2$, comparing theory versus simulations.

After verification from the simulation model with theoretical analysis from Figure 3.2 and Figure 3.3, we further conduct simulations to study the detection performance of the aerial base station under the Rice fading channel. One should note that as mentioned before theoretical results do not exist in literature for energy detection under Rice channel for $N_s > 2$, and therefore we present simulation results here. Such simulation results can be used to identify how many samples are required to perform energy based sensing in order to maintain a prescribed detection probability to be set by the regulatory body, as an example let us set the required minimum detection probability for the operation is $P_D = 0.9$ or alternatively $P_M = 0.1$. Figure 3.3 depicts the C-ROC curves under the Rice channel for various values N_s, as expected increasing the number of samples for the energy estimate improves the detection performance and the results shown in Figure 3.4 quantifies the detection performance in terms of the C-ROC curves.

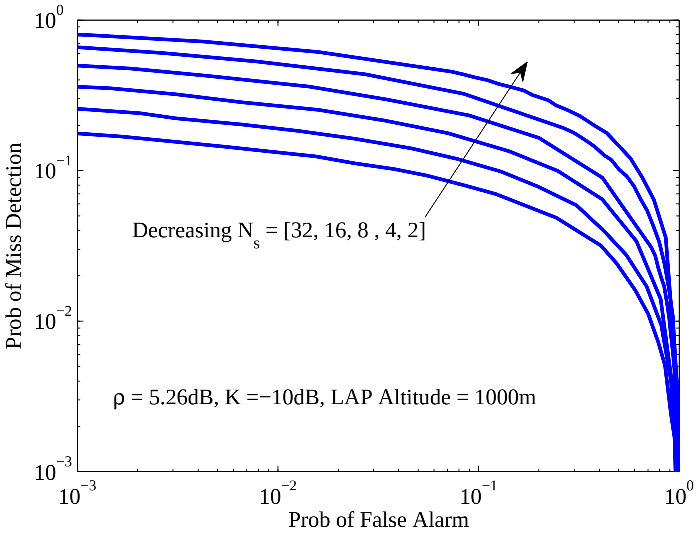


FIGURE 3.4: Complementary ROC curves for Rice envelope fading channel with varying N_s.

The detection performance was analysed for various LAP altitudes under the Rice fading channel and the corresponding results are presented in Figure 3.5. The results clearly indicate that the detection performance degrades with increasing LAP altitudes mainly due to the drop in the mean received signal to noise ratio resulting from the increased path-loss between the LAP and the terrestrial terminal. It is clearly identified that even with $N_s = 20$ the detection performance can only be met with a high value of false alarm probability. Therefore, it is desired to increase N_s to meet the required detection performance with an acceptable (low) false alarm probability.

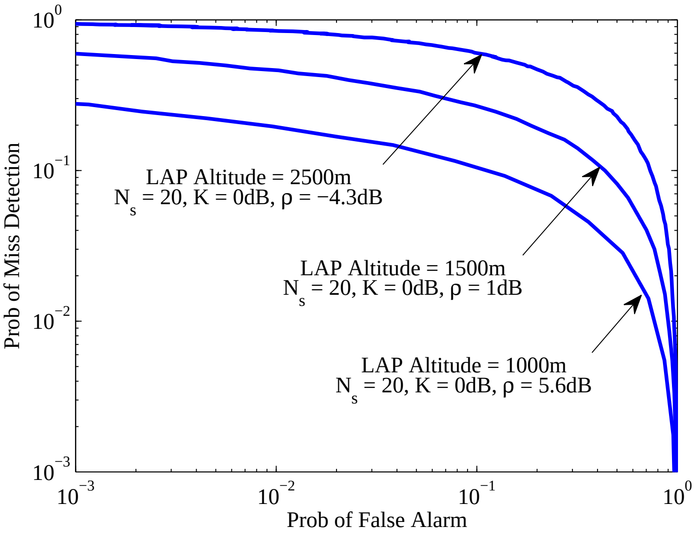


FIGURE 3.5: Complementary ROC curves for Rice envelope fading channel with varying LAP altitudes.

Figure 3.6 on the other hand, depicts the detection performance under various Rice factor values K for $N_s = 10$. Again, it is observed that when K decreases the desired detection performance cannot be achieved without sacrificing on the false alarm probability. Therefore, using the simulation results one is able to decide on the required number of samples N_s for the Rice channel for varying channel conditions.

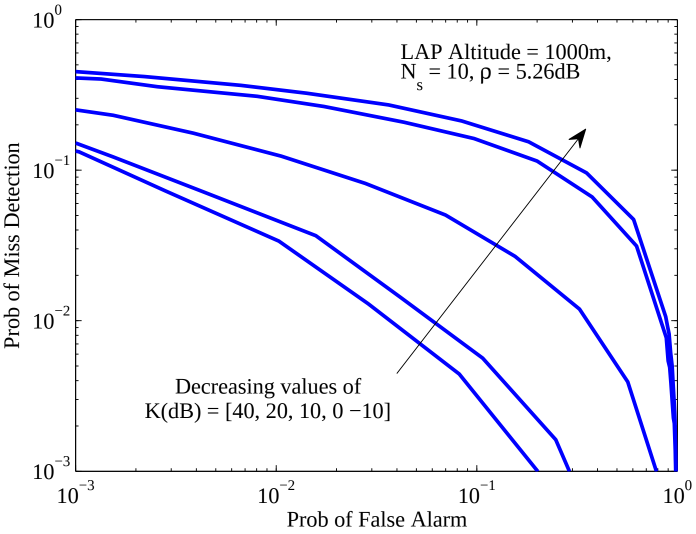


FIGURE 3.6: Complementary ROC curves for Rice envelope fading channel with varying Rice factor K.

3.8 Summary of Contribution & Conclusion

This chapter's major contribution is the set of results acquired through Monte Carlo simulations for the Rician channels with sample sizes exceeding two, where theoretical results are limited for the fading models between aerial and terrestrial communications.

The chapter concludes the issue of detecting incumbent UEs by a secondary aerial eNodeB in a disaster recovery scenario which is presented as a novel sensing approach at the aerial eNodeB for detecting the presence of the incumbent UEs in the environment considering the 3GPP-LTE specifications. In particular, the utilization of time slots for the sounding reference signal transmission from the secondary UEs to the aerial eNodeB to perform the sensing is of key concern. The energy detector is presented under a Rician fading environment and an extensive set of simulation results are presented to study the detection performance at the aerial eNodeB. Simulation results demonstrate that the number of samples play a major role in obtaining the required detection performance especially under the Rice fading channel for which no closed form theoretical solutions exist for the detection probability.

Chapter 4

Channel Access in Cognitive Radio Networks using Reinforcement Learning

CRs must be able to learn and make decisions in an autonomous manner with regards to minimal interference to the PU system. Markov models have been used to formulate the problem of interference between PU and SU transmissions at a given channel. The CR should be capable to operate in an unknown RF environment (channel knowledge), as they are unlicensed users/ SUs in any geographical area at a given time interval. Furthermore, the CR should learn the frequency of PU transmissions and determine an appropriate action, such that interference and channel wastage is kept minimal. This is done by means of unsupervised learning method denoted as RL.

4.1 Background

The basic idea of CRNs is the capability of radios to make intelligent decisions and perform DSA in an autonomous manner. CRs are aimed at improving utilization of the spectrum by applying various spectrum sharing techniques. Therefore, CRs require to be self-programming in order to learn and adapt to their radio environments. A CR should perform its cognitive tasks and be capable of learning and reasoning with the received input. This is performed through the application of machine learning algorithms which are categorized as supervised and unsupervised learning. We shall focus on an unsupervised learning method termed as Reinforcement Learning (RL). It is well known that RL algorithms such as Value Iteration and Q-Learning provide a sufficient framework for autonomous unsupervised learning, albeit its unsatisfactory performance in multi-agent models.

This chapter focuses on the use of RL algorithms in CRNs that allow the CR to learn the ongoing PU transmissions and determine the appropriate interval for SU transmission to take place. We shall consider the PU transmissions as a two-state Markov model and the respective SU transmissions given the PU state, as two-state Markov models.

4.2 Need for learning in Cognitive Radios and my Contribution

The ability of learning is an indispensable component of intelligent behaviour, hence an integral part of CRNs. There are several learning problems that are specific to CR applications due to the nature of the CRs and the operating RF environments. In this chapter, we shall focus on the noisy observations and sensing errors, CRs obtain partial information of their state variables. Hence, the learning in partially observable environments are to be addressed. Autonomous learning methods are desired for CRs to learn in unknown RF environments, as in contrast with wireless users, CR should operate in any available spectral bands, at any time and location.

4.3 Literature Review

As previous chapters have focused heavily on the literature of RL in CRNs, this section chooses to focus on some literature that is very relevant to the work contributed in this chapter. The section provides brief descriptions of similar work done and identifies the difference and gaps between our contribution and current literature.

Past research in the cognitive area consisting of RL is not heavily focused. However, there have been few papers that have considered this area of research. One paper that strongly focuses on this area is the paper by Berthold [76]. This paper focuses mainly on the detection of the PU system's allocation and usage of spectral resources. The detection is performed frequently and on all sub-channels of the PU transmission spectrum by using FFT(Fast Fourier Transform). Additionally, it focuses on the allocation and management of the frequency bands as well. One paper produced by Lo focuses on RL in co-operative cognitive Ad-Hoc networks[41]. This paper focuses on solving the overhead problem by introducing RL to the co-operative cognitive network, whose major focus is on the ability of the SU to minimize sensing delay and find an optimal set of co-operating neighbours. There are few other papers considering RL on wireless networks such as the ones mentioned [37][33]. Yau presented a paper on the application of context awareness in wireless networks. Okansen also, has presented a paper on RL in energy efficient networks using a sensing policy optimization methodology. However, our paper proposes the application of RL in the periodic sensing of the CR to detect PUs in a periodic manner and determine the presence of their transmission, before allowing SU transmission to take place. This allows us to consider the wastage of slots and the behaviour of interference between the PU and SU in the model.

4.4 System and Network Model

The scenario considered for this chapter is a single PU and SU system being operated in the same frequency bands. The model is under a centralized scheme. The system and network model is described by means of Markov models, and for further references one could refer to [62]. This section provides a methodology of the inclusion of RL in the scenario of the SU transmission to optimize the Markov model.

4.4.1 Periodic Detection of CR

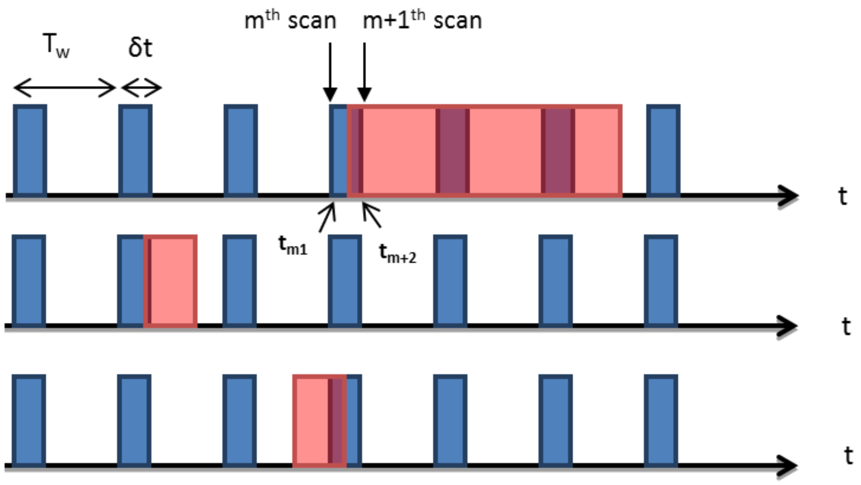


FIGURE 4.1: Different scenarios of PU detection by CR, where the CR is represented as blue boxes and PU transmission as red.

Figure 4.1 shows the scanning period of the CR to detect PU system transmissions. The CR node performs a detection task in order to detect whether the PU is currently transmitting at the given frequency band, and acquires information stating the current status of the PU. Without the use of RL, SU transmission tends to take place immediately when it decides the channel is vacant, hence causing more interference to PU transmissions. The inclusion of RL allows the SU to transmit only after successive CR scans of the absence of PU transmission in the channel. The CR node has a sensing period of T_w seconds and a fixed sensing duration δt seconds. The CR detects the presence of PU transmission only during δt. Therefore, the CR node senses the spectrum periodically for PU transmission throughout the frequency band and detects the presence of PU transmission between the duration of $(t_0 + tm_1)$ and $(t_0 + tm_2)$ for some $(t_0 \epsilon R)$ for the m^{th} scanning iteration [62][77], for all $m \epsilon N$.

A decision threshold Γ is introduced for the CR to determine whether SU should transmit or not depending on the presence or absence of PU transmission after a specified number of CR scans. After each CR scan, the cost function C is compared to the threshold value as shown:

- SU does not transmit : $\Gamma \leq C$
- SU transmits : $\Gamma \geq C$.

4.4.2 Temporal Behaviour of the PU system

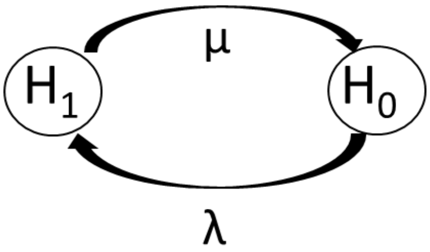


FIGURE 4.2: PU System Markov Model.

Figure 4.2 shows the interaction between the two states of the PU by means of a Markov representation. The states are represented in the following:

- H_1 - PU transmits
- H_0 - PU does not transmit

The two states or hypotheses are linked by factors such as λ (arrival rate of PU) and μ (termination rate of the PU). The temporal behaviour of the PU transmission which comprises of the arrival and termination rates, is modelled as a random Poisson Exponential Process[62]. For further references on the temporal model and the Poisson Exponential Model, one should refer to [39][65][38].

4.5 Secondary User model without Reinforcement Learning

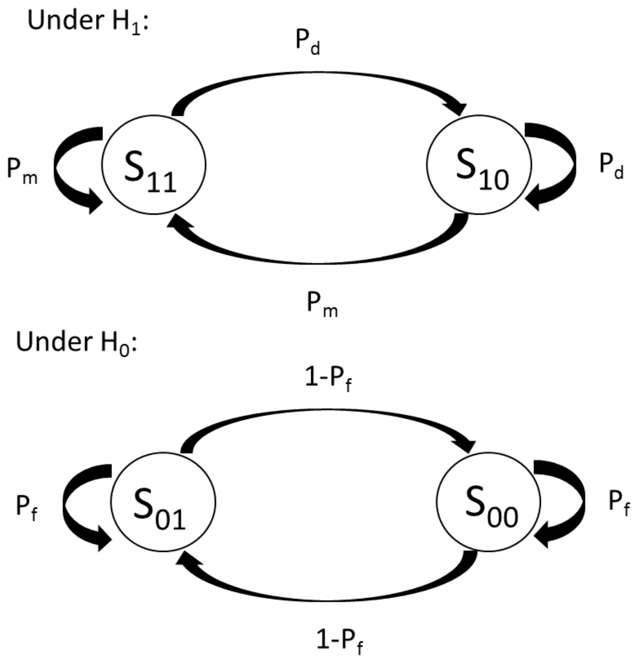


FIGURE 4.3: SU system represented in a Markov Model for H_1 & H_0 - Absence of RL.

From the stated PU model in figure 4.2, this chapter proposes the SU transmission Markov model. There are four states in the SU transmission model:

- S_{00} - SU does not transmit | H_0
- S_{01} - SU transmits | H_0
- S_{10} - SU does not transmit | H_1
- S_{11} - SU transmits | H_1

where the respective state-action models depend on the PU states H_0 and H_1. Figure 4.3 shows that the transition from state S_{11} to S_{10} with a detection probability P_d of the PU, while it loops back to S_{11} with a miss detection probability P_m of the PU, for the hypothesis H_1. The vice versa applies for the state S_{10} under the PU hypothesis H_1. When the hypothesis H_0 is taken into consideration, probability of false alarm P_f comes to play. The state S_{01} loops back with a probability P_f and moves to the state S_{00} with a transition probability $1 - P_f$. The vice versa is applicable for the state S_{01}

under the hypothesis H_0 as shown in figure 4.3. The absence of RL leads to the fact that the system does not learn from experience and prior information, but rather on the current status of the PU and SU transmission only. This leads to a very low optimization of the Markov chain model. Hence, a modified SU Markov model is proposed to be incorporated with the inclusion of RL strategy.

4.6 Reinforcement Learning Algorithm

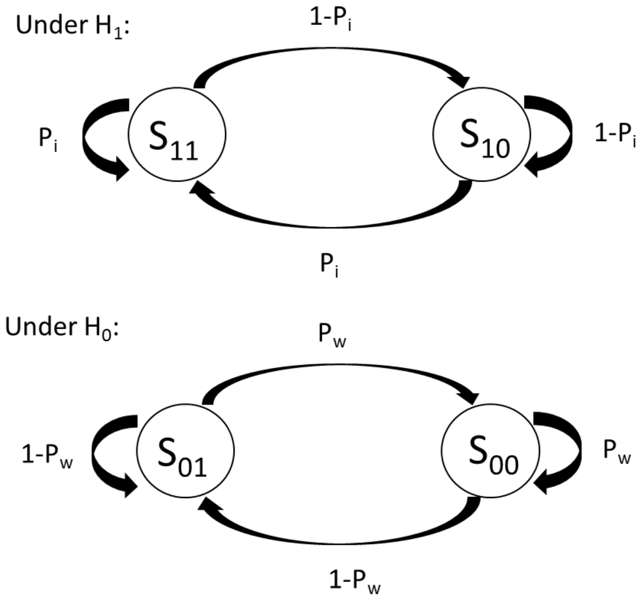


FIGURE 4.4: SU system represented in a Markov Model for H1 & H0 - Presence of RL.

Figure 4.4 shows the modified SU Markov model which is incorporated with RL strategy. The modified SU transmission model is represented for the two PU scenarios H_1 and H_0 respectively., The inclusion of RL strategy results in the use of prior information and the current status of both PU and SU transmissions. The modified Markov model follows the same transitions as that of the traditional SU model, but with different transition probabilities. They are probability of interference P_i and probability of wastage P_w that are derived from the P_m and P_f expressions, respectively.

In order to comprehend, slots are defined as the time between successive CR scans or represented as T_w. The interference P_i between the PU and SU is the measure of time or number of slots, when there is PU and SU transmission taking place simultaneously.

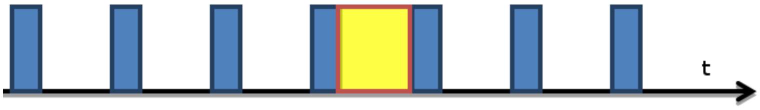


FIGURE 4.5: Possibility of PU-SU interference - Yellow box represented as SU transmission.

As to the model, such interference occurs, when the SU transmission begins at the end of CR sensing duration δ_t, and after some time prior to the next sensing period of the CR, PU transmission begins. Figure 4.5 shows the possibility of interference P_i between successive CR scans. To define interference, one must understand the terms Probability of Detection P_d and Probability of miss detection P_m which are shown as in the following.

$$P_d = \Pr(\text{PU Detection in } T_w | H_1) \tag{4.1}$$

This leads to the expression of interference which is expressed as:

$$P_i = \Pr(S_{11} \text{ in } T_w | H_1) \subseteqq P_m.$$

We define the interference probability P_i as the probability of both the SU and PU transmitting at a given time interval period of T_w, which is a consequence of the miss detection probability P_m. Thus, we express P_i as the following:

$$P_i = \sum_{n}^{N_t} \frac{\Pr(S_{11}|H_1)}{N_t}, \tag{4.2}$$

where N_t is the total number of time slots. Therefore, if $\Gamma = 1$, one concludes to the following regarding P_i:

$$P_i = P_m \tag{4.3}$$

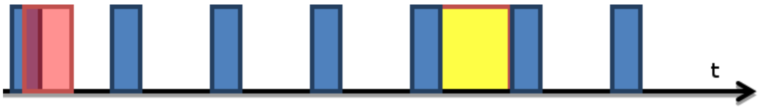


FIGURE 4.6: Wastage of Slots - Yellow box represented as SU transmission, Red box being PU transmission.

