Privacy-preserving spectrum sharing

Azza Ben-Mosbah

THÈSE DE DOCTORAT CONJOINTE avec
TÉLÉCOM SUDPARIS et l'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité

Télécommunications

École doctorale Informatique, Télécommunications et Électronique (Paris)

Présentée par

Azza BEN-MOSBAH

Pour obtenir le grade de

DOCTEUR de TÉLÉCOM SUDPARIS

Sujet de la thèse :

Un Partage de Spectre Préservant la Confidentialité

soutenue le 24 Mai 2017

devant le jury composé de :

Mme. Véronique Vèque	Rapporteuse	Université Paris-Sud
M. Jalel Ben-Othman	Rapporteur	Université Paris 13
Mme. Samia Bouzefrane	Examinatrice	CNAM, CEDRIC Lab
M. Yvon Gourhant	Examinateur	Orange Labs
M. Hossam Afifi	Directeur de thèse	Télécom SudParis
M. Timothy A. Hall	Co-directeur de thèse	NIST, CTL

Thèse N° 2017TELE0008

Université Pierre et Marie Currie
Institut Mines-Télécom
Télécom SudParis

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy in Telecommunications

Subject

Privacy-Preserving Spectrum Sharing

Presented By
Azza Ben-Mosbah

Thesis Committee

Prof. Véronique VÈQUE	Reviewer	University Paris-Sud
Prof. Jalel BEN-OTHMAN	Reviewer	University Paris 13
Prof. Samia BOUZEFRANE	Examiner	CNAM, CEDRIC Lab
Dr. Yvon GOURHANT	Examiner	Orange Labs
Prof. Michel MAROT	Invitee	Télécom SudParis
Prof. Hassine MOUNGLA	Invitee	University Paris Descartes
Prof. Hossam AFIFI	Thesis Director	Télécom SudParis
Dr. Timothy A. HALL	Thesis Co-director	NIST, CTL

Abstract

Radio frequencies, as currently allocated, are statically managed. Spectrum sharing between commercial users and incumbent users in the Federal bands has been considered by regulators, industry, and academia as a great way to enhance productivity and effectiveness in spectrum use. However, allowing secondary users to share frequency bands with sensitive government incumbent users creates new privacy threats in the form of inference attacks. Therefore, the goal of this thesis is to enhance the privacy of the incumbent while allowing secondary access to the spectrum. First, we present a brief description of different sharing regulations and privacy requirements in Federal bands. We also survey the privacy-preserving techniques (i.e., obfuscation) proposed in data mining and publishing to thwart inference attacks. Next, we propose and implement our approach to protect the operational frequency and location of the incumbent operations from inferences. We follow with research on frequency protection using inherent and explicit obfuscation to preserve the incumbent's privacy. Then, we address location protection using trust as the main countermeasure to identify and mitigate an inference risk. Finally, we present a risk-based framework that integrates our work and accommodates other privacy-preserving approaches. This work is supported by models and simulations which provide results that showcase our work, quantify the importance of evaluating privacy-preserving techniques and analyze the trade-off between privacy protection and spectrum efficiency.

Key words – Spectrum sharing, Federal incumbent, commercial secondary users, inference attack, operational security, privacy protection, obfuscation, trust management, spectrum efficiency.

Résumé

Les bandes des fréquences, telles qu'elles sont aménagées aujourd'hui, sont statiquement allouées. Afin d'améliorer la productivité et l'efficacité de l'utilisation du spectre, une nouvelle approche a été proposée : le « partage dynamique du spectre ». Les régulateurs, les industriels et les scientifiques ont examiné le partage des bandes fédérales entre les détenteurs de licences (utilisateurs primaires) et les nouveaux entrants (utilisateurs secondaires). La nature d'un tel partage peut faciliter les attaques d'inférence et mettre en péril les paramètres opérationnels des utilisateurs primaires. Par conséquent, le but de cette thèse est d'améliorer la confidentialité des utilisateurs primaires tout en permettant un accès secondaire au spectre. Premièrement, nous présentons une brève description des règles de partage et des exigences en termes de confidentialité dans les bandes fédérales. Nous étudions également les techniques de conservation de confidentialité (offuscation) proposées dans les domaines d'exploration et d'édition de données pour contrecarrer les attaques d'inférence. Ensuite, nous proposons et mettons en oeuvre notre approche pour protéger la fréquence et la localisation opérationnelles contre les attaques d'inférence. La première partie étudie la protection de la fréquence opérationnelle en utilisant une offuscation inhérente et explicite pour préserver la confidentialité. La deuxième partie traite la protection de la localisation opérationnelle en utilisant la confiance comme principale contre-mesure pour identifier et atténuer un risque d'inférence. Enfin, nous présentons un cadre axé sur les risques qui résume notre travail et s'adapte à d'autres approches de protection de la confidentialité. Ce travail est soutenu par des modèles, des simulations et des résultats qui focalisent sur l'importance de quantifier les techniques de préservation de la confidentialité et d'analyser le compromis entre la protection de la confidentialité et l'efficacité du partage du spectre.

Mots Clés – Partage du spectre, détenteur de licence fédéral, utilisateurs secondaires commerciaux, attaque d'inférence, sécurité opérationnelle, protection de la confidentialité, offuscation, gestion de confiance, efficacité du spectre.

Thesis Publications

International Journals

- **Azza Ben-Mosbah**, Timothy A. Hall, Michael Souryal, and Hossam Afifi, *Protecting the Incumbent's Frequency Against Inference Attacks in Spectrum Sharing*, Submitted to the IEEE Transactions on Cognitive Communications and Networking (TCCN).

International Conferences

- **Azza Ben-Mosbah**, Timothy A. Hall, Michael Souryal, and Hossam Afifi, *An Analytical Model for Inference Attacks on the Incumbent's Frequency in Spectrum Sharing*, the IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), March 2017. "Best Poster Award"

- **Azza Ben-Mosbah**, Timothy A. Hall, Michael Souryal, and Hossam Afifi, *Analysis of the Vulnerability of the Incumbent Frequency to Inference Attacks in Spectrum Sharing*, the 14^{th} Annual IEEE Consumer Communications & Networking Conference (CCNC), January 2017.

Acknowledgements

I would like to thank the people who have directly or indirectly supported me and contributed to my work during the years of my Ph.D studies.

Firstly, I would like to express my sincere gratitude to my supervisor and thesis director Prof. Hossam Afifi (full professor of Télécom SudParis) for offering me this opportunity. His knowledge and expertise have guided me all the way through my research and writing of this thesis.

I would like to thank as well my co-supervisor Dr. Timothy A. Hall (R&D engineer of National Institute of Standards and Technology – NIST) for his motivation, dedication, patience, and professionalism. His experience and advice have allowed me to improve the quality of my work. And I would like to acknowledge his valuable contributions and inputs to this thesis.

Besides, I would like to thank the members of the dissertation jury for accepting my invitation. I am particularly honored by their presence. I thank the dissertation reviewers Prof. Véronique Vèque (full professor of University Paris-Sud) and Prof. Jalel Ben-Othman (full professor of University Paris 13) for their insightful comments and helpful feedback. I also thank the dissertation examiners Prof. Samia Bouzefrane (associate professor of Cedric Lab from CNAM) and Dr. Yvon Gourhant (research director of Orange Labs) for their time and flexibility. I thank Prof. Michel Marot (full professor of Télécom SudParis) and Prof. Hassine Moungla (associate professor of University Paris Descartes) as well for agreeing to serve on the committee on short notice.

Many thanks go to the staff of Télécom SudParis, Télécom ParisTech and EDITE de Paris, particularly Mrs. Xayplathi Lyfoung, Mrs. Sandra Gchweinder and Mrs. Véronique Guy for their help with the administrative procedures. I thank Prof. Houda Labiod of Télécom ParisTech for her efforts as well.

My special thanks go to the Wireless Networks Division chief Dr. Nada T. Golmie and the Spectrum Sharing Project manager Dr. Michael R. Souryal, who provided me the possibility to join their team at NIST, and who gave me access to their laboratory facilities. Without their support, it would not be possible to conduct this research.

I am also indebted to all my friends and my colleagues in the Wireless Networks Division and within NIST for being supportive and thoughtful. During my stay in the United States, their knowledge and advices have contributed to this thesis in so many ways. In particular, I am thankful to Dr. Antonio Izquierdo Manzanares for his guidance and Mr. David E. Cypher for his generosity.

I would like to thank my twin sister and partner in crime Aziza Ben-Mosbah for all the fun and for all the not-so-fun we have had in the last four years. I would like also to thank my husband and best friend Aymen Khabthani for his continuous understanding and endless love. This accomplishment would not have been possible without their consideration and encouragements. And I could not have asked for a better support system.

Last but certainly not least, I would like to thank my parents and younger sister for their continuous support through all of the ups and downs of my research, and my extended family and friends for expecting nothing less than excellence from me.

To **Afifa Ben-Mosbah**, may Allah bestow his mercy and grace on her soul.

To **Moufida Demni** and **Faical Ben-Mosbah**, my beloved parents.

To **Nahla Ben-Mosbah**, my sweet baby sister.

To **Aziza Ben-Mosbah**, my twin sister and better half.

To **Aymen Khabthani**, my caring husband and the love of my life.

To the **Ben-Mosbah family**, the **Demni family**, and the **Khabthani family**.

Thank you for your continued support, kindly prayers and unconditional love.
No words can express my appreciation and gratitude.
Love you all.

Azza Ben-Mosbah Khabthani

Contents

List of Figures

List of Tables

List of Abbreviations

-	
3GPP	**3**rd **G**eneration **P**artnership **P**roject
5G	**5**th **G**eneration of Mobile Networks
A	
APT	**A**sian-**P**acific **T**elecommunity
AWG	**A**PT **W**ireless **G**roup
AWS	**A**vanced **W**ireless **S**ervices
C	
CBRS	**C**itizens **B**roadband **R**adio **S**ervice
CBSD	**C**itizens **B**roadband **S**ervice **D**evice
CIA	**C**onfidentiality **I**ntegrity **A**vailability
CRTC	**C**anadian **R**adio-television and **T**elecommunications **C**ommission
CSMAC	**C**ommerce **S**pectrum **M**anagement **A**dvisory **C**ommittee
CTIA	**C**ellular **T**elecommunications **I**ndustry **A**ssociation
CUS	**C**ollective **U**se of **S**pectrum
D	
DARPA	**D**efense **A**dvanced **R**esearch **P**rojects **A**gency
DoD	**D**epartment **o**f **D**efense
DSA	**D**ynamic **S**pectrum **A**ccess
E	
EARS	**E**nhancing **A**ccess to the **R**adio **S**pectrum
EIRP	**E**quivalent **I**sotropically **R**adiated **P**ower
ESC	**E**nvironmental **S**ensing **C**apability
ETSI	**E**uropean **T**elecommunications **S**tandards **I**nstitute
F	
FCC	**F**ederal **C**ommunications **C**ommission
FSS	**F**ixed **S**atellite **S**ervice
G	
GAA	**G**eneral **A**uthorized **A**ccess
GDB	**G**eolocation **D**ata**B**ase

I

IA	**I**ncumbent **A**ccess
IDPS	**I**ntrusion **D**etection and **P**revention **S**ystems
IoT	**I**nternet **o**f **T**hings

L

LBS	**L**ocation-**B**ased **S**ervice
LSA	**L**icensed **S**hared **A**ccess
LTE	**L**ong-**T**erm **E**volution

M

MENA	**M**iddle **E**ast and **N**orth **A**frica

N

NSC	**N**ational **S**pectrum **C**onsortium
NeTS	**N**etworking **T**echnology and **S**ystems
NSA	**N**on-**S**ensitive **A**ttribute
NSF	**N**ational **S**cience **F**oundation
NTIA	**N**ational **T**elecommunications and **I**nformation **A**dministration

O

OFCOM	**O**ffice of **C**ommunications

P

PA	**P**riority **A**ccess
PCAST	**P**resident's **C**ouncil of **A**dvisors on **S**cience and **T**echnology
PU	**P**rimary **U**ser

Q

QID	**Q**uasi-**Id**entifier
QoS	**Q**uality **o**f **S**ervice

R

RF	**R**adio **F**requency

S

SA	**S**ensitive **A**ttribute
SAS	**S**pectrum **A**ccess **S**ystem
SINR	**S**ignal-to-**I**nterference-plus-**N**oise **R**atio
SNR	**S**ignal-to-**N**oise **R**atio
SpecEES	**Spec**trum **E**fficiency, **E**nergy **E**fficiency, and **S**ecurity
SSC	**S**pectrum **S**haring **C**ommittee
SU	**S**econdary **U**ser

T

TVWS	**TV W**hite **S**pace

U

UHF	**U**ltra **H**igh **F**requency

V

VHF	**V**ery **H**igh **F**requency

W

WInnForum	**W**ireless **Inn**ovation **Forum**

Chapter 1

Introduction

1.1 Motivations of Sharing in Federal Bands

Frequency bands are statically allocated, but inefficiently utilized. Sharing has been considered to address this concern [1], for example, in unlicensed bands. Sharing in under-utilized government bands has been encouraged, as well [2]. In this chapter, we will review how this is becoming an international trend, and we will introduce the contributions to the area presented in this thesis.

1.1.1 Wireless and Mobile Usage

The fast growth of wireless broadband services has created an increased demand for spectrum. A spectrum resource is defined as the combination of allowed frequency, time and power for use over a given region. In fact, the Cellular Telecommunications Industry Association (CTIA) reported that the annual data usage in the United States has increased from 3.2 Trillion Megabytes in 2013 to 4 Trillion Megabytes in 2014 to 9.6 Trillion Megabytes in 2015, and that 60 % of this traffic is dedicated to video [3]. This shows how data consumption has doubled in one year, and how this significant rise is expected to continue for the next decade and beyond. Also, according to the President's Council of Advisors on Science and Technology (PCAST), the number of devices connected to mobile networks worldwide will be multiplied by 10 (from 5 billion to 50 billion) by 2020 [4]. Given that the population in the world is expected to reach more than 7.5 Billion people in 2020 [5], each person will own, on average, 6 to 7 mobile devices including smartphones, tablets, wearables, etc.

To accommodate such an increase in usage and devices and improve the wireless and mobile service, operators and carriers are aiming to invest more. However, investing in a congested spectrum is nor efficient neither profitable. The radio spectrum is currently a scarce resource, but measurements show that it is not been fully exploited. While sharing is already encouraged in unlicensed bands, the need for spectrum has not been satisfied yet. Licensees such as military are encouraged to optimize their access to spectrum and share their bands in order to expand the commercial use and boost the innovation.

1.1.2 American Adoption

In the United States, the Defense Advanced Research Projects Agency (DARPA) of the Department of Defense (DoD) has funded a neXt Generation (XG) communications program to enable the army to dynamically access the spectrum instead of using pre-allocated frequency bands. DARPA has also been hosting multiples challenges, that aim to develop and test methods and techniques for dynamic spectrum access using cognitive radio, artificial intelligence and machine learning [6]. The DARPA spectrum challenge is one of the events held by the DoD agency and includes multiple competitors looking for an efficient way to leverage the spectrum [7]. Its goal is to assure the seamlessness of military operations in the presence of other interfering signals and to improve the coexistence between Federal and non-Federal users.

The National Science Foundation (NSF) has also shown interest in intelligent spectrum management. It has launched the Networking Technology and Systems (NeTS) program, that encourages the research on incentives to unleash the spectrum, sharing techniques in cognitive radio networks, spectrum measurements and secondary spectrum ecosystems. The ongoing investments are called the Enhancing Access to the Radio Spectrum (EARS) program [8] and the Spectrum Efficiency, Energy Efficiency, and Security (SpecEES) program [9]. Such programs aim to boost the efficiency of the spectral allocation and the public access to the spectrum. However, NSF is still seeking proposals to optimize access to radio resources and enable sharing of unused bands by awarding $12 Million for spectrum research [10]. In addition, the National Spectrum Consortium (NSC) is launching collaborations between regulatory, industry and academia to

leverage the spectrum usage. The NSC is also awarding projects developing advanced technologies to improve the military and commercial access to the spectrum [11].

In order to satisfy the increasing need for spectrum, both the National Telecommunications and Information Administration (NTIA) and the Federal Communications Commission (FCC) have tried to reallocate some bands from Federal use to non-Federal use [12]. However, the reallocation proved to be challenging and costly, which made alternative options to reallocation to be considered more favorably for future spectrum management. So, when sharing has been proposed as an alternative, the NTIA and the FCC have worked to identify potential Federal bands that can be opened for shared access between Federal and non-Federal users [13]. For example, 3550–3700 MHz band, also known as the 3.5 GHz band, was proposed for sharing between incumbents (Navy radars and fixed satellite service earth stations) and Citizens Broadband Radio Service (CBRD) users. Federal agencies are also investigating the sharing of additional bands, including 1675–1710 MHz (meteorology), 1755–1780 MHz (military telemetry and surveillance), 4200–4220 MHz (aeronautical radionavigation), 4380–4400 MHz (aeronautical radionavigation) [14]. Similarly, the 3rd Generation Partnership Project (3GPP) newest releases are looking at techniques for LTE spectrum sharing in the TV White Space (TVWS) band, the Federal 3.5 GHz band and the unlicensed 5 GHz band [15]. So, the 5 GHz band is also expected to be open for sharing. However, sharing in the 5 GHz band will operate differently as it does not include sensitive incumbents.

1.1.3 International Adoption

Spectrum sharing is not limited to the United States. Countries around the world are rethinking their current spectrum allocation. Joint efforts are studying sharing arrangements to maintain an efficient use of the spectrum.

The Canadian Radio-television and Telecommunications Commission (CRTC) has invited comments and views on under-utilized bands, promising technologies for sharing, approaches to maximize access to spectrum, and challenges of such deployment. The result of these discussions has led to the introduction of licence-exempt devices in the land mobile frequency sub-bands 462–467 MHz and to provide a transition plan regarding the 600 MHz band (repurposed from broadcast to mobile services). Recently, the CRTC

has encouraged the development of a framework for radio local area network devices operating in the 5150–5250 MHz band (primarily allocated for fixed-satellite service and mobile-satellite service).

In Europe, the European Commission has pushed its members to rethink the current spectrum allocation policies and advised them to adopt a new paradigm for spectrum management by regulating shared access to spectrum resources [16]. The EU Framework Programs have promoted the shared use of Europe's radio spectrum by identifying the need for more flexible spectrum allocation and initiating harmonization of spectrum usage. The European Telecommunications Standards Institute (ETSI) encourages a shared access to the spectrum under two models: licensed Long-Term Evolution (LTE) in the 2.3 GHz band and unlicensed LTE/Wi-Fi in the 5 GHz band. The 2.3 GHz band has quickly become the first candidate band for a sharing experiment. In France, this band is used by the Ministry of Defense for certain aeronautical telemetry applications. The transfer of these applications to other frequency bands is not envisioned in the short term. However, sharing initiatives have been suggested at the national level in order to contribute to the European proposal. Likewise, after expressing concern on the future of spectrum management and getting responses from stakeholders, OFCOM (Office of Communications), the regulator in the United Kingdom, has recently drawn a high-level framework to assess opportunities for spectrum sharing between market access (i.e., commercial users) and public sector users (i.e., government departments and agencies) [17]. This framework is encouraging identification of unused spectrum and implementation of new protocols and technologies for sharing. The 5 GHz band, for example, is currently used by Wi-Fi. OFCOM has embraced the European initiative and introduced the 5 GHz band for sharing with other technologies, mainly unlicensed LTE.

The Asia-Pacific Telecommunity (APT) is an inter-governmental organization for the Asia Pacific region. One of the goals of this organization is to promote revised frequency allocations and potential new or alternative spectrum uses. The APT Wireless Group (AWG) has sub-groups that operate in conjunction with service providers, equipment manufacturers, and research and development organizations. One of these sub-groups

conducts research on spectrum sharing options and possible deployments in the Asia-Pacific region.

The Middle East and North Africa (MENA) countries are also keeping up with the rest of the world. They are looking at the 470–694 MHz band, L-Band, 2.6 GHz band, and C-Band as sharing opportunities. However, they are still considering the benefits and challenges associated with each of the band plan options as border coordination may interfere with spectrum sharing regulations.

While it is true that dynamic spectrum sharing has gained a worldwide interest for a common cause, the approaches employed have differed. Two paradigms are considered: Collective Use of Spectrum (CUS) and Licensed Shared Access (LSA). CUS is a license-exempt approach that does not guarantee any protection for its users. Spectrum users employ their own techniques to identify available white spaces and operate on the unused bands while using the "Listen Before Talk" method. LSA is triggered by an industry proposal to open bands for access under an individual licensing regime (i.e., shared exclusive use). This approach requires a certain collaboration from the incumbent in order to accommodate the licensee.

European countries are more interested in the LSA paradigm. Regulators in the United States, however, are opting for the CUS paradigm even though industry is encouraging LSA paradigm, thereby introducing a hybrid shared access to the spectrum in the 3.5 GHz band called Spectrum Access System (SAS). It is a database-driven approach combined with spectrum sensing. This model includes three levels of priority: incumbent access, priority access, general authorized access. The incumbents have the highest priority and require full protection. The priority access users are similar to licensees in LSA (they require protection from lower tiers but not from the incumbent). The general authorized access users have opportunistic access to the spectrum but get no protection whatsoever from higher tiers. Note that, while LSA is a two-tier model, SAS is a three-tier model. Also, LSA assumes a communication protocol between the incumbent and the licensee, however, no communication is envisioned in SAS between the incumbent and any component of the sharing environment. SAS still promotes CUS by enabling third tier access to at least 50 MHz of the band.

1.1.4 Advantages

Dynamic Spectrum Access (DSA) is a novel approach to maximize the use of under-utilized bands [18]. The Commerce Spectrum Management Advisory Committee (CSMAC) [19] stated that, "DSA's promise is to improve spectrum utilization in three dimensions: frequency, location and time. It enables a network to opportunistically use any available channel (frequency) at points in time and space when and where they are not in use, and automatically move to another channel when policy demands it or a primary user/signal appears on the current channel. Because most RF [Radio Frequency] channels are utilized only a small portion of the time and in a fraction of locations, DSA enables two or more applications/networks to share a given band."

Spectrum sharing typically uses advanced cognitive radio technologies for spectrum sensing. Their flexibility allows users (in spectrum sensing approaches) and spectrum managers (in database-driven approaches) to identify and utilize white spaces in an efficient manner. In the United States, bands previously-controlled by Federal and governmental agencies are shown to be under-utilized and poorly-managed. In fact, incumbents do not function all the time, operate only in a certain geographic regions and occupy fewer resources than assigned.

Additionally, spectrum sharing is also enabling the development of different new technologies (e.g., 5G, IoT, etc). These technologies are employed in hundreds of millions of devices that need better access to the spectrum, including connectivity and Quality of Service (QoS), in an environment where the spectrum is already crowded. And as the number of wireless services increases, a fixed spectrum will include connection and transmission/reception failures.

Moreover, spectrum sharing is helping to introduce spectrum harmonization. This is a global effort to manage the frequency allocation beyond countries borders, upcoming technologies and spectrum prices. When spectrum is made available by regulators, large scale investments will provide a better use of the spectrum which brings social, political and economic benefits. Such management will enable affordable mobile services and boost future efforts in wireless innovation.

1.1.5 Challenges

If the past experience is anything to go by, different challenges are expected. In fact, the TVWS has shown to be a long and painful process. Engaging in the TVWS without careful consideration has led to multiples problems. The lack of preliminary tests and trials have not helped identify the best way to set certification rules. For example, the FCC proposed sharing in the TVWS band in 2004 [20], and it took seven years to approve the first TVWS database administrator and devices [21]. Nonetheless, with the new planning, TVWS devices will continue to emerge to share the Ultra High Frequency (UHF) bands, as well as the Very High Frequency (VHF) bands. Furthermore, the Advanced Wireless Services AWS-3 auction is making participants invest on the infrastructure as well as the license. And even though this auction has released 65 MHz of spectrum for use, the need to acquire additional licensed spectrum is still ongoing. Additional efforts are expected to open 500 MHz of the Federal spectrum for commercial use. Therefore, the lessons learned from these two experiences make future attempts to unleash more spectrum a challenging task.

If Federal spectrum sharing is to become reality, the operational requirements of the incumbent user must be met while allowing sufficient secondary access to the spectrum. Since a network is composed by two or more users, the relationships between them are important to achieve seamless connections and successful transmissions. An obvious requirement is that the incumbent must be protected from interference when it is operating. Incumbent receivers typically have an interference level or harm level. Any measured signal-to-interference-plus-noise ratio (SINR) above that threshold is harmful to the operations of the incumbents and is considered a violation of the sharing agreement.

Other considerations exist as well. Incumbents with critical missions are suspicious about letting secondary users access their bands. So, one of the most challenging spectrum sharing issues in Federal bands is security. The system should secure the shared spectrum from two different types of attackers [22]. The first type includes malicious users, who are external intruders (exogenous attackers) and attack to disrupt primary and secondary communications without the intention to maximize their profit in terms of spectrum opportunities. The second type includes selfish users (called also greedy

users) who are insiders (intra-network attackers) and attack to degrade the performance of the shared network with the intention to gain spectrum advantage and increase their own performance. Such attackers attempt to break confidentiality (i.e., disclosure of data to unauthorized systems), integrity (i.e., damage of the accuracy, consistency, and trustworthiness of data) and availability (i.e., denial of access to the system), all of which are serious violations of the CIA (confidentiality, integrity and availability) model. Some of those attacks are general rather than specific to spectrum sharing, but still valid for its applications. Other attacks are new and specific to spectrum sharing systems. A rich literature has grown around security issues in networking, and multiple algorithms have been developed. Some allow the detection of security violations, others introduce countermeasures to avoid a threat or at least mitigate its impacts.

In this context of spectrum sharing, privacy requirements (also known as operational security requirements) are as important as they are for other networks. Sensitive parameters of the incumbent include its location, operating frequency or channel, and time of operation. Hence, Federal incumbents require assurance that their privacy will not be jeopardized and their operational parameters will not be exposed. We take into consideration the importance of such requirements, specifically if the future of spectrum sharing for Federal bands depends on it.

1.2 Contributions

Our contributions consist of the analysis and evaluation of the vulnerability of the incumbent's operational frequency, the analysis and evaluation of the vulnerability of the incumbent's operational location, and the integration of all these works in a risk-based framework.

First, we study the privacy of the incumbent's operational frequency over time by evaluating its vulnerability and proposing tunable countermeasures to mitigate the risk of exposure [23]. Several channel assignment schemes (random, ordered and semi-static) are considered under different assumptions of the spectrum load, the total number of channels and the query rate of secondaries. We look at the effect of obfuscation on the results, considering both inherent obfuscation (i.e. achieved by adjusting the system

parameters), and explicit obfuscation (i.e. attained by intentionally leaving channels vacant).

Second, we evaluate the privacy of the operational frequency using an analytical model inspired by the "coupon collector problem" [24]. This analysis calculates the number of queries needed to infer a certain number of channels which includes the incumbent's channel. Using this model, the system is able to set up the maximum query rate of secondary users in order to mitigate an inference attack.

Third, we analyze the privacy of the incumbent's location. We show that previously-used methods (mainly obfuscation) can be efficient privacy-wise but introduce spectrum loss. So, we propose a trust-based algorithm to enhance the privacy of the operational location and avoid the loss of spectrum resources. The trustworthiness level of each secondary user determines whether it is able to gain additional knowledge about the spectrum or not. If the system concludes that a secondary user is not trustworthy enough to ask for access to the spectrum, it can be immediately terminated.

Finally, we summarize our work in a risk-based framework. It does not only offer countermeasures to overcome privacy-threatening actions, but also identify and assess such risks. It accommodates other privacy-preserving techniques and analyzes the trade-off between privacy gain and spectrum loss.

1.3 Outline

The remainder of this thesis is organized as follows. Chapter 2 describes spectrum access systems and defines the threat model for an inference attack, while also surveying privacy-preserving techniques that already exist in the literature. Some metrics for privacy and spectrum utilization are also studied here. Chapter 3 and Chapter 4 include an analysis of the vulnerability of the incumbent to inference attacks. They also present privacy-preserving techniques to prevent the exposure of the frequency and the location of the incumbent in a shared environment. While Chapter 3 studies the protection of the operational frequency using inherent and explicit obfuscation, Chapter 4 studies the protection of the operational location using an obfuscation-based algorithm and a trust-based algorithm. Chapter 5 introduces our framework and presents an architecture

for privacy-preserving centrally-coordinated spectrum sharing. Metrics for quantifying privacy and spectrum availability are presented along with a constrained optimization formulation of the trade-offs between incumbent privacy and secondary usage. Chapter 6 concludes the thesis and discusses future work opportunities.

Chapter 2

Spectrum Sharing and Privacy: State of the Art

2.1 Introduction

Spectrum is becoming a scarce resource [25]. Innovative approaches to spectrum management have been suggested to satisfy the growing commercial need for spectrum. Spectrum sharing of under-utilized governmental bands is one of those approaches. Sharing allows other users, in addition to licensees, to use the spectrum. However, this rises security concerns.

The focus of this chapter is to provide a review of the current state of the art for spectrum sharing and privacy. First, we introduce spectrum sharing, its models and its threats. Second, we give insight into the state of the art of privacy-preserving models in general. Finally, we show the impact of sharing on privacy and review metrics to evaluate both.

2.2 Spectrum Sharing: Models and Challenges

2.2.1 Background

A spectrum resource is defined by time, frequency, and location; specifically it is a period of time when a given set of frequencies is available for use in a particular location. So there are three physical dimensions to share the spectrum. Frequency administration is regulated by government agencies. The spectrum allocation chart in the United States,

as shown in Appendix A, is provided by NTIA [26]. This allocation is static. While it looks compact, most of the bands remain under-utilized. Such frequency management does not provide enough white spaces to be allocated for future applications. In order to optimize the use of frequencies and accommodate the growing demand for spectrum, sharing within the same band has been proposed.

Sharing was first proposed for unlicensed or "license-exempt" bands. The only requirement was to acquire a certified radio device that complies with FCC regulations. In this case, all users are treated equally and can be subject to interference. But, this was not sufficient to satisfy the increasing demand. The interest in sharing for licensed bands as well has grown fast. Hence, the FCC has encouraged other methods of sharing. Different bands have been auctioned and licensed on a geographic area basis. When the band is already allocated to a Primary User (PU), the newcomer is considered a Secondary User (SU).

2.2.2 Approaches to Spectrum Sharing

Three sharing models were proposed in literature [27][28][29] to provide coexistence between primary and secondary users:

- The *underlay approach* allows primary and secondary users to transmit at the same time as long as the aggregate interference power level at the primary user's receiver is below a pre-defined threshold [30]. In order to mitigate interference with primary operations, secondary users usually consider a spread spectrum technique or low power signals. The secondary users are affected negatively by this approach, as it severely limits their range of communication.

- The *overlay approach* assumes that the secondary user has a-priori knowledge about the transmission parameters of the primary user. Hence, the secondary can transmit at a maximal power level, but at the same time relay some of the primary messages to maximize the signal-to-noise ratio at the primary receiver [31]. Although it is a cooperative sharing, this method introduces additional overhead for the secondaries and security threats for the primaries.

TABLE 2.1: Comparison between the underlay, overlay and interweave sharing approaches

Characteristics	**Underlay**	**Overlay**	**Interweave**
Coexistence with the primary	✓	✓	×
Cooperation with the primary	×	✓	×
SINR limitations	✓	✓	×
Transmit power control	✓	×	×
Techniques used	Spread spectrum Beamforming	Data relaying	Spectrum sensing Database coordination

- The *interweave approach* is an opportunistic approach [32]. It does not allow any coexistence between the primary and the secondary at the same time. In this case, the secondary user senses the spectrum, decides whether it is idle or not, and uses it as long as the incumbent does not need it. The secondary user is required to keep sensing the spectrum and vacate the channel once the primary user becomes active again.

The interweave approach is highly favored in literature, nonetheless it requires certain intelligence from secondary users to assure an interference-free coexistence. Hence, cognitive radio technology has been introduced to provide adaptability and performance. As shown in Table 2.1, two types of spectrum management between primary and secondary users have been identified for this approach [33]:

- The *sensing-based spectrum sharing* assumes that secondary users must have sensing capabilities to sense the spectrum and identify channels available for use. They can adapt their transmissions to the change in spectrum availability: once a primary becomes active, they abandon the channel and move to another channel.

- The *database-driven spectrum sharing* allows a third party (generally a geolocation database) to manage the spectrum. Secondary users do not have sensing capability. They request spectrum resources from the database. The latter identifies the spectrum availability for use and grants access accordingly. If no resource is available for use, the secondary users will be denied access.

Although those mechanisms have been decisive to open up some bands for secondary use, the expectation of wireless broadband systems (more spectrum) remains unmet, and new questions have been raised: How immune is the primary user against interference? How safe is the sharing environment for its primary and secondary users? Nonetheless, this thesis will only target the operational security of sensitive incumbents in spectrum sharing.

2.2.3 Interference Issues

Spectrum sharing has enabled a dynamic allocation of frequencies by allowing more users to access and utilize the shared band. However, it has also generated other issues. The main concern has traditionally been interference.

Interference is the result of two or more transmitters within the same location competing for the use of one frequency (channel) at the same time. A spectrum resource is available for secondary users because the primary users are not using the resource or because the secondary users are able to use the resource without negatively impacting the primary users. The incumbent (i.e., primary user) has a high-level priority to access the spectrum and the secondary user should ensure no harm to the primary user when accessing the spectrum. As such, spectrum sharing takes place under well-defined interference constraints, defined in terms of collisions, overlapping, dropping probability, and channel abandonment time.

Collision can occur without harming the incumbent; therefore, we need to define harmful collision. For authors in [28] and [34], harmful collision occurs when the secondary user is regulated above the "interference temperature" (i.e., interference threshold) or the noise floor of the primary user. Authors in other references have captured collision with analogous metrics. For example, d'Utra da Costa and Cardieri [35] have defined collision probability as the long term ratio of the number of corrupted packets to the number of transmitted packets. Huang et al. [36] have defined collision probability as the long term ratio of the number of collisions to the number of busy periods. Sharma and Sahoo [37] have used a similar metric (interference probability), defined as the probability that a given secondary user transmission runs into a busy period. Those three metrics capture the same impact. So, the same definition can be applied to all metrics.

Sung et al [38] have used the interference violation probability, which the ratio of the average duration of an interference event to the average duration of the effective white space. Collision probability measures generally disruption to the incumbent. Collision among secondary users can also be considered when the transmission of two or more secondary users interfered [39].

Authors in [36] and [40] have considered overlapping a more appropriate metric than collision to measure QoS for some applications. Huang et al. [36] have defined percentage overlapping time: it is the fraction of time a secondary transmission overlaps with a primary transmission. Sahoo et al. [40] have defined overlap threshold probability: assuming an interference event, the overlap between a secondary transmission and primary transmission is considered harmful when it goes above predefined overlap threshold duration. In order to describe harmful collision and overlapping, we must inquire how well the noise floor and the overlap duration are defined. Those parameters should be measured carefully, depending on the sharing approach and the channel characteristics.

The appearance of a primary user during a secondary transmission may cause, when no other channel is available, the forced termination of that transmission. So, another performance metric has been studied in [41], [42] and [43]: the dropping probability (also known as interrupting probability or forced termination probability). Certainly, a more accurate sensing and a more seamless handover can alleviate interference effects and decrease the dropping probability. Also, authors in [42] and [43] used Markov chain models that provide more resources when accessing the spectrum, and authors in [41] proposed a sub-banding process that divides secondary channels into sub-channels.

A similar concern has been addressed in related works ([34][44][45][46]). In this case, instead of being forced to terminate, when a primary user appears on a channel, the secondary user must cease transmission and abandon that channel. The time it takes to do so is called channel abandonment time or channel evacuation time. A user must know the specific amount of interference a system will be receiving to decide whether to enable access in that spectrum or not [46]. This metric should be as small as possible, so that no harmful interference can be detected.

In the presence of secondary transmissions, their impact on the primary user's link should also be studied. Low level metrics such as the signal-to-noise ratio can be used [47].

Perhaps more useful end-to-end metrics, like error rate, throughput, latency and jitter, can be also considered. Those popular metrics were studied in various evaluation studies [48][49]. Naturally, they are not specific to spectrum sharing systems, but they spotlight perfectly their performance, and the impact on user-level applications. Indeed, if transmission errors and delays turn out to be excessive, then the bandwidth may be unfairly utilized and the performance will suffer.

2.2.4 Security Threats

As we have seen, interference is already a big issue when it is unintentional. When it is intentional, it becomes a security threat to spectrum sharing and specifically to incumbents. In other words, a secondary user may use its legitimate access to the spectrum to carry out attacks against the network.

Some attacks are common to all networking applications and non-specific to spectrum sharing, but still valid for its applications. Other attacks have just appeared with the emergence of sharing architectures. Actually, sharing environments suffer from unique security issues that do not exist in conventional wireless networks. Both can be carried out by adversaries (i.e., malicious and selfish users) to disrupt or block communications on the primary network as well as the secondary network. A rich literature has grown around security issues in networking, and developed some algorithms to detect security violations and implement countermeasures to avoid or at least mitigate its impacts.

Table 2.2 summarizes different attacks against a sharing environment in different layers, then lists some countermeasures presented in literature.

TABLE 2.2: Comparison of different attacks in spectrum sharing

Layer	Attacks	Countermeasures
Physical and Link Layers	**Jamming** (e.g., *intentional jamming, receiver jamming, common control channel jamming, overlapping secondary user, spectral honeypot*): keep using licensed bands by transmitting high power signals	Directional antennas [50] Spread spectrum and frequency hopping [50] Dynamic node pairing and random frequency selection [50] Cluster-based control channel allocation [51] Shadow-fading correlation-based filter [52] Rateless coding and piecewise coding [52] Power distribution strategies [52] Random non-occupancy period [53] Trust-value indicator [54]
	Primary user (incumbent) emulation or Sensitivity amplifying: emulate the primary user by mimicking the signal characteristics of the incumbent or by replaying primary transmissions	Distance ratio test and distance difference test [55] Naïve detection and variance detection [56] Fingerprint verification [57][58] Admission control base defense approach [57] Received signal strength indication (RSSI) based location verification scheme [59][60][61][58][51] Trust-value indicator [58] Beamforming-based attack prevention [58] RF signature of the PU-CR channel [51] Wald's sequential probability ratio test [62] Neyman-Pearson composite hypothesis test [62] Securing the frontline [53]
	False feedback: (e.g., *learning engine influence, false policy provision, backoff manipulation*): hide and/or modify the truth about policy parameters, spectrum occupancy ...	Trust-value indicator [54] Punishment schemes [52] Backoff schedule publishing [51] Coin-flipping and bit-commitment [63] Incentive-based channel negotiation [63]

Layer	Attacks	Countermeasures
Physical and Link Layers	**Spectrum sensing data falsification**: diffuse false data about the presence of the incumbent and mislead sensing decision	Authentication of sensing [60] Deployment of data fusion [60] Threshold voting rules [64][51] Shadow-fading correlation-based filter [52] Reputation weight for sensing nodes [52] Trust-value indicator [65][58][54] Prefiltering the sensing data [52] Punishment strategy [64][52] Consensus-based sensing [66]
	Byzantine: create routing loop, route packets on worst paths and selectively drop packets	Trust-value indicator [54]
	Biased utility: change utility function to get more bandwidth	Trust-value indicator [54]
Network Layer	**Routing**: perform routing table overflow, routing table poisoning, packet replication, route cache poisoning, routing attack	Securing routing protocols [67] Trust-value indicator [54]
	Wormhole: receive a signal from the transmitter and tunnel it to another location	Securing routing protocols [67] Trust-value indicator [54]
	Blackhole: hinder the path finding process or intercept the transmission processes	Securing routing protocols [67] Trust-value indicator [54]
	Jamming: (e.g., *network endo-parasite, channel ecto-parasite, low cost ripple effect*): increase interference at heavy loaded or high-priority used channels, mislead channels assignments	Securing routing protocols [67] Trust-value indicator [54]

Layer	Attacks	Countermeasures
Transport Layer	**Key depletion**: exploit key repetitions to break the underlying cipher system	Counter Mode Cipher Block Chaining Message Authentication Code Protocol (CCMP) [67] Trust-value indicator [54]
	Session hijacking: take control over a session between two nodes, masquerade as one of the end nodes and hijack the session	Cooperative security mechanism [61] Trust-value indicator [54]
Application Layer	**Repudiation**: deny participation in a communication	End-to-end authentication [61] Trust-value indicator [54]
Cross-Layer	**Jellyfish**: disorder packets, drop intelligently a fraction of packets, delay packets randomly as they pass through it (performed at the network layer to affect the performance of the transport layer) or make a user switch from a channel to another (performed in the link layer to impact network and transport layers)	Authentication, authorization and accounting (AAA) process [52] Trust-value indicator [54]
	Routing information jamming: jam the exchange of routing information among neighboring nodes (performed in the physical layer to affect the performance of the network layer)	Non-parametric version of the Pages cumulative sum (CUSUM) algorithm [68] Trust-value indicator [54]

Such attacks can disrupt the basic functions of a network: transmitting, routing and receiving messages. They jeopardize not only secondary access to the spectrum, but also primary operations. However, they target mostly sensing-driven spectrum sharing. In other words, those security threats are trying to corrupt the sensed information since the sensing approach (i.e., interweave approach) is widely-adopted in spectrum sharing systems.

An additional class of threats has been introduced when database-driven spectrum sharing has been proposed [69]. Some attack the database itself by targeting protocols of access to the database (e.g., man-in-the-middle attack, denial-of-service attacks, etc). Others just attack the privacy of users (either primary or secondary users) by disclosing information about them. In this case, we note that privacy has not been looked at enough.

2.3 Spectrum Sharing in Practice: The 3.5 GHz Band in the United States

2.3.1 Architecture Adopted

The President's Council of Advisors on Science and Technology (PCAST) has recommended to identify 500 MHz of spectrum for commercial use. According to the NTIA, bands previously-controlled by Federal agencies are shown to be under-utilized [14]. In fact, incumbents are not operational all the time, occupy fewer resources than assigned and do not operate in all geographic locations either.

Following NTIA and PCAST reports, the FCC has issued a first notice of proposed rulemaking and order [70] then a second one [71] to discuss deployment of Federal-commercial sharing in the 3.5 GHz band with military, vendors, investors and other interested parties. This band has been used by military shipborne, ground-based and airborne radar systems. Appendix B presents the detailed frequency allocation of the 3550–3700 MHz band and the functional architecture proposed.

This band is managed by a spectrum access system (SAS) incorporating a dynamic geolocation database (GDB) and with the assistance of an environmental sensing capability (ESC). Fig. 2.1 illustrates a tiered model with spectrum sharing. The SAS is fed

real-time occupancy measurements from a spectrum monitoring system (i.e., ESC) in order to determine channel availability and to control spectrum access. This spectrum monitoring is in one hand sensing the activity of the primary network and in the other hand collecting information about the activity of the other tiers. A query request and response mechanism moderates the access to the spectrum between the SAS and the secondary network.

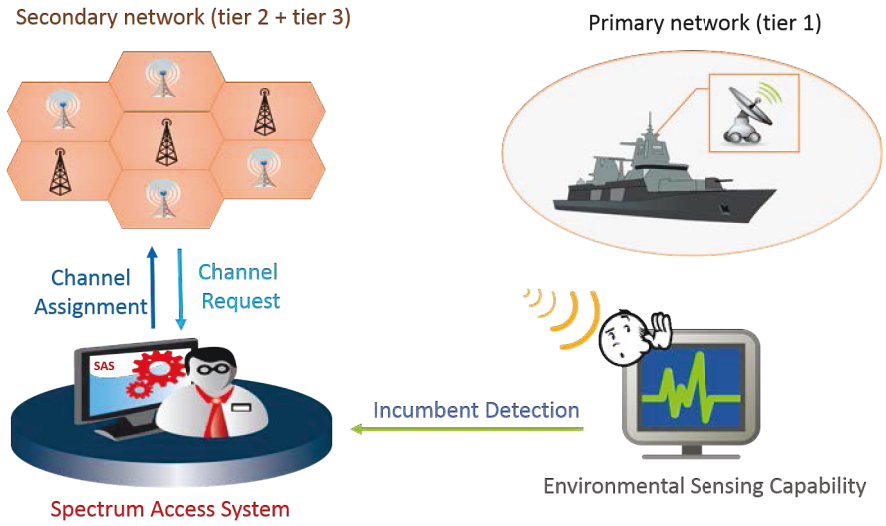


FIGURE 2.1: Architecture of spectrum sharing

Sharing is administered by a three-tiered shared access model:

- The first tier is called the Incumbent Access (IA) and it includes authorized Federal users such as telemetry, satellite, and radar. It should be protected from harmful interference caused by any other tier user.

- The second tier is called the Priority Access (PA) and it includes commercial operators for residential, business and mobile using small cell technologies. It is expected to be operating with critical QoS needs and interference protection. Preliminary approaches suggest the assignment of a license with a 10-MHz-channel in a single census tract.

- The third tier is called the General Authorized Access (GAA) and it consists of general public access based on an opportunistic non-interfering basis. However,

the PA users are supposed to share channels with the GAA users on a "use-it-or-share-it" basis.

2.3.2 Exclusion Zone vs. Protection Zone

Sharing with the Navy radars is sensitive and any interference can disrupt military operations. To protect the incumbent against interference, regulators have proposed geographic separation and frequency offsets to mitigate interference. NTIA has computed the zones where secondary users should not operate while incumbents are operating. Those zones are known as exclusion zones.

Based on a first analysis [14], NTIA recommended large exclusion zones along the coastlines. The secondary network users are authorized to operate only outside of the exclusion zone of the incumbent. In that analysis, NTIA assumed that high-power macro-cell networks are seeking access to the shared spectrum. Though it covers only a small fraction of the U.S. land mass, this area includes the entire coastal U.S., both east and west. Hence, approximately 60% of the U.S. population falls within the geographic area of an exclusion zone [72].

