A Framework for Intelligent Crowdsourced Enforcement of Access Rights in Shared Spectrum Networks

by

Debarun Das

Submitted to the Graduate Faculty of

the Department of Computer Science in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF COMPUTER SCIENCE

This dissertation was presented

by

Debarun Das

It was defended on

April 13 2024

and approved by

Dr. Taieb Znati, Department of Computer Science

Dr. Daniel Mosse, Department of Computer Science

Dr. Youtao Zhang, Department of Computer Science

Dr. Martin Weiss, Department of Informatics and Networked Systems

A Framework for Intelligent Crowdsourced Enforcement of Access Rights in Shared Spectrum Networks

Debarun Das, PhD

University of Pittsburgh, 2024

Traditional spectrum allocation policy statically grants spectrum bands to licensed primary users for exclusive access. Such a policy prevents non-primary users from accessing these spectrum bands, even when they are idle. As a result, licensed frequency bands remain underutilized for extended periods of time. The exponential increase in utilization of wireless services, however, has led to a growing demand for spectrum use. The need to address spectrum scarcity in the public domain has spurred the exploration of spectrum-sharing strategies to optimize spectrum utilization. In response, the Federal Communications Commission (FCC) announced the establishment of the Citizens Broadband Radio Service (CBRS) to facilitate shared federal and non-federal use of the 3550-3700 MHz band, which allows unlicensed secondary users to opportunistically access licensed spectrum bands when they are idle.

While spectrum sharing improves spectrum utilization, it also introduces the risk of illegitimate access to licensed spectrum. This gave rise to the need for effective access rights enforcement in shared spectrum networks. It is to be noted that timing is paramount to the applicability of spectrum enforcement, depending on whether it is applied either before or after a potentially harmful action has occurred. The former enforcement is referred to as *ex-ante*, while the latter is referred to as *ex-post* [1,2].

The focus of this dissertation is on ex-post spectrum access rights enforcement, an important component of the CBRS framework. To achieve effective ex-post enforcement, two fundamental requirements must be addressed: (i) coverage of the area of enforcement and (ii) accurate and robust detection of spectrum access violations. This dissertation develops a framework and related protocols to address the ex-post spectrum enforcement problem while fulfilling the two requirements.

The main contributions of the dissertation are the development of (i) a shared spectrum enforcement architecture, focused on a volunteer-based crowdsourced approach to achieve cost-effective and scalable spectrum monitoring and misuse detection; (ii) a methodology, that harnesses variants of the Secretary and Stable Matching algorithms, for effective selection of spectrum monitoring volunteers to ensure successful and comprehensive enforcement of spectrum access rights across the spectrum enforcement area. The selected volunteers are assigned to coverage regions based on their qualifications, as reflected by their reputation and the likelihood of their availability in a coverage region for an extended monitoring time interval; iii) a novel spectrum sampling scheme to enable accurate and robust detection of access violations, which takes into consideration the dynamically changing aspects of the volunteers' monitoring capabilities and behaviors and the intruders' misuse strategies; iv) a Machine Learning-based framework to predict volunteers' future locations and a method to estimate the sojourn time of a volunteer within a specific region.

An extensive analysis and assessment of the proposed framework's main components demonstrate the viability and effectiveness of the methodologies used to achieve ex-post enforcement of spectrum access rights, in a scalable and cost-effective manner.

Table of Contents

List of Tables

List of Figures

Preface

This Ph.D. dissertation was conducted at the Department of Computer Science, University of Pittsburgh. This research would not have been possible without the assistance and support of many individuals.

I am deeply grateful to my advisor, Dr. Taieb Znati, for his unwavering support, invaluable advice, and constructive criticism. His enthusiastic approach to my research problem has been a constant source of motivation throughout my graduate studies. I am forever thankful to him for his contributions to my life and career.

I want to thank all the faculty members and staff of the Department of Computer Science at the University of Pittsburgh. Their constant support has made the journey to a Ph.D. much smoother. I would like to offer special thanks to my committee chair, Dr. Daniel Mosse, to whom I could always reach out and seek advice on research and anything else related to academia. I am highly grateful to my other committee members — Dr. Martin Weiss and Dr. Youtao Zhang, for their consistent support and valuable feedback during my Ph.D. work. I would like to express my deepest thanks to Keena Walker, who has been like a guardian to an international student like me. She has always helped me with all sorts of queries and doubts during my entire graduate study.

I am sincerely grateful to all my friends. Their constant help and support have been invaluable throughout this journey. Finally, I wish to express my deepest gratitude and affection to my family. They have always encouraged me to pursue my dreams and believed in my decisions, even when it meant moving to the other side of the world for graduate school.

1.0 Introduction

The exponential increase in utilization of wireless services has led to a growing demand for more spectrum. To address spectrum scarcity, there is a need to implement spectrum sharing. Spectrum sharing involves the rearrangement of spectrum access rights among various stakeholders. This introduces the risk of access rights violations, which necessitates the need for spectrum monitoring. A robust spectrum access rights enforcement system is crucial to give substance to access rights and mitigate their potential violations. As spectrum sharing becomes more pervasive, there is a growing need to adopt a scalable approach to automate spectrum access rights enforcement. My dissertation focuses on the advancement of ex-post enforcement of spectrum access rights. The vulnerabilities in shared-spectrum networks are discussed, leading to the dissertation's focus on a novel spectrum enforcement approach. This approach involves a monitoring framework focused on a scalable physical infrastructure and supporting algorithms to enhance interference prevention and address various spectrum access violations.