The wastage of slots P_w is the measure of the number of slots available for the SU transmission to occur, but SU does not transmit even though there is no detection of PU transmission in the previous CR scan. This is due to the cost being higher than the decision threshold Γ which described later in this section. Therefore, the CR does not permit the SU transmission to begin. Figure 4.6 shows how wastage of slots occur for successive CR scans.

Therefore, wastage P_w is defined as:

$$P_w = \Pr(S_{00} \text{ in } T_w | H_0) \subsetneqq P_f$$

We define the wastage probability P_w as the probability of both the SU and PU not transmitting at a given time interval period of T_w, which is a consequence of the false alarm probability P_f. This leads to the following expression of P_w:

$$P_w = \sum_{n}^{N_t} \frac{\Pr(S_{00}|H_0)}{N_t}, \tag{4.4}$$

where N_t is the total number of time slots. Therefore, if $\Gamma = 1$, P_w can be stated as in the following:

$$P_w = P_f = 0 \tag{4.5}$$

The agent(SU) decides upon an action to move to the next transition based on the PU states H_1 and H_0. The transitions from one state to the other in the SU model is acquired through the cost function $C(s, a)$ which allows the RL strategy to learn the environment (PU states). The respective costs $C(s, a)$ are computed as the following:

$$\begin{aligned} &H_1 : C(s,a) = 1 \\ &H_0 : C(s,a) = \frac{1}{2}(1 + \cos(\frac{\pi\tau}{\beta}(\alpha - \frac{1-\beta}{2}))), \end{aligned} \tag{4.6}$$

where τ is the time span or number of adjacent time intervals in which H_0 lies. The cost function $C(s, a)$ to the agent(SU) from the environment varies based upon the detection or non-detection of PU transmission. The cost is the prediction of the likelihood of the presence of PU transmission in the successive scan. If the presence of PU transmission is detected by the CR within its sensing duration δ_t, then the cost is stated as 1. If there is an absence of PU transmission, during the CR detection, the cost $C(s,a)$ is then analysed by the raised cosine function. The raised cosine function includes a learning ratio α, also known as the discount factor, for convergence purposes. The learning factor α is discounted after each CR period T_w in the absence of PU transmission. This process is then restarted if PU transmission is detected. The cost $C(s,a)$ based on the raised cosine function is expressed as in the following: The learning ratio α must satisfy the criteria shown:

$$\frac{1-\beta}{2} < \alpha \leq \frac{1+\beta}{2} \tag{4.7}$$

where $\beta\epsilon(0,1)$.

The RL algorithm is represented by means of a flow chart in Figure 4.7.

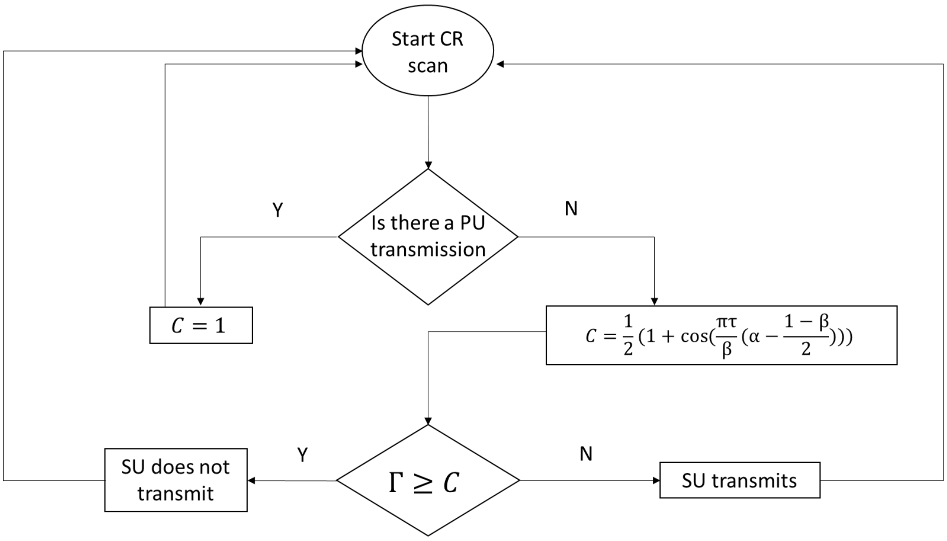

Start CR scan
Is there a PU transmission
Y
N
$C = 1$
$C = \frac{1}{2}(1 + \cos(\frac{\pi\tau}{\beta}(\alpha - \frac{1-\beta}{2})))$
$\Gamma \geq C$
Y
N
SU does not transmit
SU transmits

FIGURE 4.7: RL Algorithm for Prediction of PU Transmission.

4.7 Simulation Results & Performance Analysis

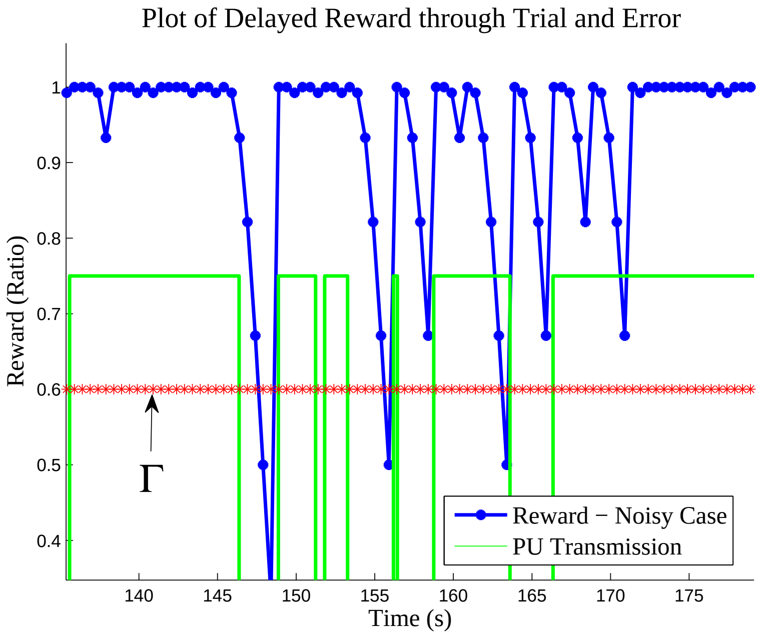


FIGURE 4.8: Reward function in prior dependence of PU Transmission for noisy case.

Figure **??** shows the overall cost or cost acquired from the algorithm for $N = 10000$ iteration scans. The output is the reward or cost to the agent (CR) which is defined as the probability of the presence of the PU transmission in the consequent CR scan. The cost state of $r = 1$ does not decrement immediately due to the absence of the PU transmission, but decreases rapidly after a few periods of CR detection when the PU transmission is absent. However, the immediate presence of the PU transmission places the cost state back to 1. The decremented slope is acquired from the cost function C.

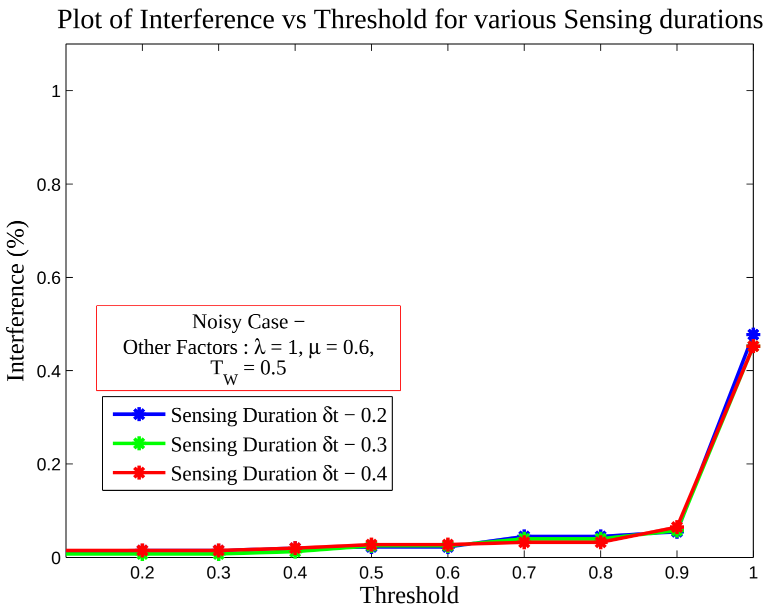


FIGURE 4.9: Plot of Interference vs. Threshold for various Sensing Durations.

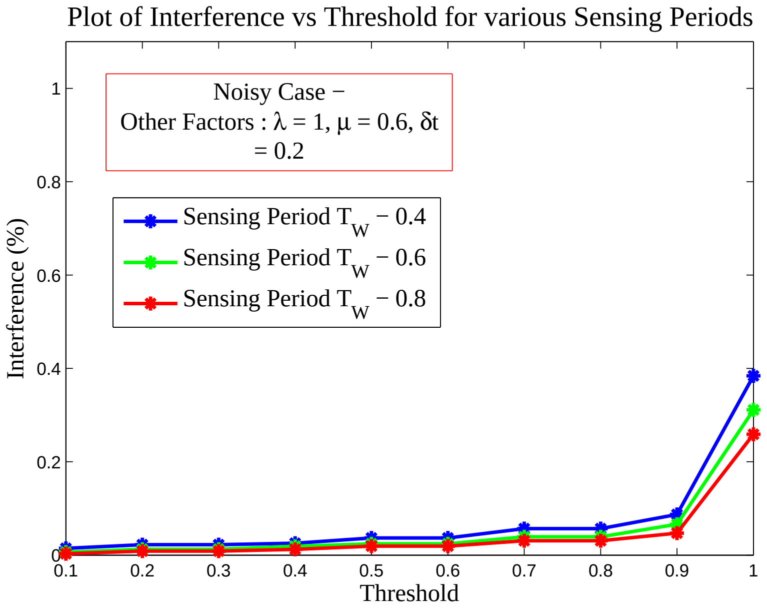


FIGURE 4.10: Plot of Interference vs. Threshold for various Sensing Periods.

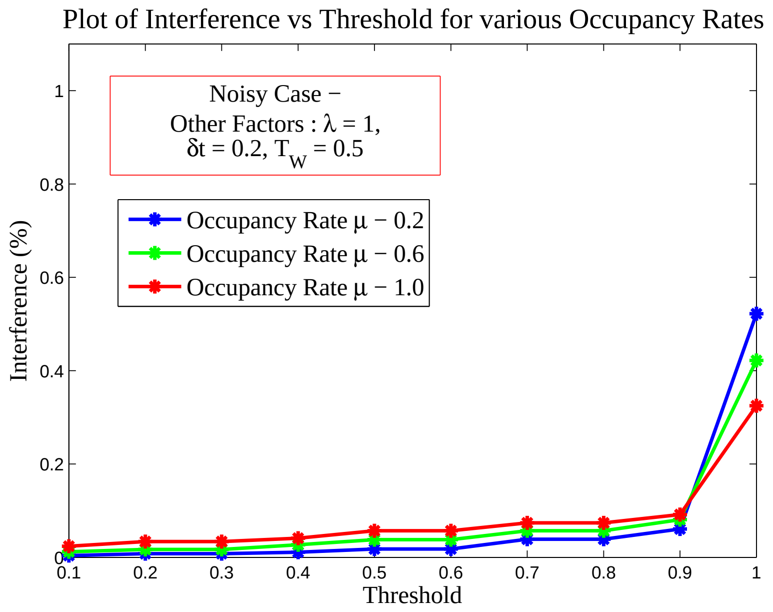


FIGURE 4.11: Plot of Interference vs. Threshold for various Occupancy Rates.

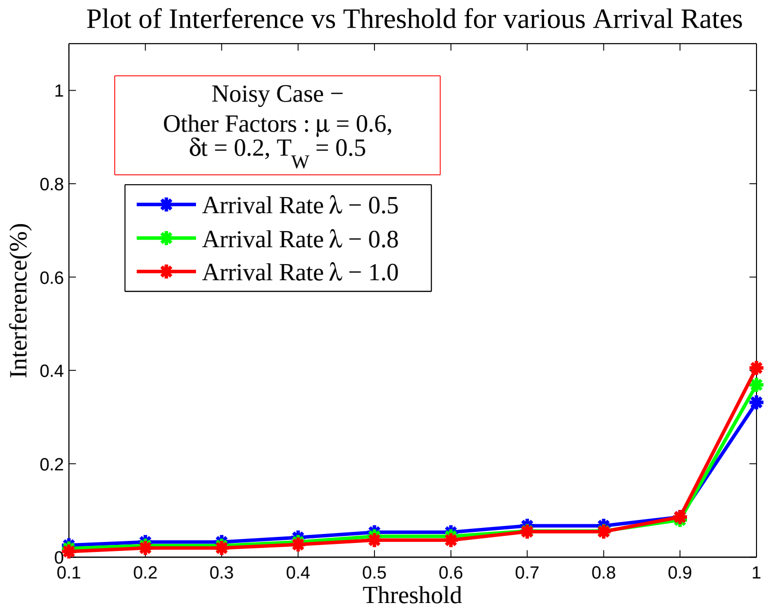


FIGURE 4.12: Plot of Interference vs. Threshold for various Arrival Rates.

Figures 4.9, 4.10, 4.11 and 4.12 exhibit the behaviour of the PU-SU interference in a percentage manner for various thresholds with one varying variable such as δ_t (Sensing Duration), T_w (Sensing Period), λ (PU arrival rates) and μ (PU Occupancy rates) respectively, while the rest of the variables are fixed respectively. This study shows us how the interference behaviour works depending on one variable such as δ_t (Sensing Duration). This plot varies accordingly since it is dependent on various factors such as PU transmission arrival times, CR sensing durations, CR periods and learning ratio of the RL algorithm. One can observe that the plot based on sensing periods, states that when the period T_w increases, the interference time decreases. This is due to the fact that the number of CR sensing blocks decrease due to the increase in the period T_w, thus resulting in such behaviour. From observations, we can notice the zero difference in interference values in specific threshold ranges. This is mainly due to the rapid decline of the cost function after a few periods of absence of PU transmission, thus, resulting in the same number of interference slots between these threshold ranges.

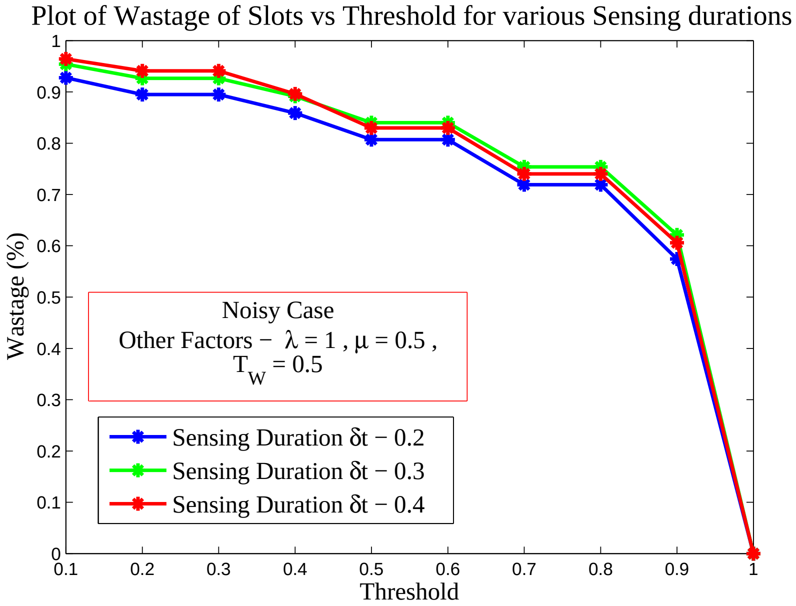


FIGURE 4.13: Plot of Wastage of Slots vs. Threshold for various Sensing Durations.

Figures 4.13, 4.14, 4.15 and 4.16 demonstrate the wastage of time slots in comparison threshold when the other variables are fixed except for one such as δ_t (Sensing Duration), T_w (Sensing Period), λ (PU arrival rates) and μ (PU Occupancy rates) respectively. The wastage of slots gradually decreases with increasing threshold values. However, one can observe a constant value in specific threshold ranges. We also noticed the constant interference value for such ranges as stated in figure 4.8. The cause of this constant value is the rapid descend of the cost value. The gradient of the slope between these ranges in the cost plot (Figure 4.9) is very high.

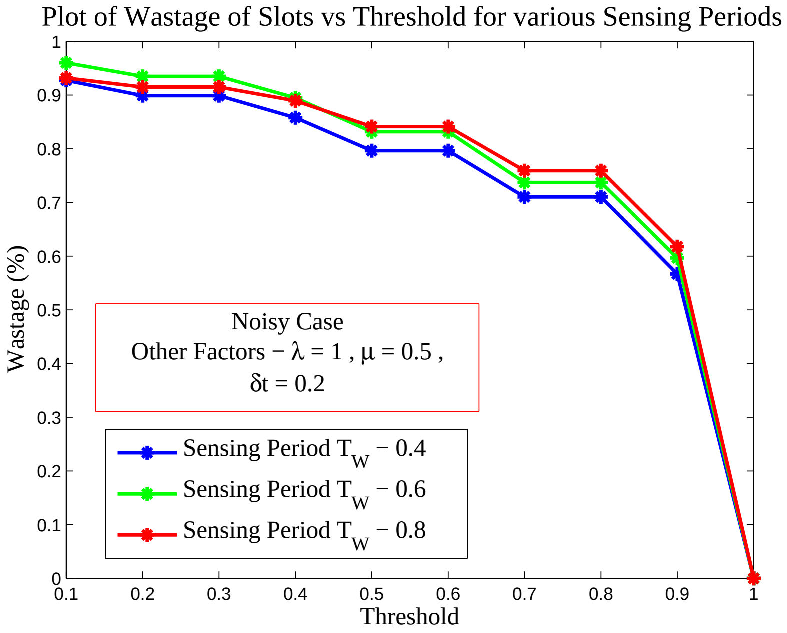


FIGURE 4.14: Plot of Wastage of Slots vs. Threshold for various Sensing Periods.

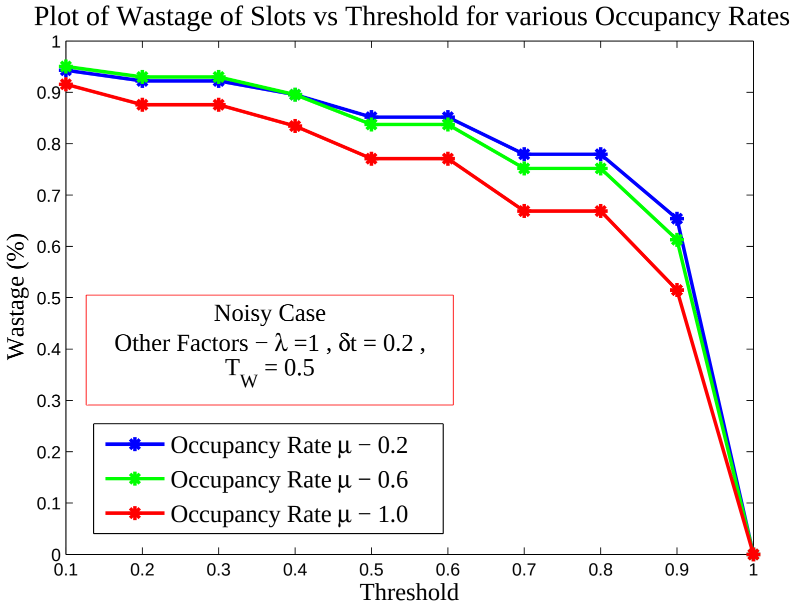


FIGURE 4.15: Plot of Wastage vs. Threshold for various Occupancy Rates.