NTIA has recently reduced the exclusion zone distances by 77% of the total geographic area impacted along the coastlines, recomputing under the assumption that small cell technology is deployed [73]. Fig. 2.2 is extracted from the NTIA report to compare between the original exclusion zones (yellow solid line) and re-defined exclusion zones (blue solid line). The redefined exclusion zones have able to reduce the total area and enhance spectrum efficiency without harming the incumbent. However, that is not enough. Those zones are still large enough to induce spectrum loss.

While initial deployments will use the exclusion zones, exclusion zones will be later replaced by protection zones. The latter ones are smaller and more dynamic. The secondary network should be allowed to operate within the primary network as long as the aggregate interference-to-noise ratio at the incumbent receivers does not exceed a defined threshold. The aggregate interference-to-noise ratio threshold has been initially set to -6 dB [14]. Once the interference threshold is reached, other secondary users can be automatically directed to other channels available for use within the same area.

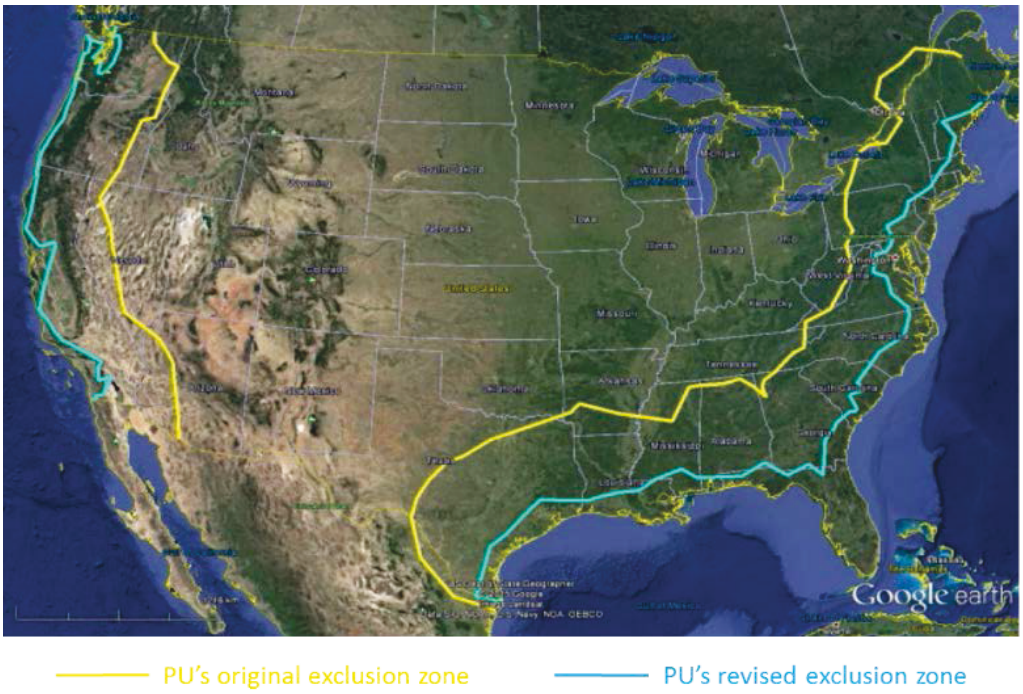


FIGURE 2.2: NTIA's exclusion zones

2.3.3 Sharing Requirements

The FCC has put together the final rules for the new Citizens Broadband Radio Service (CBRS) and Commercial Broadband Radio Service Devices (CBSDs) on April 2016 after receiving comments from regulatory, industry and academia [74]. These rules set requirements to enable an efficient use of the spectrum. Some of them are summarized below:

- *PA license terms and renewability*: The FCC has adopted a three-year license term, renewable only one time, for a total six-year licence term. This ensures flexibility of deployment while enabling competition and encouraging innovative approaches to exploit the spectrum. Also, a licensee many not hold more than 40 MHz (4 licenses) of the 70 MHz allocated for PA spectrum in each License Area.

- *Protection of PA users*: The PA licensees acquire a protection level of -80 dBm/10 MHz, which defines a protection area or a protection contour calculated to avoid interference among tier 2 users.

- *SAS and CBSD response time*: The FCC has extended the SAS-CBSD response time from 60 seconds to 300 seconds. Once the ESC detects an incumbent activity, the SAS must shutdown or reallocate its CBSDs within 300 seconds in order to protect the incumbent from interference while assuring a seemless disconnection of the service.

- *CBSD power*: Category A of CBSDs are limited to a maximum transmit power of 24 dBm and a maximum Equivalent Isotropically Radiated Power (EIRP) of 30 dBm in 10 MHz and may be deployed either indoors or outdoors. Category B of CBSDs may only be used outdoors and permitted to operate at higher power than Category A (30 dBm in 10 MHz), but only after an ESC is certified and implemented. However, there is need to increase the maximum transmit power and the EIRP in rural areas.

- *Out-of-band and adjacent channels emissions limits*: In order to ensure a shared environment without interference, the FCC aims to protect incumbent operations in adjacent channels. The limits proposed are -13 dBm/MHz from 0 to 10 MHz from the SAS assigned channel edge, -25 dBm/MHz beyond 10 MHz from the SAS assigned channel edge down to 3530 MHz and up to 3720 MHz, and -40 dBm/MHz below 3530 MHz and above 3720 MHz.

- *Location accuracy and automated geolocation*: The SAS should know the accurate location of all CBSDs in order to manage interference to the incumbent. Future developments in automated geolocation technologies are encouraged to enhance the performance of the SAS and the CBSDs. The end user devices are not required to include such feature.

- *Incumbent protection*: The main incumbents in the 3.5 GHz band are the Fixed Satellite Service (FSS) earth stations, the ground-based radars and the Navy ship-borne radars. Those incumbents are protected against harmful interference by enabling exclusion/protection zones. Also, CBSDs are required to report interference measurements to the SAS in order to re-calculate the aggregate SINR and protect incumbents against any interfering signals.

- *Opportunistic access to the spectrum*: Unused PA spectrum should be open for opportunistic access by GAA users in order to maximize the spectrum usage. However, some of the PA candidates think that this model does not help economic fairness and equity between users in the shared band.

- *Secondary markets*: Secondary markets include sub-lease, resale, exchange, etc. Those markets should be managed by a spectrum manager leasing to ensure both flexibility and control.

Other requirements and standardization matters are still being discussed by the Wireless Innovation Forum (WInnForum). Contributions are being made and trials are being held by different entities including industry and academia. The main purpose is to protect the incumbent and effectively manage the secondary/tertiary access. In this context, coordination is a key problem in a shared environment. The SAS shall be providing coordination among the different tiers by managing the aforementioned rules. It is true that all these requirements are important to achieve an efficient sharing. However, since the operational security aspect is not well defined, we will limit our work to security, mainly privacy.

2.4 Privacy Protection Applications

Data in its original form may contain sensitive information about individuals, and accessing such data will violate individual privacy. According to industry-specific legislation, regulation and self-regulation, the privacy of individuals and their data should be protected (e.g., social security number, medical diagnosis, wage, bank account, etc) from disclosure.

2.4.1 Database Inference Attacks

In general, a database inference attack occurs when an individual is able to deduce from trivial information more robust information about a database without directly accessing it [75]. In other words, an authorized individual can combine some received data from innocuous database queries with some, often publicly accessible, metadata (data providing information about one or more aspects of the data).

Inference attacks can be categorized into two types: linkage attacks and probabilistic attacks. Linkage attacks take place when an adversary is able to link a record owner to a record (record attack), a sensitive attribute (attribute attack), or the table itself (table attack). In those attacks, we assume that the adversary seeks to identify information about the victim in the dataset. Probabilistic attacks, however, allow the adversary to acquire additional facts about the database without necessarily linking them to anyone or anything; hence the difference between the a-priori knowledge and the a-posteriori knowledge is relatively large.

The question then becomes, "how can a data owner release its private data with guarantees that the individual subjects of the data cannot be identified while the data remain practically useful?" [76].

2.4.2 Privacy-Preserving Terms and Notation

Privacy-preserving models in literature fall into two general types of data treatment: data mining [77] and data publishing [78]. Data mining is an analyzing process that aims to examine data and change it from raw to useful. Data publishing is a releasing process that aims to make data available for public usage.

A very common architecture of privacy protection is presented in Fig. 2.3 and includes [79]:

- *Data owners* who own records and seek to protect their privacy;
- *Data publishers* who collect data from owners and prepare it for release;
- *Data recipients* who receive or collect information from data publishers in order to conduct data mining and/or publishing.

The data publisher gathers the information collected from the data owners and categorizes it into four categories to simplify the obfuscation process:

$$d = \{ID, QID, SA, NSA\} \tag{2.1}$$

where ID refers to the explicit identifiers, QID to the quasi-identifiers, SA to the sensitive attributes and NSA to the non-sensitive attributes.

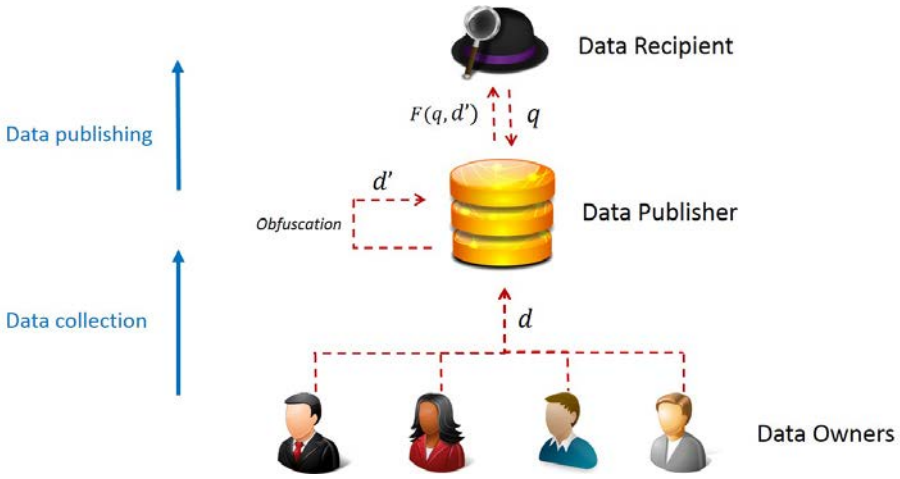


FIGURE 2.3: Privacy-preserving architecture for data mining and publishing

Explicit identifiers (i.e., key attributes) are uniquely identifying parameters, and should be always removed before release. Quasi-identifiers are non-sensitive attributes or combinations of non-sensitive attributes within a dataset that are not structurally unique but might be empirically unique and therefore uniquely identify a population unit. In other words, even though they are not sensitive, they can potentially identify record owners, e.g., 87% of the U.S. population may be uniquely identified by a combination of three quasi-identifiers: birthdate, gender and zip code [76]. All tuples with the same QID value form an equivalence class. A sensitive attribute is an attribute whose value for any particular individual must be kept secret from people who have no direct access to the original data. Non-sensitive attributes are attributes whose values are already an unrestricted record and can be accessed and/or released publicly.

In order to protect privacy, some privacy-preserving techniques have been employed. Those techniques have used obfuscation to decrease the information utility and make the inferred knowledge ambiguous or confusing to an adversary. By applying obfuscation methods to the attributes in the database, a new table is then acquired. The obfuscated table will have the following structure:

$$d' = \{QID', SA, NSA\} \tag{2.2}$$

where ID are eliminated and QID' refer to the obfuscated QID.

Sensitive attributes (SA) and non-sensitive attributes (NSA) do not change. An attacker should not be able to learn any extra information, even with the presence of a background knowledge obtained from other sources.

2.4.3 Privacy-Preserving Techniques

A rich literature introduced privacy-preserving techniques for data mining and publishing, mainly: perturbation [80], k-anonymity [76], l-diversity [81], confidence bounding [82] and differential privacy [83]. k-anonymity, and l-diversity are applied on the database itself, while differential privacy is applied on the release mechanism. Confidence bounding can be applied on both. Different derivations of those techniques have been proposed as well.

We will use the following example (Table 2.3) to explain each one of those techniques. A hospital is running a data mining and publishing operations on its patients' records. We consider the following simplified database that includes the patient's identifier, name, age, postal code and disease. The sensitive data here is the condition (disease) of the patient.

Table 2.3: Simplified example of a database

Identifier	Name	Age	Postal Code	Disease
1	Pierre	34	75000	AIDS
2	Pascale	55	75003	Cardiovascular
3	Aline	23	75000	Diabetes
4	Mary	61	75002	Cardiovascular
5	Francois	47	75003	AIDS
6	Julien	49	75001	AIDS
7	Jennifer	56	75001	Cardiovascular
8	Tom	28	75000	Cardiovascular
9	Bob	45	75002	Diabetes

Perturbation

Perturbation (or randomization) is considered as one of the simplest and most effective methods of obfuscation. Its algorithm is straightforward and intuitive: add noise to the original information [84].

Perturbation can be applied as follows:

$$d' = d + n \tag{2.3}$$

where d is the original data, n is the additive noise and d' is the perturbed data.

For example, using perturbation, we can hide the last three digits of the postal code. Table 2.3 then becomes Table 2.4.

TABLE 2.4: Perturbed database

Identifier	Age	Postal Code	Disease
1	34	75***	AIDS
2	55	75***	Cardiovascular
3	23	75***	Diabetes
4	61	75***	Cardiovascular
5	47	75***	AIDS
6	49	75***	AIDS
7	56	75***	Cardiovascular
8	28	75***	Cardiovascular
9	45	75***	Diabetes

Perturbation can also be applied to other parameters (e.g., age) in order to increase the ambiguity of the data.

k-anonymity

Anonymity is obtained by generalizing or suppressing parts of the data. No individual can be uniquely distinguished from a group of size k [76].

However, protecting the identity of an individual does not necessarily include protecting its sensitive attributes. For example, if we apply 3-anonymity to Table 2.3, it becomes Table 2.5.

Even though this table does not allow an attacker to link the disease to one person, it lacks diversity. An attacker can link a disease to a group of patients (homogeneity attack). Also, an attacker, who is targeting someone and knows his age and postal code, can deduce its disease (background knowledge attack).

TABLE 2.5: Anonymized database

Identifier	Age	Postal Code	Disease
1	< 40	75000	AIDS
3	< 40	75000	Diabetes
8	< 40	75000	Cardiovascular
5	4*	75003	AIDS
6	4*	75001	AIDS
9	4*	75002	Diabetes
2	≥ 50	75003	Cardiovascular
4	≥ 50	75002	Cardiovascular
7	≥ 50	75001	Cardiovascular

l-diversity

The l-diversity model handles some of the weaknesses in the k-anonymity model when encountering homogeneity attacks and background knowledge attacks. Hence, diversity comes to prevent sensitive attributes from appearing more frequently than others in a dataset [81].

For example, we include 2-diversity in this case by sorting the data by postal code instead of age (Table 2.6).

TABLE 2.6: Diverse database

Identifier	Age	Postal Code	Disease
1	< 40	75000	AIDS
3	< 40	75000	Diabetes
8	< 40	75000	Cardiovascular
6	≥ 40	75001	AIDS
7	≥ 40	75001	Cardiovascular
4	≥ 40	75002	Cardiovascular
9	≥ 40	75002	Diabetes
2	≥ 40	75003	Cardiovascular
5	≥ 40	75003	AIDS

Confidence Bounding

The procedure of confidence bounding aims to bound the attacker confidence of inferring the sensitive attribute in any group on to a maximum value h [78]. This technique

is similar to the l-diversity technique, the only difference that it allows flexibility in choosing the threshold h, depending on the nature and the type of sensitive attributes.

Differential Privacy

Formally, a database is ϵ-differentially private if for all datasets T_1 and T_2 differing on at most one record, $|ln\frac{Pr(F(T_1)=S)}{Pr(F(T_2)=S} \leq \epsilon$ for all $S \in Range(DB)$ where $Range(DB)$ is the set of possible responses $F(T)$ of the database DB [78]. In other words, a database ensures differential privacy if it ensured that the removal or addition of one parameter doesn't remarkably change the result of any query response.

Another approach compares the risk with and without the record's owner data in the dataset, and ensures that the difference is less than ϵ. In this case, differential privacy applies a condition on the release mechanism.

2.5 Threats in Spectrum Sharing

In order to understand how to apply those techniques in the spectrum sharing context, we need first to define the risk of inference and model the potential attacks.

2.5.1 Sensitivity of the Incumbent's Parameters

The incumbent has different sensitive parameters that should not be disclosed. Some of the most sensitive parameters are [85]:

- *Geolocation of the incumbent*: finding the exact location of the incumbent can cause military threats, and accordingly increase the predictability of its path of movement.
- *Center frequency of the incumbent*: knowing the operational frequency can engender the disclosure of the incumbent's signals.
- *Times of operation of the incumbent*: knowing the on and off times may predict the active period of the incumbent.

The SAS database must not provide any information that may, intentionally or unintentionally, compromise sensitive information or reveal operations of the Federal incumbent.

2.5.2 Diversity of Secondary Systems

The secondary network consists of CBSDs, which include PA and GAA users. The operation of all CBSDs must be coordinated by one or more authorized SAS. The number and type of secondary devices vary. Even though PA devices are known to be deploying small cell technologies [8], the GAA devices can be using other technologies (WiFi, 802.11af, 802.14.5m, etc). This diversity shall include additional privacy concerns.

2.5.3 Profile of the Attacker

In our threat model, we consider that the spectrum access system (SAS) is trustworthy; however, the trust is not extensible to the secondary users. So, we assume that the attacker is a legitimate querier that has been registered within the spectrum access system and can have access to spectrum resources. It can have some help from other secondaries (cooperative attackers) or fake different identities (identity spoofing). It also has enough computational resources to make inferences and probabilistic attacks based on received information.

Fig. 2.4 shows an inference attack on a reliable SAS carried out by a malicious secondary user. Using received non-sensitive knowledge like channel availability, allowed transmit power and time and background knowledge, the adversary can infer the incumbent sensitive data (location and path of movement, operational frequency and time, etc).

The adversary may use the inferred information to endanger the operations of the incumbent (e.g., jamming). Preventing an inference attack is a challenge, since it is hard to detect and no explicit violation is made. Different approaches have been proposed through academic papers and proprietary applications. None of them has been widely adopted, because they fall short to accommodate both the queried need for privacy and the querier need for access. Actually, each gain in privacy protection is accompanied by a loss in spectrum efficiency. Hence, not only the incumbent privacy but also the sharing effectiveness should be taken into consideration. Also, there is no "one-size-fits-all" solution that fits all scenarios. Different strategies may be used by an adversary, and the appropriate techniques should be employed in order to mitigate them.

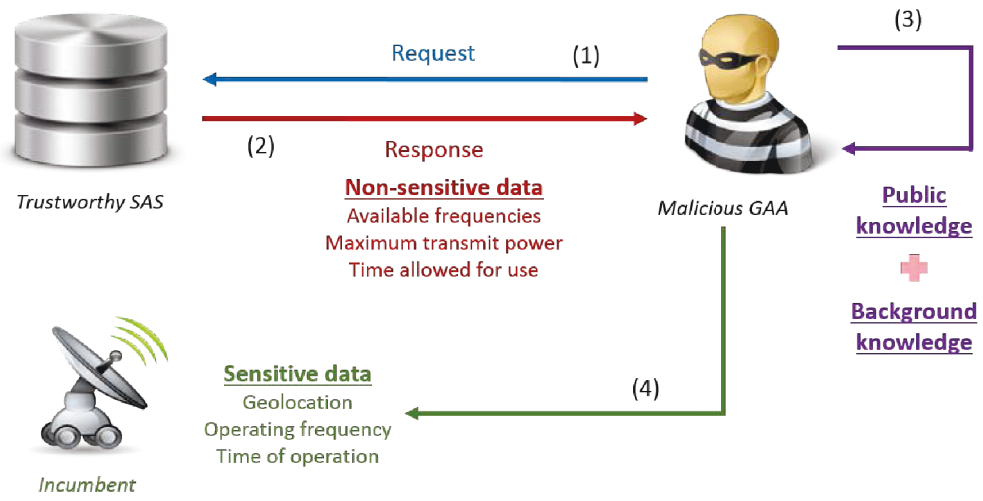


FIGURE 2.4: Database inference attack

To address this issue, the WInnForum has created a working group on security requirements [86] within the Spectrum Sharing Committee (SSC). The goal of this group is to ensure the operational privacy and security of incumbents. The WInnForum has drawn up security and privacy requirements in the 3.5 GHz band and acknowledged the importance of implementing mechanisms and algorithms to protect the location, frequency and time of operations of incumbent against inference attacks carried out by legitimate secondary users. The first requirement instructs the SAS and the ESC "not to store, retain, transmit or disclose operational information" that may put the incumbent's operations at risk [87].

2.5.4 Privacy Protection in Spectrum Sharing

Privacy is not a new subject. However, most of the attention has been given to location privacy. In the literature, location privacy issues have already arisen with the emergence of RFID systems, vehicular networks applications, and AdHoc networks routing [22]. Privacy preservation has been also studied in cognitive radio networks [88], to protect the location of licensed and unlicensed users from exposure. Privacy-preserving techniques adopted from data mining and publishing have been proposed.

Perturbation

A straightforward approach to protect the incumbent's location is to add additive noise to the radius of the protection zone. For Location-Based Services (LBSs), authors in [89] proposed three different perturbation methods: enlarge the radius, reduce the radius, or shift the center of the protection zone. A combination of these methods can be also used to enhance privacy protection of the location. In the case of spectrum sharing, reducing the radius of the protection zone would result in more interference for the incumbent unless it is combined with a reduction in the allowable transmit time and power.

The shape of protection zone of PUs can be changed, an operation called transfiguration [85]. This approach has been used in TV White Space sharing to thwart primary user emulation attacks (PUEA). In PUEA, if an attacker finds the TV location, it can emulate TV signals, act as the incumbent, and consequently claim indefinitely the spectrum. Authors in [85] give the example of transfiguration by identifying the protection zone as a polygonal region instead of a circle, and conclude that the larger the number of polygon sides, the better the privacy.

In order to anonymize the time (or power) of operation, the maximum allowed time (maximum allowed power, respectively) for use can be reduced when replying to secondary users queries. A rounding down algorithm can be employed to do so. This way, ambiguity is inserted into data. Rounding down can be applied to time or power allowed for use separately. It can be also applied for both at the same time.

k-anonymity

k-anonymity can be applied to the location of incumbents. By combining the protection zones of radars that are close to each other, a larger protection zone that operates as one single radar is created. Hence, a table is considered location k-anonymous if and only if the location information of a radar is indistinguishable from the location information of at least $k-1$ other radars [85]. This technique works well when locations are close to each other. In the context of spectrum sharing in the 3.5 GHz band, the number of operational shipborne radars in the same area is limited. Realizing location k-anonymity

in the radar context would expand the protection zone, reduce the access to spectrum opportunities, and hence decrease spectrum efficiency.

k-anonymity can be applied to the frequency of operation as well. By joining busy channels, one single busy channel is generated. The probability of inference (relating a record to an individual) cannot be greater than $1/k$. Clearly, the higher the value of k, the better the privacy. However, the problem of optimal k-anonymity is considered NP-hard. Being NP-hard is not the only problem of anonymity. One other disadvantage of k-anonymity occurs when we don't have enough adjacent busy channels to form a group of k channels and, in order to achieve k-anonymity, idle channels within that range of frequencies would be included. Such allocation results in not only an inefficient management of resources, but also an erroneous representation of spectrum occupancy.

k-clustering

k-clustering is a derivation of k-anonymity. Clustering is a solution to the limitations of anonymity. k-clustering was first introduced in [85] to protect the location privacy of the incumbent. Contrary to k-anonymity, k-clustering allows more flexibility. But, in general, k-clustering permits linking each individual to a group of size "at least" k.

k-clustering cannot however prevent linking a record to a set of individuals with the same group in the dataset. The attacker can thus infer sensitive attributes related the same group of individuals. For example, authors in [85] propose location k-clustering to replace location k-anonymity. But, even though an attacker wouldn't be able to infer the location of each individual member of the cluster, the path of movement of the whole cluster can be inferred.

l-diversity

In the context of spectrum sharing, diversity is created by providing l different responses to the same request. Therefore, all values of a given attribute are treated in a similar way irrespective of its distribution in the data. Nonetheless, even though this technique ensures that at least l distinct values exist for the sensitive attribute in each equivalence class, it assumes that sensitive attributes are uniformly distributed over their

domains, which is not always true. Also, an excessive application of diversity may be hard to achieve, meaningless and affect information utility [90].

2.6 Metrics for Spectrum Sharing and Privacy

2.6.1 Spectrum Sharing Metrics

The main goal of sharing is to meet the increasing demand for bandwidth and maximize the exploitation of the spectrum. Therefore, metrics evaluating the efficient use of spectrum as well as the impact on privacy are needed.

Spectrum Utilization

A preliminary performance metric has been presented in [34] and [44] to measure the use of spectrum: spectrum utilization. It is defined as the amount of frequency, space and time impacted.

$$U = B \times S \times T \tag{2.4}$$

where B is the frequency bandwidth (in Hz), S is the geometric space (in m^3) and T is the time (in s).

This metric can reflect how many spectrum resources have been used, but can't reflect if those spectrum resources are efficiently utilized or not.

Spectrum Utilization Efficiency

Authors in in [34], [44] and [91] have introduced a more appropriate metric: spectrum utilization efficiency. It is defined as the ratio of the information transferred to the amount of spectrum impacted, and calculated given the formula:

$$SUE = \frac{M}{U} = \frac{M}{B \times S \times T} \tag{2.5}$$

where M indicates the amount of information transferred.

The particular form and units of M vary depending upon the application. For example, data rate (in $bits/s$) is appropriate for mobile data services. Other quantities and units may be more appropriate for other applications. It measures of the quantity

of services that can be supported by a limited radio frequency bandwidth in a specific geographic area over a defined time.

Spectrum Effectiveness

Spectrum use needs a more sophisticated metric that measures effectiveness rather than simple efficiency. So, in contrast to efficiency, spectrum effectiveness includes the "communication range", i.e., how far communications go. It measures effectiveness in terms of data delivered across a range, over the spectrum, area, and time whose usage is precluded [92]. It is given by:

$$Eff_{spectrum} = \sum_{n=1}^{N} \frac{R(n)D(n)}{I^2(n)T(n)S(n)} \tag{2.6}$$

where $R(n)$ is the actual communication range of a user n, $D(n)$ is the quantity of data delivered for user n, $I(n)$ is the interference range of a user n, $T(n)$ is the time over which the spectrum is utilized by user n, $S(n)$ is the actual spectrum precluded from user n to other users, and N is the number of users within a spectrum resource and over a region.

All the previous metrics are low-level metrics. They are more suitable for physical or MAC layer performance than measuring the performance of a spectrum sharing system.

2.6.2 Privacy Metrics

In order to evaluate the privacy level, we need to compute the inferred knowledge of the attacker compared to the actual knowledge.

Uncertainty

A very traditional metric used for information theory applications is Shannon's entropy [78]. In this context, it quantifies the uncertainty of the attacker's knowledge with respect to finding a unique answer.

$$H = -\sum_{x \in X} p(x) log(p(x)) \tag{2.7}$$

where $p(x)$ denotes the probability of a particular observed value x for a sensitive attribute X.

Hence, the greater the entropy, the greater the uncertainty. This widely-used metric is not appropriate for measuring the location privacy of the incumbent. When the attacker has a high level of certainty about an incorrect distribution, it does not necessarily mean that the location of the incumbent can be compromised.

Inaccuracy

Given that the attacker's knowledge is always an estimate, another metric has been proposed to measure the performance of an obfuscation technique: inaccuracy [85]. It measures the discrepancy between an estimated distribution of a sensitive attribute and the real one.

$$IA = \sum_{x \in X} (\tilde{p}(x) - p(x))^2 \tag{2.8}$$

where $p(x)$ denotes the real distribution of a sensitive attribute x and $\tilde{p}(x)$ denotes the attacker's estimated distribution for the possible values of a sensitive attribute x.

This metric is also not sufficient for measuring location privacy. It measures the difference between the attacker's estimation of PU's location distribution and the PU's real location distribution. Nonetheless, two different estimated location distributions can have the same inaccuracy, but one can still be better than the other.

Incorrectness

In order to alleviate the limitations of both previous metrics, authors in [85] proposed a new metric called incorrectness. It measures how far inference results are from actual results.

$$IC = \sum_{x \in X} \tilde{p}(x) d(x) \tag{2.9}$$

where $\tilde{p}(x)$ denotes the attacker's estimated distribution for the possible values of a sensitive attribute x and $d(x)$ the difference between the estimated and the actual distribution of a sensitive attribute x. The larger the value, the better the privacy.

The incorrectness metric was used to quantify the location privacy of the incumbent and the time of operation privacy of the incumbent. For the former, they calculate

the expected distance between the location inferred by the attacker and the PU's true location on a specific channel. For the latter, they calculate the expected difference between the time of operation inferred by the attacker and the PU's true time of operation on a specific channel.

2.6.3 Trade-off Metrics

When more obfuscation is used to achieve a greater level of privacy, there is a loss of information utility. Likewise, when more obfuscation is used, there is less efficient use of spectrum resources. For example, when perturbation techniques over-add noise to data, valuable information is lost, and the data itself becomes useless. Also, when a k-anonymity technique is deployed, the protection zone becomes larger, limiting secondary access to the network. And when l-diversity is used, some vacant channels are omitted, which limits spectrum opportunities.

Trade-offs should be defined in order to maintain a certain balance between privacy protection, data usefulness and spectrum opportunities. Hence, some "score" metrics need to be included in order to evaluate the impact of each constraint. Authors in [78] measured the trade-off between privacy gain and information loss given this formula:

$$ILPG(a) = \frac{IL(a)}{PG(a) + 1} \tag{2.10}$$

where $IL(a)$ denotes the information loss (i.e., spectrum loss) and $PG(a)$ denotes the privacy gain by performing an obfuscation action a.

The functions $PG(a)$ and $IL(a)$ can be evaluated depending on the privacy model, the information metric and the spectrum requirement.

2.7 Conclusions

In this chapter, the privacy threat encountered by Federal incumbents and the techniques introduced in literature to thwart it have been introduced. The spectrum sharing system considered is a tiered system that includes incumbent access, priority access and

general access. Priority and general authorized devices can try to compromise the operational privacy of the incumbent through inference attacks. Obfuscation techniques, usually used in a data mining context, can alleviate those attacks.

Obfuscation techniques consist of algorithms that add a condition either on the database itself or on the release mechanism. However, we should not only focus on how to generate an obfuscated database. When explicit obfuscation is introduced, it can become spectrum-hungry. This will decrease the efficiency of the sharing. Other obfuscation methods and trade-offs should be defined in order to ensure a healthy balance between privacy protection and spectrum effectiveness. Next chapter proposes inherent obfuscation in addition to explicit obfuscation in order to protect the incumbent's frequency against inference attacks while guaranteeing a maximum secondary access to the spectrum.

Chapter 3

Protection of the Incumbent's Frequency

3.1 Introduction

Incumbents such as military and public safety users require full protection of their operations. This includes the protection of their privacy. The operational frequency of primary users is a sensitive parameter and should not be exposed. Protecting against its discovery is critical to mitigate intentional interference (e.g., jamming attacks). The activity of secondary users other than the attacker can affect the security of the incumbent.

In this chapter, the activity of secondary users is modeled via an Erlang loss queueing system. Attacks attempting to infer the incumbent's operational frequency through multiple requests are also modeled and analyzed. Solutions and metrics are proposed to evaluate and address the vulnerability of the incumbent to such attacks. Analytical and simulation results highlight the efficiency of our proposal.

3.2 State of the Art

Existing work on incumbent privacy protection in spectrum sharing is very limited. However, there are a few publications that are relevant to this work.

Bahrak et al. [85] propose obfuscation in order to enable sharing while protecting the incumbent from inference attacks. They implement privacy-preserving techniques used in data mining and publishing (perturbation, k-anonymity, and k-clustering) to secure

the location and the time of operation of the incumbent. Clark et al. [93] formulate the incumbent's location privacy problem under different attack models and estimation schemes. They also propose a metric to measure the expected time that privacy can be maintained when using obfuscation. Intuitively, more obfuscation results in better privacy but fewer opportunities to access the spectrum; however, these trade-offs are not well analyzed.

While the authors in both works develop good performance metrics to evaluate privacy, some of their assumptions are stringent and are not applicable for general cases (for example, they are unsuitable for use with the 3.5 GHz band). The authors assume the incumbent is stationary and its location is known by the SAS. However, in the 3.5 GHz band, the shipborne radar is moving and its precise location is not known by either the ESC or the SAS [87]. The ESC and SAS only know that the incumbent is somewhere within the detection zone of an ESC sensor, which can be on the order of 10 km^2. They also assume that the adversary knows the power assignment function implemented by the SAS to allocate channels for secondary use. However, the response of the SAS to a secondary user's query is fairly straightforward: a channel to use for a specific period of time [13]. Furthermore, they do not take into account the role of the spectrum load and the attacker's query rate in their approaches.

Finally, they focus on protecting the location of the incumbent and do not consider the protection of its frequency of operation. While protecting the location of an incumbent is important for its privacy, protecting its frequency can be more crucial. Once an adversary becomes aware of the presence of an incumbent, it can deduce that an incumbent is within the detection area of the ESC.

3.3 Problem Statement

In this section, we describe the shared environment by presenting a system model to access spectrum resources and a threat model to carry out an inference attack.

3.3.1 System Model

The SAS manages n channels within a specific area. Only l channels are available for use by secondary users. In other words, the incumbent is occupying $n-l$ channels. The channels occupied by the incumbent do not change during the operational period, which means that the incumbent channels will not be open for secondary sharing. Consequently, we are not interested in modeling the incumbent activity, but the sharing of the spectrum between secondary users.

To date, there are no operating SAS deployments. Therefore, in the absence of a realistic sharing environment, we model our system as an $M/M/l/l$ queue, also known as an Erlang-loss queueing system [94]. This is a queue model where the l available channels are considered servers. The servers (i.e., channels) are independent, identical, and handling a parallel service. We also model the secondaries requests with an aggregate Poisson arrival process of rate λ, and their time of transmission on a specific channel with an exponentially distributed service time of rate μ.

Since the system size is l, if a secondary requests a channel and all l channels are busy, access to the spectrum will be denied. However, when a secondary user requests spectrum resources and some channels are available for use, the SAS replies with a list of $m \leq l$ available resources for the secondary to choose from. In our analysis, we first assume that $m = 1$ (i.e., the SAS only returns a single available channel to secondary users in each response). Then, we show the effect of increasing m by simulations.

Also, we assume that the response time of the SAS is negligible compared to the service time of the secondary. Once a channel is assigned, the secondary begins transmission. We consider an unlimited number of secondaries, with an unlimited number of queries.

3.3.2 Spectrum Metrics

In a spectrum sharing system, a secondary user is denied access when no resource is available for use. This is reflected in the queuing model with the "no waiting line" property of the $M/M/l/l$ model presented in Fig. 3.1. Based on that property, an effective arrival rate λ_e and a blocking rate λ_b are defined. The effective arrival rate is the average number of active secondary users (i.e., number of occupied channels)

times the average time a secondary user spends in the system (i.e., transmission time). The blocking rate, however, is the percentage of arriving secondary users that are not allowed to use the spectrum. In our case, we assume an ideal transmission environment. Therefore, we do not consider forced termination and corrupt messages as part of the blocking rate. Generally, the blocking rate should be kept low to ensure that secondary users have access to the spectrum.

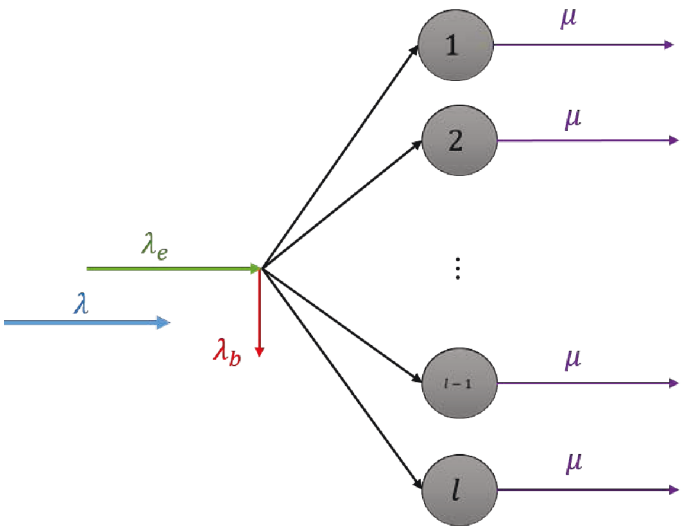


FIGURE 3.1: Modeling of an Erlang loss queueing system

The channel utilization is a Markovian birth-death process with rates:

$$\lambda_k = \begin{cases} \lambda, & \text{if } k < l \\ 0, & \text{if } k \geq l \end{cases} \tag{3.1}$$

$$\mu_k = k \cdot \mu, \text{ for } k = 1, 2, \ldots, l \tag{3.2}$$

The system is said to be in state k if there are k secondary users in the system as shown in Fig. 3.2. Let P_k be the probability that there are k secondaries in the system. Using Eq. 3.1, Eq. 3.2 and the fact that $\sum_k P_k = 1$ [95], we have the following balance equations:

$$P_k = \frac{\frac{\lambda^k}{k! \cdot \mu^k}}{\sum_{j=0}^{l} \frac{\lambda^j}{j! \cdot \mu^j}}; \; 0 \leq k \leq l \tag{3.3}$$

Let $\rho = \lambda/\mu$ be the system load. Therefore, Eq. 3.3 becomes:

$$P_k = \frac{\frac{\rho^k}{k!}}{\sum_{j=0}^{l} \frac{\rho^j}{j!}};\ 0 \leq k \leq l \tag{3.4}$$

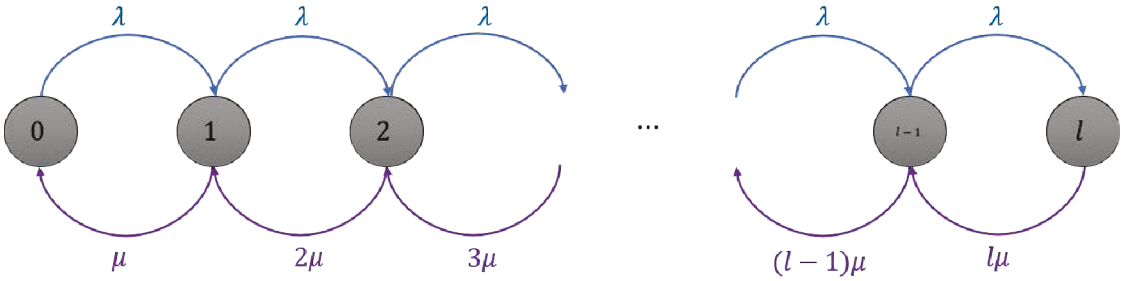


FIGURE 3.2: Rate diagram of an Erlang loss queueing system

According to Little's law [96], when in equilibrium (steady state), "the long-term average number of users in a stable system is equal to the long-term average effective arrival rate multiplied by the average time a user spends in the system." So, we get the following equations:

$$P_B = P_l \tag{3.5}$$

$$\lambda_e = \lambda(1 - P_B) \tag{3.6}$$

$$\bar{S} = \rho(1 - P_B) \tag{3.7}$$

where P_B denotes the blocking probability of a secondary, λ_e denotes the effective arrival rate of secondaries and $\bar{S}$ denotes the mean number of secondaries in the system (i.e., the mean number of busy channels).

3.3.3 Attack Model

In our model, an attacker aims to gain knowledge about the operational frequency of the incumbent. Since the SAS does not intentionally give away that information, the attacker will try to analyze the response of the SAS and deduce the channel of operation

of the incumbent. Then, an attacker is a legitimate secondary that has been registered with the SAS, and thus has access to the l available channels.

We consider the following inference attack scenario. Let ICH be the list of operational channels of the incumbent. Like any other user, the attacker sends queries requesting access to the spectrum. Its initial knowledge is a list of all potential incumbent channels (i.e., all n channels in the band of interest). Once the SAS returns a list of m available channels for use (ACH) in reply to a query, the attacker knows that each channel in ACH is not used by the incumbent. Hence, the attacker updates its knowledge by removing the channels in ACH from ICH. At any given time, the number of inferred channels is equal to the size of ICH. That is, the attacker knows that these channels are possibly being used by the incumbent. This behavior of the attacker is shown in Algorithm 1.

Input: Total number of channels n;
Output: Potential channels used by the incumbent ICH;
Initialization: $ICH = [i \text{ for } i \text{ in } [1, n]]$;
while *size of* $ICH > 1$ **do**
    Query the SAS and get allocation of channels ACH;
    for j *in* ACH **do**
        if j *in* ICH **then**
            Remove j from ICH;
        end
    end
end

Algorithm 1: Algorithm of an inference attack

After query q, the probability of inference becomes

$$Pr_{inf}(q) = \frac{n - l}{\text{size of } ICH(q)} \tag{3.8}$$

where $n - l$ is the number of channels occupied by the incumbent and *size of* $ICH(q)$ is number of channels inferred by the attacker at query q.

Even though a secondary user is allowed to ask the SAS for a specific channel [86], we don't adopt this approach in our analysis. We believe that it can help the attacker identify the incumbent channel. Therefore, the SAS denies any request for a specific

channel. Instead, it returns a list of channels based on the channel assigned scheme adopted. This spectrum management policy helps to protect the incumbent by limiting the capability of the attacker.

We are aware that other attack scenarios are possible, e.g., eavesdropping on secondary communications or sensing directly the spectrum. However, since we are examining inference attacks in particular, they are beyond the scope of this thesis.

3.3.4 Privacy Metrics

We already introduced a model of the secondary activity as a queuing system and a model of a malicious secondary using that system to infer the operational channel of the incumbent. Here, we introduce two metrics to evaluate how vulnerable the incumbent is to inference attacks.

- *Distance of Inference*: The inference process can be regarded as a discovery process of all channels available for use by secondaries. Therefore, we can evaluate privacy as a measure of "distance," that is, the number of channels remaining to be discovered.

  The distance of inference can be expressed as follows:

  $$d(q) = \text{size of } ICH(q) - n + l \tag{3.9}$$

  where $d(q)$ denotes the distance of inference at query q, size of $ICH(q)$ is the number of inferred channels at query q, n is the total number of channels, and l is the number of available channels for use (i.e., $n - l$ is the number of channels occupied by the incumbent).

- *Cost of Inference*: In spectrum sharing, the attacker invests effort to infer sensitive data. We measure the inference cost to the attacker in terms of how long the attacker takes to acquire the inferred knowledge and the number of queries to acquire that knowledge.

  The time of inference and the number of queries to inference are proportional as:

  $$q = \lambda_a \times t \tag{3.10}$$

where q is the number of queries to inference, t is the time of inference, and λ_a is the query rate of the attacker.

3.4 Obfuscation Model

In order to mitigate inference attacks, we propose obfuscation. Obfuscation can be inherently applied by tuning some of the system parameters or explicitly applied by changing the system itself.

3.4.1 Inherent Obfuscation

The system parameters include the total number of channels, the channel assignment scheme, and the number of queries allowed per secondary user. Each one of those parameters can be adjusted to fit the privacy needs of the incumbent. In a later section of this chapter, we show the effect of those parameters on the operational security.

Total Number of Channels

The total number of channels n can be defined based on the technology used by secondary users. The secondaries should not use more spectrum than their needs. In our setting, we consider the 3.5 GHz band which includes a band of 150 MHz. This band will be divided into channels of 10 MHz bandwidth. However, we know that some of the secondary users will use only 2 MHz of the 10 MHz allocated. So, in other words, if a secondary user needs only a 2-MHz channel, the SAS should not allocate a 10-MHz channel, as this allocation does not only jeopardize the incumbent privacy but also results in spectrum loss.

Channel Assignment Scheme

The channel assignment scheme has a major impact on the management of the spectrum access. We consider three channel assignment schemes that can be used by the SAS: random channel assignment, ordered channel assignment and semi-static channel assignment.

- The *random channel assignment* assigns channels to secondaries randomly from the list of available channels. For example, as shown in Fig. 3.3, consider one incumbent I operating on a channel f_2 and four secondaries S_1, S_2, S_3 and S_4, who requested channels in that order. Aside from channel f_2, which is not allowed for use by any secondary, channels are assigned randomly from the idle channels with uniform distribution.

- The *ordered channel assignment* assigns channels to secondaries in an order-wise fashion, and any particular order can be employed. For example, in (a) of Fig. 3.4, we choose to illustrate an ascending channel assignment scheme without loss of generality. In other words, if we consider one incumbent and $n - 1$ available channels, for each query, the SAS returns the lowest available channel at the time of query. Likewise, in (b) of Fig. 3.4, we illustrate a descending channel assignment scheme, i.e., the SAS returns the highest available channel at the time of query. To increase the obfuscation of this scheme, the order can change from one operational period to another. Intuitively, this scheme will increase privacy by reducing an attacker's probability of visiting all available channels.

- The *semi-static channel assignment* assigns channels to secondaries in a fixed way. A secondary will be assigned the same channel if it is not already occupied by another secondary. In the WInnForum document, WINNF-15-S-0071 [87], the requirement is "SAS providers shall implement authorization limiting techniques when assigning spectrum to users." While no specific technique is required, we refer to the suggested method in [87] as "semi-static channel assignment". This scheme requires the SAS to maintain an exhaustive channel usage database for all the secondary users requesting access to the spectrum. It is not immune against cooperative attacks, where different malicious secondary users share their channel assignments with each other in order to infer the incumbent channel. It will not also protect the SAS against smart attacks, where an attacker uses either different devices or different identities from the same device to gather more information about the availability of the spectrum.