This chapter discusses the basic premise and motivation of my research, highlighting the need for spectrum sharing and the existing methodologies to enforce it. The primary objective of the thesis is introduced, with the aim of laying the foundation for an ex-post enforcement of spectrum access rights system, enabled by crowdsourcing. This approach leverages the collective intelligence of crowdsourced agents to enhance spectrum monitoring. The primary challenges, including expansive coverage, effective volunteer selection, and robust spectrum monitoring are identified. This chapter further discusses the blueprint of the entire dissertation, with a brief summary of the major contributions of this work.

1.1 Background and Motivation

The Internet has ushered in a new era in which users' expectations are high for rapid adoption of new technologies and universal access to a steady flow of high-quality multimodal content and diverse methods of sharing knowledge anywhere and anytime. Despite unprecedented advances in network communications and wireless technology, spectrum demand continues to outstrip supply, especially in dense urban areas, where rural and in-building coverage continues to lag. Early studies in spectrum occupancy measurements and analysis revealed that, although spectrum coverage across low, mid and high bands remains scarce in dense urban areas, a significant portion of the spectrum in existing licensed bands is often underutilized [3–5].

The disproportionate use of the spectrum stems from the static method adopted by FCC to allocate frequency bands in a wireless communication environment. The method assigns specific frequency ranges to different users or services in a fixed manner. As such, the method guarantees exclusive use of frequency bands, thereby providing a level of protection against interference and significantly reducing disruptions [5,6]. The method, however, does not allow for dynamic changes or reallocation of frequencies to cope swiftly with emerging technologies, rising demands, and varying conditions. This lack of flexibility results in inefficient spectrum utilization, as licensed frequency bands are frequently underutilized or remain unused during extended periods of time. These deficiencies have led to the exploration of dynamic spectrum allocation methods to optimize spectrum utilization and enhance spectrum efficiency [7–10].

In recent years, various methods have focused on exploring the potential of spectrum sharing to optimize spectrum utilization and meet the rising demands for high-quality mobile connectivity [11–16]. These methods fall broadly under two models, namely Spectrum Property Rights (SPR) [17, 18] and Dynamic Spectrum Allocation (DSA) [19, 20]. SPR allows licensees to sell and trade their assigned spectrum in a free market, without prior regulatory approval or mandate. The basic premise of SPR methods is that the global market will drive the profitable use of the scarce and limited spectrum resource. DSA allows for dynamic sharing of spectrum among multiple communications providers and radio services, by exploiting the spatial and temporal characteristics of the traffic and the time-varying requirements of the services [21]. DSA methods harness the potential of spectrum sharing to optimize spectrum management and maximize spectrum use to address the growing demand for wireless connectivity and the increasing spectrum scarcity in dense areas.

Mindful of the high potential of spectrum sharing for optimized spectrum utilization, various developments, focused on DSA architectures [21–24], system prototypes [10, 16, 25–31], industrial standards [32–35], policies and regulations [36], have emerged to make DSA a reality. Most notable among these developments is the establishment of the Citizens Broadband Radio Service (CBRS) by the Federal Communications Committee (FCC) to facilitate shared federal and non-federal use of the 3550-3700 MHz band [37, 38]. CBRS is conceived to not only promote shared spectrum use and rural broadband deployment but, equally important, to benefit different sectors of the economy, including public safety, small businesses, education, e-commerce, and online consumers. Access and operations of CBRS shared spectrum networks will be managed by a Spectrum Access System (SAS), an automated frequency coordinator to control fundamental access to CBRS and support dense networks across operators, ranging from small in-building networks to nationally deployed large networks. SAS achieves spectrum sharing control by incorporating information provided by environmental sensing capabilities from radars and other sensing devices [39].

To meet the basic requirements of the different stakeholders and constituencies, CBRS stipulates the creation of a three-tiered access and authorization framework to accommodate the requirements of the shared federal and non-federal use of the band. The three tiers of users for this spectrum are:

- **Incumbent:** including federal agencies, Fixed Satellite Service, and grandfathered terrestrial wireless operators,
- **Priority Access License (PAL**), including, typically carriers that pay to license part of the spectrum. PAL is an authorization to use a 10-megahertz channel within a single census tract for a period of three years.
- **General Authorized Access (GAA)**, to serve mainly unlicensed users that require spectrum to establish and manage private networks, including general consumer use, and carrier-grade small network deployments. By rule, GAA use will be allowed throughout the 150 MHz band.

To manage spectrum sharing among the three tiers of spectrum sharing users, the FCC rules mandate that incumbent users be protected from GAA and PAL users' interference. The rules also mandate that GAA users are not to interfere with PAL users. GAA users will receive no interference protection from other CBRS users. The coordination of operations between and among the CBRS users in the three tiers of access authorization is carried out by SAS. To this end, a dynamic database is needed to manage access and operations across the three tiers [40].