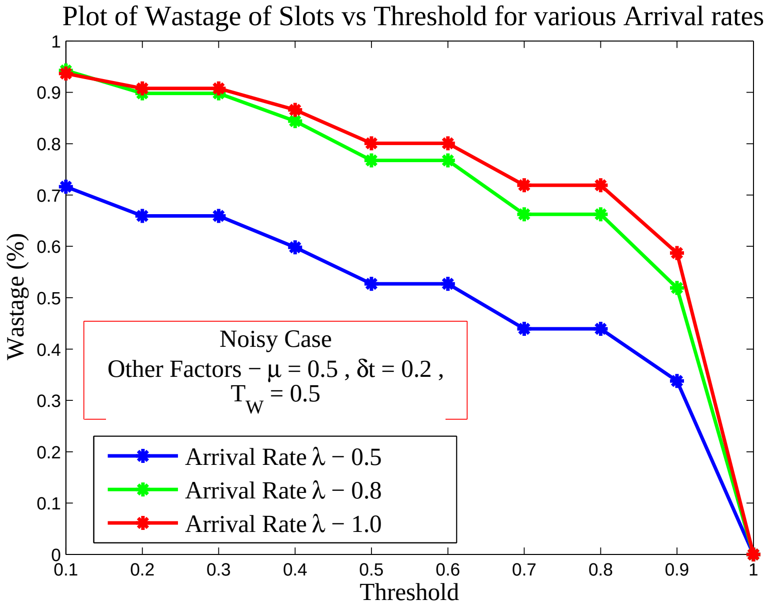


FIGURE 4.16: Plot of Wastage of Slots vs. Threshold for various Arrival Rates.

4.8 Summary of Contribution & Conclusion

The contribution of this chapter lies in the modelling of the interaction between PU and SU transmissions as Markov models, followed with a RL method that improves the overall utilization of the SU in terms of channel access, while minimizing the interference to the PU, and minimizing the wastage of time slots when the channel was available.

This chapter concludes with the presentation of the performance and methodology of detecting PU transmission in the spectrum by CRs using RL techniques. The CRs use periodic scanning to consider the temporal behaviour of the PU and provide the license from the PU to allow SU transmission in the absence of PU transmission in the spectrum. This is done by means of introducing context awareness and intelligence to the CR network. This allows the CR to observe, learn, and respond in an efficient and appropriate manner with respect to its complex dynamic environment without adhering to a set of predefined rules. This capability is of paramount importance in CR networks. The results show that RL has significantly improved the performance of PU detection and SU handling in a CRN, reduced the PU-SU interference time as well as the wastage of time slots for SU transmission. The PU-SU interference time depends on the detection threshold, sensing period, time-bandwidth product and also the temporal behaviour of the PU transmission in the spectrum. The waste of time slots for SU transmission does not depend on the temporal behaviour but rather on the detection threshold mainly and on the cost function acquired from the RL algorithm. This chapter presents closed form expressions for the cost function in the prediction of the next state of PU transmission in the cognitive sensing period, PU-SU interference time and the wastage of slots.

Chapter 5

Resilience in Secondary User Transmissions using Game Theory

CRs promise intelligence and autonomous functionality in terms of learning and adapting to the environment. In a CRN, SUs are allowed to access licensed frequency bands through means of a non-interference basis to the legacy users also termed as PU. The limited availability of spectral resources and absence of secure centralized access control results in the SUs being selfish. An exacerbated issue is when the SUs are intentionally inflicting harm upon other competing SUs. Since CRs are capable of adapting to any environment and alter communication channels when required, a significant threat is determined to be present. Game theoretic models have been considered as a viable approach to model the resilience of the physical link itself. DSA was formulated as a potential game for different model environments. Resilience by game theory lies in the access scheme, frequency hopping pattern, power allocation, and selection of control and data channels at a given time. An equilibrium is approached by iterative updates, hence a game framework was applied for selfish or jamming users to minimize their impact on other non-selfish SUs.

5.1 Background

Security, in general as been researched upon the data link layer with regards to authentication and encryption. This works efficiently for data traversing along a wireless network, but does not satisfy the demand in the fundamental operation of the physical medium itself.

CRNs are quite vulnerable to jamming attacks, due to the fact that SUs do not own the spectrum and hence they are susceptible to adversarial effects. Furthermore, highly dynamic spectral access makes security measures difficult to implement, especially when the SUs are capable of intelligently making spectral access decisions and hence, inflict further harm. Security threats are unavoidable,thus incorporating security facilities are challenging in CRNs due to its open nature. Therefore, ensuring security is important to the successful deployment of CRNs. Thus, more care and research is required in order to provide efficient security and resilience mechanisms in CRNs. Better security mechanisms ensure the trustworthiness of spectrum sensing especially for the SUs.

The detection problems arise when operating in a hostile environment as it is possible to mimic incumbent signal characteristics and pretend to be the PU. This is called as Primary User Emulation. In such cases, integrating legitimate transmitters for PUs and SUs in spectrum sensing would improve the trustworthiness of the detection process. The availability of a wide range of authorized and unauthorized signals are possible using cheap consumer devices thus leading to easy access to the creation of Denial-Of-Service (DOS) attacks which affect critical applications such as traffic control and health care. Therefore, to prevent or mitigate such events from happening in CRNs, the FCC regulations prepared the incorporation of awareness of such attacks.

In this chapter, we primarily focus on the physical (PHY) layer of resilience and security and provide an analysis of the current threats to the CR and ways to mitigate the effects of such threats. The threats in CRN are the following [20]:

- sensory input statistics can be altered
- faulty sensory input leads to belief manipulation
- manipulated individual statistics and belief may be distributed through CRNs
- behaviour algorithms based on manipulated beliefs result in sub-optimal performance or malicious behaviour

These effects can be mitigated by the following [20]:

- assume sensory input is “noisy” and subject to manipulation

- programmed to validate learned beliefs
- compare and validate learned beliefs with other devices in the CRN
- expire learned beliefs to prevent long-term attacks
- attempts to perform learning in known and accepted environments

The security aspects are researched upon the general scenario considering the secondary infrastructure and the intentional or unintentional attacks from surrounding RF equipment. There are many concerns in the security and resilience issues of the PHY layer, under the basis of CRN, where there are various types of attacks on the secondary infrastructure. To counter such issues, security concepts are considered in the cases of disaster scenarios and major temporary events. Security in the communication system is widely researched upon, especially in the network, transport and application layers. However, our focus of research is based on the physical layer resilience as a type of security in situations applicable to the ABSOLUTE architecture and scenarios. There are many concerns in the security issues of the PHY layer, under the basis of CRN, where there are various types of attacks on the secondary infrastructure.

5.2 Types of Jamming Attacks in the PHY layer in CRNs

To comprehend the level of jamming attacks, one must understand the structure of the CR. The CR comprises of the SDR and CR engine. The CR engine consists of the knowledge base, policy radio and learning radio components. The policy part consists of the policies that are executed by the CR and the learning part consists of the techniques used for investigation, exploration and exploitation of data. The knowledge base stores the data.

The CR consists of three major jamming attack types in the PHY layer. Each attack is designed to be a threat to the different component of the CR engine [30, 4, 80, 54]. They are:

- Sensory Manipulation attack - The attacker spoofs false sensory information, causing the CR to select a least optimal configuration. Radio sensors acquire digitized RF and extract useful data. However, an attacker can manipulate such RF signals and energy, one can cause false statistics to appear in its learning base. This is termed as a Policy Radio Threat.

- Belief Manipulation attack - An attacker forces the CR to learn the way it requires for the attack to take effect. This makes the attack a long term behavioural problem, thus the attack being more powerful. This is considered as a Learning Radio Threat.
- CR Viruses – This occurs in CRNs only as illustrated in 5.1, in a multi radio sensing environment. This is where one radio induces that same state or learning process from the attacker to adjacent radios, thus resulting in propagation to all radios in the network. Major destruction occurs when the network is non-cooperative.

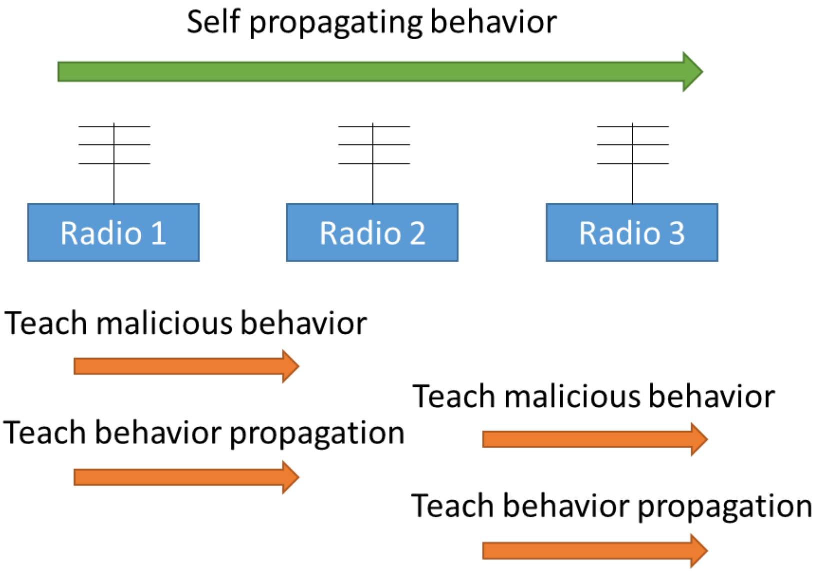

Self propagating behavior
Radio 1
Radio 2
Radio 3
Teach malicious behavior
Teach behavior propagation
Teach malicious behavior
Teach behavior propagation

FIGURE 5.1: CRN Virus[20].

In the ABSOLUTE case, the purpose of the attackers is classified as injecting interference in the channels available for the CRN, thus, the CRN selects a less optimal configuration for channel access and thus decreasing the overall performance of the secondary CRN. This type of attack could be said to fall under the class of sensory or belief manipulation attack for a long term effect, where the unauthorized users inject interference and thus forcing the SU to a least optimal configuration.

5.3 Need for Resilience and my Contribution

Significant focus has been placed on the the Third Generation Partnership Project (3GPP), where the currently introduced generation of cellular technology, known as LTE and its advanced version (LTE-A), which has approached the economic market potential, in terms of monetary and technological gain. The newer releases of LTE, specifically Release 12 and 13 [11] have introduced a revolutionary form of communication, termed as Device-to-Device (D2D)[26][35]. It promises communication features such as short range proximity services, offload traffic and performance of efficient spectrum utilization. One possible example of a D2D communication protocol was presented for example in [91, 53, 28]. The destructive effects of the unauthorized/ malicious users or jammers are measured at the location of the reference D2D receiver. All devices are assumed to be hand-held battery operated terminals using LTE-A cellular technology with limited battery capabilities. Considering all the remarkable technological improvements of miniaturized electronic equipments, we do not restrain ourselves to consider only cellular devices. Since all the devices are battery operated they have to make a judicious use of their power budget in order to maximize their utilities. From the perspective of the authorized user of interest, this means maximizing the achievable transmission rate between source and destination, whereas for the jammer this implies destroying the legitimate communication link. Such a new development brings forth strong requirements of coordination between the legitimate D2D users such as handing over control of radio access to local UEs, that in turn, require the prevention of destructive interference due to the fact that local transmissions potentially interfere with one another. A common threat is the possibility of available channels in a geographical area being interfered, which in this chapter is represented in the form of malicious users performing jamming attacks, which is possible with the increased programmability and computational processing power of the UE terminals[5].

The aim of this chapter is to demonstrate a theoretical framework and provide novel performance evaluation tools that are able to quantify the impact of jamming based on a variety of factors: number of channels jammed, power allocated per channel, distance between the jamming and legitimate users, etc. It provides an illustration of a basis of resilience for the emerging D2D paradigm against such possible adversaries.

The case considers a legitimate transmitter/receiver pair subject to jamming attacks intentionally operated via the D2D communication mode. The transmitter is termed hereafter "legitimate user" (LU) , as it is able to access only the channels leased out locally by the primary communication system (i.e., cellular network). The malicious user/jammer (MU) makes multiple attempts to interfere with the physical resource blocks (PRBs) which the D2D transmitter exploits in the LTE radio frame. From henceforth, we denote the PRBs as channels for simplicity. Hence, we formulate this situation of adversity between MU and the transmitting LU, both of similar characteristics such as cognitive capability [29],[34], to detect the available spectrum, computational power and radio propagation characteristics. Furthermore, both LU and MU are capable of making decisions upon how to plan their transmissions over each transmission time intervals (TTIs) in order to maximize their overall spectral gain. At each LTE radio frame, the LU could select a subset of channels at random according to a frequency-time hopping scheme known by the LU receiver[73, 67]. The MU, in turn, tries to hit a certain subset of channels with no prior knowledge of the hopping scheme. Therefore, both LU and MU access the maximum number of channels in order to escape/pursue the transmission of the opponent. However, under a finite power budget, both users need to simultaneously allocate a large enough power per channel to be effective in their spectral access, i.e., to transmit/interfere with enough energy per channel.

In order to deal with such a scenario, this chapter provides a performance analysis based on the fundamental trade-off that arises from combining physical layer considerations and link layer considerations, via the LU's perspective. The chapter offers three major contributions:(1) Probability of Successful Transmission under jamming, (2) Game Theoretic Approach to find the equilibrium of the LU and MU, (3) RL approach to find the best response for each MU strategy and ensure convergence in the least possible time.

5.4 Related Works in Resilience

We give a more detailed context of the mathematical tools and models used to formulate similar jamming problems in this section. This chapter provides an extensive look into this, as this is a relatively novel field of work, hence, we shall view the existing literature and identify the gaps.

The concept of resilience in the channels of the LTE system is a novel approach that provides the authorized system the capability of successfully detecting interfered channels and providing a probabilistic means of successful transmission for channels under jamming effects. There are multiple areas of work done based on radio propagation characteristics, game theory and RL that are integrated to formulate the optimal choice of the LU in terms of power, channel location, channel type selection control against jamming attacks. Areas of RL and game theory have been heavily researched independently and moderately researched dependently. In this section, we give an overall brief review of the work that has been done in the areas of game theory and learning for anti-jamming methods in CRNs. Some key papers are reviewed in this section to state the objectives and provide the basis to the difference in the objective of this paper and the formulated system model.

Under non-cooperative game theory, comes the idea of zero-sum games that involves two different players (LU and MU), where the game is strictly competitive such that the one player's gain is balanced by the loss in the utility of the other player. The total sum of the utility of both players results in a zero value. Zero-sum games are non-cooperative and hence usually solved with the NE. Colonel Blotto games fall under zero-sum games which consists of two players where the players are assigned to distribute limited resources over several entities [19]. The paper referred in [84], presents the concept of non-cooperative spectrum allocation problem in CRNs. The multi Nash Equilibrium (NE) problem of the non-cooperative game based spectrum sharing in CRNs is dealt with the variation of the utility function of the CR users and is considered to judge the stability after an intense iterative process, leading to the design of an improved non-cooperative spectrum allocation algorithm. The paper referred in [86] covers the topic of anti-jamming games where the basis lies in the non-cooperative interaction between the SU and MU. The author analyses the interaction between the players that leads to the derivation of a channel hopping defence strategy. This hopping strategy is implemented using the MDP. The proposed idea consists of two learning schemes for the SU to gain knowledge of the attacker to handle cases without complete knowledge. The two ideas are MDP based channel hopping and MDP based power allocation for the multi-channel cognitive networks. The author deals mainly with the concept of intentional jamming attacks and the methodology to counter them. One other paper as referred in [79], considers the Colonel Blotto game leading to a NE, where its aim is to illustrate the concept of the SU using a minimax strategy where it adopts to minimize the worst case damage to its network by the MU.

In the area of RL application on jamming games, we focus on the work done and objectives in the learning process of the authorized user in the following papers. Firstly, the work done in [88], shows a Competing Mobile Network Game played as a stochastic play between the cognitive networks in order to dominate the open spectrum access. A coalition scheme is considered to integrate anti-jamming and jamming methods to the stochastic model, followed by Q-learning, stating that Q learning is more suitable for an aggressive environment and provides the best solution under distributed mobile ad-hoc networking scenarios in which the availability of centralized control is not probable. Secondly, the paper in [28], investigates the security mechanism when SUs face jamming attacks, by developing a stochastic game framework as a means of an anti-jamming defense. The SUs observe the spectrum availability, quality of channels accessed and attacker's strategy from the output acquired from the jammed channels, thus, adopting a strategy based on the number of channels used for control and data message, giving a higher priority for control messages over data. The reward is updated by Q-learning mechanism and the overall objective is to reduce the number of control channels transmitted on the jammed ones and increase the number of data channels, vice versa in the unjammed channels. This is due to the fact that the reward of losing data channels is significantly lower than the control channels in the communication network. Thirdly, the paper [69] demonstrates the use of the Q-learning with value iteration (QV) and the state-action-reward-state action (SARSA) RL algorithms instead of Minimax Q-learning, although it achieves an optimal solution in anti-jamming strategies. The author indicates that QV learning performs even better than SARSA, as in QV both Q- as well as V- values of the game are updated. Additionally, the paper [69] proposed the novel concept of introducing RL as a means of solution to counter against jamming attacks in CRNs. The author considers three RL techniques to determine the most effective RL strategy to counter against MUs: QV learning, SARSA and minimax Q-learning, based on stochastic zero-sum game theory. The author describes the SU as one with capability to decide on the two actions: staying in the same channel or hopping, by different RL techniques. The paper addressed in [71] describes the idea of anti-jamming strategies for channel access in cognitive networks where the SU tries to access idle channels that are not under attack, by means of greedy search, $\epsilon-$first and random selection.

The background literature provides detailed works in the respective areas of propagation, game theory and RL individually. There are very few substantial works that relate to such areas in a combinatorial manner and dwell into the concept of jamming in CRNs. Such works are shown in the following papers [36, 42, 43].

5.5 System Model and Analysis

The scenario consists in a single user model comprising of one LU transmitter (Tx)/receiver (Rx) pair that are scheduled to transmit over the available slots in the geographical area that are not being utilized by the PU, via D2D communication mode. The receiver of interest is assumed to be in the centre for simplicity. The radial distance between the LU Tx and Rx, and MU and Rx is varied. Figure 5.2 illustrates the model described prior. Referring to the structure of LTE physical layer [67], we define a slot as one TTI (i.e. 1 ms) over time and one PRB in frequency (180 kHz). The way LU transmitter and receiver agree on the number of slots to use, and their location within the time-frequency structure of LTE, is a protocol specific choice that is out of scope of this present work. A single MU is introduced whose purpose is to disrupt the communication between the LU Tx/Rx.

5.5.1 System Model Assumptions and Parameters

The model takes into account the probability of jamming interference between the LU and MU, when the policy adopted here consists of a channel access based on a random selection of the number of slots to use and their locations within the time-frequency LTE frame. For completeness, we also remind that an LTE radio frame lasts 10 ms, or in other words it comprises 10 TTIs. For simplicity, we make the following assumptions:

- The LU Rx is located in the centre of the geographical area.
- The total transmit power of both LU and MU is fixed and finite (to fulfill the characteristics of a handheld device). The power per slot of the LU and MU is uniformly distributed. Hence, it leads to say that the LU and MU objectives depend on the total transmission power, the number of slots used and the power per slot. The total number of slots used by the legitimate pair depends on the specific application: for a video service which requires minimum guaranteed bit rate, the slots should be reserved consecutively.
- All radio transmissions are assumed affected by small-scale Rayleigh flat fading and distance-dependent path-loss.

The model is described in terms of a LU's perspective based on the following factors: distance between LU Tx/Rx, path loss exponent α, fading characteristics such as Rayleigh and probability of interference P_i between LU and MU [56].

Figure 5.2 illustrates the three boundaries of operation where the LU and MU attempt to access the spectrum, provided that the LU receiver is at the centre for simplicity. Further the distance from the LU Rx, greater the decrease in the impact of the MU, thus benefiting the LU. The selected numerical parameters are shown in the Table 5.1. Parameters are chosen based on the LTE characteristics for a $W = 1.4$ MHz system bandwidth. The parameters α and β are respectively selected to model any propagation environment and minimum required Signal-to-Interference-plus-Noise-Ratio (SINR) to decide whether the signal is successfully received.