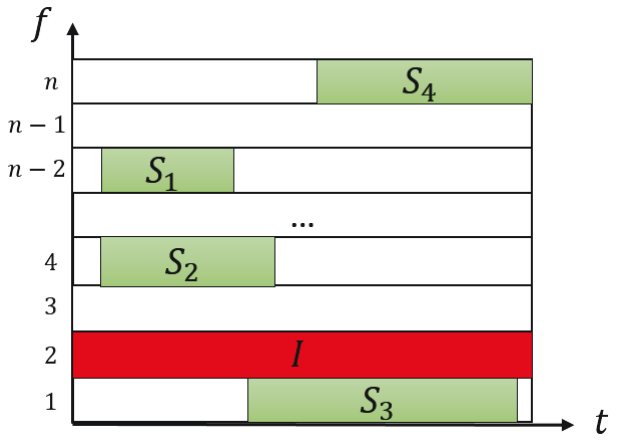


FIGURE 3.3: Example of a random channel assignment

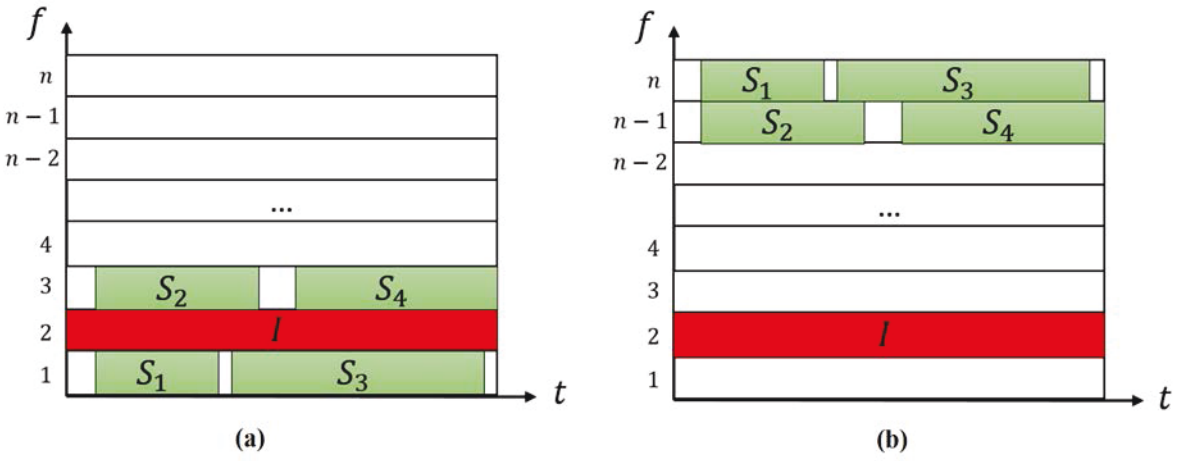


FIGURE 3.4: Examples of an ordered channel assignment:
(a) ascending order, (b) descending order

Query Rate of Secondary Users

The number of queries allowed per secondary user, or more generally the query rate of a secondary user, has an impact on the length of the inference process. By bounding that number to a certain threshold, we can limit the inferred knowledge of the attacker and slow down the acquisition of additional information about the spectrum.

3.4.2 Explicit Obfuscation

The SAS can implement explicit obfuscation in addition to inherent obfuscation in order to increase the privacy of the incumbents. In our system, obfuscation can be achieved by removing channel availability, i.e., removing additional channels from the list of idle channels. Hence, some idle channels are intentionally left vacant. This will increase both the uncertainty of the attacker (i.e., probability of inference) and the blocking probability of the system (i.e., denial of access).

Removed channels can be either adjacent channels, random non-adjacent channels or channels chosen according to some performance criteria. For example, the quality of the channel can be a selection parameter of the obfuscated channels (e.g., obfuscate the channels having the lowest QoS to mitigate significant spectrum loss).

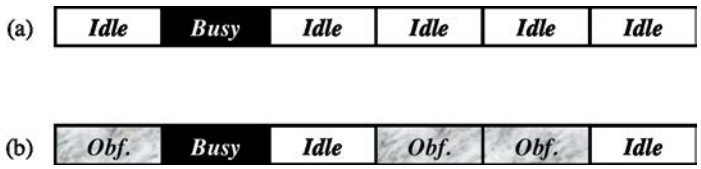


FIGURE 3.5: Example of the change in channel availability: (a) before obfuscation, (b) after obfuscation

Fig. 3.5 shows an example of obfuscation. It includes a SAS managing six channels and an incumbent operating on a single channel. Five channels are thus idle. Obfuscation can be applied by allocating only two channels for secondary use. The other three channels are not assigned to any secondary user during the operational period of the primary user. A malicious secondary user, who is trying to infer the operational channel, will assume at the end of the inference process that the incumbent is occupying four channels. In this case, the attacker's confidence about the channel used by the incumbent

(i.e., probability of inference) has decreased from 100 % to 25 %. The malicious user hence does not know which of the four channels the incumbent is using.

3.5 Analytical Model

Since we want to determine the expected number of channel requests an attacker must make to infer the operational channel of the incumbent, this problem is equivalent to the well-known *coupon collector's problem*.

3.5.1 The Coupon Collector Problem

The coupon collector problem is a well-konwn problem in probability based on a contest, where an individual purchases product boxes containing a coupon. If someone succeeds in collecting all distinct coupons from one set, he/she wins the big price offered by that product company. So, what is the expected number of boxes a coupon collector must purchase to complete a set of coupons? And what are the chances of collecting a complete set of coupons after purchasing a certain number of boxes?

Several researchers ([97][98][99][100][101][102][103][104]) have studied this problem and come up with different approaches and approximations to compute the total number of trials needed to complete a set of coupons and the probability of collecting a set of coupons for given a number of trials as well.

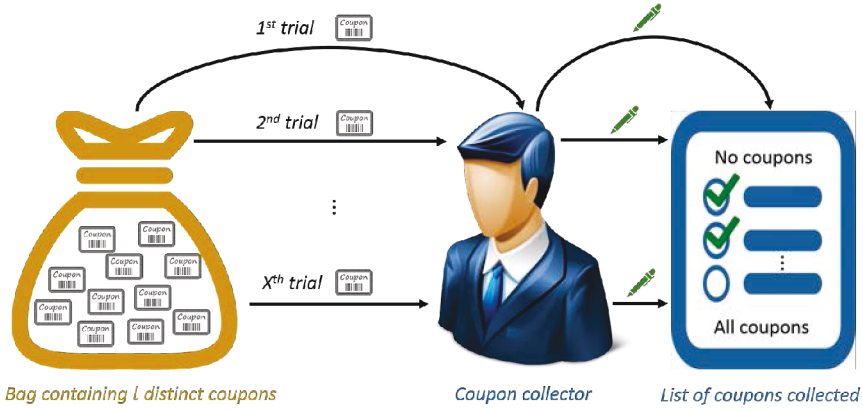


FIGURE 3.6: Example of the coupon collector problem)

In fact, this problem has several variants, but the basic form is shown in Fig. 3.6. An individual wants to collect a complete set of l different types of coupons. The coupons

arrive in sequence. The type of each coupon is a random variable where type j occurs with probability p_j. It is intuitive that the first few coupons are easy to collect. The last few coupons, however, are the most time-consuming. With each trial, the collector acquires multiple coupons of the same type and it becomes more difficult to acquire a different type.

3.5.2 The Channel Collector Problem

Fig. 3.7 shows how an adversary will proceed to infer the incumbent channel. Only two SAS channel assignment schemes are considered in our analysis: one where the SAS assigns an idle channel at random, and the other where the SAS assigns the lowest-numbered available (i.e., idle) channel. In both cases, the SAS returns only one channel per request ($m = 1$). Then, we show how to compute the expected number of channel requests to infer the incumbent's operational channel.

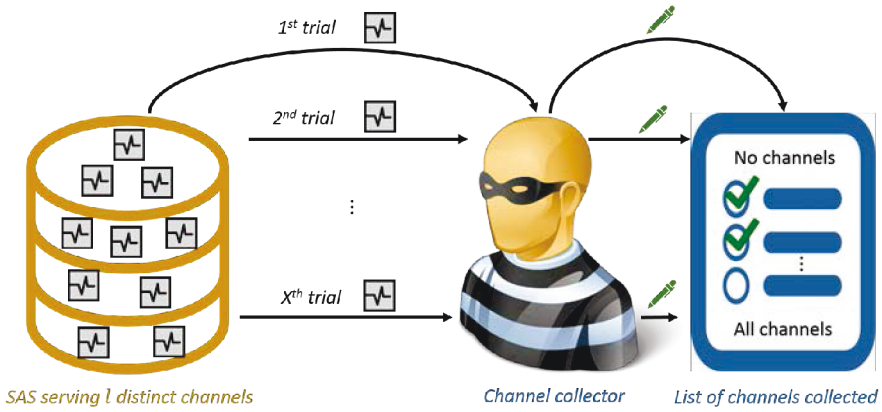


FIGURE 3.7: Example of the channel collector problem)

Analysis of the Random Channel Assignment

We first consider the case where the SAS returns a channel at random in reply to an attacker's request. That is, when an attacker requests a channel, there is an equal probability of any one of the idle channels being returned by the SAS. When there are no other secondaries besides the attacker making requests for spectrum, all l channels are idle, and each channel is returned by the SAS with probability $1/l$. The expected

number of channel requests can be calculated directly using the solution to the coupon collector's problem with equal probabilities.

The only variable is l, the number of channels available for use by secondaries. If the random variable X represents the number of requests an attacker needs to make in order to receive each of the l idle channels at least once in a response from the SAS, then

$$E[X] = l \sum_{k=1}^{l} \frac{1}{l-k+1} = l \sum_{k=1}^{l} \frac{1}{k} = lH_l \tag{3.11}$$

where H_l is the l^{th} harmonic number [97].

We can also compute the number of requests an attacker needs to make in order to receive $i \leq l$ idle channels at least once in a response from the SAS. If the random variable X_i represents the number of queries needed to collect i channels, a generalized formula is generated from Eq. 3.11 is generated as follows

$$E[X_i] = l \sum_{k=1}^{i} \frac{1}{l-k+1} = l \sum_{k=0}^{i-1} \frac{1}{l-k} \tag{3.12}$$

In order to determine the expected number of channel requests needed to infer the operational channel of the incumbent when there are other secondary users besides the attacker, we consider the $M/M/l/l$ queuing system model discussed in Section 3.3.1. The blocking probability P_B is the probability that all channels are busy at the time of a request and is calculated, using Eq. 3.4 and Eq. 3.5 as

$$P_B = \frac{\frac{\rho^l}{l!}}{\sum_{k=0}^{l} \frac{\rho^k}{k!}} \tag{3.13}$$

where $\rho = \frac{\lambda}{\mu}$ is the system load.

In the above system, when the attacker makes a request, the SAS will either return one of the available idle channels or, if all the channels are busy, will respond saying no channel is available. In this case we can express X as

$$X = X_b + X_r \tag{3.14}$$

where X_b is the number of times all channels were busy when the attacker made the request, i.e., the request was blocked, and X_r is the number of times an available idle channel was returned by the SAS. We know that

$$E[X] = E[X_b] + E[X_r] \tag{3.15}$$

and that

$$E[X_b] = P_B E[X] \tag{3.16}$$

Therefore,

$$E[X_r] = (1 - P_B)E[X] \tag{3.17}$$

and thus

$$E[X] = \frac{E[X_r]}{1 - P_B} \tag{3.18}$$

All that remains is to calculate $E[X_r]$. We observe that when $1 \leq k \leq l$ channels are idle, then each channel is idle with probability $\frac{k}{l}$, and if idle will be returned by the SAS with probability $\frac{1}{k}$. Thus each channel has probability $\frac{1}{l}$ being the one returned by the SAS. Therefore, we can calculate $E[X_r]$ using Eq. 3.11 and thus have

$$E[X] = \frac{lH_l}{1 - P_B} \tag{3.19}$$

The generalized formula to calculate the expected number of queries needed to collect $i \leq l$ channels given the blocking probability of the system P_B becomes

$$E[X_i] = \frac{l \sum_{k=0}^{i-1} \frac{1}{l-k}}{1 - P_B} \tag{3.20}$$

Analysis of the Ordered Channel Assignment

In the case where the SAS assigns the lowest available idle channel instead of a random available channel, however, it is not true that each channel is assigned to an incoming request with probability $\frac{1}{l}$. In general these probabilities are unequal. We need to find the probability p_j that the SAS will return channel j in reply to an attacker's request. This is equivalent to the probability that channel j is the lowest available idle channel

at the time of an attacker's request. In order to calculate this we use the same $M/M/l/l$ queuing system model as above. We want to find the steady-state, or equilibrium, probability that the system is in state j, defined as the state where the $j-1$ lowest index channels are busy and channel j is idle.

We assume that the attacker's request rate is sufficiently smaller than that of the aggregate request rate of all other secondaries, so that each time the attacker requests a channel, the SAS will return channel j with p_j (i.e., the steady-state probability).

In the derivation that follows, we use the approach introduced by Cooper in his paper on queues with s ordered heterogeneous servers, that is, servers with different service rates [105]. This is a more general system than ours, though he did not calculate the probabilities we are interested in.

Let B_j be the probability that an arriving request finds the first j channels busy. The conditional probability that an arriving request finding the first $j-1$ channels busy also finds channel j busy is B_j/B_{j-1}. If $\gamma_j(z)$ is the Laplace-Stieltjes transform of the distribution function of elapsed time between successive times when an arriving request finds the first $j-1$ channels busy and is assigned channel j, then

$$\gamma_j(\mu) = \frac{B_j}{B_{j-1}} \tag{3.21}$$

where $B_0 = 1$ and $\gamma_j(z)$ is defined by the recurrence relation

$$\gamma_{j+1}(z) = \frac{\gamma_j(z+\mu)}{1-\gamma_j(z)+\gamma_j(z+\mu)}, \quad j = 1, 2, \ldots \qquad \gamma_1(z) = \tfrac{\lambda}{\lambda+z} \tag{3.22}$$

Note that it follows from equation 3.21 and $B_0 = 1$ that

$$B_j = \gamma_1(\mu)\cdots\gamma_j(\mu),\ j = 1, 2, \ldots \tag{3.23}$$

Using the above, we can calculate the probabilities p_j. Let I_j be a random variable for the state of channel j, where a value of 1 means the channel is busy and 0 means

that it is idle. Then,

$$\begin{aligned} p_j &= Pr\{I_1 = 1, ..., I_{j-1} = 1, I_j = 0\} \\ &= Pr\{I_j = 0 | I_1 = 1, ..., I_{j-1} = 1\} Pr\{I_1 = 1, ..., I_{j-1} = 1\} \\ &= (1 - Pr\{I_j = 1 | I_1 = 1, ..., I_{j-1} = 1\}) Pr\{I_1 = 1, ..., I_{j-1} = 1\} \\ &= \left(1 - \frac{B_j}{B_{j-1}}\right) B_{j-1} \qquad (3.24) \\ &= B_{j-1} - B_j \qquad (3.25) \end{aligned}$$

Note that the blocking probability $P_B = B_l$ and that

$$P_B + \sum_{j=0}^{l} p_j = 1 \qquad (3.26)$$

We can find $E[X]$ by using the p_j as calculated above in the solution to the coupon collector's problem for unequal probabilities is

$$E[X] = \sum_{i=1}^{l} \frac{1}{p_i} - \sum_{i<j} \frac{1}{p_i + p_j} + \sum_{i<j<k} \frac{1}{p_i + p_j + p_k} - \cdots + (-1)^{l+1} \frac{1}{p_1 + \cdots + p_l} \qquad (3.27)$$

In order to compute the expected number of queries to collect $i \leq l$ channels, we consider the approach in [97]. Their approach assumes that the sum of the probabilities is equal to one ($\sum_j p_j = 1$). In our case, $P_B + \sum_j p_j = 1$. So, we recalculate the probabilities p_j by dividing them by $1 - P_B$.

Therefore, if X_r is the number of times an available idle channel was returned by the SAS, we have

$$E[X_r] = \sum_{k=1}^{i} \sum_{k_1 \neq k_2 \neq \cdot \neq k_{i-1}}^{l} \frac{p_{k_1} p_{k_2} \cdots p_{k_{i-1}}}{(1 - p_{k_1})(1 - p_{k_1} - p_{k_2}) \cdots (1 - p_{k_1} - \cdots - p_{k_{i-1}})} \qquad (3.28)$$

Knowing that X_i is the number requests an attacker needs to collect i channels and using Eq. 3.17, we conclude that

$$E[X_i] = \frac{E[X_r]}{1 - P_B} \qquad (3.29)$$

3.6 Simulation Experiments

Even though the analytical model provides the expected number of queries to infer the incumbent channel, the computation of some of the combinations is time-consuming, specifically for the ordered assignment scheme. So, as part of the vulnerability analysis, we also run simulations to verify and extend the analytical results.

3.6.1 Simulation Assumptions

We assume that the band is divided into n equal channels. We also assume that our system includes one incumbent, one SAS and one adversary (attacker) within the same area. The incumbent is operating on a single channel. Secondaries share the use of the remaining $n-1$ channels, and the SAS manages access to those channels. The SAS usually returns a list of m available channels per query for a secondary to choose from. In the case where the value of m is greater than the number of idle channels at the time of the query, the SAS returns the list of available channels for use. In other words, the secondary users gets a list of $min(m, size(idle\ channels))$ channels.

We implement the $M/M/l/l$ Erlang loss queuing system described above. Thus, secondaries query the SAS according to a Poisson process with aggregate rate λ, and the service time is exponentially distributed with rate μ.

The attacker is one of the secondaries, and is trying to infer the operational channel of the incumbent. Secondary requests for spectrum opportunities from the SAS are not limited. In order to address the vulnerability of the incumbent, we consider that the channels available for use by secondaries do not change over time, i.e., the incumbent is using the same channel for a long period of time. In a first analysis, the SAS replies with an available channel for each query ($m = 1$). Later, we vary the number of available channels returned by the SAS for each query ($m > 1$). The attacker does not know a priori the channel assignment scheme used by the SAS. And it only uses the information given by the SAS and does not have access to any external knowledge.

We use the models previously presented and implement an example system in Python. We simulate the query/response process using SimPy, which is a process-based discrete-event simulation framework in Python. Then, we run and average 150 attack trials.

Table 3.1 summarizes the different parameters of the simulation.

TABLE 3.1: Simulation specification

Parameter	Definition
n	Total number of channels
$n - l = 1$	Number of channels occupied by the incumbent
$l = n - 1$	Number of channels available for secondary use
m	Number of channels returned per secondary request
λ	Aggregate arrival rate of secondary users
λ_a	Aggregate arrival rate of attackers
$1/\mu$	Individual service time of secondary users
$\rho = \lambda/\mu$	System load

Next, we address the following questions:

- How long does it take for an attacker to infer the incumbent frequency?
- How does varying the system load affect the inference process?
- How does varying the attacker query rate affect the time to discovery?
- How does obfuscation affect the privacy of the incumbent?

3.6.2 Comparison Between Simulation Results and Analytical Results

Using the analytical model presented in Section 3.5.2, we can calculate the expected number of queries to discovery. We also run simulations to analyze the inference process for multiple scenario cases ($n = 10, 15, 20, 25, 30$ and $m = 1$).

In figures 3.8 and 3.9, we provide a comparison between the calculated results and the simulated results for both random and ordered channel assignment schemes. Hence, we vary the system load ρ and compute the expected number of queries to discovery (analytically) and the average number of queries to discovery (by simulation) for different values of the total number of channels n and the system load ρ. In Fig. 3.8, where channels are randomly assigned, we notice that the simulation results match the analytical model results for all values of ρ. In Fig. 3.9, where channels are assigned order-wise, simulation results match the analytical model results for high values of ρ. For low values of ρ, the simulations take too long to run. We note that all results are matching and the values are within the confidence interval.

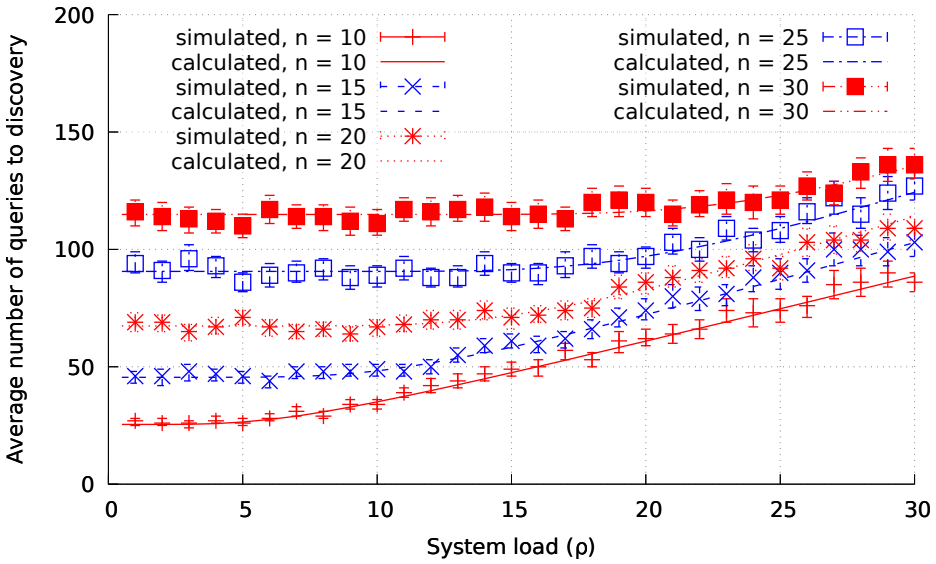


FIGURE 3.8: Spectrum load vs. Number of queries (random channel assignment scheme, $m = 1$)

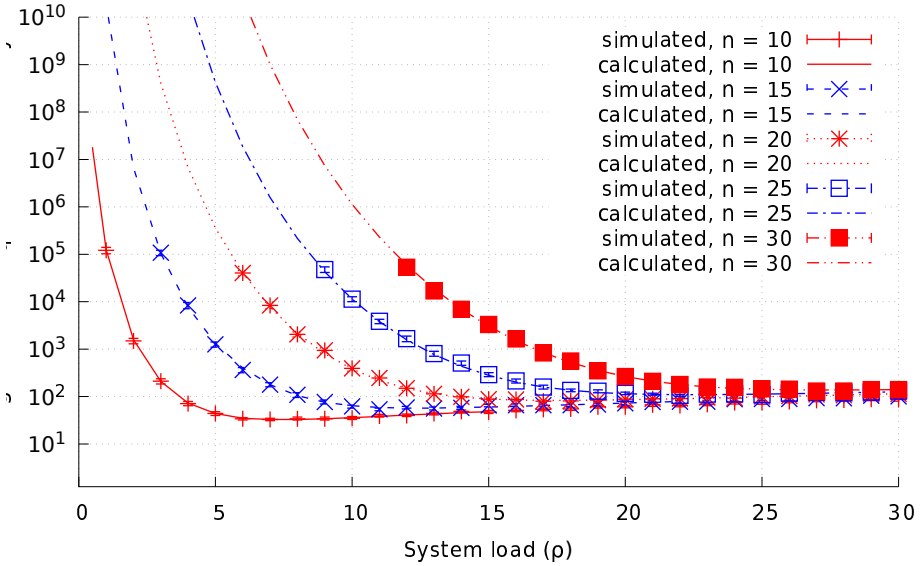


FIGURE 3.9: Spectrum load vs. Number of queries (ordered channel assignment scheme, $m = 1$)

Detailed tables comparing simulation results and analytical results for both the random and ordered channel assignment schemes for $m = 1$ are in Appendix C.

3.6.3 Results Analysis

Effect of the Channel Assignment Scheme

We consider the effect of overall system load, defined in terms of $\rho = \frac{\lambda}{\mu}$, and the channel assignment scheme. Figures 3.10 and 3.11 show the inference distance as a function of time, i.e., how close an attacker is to inferring the incumbent's channel over time, for $n = 10$ and $n = 50$ channels, respectively. The aggregate query rate of all secondaries is λ. We define the attacker's query rate λ_a as the fraction $\frac{1}{10n}$ of the aggregate rate λ. We plot results for the three assignment schemes and ρ values equal to 10%, 30% and 80% of the number of channels. Time is expressed in units of $\frac{1}{\mu}$.

Figures 3.10 and 3.11 both indicate that as the system load increases, the attacker can more quickly infer the incumbent's channel. Equivalently, it requires less time to discover the incumbent's channel when the system load is higher.

The effect of using different channel assignment schemes is even more dramatic. This is especially true for a low system load. Under the high system load, the average time of discovery is almost the same for both channel assignment schemes, though we see some separation for the last few channels in the $n = 50$ case (Fig. 3.11). The inference for the semi-static channel assignment scheme takes longer than the random and the ordered schemes. At low system load, we see that inference time is longer by orders of magnitude with ordered and the semi-static assignments. It takes much longer for an attacker to visit all channels of a lightly loaded system when they are assigned semi-statically or order-wise rather than randomly. However, when comparing the ordered and semi-static schemes, we note that the ordered channel assignment performs better than the semi-static assignment. Thus, overall, the ordered assignment scheme protects the incumbent's frequency from inference attacks significantly better than the random assignment. The semi-static assignment scheme protects the incumbent's frequency better than the ordered assignment.

In Figures 3.12, 3.13 and 3.14, we analyze the effect of increasing the number of channels returned by the SAS per query. Therefore, in each figure, we fix the system

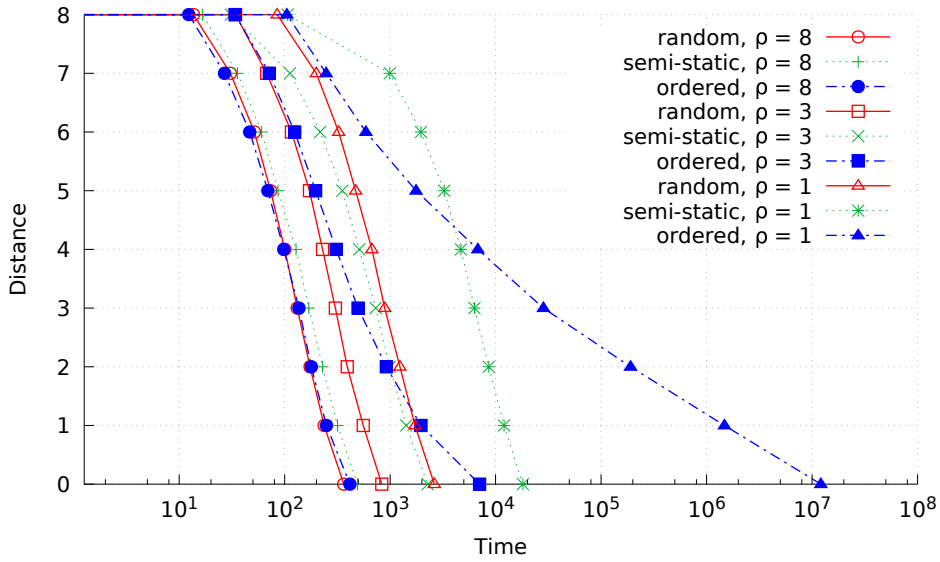


FIGURE 3.10: Inference distance vs. Time ($n = 10$, $m = 1$, different channel assignment schemes, no obfuscation implemented)

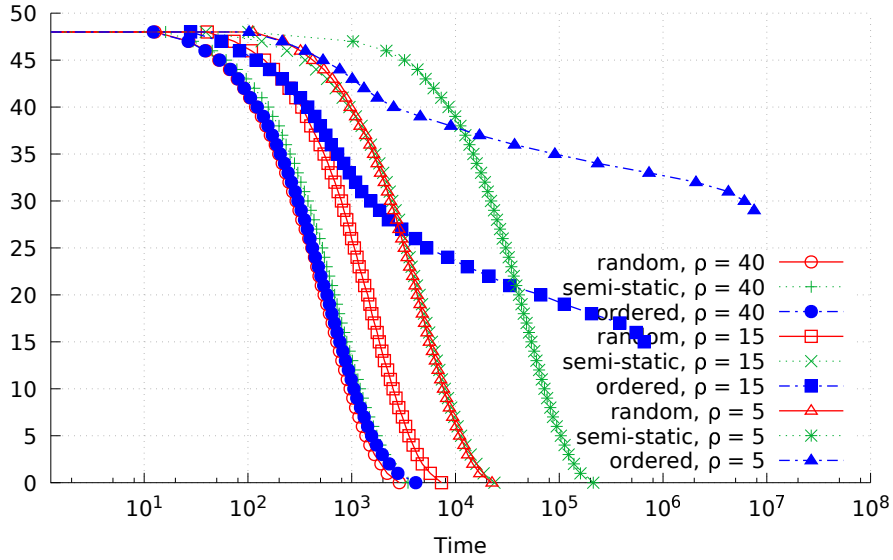


FIGURE 3.11: Inference distance vs. Time ($n = 50$, $m = 1$, different channel assignment schemes, no obfuscation implemented)

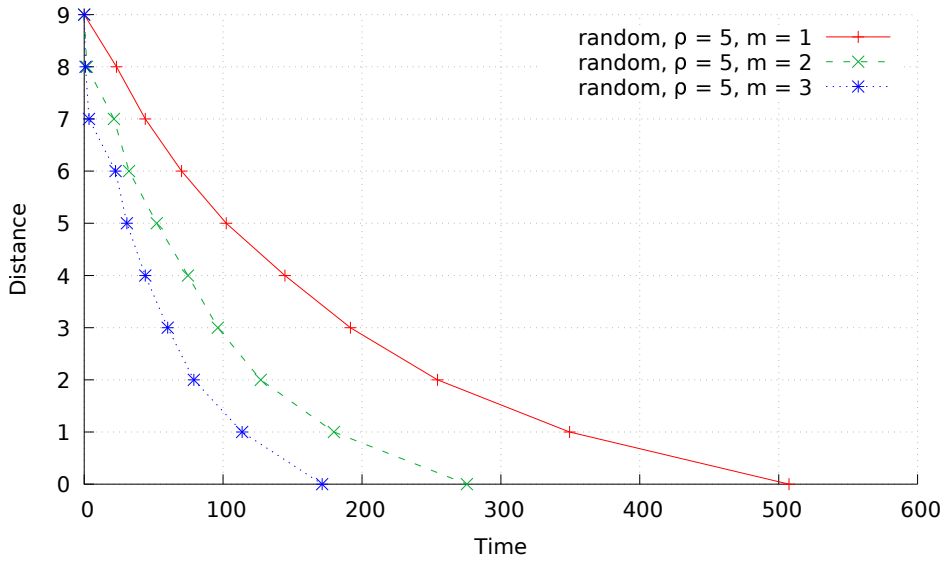


FIGURE 3.12: Inference distance vs. Time ($n = 10$, $\rho = 5$, random channel assignment scheme, no obfuscation implemented)

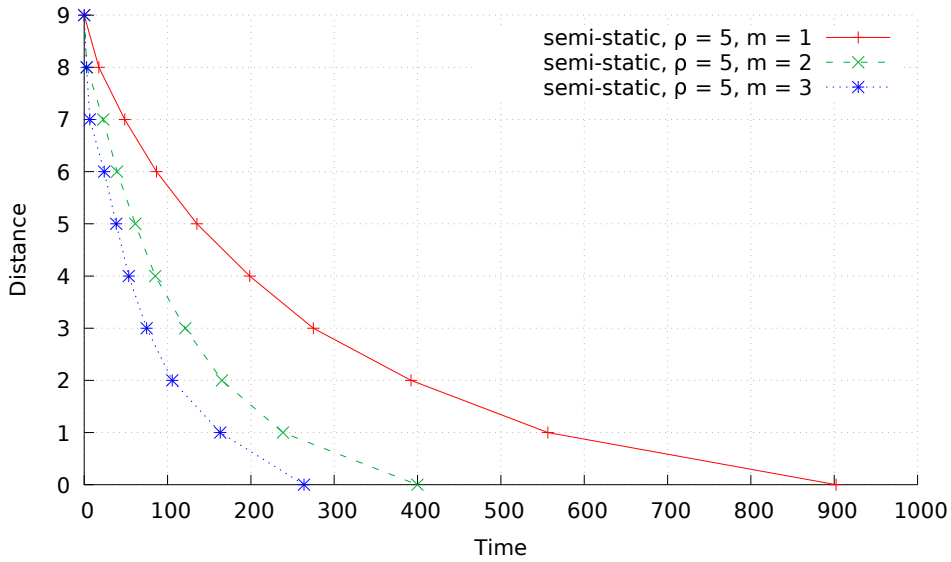


FIGURE 3.13: Inference distance vs. Time ($n = 10$, $\rho = 5$, semi-static channel assignment scheme, no obfuscation implemented)

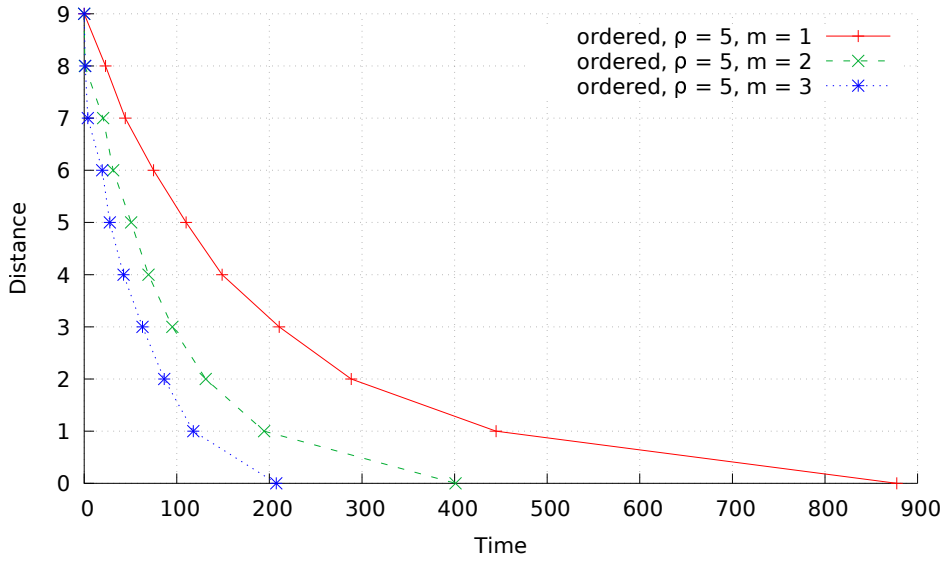


FIGURE 3.14: Inference distance vs. Time ($n = 10$, $\rho = 5$, ordered channel assignment scheme, no obfuscation implemented)

load ($\rho = 5$) and the channel assignment scheme. We conclude that the trend is similar for all assignment schemes: when m increases, the time of inference decreases. Those results also confirm the fact the semi-static assignment scheme performs better than the ordered assignment scheme for high system loads.

We conclude that the ordered channel assignment scheme guarantees better protection of the incumbent frequency for low and medium system loads by increasing the time to inference. We also conclude that returning just one channel per secondary request limits diversity, and hence increase the time to inference and privacy.

Effect of the Spectrum Load

Returning to the overall number of queries to discovery metric, we vary the system load ρ and show the average number of queries to discovery (i.e., the average number of queries needed by the attacker to infer the incumbent's channel). Fig. 3.10 provides analogous plots to those in Figures 3.8 and 3.9.

For a random assignment scheme, the average number of queries to discovery increases monotonically as the system load increases. When the system is over-loaded, the blocking probability P_B is high and the attacker's knowledge takes longer to be updated. This conduct is emphasized in Fig. 3.17.

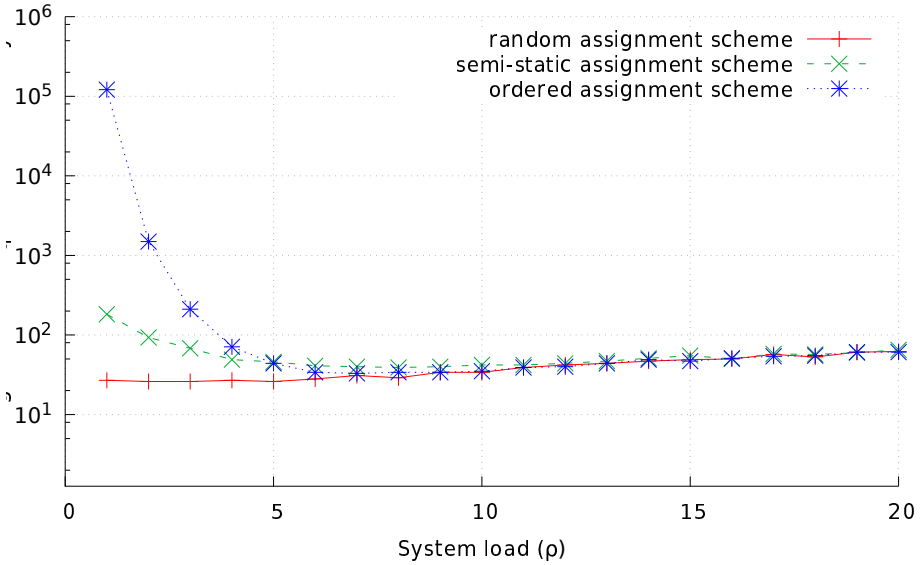


FIGURE 3.15: Average number of queries to infer the incumbent's channel vs. System load ($n = 10$, $m = 1$, different channel assignment schemes, no obfuscation implemented)

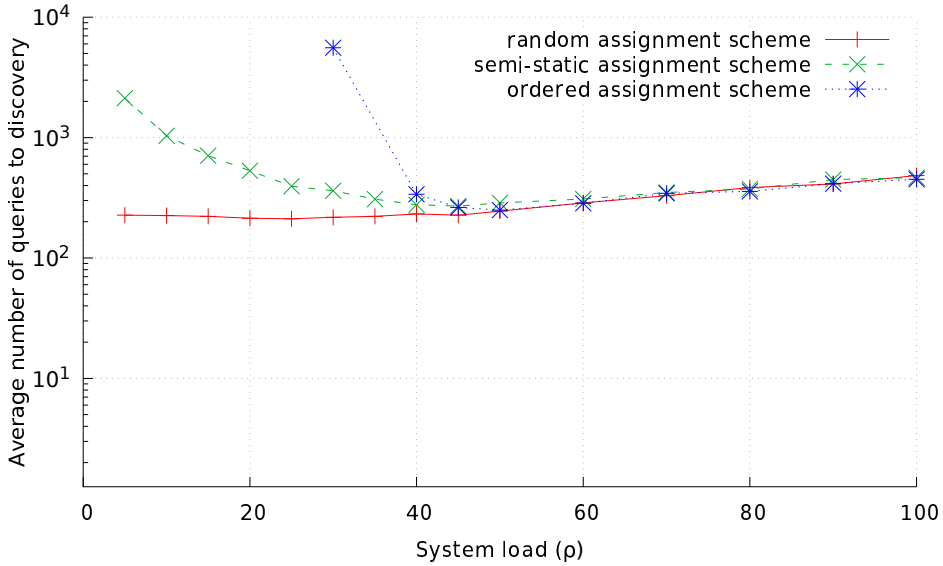


FIGURE 3.16: Average number of queries to infer the incumbent's channel vs. System load ($n = 50$, $m = 1$, different channel assignment schemes, no obfuscation implemented)

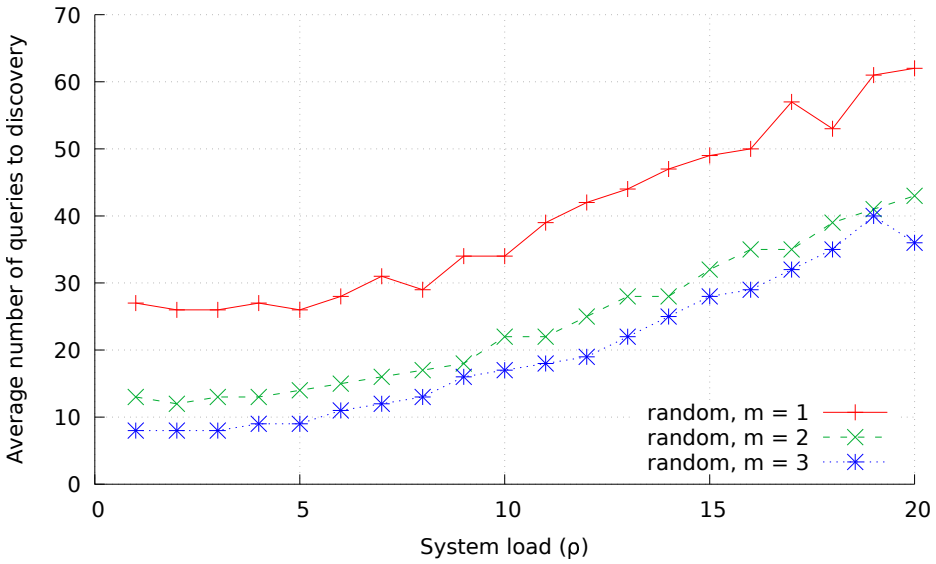


FIGURE 3.17: Average number of queries to infer the incumbent's channel vs. System load ($n = 10$, random channel assignment schemes, no obfuscation implemented)

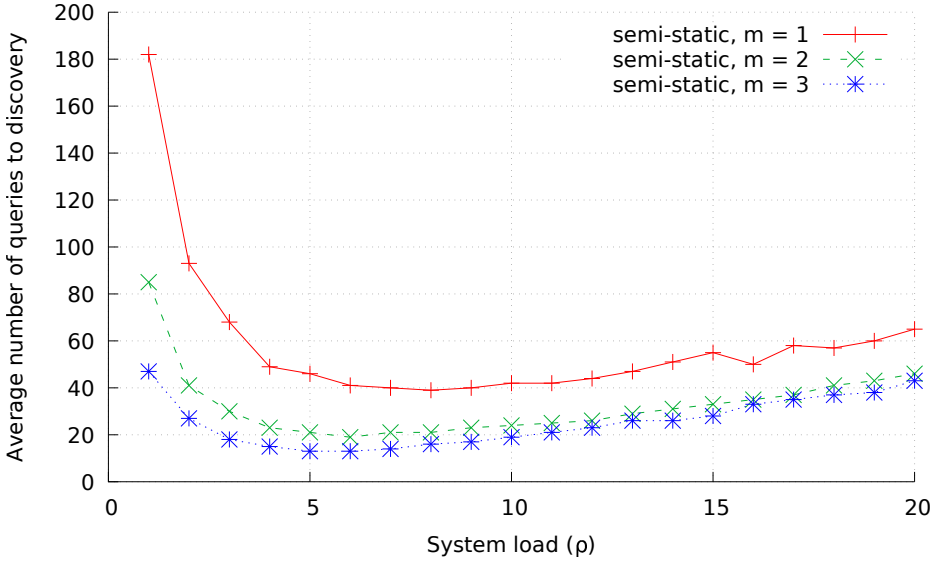


FIGURE 3.18: Average number of queries to infer the incumbent's channel vs. System load ($n = 10$, semi-static channel assignment schemes, no obfuscation implemented)

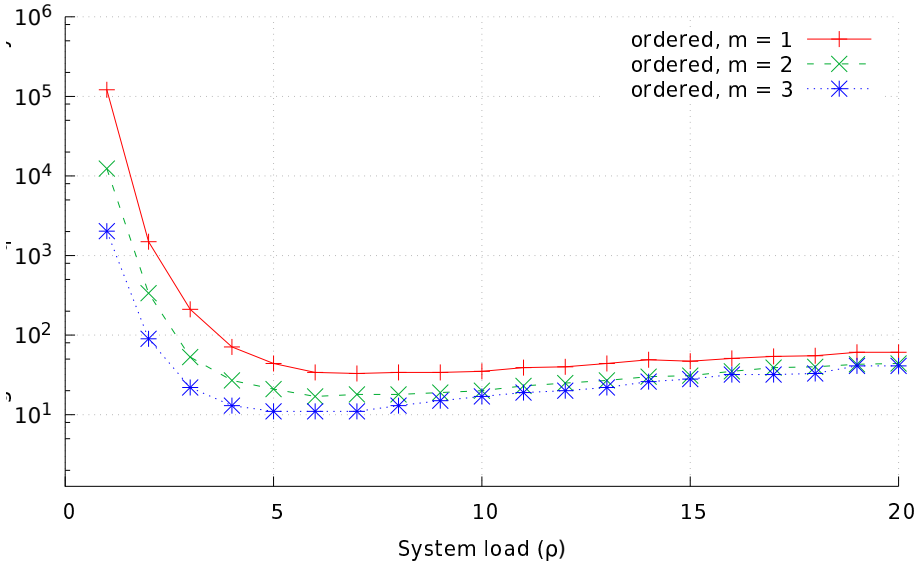


FIGURE 3.19: Average number of queries to infer the incumbent's channel vs. System load ($n = 10$, ordered channel assignment schemes, no obfuscation implemented)

The behavior of the ordered assignment scheme is significantly different. The average number of queries to discovery is very high for low values of ρ, then decreases dramatically, hits a minimum near $\rho = n$, then increases again to match the values of the random assignment scheme. Since the channel assignment is ordered, when the system load ρ is low (i.e., the number of secondaries is limited), the attacker takes longer to infer the incumbent's channel. As the value of ρ approaches n, the privacy of the incumbent declines considerably.

The semi-static assignment scheme's behavior is similar to the behavior of the ordered scheme. The average number of queries to discovery decreases when the system load increases ($\rho < n$). Then, it stabilizes after $\rho = n$ to match the values of the random and ordered schemes, i.e., the system is over-loaded, and the choice of channel assignment scheme makes no difference. However, interestingly, the attack in the semi-static scheme is slightly more costly than both the random and the ordered schemes for $\frac{n}{2} < \rho < n$.