FCC CBRS regulatory rules permit commercial broadband users to operate in the 3550-3700 MHz radio frequency spectrum, provided that the operations of federal incumbents in and adjacent to this band are interference-protected. To enforce these rules, common industry-standardized algorithms have been developed to protect a given channel occupied by a CBRS incumbent from second and third-tier users' interference. Collectively, the algorithms and associated mechanisms use a set of agreed-upon RF propagation and aggregate interference models. The main objective of these models is to identify authorized transmissions that are allowed to continue over a protected channel and those transmissions that must be either relocated to another channel if such a channel is available or suspended.

The interference prevention algorithms, coupled with radio trustworthiness and operational compliance, conceptually provide a level of protection reasonably suited to enforcing the applicable rules and policies in a shared spectrum environment. The openness of the wireless medium, however, is susceptible to other forms of spectrum misuse and abuse, that go beyond interference, especially in private and openly shared-spectrum networks [41–44]. These open wireless-induced vulnerabilities are further compounded by the ease of programmability of frequency-agile cognitive radio devices and the high cost required to equip these devices with sophisticated security features. Consequently, shared-spectrum networks are likely to be prone to unauthorized users intentionally carrying out a number of illicit activities, including violating the interference constraints established by the incumbents, transmitting aggressively, both in time and frequency, to gain disproportionate use of the spectrum, and disrupting network operations by violating pre-set spectrum access rules and policies.

The critical threats and anomalous behaviors mentioned above will continue to grow in severity and scale, with (i) the exponential growth and rapid proliferation of various mobile devices with a wide range of hardware and software capability, (ii) unprecedented advances in hardware reconfigurability and agility, and (iii) the need for new spectrum sharing policies to support future innovation and creativity. The focus of this dissertation is to investigate a fundamentally different spectrum enforcement approach, which consists of augmenting a SAS-enabled infrastructure with a pervasive spectrum access monitoring infrastructure composed of volunteers selected from the users of the protected network. The hybrid infrastructure will enhance the enforcement of the interference prevention rules among and between all stakeholders in a shared-spectrum network. It will also mitigate other misuse and spectrum access violations by malicious users.

1.2 Spectrum Enforcement Requirements

A robust spectrum enforcement scheme to ensure stakeholder rights and mitigate risks becomes critical. In order to deliver for its promises and potential, an efficient spectrum enforcement scheme must include the following basic functionalities.

- Comprehensive enforcement for improving the confidence of all stakeholders
- Automation of enforcement procedures to ensure that the outcome of the enforcement process can occur in near real-time and at scale.

Below are critical elements that must be addressed to ensure a robust and efficient spectrum enforcement framework:

1. The enforcement scheme should ensure high geographical coverage. High geographical coverage ensures that the maximum number of spectrum infraction events are detected and dealt with. Traditionally, spectrum monitoring is usually undertaken by government agencies or cellular service providers while driving around a geographical area by using specialized, state-of-the-art devices such as real-time analyzers, spectrum analyzers, and vector signal analyzers. Due to the required high cost of the devices and manpower, this approach does not scale well for real-time, large-scale spectrum monitoring. Conse-

quently, a budget limitation may lead to coverage sparsity in many geographical regions, especially where the population density is low. The lack of spectrum coverage may reduce significantly the ability of the access right enforcement infrastructure to detect spectrum misuse and access violation in poorly covered regions of the enforcement area [45].

2. The types of devices that can be used for monitoring the spectrum are varied, ranging from the expensive traditional superheterodyne swept-tuned spectrum analyzers to low-cost software-defined radios (costing as low as a few hundred dollars). Depending on the type of device used, the tradeoff is usually between the cost and the capability (in terms of bandwidth covered and accuracy) of the device used.
3. Functionalities are needed to protect the rights of authorized spectrum users while guaranteeing the expected performance requirement of the Cognitive Radio network in terms of available bandwidth and Quality of Service.
4. Collaboration among multiple stakeholders should be considered, where different stakeholders are capable of contributing to ensuring comprehensive spectrum enforcement. The enforcement framework should also be scalable and be able to handle an increasing number of infractions.
5. Capabilities must be in place to analyze and characterize the environment to optimize communication strategies based on different constraints such as accuracy, reliability, power consumption, and security.
6. Strategies that are cognizant of the concurrent conditions of the networks must be designed to make decisions efficiently and to take action for maximizing the overall performance of the Cognitive Radio network.
7. The enforcement scheme should be able to differentiate between violations caused by malicious users and access anomalies caused by technical failures.

1.3 Research Scope

Spectrum sharing requires a specification of spectrum rights and the mechanisms to enforce them. These access rights may be either explicit or implicit. Enforcement is what enables these rights to affect behavior. As the number of stakeholders increases, so do the spectrum-sharing granularity and the rate at which spectrum is dynamically allocated. These dynamics give rise to a significant increase in enforceable events, with the potential for new and perceived risks to the stakeholders.

My research work focuses on the enforcement of spectrum access misuse. Three key aspects of a spectrum enforcement regime are the timing of the enforcement action, the form of enforcement sanction, and whether the enforcement action is private or public [46, 47]. Timing is, therefore, paramount to the applicability of any spectrum enforcement action, as enforcement can be either before or after a potentially "harmful" action has occurred. The former enforcement is referred to as *ex-ante* while the latter is *ex-post* [1, 2, 47, 48]. The concept of spectrum enforcement, ex-post and ex-ante, is included in the new Citizens Broadband Radio Service (CBRS) framework, aimed at addressing access rights of the CBRS three tiers of users. The CBRS framework, however, does not specifically address misuse and malicious spectrum access behavior that is designed to distinguish the three classes of user types discussed previously [38].