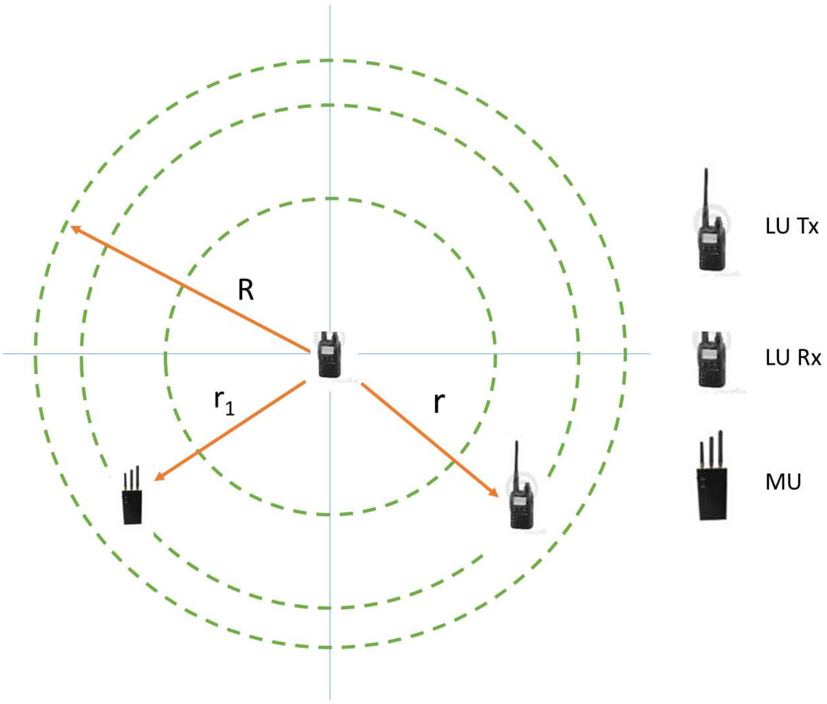


FIGURE 5.2: Three logical boundaries define the operation of the system and the impact of the LU transmitter's and MU strategies.

TABLE 5.1: Physical layer parameters.

Parameter	**Meaning**	**Value**
P_{T}	RF transmit power of LU	23 dBm
P_{J}	RF transmit power of MU	23 dBm
W	Bandwidth	1.4 MHz
β	Minimum required threshold value for SINR or SNR	3÷10 dB
α	Path-loss exp.	2.1
N_0	One-sided noise power spectral density	-174 dBm/Hz
D	Maximum coverage radius	500 km
M	Total number of Slots	60
m_1	No. of Slots accessed by LU	1 ÷ 60
m_2	No. of Slots accessed by MU	1÷ 60
r_1	Distance between LU Tx and LU Rx	1÷D
r	Distance between MU and LU Rx	1÷D

5.5.2 Collision Probability

Upon the assumption of a random selection policy of slots used by the LU pair (i.e. a non-guaranteed bit rate application is selected), the probability (P_i) that the MU can jam (i.e. interfere) the LU transmission implies that there are i jammed slots in which the MU collide with the LU for given m_1 and m_2. Clearly, if $m_1 \geq M - m_2$, there are collisions on at least $m_1 + m_2 - M$ channels. On the other hand, if the condition $m_1 > M - m_2$ holds, there is a non-zero probability of no collisions. Hence, this can be written as $i \geq \max(0, m_1 + m_2 - M)$.

Furthermore, $P_i = 0$, $\forall\, i > \min(m_1, m_2)$, which leads to the expression (5.1) that describes the number of jammed slots i between LU and MU to be modeled with a binomial distribution,

$$P_i = \binom{m_1}{i} \left(\frac{m_2}{M}\right)^i \left(1 - \frac{m_2}{M}\right)^{m_1 - i} \quad , \tag{5.1}$$

where the supports are:

$i = \max(0, m_1 + m_2 - M), \ldots, \min(m_1, m_2)$ and $P_i = 0$ elsewhere.

One proceeds to term the probability of i collision slots ranging from $i = \max(0, m_1 + m_2 - M), \ldots, \min(m_1, m_2)$, to be denoted with the set C_{Y}. The complementary of C_{Y} is denoted as C_{N} and is stated as the probability that no collisions occur.

$$C_{\mathrm{Y}} = \sum_{i=\max(0, m_1+m_2-M)}^{\min(m_1, m_2)} P_i \tag{5.2}$$

$$C_{\mathrm{N}} = 1 - C_{\mathrm{Y}}$$

5.5.3 Probability of Successful Transmission

Relying now on the analysis described in section 5.5.2, there are two possible cases for the slots accessed by the legitimate pair:

- **Case 1** The slots accessed by the LU is affected by background noise (with P_{N} denoting the in-band noise power) and fading, where P_{N} can be calculated from N_0 and W.

- **Case 2** The slot accessed by the LU is affected by interference from MU (with P_{J} denoting the power of the MU), background noise and fading (P_{N} + P_{J}).

Case 1 allows us to derive the probability of successful transmission per slot under the absence of jamming, which is denoted as P_{S0}.

$$P_{\mathrm{S0}} = \mathbb{P}(\Lambda \geq \beta),$$

where β is the threshold for detection and Λ is defined as the signal-to-noise ratio (SNR):

$$\Lambda = \frac{P_1}{P_{\mathrm{N}}}$$

and $P_1 = P_{\mathrm{T}} {r_1}^{-\alpha}$ is defined as the received power of LU at LU Rx. Here P_{T} denotes the transmit power of the legitimate transmitter.

This leads to the expression of P_{S0} which is expressed in equation (5.3), taking into account the channel power gain coefficient under the hypothesis of Rayleigh distributed fading affecting the LU denoted as h_1 and its expectation $\mathbb{E}(h_1) = 1$.

$$P_{\mathrm{S0}} = \mathbb{P}\left(h_1 \frac{P_1}{m_1 P_{\mathrm{N}}} \geq \beta\right) = \mathbb{P}\left(h_1 \geq \frac{\beta m_1 P_{\mathrm{N}}}{P_1}\right) = \exp\left(-\frac{\beta m_1}{\Lambda}\right) \tag{5.3}$$

Case 2 allows us to derive the probability of successful transmission per slot under the presence of jamming interference which is denoted as P_{S1},

$$P_{\mathrm{S1}} = \mathbb{P}(\Theta \geq \beta),$$

where Θ denotes the SINR in this case, and $P_2 = P_{\mathrm{J}} {r_2}^{-\alpha}$ is the power received at LU Rx location from the transmission of the MU.

$$\Theta = \frac{P_1}{P_2 + P_{\mathrm{N}}} \, . \tag{5.4}$$

It is now worth to remark that the SINR Θ can be computed only when the MU is assumed to be present in the scenario. Upon this assumption, assuming the MU is at any random distance r away from the LU Rx and the LU Tx is at a fixed distance r_1 from the LU Rx, we define the following equation. For more in-depth insights to this expression we rely on [56]. The main difference with respect to our previous work is that in this chapter the two cases with and without fading are studied in a more comprehensive manner in order to characterize the game theoretic approach presented in subsequent sections of this work.

Let us assume that one slot occupied by LU is interfered by the MU and denote with h the channel power gain coefficient of an MU transmission affected by Rayleigh distributed fading over the same slot.

$$\Theta = \frac{\frac{h_1 P_{\mathrm{T}}}{m_1} r_1^{-\alpha}}{\frac{h P_{\mathrm{J}}}{m_2} r^{-\alpha} + P_{\mathrm{N}}} \,.$$

Based on equation (5.4), we can rewrite the probability of success for the LU receiver as follows

$$P_{\mathrm{S1}}(r) = \mathbb{P}(h_1 \geq \beta \frac{m_1 P_{\mathrm{J}} h r^{-\alpha}}{m_2 P_{\mathrm{T}} r_1^{-\alpha}} + \frac{P_{\mathrm{N}} r_1^{-\alpha} m_1}{P_{\mathrm{T}}}) \,. \tag{5.5}$$

The expression for P_{S1} provided below is developed assuming that both legitimate pair and malicious node are deployed over a 2-dimensional area of radius R. The dependence upon the distance r of the MU from the LU Rx can be removed using the Riemann-Stieltjes integral expression as follows

$$\begin{aligned} P_{\mathrm{S1}} &= \int_{R_1}^{R_2} P_{\mathrm{S1}}(r) dF_{\mathrm{R}}(r) \qquad (5.6) \\ &= \int_{R_1}^{R_2} \exp\left(\frac{-\beta P_{\mathrm{N}} m_1 r_1^{-\alpha}}{P_{\mathrm{T}}}\right) \times \exp\left(-\left(\beta \frac{m_1 P_{\mathrm{J}}}{m_2 P_{\mathrm{T}}} h \left(\frac{r}{r_1}\right)^{-\alpha}\right)\right) \frac{2r}{R^2} dr \end{aligned}$$

where $dF_{\mathrm{R}}(r)$ is the probability that the distance between LU Rx and MU lays within the interval $[r, r + dr]$ and we assume that $R_1 = 0$ and $R_1 = R$. Performing integration by substitution, we use the change of variable $t = (\beta \frac{m_1 P_{\mathrm{J}}}{m_2 P_{\mathrm{LU}}} h) {r_1}^{\alpha} r^{-\alpha}$. We are now in the position to compute the final P_{S1} expression by removing the conditioning upon the channel fading, or in other words we compute $P_{\mathrm{S1}} = \mathbb{E}_h(P_{\mathrm{S1}})$. Hence the integral in (5.6) can be written as:

$$P_{\mathrm{S1}} = \frac{2}{\alpha}\left(\frac{r_1}{R}\right)^2\left(\beta\frac{m_1 P_{\mathrm{J}}}{m_2 P_{\mathrm{T}}}\right)^{\frac{2}{\alpha}} h^{\frac{2}{\alpha}} \exp\left(\frac{-\beta P_{\mathrm{N}} m_1 r_1^{-\alpha}}{P_{\mathrm{T}}}\right)\cdot\left[\int_{t_2}^{\infty} \exp^{-t} t^{\frac{-2}{\alpha}-1} dt - \int_{t_1}^{\infty} \exp^{-t} t^{\frac{-2}{\alpha}-1} dt\right],$$

which after simple manipulation and reminding the definition of the upper incomplete Gamma function $\Gamma(z, x) := \int_x^{\infty} y^z e^{-y} dy$, can be rewritten as follows

$$P_{\mathrm{S1}} = \exp\left(\beta\frac{-P_{\mathrm{N}} m_1 r_1^{-\alpha}}{P_{\mathrm{T}}}\right)\cdot\frac{2}{\alpha}\Big(\beta\frac{m_1 P_{\mathrm{J}}}{m_2 P_{\mathrm{T}}}\Big)^{\frac{2}{\alpha}}\left(\frac{r_1}{R}\right)^2\Gamma\Big(1+\frac{2}{\alpha}\Big)\cdot\Gamma\Big(\frac{-2}{\alpha}, \beta\frac{m_1 P_{\mathrm{J}}}{m_2 P_{\mathrm{T}}} h\left(\frac{r_1}{R}\right)^{\alpha}\Big). \tag{5.7}$$

Although the above expression provides a very general result, hereinafter we consider the specific case in which the powers received at the location of the LU receiver from the LU transmitter and MU are the same. For better explanation, referring to Figure 5.2, it can be intuitively understood that when the LU tx is very close to the LU Rx and the MU located further away, the jamming effect is not degrading significantly the communication of the legitimate pair. Vice versa in the opposite condition the jamming effect is predominant. These are two clear cases which do not require further inspection. Indeed, we decide to analyze the case just mentioned above in which the combined effect of jamming and legitimate transmission is less evident. Therefore, the computation of P_{S1} greatly simplifies and it only remains to remove the conditioning upon the fading affecting the MU transmission (again assumed independent from the fading affecting the legitimate transmitter).

We now denote with $h = h_2$ the power gain coefficient of the Rayleigh distributed fading affecting the MU transmission with first moment $\mathbb{E}(h_2) = 1$. Developing P_{S1} as before we can write:

$$\begin{aligned} P_{\mathrm{S1}} &= \mathbb{P}\left(\frac{h_1 P_1/m_1}{h_2 P_2/m_2 + P_{\mathrm{N}}} \geq \beta\right) \\ &= \mathbb{P}\left(h_1 \geq \frac{m_1}{P_1}\beta\left(h_2\frac{P_2}{m_2} + P_{\mathrm{N}}\right)\right) . \end{aligned} \tag{5.8}$$

With simple manipulations of the expression above and using that $\mathbb{E}(h_2) = 1$ by assumption, we can finally rewrite equation (5.8) as follows

$$P_{\mathrm{S1}} = \exp\left(-\beta \frac{m_1}{P_1}\left(\frac{P_2}{m_2} + P_{\mathrm{N}}\right)\right) . \tag{5.9}$$

We will use this modified version of P_{S1} for simplicity purposes, and in particular in the development of the zero-sum game.

Relying on the results developed until now for two individual cases, we can derive the probability of successful transmission per channel which incorporates for both both and is simply denoted as P_{S}. Hence P_{S} is defined as

$$P_{\mathrm{S}} := P_{\mathrm{S0}} C_{\mathrm{N}} + P_{\mathrm{S1}} C_{\mathrm{Y}} = P_{\mathrm{S0}} - (P_{\mathrm{S0}} - P_{\mathrm{S1}}) C_{\mathrm{Y}} , \tag{5.10}$$

where the last equality holds by means of equation (5.2).

We can finally write the utility function $M_{\mathrm{s}}(m_1, m_2)$ of the system composed of LU pair and MU device as the average number of channels successfully received under malicious attack.

$$M_{\mathrm{s}}(m_1, m_2) = m_1 P_{\mathrm{S}} \tag{5.11}$$

5.5.4 Analysis of Numerical Simulations

Numerical results are provided here in order to evaluate the effect of the MU on the legitimate pair in the specific case in which the powers received by the LU Rx from LU Tx and MU equal each other. All devices are assumed deployed within a circle of radius R, with the LU Rx located in the centre. In particular, we show results for the success probability P_{S} in equation (5.10). The first result is illustrated in Figure 5.3(a). This figure shows the effect on P_{S} of m_1 and m_2. We observe that P_{S} decreases as m_1 increases, due to the fact that $\frac{P_{\mathrm{T}}}{m_1}$ decreases, hence leading to a lower SNR Λ (respectively SINR Θ with interference caused by the MU) per channel, since P_{T} is assumed to be fixed and finite. Furthermore, from Figure 5.3(a) we can notice that the increase in m_2 yields an increase in P_{S}, since Θ becomes larger, due to the decrease of $\frac{P_{\mathrm{J}}}{m_2}$. We can also notice that in the absence of the MU ($m_2 = 0$), we expect P_{S} to reach the highest value P_{S0}.

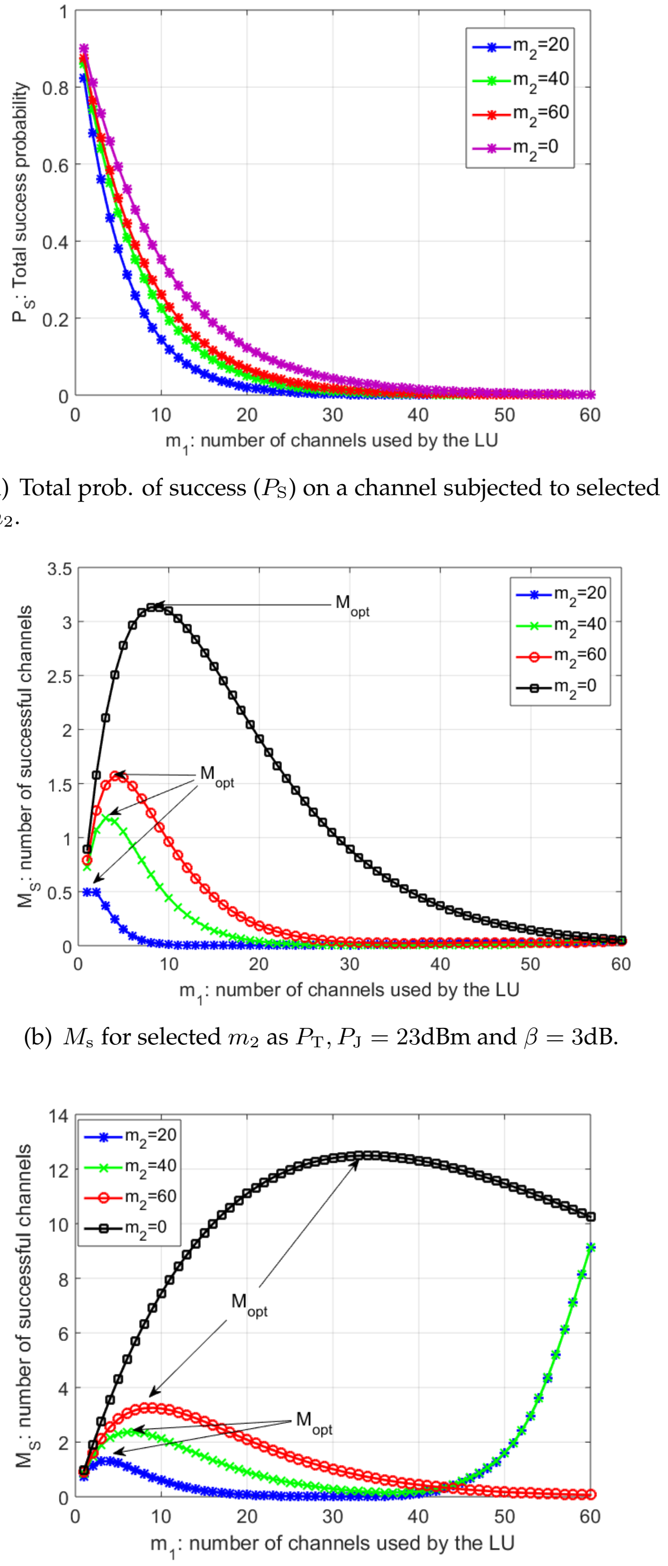
Numerical results are obtained also for M_{s} to show the best response of the LU as a function of m_1 as depicted in Figure 5.3(b) varying m_2. Figure 5.3(b) illustrates the optimal m_1 denoted as M_{opt}, when the MU is absent (i.e. $m_2 = 0$), indicating that the optimal strategy of the LU depends only on P_{S0}, hence Λ, and the optimal strategy of the LU when MU is present $(m_2 = 20, 40, 60)$, characterizing the function P_{S} to depend on Θ.

We note that at the absence of the MU $(m_2 = 0)$, M_{opt} is at $m_1 = 8$, with a successful transmission output (M_{s}) of $3 - 4$ channels. As m_1 increases, the number of successful channels decrease, since $\frac{P_{\mathrm{t}}}{m_1}$ is lower. However, under the presence of jamming attacks $(m_2 = 20, 40, 60)$ where there is an increasing m_2, the LU's optimum shifts towards the right, stating that M_{opt} increases accordingly. For values of $m_2 = 20, 40, 60$, M_{opt} shifts to $3, 5, 6$ respectively, with a successful output (M_{s}) of $1 - 3$ channels, indicating that LU's best option is to increase m_1 as m_2 increases. This is because the parameter $\frac{P_{\mathrm{J}}}{m_2}$ decreases as an increase in m_2 occurs, therefore, the Θ per channel increases. This allows the LU to consider increasing m_1 for better transmission rates that are acceptable within the threshold limit β.

Figure 5.3(c) describes the behavioral effect of M_{s} for the increase in P_{T}, P_{J} and $\beta = 7$dB. There is no single global optimum solution as shown in Figure 5.3(c), since P_i plays a major role in M_{opt}, since P_{S} is significantly higher for increasing m_1. One observes that for MU's strategy of $m_2 = 20, 40$, we observe two optimal points of the LU's strategy. The number of successful channels M_{s} is larger at $m_1 = 60$ than at m_1 is in the range of $5 - 10$ channels.