Figures 3.17, 3.18, 3.19 re-plot the average number of queries to discovery as a function of the system load but take into consideration the effect of m (the number of channels returned per query). We see that the best case is obviously when the SAS returns only one channel per query ($m = 1$). When m increases, the number of queries decreases for

all system load values. However, for the semi-static channel assignment (Fig. 3.18), the difference in the number of queries between $m = 2$ and $m = 3$ is remarkably small. For the ordered channel assignment (Fig. 3.19), the cost of inference is significantly reduced with the increase in m. These plots also provide an illustration of the impact of m for higher values of ρ. Since the system is overloaded, the SAS returns less or equal to m channels per request. Hence, the difference in the number of queries to discovery for $m = 2$ and $m = 3$ is relatively small.

The effect of the system load is important since it affects the privacy of the incumbent frequency. We conclude that when the system is overloaded (i.e., all channels are busy), the channel assignment scheme deployed does not matter anymore. Also, the results here confirm the previous conclusion about the effectiveness of returning just one channel per secondary request as it increases the number of queries to inference.

Effect of the Attacker's Query Rate

For Figures 3.20 and 3.21, we fix the system load to $\rho = 1$ and $\rho = 5$ and vary the query rate of the attacker λ_a.

We consider the random, semi-static and ordered channel assignment schemes as well. An attacker query rate of λ_a could represent a single aggressive attacker with query rate λ_a or multiple collaborating attackers with aggregate query rate λ_a. In the semi-static assignment, we only consider one attacker. Having different attackers under that condition gives results similar to the random assignment. Clearly, in the case of collaborative attackers, the semi-static assignment is no longer as efficient.

In both plots, we see how increasing the query rate accelerates the inference process. The attacker is acquiring more information in less time. These results provide insight into the limit that the SAS may set on the query rate in order to minimize the risk of inference and increase the privacy of the incumbent. For example, Figures 3.20 and 3.21 suggest that if in this system the incumbent needs $\frac{10}{\mu}$ time units to use the channel and leave, the SAS should limit a user's query rate to 3μ when using random channel assignment and to 10μ when using ordered channel assignment, under moderate load.

In Fig. 3.20, the system load is low ($\rho = \frac{n}{10}$). As a result, the gap shown between the three curves implies that a random or a semi-static assignment can facilitate the

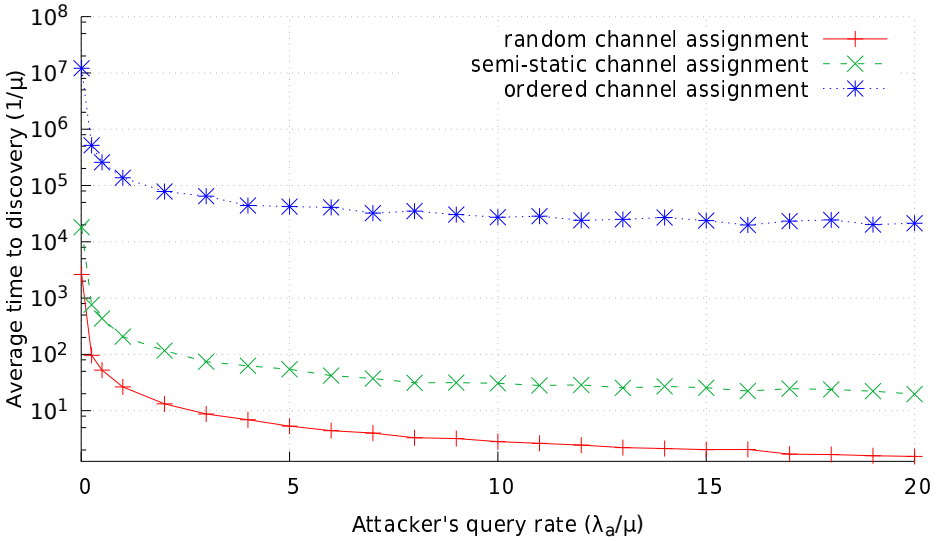


FIGURE 3.20: Average time to infer the incumbent's channel vs. Attacker's query rate ($n = 10$, $m = 1$, $\rho = 1$, no obfuscation implemented)

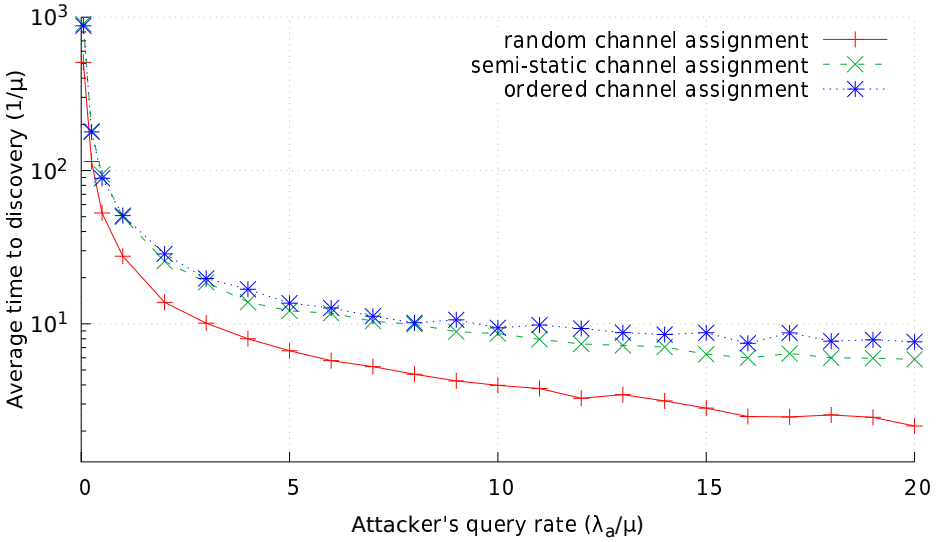


FIGURE 3.21: Average time to infer the incumbent's channel vs. Attacker's query rate ($n = 10$, $m = 1$, $\rho = 5$, no obfuscation implemented)

inference process, while an ordered assignment makes it effectively impossible.

In Fig. 3.21, the system load is relatively high ($\rho = \frac{n}{2}$). Consequently, the inference process has become faster. The average time of inference has specifically decreased for the ordered assignment.

We conclude that an attacker with multiple fake identities or other cooperative attackers can jeopardize the privacy of the incumbent channel by decreasing the time to inference, and hence optimizing the attack. The ordered channel assignment scheme can be a solution to this problem. However, bounding the secondary's query rate is also envisioned to ensure better privacy.

Effect of the Blocking Probability

For Figures 3.22 and 3.23, we implement obfuscation and analyze its impact on privacy.

In Fig. 3.22, we fix the total number of channels and vary the number of available channels (i.e., vary the blocking probability). The channel assignment scheme does not have any effect on the blocking probability. In this case, increasing the privacy comes with a price: less secondary access. For low system load (10%) and medium system load (30% and 50%), obfuscation can be extremely efficient by increasing the distance of inference without decreasing the availability of the system. So, the SAS might be able to reduce significantly the knowledge of the attacker without limiting secondary access. For high system load (80%), obfuscation has a bigger effect on the system availability. However, the SAS can still enhance the privacy by 30% when denying only 20% of the secondary's queries.

In Fig. 3.23, we fix the number of channels to $n = 10$, the system load to $\rho = 5$ and the channel assignment scheme to ordered. Then, we introduce obfuscation by varying the number of available channels (i.e., varying the blocking probability). When no obfuscation is implemented and 9 channels are available, the blocking probability P_B is equal to 0.0375 and the attacker can deduce the channel of the incumbent at the end of the inference process. But, as shown in Figure 3.23, when obfuscation is implemented by blocking 4 channels (leaving 5 channels available for the secondary users, with $P_B = 0.2849$) and blocking 6 channels (leaving 3 channels with $P_B = 0.5297$)

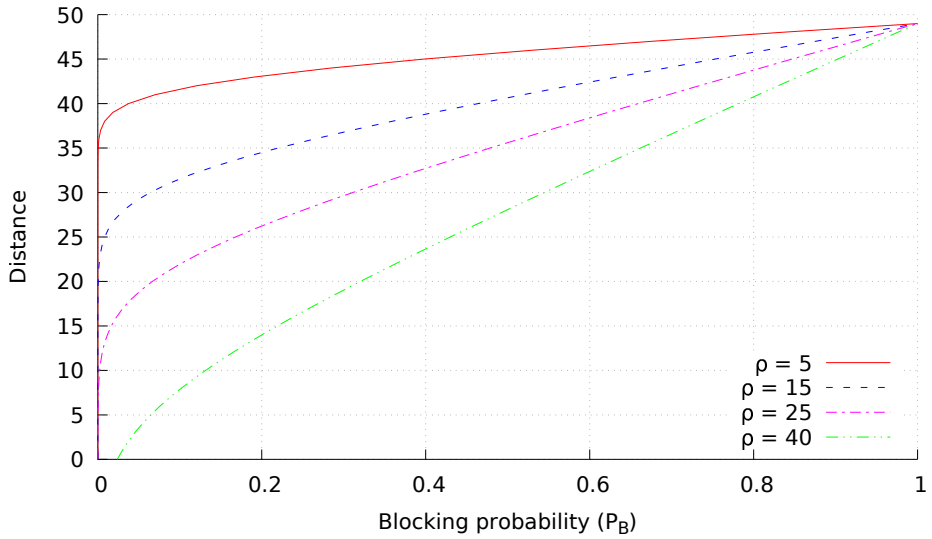


FIGURE 3.22: Inference distance vs. Blocking probability ($n = 50$, obfuscation implemented)

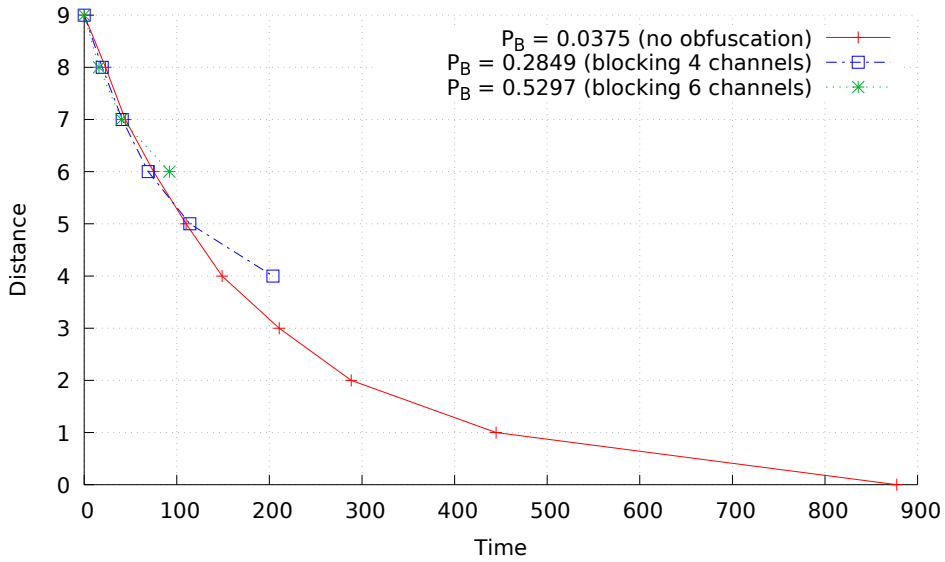


FIGURE 3.23: Inference distance vs. Time ($n = 10$, $\rho = 5$, ordered channel assignment, obfuscation implemented)

respectively, the attacker's knowledge peaks at a distance proportional to the number of channels blocked. We also note that, when obfuscation is introduced, the time of inference decreases at the beginning of the inference process but increases by the end. In fact, the channels are limited and the channel assignments from the SAS are repetitive. While this speeds up the discovery for the attacker at first, ultimately it is a barrier that slows it down.

We conclude that explicit obfuscation increases the privacy of the incumbent frequency by increasing the uncertainty of an attacker. Since this technique results in spectrum loss, it is recommended to implement it when the system resources are enough to serve all coming secondary requests.

3.7 Conclusions

This work has been conducted to analyze the vulnerability of the incumbent's frequency to inference attacks and determine inherent and explicit privacy-preserving techniques to thwart them. We design the secondary activity in the shared environment using an Erlang loss queueing system model, and we also design the attack algorithm. We conclude that the values of system parameters have a significant effect on the privacy of incumbents, either by decreasing the time to discovery or increasing the acquired knowledge of the attacker.

Implementing an ordered channel assignment scheme is effective in all cases, regardless of the spectrum load or the number of channels or even the attacker's query rate. Since randomness in the random channel assignment scheme adds diversity to the query responses, it allows the attacker to infer the incumbent's channel in less time. Hence, the use of the ordered scheme raises both the cost and the distance of inference, i.e., an attacker will take more time to infer the same information. An attacker can fake different identities, get help from other secondaries or just flood the system with queries in order to speed up the inference process. Limiting the secondary users query rate can be an efficient way to mitigate the inference of the incumbent frequency during the operational period. Introducing obfuscation will also enhance the privacy of the incumbent

by increasing the distance and the cost of inference to an attacker. Even though it limits the availability of the spectrum, it can be very efficient when the system load is low.

As a result, the SAS should implement effective privacy-preserving techniques by calibrating some adjustable parameters (channel assignment scheme, query rate of secondaries, number of available channels) to the system load or removing some channels from the table of available channels, taking into consideration its impact on the secondary access and the spectrum efficiency. The results obtained show that the effect of modifying these parameters is predictable, and this will allow the SAS to tune them according to the expected privacy level.

Inferring the operational frequency of the incumbent is the first step to infer other sensitive parameters (e.g., location). In fact, by doing so, the attacker knows that it is in the operational zone of the incumbent. All that remains is to find the boundaries of that zone. So, in the next chapter, we analyze the vulnerability of the incumbent's location to inference attacks given that the operational frequency is already known.

Chapter 4

Protection of the Incumbent's Location

4.1 Introduction

Protecting the frequency of operation of the incumbent is a first step to protect the incumbent from exposure. By just adjusting some spectrum parameters or including straightforward obfuscation features, the privacy can significantly increase. However, this is not enough. The location of the incumbent must be protected as well in order to increase its privacy.

The focus of this chapter is on the analysis of the vulnerability of the incumbent location to inference attacks and measuring the impact of some obfuscation approaches on privacy.

4.2 Background

The 3.5 GHz band (3550–3700 MHz) is exclusively allocated to authorized Federal and grandfathered FSS users. This allocation results in an inefficient spectrum usage as no one can use that band nationwide even when no incumbent is operating. In other words, no one can use the band anywhere within the boundaries even if the incumbent only operates in specific geographic locations. When sharing has been proposed, NTIA has defined exclusion zones to protect the incumbent against interference generated by secondary users. As stated in Chapter 2, Section 2.3.2, their approach is straightforward:

no secondary users are allowed in the exclusion zones. The spectrum access system (SAS) thus only serves and manages the secondaries outside the exclusion zones.

The exclusion zones are computed based on a fixed deployment of secondary users [73]. However, the network is dynamic because of the mobility of the devices and the changeability of their parameters. Also, the exclusion zones are employed along the coastlines and have been already reduced by 60 % as revised by NTIA and stated in Chapter 2. However, the current configuration still excludes a large number of users as the U.S. population density is high in coastal areas. In fact, 39 % of the total population in 2010 is concentrated along the coast, and this number is expected to increase by 8 % between 2010 and 2020 [106]. Fig. 4.1 and Fig. 4.2 display the U.S. population and are generated using data provided by the U.S. Census Bureau [107]. Fig. 4.1 shows the population density in the U.S. (people per square km of land area). Fig. 4.2 confirms that almost 40 % of the U.S. population lives within 100 km of the coast. Therefore, implementing exclusion zones is too conservative to meet the expectations of sharing in the 3.5 GHz band.

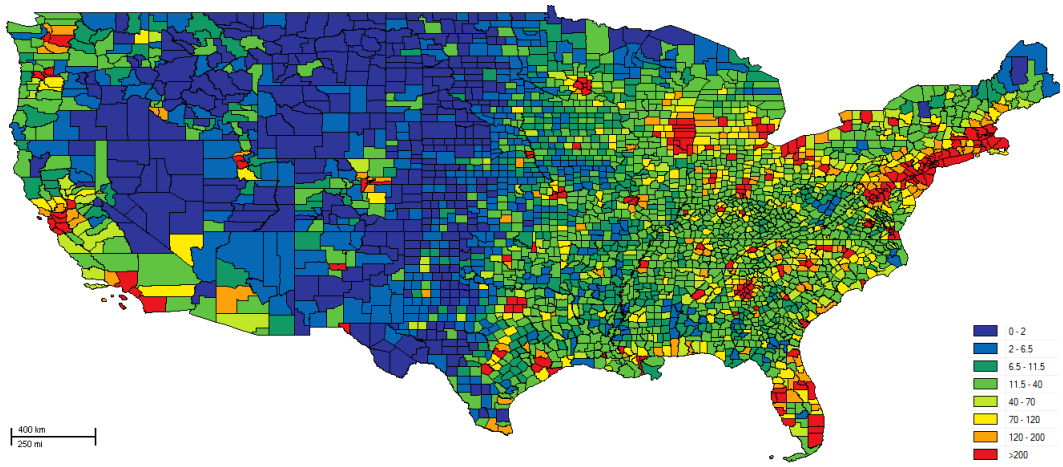


FIGURE 4.1: Population density in the United States (people per sq. km of land area)

To allow more spectrum availability, protection zones have been proposed. The deployment of the ESC reduces the exclusion zone to a protection zone. In fact, the ESC is a system associated to the SAS and is designed to increase spectrum availability for secondary usage by determining the incumbent activity and identifying the operational channel. The SAS decides then whether a secondary user within the protection zone

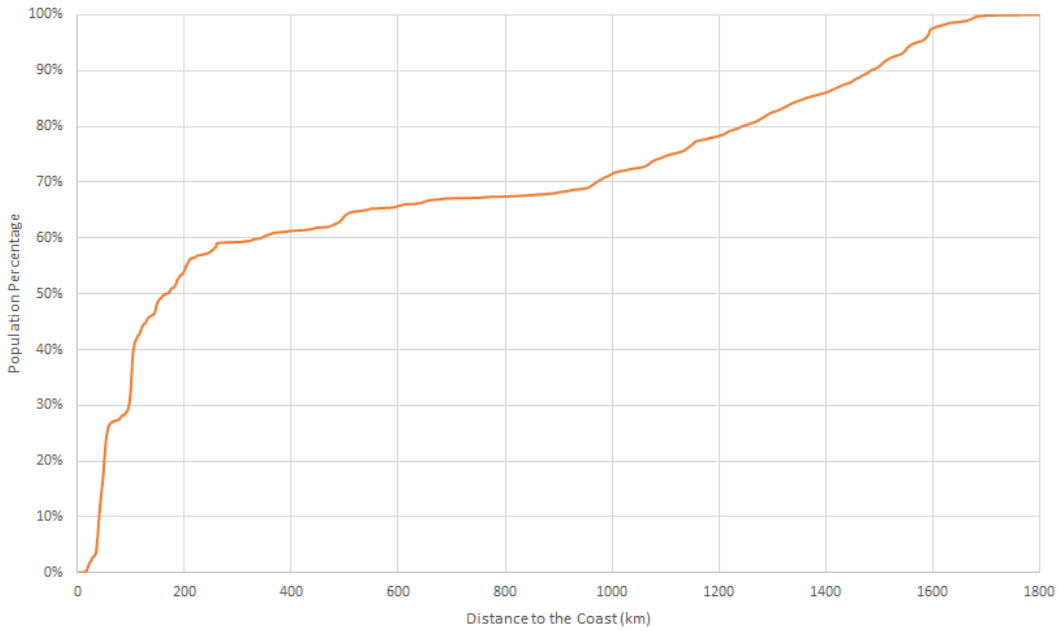


FIGURE 4.2: Distance to the coast vs. Population percentage

can operate or not. In this case, the secondary users are able to operate as long as their operations does not cause interference for the incumbent. By including the ESC, the white spaces in the 3.5 GHz band are efficiently utilized, and the incumbent is protected from harmful interference. Fig. 4.3 shows how protection zones are envisioned. We consider an exclusion zone within an area of interest. Based on the ESC sensors location and the secondaries deployment, the SAS divides the exclusion zone into multiple protection zones. Note that not all protection zones have the same size. When the ESC sensor determines that an incumbent is operating, the SAS shuts down one or multiple protection zones associated with the operational zone of the incumbent. In other words, secondaries operating on the incumbent's channel will be directed to other channels or denied access to the spectrum.

For the sake of clarity, we also define the detection zone. It is the area of detection of the incumbent and is associated with a set of protection zones. In Fig. 4.4, we present an example of the detection zone. In this example, each detection zone (violet area) is associated to one protection zone (green area) and each protection zone is monitored by one ESC. Once the incumbent enters the detection zone, the corresponding protection zones are shut down to avoid any interference at the incumbent's receiver.

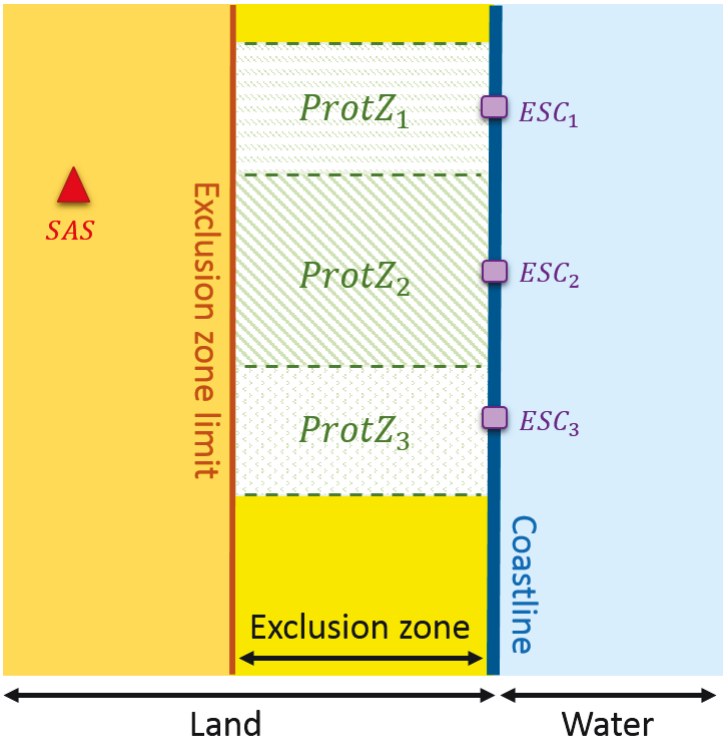


FIGURE 4.3: Exclusion zones vs. Protection zones

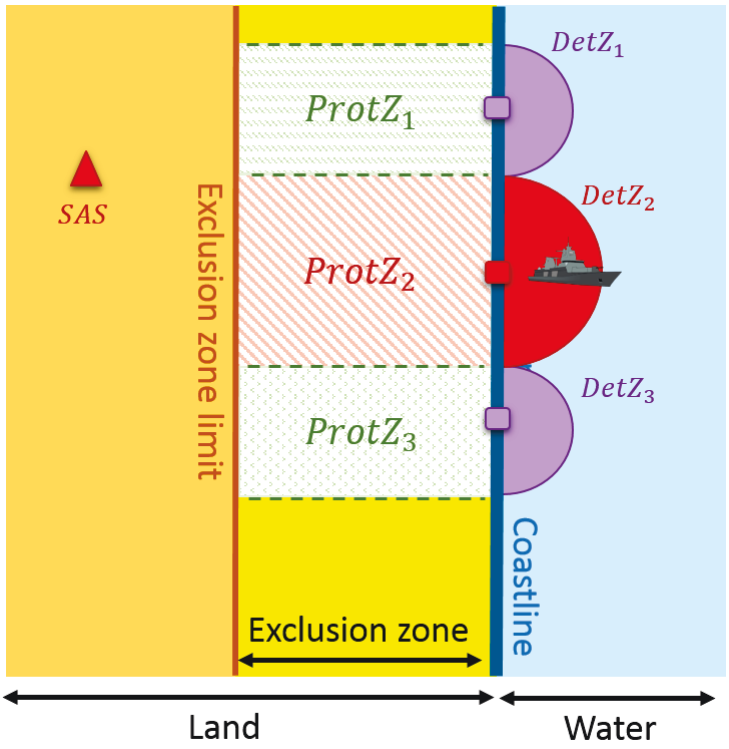


FIGURE 4.4: Detection zone of the incumbent

For the remainder of this chapter, we assume that fixed protection zones are deployed to protect the incumbent from possible interference while allowing secondary access to the spectrum. Nonetheless, such implementation does not guarantee the security of the incumbent operations, and it is essential to ensure the operational security as well.

4.3 State of the Art

The WInnForum working group on security has set requirements about the uncertainty of the incumbent location. The operational location should be protected against inferences and disclosure. In fact, its uncertainty should not fall below some threshold (e.g., 110 km) for all entities in the shared environment [86].

Bahrak et al. [85], Clark et al. [93] and work on the previous chapter have adopted obfuscation techniques to protect the incumbent's privacy. Their work was inspired by existing work on data mining and data publishing [77][78], which presents obfuscation as a technique to increase the anonymity of data owners and hence mitigate inference attacks.

Bahrak et al. [85] propose obfuscation in order to enable sharing while protecting the location and time of operation of the incumbent. Clark et al. [93] investigate different location inference attack strategies and also propose obfuscation to maintain a critical level of privacy for the incumbent. Both works recommend techniques that increase the uncertainty of an adversary, by enlarging the protection zone, incorporating false information, or adding noise to the incumbent's parameters. Such approaches result in the loss of spectrum for other non-malicious secondary users. Then, sharing will be limited and spectrum efficiency decreased. Also, their approaches consider an inland incumbent. However, for the 3.5 GHz band, the incumbent is the shipborne radar (currently, the AN/SPN-43 [73]), which is the Navy's air traffic control radar system. In their models, authors assume that the attacker knows the power assignment function of the SAS. They also assume that the incumbent provides its location to the SAS. In the 3.5 GHz, based on privacy requirements [86], the location should remain unknown for both the sensing system and the management system to prevent internal breaches.

Chapter 3 evaluates the vulnerability of the frequency of the incumbent to inference attacks analytically and through simulation, then proposes inherent, in addition to explicit, obfuscation measures in order to mitigate such attacks. This work is different from others as the authors demonstrated that tuning some of the environment's parameters such the query rate and the channel assignment scheme can highly increase the privacy without impacting spectrum resources.

Another aspect that has not been studied in previous work is identifying a risk and assessing it. Intrusion Detection and Prevention Systems (IDPS) have been presented in the context of network or computer security to monitor and manage emerging threats (i.e., suspicious activities). A guide to detecting and preventing intrusions has been developed by the National Institute of Standards and Technology (NIST) [108]. It discusses the capabilities of IDPS and provides recommendations for its use for computer security in enterprises. In network security, one way to identify a threat is to evaluate the trust of network participants and hence secure the network from malicious behaviors. Specifically, trust management has been proposed in networks that require collaboration between different entities, such as distributed networks. IDPS were also proposed for cognitive radio systems to ensure threat-free sharing between primary and secondary users [109] [110]. The following specific attacks were studied: Primary User Emulation (PUE) attacks [55] [111] [112], attacks to cooperative sensing mechanisms [113] [114], and Objective Function (OF) attacks [93]. Such manipulative attacks are carried out by insiders and aim to corrupt the behavior of a cognitive radio system by altering sensing results and disrupting communications. IDPS solutions are usually statistical as they depend on observations from cooperating network nodes.

4.4 Problem Statement

This section presents the spectrum access model and the threat model used in our analysis to address the vulnerability of the incumbent location to inference attacks.

4.4.1 System Model

In this chapter, we consider the work done by NTIA [73] in order to calculate the exclusion zone while taking into account the deployment of and the technology used by secondary users. Since the approach used by NTIA is published, an attacker can try to infer the location of the incumbent by combining that public information with some inferred knowledge.

When the incumbent is operating, a set of protection zones will be activated (i.e., no secondary user will be able to operate inside that area). When a secondary user requests access to the spectrum, the SAS will decide depending on the secondary user's location. If the location of the secondary user is within the operational area, the use of the incumbent's frequency will be denied, as it may cause interference to the incumbent. However, when it is outside the operational zone, the use of the incumbent's frequency will be granted.

As a first step to the SAS et ESC deployments, the exclusion zone drawn by NTIA will be divided into fixed protection zones. An attacker would hence try to infer the boundaries of the protection zone, then deduce the location of the incumbent.

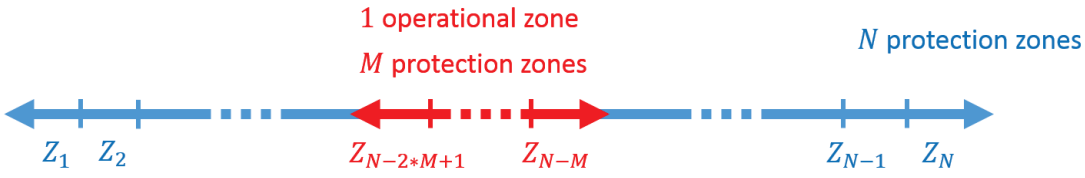


FIGURE 4.5: Division of the exclusion zone into protection zones

For the sake of simplicity and without loss of generality, we consider the exclusion zone to be a straight line for the purpose of this analysis. Let N be the number of protection zones and Z_n the n^{th} protection zone, where $1 \leq n \leq N$. When an incumbent is operating, M protection zones are activated. The set of M protection zones is refereed to as the operational zone. So, the operational zone is spread over M protection zones, where $M \geq 1$. Fig. 4.5 shows the division of the exclusion zone into protection zones and an example of the activation of the operational zone.

4.4.2 Attack Model

In this chapter, we adopt a different threat model compared to the one adopted in Chapter 3. We assume that the attacker already knows the operational frequency of the incumbent, either using the algorithm presented in Chapter 3 or a different approach. Such information will include that, within that area, the incumbent is on and an operational zone is activated. We also assume that the SAS allows a secondary to request a specific channel as expected in the 3.5 GHz band and mentioned in [86]. In order to find the location of the incumbent, the attacker will try to infer the boundaries of the operational zone. Fig. 4.6 shows a scenario of an attack. It will be carried out by two attackers: one attacker moving up the line and another attacker moving down the line. It is known that no secondary user is allowed to use the channel occupied by the incumbent within the operational zone. However, both attackers are querying the database about the incumbent's channel. Every time they are denied access, they keep moving. Once access is granted, the attackers know that they are, at that moment, out of the operational zone.

The attackers are relocating either up or down following a specific path of movement. We consider two paths of movement:

- Using a fixed step: The attack system deploys a static path of movement between one query and another, as described in Algorithm 2. Both attackers move with a fixed step until the boundary of the operational zone is crossed.

- Using an adaptive step: The attack system deploys an adaptive path of movement between one query and another by doubling the step if the boundary is not crossed and halving the step if the boundary is crossed (i.e., executing a binary search), as described in Algorithm 3. Intuitively, the "adaptive-step" inference algorithm allows the attacker to acquire more knowledge in fewer queries.

In Algorithm 2, both attackers are initially co-located and know the incumbent channel. When they query the SAS for the incumbent channel and are denied, they agree on a fixed value of the step and start moving away from the initial location. One attacker takes a step backward, and the other attacker takes a step forward. Next, they query the SAS again for the incumbent channel. Every time an attacker is denied, it keeps

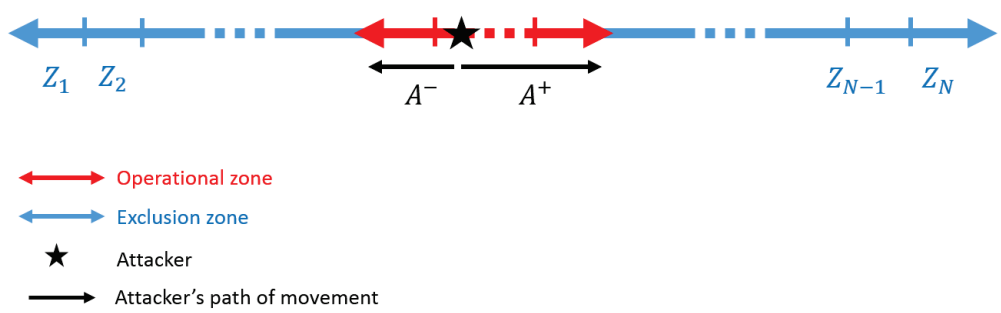


FIGURE 4.6: Initial location of the attacker and its path of movement

moving a step either backward or forward based on its approach. And once an attacker is granted access, it stops moving and concludes that the last location is the knowledge needed to infer the boundaries of the operational zone.

Input: Channel of the incumbent ch_I, Initial location loc_A of the two attackers A^- and A^+, Distance of movement of both attackers *step*;

Output: Boundaries of the operational zone $bound^-$ and $bound^+$;

Initialization: $loc_{A^-} = loc_A$, $loc_{A^+} = loc_A$;

for A *in* $[A^-, A^+]$ **do**
  while *access to* ch_I *is denied* **do**
    if $A = A^-$ **then**
      $loc_{A^-} = loc_{A^-} - step$;
    end
    if $A = A^+$ **then**
      $loc_{A^+} = loc_{A^+} + step$;
    end
    request access to ch_I;
  end
end
$bound^- = loc_{A^-} + step$;
$bound^+ = loc_{A^+} - step$;

Algorithm 2: Algorithm of an inference attack - Fixed step

In Algorithm 3, both attackers use a smarter inference algorithm by executing two processes: learning and refining. The "learning" process consists of acquiring as much as possible of knowledge. The "refining" process consists of improving the accuracy of the acquired knowledge. First, the attackers start at the same location, use the same initial step value and begin the learning process. One attacker takes a step backward and requests access to the incumbent channel. If it is denied access, it doubles its step value,

Input: Channel of the incumbent ch_I, Initial location loc_A of the two attackers A^- and A^+, Initial distance of movement of both attackers *step*;

Output: Boundaries of the operational zone $bound^-$ and $bound^+$;

Initialization: $loc_{A^-} = loc_A$, $loc_{A^+} = loc_A$, $step^- = step$, $step^+ = step$;

for A *in* $[A^-, A^+]$ **do**
  while *access to* ch_I *is denied* **do**
    if $A = A^-$ **then**
      if *first attempt* **then**
        $step^- = step$;
      end
      else
        $step^- = step^- \times 2$;
      end
      $loc_{A^-} = loc_{A^-} - step^-$;
    end
    if $A = A^+$ **then**
      if *first attempt* **then**
        $step^+ = step$;
      end
      else
        $step^+ = step^+ \times 2$;
      end
      $loc_{A^+} = loc_{A^+} + step^+$;
    end
    request access to ch_I;
  end
  while *access to* ch_I *is granted* **do**
    if $A = A^-$ **then**
      $step^- = step^-/2$;
      $loc_{A^-} = loc_{A^-} + step^-$;
    end
    if $A = A^+$ **then**
      $step^+ = step^+/2$;
      $loc_{A^+} = loc_{A^+} - step^+$;
    end
    request access to ch_I;
  end
end
$bound^- = loc_{A^-} - step^-$;
$bound^+ = loc_{A^+} + step^+$;

Algorithm 3: Algorithm of an inference attack - Adaptive step

moves backward again and requests access. The other attacker takes a step forward and requests access to the incumbent channel. If it is denied access, it doubles its step value, moves forward again and requests access. Both attackers keep doubling their step values and moving until the access is granted. Once the access is granted, the attackers know then that they are outside the operational zone, and thus the boundary is crossed. So, each one of them starts the refining process by splitting the last value of the step in two, moving in the opposite direction and requesting access to the incumbent channel again. They keep doing so until the access is denied, i.e., the boundary is crossed again. An attacker can keep "learning" and "refining" until its knowledge is no longer updated. However, the goal of our analysis is to evaluate the vulnerability of the incumbent, so we limit our implementation to the first iteration of learning/refining.

4.5 Proposed Techniques and Metrics

In this section, we propose two privacy-preserving techniques: obfuscation and trustworthiness. While obfuscation is applied to the system itself, trustworthiness is applied to the secondary users in the system.

4.5.1 Obfuscation

Obfuscation is widely used in literature in order to introduce uncertainty into the attacker's knowledge [78][77][85][93][23]. It is applied on the database itself. In this case, we use perturbation to obfuscate the boundaries of the operational zone as described in Fig. 4.7 and Algorithm 4. In other words, the boundaries of the operational zone are extended to mislead the adversary and tamper with its knowledge.

Therefore, the boundaries of the operational zone will be replaced in the database by the corrupted (i.e., obfuscated) boundaries. So, if a secondary user is within the obfuscated area and requests access to the incumbent's frequency, it will be denied access, even though no possible interference is envisioned.

The advantage of such an algorithm is that it does not depend on the secondary user. However, one major inconvenience is that it will impact not only malicious secondary users but also innocuous secondary users by denying them access outside of the

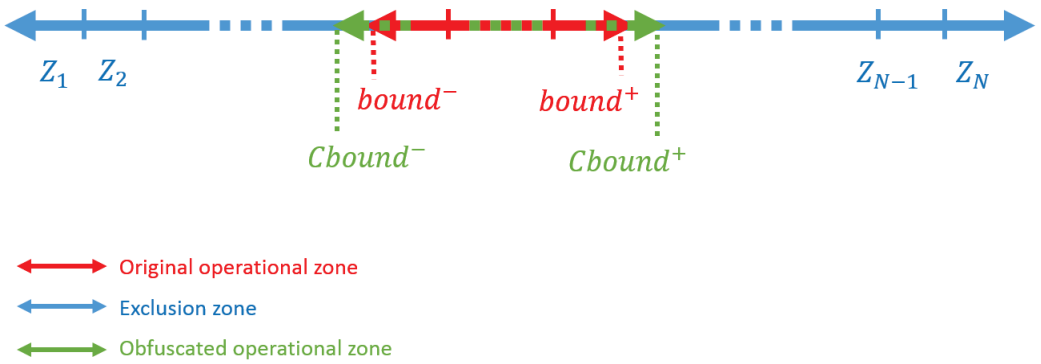


FIGURE 4.7: Adding perturbation to the operational zone

Input: Predefined noise level $noise$, Boundaries of the operational zone $bound^-$ and $bound^+$;

Output: Corrupted boundaries of the operational zone $Cbound^-$ and $Cbound^+$;

Initialization: $Cbound^- = bound^- - \frac{noise}{2}$, $Cbound^+ = bound^+ + \frac{noise}{2}$;

while S *requests access to* ch_I **do**
  if $Cbound^- \leq loc_S \leq Cbound^+$ **then**
    deny request access to ch_I;
  end
  else
    grant request access to ch_I;
  end
end

Algorithm 4: Perturbation algorithm

operational zone, which will affect spectrum efficiency. In other words, some available spectrum resources are being withheld by the SAS and hence unable to be used by the secondary network.

4.5.2 Trustworthiness

It is true that obfuscation helps mitigating an attack, but it does not help identifying an attacker. To the best of our knowledge, no work has been done to recognize and stop an inference attack in spectrum sharing. Therefore, in order to fulfill a threat-free spectrum sharing, the SAS needs also to assess any act of maliciousness.

A trust-based mechanism can be implemented and periodically updated in order to evaluate the trustworthiness of secondary users. In particular, when a secondary user i, with a trust value T_i, sends a query Q_i, the SAS sends a response $R_i = f(Q_i, T_i, L_{ch})$, where f is a decision making function of the SAS.

The query Q_i is a channel request that can be defined as follows:

$$Q_i = \{i, ch_q\} \tag{4.1}$$

where i is the identifier of the secondary user and ch_q is the channel requested. The secondary user i can request any channel available for use ($ch_q = \emptyset$) or a specific channel ζ available for use ($ch_q = \zeta$).

The response R_i is a channel assignment and can be formulated as follows:

$$R_i = \{i, r_i, ch_r\} \tag{4.2}$$

where r_i is a binary response of access to the spectrum and ch_r is the channel assigned if the access is granted ($r_i = 1$).

The SAS refers back to the list of available channels L_{ch} and decides whether user i can be granted access or not. Then,

$$r_i = \begin{cases} \{T_i \geq thr_T\} \otimes \{\zeta \in L_{ch}\} & \text{if } ch_q = \zeta \\ \{T_i \geq thr_T\} \otimes \{L_{ch} \neq \emptyset\} & \text{if } ch_q = \emptyset \end{cases} \tag{4.3}$$

where T_i is the trust value of the secondary user i, thr_T is the trust threshold predetermined by the SAS, ch_q is the channel requested (either ζ or any channel available), L_{ch} is the list of channels available for use. $\otimes$ is the binary operator (associative with neutral element 1 and absorbing element 0) used to calculate the binary response.

Different scenarios of response are envisioned depending on the request, the trust value and the availability of the spectrum. This consists of four cases:

- If the user i requests any channel available and his trust value T_i is above the trust threshold thr_T, the SAS grants access based on the channel assignment scheme employed.
- If the user i requests any channel available but his trust value T_i is below the trust threshold thr_T, the SAS denies access by returning that all channels are occupied at the moment.
- If the user i requests a specific channel to use, his trust value T_i is above the trust threshold thr_T, the SAS grants access to the channel requested if available or another channel if not.
- If the user i requests a specific channel to use, his trust value T_i is below the trust threshold thr_T, the SAS denies access by returning that all channels are occupied at the moment.

In this approach, a trust metric is assigned to every secondary user registered with the SAS. This applies a condition on the release mechanism of the information, i.e., it includes a condition on the response to every request. The system keeps a database of all secondary users and evaluates their trustworthiness by updating the trust metric with every request. Generally, a third party establishes trust using either direct observations by evaluating the conduct of users or indirect observations by relying on the recommendations of other users. In our case, secondaries have a direct relationship with the SAS and no relationships with other secondaries. Therefore, we introduce the notation: {Trustor, Trustee, Trust} as shown in Fig. 4.8. The trustor (i.e., SAS) evaluates a trustee (i.e., secondary user) when the latter sends a query by updating its trust value based on its distance to the boundaries of the operational zone.

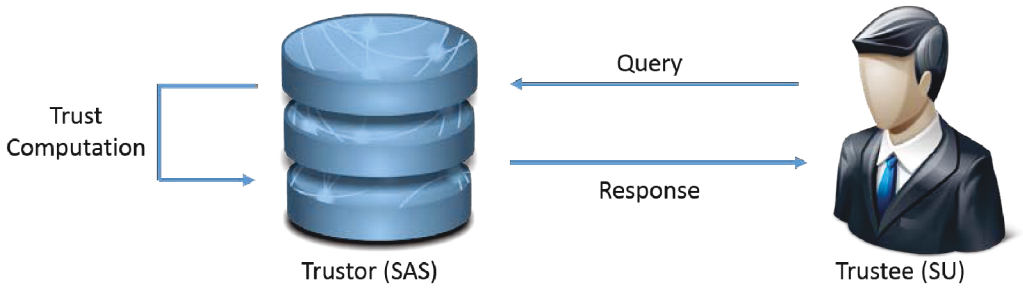


FIGURE 4.8: Trustworthiness computation and update in the 3.5 GHz

Algorithm 5 describes how such an approach will be deployed. Two distance thresholds (thr_{D^-} and thr_{D^+}) are included to allow certain flexibility. The SAS can thus adjust the values of those thresholds in order to adapt to the system needs and the incumbent operations. If the risk of an inference is high, the SAS increases the values of the distance thresholds. If the risk of an inference is low, the SAS decreases the values of the distance thresholds. Fig. 4.9 illustrates an example of the distance thresholds. When a secondary user is requesting access to the incumbent's channel and is within thr_{D^+} of one of the operational zone boundaries, it is considered a threat to the privacy of the incumbent. Hence, its trust value is lowered by τ. But, when it is requesting access to the incumbent's channel and is within thr_{D^-} of one of the operational zone boundaries, the threat is more serious. So, the trust value is lowered by 2τ.

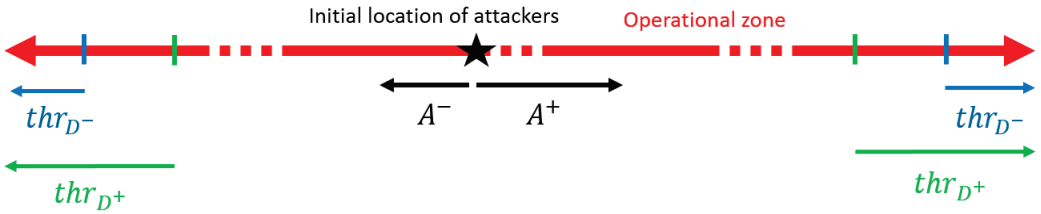


FIGURE 4.9: Distance thresholds illustration for the trustworthiness algorithm

This algorithm can be summarized as follow:

$$T_i(q) = \begin{cases} T_i(q - \Delta q) & \text{if } D > thr_{D^+} \\ T_i(q - \Delta q) - \tau & \text{if } thr_{D^-} < D \leq thr_{D^+} \\ T_i(q - \Delta q) - 2\tau & \text{if } D \leq thr_{D^-} \end{cases} \tag{4.4}$$

where q refers to the query number, Δq refers to the trust update interval, τ is update factor, D is the distance from the attacker location to the boundary of the operational zone, and thr_{D^-} and thr_{D^+} are the distance thresholds as defined above.