The ex-ante and ex-post enforcement effects are closely interconnected. For instance, robust ex-ante rules and processes can help prevent ex-post harms before they occur. Additionally, some ex-ante rules may be more straightforward to monitor, potentially reducing enforcement costs. However, even well-designed ex-ante rules may still necessitate ex-post enforcement. For example, while licensing approval is typically granted based on a prototype or pre-production unit, ongoing compliance of production units often requires additional policing [47].

A considerable body of work, carried out by members of the industry and academic communities, has focused on developing mechanisms to enforce access rights in shared-spectrum networks. For example, in its suite of CBRS products and services, Google has provisioned support for interference management of its SAS infrastructure. No support for misuse and malicious behavior has been explicitly defined. Several research works have shown that the social cost (i.e., the opportunity cost of spectrum use) can be quite high and is not internalized appropriately by decision-makers. Further, static ex-ante approaches (such as exclusion or protection zones) are not easily adaptable to policy changes. For example, TV

White Spaces database systems function by restricting users with subordinate rights from accessing spectrum when and where other users with superior rights are active [48, 49].

It is to be noted that the emphasis of research in academia, commercial and government spectrum sharing has focused mainly on preventative (ex-ante) approaches and measures to ensure the protection of the incumbent from spectrum interference. The scope of this dissertation covers ex-post enforcement with the focus on developing new algorithmic approaches and new techniques for automating the enforcement of events after they occur.

Ex-post enforcement consists of corrective measures to be undertaken after an event has occurred. The corrective measures may depend on the access rights being infringed and the ensuing economic losses these infringements may entail. These measures may include penalties, revocation and potentially complete denial of spectrum access over a specified period of time. Automating ex-post spectrum enforcement is challenging, with potential impact on the wider deployment of large-scale dynamic spectrum sharing networks.

The detection of enforceable events includes the detection of "interference events", or RF signal energy that is not consistent with the transmission rights structure. These may include energy from general users to interference protected users, but can also include energy generated by protected users and energy generated among general users. Additionally, when a potentially enforceable event has occurred, forensics must be performed to determine if a claim can be adjudicated and to build evidence in support of an adjudication process. At a high level, the process of spectrum forensics consists of identifying the source of the interference, identifying the time and location of the interference, and assessing the impact of the event. Identifying the source of the interference may involve the individual radio or alternatively the operator/service provider. Other information that could prove valuable to an automated adjudication system may be the related actions, decisions, results, entries, etc. from the Spectrum Access System as well as an after-the-fact statement of transmitter or receiver compliance. This thesis, however, does not focus on the spectrum forensics but instead on detecting spectrum infraction events to automate ex-post enforcement. Since there is a large literature on signal detection and transmitter identification methods, this thesis does not consider the physical aspects of signal detection either. This thesis is instead focused on the development of an effective ex-post spectrum enforcement framework and addressing the relevant challenges.

1.4 Research Goal and Questions

A centralized approach to effective ex-post spectrum enforcement, using collaborative effort of volunteer users, lends itself to the concept of crowdsourcing. Broadly defined, crowdsourcing is the act of assigning a task traditionally undertaken by a designated single agent, to be performed by typically a large group of agents, recruited through an open call [50]. This concept has been used in many contexts including software testing by a large group of software engineers, collaborative microtasking on a large scale, online opinion sharing and brainstorming of complex topics among a large group of people. The immense versatility demonstrated by the use of crowdsourcing across diverse domains and in support of different applications leads to the following thesis statement:

Thesis Statement

”Crowdsourcing is a viable approach to develop a robust and cost-effective ex-post spectrum enforcement framework.”

Crowdsourcing, in this context, offers not only cost-effectiveness compared to traditional static enforcement mechanisms but also ensures broader coverage through extensive data collection and seamless scalability. This approach would involve recruiting authorized transmitters who voluntarily engage in spectrum monitoring.

In order to achieve the goal of effective enforcement of spectrum access rights, algorithmic approaches and mechanisms must be developed to address the following set of questions:

- Can this approach effectively cover both the spectrum and the geographical area of enforcement?
- Given the inherent trust issues with crowdsourced agents, how is it possible to establish reliable detection of spectrum misuse events over prolonged durations?

- Furthermore, what strategies can be employed to guarantee the effective recruitment and selection of the crowdsourced agents?
- Lastly, how is it possible to ensure that spectrum monitoring by crowdsourced agents remains resilient and successful over the long term?

In the following, we explore the approaches to the challenges that need to be addressed to develop an effective ex-post spectrum enforcement framework.

1.5 Ex-Post Spectrum Enforcement Framework

This section gives a blueprint of the proposed framework for ex-post enforcement of spectrum access rights.

1.5.1 Enforcement Area Coverage

Effective spectrum enforcement requires consistent coverage of the enforcement area and all its channels. High coverage ensures that the enforcement scheme is not only responsive to an increasing number of spectrum access infractions but also equipped to take near-real-time actions against perpetrators of spectrum misuse. Traditional static ex ante approaches (such as exclusion or protection zones) are neither easily adaptable to policy changes nor do they ensure high coverage of the geographical area of enforcement. Additionally, they are quite expensive, making it usually infeasible to enhance coverage by increasing the number of deployments of such static infrastructure.