For increasing m_1 and the condition $m_1 \geq m_2$ is satisfied, we can state the existence of two optimums due to P_i playing a significant role, due to $\frac{P_{\mathrm{T}}}{m_1}$ being sufficient enough for successful transmission for the condition $P_{\mathrm{S}} \geq \beta$ to be satisfied.

As seen in Figure 5.4, M_{s} decreases and M_{opt} shifts to the left indicating an overall disadvantage for the LU as P_{J} increases, for a selected m_2. Moreover, Figure 5.5 highlights that M_{s} increases and M_{opt} shifts to the right of the figure indicating an overall advantage for the LU as r_2 increases, for a selected m_2.



(a) Total prob. of success ($P_{\rm S}$) on a channel subjected to selected m_2.

(b) $M_{\rm s}$ for selected m_2 as $P_{\rm T}, P_{\rm J} = 23$dBm and $\beta = 3$dB.

(c) $M_{\rm s}$ for selected m_2 as $P_{\rm T}, P_{\rm J} = 33$dBm and $\beta = 7$dB.

FIGURE 5.3: Theoretical Graphical Analysis of $P_{\rm S}$ and $M_{\rm s}$

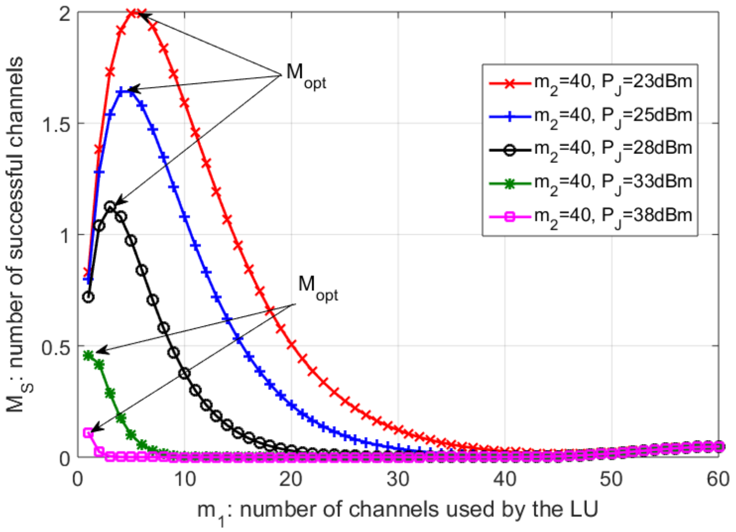


FIGURE 5.4: M_s for varying P_J with $P_\mathrm{T} = 23\mathrm{dBm}$ and $\beta = 3\mathrm{dB}$.

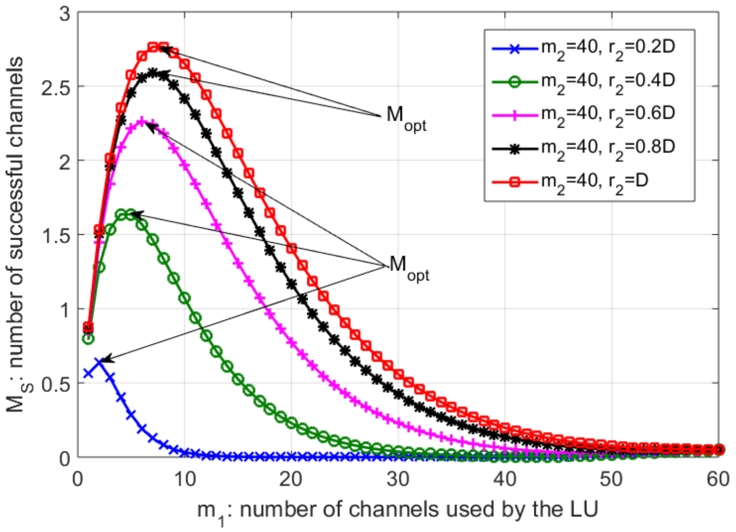


FIGURE 5.5: M_s for varying r_2 with $P_\mathrm{T} = 23\mathrm{dBm}$ and $\beta = 3\mathrm{dB}$.

5.6 A game theoretic approach with jamming history

The aim of the game-theoretic model is the description and prediction of the resulting behaviour between the two competing players (LU & MU). The objective of the LU is to achieve the highest M_s, while the player MU's objective is the complementary which is to thwart the LU's transmissions and minimize the M_s[56, 58]. This leads to the use of the context: zero-sum game as a natural model [24], since the MU aims at minimizing, and LU at maximizing, from which arose the games' name "zero-sum".

Zero-sum games are proven to have the existence of a NE by using Von Neumann's theorem [50]. Hence, we can say that our game model has an existing NE. Both the

LU and MU are stated to be *mixers*, i.e., which means they adopt *mixed strategies* for transmission.

Denote $\pi_{m_1}^{(LU)}(\pi_{m_2}^{(J)}$, resp.) as the probability that LU transmits on $m_1(m_2)$ channels and $\pi_{m_2}^{(J)}(\pi_{m_1}^{(LU)}$, resp.) as the probability that MU transmits on $m_2(m_1)$ channels. Therefore, $M_{\mathrm{s}}(\pi^{(LU)}, \pi^{(J)})$ is the expected return with respect to the independent transmission strategies of LU and MU.i.e.,

Definition 1 *The pair of transmission strategies* $(\pi^{(LU)*}, \pi^{(J)*})$ *is said to be at NE:*

$$M_{\mathrm{s}}(\pi^{(LU)}, \pi^{(J)*}) \leq M_{\mathrm{s}}(\pi^{(LU)*}, \pi^{(J)*}) \leq M_{\mathrm{s}}(\pi^{(LU)*}, \pi^{(J)})$$

for all transmission strategies $\pi^{(LU)}, \pi^{(J)}$.

The concept of NE is the existence of an equilibrium point where the strategies used by LU and MU do not need to change, as there is no further benefit by either player changing their own strategies $\pi^{(LU)}$, $(\pi^{(J)})$ respectively. In zero-sum games, pure strategies at NE exist among mixed strategies [50, 48].

5.6.1 Fictitious Play

This addresses a dynamic case where the LU transmission occurs over multiple frames. A key assumption is that the LU and MU keep track of the number of channels and not the channel locations as they are accessed in a random manner. The LU and MU keep track of the strategies played by their opponent denoted as (π^{m_2}, π^{m_1}) respectively, at several turns of the game. At each play, the LU acts to maximize M_{s} against MU's observed empirical probability distribution.

This is followed with a learning mechanism termed as *Fictitious Play* and is defined as the following,

Definition 2 (Fictitious Play) *[16, 15] Let $f^{(LU)}_{m_1}(t)$ ($f^{(J)}_{m_2}(t)$, resp.) be the frequency up to time t with which LU(MU) has transmitted on m_1 (m_2) channels, i.e.,*

$$f^{(LU)}_{m_1}(t) = \sum_{k=1}^{t} \Pi(m_1(k) = m_1)/t$$
$$f^{(MU)}_{m_2}(t) = \sum_{k=1}^{t} \Pi(m_2(k) = m_2)/t,$$

where Π defines the indicator function.

By convention, we initialize $f^{(LU)}_{m_1}(0) = f^{(J)}_{m_2}(0) = 1/M$. From frame $t = 1$ onwards, it is well known [55] that the frequencies f of FP converge to a (in general, mixed) equilibrium of the game. At frame t, both LU and MU know $m_2(k)$ and $m_1(k)$ utilized for transmission at frame $k = 1, \dots, t-1$ by the opponent MU and LU, respectively. The action at time t by the players LU and MU are denoted as $m_1(t)$ and $m_2(t)$, respectively.

$$m_1(t) = \arg\max_{m_1} \; M_{\mathrm{s}}\big(m_1, f^{(MU)}_{m_2}(t-1)\big)$$
$$m_2(t) = \arg\min_{m_2} \; M_{\mathrm{s}}\big(f^{(LU)}_{m_1}(t-1), m_2\big)$$

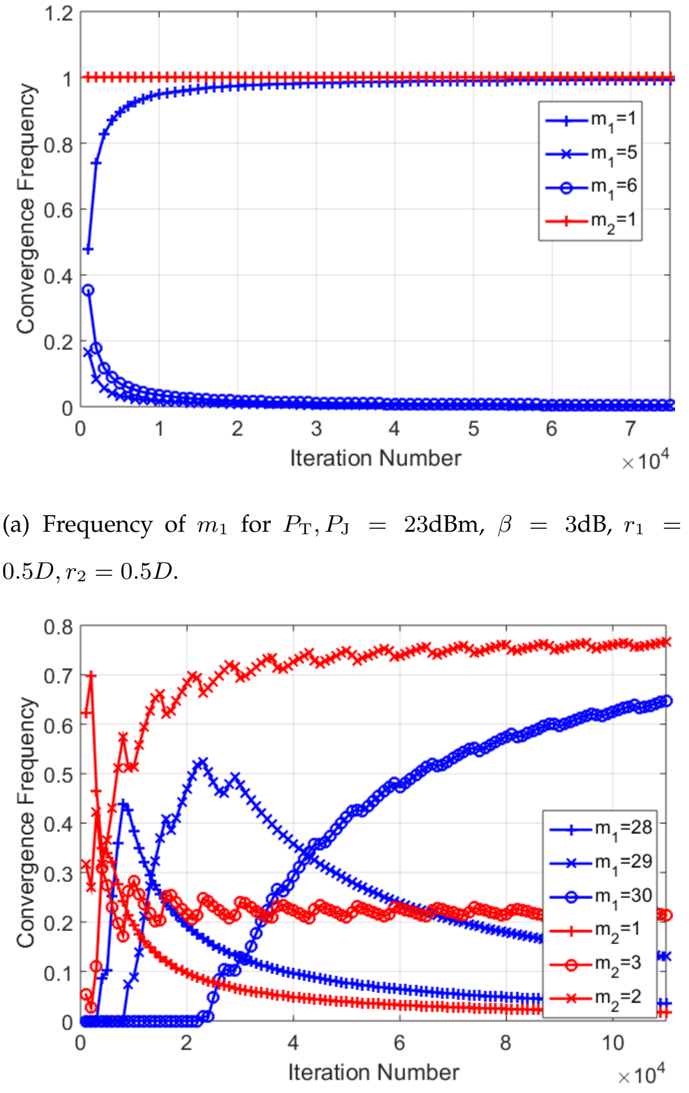
The evolution of the FP game results in a NE, based on the respective opponent's strategy and the corresponding frequency of play [55],

$$\lim_{t\to\infty} f^{(LU)}(t) = \pi^{(LU)*}$$
$$\lim_{t\to\infty} f^{(J)}(t) = \pi^{(J)*} \quad .$$

5.6.2 Simulation Results & Performance Analysis

The most effective strategies with relation to M_{s} are acquired by the players in the FP model from empirical observation of their respective opponent's strategies. Figure 5.6(a) illustrates the strategies of both players LU and MU which are m_1, m_2 respectively, under the constraints: P_{T} = 23dBm, $r_1 = 0.5D, r_2 = 0.5D$ and β = 3dB. An equilibrium point (m_1,m_2=1,1) is reached as noted in the Figure 5.6(a) indicating that the game $M_{\mathrm{s}}(\pi^{(LU)*}, \pi^{(J)*})$ plays a pure strategy. Hence, the game converges to an equilibrium comprising of the minimum number to conserve all its power on one channel. This is mainly due to the condition $P_{\mathrm{S}0} \geq \beta$ not being satisfied.

Figure 5.6(b) illustrates that the LU and MU employ mixed strategies for the radii $r_1 = 0.5D, r_2 = 0.3D$. It illustrates the fact that the LU and MU use mixed strategies with corresponding frequency of play. One should note that the strategies employed by LU and MU (π^{m_1}, π^{m_2}) are adjacent values in the discrete domain.



(a) Frequency of m_1 for $P_{\mathrm{T}}, P_{\mathrm{J}}$ = 23dBm, β = 3dB, $r_1 = 0.5D, r_2 = 0.5D$.

(b) Frequency of m_1 for for $P_{\mathrm{T}}, P_{\mathrm{J}}$ = 23dBm, β = 3dB, $r_1 = 0.5D, r_2 = 0.3D$.

FIGURE 5.6: Fictitious Play Results of the LU and MU

5.6.3 Uniqueness of NE

In this section, we aim to prove the existence of a unique NE in zero-sum games under the following conditions: $P_{\mathrm{T}}, P_{\mathrm{J}} = 23$dBm, $\beta = 3$dB, by use of the analogy of convex-concave games. The uniqueness is proven by means of a minimum and maximum solution [14, 70, 49], where the convexity and concavity of the component functions of the utility function M_{s} are taken into consideration. First of all, we shall assume the utility $M_{\mathrm{s}}(m_1, m_2)$ to be continuous rather than a discrete set of values, to apply the above theorem.

The theorem is defined as in the following:

Theorem 1 ([31]) *Let X, Y be a compact, convex subset of some finite dimensional Euclidean space, where $m_1 \in X$, $m_2 \in Y$, and $M_{\mathrm{s}} : X \times Y \leftarrow \mathbb{R}$ be stated as a continuous optimal function, i.e., concave in m_2 and convex in m_1. This characterizes the two-person zero-sum game where M_{s} is the payoff function and X(resp.Y) is the strategy set of the maximizer (resp.minimizer). It is known and expected to be the following [70],[31]:*

$$A(m_2) = \max_{m_1 \in X} M_{\mathrm{s}}(m_1, m_2) \tag{5.12}$$

$$B(m_1) = \min_{m_2 \in Y} M_{\mathrm{s}}(m_1, m_2)$$

Therefore, A is convex and continuous on Y, B is concave and continuous on X, by using the Maximum theorem. One should take into account that the following condition: $A(m_2) \geq M_{\mathrm{s}}(m_1, m_2) \geq B(m_1)$, for all m_1, m_2 in $X \times Y$ is satisfied, thus, we can state the following:

$$\max_{m_1 \in X} B(m_1) = \max_{m_1 \in X} \min_{m_2 \in Y} M_{\mathrm{s}}(m_1, m_2) \leq \min_{m_2 \in Y} \max_{m_1 \in X} M_{\mathrm{s}}(m_1, m_2) = \min_{m_2 \in Y} A(m_2). \tag{5.13}$$

Hence, we can further state that if the above conditions based on M_{s} is satisfied, the game is said to have a single NE. Weaker assumptions of convex-concave games such as quasi-convex and quasi-concave games suffice to prove that the game suffice as well due to the Hausdorff topological space property[70, 31].

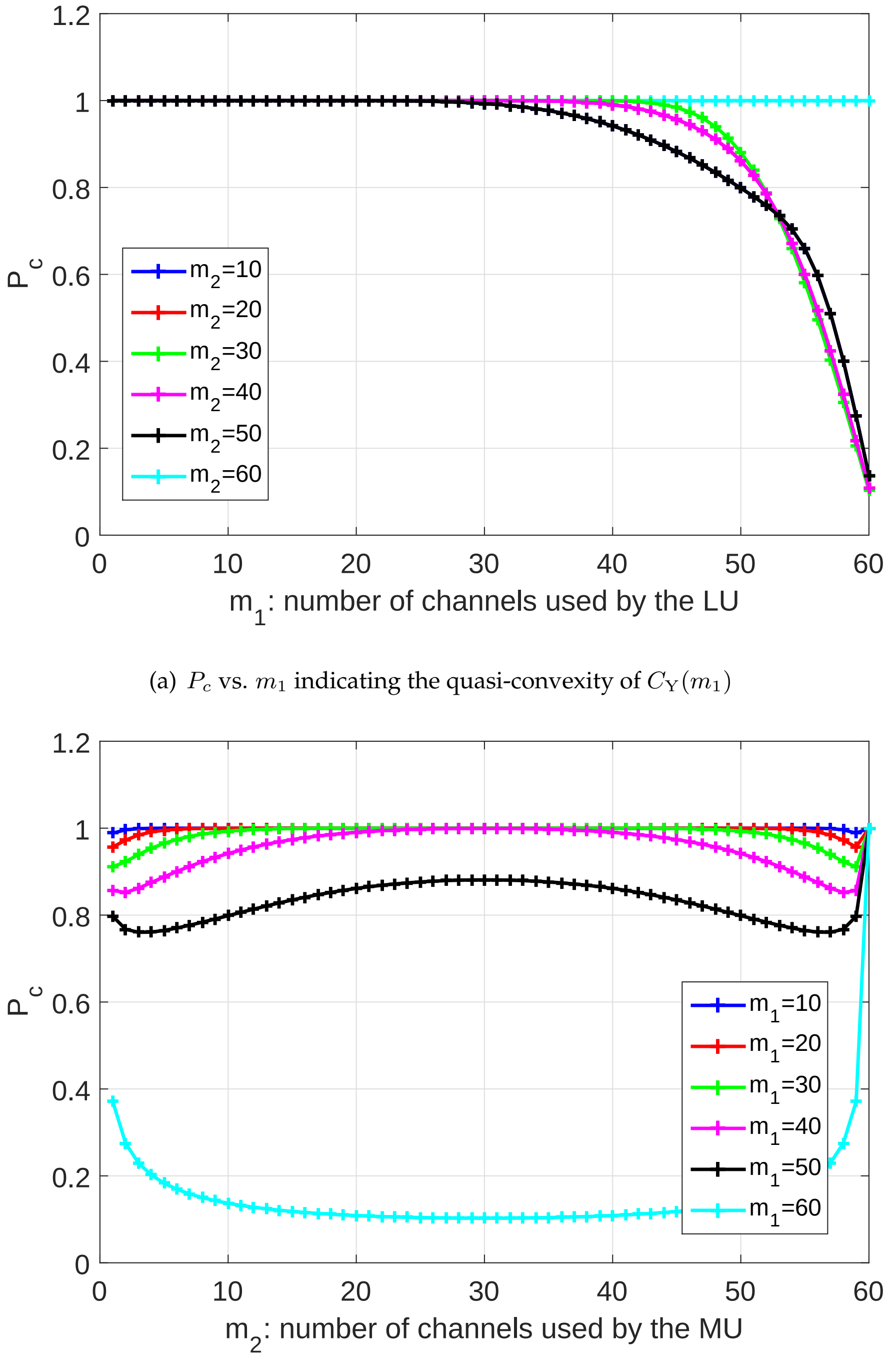
By applying the above definition, we aim to prove that our game is a convex-concave (weaker basis as quasi-convex and quasi-concave) through identifying the individual components: $P_{\mathrm{S0}}, P_{\mathrm{S1}}$ and P_i of the function M_{s}. Firstly, we shall prove the convexity of $M_{\mathrm{s}}(m_1)$. Secondly, we prove the concavity of $M_{\mathrm{s}}(m_2)$. Once this is proven under certain parametric constraints such as: $P_{\mathrm{T}}, P_{\mathrm{J}} = 23$dBm, $\beta = 3$dB and for values of r_1, r_2, we can state the game has a unique NE as the game displays the behaviour of a convex-concave one.

The following mechanism proves the convexity and concavity of M_{s}.

- $P_{\mathrm{S0}}(m_1)$ and $P_{\mathrm{S1}}(m_1)$ are convex functions due to their exponential property. Therefore, one states $(P_{\mathrm{S0}} - P_{\mathrm{S1}})$ is a convex function as well.
- $P_{\mathrm{S1}}(m_2)$ is a concave function as proven in equation (5.14), where $Y = \frac{\beta m_1}{P_1}$.
- We, then consider the discrete function $C_{\mathrm{Y}}(m_1, m_2)$. We state that $C_{\mathrm{Y}}(m_1)$ exhibits a quasi-convex and $C_{\mathrm{Y}}(m_2)$ displays a quasi-concave behaviour. This is shown in Figure 5.7(a). One should note that the quasi-concavity does not hold for $m_2 = 60$ in Figure 5.7(b), as all the channels are interfered.

$$\frac{dP_{\mathrm{S1}}}{dm_2} = YP_2 m_2^{-2} \exp\left(-Y\left(\frac{P_2}{m_2} + P_{\mathrm{N}}\right)\right) \tag{5.14}$$

$$\frac{d^2P_{\mathrm{S1}}}{dm_2^2} = Y^2P_2^2 m_2^{-4} \exp\left(-Y\left(\frac{P_2}{m_2} + P_{\mathrm{N}}\right)\right) - 2YP_2 m_2^{-3} \exp\left(-Y\left(\frac{P_2}{m_2} + P_{\mathrm{N}}\right)\right) \leq 0$$



(a) P_c vs. m_1 indicating the quasi-convexity of $C_{\mathrm{Y}}(m_1)$

(b) P_c vs. m_2 indicating the quasi-concavity of $C_{\mathrm{Y}}(m_2)$

FIGURE 5.7: Plots of the probability of collision $C_{\mathrm{Y}}(m_1, m_2)$

This indicates that $M_{\mathrm{s}}(m_1, m_2)$ follows a quasi-convex and quasi-concave behaviour, thus leading to the possibility of a unique NE, provided the absence of a maxima at $m_1 = M$ as shown in equation (5.15) which defines a general case where parametric constraints are not considered.