Input: Channel of the incumbent ch_I, Predefined thresholds for distance thr_{D^-} and thr_{D^+} and for trust thr_T, Boundaries of the operational zone $bound^-$ and $bound^+$, Location of the secondary user loc_S, Initial trust level of the secondary user T_S;

Output: Updated trust level of the secondary user T_S, Termination of the connection *terminate*;

Initialization: $terminate =$ **False**, $T_S = T_0$;

while S *requests access to* ch_I ***and*** $T_S \geq thr_T$ **do**
  if $dist(loc_S, bound^-) < thr_{D^+}$ ***or*** $dist(loc_S, bound^+) < thr_{D^+}$ **then**
    $T_S = T_S - \tau$;
  end
  else
    if $dist(loc_S, bound^-) < thr_{D^-}$ ***or*** $dist(loc_S, bound^+) < thr_{D^-}$ **then**
      $T_S = T_S - 2\tau$;
    end
  end
end
if $T_S < thr_T$ **then**
  terminate = **True**;
end

Algorithm 5: Trustworthiness algorithm

The trust evaluation depends on the secondary user's request and allows the SAS to evaluate the maliciousness of a secondary user. A secondary user is considered untrustworthy and can be terminated if its trust crosses the trust threshold. We assume that, in this case, the trust can only decrease to avoid malicious use of the metric. If the trust value can be improved and an adversary is aware of that, it will try to keep its trust value below the predefined threshold.

In the remainder of our analysis, we consider the trust as a measure of percentage and use $T_0 = 100$ % and $\tau = 10$ %.

4.5.3 Metrics

We need to quantify both the attack and the countermeasure in order to evaluate the privacy of the incumbent and utilization efficiency of the spectrum. Similarly to

Chapter 3, we quantify the uncertainty of an attacker about the inferred knowledge, the effort invested in the attack, and the spectrum resources lost after applying a privacy-preserving technique.

Distance

The distance of an inferred location to the original location is measured in terms of uncertainty, i.e., how uncertain an attacker is about the boundaries of the operational zone.

We define the uncertainty U of an attacker about the inferred zone as

$$U = d(Ibound^{-} - bound^{-}) + d(Ibound^{+} - bound^{+}) \tag{4.5}$$

where d is the geometric distance, $Ibound^{-}$ and $Ibound^{+}$ are the inferred lower and upper bounds respectively, and $bound^{-}$ and $bound^{+}$ are the original lower and upper bounds respectively.

As explained in the algorithms above, we assume that the attacker always considers the last location as the inferred boundary. Ultimately, using this approach, we have

$$L_{OpZ} \geq L_{inf} \tag{4.6}$$

and Eq. 4.5 becomes

$$U = L_{OpZ} - L_{inf} \tag{4.7}$$

where L_{OpZ} is the actual length of the original operational zone and L_{inf} is the length of the inferred operational zone.

We also define the normalized uncertainty NU of an attacker as

$$NU = \frac{U}{L_{OpZ}} \tag{4.8}$$

and thus, Eq. 4.8 becomes

$$NU = 1 - \frac{L_{inf}}{L_{OpZ}} \tag{4.9}$$

The normalized metric is more generic since it does not depend on the operational zone and allows to compare different attacks and different countermeasures.

Cost

The cost of an attack is measured in terms of the number of queries, i.e., how many queries it takes for an attacker to acquire certain knowledge.

The number of queries is the sum of the number of queries of both attackers as we consider them a single system. So, it is defined as

$$Q = Q^- + Q^+ \tag{4.10}$$

where Q^- is the number of queries by the first attacker and Q^+ is the number of queries by the second attacker.

We can also define the contribution of each attacker to the inference cost when dividing the number of queries from each attacker by the total number of queries, as follows. The contribution of the first attacker becomes

$$NQ^- = \frac{Q^-}{Q} = \frac{Q^-}{Q^- + Q^+} \tag{4.11}$$

and the contribution of the second attacker becomes

$$NQ^+ = \frac{Q^+}{Q} = \frac{Q^+}{Q^- + Q^+} \tag{4.12}$$

Spectrum Loss

Some countermeasures that are introduced to protect the privacy of the incumbent, are accompanied by a loss in spectrum opportunities for non-malicious secondary users. Such loss results in low spectrum efficiency.

In this case, we define spectrum loss as the loss in space:

$$SL = d(Abound^- - bound^-) + d(Abound^+ - bound^+) \tag{4.13}$$

where d is the geometric distance, $Abound^-$ and $Abound^+$ are the altered lower and upper bounds respectively, and $bound^-$ and $bound^+$ are the original lower and upper bounds respectively.

Since obfuscation enlarges the size of the operational zone, we have

$$L_{alt} \geq L_{OpZ} \tag{4.14}$$

and using our approach, Eq. 4.13 becomes

$$SL = L_{alt} - L_{OpZ} \tag{4.15}$$

where L_{OpZ} is the length of the original operational zone and L_{alt} is the length of the altered operational zone.

The normalized spectrum loss will be defined as

$$NSL = \frac{SL}{L_{alt}} \tag{4.16}$$

Then, Eq. 4.16 becomes

$$NSL = 1 - \frac{L_{OpZ}}{L_{alt}} \tag{4.17}$$

4.6 Evaluation Study

This sections studies the analytical model and the simulation model of the inference attacks and evaluates the incumbent privacy under different assumptions.

4.6.1 Analytical Model

In this section, we try to mathematically model the inference algorithm of the attacker by quantifying the uncertainty and the number of queries.

Fixed-Step Inference Algorithm

During the fixed-step inference process, both attackers move with a static path of movement as explained in Algorithm 2. The uncertainty $U = U^- + U^+$ and the cost

$Q = Q^- + Q^+$ depend on the initial location d_0, where U^- and Q^- are the uncertainty and the cost of A^- respectively, and U^+ and Q^+ are the uncertainty and the cost of A^+ respectively. Let $d = L_{OpZ}$ be the length of the operational zone, d_0 the initial location of the attackers A^- and A^+, and *step* the value of the step of the inference algorithm. We assume that d and *step* are integers.

The fixed-step inference algorithm is explained in 4.4.2 and illustrated in Fig. 4.10. Note that, for next equations, we use **div** as the division operator which discards the remainder of the division and results in a value of integral type.

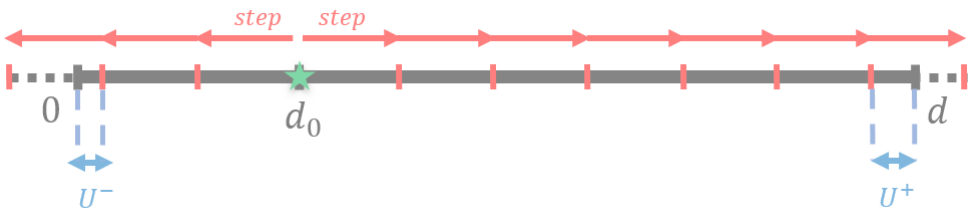


FIGURE 4.10: Illustration of the fixed-step algorithm

The uncertainty of an attacker can be calculated as follows. The uncertainty of A^- about the lower boundary is

$$U^- = d_0 - step \times (d_0 \ \mathbf{div} \ step) \tag{4.18}$$

and the uncertainty of A^+ about the upper boundary is

$$U^+ = (d - d_0) - step \times ((d - d_0) \ \mathbf{div} \ step) \tag{4.19}$$

Therefore, using Equations 4.18 and 4.19, the total uncertainty of the attack system about the boundaries for the operational zone becomes

$$U = d - step \times (d_0 \ \mathbf{div} \ step \ + \ (d - d_0) \ \mathbf{div} \ step) \tag{4.20}$$

The number of queries needed to infer the boundaries of the operational zone can be calculated as follows. The number of queries needed by A^- to cross the lower boundary is

$$Q^- = d_0 \ \mathbf{div} \ step \ + \ 1 \tag{4.21}$$

and the number of queries needed by A^+ to cross the upper boundary is

$$Q^+ = (d - d_0) \textbf{ div } step \ + \ 1 \tag{4.22}$$

Therefore, using Equations 4.21 and 4.22, the total number of queries needed by the attack system to cross the boundaries of the operational zone becomes

$$Q = d_0 \textbf{ div } step \ + \ (d - d_0) \textbf{ div } step \ + \ 2 \tag{4.23}$$

Adaptive-Step Inference Algorithm

The adaptive-step inference process is divided into phases as explained in 4.4.2: the attacker acquires a maximum of information in the learning phase by moving further from the initial location, then improves its knowledge in the refining phase by moving back to the initial location.

We adopt the same notation used for the fixed-step inference algorithm: d is the length of the operational zone, d_0 is the initial location of the attackers A^- and A^+, *step* is the value of the step of the inference algorithm. Note that, for next equations, we use $\lfloor \cdot \rfloor$ as the function that maps a real number to the greatest preceding integer.

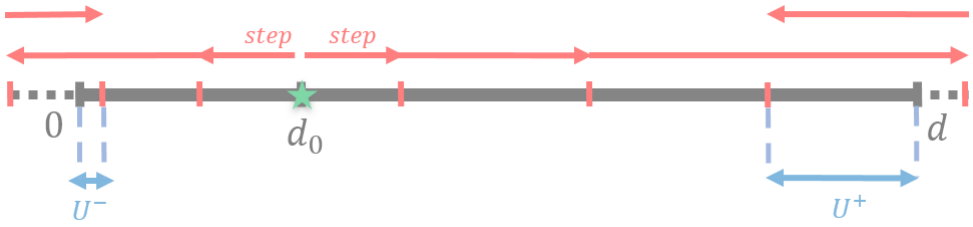


FIGURE 4.11: Illustration of the adaptive-step algorithm

The number of queries needed by attacker A^- in the learning phase (i.e., to cross the lower boundary) is

$$Q_l^- = \lfloor \ log_2(d_0 \textbf{ div } step \ + \ 1) \ \rfloor \ + \ 1 \tag{4.24}$$

and the number of queries needed by attacker A^+ in the learning phase (i.e., to cross the upper boundary) is

$$Q_l^+ = \lfloor\ log_2((d - d_0)\ \textbf{div}\ step\ +\ 1)\ \rfloor\ +\ 1 \tag{4.25}$$

Therefore, using Equations 4.24 and 4.25, the total number of queries needed by the attack system in the learning phase (i.e., to cross the boundaries of the operational zone) becomes

$$Q_l = \lfloor\ log_2(d_0\ \textbf{div}\ step\ +\ 1)\ \rfloor\ +\ \lfloor\ log_2((d - d_0)\ \textbf{div}\ step\ +\ 1)\ \rfloor\ +\ 2 \tag{4.26}$$

The mathematical model here does not provide an exhaustive analysis of the incumbent privacy. If the attack system is using an adaptive-step algorithm, the number of queries and the uncertainty for the learning phase and the refining phase depend on the initial location of the attack system which we do not model mathematically. So, in next section, we implement and run Monte Carlo simulations to study the vulnerability of the incumbent and show the impact of the proposed techniques.

4.6.2 Simulation Model

Our simulations include one SAS, one operational incumbent, and one attack system (two collaborating attackers). Based on the location of the secondary user, the SAS decides whether to grant or deny access. We assume that the attackers know the NTIA's exclusion zone and are moving along its boundaries. We also assume that the attackers are within the operational zone and know the operational channel of the incumbent, but no further knowledge is given. Both attackers query the channel occupied by the incumbent while moving in order to infer the boundaries of the operational zone.

In this section, we evaluate the privacy of the incumbent by analyzing the protection of an operational zone from inferences in three cases:

- a baseline system where no privacy-preserving technique is deployed;
- a system where obfuscation is added to the operational zone;
- a system where trustworthiness is defined for each secondary user.

The last two cases present the results when two privacy-preserving techniques are implemented, while the first case presents results about how vulnerable the privacy is to inference attacks, which will be used as a baseline for our results. For each case, we deploy the inference algorithms described by Algorithm 2 and Algorithm 3 and simulate them in Python using a Monte Carlo approach. We run and average 150 attack trials with random initial locations using the different parameters summarized in Table 4.1.

TABLE 4.1: Simulation specification

Parameter	**Definition**
L_{OpZ}	Length of the operational zone
$step$	Step of movement
$noise$	Noise added to the operational zone
thr_T	Trust threshold
$T_0 = 100\%$	Initial trust threshold
$\tau = 10\%$	Factor of trust decrease
$thr_D+ = 0.2L_{OpZ}$	Distance threshold of tolerance
$thr_D- = 0.05L_{OpZ}$	Distance threshold of intolerance

No Privacy-Preserving Techniques Implemented

When no privacy-preserving techniques are implemented, the system is obviously more vulnerable to inference attacks. Figures 4.12 and 4.13 show how the attack system gains knowledge with each additional query. By the end of the inference attack, the attack system is able to get as close as the *step* value allows it. We consider 7 values of steps for comparison purposes in order to evaluate how the step value influences the inference process and the privacy of the incumbent. Let *step* be 0.01, 0.025, 0.05, 0.075, 0.1, 0.2, and 0.3 of the length of the operational zone. In Fig. 4.12, the attackers move with a fixed distance equal to *step*. In Fig. 4.13, the attackers update their value of *step* by executing more or less a binary search.

We note that, for the fixed-step algorithm, the uncertainty of the attacker is declining slowly but steadily. The attackers using this algorithm advances to the boundaries of the operational zone with a fixed pace. Therefore, their knowledge is statically updated with each query. For all values of *step*, the same trend is observed as the attack results matches the analytical model in 4.6.1.

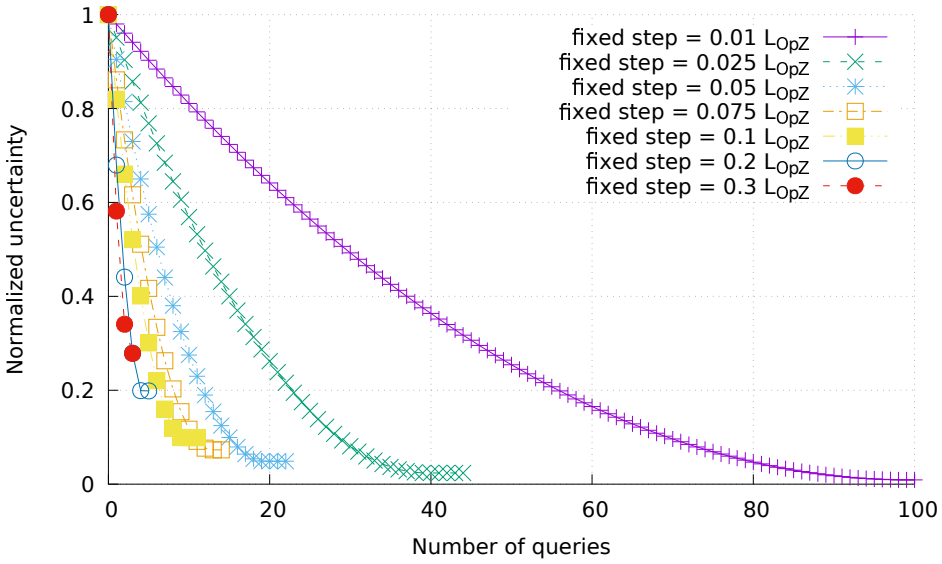


FIGURE 4.12: Impact of the baseline system on privacy for a "fixed-step" attack

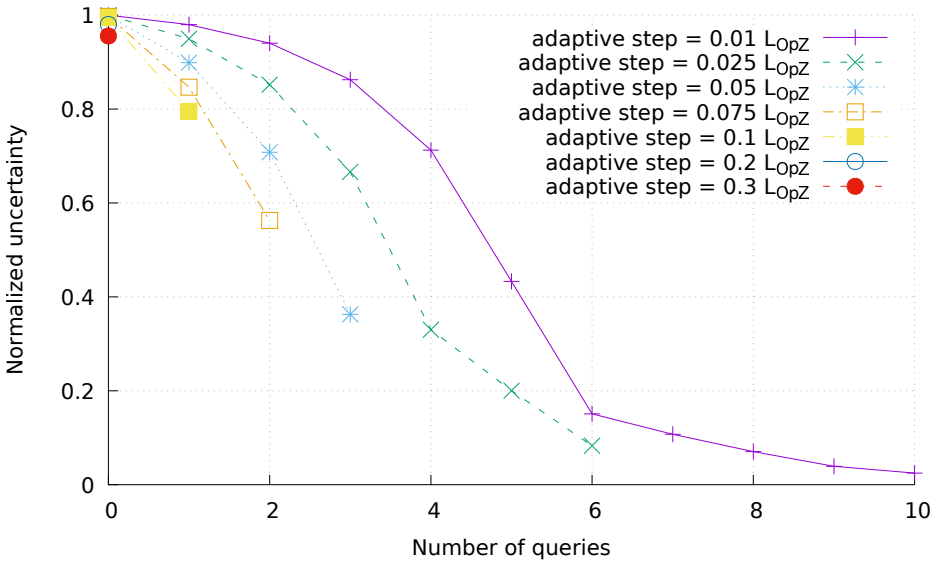


FIGURE 4.13: Impact of the baseline system on privacy for an "adaptive-step" attack

We also note that, for the adaptive-step algorithm, the uncertainty of the attacker dramatically decreases, hits an inflection point, and gradually decreases again. That inflection point represents the transition point from the "learning" phase to the "refining" phase as explained in Algorithm 3. Thanks to this two-phase process, the adaptive-step algorithm gives the best cost for the attacker since it cuts down the number of queries needed in the fixed-step algorithm by almost 80 %. However, for higher values of *step*, the accuracy of an adversary using the fixed-step algorithm is higher than that of an adversary using the adaptive-step algorithm.

The trade-off between cost and accuracy is important when both the adversary and the SAS aim to optimize their algorithms. At the end of the inference process, for small values of *step*, the uncertainty of the fixed-step algorithm is almost zero. However, the cost of that approach (i.e., the number of queries) is high. In the case of the adaptive-step algorithm, the attack system acquires more knowledge in a fewer number of queries thereby offering better attack performance overall.

For next simulations, we implement two privacy-preserving techniques for two *step* values: $0.01 \times L_{OpZ}$ and $0.05 \times L_{OpZ}$. The first technique is obfuscation, applied to the operational zone. The second technique is trustworthiness, applied to secondary users.

Implementation of Obfuscation (Perturbation)

By implementing perturbation, the SAS can enlarge the operational zone. We plot the normalized uncertainty of an attacker vs. the number of queries of an attacker for different values of the noise in Figures 4.14, 4.15, 4.16, and 4.17. An evaluation of this privacy-preserving technique requires also the evaluation of its repercussions as well. Therefore, for each simulation, we calculate the spectrum loss associated with each privacy gain. The spectrum loss generated by the value of *noise* is illustrated in each figure by the horizontal solid lines.

We consider 5 values of *noise*: adding 0.01, 0.05, 0.1, 0.2, 0.3 of area to the operational zone, i.e., adding $\frac{noise}{2}$ to the first boundary of the operational zone and $\frac{noise}{2}$ to the second boundary of the operational zone as explained in 4.5.1. Hence, the obfuscated operational zone is enlarged by the value of *noise*. Compared to the results of the

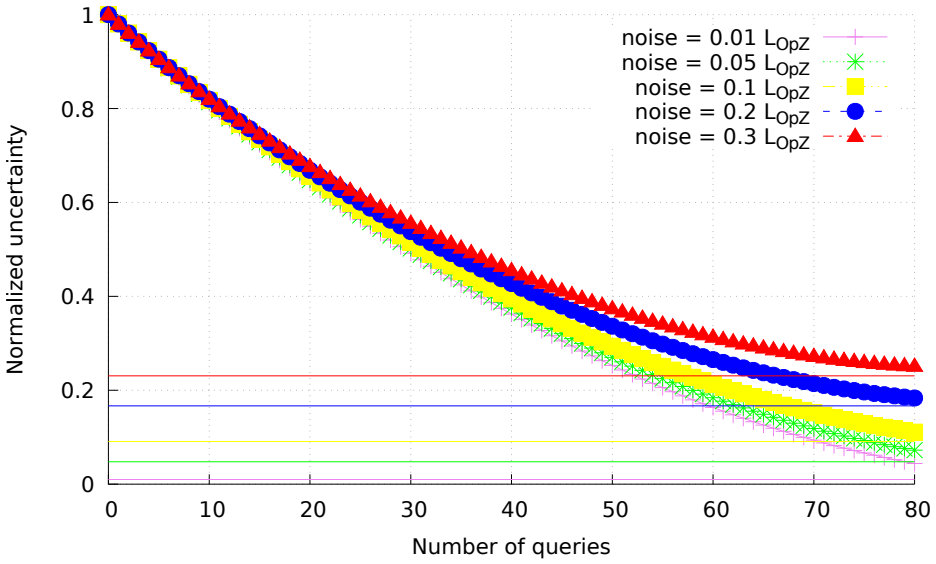


FIGURE 4.14: Impact of the obfuscation (perturbation) algorithm on privacy for a "fixed-step" attack ($step = 0.01\ L_{OpZ}$)

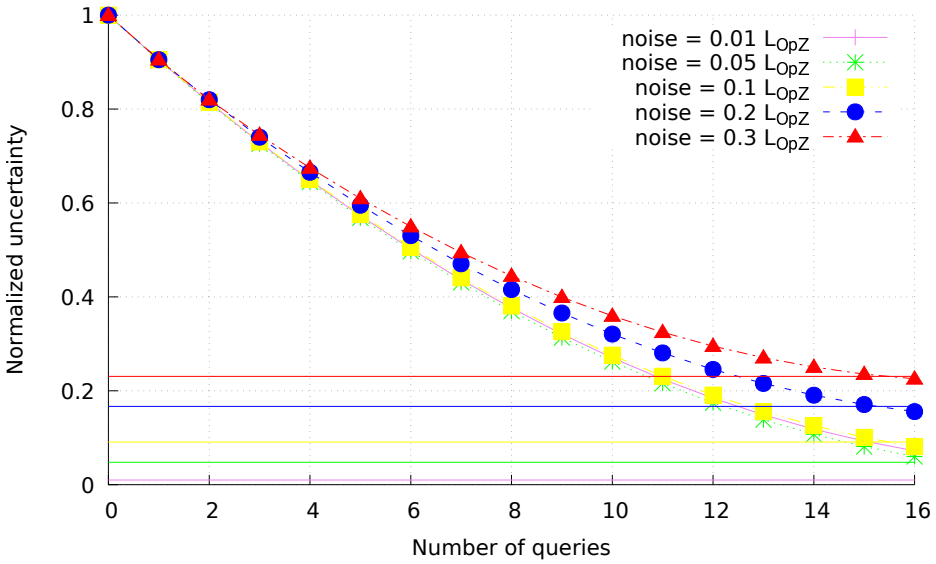


FIGURE 4.15: Impact of the obfuscation (perturbation) algorithm on privacy for a "fixed-step" attack ($step = 0.05\ L_{OpZ}$)

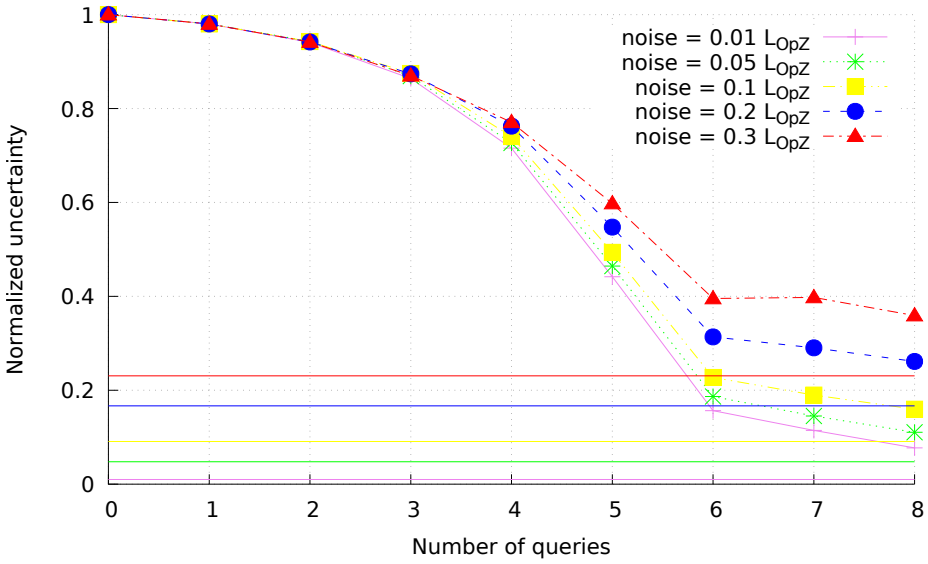


FIGURE 4.16: Impact of the obfuscation (perturbation) algorithm on privacy for an "adaptive-step" attack ($step = 0.01\ L_{OpZ}$)

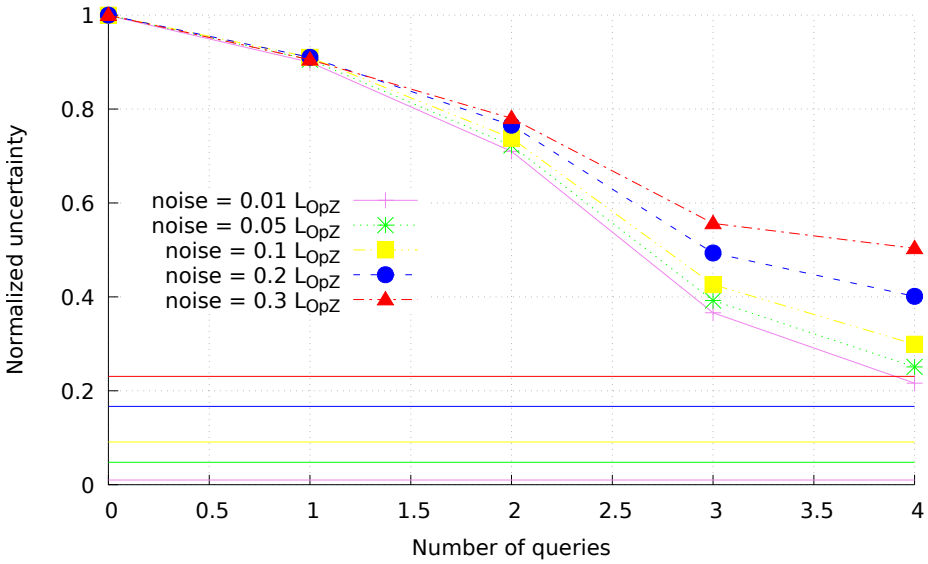


FIGURE 4.17: Impact of the obfuscation (perturbation) algorithm on privacy for an "adaptive-step" attack ($step = 0.05\ L_{OpZ}$)

baseline system (Figures 4.12 and 4.13), we note an improvement of the privacy of the incumbent. The accuracy of the adversary is affected by the obfuscated boundaries: the greater the perturbation, the greater the privacy.

For the fixed-step algorithm (Figures 4.14 and 4.15), we note the same trend. The decrease in privacy is gradual. Nevertheless, the knowledge of the attacker hits a maximum. We also note that the increase in the value of the step can drastically decline the privacy by reducing the cost of the attack. In fact, the number of queries for $step = 0.05 \times L_{OpZ}$ is nearly five times less than the number of queries for $step = 0.01 \times L_{OpZ}$.

For the adaptive-step algorithm (Figures 4.16 and 4.17), we note a different trend. The decrease in privacy is dramatic at first, then hits a maximum and stabilizes. As explained above, this is caused by the two-phase inference process; the attacker is first learning then refining. Moreover, the increase in the value of the step slightly decreases the cost of the attack. Compared to the fixed-step algorithm, the cost of the attack system is nearly ten times less for $step = 0.01 \times L_{OpZ}$ and four times less for $step = 0.05 \times L_{OpZ}$. Overall, the adaptive-step algorithm shows a better performance for the attacker. However, we conclude that no matter how smart the attacker is, this privacy-preserving technique can ultimately stop it.

Figures 4.14, 4.15, 4.16, and 4.17 show the spectrum loss for each value of the noise (horizontal solid lines). The uncertainty of an adversary depends on the sharing loss as the latter is a lower bound for the former. Those plots prove that a basic trade-off between privacy protection and spectrum efficiency exists: the better the privacy, the poorer the efficiency. However, in order to optimize the usage of such a technique, the SAS can implement obfuscation when the spectrum load is low and the loss in resources would not affect the secondary users already operating. In this case, the existing spectrum resources are enough to accommodate all secondary requests and attackers are fed inaccurate information about the operational zone, thereby providing both spectrum efficiency and privacy protection.

Implementation of Trustworthiness

The implementation of trust requires defining thresholds for distance thr_{D^-} and thr_{D^+} and for trust thr_T. As explained in 4.5.2, a trust value of T means that the

secondary user is T % trustworthy. In Figures 4.18, 4.19, 4.20 and 4.21, we plot the uncertainty of an attack vs. the cost of an attack. We present a comparison between different trust levels and show how they affect the privacy for a fixed-step algorithm and an adaptive-step algorithm. In this case, we fix the values used for distance: $thr_{D^-} = 0.05 \times L_{OpZ}$ and $thr_{D^+} = 0.2 \times L_{OpZ}$. These values provide diverse results and allow better evaluation of the efficiency of this technique. Nonetheless, the SAS can implement other values of the distance thresholds as well. The values considered for thr_T are 0 %, 10 %, 30 %, 50 %, 80 %, and 100 %.

Overall, when the trust threshold value is higher, the number of queries needed to infer the same knowledge is higher. This is due to the fact that the attacker is bounded by the trust value. Depending on the trust threshold thr_T, after certain number of queries, the adversary's knowledge reaches a maximum and is no longer updated.

Some trust values are reasonable and can be implemented without using extreme caution ($thr_T = 100$ %) or allowing extreme freedom ($thr_T = 0$ %). When the trust threshold thr_T is equal to 100 %, every move from the secondary user is considered a threat. In other words, every secondary user must keep its trust value at its initial value $T_0 = 100\%$. The first suspicious action (i.e., moving in the direction of the operational zone boundary and requesting the incumbent channel) results in the decrease of its trust value and therefore in its termination. When the trust threshold thr_T is equal to 0 %, the system takes more time to recognize a malicious secondary user (i.e., it takes more secondary queries and hence more system updates for the trust value to reach 0 %). Neither approach is recommended since they provide the worst trade-off: one assumes that all secondary users are trustworthy, the other assumes that they are not.

We note that all algorithms show the same trend for two values of the step (0.01 L_{OpZ} and 0.05 L_{OpZ}). When the step value is small (0.01 L_{OpZ}), the attacker is shortly detected and stopped, so the uncertainty varies between 0.3 and 0.7. So, even for low trust thresholds, the trust algorithm keeps the adversary at relatively high uncertainty. When the step value is higher (0.05 L_{OpZ}), the detection process takes more time and the knowledge of the attacker can be significant if the trust threshold is not properly defined. For example, for $thr_T \leq 50$, the uncertainty of the attacker can be as low as 0.1. In order to remedy that, the SAS can adjust the distance thresholds.

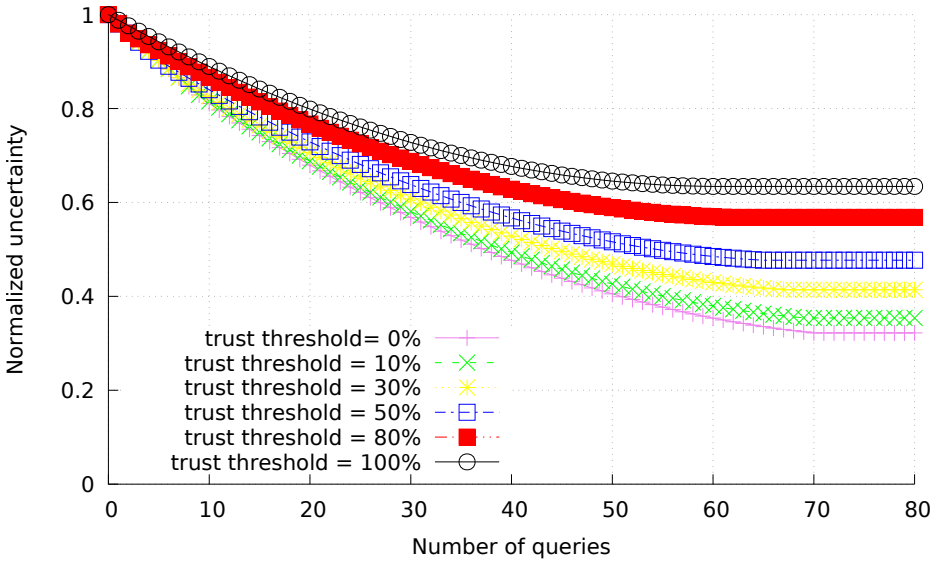


FIGURE 4.18: Impact of the trustworthiness algorithm on privacy for a "fixed-step" attack ($step = 0.01\ L_{OpZ}$)

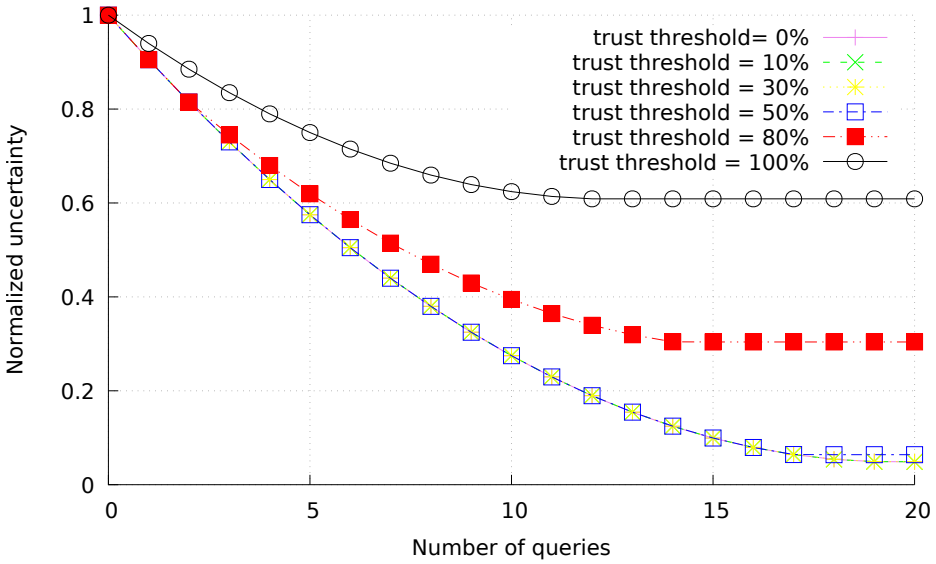


FIGURE 4.19: Impact of the trustworthiness algorithm on privacy for a "fixed-step" attack ($step = 0.05\ L_{OpZ}$)

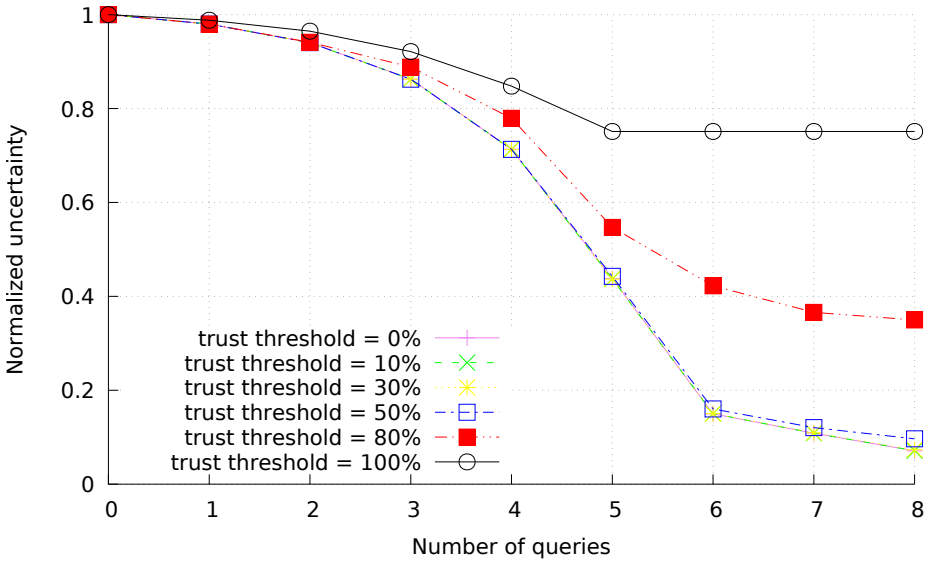


FIGURE 4.20: Impact of the trustworthiness algorithm on privacy for an "adaptive-step" attack ($step = 0.01\ L_{OpZ}$)

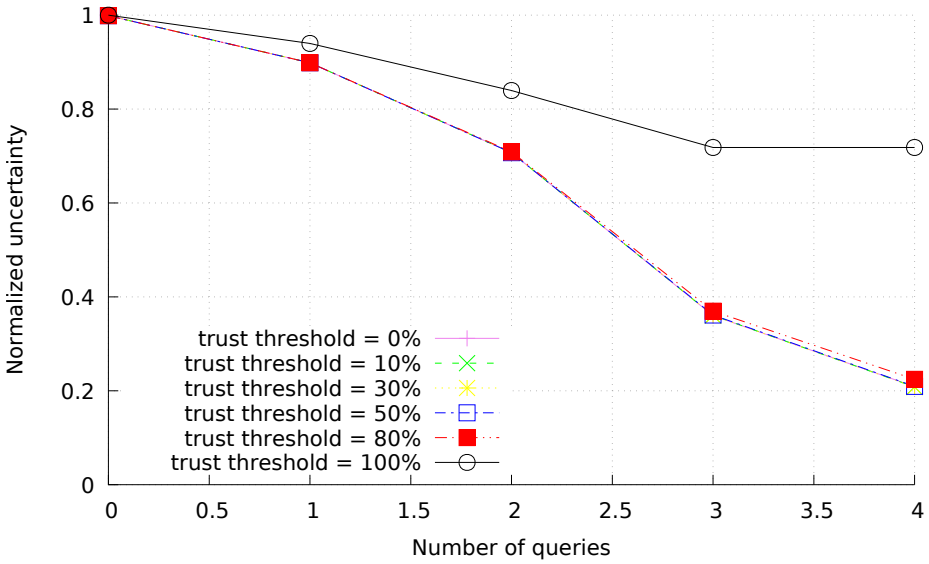


FIGURE 4.21: Impact of the trustworthiness algorithm on privacy for an "adaptive-step" attack ($step = 0.05\ L_{OpZ}$)

The fixed-step attack algorithm in Figures 4.18 and 4.19 present the same trend for the inference process. The privacy is decreasing slowly but steadily with each query. After a certain number of queries, the trust threshold is crossed and the attacker is terminated. Moreover, when $step = 0.01\ L_{OpZ}$, the plots of trust thresholds are close to each other, but distinctive for all values of thr_T. When $step = 0.05\ L_{OpZ}$, the results of the fixed-step algorithm are similar for $thr_T \leq 50$. In fact, for the latter case, the step value of the attack system is high enough to trigger the trust mechanism and the trust value quickly declines to reach the threshold.

The adaptive-step algorithm in Figures 4.20 and 4.21 shows a different trend. The privacy rapidly decreases and the attacker is still able to acquire additional knowledge before being terminated. However, when $step = 0.01\ L_{OpZ}$, the minimum uncertainty is 0.1 and when $step = 0.05\ L_{OpZ}$, the minimum uncertainty is 0.2. For a trust threshold of 80 %, the uncertainty of the attack is still high enough to mitigate and detect an attack, as it limits the knowledge of the attacker to less than 65 % when $step = 0.01\ L_{OpZ}$ and less than 80 % when $step = 0.05\ L_{OpZ}$.

In Fig. 4.22 and Fig. 4.23, we compare the performance of both attack algorithms in the presence of the trust metric. So, in order to choose the appropriate trust threshold, we plot the normalized uncertainty vs. the trust threshold. The SAS decides thus the minimum allowable uncertainty of an attack and matches it with the corresponding trust threshold. As expected, we see that higher trust threshold values allow better protection of the incumbent. The fixed-step algorithm presents the highest uncertainty while the adaptive-step algorithm presents the lowest uncertainty. If the adversary is using an adaptive-step algorithm, the uncertainty values seem static then quickly increase for the last few values of the trust threshold. If the adversary is using a fixed-step, the uncertainty values keep increasing with the increase in trust threshold values.

If an attacker is moving with a fixed step of $0.01\ L_{OpZ}$ and the SAS is trying to keep the attacker's uncertainty at 0.3, the trust threshold implemented can be as low as 0 %. For a fixed step of $0.05\ L_{OpZ}$, the trust threshold implemented should be greater than 75 %. This shows how the step value of the attacker can be crucial for the evaluation of the trust performance. In the case where the attacker is moving using an adaptive step, the trust threshold should be greater than 75 % for $step = 0.01\ L_{OpZ}$ and greater than

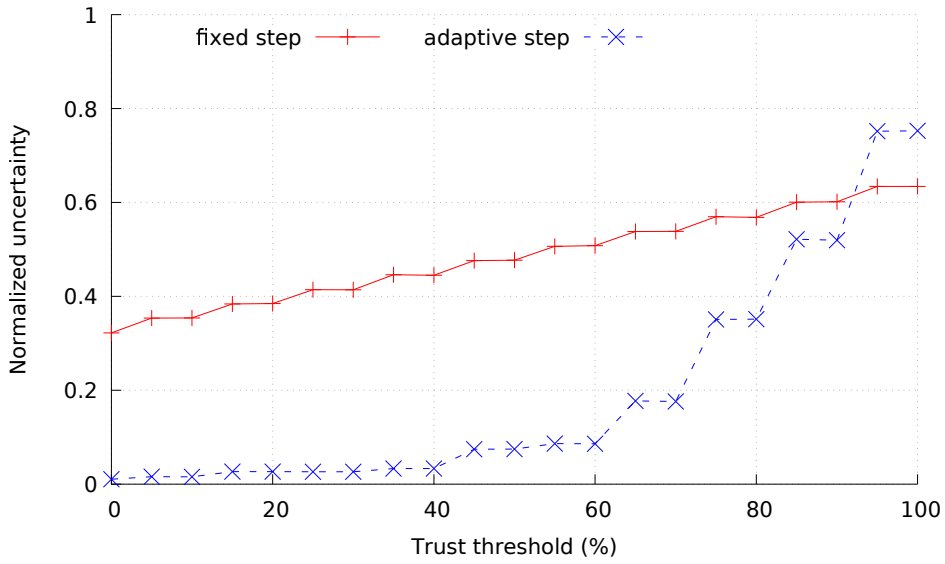


FIGURE 4.22: Comparison of the impact of the trustworthiness algorithm on privacy for all attack algorithms ($step = 0.01\ L_{OpZ}$)

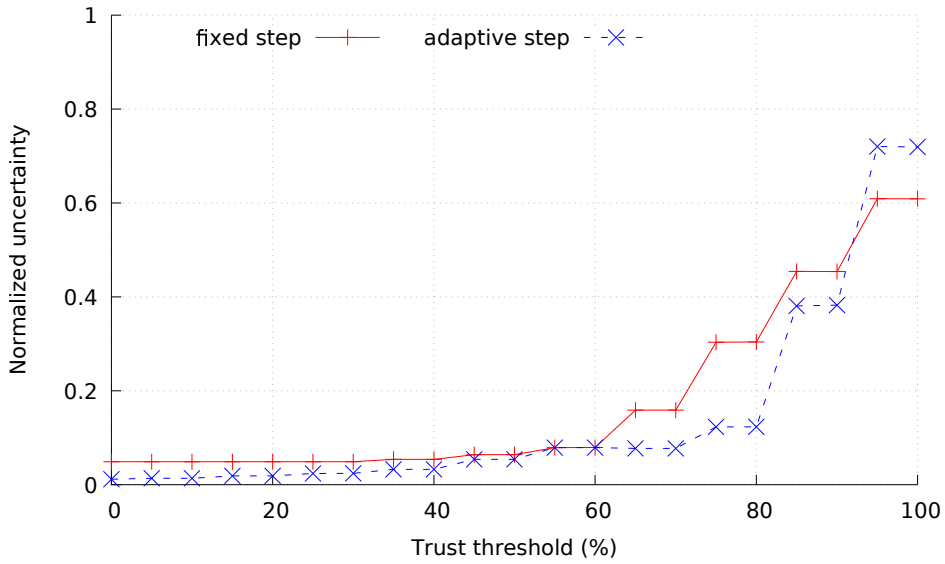


FIGURE 4.23: Comparison of the impact of the trustworthiness algorithm on privacy for all attack algorithms ($step = 0.05\ L_{OpZ}$)

85 % for $step = 0.05\ L_{OpZ}$.

We also note that the uncertainty of the attacker converges independently of the step. When the trust threshold thr_T is equal to 100 %, the uncertainty is equal to 0.65 for the fixed-step and equal to 0.75 for the adaptive-step algorithm. In fact, no matter how the attacker is advancing to the boundary of the operational zone, when the trust threshold is greater than 95 %, any move is considered a threat to the operational privacy and the querying secondary user is immediately terminated.

4.7 Conclusions

In this chapter, the vulnerability of the incumbent location to inference attacks is assessed. Initially, we show that the first step of inferring the incumbent's location is to infer the boundaries of the operational zone (one or more protection zones). Then, we evaluate the vulnerability of that zone to inference attacks under two attack scenarios (fixed-step algorithm and adaptive-step algorithm). The attacker using the fixed-step algorithm moves with a fixed pace. However, the attacker using the adaptive-step algorithm adapts its pace to each query's response. In order to mitigate the inference risk, two privacy-preserving techniques (obfuscation and trustworthiness) are proposed. While obfuscation is applied to the operational zone by enlarging its boundaries, trustworthiness is applied to secondary users by examining their honesty. Obfuscation presents solid results and prevent inference by increasing the uncertainty of the attacker. Nonetheless, it is accompanied with spectrum efficiency loss. The SAS can deploy obfuscation when it does not affect the available resources for use by existing secondary users in the shared environment. Trustworthiness includes more assumptions about different thresholds, e.g., distance to the boundaries of the operational zone. However, it allows the SAS to directly identify a threat and ultimately eliminate malicious secondary users. We conclude that the SAS can use both methodologies depending on the system state to achieve an optimal trade-off between privacy preservation and spectrum efficiency.