To address this challenge, a hybrid physical infrastructure is employed to utilize crowdsourcing for achieving high coverage. A ”divide and conquer” approach is undertaken by dividing the area of enforcement into smaller zones using the Voronoi and Lloyd’s clustering algorithms [47, 48, 51, 52]. Enforcement is then independently performed in these zones to ensure effective coverage in monitoring the spectrum.

In the proposed framework, authorized transmitters access available channels through a local access point in a geographical region where they currently reside. Unauthorized spectrum access by a transmitter represents a violation of spectrum access rights. The fundamental principle of this work is that unauthorized spectrum access can be effectively detected using crowdsourced spectrum monitoring agents. These agents are recruited and subsequently selected by a cloud-based Spectrum Control System to monitor the spectrum across all the regions in the enforcement area.

The Spectrum Control System consists of three main components: 1) a spectrum access database, 2) a portal that primarily serves as a platform for submitting spectrum access requests, and 3) an access enforcement computational infrastructure that is primarily responsible for registration and selection of crowdsourced spectrum monitoring agents.

1.5.2 Volunteer Selection

Any authorized transmitter can volunteer to monitor the spectrum. An effective volunteer selection process should address all forms of collusion among volunteers to illegally utilize spectrum. Additionally, it is desirable to avoid selecting volunteers who are free riders and dishonest. To achieve this, the following challenges must be addressed:

1. The area of enforcement can be wide and exposed to various factors leading to interference, complicating effective spectrum sensing.
2. Volunteers may behave unreliably.
3. It is essential to achieve near-optimal coverage, considering different volunteer attributes, including geographical location and sensing device characteristics that may constrain monitoring certain channels.

To address these challenges, criteria for volunteer selection need to be determined. Crucial criteria include the hardware capabilities of the volunteer's device, the volunteer's location, and their trustworthiness. Enforcing these criteria during volunteer selection is necessary, and an effective methodology should ensure the selection of only the most qualified volunteers.

The success of a crowdsourced infrastructure depends primarily on the performance of recruited crowdsourced agents. A recruited agent monitoring a geographical region should have a high likelihood of residing in that region. Trustworthiness is another crucial criterion, and failure to consistently report spectrum misuse should be penalized. However, failures can also occur due to 'acts of God,' such as sensing device failure and signal reception. Therefore, effective recruitment should be immune to such challenges.

A methodology is developed to determine the qualification of a volunteer, enabling their selection in a spectrum monitoring epoch. Successful volunteer recruitment should strive for mutual satisfaction between crowdsourced agents and the spectrum enforcement platform. Considering factors like monitoring device characteristics, ensuring the satisfaction of crowdsourced agents will result in fewer incentives for an agent to deviate from the current monitoring channel, reducing overhead.

1.5.3 Robust Spectrum Monitoring For Consistent Detection of Acess Right Violations

Accurate detection of spectrum access violations is required for an effective spectrum enforcement scheme. The challenge of maintaining robustness and reliability of the spectrum enforcement scheme to consistently detect unauthorized access over prolonged intervals needs to be addressed. This is necessary to ensure consistent performance and system persistence against varying physical and environmental adversities.

This need gives rise to the requirement for a system that can perform effectively against not only the different behaviors and actions of intruders but also against crowdsourced spectrum monitoring agents. The system should withstand physical and environmental factors, such as poor signal reception or sensing device failure, which essentially act as unforeseen challenges.

The existing physical layer mechanisms are built upon by utilizing the signature of a device during signal transmission to identify spectrum intruders. An effective spectrum sampling strategy is employed to ensure robust and accurate detection of spectrum misuse, thereby ensuring high-quality sensing. Additionally, a reputation management strategy for crowdsourced spectrum monitoring agents is utilized to enhance enforcement against corrupt agents.

Figure 1 gives an outline of the primary components of this dissertation. In the following section, the primary contributions of this dissertation are discussed.

1.6 Contributions

This dissertation consists of the following main contributions:

- A shared spectrum enforcement infrastructure is proposed that can detect potentially enforceable events effectively. Effective prevention of spectrum misuse and access right violations depends on an infrastructure that includes a few dedicated devices with advanced trust and authentication capabilities, complemented by an opportunistic network of peer wireless devices with diverse software and hardware capabilities.
- A mechanism is established to determine the qualification of crowdsourced spectrum monitoring agents to monitor spectrum in a geographical area of enforcement. This is primarily determined by the trustworthiness of the spectrum monitoring agents and their likelihood to be in a geographical area of enforcement [51].
- Volunteer selection mechanisms are proposed to ensure high coverage, detection accuracy, and mutual satisfaction of the crowdsourced spectrum monitoring agents and the enforcement platform. By combining the advantages of a Secretary-based algorithm — an online algorithm that optimizes the probability of selecting the most qualified volunteers and can serve as a framework for online auctions — with variants of the stable matching algorithm, two hybrid algorithms are developed to address the limitations of the individual vanilla algorithms [53].
- A spectrum sampling strategy by crowdsourced spectrum monitoring agents is proposed, ensuring effectiveness and accuracy in detecting spectrum access violations. This strategy, when coupled with a reputation management methodology, enables the development of a robust system to detect spectrum infractions against varying types of intruders, crowdsourced agents, and environmental factors [54].
- AI-based methodologies are utilized to develop a volunteer mobility model. RNN-based models and Transformers are explored to predict the future location of volunteers to enhance effective volunteer selection for spectrum monitoring.