Equation (5.15) provides the basis for the existence or non-existence of a minima at $m_1 = M$, by use of the second order derivative of P_{S}, where K is defined as $C_{\mathrm{Y}}(m_1 = M, m_2)$ and is a constant as a function of m_1.

$$\frac{d^2 P_{\mathrm{S}}}{dm_2^2} = \left(\frac{\beta P_{\mathrm{N}}}{P_1}\right)^2 \exp\left(\frac{-M\beta P_{\mathrm{N}}}{P_1}\right) + \left(K\left(\left(\frac{\beta}{P_1}\left(\frac{P_2}{m_2} + P_{\mathrm{N}}\right)\right)\right)\exp\left(-M\left(\frac{P_2}{m_2} + P_{\mathrm{N}}\right)\right)\right) - K\left(\frac{\beta P_{\mathrm{N}}}{P_1}\right)^2 \exp\left(\frac{-M\beta P_{\mathrm{N}}}{P_1}\right) \tag{5.15}$$

From equation (5.15), let us denote
$C_1 = \left(\frac{\beta P_{\mathrm{N}}}{P_1}\right)^2 \exp\left(\frac{-M\beta P_{\mathrm{N}}}{P_1}\right)$, $C_2 = \left(K\left(\left(\frac{\beta}{P_1}\left(\frac{P_2}{m_2} + P_{\mathrm{N}}\right)\right)\right)\exp\left(-M\left(\frac{P_2}{m_2} + P_{\mathrm{N}}\right)\right)\right)$
and $C_3 = -K\left(\frac{\beta P_{\mathrm{N}}}{P_1}\right)^2 \exp\left(\frac{-M\beta P_{\mathrm{N}}}{P_1}\right)$.
In the case of $C_1 + C_2 \leq C_3$, we state there is a single maximum point at $m_1 \leq M$, thus leading to a single equilibrium point. If the condition $C_1 + C_2 \geq C_3$ is satisfied at $m_1 = M$, there is a minima at $m_1 \geq M$, thus indicating the possibility of two maximum points, which leads to two or more equilibrium points. Thus, we observe two optimal points enveloping into a mixed value rather than a unique one. Therefore, we can give the bound to determine whether there is a unique single optimum or two optimal points. Hence, we can prove the existence of a unique NE under different conditions.

5.7 Multi-Armed Bandit Model with Q-Learning

In this section, we aim to identify the LU's play m_1 in order to get the best response (M_{s}), given m_2 is fixed, but the LU and MU's spectral locations are random. The game theoretic approach through the FP model resulted in a global optimum while the RL model aims to provide the local optimum for specific strategies of MU (m_2). We take into consideration, the total power of the LU and MU ($P_{\mathrm{T}}, P_{\mathrm{J}}$) are the same for our current set of workings but is not necessarily constrained to be, as the learning can be applied regardless of power level consumption of the users. However, the power allocated per channel by the LU and MU ($\frac{P_{\mathrm{T}}}{m_1}$), ($\frac{P_{\mathrm{J}}}{m_2}$) may differ respectively. The LU is unaware of any threats and begins to learn the environment (interfered channels/ successful channels), while the MU randomly interferes m_2 channel locations. Therefore, one would desire the LU to have a best response m_1. Therefore, the LU learns the jammed spectral locations and must decide to reduce the number of channels and alter the spectral locations at the next time interval. This is done by means of a RL approach [57, 40].

The model used is the multi-armed bandit enforced with the Q-learning algorithm for a dynamic reward system. The multi-armed bandit has a set of real bernoulli distributions for the rewards ($B = R_1, R_2, ...R_M$) with the corresponding actions ($m_1 = 1, 2, ..., M$). The model has mean rewards denoted as ($\mu = \mu_1, \mu_2, ..., \mu_k$) for actions ($m_1 = 1, 2, ..., M$). The reward function at time t denoted as R_t is expressed in equation (5.16), where M_{RL} is the number of successful channels at a particular time in the learning scheme.

$$R_t = M_{\text{RL}}(m_1, m_2) \tag{5.16}$$

A MDP contains four important components: finite set of states ($s \in S$ where $s \in 1, .., M$), finite set of actions ($a \in A$ where $a \in 1, .., M$), transitional probabilities, and immediate payoffs/rewards. Q-learning is an off-policy temporal difference (TD) control algorithm. The transition between the current and next state is simply the action $a = m_1$ taken at the current time interval, which may differ at adjacent time intervals. Simply, the action a performed leads to the next state of the LU at time $(t+1)$ which is denoted as s'. The action at the next time interval is denoted as a'. There are M states in total, thus, the transitional probabilities $p(s'|s, m_1)$ to move from one state to another at a given time frame, is expressed as: $p(s'|s, m_1) = \frac{1}{M}$, initially.

The Q-learning algorithm comprises of state-action matrix denoted as $Q(s, a)$, where the learned action-value function denoted as Q, approximates to Q^* the optimal action-value function, without depending on any policy to be followed. This dramatically simplifies the analysis of the algorithm and enables early convergence proofs. The policy has an effect where it determines which state-action pairs $Q(s, a)$ are visited and updated. However, all that is required for correct convergence is that all pairs continue to be updated.

One should understand that at the end of the time frame t, the LU observes the current state s_t, hence it chooses an action a_t which increments or decrements m_1, based on the reward of the previous time frame R_{t-1}.

$$Q_{t+1}(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \alpha_{RL}.(R_{t+1} + \gamma \max_{a'}(Q_t(s_{t+1}, a)) - Q(s_t, a_t)) \tag{5.17}$$

The commonly-known equation of Q-learning is shown in equation (5.17), states how the Q-learning algorithm updates the Q matrix over a number of iterations/time frames, where α_{RL} is the learning rate, γ is the discount factor that is required for convergence and the conditions ($\alpha_{RL} < 1$) and ($\gamma < 1$) are satisfied. The algorithm converges once the condition based on Υ is met where Υ is the deciding threshold for convergence of the Q-matrix and is set so that the Q-learning matrix does not frequently incorporate the greedy search. The $\epsilon-$greedy search strategy ensures that the best action a is selected based on the previous reward with a probability $\epsilon = 0.9$, and a random action a is selected with a probability $(1 - \epsilon)$. Such a value is taken to ensure that the explorative phase of the learning is kept minimal and the exploitative phase is pursued with a higher need. It is further extended by applying the selective $\epsilon-$greedy search where the subspace of the matrix is controlled by the factor of our learning method as based on R.

Algorithm 1 Q-Learning of the multi-armed bandit.

```
procedure CH–Q-LEARNING ALGORITHM
    for $t = 1, 2, 3, ....$ do
        for every state $s' \in m_1 = 1, 2, 3, ....., M$ do
            for every action $'a' \in m_1 = 1, 2, 3, ....., M$ do
                $Q(s, a) \leftarrow Q(s, a) + \alpha.[R(s, a) + \gamma \max_{a'}(Q(s', a')) - Q(s, a)]$
                $s'$ is the next state and $a'$ is the action performed at the next time step.
                $s' \leftarrow a$
                $s \leftarrow s'$
            end for
            if $\max(Q(s, a)_{t-1}) - \max(Q(s, a)) \leq \Upsilon$ then
                The outer loop is terminated
            end if
        end for
    end for
Normalize the $Q(s, a) \rightarrow \bar{Q}$
After convergence, $\max(\bar{Q}_t(s, a))$ is the optimal value or best state $s$ to be and perform action $a$.
end procedure
```

Algorithm 1 describes the Q-learning approach taken by the LU to determine the optimal state Q^*, that defines the best response state s and action a for given m_2, though it is the action $a^* = m_1$ that is significant. The matrix $Q(s,a)$ represents the learning paths of the LU while being in one state and performing an action out of the action set, where the initial state s and action a is randomly selected. At each time frame, the reward R is calculated as shown in equation (5.16). Hence, the Q value is updated as referred upon in (5.17).

Furthermore, the algorithm incorporates an $\epsilon-$greedy search to explore all possible actions as much as possible. This is necessary to ensure that the Q-learning has taken into account the possibility of two or more best responses.

Figure 5.8 depicts the optimal strategies of the LU as obtained from Q-learning for $m_2 = 60$ strategy of the MU, which can be extended to the different strategies of m_2. The blue trend depicts the m_1 used by the LU and the red trend depicts the acquired M_s. The optimum M_opt is observed from the blue trend as it peaks at $6-7$ channels. Figure 5.9 compares the M_opt of LU acquired via numerical analysis and Q^* obtained through Q-learning algorithm by Monte-Carlo simulations. Therefore, one can state that the Q-Learning converges to the theoretical best response M_opt of the LU for each m_2.

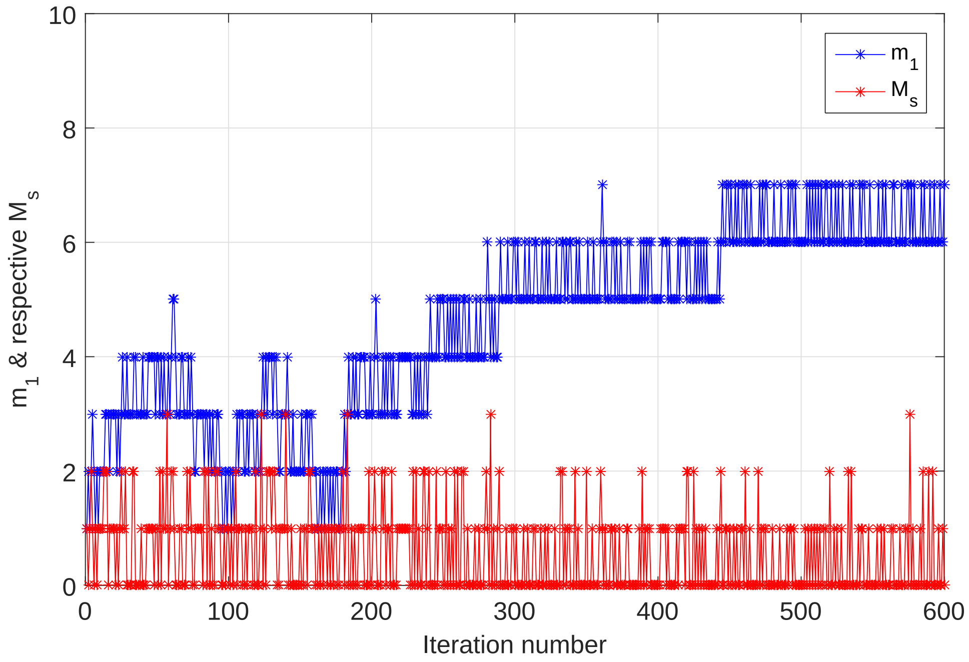


FIGURE 5.8: Reinforcement Learning for $m_2 = 60$ detailing the reward M_s for the corresponding m_1.

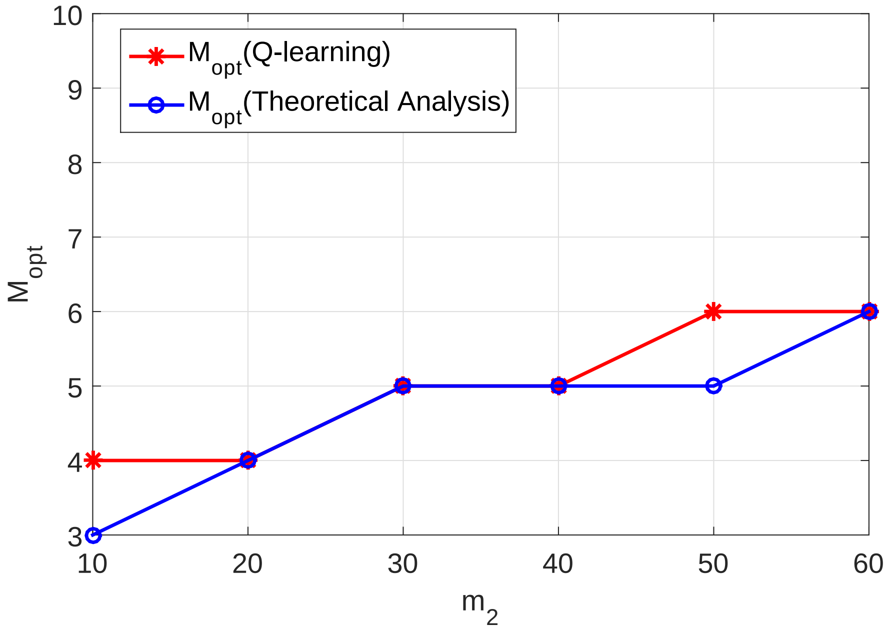


FIGURE 5.9: Numerical analysis from theoretical curves vs Q-Learning simulations.

5.8 Summary of Contribution and Conclusion

The main contribution of this chapter lies in the method of successful transmission given jamming takes place, hence, provide a game theoretical framework in the form of zero-sum games with a NE for a single time interval, and a FP model for many time intervals, with the prior assumption that both LU and MU have past knowledge of each other. The other contribution is a multi-armed bandit model with Q-learning to allow the LU to learn the MU's behaviour and determine a best response in terms of number of channels to access against the MU's local strategies. This chapter concludes the investigation of resilience of the CRN with multiple available channels leased by the primary communication system, by considering the model interaction between the secondary users: LU and MU, hence study the optimal strategy, and the global and local equilibrium of the games, for the scenarios modelled in the ABSOLUTE project [7]. Both the users are equipped with a single radio and are able to access multiple channels simultaneously. Therefore, the play between the LU and MU are based on the power allocated per channel and the number of channels accessed at one instance. Additionally, Q-learning supports the numerical analysis of the model and provides a local optimum for selective malicious strategies: based on the number of channels accessed, while spectral locations are randomized. Furthermore, we investigate the global optimum for the users considering a zero-sum game model. The resilience strategy of the LU shown on the NE is optimal since it minimizes the worst possible damage to the legitimate user caused by the attacks of the malicious users. We demonstrate and prove the uniqueness of the game by applying the concept of convex-concave games, where there exists a maxmima/minima solution.

Chapter 6

Optimal Decision Making for Opportunistic Spectral Access

Several SUs within a geographical space are competing for channel availability at a given time interval. A MDP is formulated to determine the optimal state for a user of interest to be in, with consideration to minimal interference and wastage of the channel. This chapter describes the formulation of the MDP to the stated problem, and compares the performance of other models such as POMDP and Incomplete Markov Model (IMM) in terms of utilization and interference.

6.1 Background

CRNs comprise of several unauthorized users attempting opportunistic access at the available spectrum in a given geographical area. This results in an unreliable form of communication for such users as they are not leased to spectral information. Thus, they are identified as SUs. The loss in such wireless networks are mainly due to interference between other SUs and transmission error. Although transmission error depends on the radio propagation environment and is difficult to be predicted upon, interference probability can be predicted by inference from the acquisition of information regarding the transmission environment.

The application of this proposal is heavily considered in Vehicular Communications. Vehicular Communication is proposed as a functional concept due to the rising potential in automotive-communication technology. During the last decade, a new approach known as cooperative driving based on Vehicle-to-Vehicle (V2V) was proposed, as a means of overcoming potential dangerous situations. It comprises of an intelligent transportation system to reduce probability of road accidents. Different type of such systems are proposed, and among them CR based systems are becoming more prominent due to their capability of conforming to the bandwidth demand of highly congested areas. Such an issue in demand is capable of being solved in CR based systems.

In this chapter, we propose a problem formulation through a fully observable MDP approach to determine whether a user/ device of interest should transmit or not at the given time transmission time interval (TTI). Hence, the tradeoff between interference and wastage is studied based on the decisions made over time. We compare the degree of performance on the mentioned parameters with the Incomplete Markov Model (IMM) that incorporated RL[84] and the Partially Observable Markov Decision Process (POMDP) from existing literature [11].

6.2 Literature Review

The area of CR communications have been rising slowly in the past decade as commercialization based firms consider such an area a noteworthy investment. This is confirmed further by the alarming need of communications between vehicles for public safety and road assistance. This leads to an exponential rise in the number of connected devices. Thus, leading to a boom in inanimate objects using the upcoming 5G network - known as the "Internet of things". The need for efficient opportunistic access is of dire concern and thus, there have been many significant works on it. We shall focus on the following works [78, 82, 75, 83]. The paper [78] refers to the use of Markov process incorporating myopic sensing for opportunistic access. Myopic sensing maximizes the immediate reward but ignores the impact of the current strategy on the future reward. Thus, an analytical study was conducted on the optimality of the myopic policy under imperfect sensing. The second paper [82] refers to the use of a channel learning algorithm for opportunistic access between random packet burst, where the CR system is modelled as a Hidden Markov model (HMM). A gradient method to find the underlying PU traffic pattern to facilitate the algorithm's function of optimal opportunistic access. An analysis was conducted on the degree of performance between the proposed algorithm and the generic listen-before-talk algorithm. Thirdly, the paper [75] refers to dynamic bargaining solutions for opportunistic spectrum access where comparison is made with global optimum and non-cooperative solutions in a strategic setting. The cost of bargaining and benefit of bargaining in the stochastic model is examined, in which each user has its own state MDP. The states and actions of the users that sense the similar channel, determine the instantaneous payoffs. They also determine the transition probabilities to move to the next states. The bargaining outcomes are characterized in the short and long term. There have been extensive work in the various methods applied for opportunistic access in CRNs. The novelty in our paper lies in the fully observable MDP for opportunistic access in terms of detection and false alarm probabilities and an analysis of the possible interference and wastage is conducted with comparison to the models referred in [82],[57]. The possible use of CR based autonomous vehicular communications are shown in the following existing literature [61],[23],[60].

6.3 System Model

The system comprises of a single user of interest with CR functionality trying to access one specific channel at a given instance in a distance limited geographical area and be aware of other users intending to access the same channel. The factor to be learned upon is the occupancy of the channel with regard to the presence of ongoing dynamic transmissions. A decision on whether to transmit or not for the user of interest is made upon the acquired knowledge of the channel's occupancy and other nearby CRs. The next assumption is that the CR in the vehicle performs a CSMA (Carrier Sense Multiple Access) method for listening to the channel.

The channel occupancy is described by the use of the two states: H_0 (Channel is not occupied) and H_1 (Channel is occupied), and the transitional probabilities channel occupancyλ (Rate of channel being occupied) and channel non-occupancy μ (Rate of channel not being occupied).

The sensing scheme of each CR is illustrated in figure 6.1. The sensing duration is denoted as δ_t and sensing period is known as T_w. This leads to the analysis of MDP

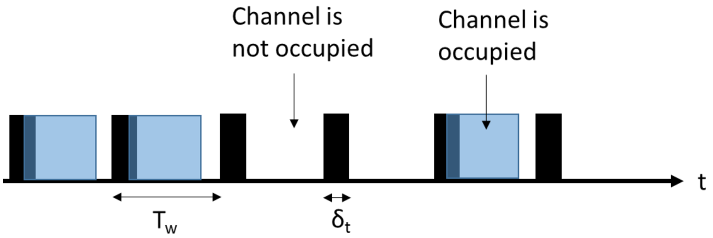


FIGURE 6.1: Channel Occupancy Sensing Scheme of CR

for the user of interest. This is important as it characterizes the method in which the user acquires information regarding the environment. It is also important to note the following assumption: the presence of multiple CRs leading to decisions being made in a distributive manner results in the requirement of the convergence of the system.