The work presented here opens the possibility to further studies and analyses, such as automatically adjusting the variables and the thresholds to the system requirements (privacy protection and spectrum availability) and the threat envisioned. Also, ongoing

discussions suggest unleashing more spectrum by opening more bands for sharing between Federal and commercial users. The 3500–3550 MHz and the 3700–4200 MHz are bands assigned for radiolocation and FSS space-to-Earth. Other bands are being considered as well. Those SAS-able bands require the protection of the operational security of incumbents. A more generic framework should be envisioned to allow scalability. This would be discussed in detail in the next chapter.

Chapter 5

Risk-Based Privacy-Preserving Framework

5.1 Introduction

Privacy-preserving algorithms have been proposed to protect sensitive incumbents in spectrum sharing against inferences. However, they are limited in scope and do not present a method to determine the appropriate type and extent of privacy protection. In fact, implementing privacy-preserving techniques can be efficient, yet insufficient to assess and manage an attack targeting the privacy of incumbents in spectrum sharing systems.

In this chapter, we propose a framework to identify privacy threats and monitor the shared environment in order to provide the best trade-off between privacy protection and spectrum efficiency. This trade-off can be tuned based on the parameters of the spectrum and the requirements of the sharing manager.

5.2 Framework Overview

Identifying risk and responding appropriately are key to protecting the incumbent against inference attacks. A five-step model borrowed from project management is presented (Fig. 5.1) that incorporates risk monitoring, identification, assessment, analysis and management to provide comprehensive protection of the incumbent [115]. General guidelines can be set to detect and mitigate a threat, thereby providing a safe shared spectrum for incumbents and guaranteed access for secondary users.

The spectrum manager will engage in risk monitoring as part of its normal operations. Generally, this will consist of looking for anomalous patterns in secondary spectrum requests. A suspicious action or threat event triggers the next step, identification of risk. Here, the likelihood of an inference attack and its impact are calculated. Factors such as the system load (overall secondary user activity) or local risk factors associated with a particular geographic region may be used in addition to the details of the specific threat event observed. The likelihood and impact are then used as inputs for the risk assessment step, which determines a system risk level. Calculation of the inference risk likelihood and impact and of the overall risk level are discussed in more detail in Section 5.3.2.

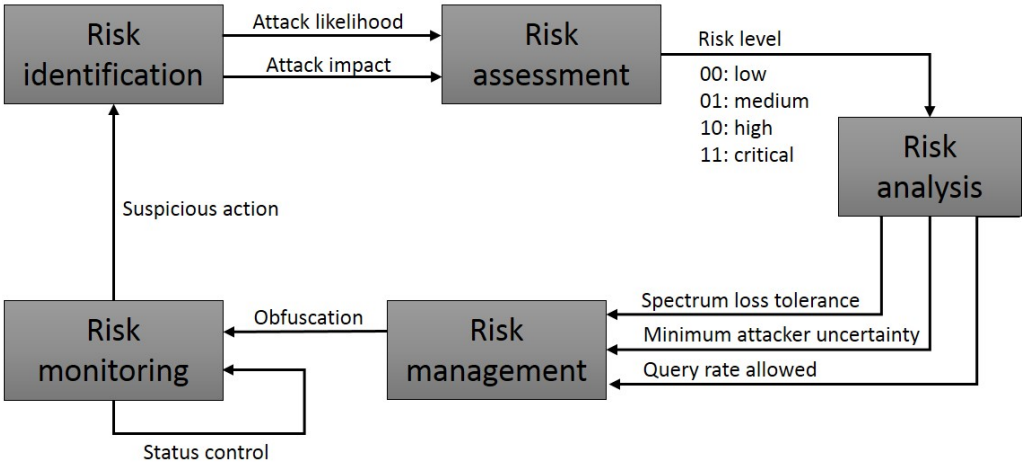


FIGURE 5.1: Risk-based privacy-preserving framework

The risk analysis step uses the risk level to select appropriate privacy measures, which we discuss in Section 5.3.5. In risk management, the desired privacy measures and spectrum availability are used in a constrained optimization to determine the trade-offs between privacy protection and spectrum utilization. The resulting privacy measures are applied, and the system manager enters the risk monitoring stage again.

5.3 Risk Mitigation Strategy

In this section, we present the different steps of the risk-based framework and introduce the handling options of a threat.

5.3.1 Risk Monitoring

The spectrum manager is regularly checking the system and tracking the sharing process by surveilling the activity of secondary users, the quality of channels, the functioning of the sensors, and the performance of other entities in the shared environment. Monitoring the privacy level of the incumbent can be easily incorporated in the regular proceedings of the spectrum manager and thus organized according to a specific and timely routine.

The surveillance of secondary users, their queries and their moves is one of the best approaches to identify a risk. The system manager is placed on alert while managing access to the spectrum, and if any privacy-threatening move is detected, the next step is triggered; risk identification and assessment.

Even a secondary query can be considered a suspicious action and should be controlled. The spectrum manager shall engage in special efforts to record secondary actions when they includes one of the following:

- request access to the operational channel of the incumbent,
- request access to multiple channels at the same time,
- request access to more resources in terms of channels or transmit power,
- cease the use of assigned spectrum resources,
- increase the query rate to access the spectrum,
- etc.

The spectrum manager can define a threshold of tolerance, i.e., the maximum number of times a suspicious action is tolerated by one secondary or a group of secondaries. When that threshold is crossed, the system reacts by entering the risk identification phase.

5.3.2 Risk Identification

After acknowledging a risk, we need to measure the likelihood of occurrence of an inference attack and its impact on the incumbent's operations. Computing the likelihood and the impact of a risk falls within the risk identification phase in Fig. 5.1.

The inference risk likelihood can be considered the probability that an inference attack is occurring. This probability can be defined using *Bayes' rule* [116]: (*i*) computing the initial probability of inference, (*ii*) analyzing attack history and identifying suspicious query rates, unbalanced spectrum load and other risk events, and (*iii*) updating the probability of inference. In other words, let h be the hypothesis that an inference attack is occurring and e be the observed event (i.e., an indicator of an inference attack), the updated inference probability becomes:

$$L(e) = P(h|e) = \frac{P(e|h)}{P(e)} P(h) \tag{5.1}$$

where $P(h|e)$ denotes the probability of inference h given an event e (*posterior probability*), $P(h)$ denotes the probability of an inference h before observing an event e (*prior probability*), $P(e)$ denotes the probability of observing an event e, and $P(e|h)$ denotes the probability of observing an event e given a hypothesis h.

In order to update the likelihood of an inference attack, we adopt an iterative Bayesian analysis by using the posterior probability as a prior probability after observing new evidence of an inference attack. An output of the risk monitoring phase triggers the re-evaluation of the risk in the risk identification phase by updating the probability of inference and its impact.

Since the operational location, frequency and time of the incumbent are the parameters that usually need privacy protection, the impact function I should be a function of those three factors. Thus, the overall impact I can be written as the vector of those factors.

$$I(e) = \begin{bmatrix} I_l(e) & I_f(e) & I_t(e) \end{bmatrix}^T \tag{5.2}$$

where $I_l(e)$ (respectively, $I_f(e)$ and $I_t(e)$) is the impact of an event e on location (respectively, frequency and time).

To simplify the impact vector, we only consider binary values. In other words, the impact of an inference attack evaluates the impact on location, frequency and time as a binary measure: 1 when there is an impact and 0 otherwise. For example, if an event e_1 has an impact only on location and frequency, $I(e_1)$ becomes $[1\ 1\ 0]^T$. If another event e_2 has an impact only on frequency, $I(e_2)$ becomes $[0\ 1\ 0]^T$.

5.3.3 Risk Assessment

Once the likelihood and impact functions are evaluated, the system enters the risk assessment phase by quantifying the risk level. This quantification will help the system determine a better understanding of the risk and its consequences on the privacy of the incumbent.

The inference risk is a metric that has been proposed in [117] to measure the likelihood of an inference attack and its impact on operational privacy after observing a suspicious event. It is expressed as follows:

$$IR(e) = L(e) \cdot I(e) \tag{5.3}$$

where $IR(e)$ is the inference risk, $L(e)$ is the likelihood of an inference attack, and $I(e)$ is the impact of an inference attack, after observing an event e.

An example of the levels of the risk on the operational location of the incumbent is shown in table 5.1. The spectrum manager may adopt different levels of risk for each sensitive parameter.

TABLE 5.1: Example of inference risk levels on location

Definition	State
$0\% \leq IR_l(e) < 25\%$	Low
$25\% \leq IR_l(e) < 50\%$	Medium
$50\% \leq IR_l(e) < 75\%$	High
$75\% \leq IR_l(e) \leq 100\%$	Critical

Although the inference risk is a useful metric for optimizing privacy-preserving algorithms, finding well-defined likelihood and impact functions can be challenging.

5.3.4 Risk Analysis

Quantifying the risk helps select an obfuscation technique. Thus, the level of obfuscation needed to prevent inference changes according to the level of privacy risk. For example, an inference risk of 20% results in less obfuscation than an inference risk of 60%.

In this phase, the system takes into consideration the system status as well as the sharing policy (incumbent requirements and secondary preferences). By doing so, the system decides on a number of thresholds to be maintained while choosing the right privacy-preserving technique to be implemented:

- blocking rate or spectrum availability,
- attacker's uncertainty,
- secondary query rate,
- etc.

5.3.5 Risk Management

The obfuscation techniques from data mining and data publishing can be used to protect against inference attacks. Data obfuscation enables databases to add ambiguity while preserving their usefulness, thereby adding uncertainty to the inference process.

Although in this chapter we primarily consider obfuscation to increase attacker uncertainty, there are other obfuscation techniques that increase the attacker's inference cost, which we measure as the number of queries or total time the attacker needs to perform an inference attack. An example of an obfuscation technique that increases the attacker's cost is limiting the query (request) rate of secondary users as explained in Chapter 3. There are also techniques that target a specific individual attacker, such as limiting or denying spectrum requests for a period of time as shown in Chapter 4.

The risk analysis and risk management phases in Fig. 5.1 are key to choosing the right obfuscation technique to implement subject to the limitations of the system. These two phases will be discussed in detail in next sections.

5.4 Privacy-Preserving Architecture

As stated earlier in Chapter 2, privacy-preserving models have been studied in the context of data mining [77] and data publishing [78]. A common architecture of privacy-preserving models includes data owners who own records and seek privacy protection, data publishers who collect data from record owners and prepare it for release and

data recipients who receive information from data publishers to conduct data mining. Traditionally, obfuscation is applied at the data publisher level.

We can use a similar architecture for centrally-coordinated spectrum sharing systems. Secondary users correspond to the data recipients. The spectrum manager corresponds to the data publisher. We have added an anonymizer function [118] as the entity that provides obfuscated data extracted from the original data. Primary users (incumbents) correspond to the data owners. In some spectrum sharing systems, there is no direct communication between the spectrum manager and incumbents, and the data is collected through sensors and analyzed by a decision system. Fig. 5.2 illustrates this architecture.

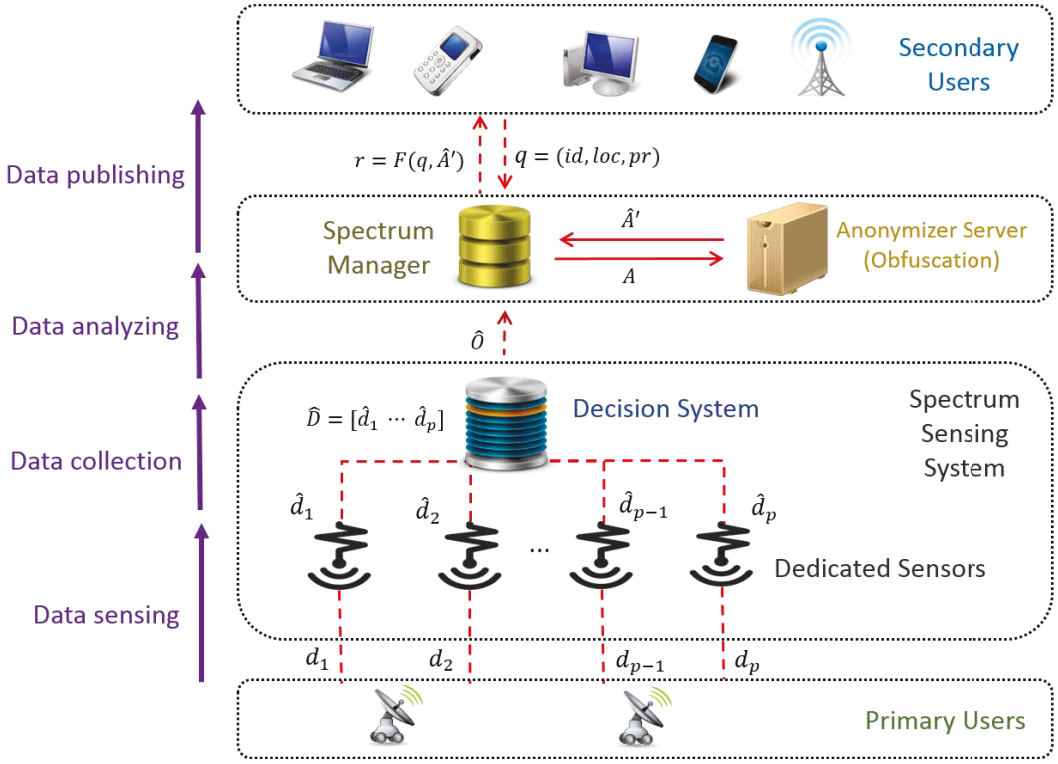


FIGURE 5.2: Privacy-preserving architecture for spectrum sharing

In the data sensing phase, the data collectors report in real time the activity of the incumbents to the data analyzer (i.e., the decision system). This data may include estimation errors. Thus, it is denoted $\hat{d}_j = d_j + e_j$, where d_j and e_j are respectively the true activity and the estimation error of a sensor j.

In the data collection phase, the decision system collects data from dedicated sensors, combines sensing data in $\hat{D}$ and determines whether incumbents are operating, and if

so, what frequencies they are operating on.

$$\hat{D} = \begin{bmatrix} \hat{d}_1 & \hat{d}_2 & \dots & \hat{d}_p \end{bmatrix} \tag{5.4}$$

Next, the decision system computes a matrix of occupancy $\hat{O}$ and delivers those calculations to the spectrum manager. The matrix of occupancy is a 3-D matrix (location, frequency, time) [119]: the i^{th} matrix row refers to the i^{th} channel, the j^{th} matrix column refers to the j^{th} sensor, and the k^{th} matrix column refers to the k^{th} time interval. The dimensions of the matrix are $n \times p \times q$.

$$\hat{O} = \begin{bmatrix} \begin{bmatrix} \hat{o}_{111} & \hat{o}_{121} & \dots & \hat{o}_{1p1} \\ \hat{o}_{211} & \hat{o}_{221} & \dots & \hat{o}_{2p1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{o}_{n11} & \hat{o}_{n21} & \dots & \hat{o}_{np1} \end{bmatrix} \\ \begin{bmatrix} \hat{o}_{112} & \hat{o}_{122} & \dots & \hat{o}_{1p2} \\ \hat{o}_{212} & \hat{o}_{222} & \dots & \hat{o}_{2p2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{o}_{n12} & \hat{o}_{n22} & \dots & \hat{o}_{np2} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \hat{o}_{11q} & \hat{o}_{12q} & \dots & \hat{o}_{1pq} \\ \hat{o}_{21q} & \hat{o}_{22q} & \dots & \hat{o}_{2pq} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{o}_{n1q} & \hat{o}_{n21q} & \dots & \hat{o}_{npq} \end{bmatrix} \end{bmatrix} \tag{5.5}$$

This gives a visual representation of the spatio-temporal channel occupancy. It shows where in time and space a channel is occupied. It also simplifies the understanding of the interactions between different entities of the privacy-preserving architecture.

In the data analyzing phase, the spectrum manager generates a matrix of availability referred to as $\hat{A}$ based on $\hat{O}$.

$$\hat{A} = \begin{bmatrix} \begin{bmatrix} \hat{a}_{111} & \hat{a}_{121} & \dots & \hat{a}_{1p1} \\ \hat{a}_{211} & \hat{a}_{221} & \dots & \hat{a}_{2p1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{a}_{n11} & \hat{a}_{n21} & \dots & \hat{a}_{np1} \end{bmatrix} \\ \begin{bmatrix} \hat{a}_{112} & \hat{a}_{122} & \dots & \hat{a}_{1p2} \\ \hat{a}_{212} & \hat{a}_{222} & \dots & \hat{a}_{2p2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{a}_{n12} & \hat{a}_{n22} & \dots & \hat{a}_{np2} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \hat{a}_{11q} & \hat{a}_{12q} & \dots & \hat{a}_{1pq} \\ \hat{a}_{21q} & \hat{a}_{22q} & \dots & \hat{a}_{2pq} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{a}_{n1q} & \hat{a}_{n21q} & \dots & \hat{a}_{npq} \end{bmatrix} \end{bmatrix} \tag{5.6}$$

An anonymizer function generates an altered (obfuscated) availability matrix referred to as $\hat{A}' = G(\hat{A})$, where $G(\cdot)$ is an obfuscation function. The matrix $\hat{A}'$ is generated by applying obfuscation to one or more rows (i.e., location obfuscation), one or more columns (i.e., frequency obfuscation), or a combination of both (i.e., location/frequency obfuscation). Obfuscation can be also applied to the time column, representing obfuscation of the operational time intervals of incumbents.

In the data publishing phase, the data recipient (i.e., the secondary user) sends a request q to the spectrum manager requesting spectrum resources. The query q includes the identifier id, the location loc and the profile (antenna parameters, technology used, other operational parameters, etc.) pr of the user. The spectrum manager replies with a list of available resources from the altered matrix, referred to as $r = F(q, \hat{A})$, where $F(\cdot)$ is a function of the spectrum manager that returns a response to a querier depending on its characteristics and the status of the spectrum.

For the remainder of this chapter, we neglect the data estimation error e at the sensors and assume the decision system has perfect knowledge about the operational

incumbents, i.e., $\hat{d}_j = d_j$ for $j = 1, \ldots, p$ and consequently $\hat{A} = A$. For the sake of readability, we show the real-time matrix of availability (i.e., 2-D matrix with only location and frequency) in the examples that follow. The extensions to 3-D are trivial. We also assume that the sensors $\{1, \ldots, p\}$ are placed in different locations to monitor the activity of incumbents, and that they return binary values about the occupancy of the spectrum. Thus, $a_{ij} = 1$ when a channel i is available for use by a secondary user at location j; $a_{ij} = 0$ otherwise. The same matrix representation can be used in spectrum sharing system where the primary notifies the spectrum manager directly that it is operating. In this case, the location j refers to an area of operation.

5.5 Privacy-Preserving Techniques

Privacy-preserving techniques insert ambiguity into the dataset itself (i.e., the matrix of availability) by increasing the uncertainty of an attacker or into the release mechanism (i.e., the reply to a query) by increasing the cost to an attacker.

5.5.1 Obfuscation of the Matrix of Availability

In data mining, obfuscation is obtained by generalizing and/or suppressing parts of the data so that no individual can be uniquely distinguished [76]. In spectrum sharing, obfuscation is applied to the incumbent's location, frequency and time of operation.

Obfuscation of the Location

We can apply obfuscation to location by creating a larger area of operation and injecting ambiguity into the attacker's estimated incumbent's location.

For example, if we have five channels, five locations and the following matrix of availability A, where an incumbent is operating on channel N^o 3 in location N^o 2

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

We enlarge the area of operation, and the anonymized matrix of availability becomes

$$A' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

It adds anonymity to the location and increases the uncertainty of an attacker. The latter does not know if there is only one incumbent or two operating in two different locations.

Obfuscation of the Frequency

We can also apply obfuscation to frequency. Frequency obfuscation can be achieved by removing channels availabilities from a location. Obfuscated channels can be either adjacent channels or random non-adjacent channels. Also, frequency obfuscation can be applied in one or more locations.

If we use the same example from above, the matrix of availability A is

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

And we apply frequency obfuscation by removing two available channels in addition to the incumbent's channel, the matrix of availability becomes

$$A' = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Obfuscation of the Time

Authors in [85] suggested time obfuscation by combining successive operational interval times. This is possible in a system where incumbents inform the spectrum manager beforehand about their operations. The spectrum manager cannot do this when spectrum resource allocation is based on real-time sensing. However, the spectrum manager can apply time obfuscation in a sensed system by extending the time that a channel is marked as occupied after the incumbent stops operating. The spectrum manager can also use time obfuscation in either system by simulating an incumbent activity when none is present.

Obfuscation of the Location, Frequency and Time

In order to increase the uncertainty of the attacker, obfuscation can be applied to more than one sensitive parameter: either combine location and frequency, or location and time, or frequency and time, or location and frequency and time.

For example, we can make two additional channels unavailable for use by secondaries in two different locations, even though the incumbent is operating on one channel in one location as follows:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

where A is the original matrix of availability.

$$A' = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

where A' is the obfuscated matrix of availability.

One of the advantages of the matrix representation of the spatio-temporal channel availability is that it makes such a technique easy to implement and monitor.

5.5.2 Obfuscation of the Reply

Obfuscation of the reply prevents some parameters from appearing more frequently than others in reply to a query [81].

In the context of spectrum sharing, the spectrum manager returns a list of m availabilities. We apply obfuscation by providing fewer availabilities, i.e., instead of returning m channels available for use, the spectrum manager will return only $m' < m$ channels available for use. For example, a spectrum manager that returns three channels available per use per query without obfuscation may return only one channel with obfuscation.

Channel assignment schemes can also be used to apply obfuscation to the reply. For example, a spectrum manager can assign the same channel to the same secondary on successive queries, if available. Another effective scheme is for the spectrum manager to always assign the lowest available channel [23]. The spectrum manager can bound the query rate of a user to a maximum threshold as well. Hence, the cost of an attack will increase. While an attacker may still be able to infer incumbent channels, it will take longer to do so.

5.6 Quantification Metrics

In this section, we propose different metrics to evaluate both the availability of the spectrum and the privacy of the incumbent. We also introduce mathematical optimization as a solution to the achieve the optimal trade-off.

5.6.1 Spectrum Availability

Obfuscated data is distorted data and therefore incurs information loss [78]. In spectrum sharing systems, this translates to loss of spectrum. Since the main goal of sharing is to meet the increasing demand for bandwidth and maximize the use of spectrum, we need to quantify the impact of obfuscation on spectrum utilization. Some metrics have been proposed for doing this, including spectrum efficiency. Spectrum efficiency measures the data rate for a given bandwidth [85]. This is a low-level metric more suitable for physical or MAC layer performance.

We would like a metric that reflects system performance at a higher level, such as one that measures whether or not a secondary user's request for spectrum will be granted. A simple, well-understood and meaningful metric of spectrum availability is the blocking probability $P_b(j)$, the probability that a secondary user's request to use the spectrum is denied in a location j. In many cases, we are interested in the absolute blocking probability. Other times, we might be more interested in the resulting change in the blocking probability, i.e.,

$$\Delta P_b(j) = P_b(j) - P_b'(j) \tag{5.7}$$

where $P_b(j)$ and $P_b'(j)$ denote the blocking probability in a location j before and after obfuscation, respectively.

5.6.2 Privacy

The assessment of privacy requires the definition of a proper metric. The Hamming distance is a measure of the distance between two binary strings, vectors or matrices. In this case, we use it to measure the difference between the obfuscated availability matrix and the original availability matrix. That is,

$$P_r = D_H(A', A) \tag{5.8}$$

where $D_H(\cdot)$ is the Hamming distance.

The normalized privacy due to obfuscation is defined as the Hamming distance divided by the size of the matrix of availability.

$$\tilde{P}_r = \frac{D_H(A', A)}{n \times p \times q} \tag{5.9}$$

where n is the number of channels, p is the number of locations, q is the number of time intervals considered by a spectrum manager.

We may also want the normalized privacy in a single location j by comparing the original vector of availability z_j and the obfuscated vector of availability z_j'.

$$\tilde{P}_r(j) = \frac{D_H(z_j', z_j)}{n} \tag{5.10}$$

5.6.3 Trade-off: Optimization

Applying obfuscation incurs loss in spectrum resources, so we need to measure the trade-off between spectrum availability and privacy protection. However, having a single metric for the trade-off, such as the ratio of the two, is of limited value in the design and operation of a spectrum sharing system. More useful is framing the competing goals of incumbent privacy and spectrum availability as a constrained optimization problem. In constrained optimization, an objective function is maximized or minimized subject to a set of constraints on the variables in the objective function or on quantities related in some way to the variables in the objective function. The constraints can be hard constraints or soft constraints. A hard constraint is one that is required to hold; a soft constraint is one that is not. Soft constraints are usually assigned a penalty or weight according to the degree that they do not hold. Constraints may also be ranked in priority, where higher priority constraints are satisfied before lower priority ones are considered.

Constrained optimization is a broad and well-studied field. There are many ways to formulate the problems and a wide variety of solution techniques and strategies [120][121]. It is beyond the scope of this chapter for us to give a comprehensive overview. Rather, we are demonstrating the usefulness of the approach.

Using the notation and metrics developed in this chapter, we can formulate the spectrum sharing system performance criteria for privacy and spectrum utilization in useful and quantitative ways. For example, if we want to maximize the total incumbent privacy while requiring that the probability that a secondary's request to transmit is granted is above 95% in all locations, we can express it as

$$\begin{aligned} \max \quad & D_H(A', A) \\ \text{subject to} \quad & P_b(j) < 0.05 \quad \text{for } j = 1, ..., p \end{aligned} \tag{5.11}$$

where $D_H(\cdot)$ is the Hamming distance between two matrices and $P_b(j)$ is the blocking probability in location j. Since the $P_b(j)$ values depend on the system load in location j as well as the level of obfuscation, the constraints on them should be soft constraints.

More complex and specific requirements can easily be expressed as well. For example, suppose we want to maximize the total privacy while ensuring that the frequency

in location 3 is obfuscated by making two additional channels unavailable for use by secondaries due to an elevated risk at that location, with a desire that the probability that a secondary's request to transmit is decreased by no more than 4% in any location. Here we have a hard privacy requirement (constraint) that is local to one location and a soft constraint on the impact that the obfuscation added has on spectrum availability. This can be expressed as

$$\begin{aligned} \max \quad & D_H(A', A) \\ \text{subject to} \quad & W_H(z'(j)) \geq 3 \quad \text{for } j = 1, ..., p \\ & \Delta P_b(j) < 0.04 \quad \text{for } j = 1, ..., p \end{aligned} \tag{5.12}$$

where $W_H(\cdot)$ is the Hamming weight of a vector, and $z'(j)$ is the j^{th} column in A' and thus represents the modified (obfuscated) location j. $\Delta P_b(j)$ is the difference in the blocking probability in location j after obfuscation and before. $D_H(\cdot)$ is as defined above.

5.7 Example Evaluation

The current work on privacy for spectrum sharing lacks a generic framework to detect a threat and implement the corresponding privacy-preserving technique to mitigate its impact. Therefore, we can not evaluate our proposal against other proposals. In this case and for this section, we showcase an example to analyze the trade-off between spectrum availability and privacy preservation.

To illustrate the trade-offs, we look at an example using the system model of [23]. For a location j, the system is an $M/M/l_j/l_j$ queuing system. There are n channels in total, with l_j channels available for use by secondaries. The remaining $n - l_j$ channels are used by incumbents. P_s is the probability that there are s secondary users active in the system.

$$P_s = \frac{\frac{\rho^s}{s!}}{\sum_{i=0}^{l_j} \frac{\rho^i}{i!}};\ 0 \leq s \leq l_j \tag{5.13}$$

where ρ is the system load.

Blocking occurs when the system has no free channels to assign. Therefore, the blocking probability is equal to the probability that there are l_j users in the system, $P_b(j) = P_{l_j}$.

To protect the incumbent's privacy, the spectrum manager implements k-anonymity [76] for frequency. This technique removes $k-n+l_j$ channels in addition to the incumbent channels from the vector of availability in location j, so that only $n-k$ channels are available for use by secondaries.

Therefore, the change in spectrum availability is

$$\Delta P_b(j) = P_{l_j} - P_{l'_j} = \frac{\frac{\rho_j^l}{l_j!}}{\sum_{i=0}^{l_j} \frac{\rho^i}{i!}} - \frac{\frac{\rho_j^{l'_j}}{l'_j!}}{\sum_{i=0}^{l'_j} \frac{\rho^i}{i!}} \tag{5.14}$$

where P_{l_j} denotes the blocking probability of a secondary before obfuscation and $P_{l'_j}$ denotes the blocking probability of a secondary after obfuscation in j.

The Hamming distance between the original vector of availability z_j and the obfuscated vector of availability z'_j is equal to $k - n + l_j$. Hence, the normalized privacy in location j is

$$\tilde{P}_r(j) = \frac{D_H(z'_j, z_j)}{n} = \frac{k - n + l_j}{n} \tag{5.15}$$

Consider the case where the location j serves 100 channels ($n = 100$). There is one incumbent occupying one channel, and $l_j = n - 1 = 99$. We vary the values of k from 1 (no obfuscation) to n (full obfuscation). We also vary the system load ρ to analyze its impact on privacy. The system load here refers to the average number of requests multiplied by the average time of channel occupancy. Figures 5.3 and 5.4 show the trade-off between spectrum availability and privacy.

In Fig. 5.3, we plot the normalized privacy (i.e., normalized Hamming distance) against the blocking probability for different system loads. First, we note that for all values of the system load ρ, normalized privacy increases monotonically as the blocking probability increases. Next, we see that the normalized privacy at a given value of the blocking probability decreases as the system load increases. For example, at a blocking probability of 0.2, the normalized privacy is just above 0.7 for $\rho = 20$, about 0.55 for $\rho = 50$ and about 0.15 for $\rho = 100$.

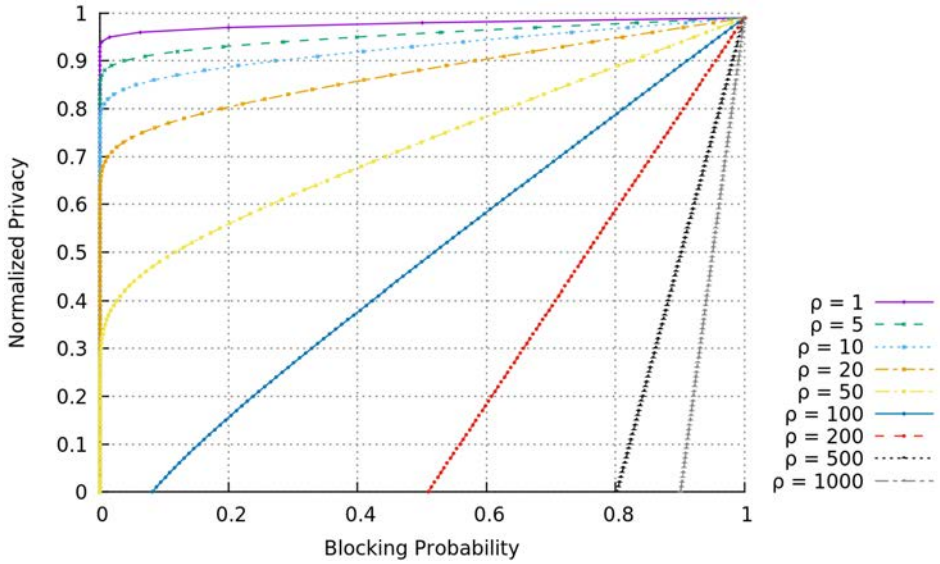


FIGURE 5.3: Blocking probability vs. Hamming distance ($n = 100$)

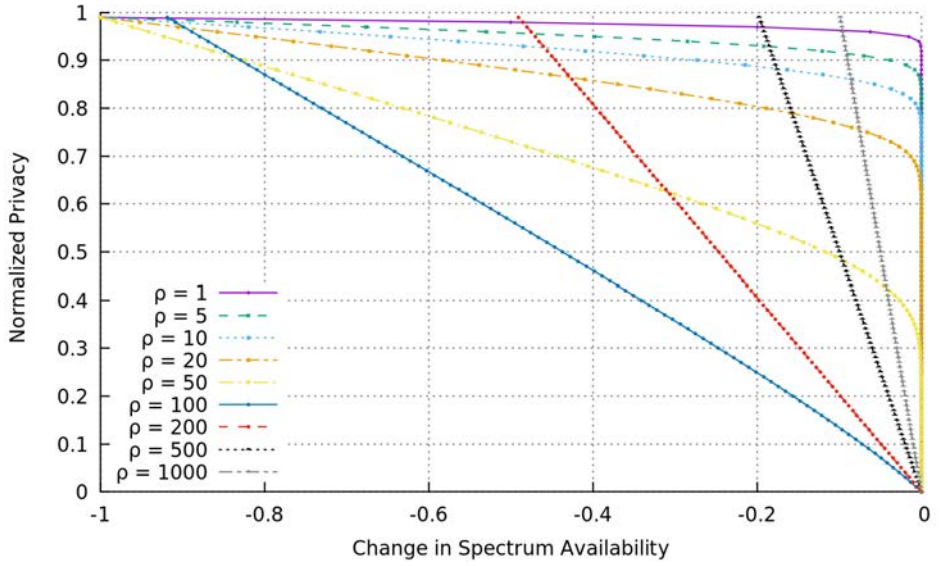


FIGURE 5.4: Change in spectrum availability vs. Normalized privacy ($n = 100$)

For low and very low system loads ($1 \leq \rho \leq 20$), significant levels of obfuscation can be applied with almost no impact on spectrum availability, i.e., $P_b(j) \approx 0$. For example, for $\rho = 20$, a normalized privacy of just under 0.7 can be achieved with no noticeable loss of spectrum availability. The number of channels available for use after obfuscation is still large enough to satisfy the demand of secondary users.

At a medium system load of $\rho = 50$ (i.e., $\rho = \frac{n}{2}$), a normalized privacy of about 0.3 can be achieved with a blocking probability close to zero. For a high system load of $\rho = 100$ ($\rho = n$), however, the blocking probability is non-negligible without any obfuscation applied. The trade-off between privacy and blocking probability becomes linear with a slope of $\frac{\rho}{n} = 1$ as blocking probability increases.

For very high system loads where $\rho \geq 200$, the blocking probability is high with no obfuscation applied, and any obfuscation increases both the normalized privacy and the blocking probability. The relationship between normalized privacy and blocking probability is linear with a slope of $\frac{\rho}{n}$.

In Fig. 5.4, we plot the normalized privacy as a function of the change in spectrum availability. The change in spectrum availability is negative, because the blocking probability increases with obfuscation. This plot highlights the impact that increasing privacy has on spectrum availability for various system load levels.

5.8 Conclusions

In this chapter, we presented a risk-based privacy-preserving framework for spectrum sharing. We introduced an architecture for centrally-coordinated spectrum-sharing systems based on those used in data publishing and data mining. We developed a notation for representing incumbent occupancy in frequency, location and time at each level of the architecture.

In the framework, the spectrum manager performs risk monitoring to detect anomalous behavior in secondary request patterns, which trigger a threat event. A threat event moves the system into the risk identification and assessment stages, where a risk level is computed using the inference risk metric. Risk analysis uses the resulting risk level to

determine the appropriate privacy measures. In risk management, a constrained optimization is solved to determine the trade-off between the desired privacy and spectrum availability. We also use the normalized Hamming distance between the binary obfuscated and original (unobfuscated) spectrum availability matrices as a privacy measure and the blocking probability as a system-level metric for spectrum utilization.

In literature, many privacy-preserving techniques have been proposed to alleviate inferences and ensure privacy. They deliver specific results for specific cases, and most of their assumptions are invalid in a real-world scenario. In reality, it is difficult to implement those techniques without prior knowledge about the protectee or even the risk encountered. Our framework is an important baseline for future work in risk detection and mitigation for spectrum sharing purposes, as it provides the best trade-off between privacy preservation and spectrum efficiency. Instead of developing a static standalone algorithm, this framework allows to dynamically adjust the privacy-preserving technique based on the system requirements. It also enables real-time interaction between the multiple entities of the spectrum, and can be easily incorporated into a functional system in order to deliver a better performance overall.

Chapter 6

Conclusions and Perspectives

6.1 Conclusions

Given the explosive growth in mobile broadband traffic, the world has engaged in innovative frequency allocation approaches instead of exclusive and high cost static spectrum. As spectrum assigned to governmental agencies have been shown to be underutilized, regulatory agencies have considered the sharing option. However, the sensitivity and limited mobility of the incumbent as well as the risk of insider attacks urge the need for privacy protection. In particular, the location, frequency and time of operation of the incumbent (i.e., primary user) should not be revealed. Authorized spectrum users (i.e., secondary users) can exploit the shared environment and acquire unauthorized access to sensitive information using legitimate queries. Therefore, an adversary may be able to infer the operational parameters of the incumbent by combining a-priori knowledge with acquired knowledge. This is known as an inference attack, and such an attack is hard to detect and to defend against. In this thesis, we are mainly concerned with the protection of two operational parameters: frequency and location. We first analyzed the vulnerability of each parameter to inference attacks and then proposed privacy-preserving techniques.

Initially, we examined the requirements of Federal incumbents in spectrum sharing. The success of the sharing between Federal and commercial users imply interference management and privacy preservation. Even though interference mitigation is still a current concern, researchers are taking interest in operational security. We also surveyed the state of the art of privacy-preserving techniques, mainly obfuscation. Obfuscation is widely used in data mining and data publishing. Its purpose is to alter the data

and hence increase the uncertainty of an adversary. Some of those techniques can be implemented in the 3.5 GHz band. However, even though they enhance the privacy of the incumbent, they result in spectrum loss for secondary users. Quantification is then important to evaluate a privacy-preserving technique. Multiple spectrum utilization and privacy metrics exist to evaluate privacy-preserving techniques. However, not all necessarily apply for the 3.5 GHz band case. We argue that the need for more appropriate metrics is still persistent.

Next, we evaluate the privacy of the frequency of the incumbent to inference attacks and propose inherent and explicit obfuscation to ensure its protection. Using basic parameters (random channels assignment scheme, unrestricted secondary query rate and unlimited spectrum resources), the adversary is able to infer quickly and accurately the operational channel. By implementing other channel assignment schemes (ordered and semi-static) and limiting the secondary query rate, we enhance the privacy of incumbent. The adversary takes longer to infer the the operational channel under low and moderate system loads. Under high system load, the semi-static channel assignment scheme performs better than the ordered channel assignment scheme. In order to decrease the knowledge acquired by the adversary, we intentionally keep some spectrum resources idle. By the end of the inference attack, the adversary is not able to uniquely decide which channel is the operational channel of the incumbent. Even though such an approach makes some spectrum unavailable for secondary users, we conclude that the gain in privacy is always greater than the loss in spectrum. For the best results, we combine inherent and explicit obfuscation. This increases both the cost and the uncertainty of an inference attack.

Afterwards, we evaluate the privacy of the location of the incumbent to inference attacks and propose obfuscation and trustworthiness to ensure its protection. We show that the operational location of the incumbent is threatened when the adversary is able to infer the boundaries of the operational zone. Therefore, we focus on securing those boundaries. We first suggest obfuscation by adding noise to the operational zone. This technique allows to include perturbation at the boundaries of the operational zone. The operational zone becomes larger. The adversary is hence unable to infer correctly the boundaries of the operational zone. This technique is simple to implement and effective

privacy-wise. However, it induces spectrum resources loss and affects other secondary users. In other words, the spectrum is no longer available for use in the region impacted by obfuscation, even though it does not generate any interference for the incumbent. So, we can use trustworthiness of secondary users to remedy the drawbacks of obfuscation. In fact, trust can be defined by assigning a reputation metric to each secondary user. Note that we don't add any perturbation to the operational zone when using the trust metric. Its implementation results provide high protection of the incumbent by detecting a potential adversary and ultimately eliminating it from the sharing environment. This technique has shown to be effective since it does not result in any loss in spectrum resources.

Finally, we proposed a generic privacy-preserving architecture. It applies to any incumbent in the shared environment seeking privacy protection. It includes a three-dimension matrix that represents the spatio-temporal channel occupancy. The implementation of a privacy-preserving technique becomes, therefore, easier to manage and evaluate. We also proposed a generalized privacy-preserving framework for spectrum sharing. It is a five-step risk management framework that does not depend on the nature of the incumbent or the type of the countermeasure. It depends mainly on the detected risk. The spectrum manager quantifies the risk using thresholds and other likelihood and impact functions. Once the risk level is evaluated, a criteria of obfuscation is selected. This phase includes the determination of the number of allowed queries per secondary user, the spectrum loss tolerance and the attacker uncertainty tolerance. Those measures addresses the trade-off between spectrum efficiency and privacy protection. Then, the spectrum manager chooses a privacy-preserving technique that meets those criteria, and implements it.

This work adds a flexibility and easiness to deal with such threats, reduces computational time, saves energy, and provides better privacy by either slowing down the inference process or making the adversary uncertain about the inferred information. The privacy-preserving technique does not need to be implemented beforehand. First, the spectrum manager assesses a threat, selects the spectrum loss tolerance based on the system load (i.e., number of secondaries), and sets the maximum knowledge that can be acquired by the attacker. Then, it chooses and implements the best methodology

to protect against that threat. For example, if the system load is low, it is easier and more efficient to implement perturbation instead of trustworthiness. In fact, the loss in spectrum will not affect other operating secondaries as there will be enough spectrum resources to accommodate each secondary. Also, the spectrum manager does not need to keep updating the trustworthiness of each secondary, which avoids additional computation. The spectrum manager is therefore able identify the best approach to be implemented based on the risk level to protect the operational parameters of the incumbent and guarantee a safe and secure spectrum sharing.

6.2 Perspectives

This thesis has mainly focused on protecting the operational security of Federal operations in the shared environment. In the next decade, additional bands will be opened for sharing as well. So, many opportunities for extending this work remain open.

There are a number of challenges related to the models proposed in this thesis. The operational time of an incumbent was not explicitly protected against inference attacks. While protecting the operational frequency and location, the system is able to provide some protection against the inference of the operational time. However, more straightforward privacy-preserving approaches should be applied. Furthermore, in the absence of a real deployed system operating in a shared environment, our evaluation is simulation-based. Once the first spectrum access systems and sensors are approved and certified, we envision implementing a real-time evaluator to address the vulnerability of the incumbent to inference attacks. The privacy-preserving techniques proposed in this thesis will be incorporated into a real system, and their resiliency and sustainability will be put to test. It is essential to examine the efficiency of our design, and to ensure that it matches simulation results.

Future work also includes spectrum regulation. In order to keep the shared environment secure, some policy enforcement mechanisms should be implemented. They can be either preventive (ex ante) or punitive (ex post). We already discussed in this thesis the preventive mechanisms, such as privacy-preserving techniques (e.g., inherent and explicit obfuscation). However, those techniques does not perfectly guarantee the protection of

the incumbent, as the capabilities of an adversary are unforeseeable. As result, punitive mechanisms are needed to provide full control of the system. We introduced a punitive method by assigning a reputation metric to secondary users and eliminating untrustworthy secondary users. Other solutions to identify, localize and punish an adversary can be envisioned, and they can be implemented by the spectrum manager or the regulatory authority or both. Data collection and investigation is considered to detect and address intrusions, and punishments can be regulatory or economic or even legal.

Bibliography

[1] *Facilitating Opportunities for Flexible, Efficient and Reliable Spectrum Use Employing Cognitive Radio Technologies, Notice of Proposed Rulemaking and Order.* ET Docket no. 03-322. Federal Communications Commission (FCC), Dec. 2003.

[2] *Report of the Spectrum Efficiency Working Group.* Spectrum policy task force, ET Docket No. 02-135. Federal Communications Commission (FCC), Nov. 2002.

[3] Cellular Telecommunications Industry Association (CTIA). *Wireless Quick Facts.* `http://www.ctia.org/industry-data/wireless-quick-facts`. [Online; accessed 03-March-2017]. 2016.

[4] *Realizing the Full Potential of Government-Held Spectrum to Spur Economic Growth.* Tech. rep. President's Council of Advisors on Science and Technology (PCAST), July 2012.

[5] U.S. Census Bureau. *World Population: 1950–2050, International Data Base.* `https://www.census.gov/population/international/data/idb/worldpopgraph.php`. [Online; accessed 03-March-2017]. 2016.

[6] Defense Advanced Research Projects Agency (DARPA). *Dr. Preston Marshall Speech.* `http://archive.darpa.mil/DARPATech2002/presentations/ato_pdf/speeches/MARSHALL.pdf`. [Online; accessed 03-March-2017].

[7] Defense Advanced Research Projects Agency (DARPA). *Spectrum Challenge.* `http://archive.darpa.mil/spectrumchallenge/`. [Online; accessed 03-March-2017].

[8] *Enhancing Access to the Radio Spectrum (EARS) – Addressing Future Challenges.* Program Solicitation NSF 16-537. National Science Foundation (NSF), 2016.

[9] *Spectrum Efficiency, Energy Efficiency, and Security (SpecEES): Enabling Spectrum for All.* Program Solicitation NSF 16-616. National Science Foundation (NSF), 2016.

[10] National Science Foundation (NSF). *Press Release 16-121.* `https://www.nsf.gov/news/news_summ.jsp?cntn_id=189863&WT.mc_ev=click`. [Online; accessed 03-March-2017].

[11] National Spectrum Consortium (NSC). `http://www.nationalspectrumconsortium.org/`. [Online; accessed 03-March-2017].

[12] *Federal Relocation Costs and Auction Revenues, a Report to the Committee on Armed Services, U.S. Senate.* GAO-13-472. Government Accountability Office (GAO), May 2013.

[13] *Report and Order and Second Further Notice of Proposed Rulemaking.* Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550-3650 MHz Band, GN Docket No. 12-354. Federal Communications Commission (FCC), Apr. 2015.

[14] *An Assessment of the Near-Term Viability of Accommodating Wireless Broadband Systems in the 1675-1710 MHz, 1755-1780 MHz, 3500-3650 MHz, and 4200-4220 MHz, 4380-4400 MHz Band.* Fast Track Report. National Telecommunications and Information Administration (NTIA), Oct. 2010.