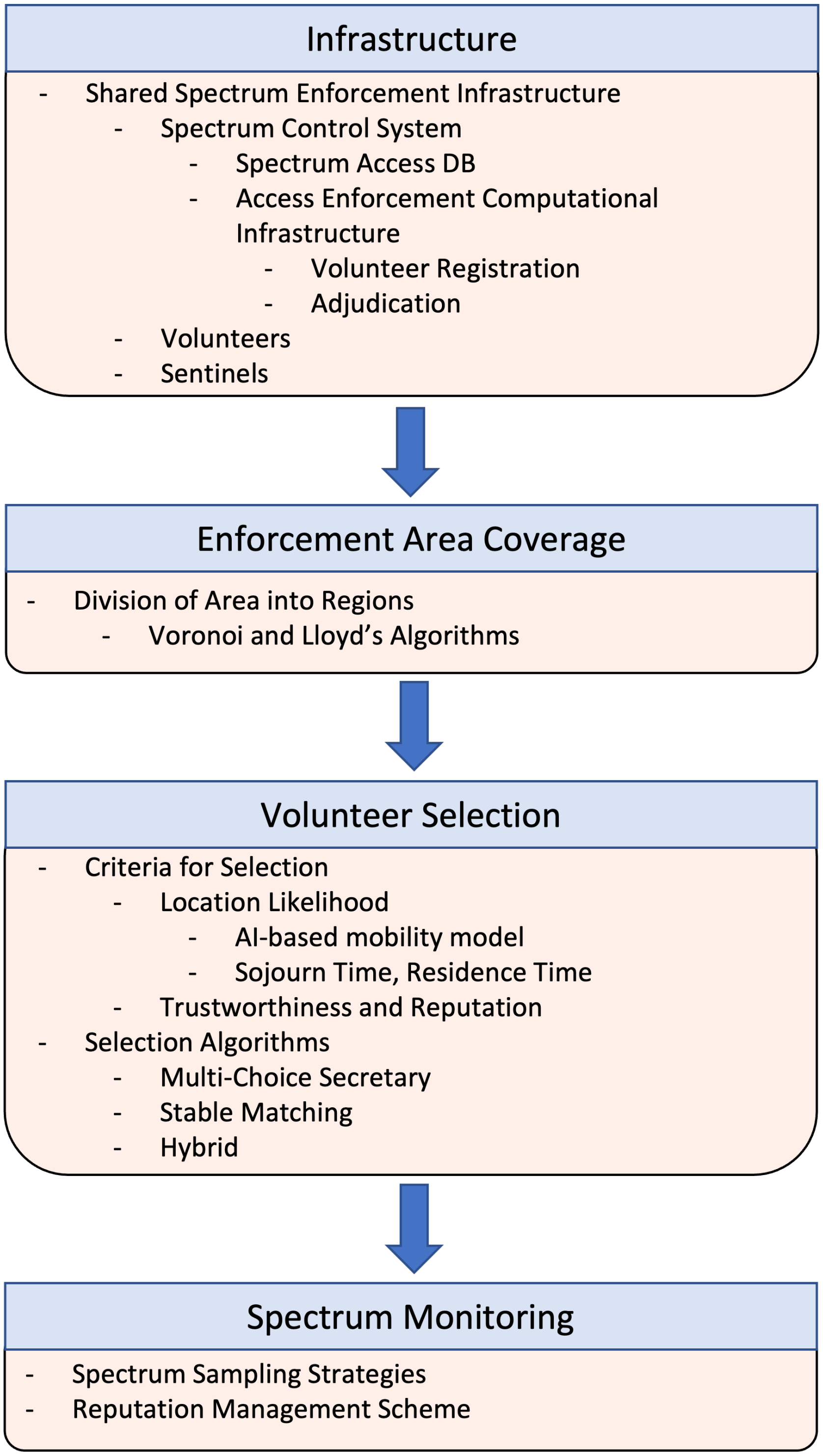

Infrastructure
- Shared Spectrum Enforcement Infrastructure
- Spectrum Control System
- Spectrum Access DB
- Access Enforcement Computational Infrastructure
- Volunteer Registration
- Adjudication
- Volunteers
- Sentinels
Enforcement Area Coverage
- Division of Area into Regions
- Voronoi and Lloyd's Algorithms
Volunteer Selection
- Criteria for Selection
- Location Likelihood
- AI-based mobility model
- Sojourn Time, Residence Time
- Trustworthiness and Reputation
- Selection Algorithms
- Multi-Choice Secretary
- Stable Matching
- Hybrid
Spectrum Monitoring
- Spectrum Sampling Strategies
- Reputation Management Scheme

Figure 1: Dissertation Outline

1.7 Dissertation Outline

The rest of this thesis is structured as follows: Chapter 2 centers on literature reviews, aiming to provide a deeper understanding of existing DSA enforcement solutions. In Chapter 3, the spectrum enforcement framework and the physical infrastructure are discussed. Chapter 4 explores strategies and algorithms for selecting crowdsourced agents to monitor the spectrum. Chapter 5 delves into a spectrum sampling strategy, which, when coupled with a reputation management scheme, ensures robust and consistent detection of spectrum access violations over extended periods. Chapter 6 examines AI-based methodologies employed to develop a volunteer mobility model for predicting the location likelihood of volunteers. Finally, in Chapter 7, the thesis is concluded, and potential directions for future research are explored.