6.4 Markov Decision Process Formulation for Optimal Channel Access

The complete MDP for the CR environment is illustrated in figure 6.2. The completely observable MDP comprises of states, actions leading to the next state and the transitional probabilities.

6.4.1 MDP model formulation

The states are based on the channel occupancy and the decision of the user of interest. They are denoted as $H_{\text{i,j}}$ where i denotes the channel occupancy and j denotes the transmission status of the user. Hence, there are four states ($s \in S$): $H_{0,0}$ - channel is not occupied by any user, $H_{0,1}$ - channel is occupied by the user of interest, $H_{1,0}$ - channel is occupied by other user, and $H_{1,1}$ - channel is occupied by both users. There are two possible actions ($a \in A$) that the user of interest may take: $a = 0$ (Does not transmit) and $a = 1$ (Does transmit). The user of interest may decide to transmit with a probability p and not transmit with $1-p$, where $p = (1-P_F)(1-\lambda)+\lambda(1-P_D)$. The value of p is assumed to be known by the user as prior learning would have already taken place. The transition probability between the states are defined by the parameters:λ, probability of detection P_D and probability of false alarm P_F. The transition from the states $H_{\mathrm{i,j}}$ where i $= 0$ to i $= 1$ is represented by $\lambda(1-P_D)$ and λP_D. This allows one to note that with a miss detection probability $1-P_D$, the value of j at state $H_{\mathrm{i,j}}$ shifts to 1 or stays at 0 with P_D. Meanwhile, the transition from the states $H_{\mathrm{i,j}}$ where i $= 1$ to i $= 0$ is represented by $(1-P_F)(1-\lambda)$ and $P_F(1-\lambda)$.

Figure 6.2 illustrates the possible states that the user of interest could take. Let us assume the user lies in $H_{0,0}$ at time t, the next action of transmitting or not allows the transition to $H_{0,1}$ or remain in $H_{0,0}$ at time $(t+1)$, respectively. The transitional probabilities are denoted as $(1-P_F)(1-\lambda)$ and $P_F(1-\lambda)$ respectively. This is the case only if the next state has a low channel occupancy likelihood. If not, the user of interest may jump to the state $H_{1,0}$ or $H_{1,1}$, depending on its action of transmitting or not.

6.4.2 Cost

The cost C depends on the state s_t and the action a_t for a given time interval t. The cost C is used to determine the value of being in the current state and provides an idea of the channel status and how long the channel is not occupied over a period of time. This allows the user to learn about the state and determine the action and state pair for a sequence of intervals. The function c gives an estimate value of the cost at $H_{0,0}$ and tends to decrease over repeated intervals, where the rate of decrease depends on β where $\beta \in (0,1)$, and τ is the number of adjacent time intervals where s_t lies at $H_{0,0}$. In case of complete and incomplete knowledge, the cost C acts as a function to determine the degree of interference and wastage in the model.

$$c = \frac{1}{2}(1+\cos(\frac{\pi\tau}{\beta}(\alpha - \frac{1-\beta}{2}))). \tag{6.1}$$

Therefore, the cost C of being in each state at a given time interval t is defined as the following:

$$C = \begin{cases} c & \text{if } s_t = H_{0,0} \\ 0 & \text{if } s_t = H_{0,1} \\ 1 & \text{if } s_t = H_{1,0} \\ 1 & \text{if } s_t = H_{1,1} \end{cases}$$

The objective policy of the MDP is to minimize the cost of being in state S_{t}. Therefore, the goal G of the MDP is to be in $G \rightarrow s = H_{01}$.

6.4.3 MDP Solution

The Bellman equation for the MDP $< S, A, \Pr, C, G >$ where $J_t(s_t)$ is defined as the optimal cost or the minimum cost to reach G from the current state s_t.

$$J_t(s_t) = 0 \text{ if } s_t \in G \tag{6.2}$$

$$J_t(s_t) = \min_{a_t \in A} \sum_{s_{t+1} \in S} \Pr(s_{t+1}|s_t, a_t)[C(s_t, a_t, s_{t+1}) + J_t(s_{t+1})] \tag{6.3}$$

The summation of such transition probabilities are simplified by taking into account the expectation.

$$J_t(s_t) = \min_{a_t \in A} \Bigg(C(s_t, a_t, s_{t+1}) + \mathbb{E}\ J_t(s_{t+1}) \Bigg) \tag{6.4}$$

The following MDP links the possibility of a particular state at time t denoted as $s_t = s$ to transit to the next state denoted as $s_{t+1} = s'$, which is any one of the above mentioned four. Due to the property of the MDP, the transitional probabilities depend on the action taken and the next state chosen. The transitional probabilities are defined through the use of detection probability P_D, false alarm probability P_F, and channel occupancy λ. The values of the stated parameters are acquired from empirical observations. Therefore, the transition matrix can be written in the form as shown in the following:

$$\begin{bmatrix} . \\ J_t(s_t) \\ . \end{bmatrix} = \min_{a_t \in A} \left(\begin{bmatrix} . \\ C(s_t) \\ . \end{bmatrix} + \begin{bmatrix} . & . & . \\ . & \Pr(s_{t+1}) & . \\ . & . & . \end{bmatrix} \begin{bmatrix} . \\ J_t(s_{t+1}) \\ . \end{bmatrix} \right) \tag{6.5}$$

It is eventually expressed as:

$$J_t(s_t) = \min_{a_t}(C(s_t) + \Pr(s_{t+1}) J_t(s_{t+1})) \tag{6.6}$$

The result is an optimal decision vector ${a_t}^* = {a_t}^*(s)$ in the domain $s \in S$, which is equivalent to determining the best policy. Therefore, it is easiest to envision the above optimality as a scalar, but in many applications ${a_t}^*$ is itself a vector.

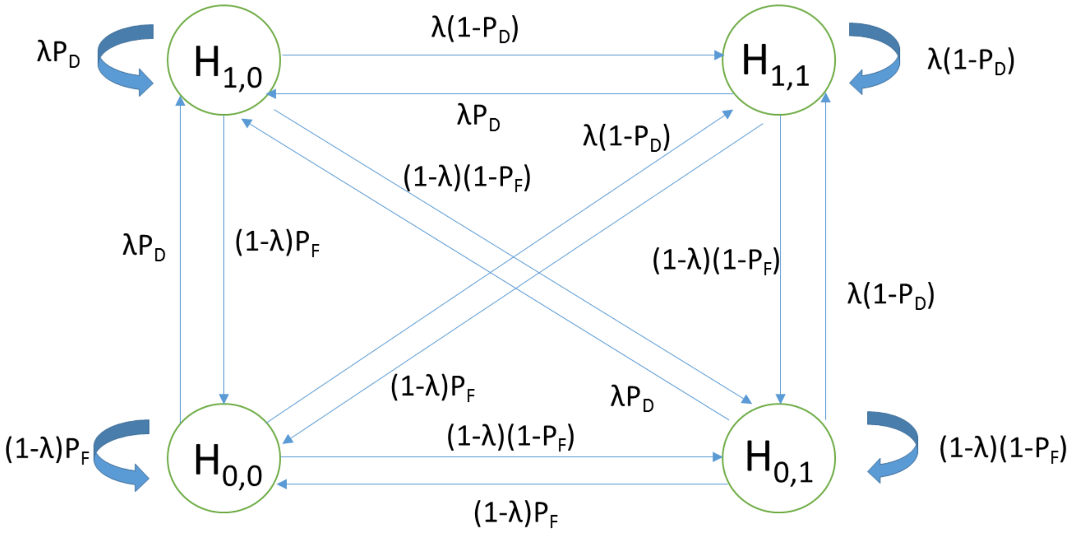


FIGURE 6.2: MDP for User of Interest after learning

6.5 Analysis of Learning and Performance

Simulation analysis is done to compare the degree of performance of the three models described in the chapter and literature [82], [57]. Figure 6.3 illustrates the system performance analysis of our model and the existing models as described in prior. Each model tends to learn or is assumed to acquire the information, resulting in the partial or full knowledge of the robust environment. The POMDP model described in [82] characterizes the PU system behaviour at the beginning and end of a time slot, where the state transitional probabilities are defined as the poisson exponential process with λ and μ. Meanwhile, the IMM model described in [57] and the MDP described in this chapter characterizes the interactions between the PU and SU behaviours. The state transitional probabilities are characterized by the parameters: $P_D = Q_{N_s/2}(\sqrt{N_s\rho}, \sqrt{\lambda})$ and $P_F = \Gamma(N_s/2, \Lambda/2)$, where Λ is the detection threshold of an energy detector.

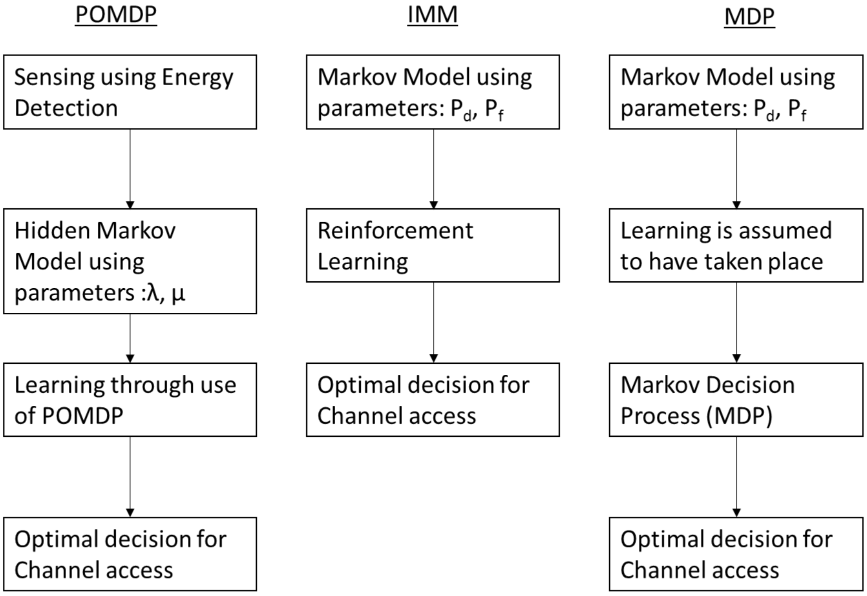

POMDP
IMM
MDP
Sensing using Energy Detection
Hidden Markov Model using parameters :λ, μ
Learning through use of POMDP
Optimal decision for Channel access
Markov Model using parameters: P_d, P_f
Reinforcement Learning
Optimal decision for Channel access
Markov Model using parameters: P_d, P_f
Learning is assumed to have taken place
Markov Decision Process (MDP)
Optimal decision for Channel access

FIGURE 6.3: Performance Analysis of the three System Models

6.5.1 Wastage and Interference Analysis

After the formulation of the MDP, an analysis of how well the MDP performed in terms of overall interference and wastage is implemented. The probability of interference (P_i) and probability of wastage (P_w) is defined as:

$$P_i = \sum_{n=1}^{N_t} \frac{Pr[s_t = (1,1), a_t = 1]}{N_t} \tag{6.7}$$

$$P_w = \sum_{n=1}^{N_t} \frac{Pr[s_t = (0,0), a_t = 0]}{N_t} \tag{6.8}$$

where N_t is the total number of time slots. These two parameters are used as a measure of degree of performance in this chapter.

6.5.2 Simulation Analysis

Simulations were conducted in order to analyze the performance of the three models including [11] in terms of interference or collision probability P_i under different utilization values, which is illustrated in Figure 6.4. It is widely accepted that the complete MDP has a superior performance than the POMDP and IMM. The major concern is the behaviour of these models for varying Λ and μ, which are the underlying basic constraints in all three models. Figure 6.4 illustrates the behaviour of the IMM and MDP in terms of P_i and P_w for varying λ and μ values. As λ increases, P_w decreases prior to an initial increase due to the user of interest having sensing model parameters such as δ_t not altered, hence leading to an initial increase in the waste. The MDP follows a similar behaviour but with a lower bound. On the other hand, P_i tends to increase for increasing values of λ, where the difference between the models is negligible compared to the P_w.

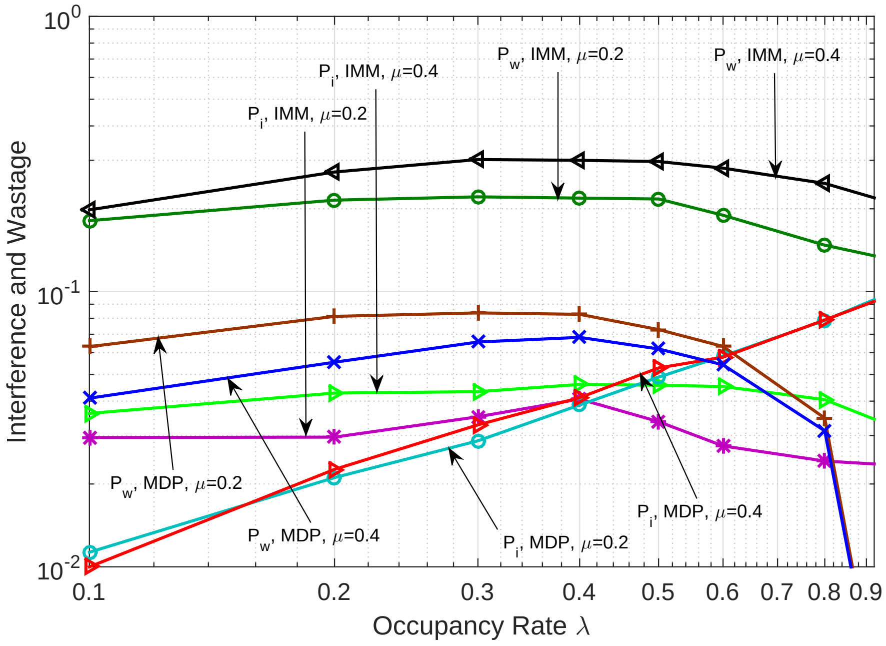


FIGURE 6.4: Interference & Wastage Probability of IMM and MDP

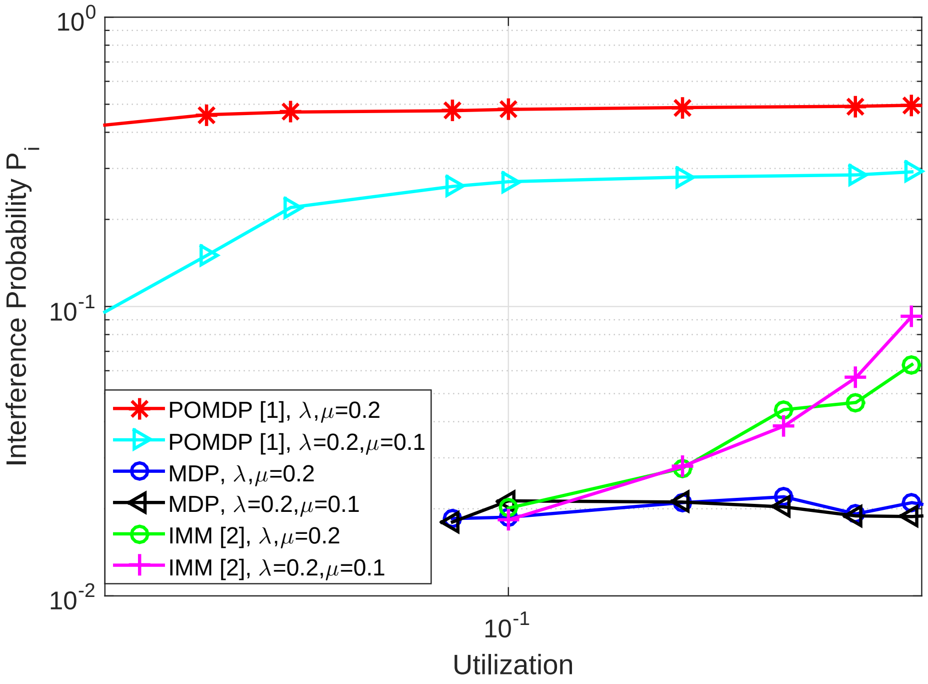


FIGURE 6.5: Interference Probability of the three System Models

Figure 6.5 illustrates the utilization behaviour of the three models: IMM, POMDP, and MDP for varying interference probabilities. It shows that the gradual increase in utilization results in an increase in interference. The MDP model provides the least interference, while the IMM and POMDP are more significant.

One should note that the three models that are compared, tends to have different system and Markov model formulation. Hence, one cannot completely acknowledge the performance of each model as completely valid. This provides a novel concept of the MDP formulation and determines the degree of performance with the IMM and existing POMDP of similar relevance.

6.6 Summary of Contribution & Conclusion

The main contribution of this chapter is the channel occupancy model behavior between competing users as a four-state MDP process, taking into account of the Receiver Operating Characteristics (ROCs) which comprise of $P_{\rm D}$ and probability of false alarm $P_{\rm F}$. The other contributions are the analysis of the trade-off between interference and wastage probabilities of the channel, and the comparison of the degree of performance in utilization and interference against other models such as IMM and POMDP.

This chapter describes a decision making strategy through use of a Markov Decision Process for an environment comprising of the ongoing transmissions in the channel. The decision-making scheme is performed and analysed in terms of interference and wastage probabilities for various occupancy rates. Furthermore, comparisons are made between the IMM, POMDP and MDP in terms of utilization.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this thesis, we explore the three key areas of optimal decision making of Cognitive Radio in Cognitive Radio Networks as research objectives, and present some significant contributions. The thesis clearly identifies some of the objectives that are addressed in cognitive radio networks, and have provided some novel research material to address the relevant research objectives. The research objectives and the relevant contributions of this thesis are addressed individually in chapters $3-6$. The three research objectives are based on some areas of cognitive radio networks, where some gaps were found in current literature. They are in threefold: *research objective 1* sensing and learning in robust aerial-to-terrestrial communications, *research objective 2* resiliency of authorized cognitive radio in cognitive radio networks against jamming attacks from unauthorized cognitive radios, and *research objective 3* optimal decision making of cognitive radio to improve spectral efficiency.

The first contribution in *research objective 1* was an extensive model featuring aerial-to-terrestrial communications which considered fading channels from a theoretical perspective, followed by a data set obtained from simulation where theory was limited (especially in the Rician case). The second contribution was the modelling of Primary user transmissions as a simple Markov model, followed with Markov models for the Secondary user transmissions; which incorporated a Reinforcement Learning algorithm to determine the performance of the Secondary user transmissions in terms of interference and wastage of the channel.

The contribution in *research objective 2* was the proposal of a means of successful transmission for the authorized cognitive radio under jamming attacks from unauthorized cognitive radios, which is formulated as a zero-sum game to model the resilience of the authorized cognitive radio. The framework was then extended by incorporating Fictitious Play learning for both users, to reach a Nash equilibrium. Additionally, we model the action set of the authorized cognitive radio as a multi-armed bandit model that uses Q-learning to learn the behaviour of the jamming attacks and decide upon a suitable strategy against the jamming strategy.

The contribution in *research objective 3* was modelling the interaction between primary user and secondary user transmissions as a complete four-state Markov Decision Process. The utilization performance of the secondary user was compared with other existing models such as Incomplete Markov model and Partial Observable Markov Decision Process. It is further extended by the performance analysis on the secondary user 's utilization against wastage and interference of the channel of interest and compared with the Incomplete Markov model.

Throughout this timespan of the conducted research for this dissertation, we heavily relied upon verifying the analytical results and formulas against computer simulations using Monte-Carlo technique. We addressed practical numerical examples to reflect the usefulness of the stated methodologies. The vast majority of the work presented in this thesis was published in-part or as a whole in peer-reviewed journals, conference proceedings, book chapters, or otherwise currently undergoing a peer review process.

7.2 Future Work

We shall draw out some future research paradigms based on the work accomplished in this thesis, with possible extension in analysis of decision making in cognitive radio networks. We shall discuss the future work with reference to the research objectives mentioned in this thesis.