[15] Y. Ye, D. Wu, Z. Shu, and Y. Qian. "Overview of LTE Spectrum Sharing Technologies". In: *IEEE Access* 4 (2016), pp. 8105–8115.

[16] *A Regulatory Framework for Radio Spectrum Policy in the European Community (Radio Spectrum Decision).* Decision No 676/2002/EC of the European Parliament and of the Council. European Commission (EC), Mar. 2002.

[17] *A Framework for Spectrum Sharing.* Statement. Office of Communications (OFCOM), Apr. 2016.

[18] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty. "NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A survey". In: *Computer networks* 50.13 (2006), pp. 2127–2159.

[19] *Final Report*. Interference and Dynamic Spectrum Access Subcommittee. Commerce Spectrum Management Advisory Committee (CSMAC), Nov. 2011.

[20] *Notice of Proposed Rulemaking*. Unlicensed Operation in the TV Broadcast Bands, ET Docket No. 04-186. Federal Communications Commission (FCC), May 2004.

[21] *Order*. Unlicensed Operation in the TV Broadcast Bands, ET Docket No. 04-186. Federal Communications Commission (FCC), Jan. 2011.

[22] L. Buttyan and J.-P. Hubaux. *Security and Cooperation in Wireless Networks: Thwarting Malicious and Selfish Behavior in the Age of Ubiquitous Computing*. Cambridge University Press, 2007.

[23] A. Ben-Mosbah, T. A. Hall, M. Souryal, and H. Afifi. "Analysis of the Vulnerability of the Incumbent Frequency to Inference Attacks in Spectrum Sharing". In: *IEEE Consumer Communications and Networking Conference (CCNC'17)*. Las Vegas, NV, Jan. 2017.

[24] A. Ben-Mosbah, T. A. Hall, M. Souryal, and H. Afifi. "An Analytical Model for Inference Attacks on the Incumbent's Frequency in Spectrum Sharing". In: *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN'17)*. Baltimore, MD, Mar. 2017.

[25] J. M. Peha and S. Panichpapiboon. "Real-time Secondary Markets for Spectrum". In: *Telecommunications Policy* 28.7 (2004), pp. 603–618.

[26] *Text Equivalent Version of the U.S. Frequency Allocation Chart*. 2013 Edition (Rev. 5/2014). National Telecommunications and Information Administration (NTIA), May 2014.

[27] M. Song, C. Xin, Y. Zhao, and X. Cheng. "Dynamic Spectrum Access: From Cognitive Radio to Network Radio". In: *IEEE Wireless Communications* 19.1 (2012), pp. 23–29.

[28] Q. Zhao and B. M. Sadler. "A Survey of Dynamic Spectrum Access". In: *IEEE signal processing magazine* 24.3 (2007), pp. 79–89.

[29] S. Srinivasa and S. A. Jafar. "Cognitive Radios for Dynamic Spectrum Access - The Throughput Potential of Cognitive Radio: A Theoretical Perspective". In: *IEEE Communications Magazine* 45.5 (2007).

[30] L. Luo, P. Zhang, G. Zhang, and J. Qin. "Outage Performance for Cognitive Relay Networks with Underlay Spectrum Sharing". In: *IEEE Communications letters* 15.7 (2011), pp. 710–712.

[31] R. Blasco-Serrano, J. Lv, R. Thobaben, E. Jorswieck, A. Kliks, and M. Skoglund. "Comparison of Underlay and Overlay Spectrum Sharing Strategies in MISO Cognitive Channels". In: *Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), 2012 7th International ICST Conference on*. IEEE. 2012, pp. 224–229.

[32] M. Abdelraheem, M. El-Nainay, and S. F. Midkiff. "Spectrum Occupancy Analysis of Cooperative Relaying Technique for Cognitive Radio Networks". In: *Computing, Networking and Communications (ICNC), 2015 International Conference on*. IEEE. 2015, pp. 237–241.

[33] S. Bhattarai, J.-M. J. Park, B. Gao, K. Bian, and W. Lehr. "An Overview of Dynamic Spectrum Sharing: Ongoing Initiatives, Challenges, and a Roadmap for Future Research". In: *IEEE Transactions on Cognitive Communications and Networking* 2.2 (2016), pp. 110–128.

[34] I. C. Society, I. S. C. C. 4. on Dynamic Spectrum Access Networks, I. of Electrical, E. Engineers, and I.-S. S. Board. *IEEE Standard Definitions and Concepts for Dynamic Spectrum Access: Terminology Relating to Emerging Wireless Networks, System Functionality, and Spectrum Management*. IEEE std. Institute of Electrical and Electronics Engineers. ISBN: 9780738157771. URL: `https://books.google.com/books?id=bk-3jwEACAAJ`.

[35] M. d. da Costa and P. Cardieri. "Collision Probabilities for Dynamic Spectrum Access with Cognitive Radios". In: *Microwave and Optoelectronics Conference (IMOC), 2009 SBMO/IEEE MTT-S International*. IEEE. 2009, pp. 272–276.

[36] S. Huang, X. Liu, and Z. Ding. "Opportunistic Spectrum Access in Cognitive Radio Networks". In: *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*. IEEE. 2008.

[37] M. Sharma and A. Sahoo. "A Comprehensive Methodology for Opportunistic Spectrum Access based on Residual White Space Distribution". In: *Proceedings of the 4th International Conference on Cognitive Radio and Advanced Spectrum Management*. ACM. 2011, p. 46.

[38] K. W. Sung, S.-L. Kim, and J. Zander. "Temporal Spectrum Sharing based on Primary User Activity Prediction". In: *IEEE Transactions on Wireless Communications* 9.12 (2010), pp. 3848–3855.

[39] Y. Pei, A. T. Hoang, and Y.-C. Liang. "Sensing-throughput Tradeoff in Cognitive Radio Networks: How Frequently Should Spectrum Sensing be Carried Out?" In: *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE. 2007, pp. 1–5.

[40] A. Sahoo, M. Souryal, and M. Ranganathan. "Implementation of an Opportunistic Spectrum Access Scheme with Disruption QoS". In: *2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*. IEEE. 2014, pp. 335–340.

[41] S. Homayouni, S. A. Ghorashi, and F. Azizzadeh. "Spectrum Sharing by Bubbanding in Cognitive Radio Networks". In: *Computer and Knowledge Engineering (ICCKE), 2011 1st International eConference on*. IEEE. 2011, pp. 322–326.

[42] H. Qing, X. Shaoyi, and J. Xiaojun. "Performance Evaluation of Secondary Users in Dynamic Spectrum Access system". In: *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE. 2011, pp. 710–714.

[43] Z. Yan, X. Zhang, and W. Wang. "Performance Analysis of Secondary Users in Dynamic Spectrum Access under Interference Temperature Constraints". In: *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*. IEEE. 2010, pp. 2655–2661.

[44] Y. Zhao, S. Mao, J. O. Neel, and J. H. Reed. "Performance Evaluation of Cognitive Radios: Metrics, Utility Functions, and Methodology". In: *Proceedings of the IEEE* 97.4 (2009), pp. 642–659.

[45] M. Mchenry, E. Livsics, T. Nguyen, and N. Majumdar. "XG Dynamic Spectrum Access Field Test Results [Topics in Radio Communications]". In: *IEEE Communications Magazine* 45.6 (2007), pp. 51–57.

[46] J. Boksiner, K. Chang, P. Costello, R. Huck, T. Leising, M. Zankel, Y. Posherstnik, M. Totaro, T. Mai, and J. Molnar. "Testing of Policy-based Dynamic Spectrum Access Radios". In: *MILITARY COMMUNICATIONS CONFERENCE, 2010-MILCOM 2010*. IEEE. 2010, pp. 773–778.

[47] J. M. Peha. "Sharing Spectrum Through Spectrum Policy Reform and Cognitive Radio". In: *Proceedings of the IEEE* 97.4 (2009), pp. 708–719.

[48] G. D. Nguyen, S. Kompella, J. E. Wieselthier, and A. Ephremides. "Channel Sharing in Cognitive Radio Networks". In: *MILITARY COMMUNICATIONS CONFERENCE, 2010-MILCOM 2010*. IEEE. 2010, pp. 2268–2273.

[49] S. V. Alagesan, I.-J. Wang, and X. Liu. *A Cognitive MAC Protocol to Coexist with Reactive Primary Users*.

[50] R. Di Pietro and G. Oligeri. "Jamming Mitigation in Cognitive Radio Networks". In: *IEEE Network* 27.3 (2013), pp. 10–15.

[51] Y. Zhang and L. Lazos. "Vulnerabilities of Cognitive Radio MAC Protocols and Countermeasures". In: *IEEE Network* 27.3 (2013), pp. 40–45.

[52] A. Attar, H. Tang, A. V. Vasilakos, F. R. Yu, and V. C. Leung. "A Survey of Security Challenges in Cognitive Radio Networks: Solutions and Future Research Directions". In: *Proceedings of the IEEE* 100.12 (2012), pp. 3172–3186.

[53] T. R. Newman, T. C. Clancy, M. McHenry, and J. H. Reed. "Case Study: Security Analysis of a Dynamic Spectrum Access Radio System". In: *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*. IEEE. 2010, pp. 1–6.

[54] Q. Pei, L. Li, H. Li, and B. Yuan. "Adaptive Trust Management Mechanism for Cognitive Radio Networks". In: *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE. 2012, pp. 826–831.

[55] R. Chen and J.-M. Park. "Ensuring Trustworthy Spectrum Sensing in Cognitive Radio Networks". In: *Networking Technologies for Software Defined Radio Networks, 2006. SDR'06.1 st IEEE Workshop on*. IEEE. 2006, pp. 110–119.

[56] Z. Chen, T. Cooklev, C. Chen, and C. Pomalaza-Ráez. "Modeling Primary User Emulation Attacks and Defenses in Cognitive Radio Networks". In: *2009 IEEE 28th International Performance Computing and Communications Conference*. IEEE. 2009, pp. 208–215.

[57] R. Yu, Y. Zhang, Y. Liu, S. Gjessing, and M. Guizani. "Securing Cognitive Radio Networks against Primary User Emulation Attacks". In: *IEEE Network* 30.6 (2016), pp. 62–69.

[58] Ö. Cepheli and G. K. Kurt. "Physical Layer Security in Cognitive Radio Networks: A Beamforming Approach". In: *Communications and Networking (BlackSeaCom), 2013 First International Black Sea Conference on*. IEEE. 2013, pp. 233–237.

[59] R. Chen, J.-M. Park, and J. H. Reed. "Defense Against Primary User Emulation Attacks in Cognitive Radio Networks". In: *IEEE Journal on selected areas in communications* 26.1 (2008), pp. 25–37.

[60] R. Chen, J.-M. Park, Y. T. Hou, and J. H. Reed. "Toward Secure Distributed Spectrum Sensing in Cognitive Radio Networks". In: *IEEE Communications Magazine* 46.4 (2008), pp. 50–55.

[61] R. Dubey, S. Sharma, and L. Chouhan. "Secure and Trusted Algorithm for Cognitive Radio Network". In: *2012 Ninth International Conference on Wireless and Optical Communications Networks (WOCN)*. IEEE. 2012, pp. 1–7.

[62] Z. Jin, S. Anand, and K. Subbalakshmi. "Mitigating Primary User Emulation Attacks in Dynamic Spectrum Access Networks Using Hypothesis Testing". In: *ACM SIGMOBILE Mobile Computing and Communications Review* 13.2 (2009), pp. 74–85.

[63] K. Bian, J.-M. Park, X. Du, and X. Li. "Enabling Fair Spectrum Sharing: Mitigating Selfish Misbehaviors in Spectrum Contention". In: *IEEE Network* 27.3 (2013), pp. 16–21.

[64] S. T. Zargar, M. B. Weiss, C. E. Caicedo, and J. B. Joshi. "Security in Dynamic Spectrum Access Systems: A Survey". In: (2009).

[65] W. Wang, H. Li, Y. Sun, and Z. Han. "Attack-proof Collaborative Spectrum Sensing in Cognitive Radio Networks". In: *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*. IEEE. 2009, pp. 130–134.

[66] F. R. Yu, H. Tang, M. Huang, Z. Li, and P. C. Mason. "Defense Against Spectrum Sensing Data Falsification Attacks in Mobile AdHoc Networks with Cognitive Radios". In: *MILCOM 2009-2009 IEEE Military Communications Conference*. IEEE. 2009, pp. 1–7.

[67] C. N. Mathur and K. Subbalakshmi. "Security Issues in Cognitive Radio Networks". In: *Cognitive networks: towards self-aware networks* (2007), pp. 284–293.

[68] C. Sorrells, L. Qian, and H. Li. "Quickest Detection of Denial-of-Service Attacks in Cognitive Wireless Networks". In: *Homeland Security (HST), 2012 IEEE Conference on Technologies for*. IEEE. 2012, pp. 580–584.

[69] J.-M. Park, J. H. Reed, A. Beex, T. C. Clancy, V. Kumar, and B. Bahrak. "Security and Enforcement in Spectrum Sharing". In: *Proceedings of the IEEE* 102.3 (Mar. 2014), pp. 270–281.

[70] *Notice of Proposed Rulemaking and Order*. Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550- 3650 MHz Band, GN Docket No. 12-354. Federal Communications Commission (FCC), Dec. 2012.

[71] *Further Notice of Proposed Rulemaking and Order*. Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550- 3650 MHz Band, GN Docket No. 12-354. Federal Communications Commission (FCC), Apr. 2014.

[72] M. Altamimi, M. B. Weiss, and M. McHenry. "Enforcement and Spectrum Sharing: Case Studies of Federal-commercial Sharing". In: *Available at SSRN 2310883* (2013).

[73] *3.5 GHz Exclusion Zone Analyses and Methodology*. Technical Report. National Telecommunications and Information Administration (NTIA), June 2015.

[74] *Order on Reconsideration and Second Report and Order*. ET Docket No. 12-354. Federal Communications Commission (FCC), Apr. 2016.

[75] G. Shafer. "Detecting Inference Attacks Using Association Rules". In: (2001), pp. 270–281.

[76] L. Sweeney. "k-anonymity: A Model for Protecting Privacy". In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.5 (Oct. 2002), pp. 557–570.

[77] J. Vaidya, C. W. Clifton, and Y. M. Zhu. *Privacy Preserving Data Mining*. Vol. 19. Springer Science & Business Media, 2006.

[78] B. Fung, K. Wang, R. Chen, and P. S. Yu. "Privacy-Preserving Data Publishing: A Survey of Recent Developments". In: *ACM Computing Surveys (CSUR)* 42.4 (June 2014).

[79] B.-C. Chen, D. Kifer, K. LeFevre, A. Machanavajjhala, et al. "Privacy-Preserving Data Publishing". In: *Foundations and Trends® in Databases* 2.1–2 (2009), pp. 1–167.

[80] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar. "Random-Data Perturbation Techniques and Privacy-Preserving Data Mining". In: *Knowledge and Information Systems* 7.4 (2005), pp. 387–414.

[81] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. "l-diversity: Privacy Beyond k-anonymity". In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1.1 (Mar. 2007).

[82] R. J. Bayardo, R. Agrawal, and D. Gunopulos. "Constraint-Based Rule Mining in Large, Dense Databases". In: *Data Engineering, 1999. Proceedings., 15th International Conference on*. IEEE. 1999, pp. 188–197.

[83] A. Friedman and A. Schuster. "Data Mining with Differential Privacy". In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2010, pp. 493–502.

[84] A. Evfimievski. "Randomization in Privacy Preserving Data Mining". In: *ACM Sigkdd Explorations Newsletter* 4.2 (2002), pp. 43–48.

[85] B. Bahrak, S. Bhattarai, A. Ullah, J.-M. Park, J. Reed, and D. Gurney. "Protecting the Primary Users' Operational Privacy in Spectrum Sharing". In: *IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN'14)*. McLean, VA, Apr. 2014, pp. 236–247.

[86] Wireless Innovation Forum. *Spectrum Sharing Committee Working Group 2: Security Requirements*. `http://www.wirelessinnovation.org/ssc-wg2`. [Online; accessed 07-July-2016]. 2016.

[87] *CBRS Operational Security*. Document WINNF-15-S-0071. Wireless Innovation Forum, May 2016.

[88] W. Wang and Q. Zhang. *Location Privacy Preservation in Cognitive Radio Networks*. Springer, 2014.

[89] C. A. Ardagna, M. Cremonini, E. Damiani, S. D. C. Di Vimercati, and P. Samarati. "Location Privacy Protection through Obfuscation-based Techniques". In: *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer. 2007, pp. 47–60.

[90] N. Li, T. Li, and S. Venkatasubramanian. "t-closeness: Privacy beyond k-anonymity and l-diversity". In: *2007 IEEE 23rd International Conference on Data Engineering*. IEEE. 2007, pp. 106–115.

[91] J. Xiang, Y. Zhang, and T. Skeie. "Medium Access Control Protocols in Cognitive Radio Networks". In: *Wireless Communications and Mobile Computing* 10.1 (2010), pp. 31–49.

[92] P. Marshall. *Scalability, Density, and Decision Making in Cognitive Wireless Networks*. Cambridge University Press, 2012.

[93] M. Clark and K. Psounis. "Can the Privacy of Primary Networks in Shared Spectrum be Protected?" In: *IEEE International Conference on Computer Communications (INFOCOM'16)*. San Francisco, CA, Apr. 2016.

[94] J. M. Chaiken and E. Ignall. "An Extension of Erlang's Formulas which Distinguishes Individual Servers". In: *Journal of Applied Probability* 9.1 (Mar. 1972), pp. 192–197.

[95] D. L. Jagerman. "Some Properties of the Erlang Loss Function". In: *Bell System Technical Journal* 53.3 (Mar. 1974), pp. 525–551.

[96] J. D. Little and S. C. Graves. "Little's Law". In: *Building intuition*. Springer, 2008, pp. 81–100.

[97] M. Ferrante and M. Saltalamacchia. "The Coupon Collector's Problem". In: *Materials matemàtics* (2014), pp. 1–35.

[98] J. S. Croucher et al. "Collecting Coupons - A Mathematical Approach". In: (2006).

[99] M. Brown, E. A. Peköz, and S. M. Ross. "Coupon Collecting". In: *Probability in the Engineering and Informational Sciences* 22.02 (2008), pp. 221–229.

[100] A. N. Myers and H. S. Wilf. "Some New Aspects of the Coupon Collector's Problem". In: *SIAM review* 48.3 (2006), pp. 549–565.

[101] I. Adler, S. Oren, S. M. Ross, et al. "The Coupon Collector's Problem Revisited". In: *Journal of Applied Probability* 40.2 (2003), pp. 513–518.

[102] P. Neal. "The Generalised Coupon Collector Problem". In: *Journal of Applied Probability* (2008), pp. 621–629.

[103] A. V. Doumas and V. G. Papanicolaou. "The Siblings of the Coupon Collector". In: *arXiv preprint arXiv:1412.4346* (2014).

[104] A. Boneh and M. Hofri. "The Coupon Collector Problem Revisited — A Survey of Engineering Problems and Computational Methods". In: *Stochastic Models* 13.1 (1997), pp. 39–66.

[105] R. B. Cooper. "Queues with Ordered Servers that Work at Different Rates". In: *Opsearch* 13.2 (1976), pp. 69–78.

[106] *National Coastal Population Report*. Population Trends from 1970 to 2020. National Oceanic and Atmospheric Administration (NOAA), Mar. 2013.

[107] U.S. Census Bureau. *U.S. Census 2010 TIGER / Line Shapefiles*. `https://www.census.gov/geo/maps-data/data/tiger-line.html`. [Online; accessed 03-February-2017]. 2010.

[108] K. Scarfone and P. Mell. "Guide to Intrusion Detection and Prevention Systems (idps)". In: *NIST special publication* 800.2007 (2007), p. 94.

[109] O. León, R. Román, and J. Hernández-Serrano. "Towards a Cooperative Intrusion Detection System for Cognitive Radio Networks". In: *International Conference on Research in Networking*. Springer. 2011, pp. 231–242.

[110] Z. M. Fadlullah, H. Nishiyama, N. Kato, and M. M. Fouda. "Intrusion Detection System (IDS) for Combating Attacks Against Cognitive Radio Networks". In: *IEEE network* 27.3 (2013), pp. 51–56.

[111] O. Leon, J. Hernandez-Serrano, and M. Soriano. "A New Cross-layer Attack to TCP in Cognitive Radio Networks". In: *Cross Layer Design, 2009. IWCLD'09. Second International Workshop on*. IEEE. 2009, pp. 1–5.

[112] J. Hernandez-Serrano, O. León, and M. Soriano. "Modeling the Lion Attack in Cognitive Radio Networks". In: *EURASIP Journal on Wireless Communications and Networking* 2011 (2011), p. 2.

[113] C. Song and Q. Zhang. "Achieving Cooperative Spectrum Sensing in Wireless Cognitive Radio Networks". In: *ACM SIGMOBILE Mobile Computing and Communications Review* 13.2 (2009), pp. 14–25.

[114] S. M. Mishra, A. Sahai, and R. W. Brodersen. "Cooperative Sensing Among Cognitive Radios". In: *2006 IEEE International Conference on Communications*. Vol. 4. IEEE. 2006, pp. 1658–1663.

[115] P. R. Garvey. *Analytical Methods for Risk Management: A Systems Engineering Perspective*. CRC Press, 2008.

[116] D. M. Grether. "Bayes Rule as a Descriptive Model: The Representativeness Heuristic". In: *The Quarterly Journal of Economics* 95.3 (1980), pp. 537–557.

[117] *Privacy Risk Management for Federal Information Systems*. NISTIR 8062 (Draft). National Institute of Standards and Technology (NIST), May 2015.

[118] F. Liu, K. A. Hua, and Y. Cai. "Query l-diversity in Location-based Services". In: *Tenth International Conference on Mobile Data Management: Systems, Services and Middleware (MDM'09)*. IEEE. Taipei, May 2009, pp. 436–442.

[119] A. Solo. "Multidimensional Matrix Mathematics: Notation, Representation, and Simplification, Part 1 of 6". In: *Proceedings of the world congress on engineering*. Vol. 3. 2010, pp. 1824–1828.

[120] M. J. Powell. "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations". In: *Numerical analysis*. Springer, 1978, pp. 144–157.

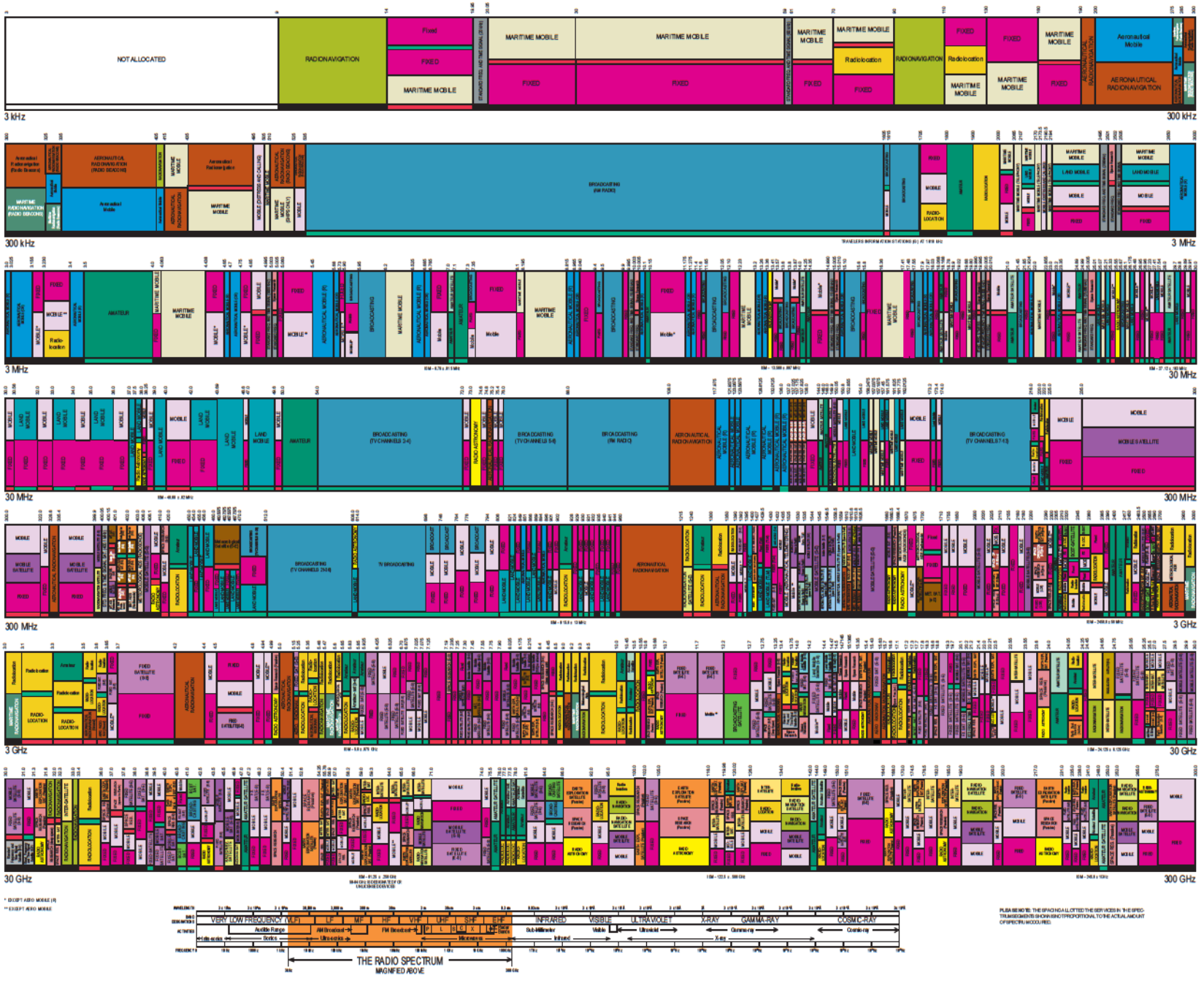
[121] S. Uryasev. *Probabilistic Constrained Optimization: Methodology and Applications*. Vol. 49. Springer Science & Business Media, 2013.

Appendix A

United States Frequency Allocations: The Radio Spectrum

UNITED STATES FREQUENCY ALLOCATIONS

THE RADIO SPECTRUM



RADIO SERVICES COLOR LEGEND

- AERONAUTICAL MOBILE
- AERONAUTICAL MOBILE SATELLITE
- AERONAUTICAL RADIONAVIGATION
- AMATEUR
- AMATEUR SATELLITE
- BROADCASTING
- BROADCASTING SATELLITE
- EARTH EXPLORATION SATELLITE
- FIXED
- FIXED SATELLITE
- INTER-SATELLITE
- LAND MOBILE
- LAND MOBILE SATELLITE
- MARITIME MOBILE
- MARITIME MOBILE SATELLITE
- MARITIME RADIONAVIGATION
- METEOROLOGICAL AIDS
- METEOROLOGICAL SATELLITE
- MOBILE
- MOBILE SATELLITE
- RADIO ASTRONOMY
- RADIODETERMINATION SATELLITE
- RADIOLOCATION
- RADIOLOCATION SATELLITE
- RADIONAVIGATION
- RADIONAVIGATION SATELLITE
- SPACE OPERATION
- SPACE RESEARCH
- STANDARD FREQUENCY AND TIME SIGNAL
- STANDARD FREQUENCY AND TIME SIGNAL SATELLITE

ACTIVITY CODE

- GOVERNMENT EXCLUSIVE
- GOVERNMENT/NON-GOVERNMENT SHARED
- NON-GOVERNMENT EXCLUSIVE

ALLOCATION USAGE DESIGNATION

SERVICE	EXAMPLE	DESCRIPTION
Primary	FIXED	Capital Letters
Secondary	Mobile	1st Capital with lower case letters

This chart is a graphic single-point-in-time portrayal of the Table of Frequency Allocations used by the FCC and NTIA. As such, it does not completely reflect all aspects, i.e., footnotes and recent changes made to the Table of Frequency Allocations. Therefore, for complete information, users should consult the Table to determine the current status of U.S. allocations.

U.S. DEPARTMENT OF COMMERCE
National Telecommunications and Information Administration
Office of Spectrum Management

October 2003

* EXCEPT AERO MOBILE (R)

** EXCEPT AERO MOBILE

PLEASE BE NOTED: THE SPACING ALLOTTED THE SERVICES IN THE SPECTRUM SEGMENTS IS SHOWN IS NOT PROPORTIONAL TO THE ACTUAL AMOUNT OF SPECTRUM ALLOCATED.

Appendix B

Overview of The 3.5 GHz Band

Source:

Wireless Innovation Forum Webinar Series

Webinar Nº 16 – Understanding the New U.S. 3.5 GHz Band

17 June 2015

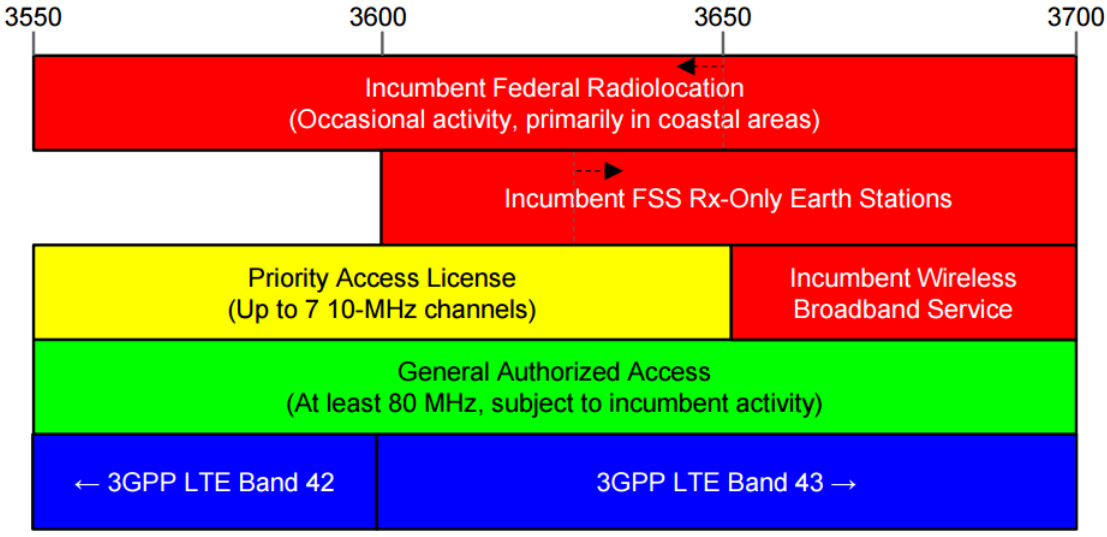


FIGURE B.1: Frequency allocation in the 3.5 GHz band

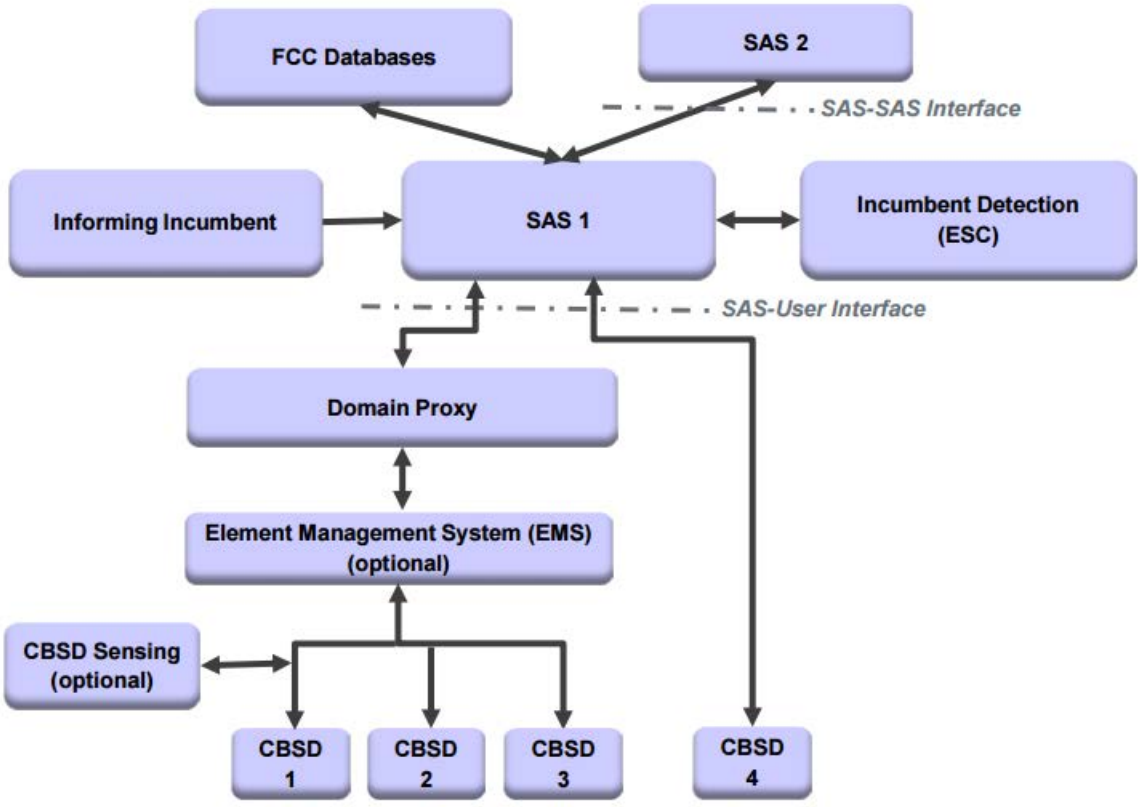


FIGURE B.2: Functional architecture for the 3.5 GHz band

Appendix C

The Channel Collector Problem: Calculation vs. Simulation

TABLE C.1: Comparison table between calculation and simulation for the random assignment scheme ($m = 1$)

$n = 10$ and $l = 9$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	26	26	24	28
2.0	0.0002	26	27	25	29
3.0	0.0027	26	27	25	29
4.0	0.0133	26	27	25	29
5.0	0.0375	27	27	25	28
6.0	0.0751	28	28	26	30
7.0	0.1221	30	29	27	30
8.0	0.1731	31	31	29	33
9.0	0.2243	33	34	31	37
10.0	0.2732	36	35	33	38
11.0	0.3187	38	37	35	40
12.0	0.3604	40	41	38	43
13.0	0.3984	43	44	41	47
14.0	0.4328	45	45	42	49
15.0	0.4639	48	46	43	49

$n = 15$ and $l = 14$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	46	46	43	48
2.0	0.0000	46	46	42	49
3.0	0.0000	46	48	44	51
4.0	0.0001	46	47	44	49
5.0	0.0005	46	46	43	48
6.0	0.0022	46	44	41	46
7.0	0.0071	46	48	45	50
8.0	0.0172	47	48	45	50
9.0	0.0338	48	48	46	51
10.0	0.0568	49	49	46	51
11.0	0.0852	50	48	46	51
12.0	0.1172	52	50	47	53
13.0	0.1512	54	55	52	58
14.0	0.1858	56	59	56	62
15.0	0.2200	59	61	58	64
16.0	0.2531	61	59	56	63
17.0	0.2847	64	62	58	65
18.0	0.3147	67	66	62	70
19.0	0.3429	70	71	67	75
20.0	0.3694	73	74	70	79
21.0	0.3942	76	80	75	85
22.0	0.4174	79	79	74	84

$n = 20$ and $l = 19$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	68	69	65	73
2.0	0.0000	68	69	65	72
3.0	0.0000	68	65	61	68
4.0	0.0000	68	67	64	71
5.0	0.0000	68	71	68	74
6.0	0.0000	68	67	63	70
7.0	0.0001	68	65	62	68
8.0	0.0004	68	66	63	69
9.0	0.0014	68	64	61	67
10.0	0.0037	68	67	63	70
11.0	0.0085	68	68	65	72
12.0	0.0165	69	70	66	74
13.0	0.0284	70	70	66	74
14.0	0.0442	71	74	70	78
15.0	0.0637	72	71	66	75
16.0	0.0861	74	72	69	76
17.0	0.1105	76	74	70	78
18.0	0.1362	79	75	71	79
19.0	0.1625	81	84	79	89
20.0	0.1889	84	86	81	90
21.0	0.2149	86	88	84	93
22.0	0.2403	89	91	86	96
23.0	0.2648	92	92	88	97
24.0	0.2884	95	96	91	102
25.0	0.3109	98	92	87	97
26.0	0.3325	101	103	97	109
27.0	0.3530	105	104	98	110
28.0	0.3725	108	104	99	109
29.0	0.3911	111	109	104	115
30.0	0.4087	115	109	104	114

$n = 25$ and $l = 24$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	91	94	90	99
2.0	0.0000	91	91	86	95
3.0	0.0000	91	96	91	102
4.0	0.0000	91	93	88	97
5.0	0.0000	91	86	82	90
6.0	0.0000	91	89	84	94
7.0	0.0000	91	90	86	95
8.0	0.0000	91	92	87	96
9.0	0.0000	91	88	83	93
10.0	0.0001	91	89	84	93
11.0	0.0003	91	92	87	97
12.0	0.0008	91	88	84	92
13.0	0.0020	91	88	84	93
14.0	0.0043	92	94	89	99
15.0	0.0084	92	90	86	94
16.0	0.0147	92	90	85	94
17.0	0.0236	93	93	88	99
18.0	0.0353	94	97	92	102
19.0	0.0495	96	94	90	99
20.0	0.0661	98	97	92	101
21.0	0.0845	99	103	98	109
22.0	0.1044	102	100	95	105
23.0	0.1252	104	109	104	114
24.0	0.1465	107	104	99	109
25.0	0.1680	109	108	103	114
26.0	0.1894	112	116	111	122
27.0	0.2105	115	122	115	129
28.0	0.2312	118	115	109	122
29.0	0.2514	122	124	117	131
30.0	0.2709	125	127	121	134

$n = 30$ and $l = 29$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	115	116	110	121
2.0	0.0000	115	114	108	120
3.0	0.0000	115	113	107	118
4.0	0.0000	115	112	107	117
5.0	0.0000	115	110	105	115
6.0	0.0000	115	117	111	123
7.0	0.0000	115	114	109	119
8.0	0.0000	115	114	108	119
9.0	0.0000	115	112	106	1184
10.0	0.0000	115	111	106	117
11.0	0.0000	115	117	112	122
12.0	0.0000	115	116	110	122
13.0	0.0001	115	117	112	123
14.0	0.0002	115	118	112	124
15.0	0.0004	115	114	108	120
16.0	0.0011	116	115	109	121
17.0	0.0023	116	113	108	118
18.0	0.0044	116	120	114	126
19.0	0.0078	116	121	114	127
20.0	0.0128	117	120	114	126
21.0	0.0197	118	115	110	121
22.0	0.0286	119	119	114	125
23.0	0.0395	120	121	115	128
24.0	0.0522	122	120	113	127
25.0	0.0666	124	121	115	127
26.0	0.0823	126	127	121	133
27.0	0.0991	128	124	119	129
28.0	0.1166	131	133	126	139
29.0	0.1345	133	136	129	1433
30.0	0.1527	136	136	130	143

TABLE C.2: Comparison table between calculation and simulation for the ordered assignment scheme ($m = 1$)

$n = 10$ and $l = 9$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	125076	122889	106390	139388
2.0	0.0002	1579	1689	1473	1905
3.0	0.0027	208	188	164	211
4.0	0.0133	73	78	70	87
5.0	0.0375	44	44	39	48
6.0	0.0751	35	38	35	42
7.0	0.1221	33	33	31	35
8.0	0.1731	33	33	31	36
9.0	0.2243	34	34	32	37
10.0	0.2732	36	37	34	40
11.0	0.3187	38	41	38	44
12.0	0.3604	41	43	40	46
13.0	0.3984	43	42	39	45
14.0	0.4328	46	47	44	50
15.0	0.4639	48	46	43	49

$n = 15$ and $l = 14$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	1.83303e+10	–	–	–
2.0	0.0000	6.69166e+06	–	–	–
3.0	0.0000	104459	89483	77939	101058
4.0	0.0001	7670	7483	6381	8584
5.0	0.0005	1329	1281	1131	1430
6.0	0.0022	400	393	345	441
7.0	0.0071	178	183	162	203
8.0	0.0172	105	104	94	114
9.0	0.0338	76	76	68	84
10.0	0.0568	64	64	60	69
11.0	0.0852	59	57	53	62
12.0	0.1172	57	57	54	61
13.0	0.1512	57	59	55	62
14.0	0.1858	59	58	54	62
15.0	0.2200	60	60	56	63
16.0	0.2531	62	64	60	68
17.0	0.2847	65	64	60	68
18.0	0.3147	67	66	61	70
19.0	0.3429	70	73	69	77
20.0	0.3694	73	75	71	80
21.0	0.3942	76	78	74	83
22.0	0.4174	79	78	73	83

$n = 20$ and $l = 19$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	1.84244e+16	–	–	–
2.0	0.0000	2.03971e+11	–	–	–
3.0	0.0000	4.03828e+08	–	–	–
4.0	0.0000	6.71685e+06	–	–	–
5.0	0.0000	360128	387427	323065	451790
6.0	0.0000	40651	51596	44764	58428
7.0	0.0001	7703	7873	6906	8840
8.0	0.0004	2143	2107	1825	2389
9.0	0.0014	803	777	700	853
10.0	0.0037	381	386	343	428
11.0	0.0085	220	231	209	253
12.0	0.0165	149	158	143	173
13.0	0.0284	115	123	111	134
14.0	0.0442	97	101	93	108
15.0	0.0637	89	85	79	90
16.0	0.0861	84	85	79	90
17.0	0.1105	83	81	76	87
18.0	0.1362	83	85	80	90
19.0	0.1625	84	87	82	92
20.0	0.1889	86	82	77	87
21.0	0.2149	88	90	85	94
22.0	0.2403	90	88	83	94
23.0	0.2648	93	90	84	96
24.0	0.2884	96	98	92	104
25.0	0.3109	99	108	102	115
26.0	0.3325	102	101	95	106
27.0	0.3530	105	109	103	114
28.0	0.3725	108	117	111	123
29.0	0.3911	111	110	104	117
30.0	0.4087	115	111	105	117

$n = 25$ and $l = 24$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	7.34618e+22	–	–	–
2.0	0.0000	2.50161e+16	–	–	–
3.0	0.0000	6.39993e+12	–	–	–
4.0	0.0000	2.4707e+10	–	–	–
5.0	0.0000	4.23e+08	–	–	–
6.0	0.0000	1.8618e+07	–	–	–
7.0	0.0000	1.57539e+06	–	–	–
8.0	0.0000	215533	246006	211993	280019
9.0	0.0000	42656	42302	37150	47454
10.0	0.0001	11323	11182	9873	12490
11.0	0.0003	3821	3898	3404	4393
12.0	0.0008	1577	1675	1451	1898
13.0	0.0020	772	792	706	878
14.0	0.0043	437	417	380	455
15.0	0.0084	281	285	257	314
16.0	0.0147	202	197	178	216
17.0	0.0236	159	154	143	165
18.0	0.0353	136	142	130	153
19.0	0.0495	122	117	109	125
20.0	0.0661	115	116	108	124
21.0	0.0845	111	114	107	121
22.0	0.1044	110	111	105	118
23.0	0.1252	110	112	105	118
24.0	0.1465	111	114	108	120
25.0	0.1680	113	112	106	118
26.0	0.1894	115	117	111	124
27.0	0.2105	117	112	106	117
28.0	0.2312	120	125	118	132
29.0	0.2514	123	121	114	128
30.0	0.2709	126	125	118	131

$n = 30$ and $l = 29$

	Calculated		Simulated		
ρ	P_B	$\lceil E[X] \rceil$	$E[X]$	$E[X]^-$	$E[X]^+$
1.0	0.0000	8.59431e+29	–	–	–
2.0	0.0000	9.05732e+21	–	–	–
3.0	0.0000	3.01704e+17	–	–	–
4.0	0.0000	2.72843e+14	–	–	–
5.0	0.0000	1.50843e+12	–	–	–
6.0	0.0000	2.62449e+10	–	–	–
7.0	0.0000	1.00851e+09	–	–	–
8.0	0.0000	6.92924e+07	–	–	–
9.0	0.0000	7.42912e+06	–	–	–
10.0	0.0000	1.13277e+06	1125965	969857	1282074
11.0	0.0000	229954	245971	216776	275165
12.0	0.0000	59207	57706	50548	64865
13.0	0.0001	18640	18736	16735	20737
14.0	0.0002	6976	7992	6962	9022
15.0	0.0004	3035	2973	2591	3356
16.0	0.0011	1508	1859	1623	2096
17.0	0.0023	843	850	749	951
18.0	0.0044	524	534	483	584
19.0	0.0078	357	354	319	389
20.0	0.0128	265	283	261	305
21.0	0.0197	211	213	193	232
22.0	0.0286	180	180	165	196
23.0	0.0395	160	166	154	178
24.0	0.0522	149	150	140	160
25.0	0.0666	143	147	137	156
26.0	0.0823	139	145	137	153
27.0	0.0991	138	136	128	144
28.0	0.1166	138	137	129	146
29.0	0.1345	139	137	130	143
30.0	0.1527	140	142	134	150

Appendix D

Résumé de la thèse

D.1 Introduction Générale

D.1.1 Partage du Spectre aux États-Unis

Aux États-Unis, l'administration nationale des télécommunications et de l'information (NTIA) et la commission fédérale de Communications (FCC) ont réaffecté certaines bandes de l'utilisation fédérale à l'utilisation non-fédérale. Toutefois, la réaffectation s'est révélée difficile et coûteuse comme solution pour la gestion future du spectre. Ainsi, lorsque le partage a été proposé comme une alternative, la NTIA et la FCC ont travaillé pour identifier les bandes fédérales potentielles qui peuvent être ouvertes pour un accès partagé entre les utilisateurs fédéraux et les utilisateurs non-fédéraux. Par exemple, la bande de 3.5 GHz (3550–3700 MHz) a été proposée pour le partage entre les opérateurs historiques (radars de la Marine et stations terriennes de service fixe par satellite) et les utilisateurs des services de radiodiffusion à large bande (CBRD). Mais, le partage du spectre ne se limite pas aux États-Unis. Les pays du monde entier sont en train de repenser leur répartition actuelle du spectre. Des efforts conjoints sont en cours pour étudier les modalités de partage afin de maintenir une utilisation efficace du spectre.