2.0 Related Works

This chapter focuses on a literature review of the works related to my enforcement of spectrum access rights. The primary focus has been given to the works related to varied practices on the detection of spectrum access rights violations and their enforcement.

2.1 Spectrum Sharing

In recent years, various methods have focused on exploring the potential of spectrum sharing to optimize spectrum utilization and meet the rising demands for high-quality mobile connectivity [11–16]. These methods fall broadly under two models, namely Spectrum Property Rights (SPR) [17, 18] and Dynamic Spectrum Allocation (DSA) [19, 20]. SPR allows licensees to sell and trade their assigned spectrum in a free market, without prior regulatory approval or mandate. The basic premise of SPR methods is that the global market will drive the profitable use of the scarce and limited spectrum resource. DSA allows for dynamic sharing of spectrum among multiple communications providers and radio services, by exploiting the spatial and temporal characteristics of the traffic and the time-varying requirements of the services [21]. DSA methods harness the potential of spectrum sharing to optimize spectrum management and maximize spectrum use to address the growing demand for wireless connectivity and the increasing spectrum scarcity in dense areas.

Most notable among the developments in shared spectrum networks is the establishment of the Citizens Broadband Radio Service (CBRS) by the Federal Communications Committee (FCC) to facilitate shared federal and non-federal use of the 3550-3700 MHz band [37,38]. CBRS is conceived to not only promote shared spectrum use and rural broadband deployment but, equally important, to benefit different sectors of the economy, including public safety, small businesses, education, e-commerce, and online consumers. Krishnan et al. study the coexistence of a naval radar (incumbent) with a wide area wireless communication network composed of CBRS devices (CBSDs) and analyze the interference caused by the CBSDs for different protection distances of the incumbent [55]. Kuo et al. develop a Maximum Transmission Continuity scheme to maximize the continuity of CBRS channel access for improving channel utilization and reliability of data transmission among the CBSDs [56]. Access and operations of CBRS shared spectrum networks are managed by a Spectrum Access System (SAS), an automated frequency coordinator to control fundamental access to CBRS and support dense networks across operators, ranging from small in-building networks to nationally deployed large networks. SAS achieves spectrum sharing control by incorporating information provided by environmental sensing capabilities from radars and other sensing devices [39]. Ying et al. propose a dynamic assignment of channels that is assisted by SAS for Priority Access License and General Authorized Access tier users [57]. SAS, being centralized can be prone to failure and may not be trustworthy. To address these issues, Xiao et al. propose a blockchain-based decentralized SAS architecture called BD-SAS that addresses the shortcomings of a centralized SAS and ”provides SAS services securely and efficiently, without relying on the trust of each individual SAS server for the overall system trustworthiness” [58]. On a similar note, Griss et al. propose a trustworthy framework for SAS that utilizes cryptographic and blockchain-based approaches to address the privacy issues of Secondary Users using SAS to access spectrum [59].

To manage spectrum sharing among the three tiers of spectrum sharing users in CBRS, the FCC rules mandate that incumbent users be protected from GAA and PAL users' interference. The rules also mandate that GAA users are not to interfere with PAL users. GAA users will receive no interference protection from other CBRS users. Rochman et al. measure the impact of interference and the degradation in transmission throughput in adjacent channels between CBRS (3.55 - 3.7 GHz) and C-band (3.7-3.98 GHz) [60]. The coordination of operations between and among the CBRS users in the three tiers of access authorization is carried out by SAS. To this end, a dynamic database is needed to manage access and operations across the three tiers.

2.2 Rights and their Enforcement in Spectrum

All Dynamic Spectrum Access (DSA) systems involve the delineation of spectrum rights, comprising both explicit and implicit components. The efficacy of these rights in shaping behavior is contingent upon enforcement. Scholars such as Coase and Hazlett advocate for a ”property rights” framework in the management of spectrum resources, with property rights seen as a social construct emerging from the necessity to separate ownership and decision-making due to economic specialization. This separation facilitates coordination among economic actors by delineating usage and decision-making rights for goods and services within the economy [61].

In the context of spectrum management, ”property rights” are conceptualized as bundles of rights, as elucidated by scholars in previous literature [62, 63]. Constructing a comprehensive enforcement framework necessitates the unpacking of these bundles.

The Common Pool Resource (CPR) literature refines the classification of rights regimes, recognizing that resources exist on a continuum concerning rivalrous consumption and the ability to exclude users. Enforcement is considered an integral element of resource governance in the CPR literature, categorizing users based on the bundles of rights they hold [64, 65].

In adapting the rights structure proposed by Schlager and Ostrom to the realm of spectrum, the composition of rights bundles is contingent on the prevailing usage paradigm. In an exclusive use paradigm, license holders are endowed with ”use” and ”collective action” rights, aligning them with the designation of ”Full Owners” in the Schlager/Ostrom framework. The specifics of reception rights vary based on application types; for example, in broadcasting, certain users may possess ”receive” rights without corresponding ”transmit” rights. The introduction of spectrum sharing introduces additional classes of users, authorized either by government policy or explicit consent through mutual agreement [63, 66].