We begin with *research objective 1* where the possibility of extending the work done on sensing to Rician fading channels for aerial-to-terrestrial communications in a theoretical perspective for sample sizes greater than 2. Additionally, learning of PU transmissions can be looked upon at the beginning and end of time slots as a Markov model and hence understand the temporal behaviour of the transmission itself, or look upon many PUs with transmissions of fixed durations. We can incorporate a Reinforcement learning to determine the impact of multiple users on the system.

Secondly, we look upon the possibility of extending *research objective 2* by considering cooperative games and auction games to provide a framework for channel resilience in the system. Furthermore, we could extend the possibility of multiple authorized users and unauthorized users with single channel and multiple channel access.

Thirdly, we can address the extension of *research objective 3* by considering a Markov decision process for the multi-user environment rather than a single one trying to access one channel. Later, we can address a Markov model of a single agent model trying to access multiple channels at a time.

Bibliography

[1] 3GPP. "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation (3GPP TS 36.211)". In: *http://www.3gpp.org/LTE* 10 ().

[2] 3GPP. "LTE Technical Specifications". In: *http://www.3gpp.org/LTE* ().

[3] S. Chandrasekharan A. Al-Hourani V. Trajkovic and S. Kandeepan. "Spectrum occupancy measurements for different urban environments". In: *Networks and Communications (EuCNC)* (2015), pp. 97–102.

[4] R. S. Thakur A. Khare M. Saxena and K. Chourasia. "Attacks & Preventions of Cognitive Network - A Survey". In: *IJARCET* 2.3 (2013), pp. 1002–1006.

[5] R. S. Thakur A. Khare M. Saxena and K. Chourasia. "Attacks & Preventions of Cognitive Networks - A Survey". In: *IJARCET* 2.3 (2013), pp. 1002–1006.

[6] A. Giorgetti A. Mariani S. Kandeepan and M. Chiani. "Cooperative Weighted Centroid Localization for Cognitive Radio Networks". In: *IEEE ISCIT* (Oct. 2012).

[7] ABSOLUTE. "EU-FP7 IP Project ABSOLUTE". In: *www.absolute-project.eu* (2008).

[8] A.Ghasemi and E. Sousa. "Optimization of spectrum sensing for opportunistic spectrum access in cognitive radio networks". In: *Proc. IEEE Consumer Commun. and Networking Conf.* (Jan. 2007), pp. 1022–1026.

[9] S. Kandeepan et. al. "Experimentally Detecting IEEE 802.11n Wi-Fi Based on Cyclostationarity Features for Ultra-Wide Band Cognitive Radios". In: *IEEE PIMRC* (2009).

[10] S. Kandeepan et. al. "Periodic Spectrum Sensing Performance and Requirements for Detecting Legacy Users with Temporal and Noise Statistics in Cognitive Radios". In: *IEEE BWA-WS Globecom* (Dec. 2009).

[11] 4G Americas. "Understanding 3GPP Release 12 Standards". In: *IEEE CrownComm* (Feb. 2015).

[12] K. J. Ray Liu B. Wang Y. Wu and T. C. Clancy. "An anti-Jamming Stochastic Game for Cognitive Radio networks". In: *IEEE COMM* 29.4 (2011), pp. 1–14.

[13] Y. Wu B. Wang and K. J. Ray Liu. "Optimal Power Allocation Strategy against Jamming attacks using the Colonel Blotto Game". In: *IEEE Globecom Proceedings* (2009), pp. 1–5.

[14] C. Berge. "Sur une convexite reguliere et ses applications la theorie des jeux". In: *Bull Soc* 82 (1954), pp. 301–319.

[15] G. W. Brown and J. von Neumann. "Solutions of games by differential equations". In: *Contributions to the Theory of Games I, Annals of Mathemathical Studies* 24 (1950), pp. 73–79.

[16] G.W. Brown. "Iterative Solutions of Games by Fictitious Play In Activity Analysis of Production and Allocation". In: *T.C. Koopmans (Ed.), Wiley-Activity Analysis of Production and Allocation* ().

[17] Y. Wu B.Wang and K.J. Ray Liu. "Game theory for cognitive radio networks : An overview". In: *Computer Networks Elseiver* 54 (2010), pp. 2537–2561.

[18] C. Xin C. Chen M. Song and J. Backens. "A game-theoretical anti-jamming scheme for cognitive radio networks". In: *IEEE Networks* 27.3 (2013), pp. 22–27.

[19] C. Chen et al. "A Game-theoretical Anti-jamming Scheme for Cognitive Radio Networks". In: *IEEE Networks* 27.3 (2013), pp. 22–27.

[20] T.C. Clancy and N. Goergen. "Security in Cognitive Radio Networks: Threats and Mitigation". In: *IEEE CrownComm* (2008), pp. 1–8.

[21] Federal Communications Commission. "Facilitating Opportunities for Flexible, Efficient, and Reliable Spectrum Use Employing Cognitive Radio Technologies". In: *NPRM and Order* 23 (Dec. 2003), pp. 03–322.

[22] S. Mishra D. Cabric and R. Brodersen. "Implementation issues in spectrum sensing for Cognitive Radios". In: *Proc. Asilomar Conf. on Signals, Systems and Computers* 1 (Nov. 2004), pp. 772–776.

[23] et al. D. Guoru. "Joint exploration and exploitation of spatial-temporal spectrum hole for cognitive vehicle radios". In: *IEEE International Conference on Signal Processing* (2011), pp. 1–4.

[24] K. Avrachenkov E. Altman and A. Garnaev. "Fair resource allocation in wireless networks in the presence of a jammer". In: *IEEE Performance Evaluation* 67.4 (2010), pp. 338–349.

[25] M. K. Simon F. Digham M. Alouini. "On the Energy Detection of Unknown Signals over Fading Channels". In: *IEEE ICC* 5 (May 2003), pp. 3575–3579.

[26] G. Mildh S. Parkvall N. Reider G. Miklos G. Fodor E. Dalman and Z. Turanyi. "Design Aspect of Network Assisted Device-to-Device Communications". In: *IEEE Wireless Commun.* 50.3 (Mar. 2012), pp. 170–177.

[27] G. Ganesan and Y. Li. "Agility improvement through cooperative diversity in Cognitive Radio". In: *Proc. IEEE Globecom* (Nov. 2005), pp. 2505–2509.

[28] L. Goratti et al. "A Novel Device-to-Device Communication Protocol for Public Safety Applications". In: *IEEE* (2013).

[29] S. Haykin. "Cognitive radio: brain-empowered wireless communications". In: *IEEE Journal on Selected Areas of Communications* 23 (2005), pp. 201–220.

[30] D. Hlavacek and J. M. Chang. "A Layered Approach to Cognitive Radio Network Security : A Survey". In: (2013).

[31] J. Hofbauer and S. Sorin. "Best response dynamics for continuous zero-sum games". In: *Discrete and continuous dynamical systems - Series B* 6.1 (2006), pp. 215–224.

[32] V. Koivunen J. Lunden S.R. Kulkarni and H.V. Poor. "Multi-agent Reinforcement Learning Based Spectrum Sensing Policies for Cognitive Radio Networks". In: *Selected Topics in Signal Processing* 7.5 (2013), pp. 858–868.

[33] J. Lunden J. Okansen and V.Koivunen. "Reinforcement Learning based sensing policy for energy efficient cognitive radio networks". In: *NeuroComputing, ELSEVIER* (Mar. 2012).

[34] J.Mitola and G. Maguire Jr. "Cognitive Radio: Making Software Radios more Personal". In: *IEEE Personal Comms* 6.4 (Aug. 1999), pp. 13–18.

[35] C. Wijting C. B. Ribeiro K. Doppler M. Rinne and K. Hugl. "Device-to-Device Communication as an Uderlay to LTE-Advanced Networks". In: *IEEE Wireless Commun.* 47.12 (Dec. 2012), pp. 42–49.

[36] G. Noubir K. Firouzbakht and M. Salehi. "On the capacity of rate-adaptive packetized wireless communication links under jamming". In: *Proceedings of the ACM WISEC Conference* (2012).

[37] P. Komisarczuk K. Yau and P. Teal. "Reinforcement Learning for context awareness and intelligence in wireless networks: Review, new features and open issues". In: *Journal of Network and Computer Applications, ELSEVIER* (Jan. 2012).

[38] S. Kandeepan and A. Giorgetti. "Cognitive Radio Techniques – Spectrum Sensing, Interference Mitigation and Localization". In: *Cognitive Radio Techniques – Spectrum Sensing, Interference Mitigation and Localization* (2013).

[39] S. Kay. "Intuitive Probability and Random Processes using MATLAB". In: *Springer, New York* (2006).

[40] M. Littman L. Kaelbling and A. Moore. "Reinforcement Learning: A Survey". In: *Journal of Artificial Intelligence Research* 4 (May 1996), pp. 237–285.

[41] B. Lo and I. Akyildiz. "Reinforcement Learning-based Cooperative Sensing in Cognitive Ad Hoc Networks". In: *IEEE PIMRC* (Sept. 2010).

[42] M. J. Abdel-Rahman M. K. Hanawal and M. Krunz. "Game theoretic anti-jamming dynamic frequency hopping and rate adaptation in wireless systems". In: *Proceedings of the WiOpt Conference* (2014).

[43] M. J. Abdel-Rahman M. K. Hanawal and M. Krunz. "Joint adaptation of frequency hopping and transmission rate for anti-jamming wireless systems". In: *IEEE Transactions on Mobile Computing* (2015).

[44] Y. Li M.Bkassiny and S. Jayaweera. "A Survey on Machine-Learning Techniques in Cognitive Radios". In: *IEEE Commun. Surveys and Tutorials* 99 (Oct. 2012), pp. 1–24.

[45] T. Alpcan M.Bloem and T. Basar. "A Stackelberg Game for Power Control and Channel". In: *GAMECOMM* (2007).

[46] K. Yau Mee Hong Ling and L.A. “Reinforcement Learning-Based Trust and Reputation Model for Spectrum Leasing in Cognitive Radio Networks”. In: *IT Convergence and Security (ICITCS)* (2013), pp. 1–6.

[47] L.A. Mee Hong Ling Yau. K. “Reinforcement learning-based trust and reputation model for cluster head selection in cognitive radio networks”. In: *Internet Technology and Secured Transactions (ICITST)* (2014), pp. 256–261.

[48] R.B. Myerson. “Game Theory: Analysis of Conflict”. In: *Harvard University Press* (1997).

[49] J. V. Neumann and O. Morgenstern. “Theory of Games and Economic Behavior”. In: *Princeton University Press* (1947).

[50] J. Von Neumann. “Fair resource allocation in wireless networks in the presence of a jammer”. In: *Zur Theorie der Gesellschaftsspiele Math. Annalen.* 100 (1928), pp. 295–320.

[51] G. J. Janssen P. Pawelczak and R. V. Prasad. “Performance measures of dynamic spectrum access networks”. In: *Proc. IEEE Globecom* (Dec. 2006).

[52] W. Jun P. Qihang Z. Kun and L. Shaoqian. “A distributed spectrum sensing scheme based on credibility and evidence theory in cognitive radio context”. In: *Proc. IEEE Consumer Commun. and Networking Conf.* (Jan. 2007), pp. 1022–1026.

[53] B. Raghothaman et al. “Architecture and Protocols for LTE-based Device to Device Communication”. In: *Computing, Networking and Communications (ICNC), 2013 Intl. Conf. on*. 2013, pp. 895–899.

[54] Y. B. Reddy. “Security Issues and Threats in Cognitive Radio Networks”. In: *AICT* (2013).

[55] J. Robinson. “An Iterative Method of Solving a Game”. In: *Annals of Mathemathical Studies* 54 (1951), pp. 296–301.

[56] L. Maggi-F. de Pellegrini S. Arunthavanathan L. Goratti and S. Kandeepan. “On the Achievable Rate in a D2D Cognitive Secondary Network Under Jamming Attacks”. In: *IEEE Crowncomm* (June 2014), pp. 39–44.

[57] S. Kandeepan S. Arunthavanathan and R. Evans. "Reinforcement learning based secondary user transmissions in cognitive radio networks". In: *Globecom Workshops (IEEE GC Wkshps)* 6.1 (2013), pp. 374–379.

[58] S. Kandeepan S. Arunthavanathan and R. Evans. "Spectrum Sensing and Detection of Incumbent-UEs in Secondary-LTE based Aerial-Terrestrial Networks for Disaster Recovery". In: *IEEE CAMAD* (2013), pp. 201–206.

[59] and H. Jiang S. Atapattu C. Tellambura. "Performance of an Energy Detector over Channels with both Multipath Fading and Shadowing". In: *IEEE Trans. Wireless Comms.* (Jan. 2010), pp. 3662–3670.

[60] A. M. Wyglinski S. Chen R. Vuyyuru and O. Altintas. "On Optimizing Vehicular Dynamic Spectrum Access Networks: Automation and Learning in Mobile Wireless Environments". In: *IEEE Vehicular Networking Conference (VNC)* (2011).

[61] F. Althoff S. Hoch M. Schweigert and G. Rigoll. "The BMW surf project: A contribution to the research on cognitive vehicles". In: *Proc. Intell. Veh. Symp.,* (2007), pp. 692 –697.

[62] A. Giorgetti S. Kandeepan and M. Chiani. "Periodic Spectrum Sensing Performance and Requirements for Legacy Users with Temporal and noise Statistics in Cognitive Radios". In: *IEEE GLOBECOM Workshops* (Dec. 2009), pp. 1–4.

[63] D. Lowe T.C. Aysal S. Kandeepan R. Piesiewicz and S. Reisenfield. "Bayesian Tracking in Cooperative Localization for Cognitive Radio Networks". In: *IEEE VTC* (Dec. 2009), pp. 26–29.

[64] M. G. Benedetto S. Kandeepan L. Nardis and Alessandro G. Corazza. "Cognitive Satellite Terrestrial Radios". In: *IEEE Globecom* (Dec. 2010).

[65] T. Aysal A. Biswas S. Kandeepan R. Piesiewicz and I. Chlamtac. "Spectrum Sensing for Cognitive Radios with Primary User Transmission Statistics: with Linear Frequency Sweeping". In: *EURASIP Journal or Wireless and Communication Networks* 6 (Jan. 2010).

[66] T. Rasheed S. Kandeepan G. Karina and R. Laurent. "Energy Efficient Cooperative Strategies in Hybrid Aerial-Terrestrial Networks for Emergencies". In: *IEEE PIMRC* (Sept. 2011).

[67] I. Toufik S. Sesia and M. Baker. “The UMTS Long Term Evolution From Theory to Practice”. In: *Wiley 2nd edition* (Feb. 2009).

[68] C. Cordeiro S. Shankar and K. Challapali. “Spectrum agile radios: utilization and sensing architectures”. In: *Proc. IEEE Int. Symposium on New Frontiers in Dynamic Spectrum Access Networks* (Nov. 2005), pp. 160–169.

[69] Sangeeta Singh and Aditya Trivedi. “Anti-jamming in Cognitive Radio Networks Using Reinforcement Learning Algorithms”. In: *IEEE Journal on Selected areas in Communciations* (Sept. 2012), pp. 1–5.

[70] M. Sion. “On general minimax theorems”. In: *Pacific Journal of Mathemathics* 8.1 (1958), pp. 301–319.

[71] S.Sodagari and T. Charles Clancy. “An Anti-jamming Strategy for Channel Access in Cognitive Radio Networks”. In: *IEEE ICNP* (2011).

[72] S. Kandeepan T. Aysal and P. Radoslow. “Cooperative Spectrum Sensing with Noisy Hard Decision Transmissions”. In: *International Conference on Communications (ICC)* (June 2009).

[73] S. Tatesh A. Casati G. Tsirtsis K. Anchan T. Doumi M. F. Dolan and D. Flore. “LTE for Public Safety Networks”. In: *IEEE Wireless Commun.* 51.2 (Feb. 2013), pp. 106–112.

[74] Y. C. Liang T. J. Lim R.Zhang and Y. Zeng. “GLRT-Based Spectrum sensing for Cognitive Radio”. In: *Proc. IEEE Globecom* (2008).

[75] H. Tembine. “Dynamic bargaining solutions for opportunistic spectrum access”. In: *Wireless Days (WD), 2nd IFIP,* (2009), pp. 1–6.

[76] M. Schaar U. Berthold F. Fu and F. Jondral. “Detection of Spectral Resources in Cognitive Radios using Reinforcement Learning”. In: *IEEE DySPAN* (Oct. 2008), pp. 1–5.

[77] H. Urkowitz. “Energy Based Detection of Unknown Deterministic Signals”. In: *IEEE Proceedings* 55.4 (Apr. 1967), pp. 523–531.

[78] L. Quan W. Kehao C. Lin and K. Al Agha. “On Optimality of Myopic Sensing Policy with Imperfect Sensing in Multi-Channel Opportunistic Access”. In: *IEEE Transactions on Communications* 61.9 (2013), pp. 3854–3862.

[79] B. Wang, Y. Wu, and K. J. Ray Liu. “Optimal Power Allocation Strategy against Jamming Attacks using the Colonel Blotto Game”. In: *IEEE GLOBECOM Proc.* (2009), pp. 1–5.

[80] F. Wang. “Cognitive Radio Networks and Security: A Survey”. In: *Journal of Network and Computer Applications* 35.6 (Nov. 2012), 1691–1708.

[81] C. Watkins. “Learning from delayed rewards,” Ph.D. dissertation”. In: *University of Cambridge* (1989).

[82] C. Kae Won and E. Hossain. “Opportunistic Access to Spectrum Holes Between Packet Bursts: A Learning-Based Approach”. In: *IEEE Transactions on Wireless Communications* 10.8 (2011), pp. 2497 –2509.

[83] D. Lili X. Qin L. Jiping and L. Shouyin. “Optimization of PHY-layer sensing and Mac-layer access based on adaptive threshold schedule for opportunistic spectrum access”. In: *Signals Systems and Electronics (ISSSE)* (2010), pp. 1–4.

[84] R. Yang Y. Li and F. Ye. “Non-Cooperative Spectrum Allocation based on Game Theory in Cognitive Radio Networks”. In: *IEEE BIC-TA* (2010), pp. 1134–1137.

[85] F. Niu C. Dai Y. Teng Y. Zhang and M. Song. “Reinforcement Learning Based Auction Algorithm for Dynamic Spectrum Access in Cognitive Radio Networks”. In: *Proceedings of the IEEE 72nd Vehicular Technology Conference* (2010), pp. 1–5.

[86] K.J. Ray Liu Y. Wu B. Wang and T. C. Clancy. “Anti-jamming Games in Multichannel Cognitive Radio Networks”. In: *IEEE GLOBECOM Proc.* 30.1 (2012), pp. 1–12.

[87] R. Chandr P. A. Chou J. I. Ferrell T. Moscibroda S. Narlanka Y. Yuan P. Bahl and Y.Wu. “KNOWS:Cognitive radio networks over white spaces”. In: *Proc . IEEE Int. Symposium on New frontiers in Dynamic Spectrum Access Networks* (2007), pp. 416–427.

[88] Carl Fossa Youngjune Gwon Siamak Dastangooand and H. T. Kung. “Competing Mobile Network Game: Embracing Antijamming and Jamming Strategies with Reinforcement Learning”. In: *IEEE Conference on Communications and Network Security (CNS)* (Feb. 2013).

[89] T. Yucek and H. Arslan. “A Survey of Spectrum Sensing Algorithms for Cognitive Radio Applications”. In: *IEEE Comms* 11.1 (Mar. 2009), pp. 116–130.

[90] T. Yucek and H. Arslan. “Spectrum Characterization for opportunistic cognitive radio systems”. In: *Proc. IEEE Military Commun. Conf.* (Oct. 2006), pp. 1–6.

[91] Bin Zhou et al. “Intracluster Device-to-Device Relay Algorithm with Optimal Resource Utilization”. In: *IEEE Trans. on Veh. Technol.* 62.5 (2013), pp. 2315–2326.