L'accès dynamique au spectre (DSA) est une nouvelle approche visant à maximiser l'utilisation des bandes sous-utilisées. Le comité consultatif de la gestion du spectre commercial (CSMAC) a déclaré que le DSA promet d'améliorer l'utilisation du spectre en trois dimensions: la fréquence, la localisation et le temps. Il permet à un réseau secondaire d'utiliser un canal (une fréquence) de façon opportuniste lorsqu'il n'est pas utilisé, et de passer automatiquement à un autre canal lorsqu'un utilisateur primaire ou

un signal principal apparaît sur le canal en question. Le DSA permet ainsi à deux ou plusieurs applications ou réseaux de partager une bande de fréquences.

Le partage du spectre utilise généralement des technologies radio cognitives avancées pour la détection du spectre. Leur flexibilité permet aux utilisateurs (dans les approches de détection du spectre) et aux gestionnaires du spectre (dans les approches basées sur des bases de données) d'identifier et d'utiliser efficacement les espaces blancs (fréquences non-utilisées). Aux États-Unis, on constate que les bandes précédemment contrôlées par les organismes fédéraux et gouvernementaux sont sous-utilisées et mal gérées. En fait, les titulaires ne fonctionnent pas tout le temps, ne fonctionnent que dans certaines régions géographiques et occupent moins de ressources que celles assignées. Le partage du spectre permet également le développement de différentes nouvelles technologies (par exemple, 5G, internet des objets, etc.). Ces technologies sont utilisées dans des centaines de millions d'appareils qui ont besoin d'un meilleur accès au spectre, y compris la connectivité et la qualité de service (QoS). Et comme le nombre de services sans fil augmente, un spectre fixe engendrera des échecs de connexion et d'émission/réception. De plus, le partage du spectre aide à introduire l'harmonisation du spectre. Il s'agit d'un effort mondial pour gérer la répartition des fréquences au-delà des frontières des pays, des futures technologies et des prix du spectre. Lorsque le spectre est mis à disposition par les régulateurs, les investissements à grande échelle permettront une meilleure utilisation du spectre, ce qui apporte des avantages sociaux, politiques et économiques. Une telle gestion permettra des services mobiles abordables et stimulera les efforts d'innovation.

D.1.2 Défis du Partage du Spectre

Pour que le partage du spectre devienne réalité, il faut satisfaire aux exigences opérationnelles de l'utilisateur primaire tout en permettant un accès secondaire au spectre. Une exigence évidente est que l'utilisateur primaire doit être protégé contre les interférences lorsqu'il fonctionne. Les récepteurs ont généralement un seuil d'interférence ou un niveau de dommage. Tout SINR (rapport de signal sur interférence plus bruit) mesuré au-dessus de ce seuil est nuisible aux opérations fédérales et est considéré comme une violation de l'accord de partage.

D'autres considérations existent également. Les utilisateurs primaires avec des missions critiques sont vigilants; ils ne veulent pas laisser les utilisateurs secondaires accéder à leurs bandes. Ainsi, l'un des problèmes de partage du spectre les plus difficiles dans les bandes fédérales est la sécurité. Le système doit protéger le spectre partagé contre deux types d'attaquants. Le premier type inclut les utilisateurs malveillants, qui sont des intrus externes (attaquants exogènes) et attaquent pour perturber les communications primaires et secondaires sans l'intention de maximiser leur profit en termes d'opportunités spectrales. Le second type inclut les utilisateurs égoïstes (appelés aussi gourmands) qui sont internes au réseau (attaquants intra-réseau) et attaquent pour dégrader les performances du réseau partagé avec l'intention d'obtenir un avantage spectral et d'augmenter leurs propres performances. Ces types d'attaquants tentent de briser la confidentialité, l'intégrité et la disponibilité du réseau. Certaines attaques sont plutôt générales, mais toujours valables aux applications du partage du spectre. D'autres attaques sont nouvelles et spécifiques aux systèmes de partage du spectre. Une riche littérature a étudié les questions de sécurité dans le réseautage, et de multiples algorithmes ont été développés. Certains permettent de détecter les violations de la sécurité, d'autres introduisent des contre-mesures pour éviter une menace ou au moins atténuer ses impacts. Dans le contexte du partage du spectre, les exigences en matière de confidentialité (aussi appelées exigences de sécurité opérationnelle) sont aussi importantes. Les paramètres sensibles des utilisateurs primaires comprennent la localisation, la fréquence ou le canal de fonctionnement et le temps d'opération. Par conséquent, les utilisateurs primaires fédéraux doivent s'assurer que leur confidentialité ne sera pas compromise et que leurs paramètres opérationnels ne seront pas exposés. Nous prenons en considération l'importance de telles exigences, en particulier si l'avenir du partage du spectre pour les bandes fédérales en dépend.

D.1.3 Modèle de Menace

Dans notre modèle de menace, nous considérons que le système d'accès au spectre (SAS) est digne de confiance. Cependant, la confiance n'est pas extensible aux utilisateurs secondaires. Ainsi, nous supposons que l'attaquant est un requérant légitime qui a été enregistré dans le système d'accès au spectre et peut avoir accès aux ressources du

spectre. Il peut avoir de l'aide des autres utilisateurs secondaires (attaquants coopératifs) ou falsifier de différentes identités (usurpation d'identité). Il dispose également de ressources informatiques suffisantes pour conduire des inférences et des attaques probabilistes basées sur des informations reçues. En d'autres mots, en utilisant des connaissances non-sensibles reçues telles que la disponibilité des canaux, la puissance d'émission autorisée et le temps d'opération, l'adversaire peut déduire des données sensibles conceranant les utilisateurs primaires (localisation et chemin de déplacement, fréquence et temps opérationnels, etc).

L'attaquant peut utiliser les informations déduites pour mettre en danger les opérations de l'utilisateur primaire (par exemple, brouillage). Prévenir une inférence est un défi, car elle est difficile à détecter et aucune violation explicite n'est faite. Différentes approches ont été proposées. Aucune d'entre eux n'a été largement adoptée, car elles ne répondent pas aux exigences de confidentialité de l'utilisateur primaire et au besoin d'accès de l'utilisateur secondaire. En fait, chaque gain en protection de confidentialité s'accompagne d'une perte d'efficacité du spectre. Par conséquent, non seulement la confidentialité, mais aussi l'efficacité du partage, doivent être prises en considération. En outre, il n'existe pas de solution qui s'adapte à tous les scénarios. Différentes stratégies peuvent être utilisées par un attaquant, et les techniques appropriées doivent être employées afin d'atténuer leurs dégâts.

D.2 Protection de la Fréquence des Utilisateurs Primaires

Les utilisateurs primaires comme les militaires et les utilisateurs de sécurité publique ont besoin d'une protection complète de leurs opérations. Cela inclut la protection de leur confidentialité. La fréquence opérationnelle des utilisateurs primaires est un paramètre sensible et ne doit pas être exposée. Protéger contre sa découverte est essentiel pour atténuer les interférences intentionnelles (par exemple, les attaques de brouillage).

D.2.1 Étude Analytique de Vulnérabilité

Scénario d'Attaque

Le SAS gère n canaux dans une zone donnée. Seuls l canaux sont disponibles pour les utilisateurs secondaires. Nous modélisons l'activité des utilisateurs secondaires à l'aide d'une file d'attente $M/M/l/l$ et supposons que le système est en équilibre. Lorsqu'un utilisateur secondaire demande accès aux ressources du spectre, le SAS répond avec un canal disponible. Si aucun canal n'est disponible, la demande est rejetée.

Nous considérons un scénario d'attaque par inférence simple et intuitive. Soit n le nombre de canaux dans la bande d'intérêt et ICH la liste des canaux opérationnels de l'utilisateur primaire. Comme tout autre utilisateur, l'attaquant envoie des requêtes demandant l'accès au spectre. Sa connaissance initiale est une liste de tous les canaux potentiels de l'utilisateur primaire (c'est-à-dire tous les canaux de la bande). Une fois que le SAS retourne un canal j en réponse à une requête, l'attaquant sait que le canal j n'est pas utilisé par l'utilisateur primaire. Par conséquent, l'attaquant met à jour ses connaissances en supprimant le canal j de la liste ICH. L'attaquant répète cette méthode jusqu'à ce que la taille de ICH devienne finalement 1.

Puisque nous voulons déterminer le nombre attendu de requêtes de canaux qu'un attaquant doit faire pour déduire la fréquence opérationnelle de l'utilisateur primaire, ce problème est équivalent au problème du collecteur de coupons.

Affectation Aléatoire du Canal

Dans le cas de l'affectation aléatoire, le SAS renvoie un canal aléatoirement en réponse à une requête d'un attaquant. La probabilité de blocage P_B est la probabilité que tous les canaux soient occupés au moment de la demande et est calculée comme suit

$$P_B = \frac{\rho^l}{l!}\left(\sum_{k=0}^{l}\frac{\rho^k}{k!}\right)^{-1}, \tag{D.1}$$

où $\rho = \lambda/\mu$ est la charge du système, λ est le taux agrégé d'arrivée, et $1/\mu$ est le temps individuel du service.

Dans le système ci-dessus, lorsque l'attaquant fait une requête, le SAS renvoie l'un des canaux inactifs disponibles ou, si tous les canaux sont occupés, réagit en disant qu'aucun canal n'est disponible. Dans ce cas, nous pouvons exprimer $E[X]$ comme

$$E[X] = E[X_b] + E[X_r]\,, \tag{D.2}$$

où X_b est le nombre de requêtes effectuées lorsque tous les canaux sont occupés, c'est-à-dire que la demande est bloquée, and X_r est le nombre de requêtes pour lesquelles un canal disponible a été renvoyé par le SAS.

Nous savons que

$$E[X_b] = P_B E[X]\,. \tag{D.3}$$

Alors,

$$E[X] = \frac{E[X_r]}{1 - P_B}\,. \tag{D.4}$$

Il ne reste plus qu'à calculer $E[X_r]$. Lorsque les canaux $1 \leq k \leq l$ sont inactifs, chaque canal est inactif avec la probabilité $\frac{k}{l}$, et si il est inactif, il est renvoyé par le SAS avec une probabilité $\frac{1}{k}$. Ainsi, chaque canal a une probabilité d'être renvoyé $\frac{1}{l}$. Le nombre attendu de requêtes $E[X_r]$ peut alors être calculé en utilisant la solution au problème du collecteur de coupons avec des probabilités égales.

$$E[X_r] = l \sum_{k=1}^{l} \frac{1}{k}\,. \tag{D.5}$$

Ainsi, nous avons

$$E[X] = \frac{l H_l}{1 - P_B}\,. \tag{D.6}$$

Affectation Ordonnée du Canal

Dans le cas de l'affectation ordonné, les canaux sont affectés avec probabilités inégales. Nous devons trouver la probabilité p_j que le SAS renvoie le canal j en réponse à la requête d'un attaquant. Cela est équivalent à la probabilité que le canal j soit le plus bas canal disponible au moment de la demande d'un attaquant.

Soit B_j la probabilité qu'une requête arrivée trouve les j premiers canaux occupés. La probabilité conditionnelle qu'une requête arrivée trouvant les premiers $j-1$ channels occupés trouve également le canal j occupé est $\frac{B_j}{B_{j-1}}$.

Si $\gamma_j(z)$ est la transformée de Laplace-Stieltjes de la fonction de distribution du temps écoulé entre les instants successifs où une demande d'arrivée trouve les premiers canaux $j-1$ occupés et est affectée au canal j, on a

$$\gamma_j(\mu) = \frac{B_j}{B_{j-1}}\,, \tag{D.7}$$

où $B_0 = 1$ et $\gamma_j(z)$ sont définis par la relation de récurrence.

$$\gamma_{j+1}(z) = \frac{\gamma_j(z+\mu)}{1-\gamma_j(z)+\gamma_j(z+\mu)}, \quad j = 1, 2, \ldots \tag{D.8}$$
$$\gamma_1(z) = \tfrac{\lambda}{\lambda+z}\,.$$

Notons qu'il résulte de l'équation (D.7), $B_0 = 1$ et

$$B_j = \gamma_1(\mu)\cdots\gamma_j(\mu),\ j = 1, 2, \ldots \tag{D.9}$$

En utilisant ce qui précède, nous calculons les probabilités p_j. Soit I_j une variable aléatoire représentant l'état du canal j, où 1 signifie que le canal est occupé et 0 signifie qu'il est inactif. Alors,

$$\begin{aligned} p_j &= Pr\{I_1 = 1, ..., I_{j-1} = 1, I_j = 0\} \\ &= Pr\{I_j = 0 | I_1 = 1, ..., I_{j-1} = 1\} \ \times \ Pr\{I_1 = 1, ..., I_{j-1} = 1\} \\ &= (1 - Pr\{I_j = 1 | I_1 = 1, ..., I_{j-1} = 1\}) \ \times \ Pr\{I_1 = 1, ..., I_{j-1} = 1\} \\ &= \left(1 - \frac{B_j}{B_{j-1}}\right) B_{j-1} \\ &= B_{j-1} - B_j\,. \end{aligned} \tag{D.10}$$

Notez que $P_B = B_l$ et que

$$P_B \ + \ \sum_{j=0}^{l} p_j \ = \ 1. \tag{D.11}$$

Nous pouvons trouver $E[X]$ en utilisant la probabilité p_j comme calculée ci-dessus dans la solution au problème du collecteur de coupons pour les probabilités inégales.

$$E[X] = \sum_{i=1}^{l} \frac{1}{p_i} - \sum_{i<j} \frac{1}{p_i + p_j} + \sum_{i<j<k} \frac{1}{p_i + p_j + p_k} - \cdots + (-1)^{l+1} \frac{1}{p_1 + \cdots + p_l} . \quad \text{(D.12)}$$

D.2.2 Offuscation et Contre-mesures

Offuscation Inhérente

Le système de répartition des canaux peut apporter une contribution significative à la confidentialité. En regardant les résultats analytiques et les simulations, le mode ordonné est la meilleure approche pour ralentir le processus d'inférence. Les diverses réponses de requêtes dans le mode d'affectation aléatoire du canal facilitent le processus d'inférence, ce qui ne fournit aucune protection pour la fréquence opérationnelle. Bien que le mode d'affectation semi-statique de canaux fonctionne mieux que l'ordonné pour des charges de système plus élevées, il n'est pas immunisé contre les attaques collaboratives.

En plus du mode d'affectation des canaux, le nombre de canaux retournés par requête m est un facteur important pour assurer la confidentialité. Pour tous les cas et sous différents scénarios d'attaque, allouer un seul canal par requête ($m = 1$) est la meilleure option. Une fois $m > 1$, le SAS commence à aider l'adversaire à accélérer le processus d'inférence en fournissant plus d'informations sur l'état du spectre. Cela est plus important dans certains cas que dans d'autres. En fait, lorsque la charge du système est faible ou moyenne, l'adversaire tire grand bénéfice de la diversité au sein de la réponse du SAS. Cependant, si la charge du système est élevée, l'effet de m n'est pas aussi important, puisque les canaux disponibles pour l'utilisation ne couvrent pas la valeur de m ainsi que la demande des utilisateurs secondaires arrivants.

Le SAS peut également définir une limite maximale du nombre de requêtes autorisées par unité de temps pour un utilisateur secondaire (c'est-à-dire le taux de requête d'un utilisateur secondaire). Si on considère que l'utilisateur primaire est opérationnel pendant une période de temps ΔT, le SAS peut limiter le nombre de requêtes d'un utilisateur secondaire λ_s afin qu'il ne permette pas l'inférence de la fréquence opérationnelle. Par exemple, si $\Delta T = 10$ unités de temps et le système est à moitié chargé ($\rho = 50$ %), un

attaquant agressif avec $\lambda_s \geq 7$ peut inférer le canal de l'utilisateur primaire à la fin de la période opérationnelle même si le mode ordonnée est mis en place. Pour tenir compte des attaquants collaboratifs, le SAS doit envisager un λ_s plus strict. Cette limite peut également être ajustée en fonction de la charge du système, du nombre de canaux et du mode d'attribution du canal.

Offuscation Explicite

Le SAS peut mettre en œuvre de l'offuscation explicite en plus de l'offuscation inhérente afin d'assurer la confidentialité des utilisateurs primaires. Dans notre système, l'offuscation peut être obtenue en modifiant la disponibilité de canaux, c'est-à-dire en supprimant des canaux de la liste de canaux inactifs. Par conséquent, certains canaux inactifs sont délibérément laissés vacants. Cela augmenter à la fois l'incertitude de l'attaquant (c'est-à-dire diminuer la probabilité d'inférence) et la probabilité de blocage du système (c'est-à-dire augmenter le rejet d'accès). Les canaux supprimés peuvent être des canaux adjacents, des canaux non adjacents aléatoires ou des canaux choisis en fonction de certains critères de performance. Par exemple, la qualité du canal peut être un paramètre de sélection des canaux obscurcis (par exemple, obscurcir les canaux ayant la QoS la plus faible pour atténuer la perte significative du spectre).

Cette méthode est plus efficace pour augmenter la distance d'inférence (c'est-à-dire l'incertitude de l'attaquant). En gardant intentionnellement certains canaux inactifs, le SAS inclut de l'offuscation dans l'état du spectre, et donc les informations fournies dans chaque requête. En fait, à un certain moment, l'offuscation empêche la mise à jour de la connaissance de l'attaquant, en évitant une estimation précise de la fréquence opérationnelle.

Cette technique de conservation de la confidentialité présente toutefois des inconvénients puisqu'elle ne permet pas un accès secondaire efficace dans certains cas. En cas de charge élevée du système, un plus grand nombre d'utilisateurs secondaires est privé d'accès au spectre, même si certains canaux sont encore disponibles. Bien que le risque d'inférence soit atténué, l'utilisation du spectre est réduite. Lorsque la charge du système est faible ou moyenne, l'offuscation peut être très efficace. Elle augmente

la distance d'inférence sans affecter le partage du spectre, en particulier dans le cas des attaquants agressifs.

Un remède pour atténuer les effets secondaires de l'offuscation est d'augmenter le nombre de canaux en divisant les canaux d'origine en sous-canaux. Cela permet au SAS de bloquer plus de canaux (c'est-à-dire, d'augmenter l'offuscation) sans compromettre l'accès au spectre. En fait, la technologie utilisée par la plupart des utilisateurs secondaires dans les bandes partagées (3.5 GHz) nécessite de bandes de fréquences réduites (1 MHz – 2 MHz). Par conséquent, l'offuscation peut augmenter l'incertitude d'un attaquant avec une probabilité de blocage minimale. Il s'agit d'un bon compromis entre la confidentialité de l'utilisateur primaire et l'efficacité du partage entre utilisateurs secondaires.

D.3 Protection de la Localisation des Utilisateurs Primaires

Protéger la fréquence opérationnelle de l'utilisateur primaire est une première étape pour le protéger contre l'exposition. Nous avons montré qu'en ajustant simplement certains paramètres du spectre ou en incluant des caractéristiques mineures d'offuscation, la confidentialité peut considérablement augmenter. Mais, cela n'est pas sufisant. La localisation de l'utilisateur primaire doit également être protégée.

D.3.1 Scénario d'Attaque

Lorsque l'utilisateur primaire est opérationnel, toute la zone d'opération est activée (c'est-à-dire qu'aucun utilisateur secondaire ne peut fonctionner à l'intérieur de cette zone). Lorsqu'un utilisateur secondaire demande l'accès au spectre, le système d'accès au spectre décidera en fonction de son localisation. Si l'emplacement de l'utilisateur secondaire se trouve dans la zone opérationnelle, l'utilisation de la fréquence de l'utilisateur primaire sera refusée, car cela peut lui causer des interférences. Toutefois, lorsqu'il se trouve en dehors de la zone opérationnelle, l'utilisation de la fréquence de l'utilisateur primaire est accordée.

Nous supposons que l'attaquant connaît déjà la fréquence opérationnelle de l'utilisateur primaire, soit en utilisant l'algorithme présenté auparavant, soit en utilisant une différente approche. Lorsqu'une zone est activée, l'utilisateur primaire est opérationnel. Afin de trouver la localisation de l'utilisateur primaire, l'attaquant essaye d'inférer les limites de la zone opérationnelle en suivant une ligne droite. Le scénario d'attaque sera exécuté par deux attaquants: un attaquant remontant la ligne et un autre attaquant descendant la ligne. On sait qu'aucun utilisateur secondaire n'est autorisé d'utiliser le canal occupé par l'utilisateur primaire dans la zone opérationnelle. C'est pourquoi, les deux attaquants demandent accès au canal opérationnel de l'utilisateur primaire. Chaque fois qu'ils obtiennent un rejet, ils continuent à se déplacer. Une fois que l'accès est accordé, les attaquants savent qu'ils sont, à ce moment, hors de la zone opérationnelle.

Les attaquants se déplacent soit vers le haut, soit vers le bas suivant un chemin spécifique de mouvement. Nous considérons deux algorithmes:

- algorithme d'inférence à pas fixe: le système d'attaque déploie un chemin statique de déplacement entre une requête et une autre.
- algorithme d'inférence à pas adaptif: le système d'attaque déploie un chemin de déplacement adaptatif entre une requête et une autre en doublant le pas si la limite n'est pas croisée ou en divisant le pas par deux si la limite est croisée (c'est-à-dire exécuter une recherche binaire).

Intuitivement, l'algorithme d'inférence à pas adaptatif permet à l'attaquant une meilleure performance car il permet d'acquérir plus de connaissances en moins de requêtes. Lorsqu'aucune technique de préservation de la confidentialité n'est mise en œuvre, le système est évidemment plus vulnérable aux attaques d'inférence. Le système d'attaque gagne des connaissances avec chaque requête supplémentaire. À la fin de l'attaque d'inférence, le système d'attaque est capable d'estimer les limites de la zone opérationnelle suivant la valeur du pas.

Nous notons que l"algorithme à pas adaptatif donne le meilleur coût pour l'attaquant, car il réduit à peu près 80 % le nombre de requêtes nécessaires en le comparant à l'algorithme à pas fixe. Cependant, nous remarquons également que la précision d'un adversaire qui utilise l'algorithme à pas fixe est plus élevée que celle d'un adversaire qui

utilise l'algorithme à pas adaptif pour des valeurs plus élevées du pas. En effet, le système d'attaque qui utilise l'algorithme à pas adaptif perd de la précision lors de l'utilisation de valeurs supérieures du pas.

Le compromis entre coût et précision est important lorsque l'adversaire et le système d'accès au spectre visent à optimiser leurs algorithmes. À la fin du processus d'inférence, dans l'algorithme à pas fixe, l'incertitude de l'adversaire est presque nulle, pour les petites valeurs du pas. Cependant, le coût de cette approche (c'est-à-dire le nombre de requêtes) est élevé. L'adversaire peut sacrifier la précision du coût en employant l'algorithme à pas adaptif, c'est-à-dire, pour moins de requêtes, un adversaire est capable d'inférer plus de connaissances, offrant ainsi une meilleure performance d'attaque.

D.3.2 Offuscation: Bruit Additif

L'offuscation est largement utilisée dans la littérature afin d'inclure une incertitude dans la connaissance de l'attaquant. Elle est appliquée sur la base de données elle-même. Dans ce cas, nous déployons du bruit additif pour masquer les limites de la zone opérationnelle. En d'autres termes, les limites de la zone sont étendues afin de tromper l'adversaire et altérer ses connaissances. Par conséquent, les limites de la zone opérationnelle seront remplacées dans la base de données par des limites corrompues (c'est-à-dire obscurcies). Ainsi, si un utilisateur secondaire se trouve dans la zone obscurcie et demande accès à la fréquence de l'utilisateur primaire, il sera rejeté, même si aucune interférence n'est envisagée. L'offuscation de la zone opérationnelle peut être appliquée comme suit:

$$Climite^{-} = limite^{-} - \frac{bruit}{2} \tag{D.13}$$

$$Climite^{+} = limite^{+} + \frac{bruit}{2} \tag{D.14}$$

où $limite^{-}$ et $limite^{+}$ sont les limites originaux de la zone opérationnelle, $Climite^{-}$ et $Climite^{+}$ sont les limites corrompues de la zone opérationnelle, $bruit$ est le bruit additif ajouté aux limites de la zone opérationnelle.

En mettant en œuvre un bruit additif, nous notons une amélioration de la confidentialité de l'utilisateur primaire. La précision de l'adversaire est affectée par les limites corrompues de la zone opérationnelle. Pour l'algorithme à pas fixe, nous constatons que la diminution de la confidentialité est progressive. Néanmoins, la connaissance de l'attaquant atteint un maximum. Nous notons également que l'augmentation de la valeur du pas peut réduire considérablement la confidentialité en réduisant le coût de l'attaque. Pour l'algorithme à pas adaptif, nous notons une tendance différente. La diminution de la confidentialité est dramatique au début, puis atteint un maximum et se stabilise. De plus, l'augmentation de la valeur du pas diminue légèrement le coût de l'attaque. Dans l'ensemble, cet algorithme montre une meilleure performance pour l'attaquant. Cependant, nous concluons que peu importe l'intelligence de l'attaquant, cette technique de préservation de confidentialité peut l'arrêter.

L'avantage d'un tel algorithme est qu'il ne dépend pas de l'utilisateur secondaire. Cependant, un inconvénient majeur est qu'il a un impact non seulement sur les utilisateurs secondaires malveillants mais aussi sur les utilisateurs secondaires inoffensifs en rejetant leur accès en dehors de la zone opérationnelle, ce qui affectera l'efficacité du spectre. En d'autres termes, certaines ressources spectrales disponibles sont retenues par le SAS et donc intentionnellement inutilisées par le réseau secondaire. Pourtant, le SAS peut utiliser l'offuscation lorsque la charge du spectre est faible et la perte de ressources n'affecte pas les utilisateurs secondaires déjà actifs.

D.3.3 Gestion de Confiance

La métrique de confiance est attribuée à chaque utilisateur secondaire enregistré dans le système d'accès au spectre. Une telle approche applique une condition sur le mécanisme de libération de l'information, c'est-à-dire une condition sur la réponse à chaque requête. Ainsi, le système doit conserver une base de données de tous les utilisateurs secondaires et évalue leur fiabilité en mettant à jour la métrique de confiance lors de chaque demande. Généralement, un tier peut établir une observation en évaluant les comportements des utilisateurs et/ou en utilisant des observations indirectes en s'appuyant sur les recommandations d'autres utilisateurs. Dans notre cas, les secondaires ont une relation directe avec le SAS et aucune relation avec les autres secondaires. Ainsi, l'évaluation

de confiance dépend de la demande de l'utilisateur secondaire et permet d'évaluer sa malveillance. Un utilisateur secondaire est considéré comme non fiable et peut être résilié si sa confiance franchit le seuil thr_T.

Deux seuils de distance (thr_{D^-} et thr_{D^+}) sont inclus pour permettre une certaine souplesse. Lorsqu'un utilisateur secondaire demande l'accès au canal de l'utilisateur primaire et se trouve à thr_{D^+} d'une des limites de la zone opérationnelle, il est considéré comme une menace pour la confidentialité de l'utilisateur primaire. Par conséquent, sa valeur de confiance est abaissée. Mais lorsqu'il demande l'accès au canal de l'utilisateur primaire et se trouve à thr_{D^-} d'une des limites de la zone opérationnelle, la menace est plus grave. Ainsi, la valeur de la confiance est doublement abaissée. Par conséquent, le SAS évalue un utilisateur secondaire i lorsque ce dernier envoie une requête q en mettant à jour sa valeur de confiance $T_i(q)$ comme suit:

$$T_i(q) = \begin{cases} T_i(q - \Delta q) & \text{si } D > thr_{D^+} \\ T_i(q - \Delta q) - \tau & \text{si } thr_{D^-} < D \leq thr_{D^+} \\ T_i(q - \Delta q) - 2\tau & \text{si } D \leq thr_{D^-} \end{cases} \tag{D.15}$$

où q est le numéro de la requête, Δq est l'intervalle de mise à jour de confiance, τ est un facteur de mise à jour, D est la distance de l'attaquant par rapport aux limites la zone opérationnelle, and thr_{D^-} and thr_{D^+} sont les seuils de distance définis par le système pour évaluer une requête.

Nous supposons que, dans ce cas, la confiance ne peut que s'aggraver pour éviter l'utilisation malveillante de cette métrique. Si la valeur de confiance peut être améliorée et qu'un attaquant en est conscient, il essaiera de garder sa valeur de confiance en dessous du seuil prédéfini.

La mise en œuvre de la confiance requiert la définition des seuils pour la distance et la confiance. Une valeur de confiance T signifie que l'individu est T % digne de confiance. Lorsque la valeur du seuil de confiance est élevée, le nombre de requêtes nécessaires pour déduire les mêmes connaissances est plus élevé. Après un certain nombre de requêtes, les connaissances de l'adversaire atteignent un maximum et ne sont plus mises à jour. Certaines valeurs de confiance sont raisonnables et peuvent être mises

en œuvre sans adopter une prudence exagérée ($thr_T = 100$ %) ou permettre une liberté extrême ($thr_T = 0$ %). En fait, lorsque le seuil de confiance thr_T est égal à 100 %, chaque déplacement de l'utilisateur secondaire est considéré comme une menace, ce qui entraîne sa résiliation. Lorsque le seuil de confiance thr_T est égal à 0 %, le système prend plus de temps pour identifier un utilisateur secondaire malveillant. Aucune des deux approches n'est recommandée, car elles fournissent le pire compromis: l'une suppose que tous les utilisateurs secondaires sont dignes de confiance, l'autre suppose qu'ils ne le sont pas.

Comme attendu, nous constatons que des seuils de confiance plus élevés permettent une meilleure protection de l'utilisateur primaire. L'algorithme à pas fixe présente l'incertitude la plus élevée alors que l'algorithme à pas adaptatif présente l'incertitude la plus faible. Si l'adversaire utilise l'algorithme adaptatif, les valeurs d'incertitude semblent statiques puis augmentent rapidement pour les dernières valeurs du seuil de confiance. Si l'adversaire utilise l'algorithme à pas fixe, l'incertitude continue à augmenter avec l'augmentation de ces valeurs.

D.4 Framework Général pour la Protection de la Confidentialité

La mise en œuvre d'algorithmes de préservation de confidentialité peut être efficace, mais insuffisante pour évaluer et gérer une attaque ciblant les utilisateurs primaires des systèmes de partage du spectre.

D.4.1 Vue d'Ensemble du Framework

Identifier le risque et réagir de façon appropriée est essentiel pour protéger l'utilisateur primaire contre les attaques par inférence. Nous proposons un modèle à cinq étapes intègrant la surveillance, l'identification, l'évaluation, l'analyse et la gestion des risques pour assurer une protection complète de l'utilisateur primaire.

Le gestionnaire du spectre s'engage dans la surveillance des risques dans le cadre de ses activités normales. Généralement, cela consiste à chercher des activités anormales dans les demandes du réseau secondaire. Une action suspecte ou une menace déclenche l'étape suivante, l'identification du risque. Ici, la probabilité d'une attaque par inférence

et son impact sont calculés. Des facteurs tels que la charge du système (activité des utilisateurs secondaires) ou les facteurs de risque locaux associés à une région géographique particulière peuvent être utilisés en plus des détails de la menace observée. La probabilité et l'impact sont ensuite utilisés comme intrants pour l'étape d'évaluation du risque, qui détermine le niveau de risque du système. L'étape d'analyse des risques utilise le niveau de risque pour sélectionner des mesures appropriées pour la protection de la confidentialité. Dans l'étape de la gestion des risques, les mesures de confidentialité et la disponibilité du spectre sont utilisées pour déterminer les compromis entre la protection de la confidentialité et l'utilisation efficace du spectre. Les mesures de protection qui en résultent sont appliquées et le gestionnaire du système entre à nouveau dans l'étape de surveillance des risques.

D.4.2 Architecture Proposée

Les modèles de conservation de la confidentialité ont été étudiés dans le contexte de l'exploration de données. Une architecture commune comprend:

- les propriétaires de données qui possèdent des données sensibles et cherchent la protection de leur confidentialité,
- les éditeurs de données qui recueillent les données des propriétaires et les préparent pour publication,
- les destinataires de données qui reçoivent des informations des éditeurs de données pour mener des opérations d'exploration de données.

Traditionnellement, l'offuscation est appliquée au niveau de l'éditeur de données. Nous pouvons utiliser une architecture similaire pour des systèmes de partage centralement coordonnés. Les utilisateurs secondaires correspondent aux destinataires des données. Le gestionnaire de spectre correspond à l'éditeur de données. Nous avons ajouté une fonction dans le gestionnaire de spectre fournissant des données obscurcies extraites des données originales. Les utilisateurs primaires correspondent aux propriétaires de données. Dans certains systèmes de partage du spectre, il n'y a pas de communication directe entre le gestionnaire du spectre et les utilisateurs primaires, et les données sont collectées par des capteurs et analysées par un système de décision.

D.4.3 Techniques Proposées

Offuscation de la Matrice de Disponibilité

Le système de décision calcule une matrice d'occupation et transmet ces calculs au gestionnaire de spectre. La matrice d'occupation est une matrice tridimensionnelle (localisation, fréquence, temps): la i^{eme} rangée matricielle fait référence à l'i^{eme} canal, la j^{eme} colonne matricielle fait référence au j^{eme} capteur et la k^{eme} colonne matricielle fait référence au k^{eme} intervalle de temps. Les dimensions de la matrice sont n, p, q. Ceci donne une représentation visuelle de l'occupation spatio-temporelle des canaux. Elle montre où dans le temps et l'espace un canal est occupé. Elle simplifie également la compréhension des interactions entre les différentes entités de l'architecture de conservation de la confidentialité.

Dans l'exploration de données, l'offuscation est obtenue en généralisant et/ou en supprimant des parties des données afin qu'aucun individu ne puisse être distingué de manière unique. Dans le partage du spectre, l'offuscation est appliqué à la localisation, à la fréquence et au temps de fonctionnement de l'utilisateur primaire.

Nous pouvons appliquer l'offuscation à la localisation en créant une plus grande zone d'opération pour l'utilisateur primaire et en injectant de l'ambiguïté dans la localisation opérationelle estimée par l'attaquant.

Dans ce cas, nous pouvons aussi appliquer l'offuscation à la fréquence. L'offuscation de la fréquence peut être obtenue en supprimant des canaux déjà disponibles dans une zone bien déterminée. Les canaux obscurcis peuvent être des canaux adjacents ou des canaux non adjacents aléatoires. En outre, l'offuscation de fréquence peut être appliquée dans une ou plusieurs zones.

L'offuscation du temps peut être faite en combinant les intervalles de temps successifs. Cela est possible dans un système où les utilisateurs primaires informent préalablement le gestionnaire du spectre de leurs opérations. Le gestionnaire du spectre ne peut pas considérer cette technique lorsque l'allocation de ressources est basée sur la détection de l'utilisateur primaire en temps réel. Cependant, le gestionnaire de spectre peut appliquer une autre technique en prolongeant le temps d'opération d'un utilisateur primaire même si ce dernier cesse de fonctionner. Le gestionnaire de spectre peut également utiliser

l'offuscation du temps en simulant une activité primaire en cours même si aucune n'est présente.

Offuscation de la Réponse du Système

L'offuscation de la réponse empêche certains paramètres d'apparaître plus fréquemment que d'autres en réponse à une requête. Dans le cadre du partage du spectre, le gestionnaire du spectre renvoie une liste de m disponibilités. Nous appliquons l'offuscation en fournissant moins de disponibilités, c'est-à-dire qu'au lieu de renvoyer m canaux disponibles pour utilisation, le gestionnaire de spectre retourne seulement $m' < m$ canaux disponibles pour utilisation. Par exemple, un gestionnaire de spectre, qui renvoie trois canaux par requête auparavant, retourne un seul canal.

Les modes d'affectation des canaux peuvent également être utilisés pour appliquer l'offuscation à la réponse. Par exemple, un gestionnaire de spectre peut affecter le même canal au même secondaire sur des requêtes successives, si disponible. Un autre mode efficace est de toujours affecter le canal le plus bas disponible.

Le gestionnaire de spectre peut également limiter le taux de requêtes d'un utilisateur à un seuil maximal. Par conséquent, le coût d'une attaque augmentera. Alors qu'un attaquant peut encore être en mesure d'inférer les canaux de l'utilisateur primaire, il lui faudra plus de temps pour réussir.

Compromis entre Confidentialité et Disponibilité du Spectre

L'application de l'offuscation entraîne une perte de ressources spectrales, nous devons donc mesurer le compromis entre la disponibilité du spectre et la protection de la confidentialité. Cependant, avoir une seule métrique pour le compromis, tel que le rapport des deux, est de valeur limitée dans la conception et le fonctionnement d'un système de partage du spectre. Il est plus utile de le formuler comme un problème d'optimisation sous contrainte. Dans l'optimisation sous contrainte, une fonction-objectif est maximisée ou minimisée sous réserve d'un ensemble de contraintes sur ses variables ou sur des grandeurs liées d'une manière ou d'une autre à ses variables. Les contraintes peuvent être des contraintes "hard" ou des contraintes "soft". Une contrainte "hard" est

celle qui est nécessaire pour tenir; une contrainte "soft" est celle qui n'est pas. Les contraintes "soft" sont habituellement assignées une pénalité ou un poids. Les contraintes peuvent également être classées en priorité, où des contraintes de priorité plus élevée sont satisfaites avant que celles de priorité inférieure ne soient prises en considération. L'optimisation sous contrainte est un champ large et bien étudié. Il existe de nombreuses façons de formuler ses problèmes et une grande variété de techniques et de stratégies pour trouver ses solutions.

En utilisant la notation et les métriques développées dans cette thèse, nous pouvons formuler des critères de performance de la confidentialité et de l'utilisation du spectre de manière utile et quantitative. Par exemple, si nous voulons maximiser la confidentialité totale des utilisateurs primaires, tout en exigeant que la probabilité de transmission d'un utilisateur secondaire soit supérieure à 95 % dans toutes les zones, nous pouvons l'exprimer comme suit:

$$\begin{aligned} \max \quad & D_H(A', A) \\ \text{avec} \quad & P_b(j) < 0.05 \quad \text{pour } j = 1, ..., p \end{aligned} \tag{D.16}$$

où $D_H(\cdot)$ est la distance de Hamming entre deux matrices et $P_b(j)$ est la probabilité de blocage à la zone j. Dans ce cas d'optimisation sous contrainte, puisque les valeurs $P_b(j)$ dépendent de la charge du système à la zone j ainsi que du niveau d'offuscation, les contraintes doivent être soft.

Des exigences plus complexes et spécifiques peuvent être également exprimées. Par exemple, supposons que nous voulons maximiser la confidentialité totale tout en veillant à ce que la fréquence à la zone j soit occultée en rendant deux canaux supplémentaires indisponibles à cause d'un risque élevé à cette zone, avec une probabilité de transmission diminuée de pas plus que 4 % dans chaque zone. Nous avons ici une exigence de confidentialité (contrainte) qui est locale à une zone et une contrainte soft sur l'impact que

l'offuscation ajoutée sur la disponibilité du spectre. Cela peut s'exprimer comme suit:

$$\begin{aligned} \max \quad & D_H(A', A) \\ \text{avec} \quad & W_H(z'(j)) \geq 3 \quad \text{pour } j = 1, ..., p \\ & \Delta P_b(j) < 0.04 \quad \text{pour } j = 1, ..., p \end{aligned} \tag{D.17}$$

où $W_H(\cdot)$ est le poids de Hamming d'un vecteur, et $z'(j)$ est la j^{eme} colonne dans la matrice de disponibilté modifiée et représente donc la zone modifiée (obscurcie) j. $P_b(j)$ est la différence dans la probabilité de blocage dans la zone j avant et après offuscation. $D_H(\cdot)$ est telle que définie ci-dessus.

D.5 Conclusions et Perspectives

D.5.1 Conclusions

Compte tenu de la croissance explosive du trafic mobile haut débit, le monde s'est engagé dans des approches novatrices de répartition des fréquences au lieu du spectre statique exclusif et à coût élevé. Étant donné que le spectre attribué aux organismes gouvernementaux s'est révélé sous-utilisé, les organismes de réglementation ont examiné l'option de partage. Cependant, la sensibilité et la mobilité limitée des utilisateurs primaires ainsi que le risque d'attaques incitent à la protection de leur confidentialité. En particulier, la localisation, la fréquence et le temps de fonctionnement de l'utilisateur primaire ne doivent pas être révélés ou même inférés. En fait, les utilisateurs secondaires peuvent exploiter l'environnement partagé et obtenir un accès non autorisé à des informations sensibles à l'aide de requêtes légitimes. Par conséquent, un adversaire peut être capable d'inférer les paramètres opérationnels de l'utilisateur primaire en combinant les connaissances initiales avec les connaissances acquises. Ceci est connu comme une attaque d'inférence. Une telle attaque est difficile à détecter. Dans cette thèse, nous nous intéressons principalement à la protection de deux paramètres opérationnels: la fréquence et la localisation.

Dans un premier temps, nous avons examiné les exigences des utilisateurs primaires fédéraux dans le partage du spectre. Ensuite, nous avons évalué la confidentialité de la

fréquence de l'utilisateur primaire à des attaques par inférence et nous proposons des techniques d'offuscation inhérente et explicite pour assurer sa protection. Aussi, nous avons évalué la confidentialité de la localisation de l'utilisateur primaire à des attaques par inférence et nous proposons des techniques d'offuscation et de fiabilité pour assurer sa protection. Enfin, nous avons proposé une architecture générique de préservation de la confidentialité, et un modèle fondé sur les riques pour protéger les utilisateurs primaires et garantir un accès efficace au spectre par les utilisateurs secondaires.

Ce travail ajoute une flexibilité pour faire face à de telles menaces, réduit le temps de calcul, économise de l'énergie et fournit une meilleure protection en ralentissant le processus d'inférence ou en rendant l'adversaire incertain concernant l'information inférée. Le gestionnaire de spectre évalue une menace, sélectionne les pertes de spectre tolérées en fonction de la charge du système (c'est-à-dire le nombre des utilisateurs secondaires) et définit les connaissances maximales qui peuvent être acquises par l'attaquant. Ensuite, il choisit et met en œuvre la meilleure méthodologie pour se protéger contre cette menace. Par exemple, si la charge du système est faible, il est plus facile et plus efficace de mettre en œuvre des méchanismes de perturbation au lieu des méchanismes de fiabilité. En fait, la perte de spectre n'aura pas d'effet sur les autres utilisateurs secondaires, car il y aura suffisamment de ressources spectrales pour accommoder chaque individu. En outre, le gestionnaire de spectre n'a pas besoin de sauvegarder les requêtes de chaque utilisateur secondaire, ce qui évite un effort de calcul supplémentaire. Le gestionnaire du spectre est donc en mesure d'identifier la meilleure approche à mettre en œuvre en fonction du risque et son niveau. Cela permet de protéger les paramètres opérationnels des utilisateurs primaires et de garantir un partage efficace du spectre.

D.5.2 Perspectives

Cette thèse a principalement porté sur la protection de la sécurité opérationnelle des opérations fédérales dans un spectre partagé. Dans la prochaine décennie, d'autres bandes seront ouvertes pour le partage. Ainsi, de nombreuses possibilités pour étendre ce travail existent. Il y a un certain nombre de défis liés aux modèles proposés dans cette thèse. Le temps de fonctionnement d'un utilisateur primaire n'a pas été explicitement

protégé contre les attaques par inférence. Tout en protégeant la fréquence et la localisation opérationnelles, le système est capable de fournir une certaine protection contre l'inférence du temps opérationnel. Cependant, des approches plus simples de préservation de la confidentialité doivent être appliquées. En outre, suite à l'absence d'un système réel fonctionnant dans un spectre partagé, notre évaluation est fondée sur des simulations. Une fois que les premiers systèmes d'accès au spectre et les capteurs sont approuvés et certifiés, nous envisageons la mise en place d'un évaluateur en temps réel pour mesurer la vulnérabilité des utilisateurs primaires aux attaques par inférence. Les techniques proposées de préservation de la confidentialité dans cette thèse seront intégrées dans un système réel et leur résilience et durabilité seront mises à l'épreuve. Il est essentiel d'examiner l'efficacité de notre conception et de nous assurer qu'elle correspond aux résultats de la simulation.

Les travaux futurs comprennent également la réglementation du spectre. Afin de maintenir un environnement partagé et sécurisé, certains mécanismes doivent être mis en œuvre. Ils peuvent être préventifs ou punitifs. Nous avons déjà discuté dans cette thèse les mécanismes préventifs, tels que les techniques de conservation de la confidentialité (par exemple, offuscation inhérente et explicite). Ces techniques peuvent ralentir le process d'inférence ou injecter de l'incertitude dans les connaissances de l'attaquant. Toutefois, elles ne garantissent pas une protection totale. Par conséquent, des mécanismes punitifs sont nécessaires pour contrôler le système. Nous avons déjà introduit une méthode punitive en attribuant une métrique de réputation (confiance) aux utilisateurs secondaires et en éliminant les utilisateurs secondaires non fiables. D'autres solutions pour identifier, localiser et punir un adversaire peuvent être envisagées, et les sanctions peuvent être réglementaires ou économiques, voire même légales.