Demsetz underscores enforcement as a pivotal component of any property rights regime, a sentiment echoed by the CPR literature. Enforcement makes the specification of rights both credible and effective by imposing constraints on the behavior of the involved entities. The overarching objective of enforcement is to foster socially optimal behavior, particularly when individually optimal behavior deviates due to externalities, mistakes, or market failures

[67–69].

Shavell identifies three key dimensions of any enforcement regime: timing (ex-ante or ex-post), the form of sanctions (monetary or otherwise, encompassing potential criminal consequences), and the public or private nature of the enforcement action [1]. The effectiveness of an enforcement mechanism is deeply connected to the rights regime and the economic context. Striking a balance between the costs of inducing desirable behavior and the social costs and benefits is paramount, considering factors such as evidence collection, the challenges of detecting undesirable behaviors, liability establishment, claims adjudication, and the subsequent imposition of sanctions [70].

Ex-post enforcement through penalties or rewards is inherently uncertain, given the often unpredictable responses of entities to such incentives. Remedies may include license revocation, financial penalties, product recalls, or adjustments to operating rights. To serve as an effective deterrent, the expected incidence of penalties must be sufficiently substantial to counterbalance the private benefits derived from engaging in undesirable behavior [70].

2.3 Ex-Ante Enforcement

Ex ante enforcement of spectrum access rights operates by attempting to prevent interference [71]. Regulatory authorities in both the United States and internationally commonly employ Exclusion Zones (EZs) as a primary ex-ante spectrum enforcement strategy to safeguard incumbents, also known as Primary Users (PUs), in scenarios involving spectrum sharing [71]. Exclusion Zones are designated spatial separation regions aimed at shielding incumbents from interference generated by Secondary Users (SUs). Establishing the boundaries of EZs, within which incumbents enjoy exclusive spectrum access rights, poses a challenging problem in spectrum sharing due to conflicting requirements. The defined EZ area must be sufficiently large to protect PUs from interference caused by SUs, yet not overly large to avoid unnecessarily restricting SUs' spectrum access opportunities and reducing incentives for new entrants [71, 72].

The computation of EZ boundaries typically relies on assessing the interference likely to be experienced by a PU, considering the aggregate interference from all co-existing unlicensed users in a Dynamic Spectrum Sharing (DSS) system [73]. The dynamic nature of SU dynamics introduces variations in the statistics of aggregate interference. Moreover, in computing EZ boundaries, the impact of irregular terrain on path loss computations must be considered, adding complexity to an already challenging problem [73]. Existing methods, such as F curves [74], often adopt a worst-case interference scenario and define a conservative static protection boundary for PUs, emphasizing protection against harmful interference [72, 75, 76].

Addressing these challenges, Bhattarai et al. propose a novel framework called Multi-Tiered Dynamic Incumbent Protection Zones (MIPZ) [72, 77]. This systematic framework can be utilized by geolocation database (GDB) systems to dynamically prescribe protection boundaries for PUs in real-time. Unlike traditional approaches with static and overly conservative EZ boundaries, MIPZ allows for the dynamic adjustment of PU protection boundaries based on the evolving radio interference environment. The MIPZ framework guarantees protection for primary users (PUs) against harmful interference, offering a probabilistic assurance of interference protection [72, 77].

However, ex-ante enforcement methodologies lack the flexibility to adapt to rapidly changing technological advancements, market dynamics, and unforeseen developments in shared spectrum networks. Exclusion zones cannot always assure co-channel interference avoidance and its opportunity cost is high [71]. The inherent unpredictability of propagation introduces the possibility of uplink signals occasionally exceeding anticipated distances. Moreover, exclusion zones do not explicitly consider the presence of tall features such as buildings and mountains, which can lead to propagation distances longer than anticipated [71]. Therefore, ex-post approaches are needed to increase coverage and ”provide data to PUs and SUs to further tune the system for future interference avoidance” [71]. Furthermore, the openness of the wireless medium, however, is susceptible to other forms of spectrum misuse and abuse, that go beyond interference, especially in private and openly shared-spectrum networks [41–44]. These open wireless-induced vulnerabilities are further compounded by the ease of programmability of frequency-agile cognitive radio devices and the high cost required to equip these devices with sophisticated security features. Consequently, shared-spectrum networks are likely to be prone to unauthorized users intentionally carrying out several illicit activities, including violating the interference constraints established by the incumbents, transmitting aggressively, both in time and frequency, to gain disproportionate use of the spectrum, and disrupting network operations by violating pre-set spectrum access rules and policies. This motivates the exploration of ex-post spectrum enfocement.

2.4 Ex-Post Enforcement

2.4.1 Detection and Localization of Spectrum Misuse

Numerous works have been proposed to detect misbehavior in wireless networks. Tang et al. [78] propose an adaptive approach for real-time selfish misbehavior detection in IEEE 802.11-based wireless networks. A basic misbehavior detector is designed based on the non-parametric cumulative sum (CUSUM) test. An adaptive version augments the performance of the basic detector by utilizing the Markov Decision Process for faster and more efficient misuse detection. However, this model is not suitable for multi-hop networks and is not scalable because of the sequential monitoring of nodes by a single detector. Liu et al. [79] investigate the problem of detecting unauthorized spectrum usage in DSA networks. The detection of anomalous spectrum usage is formulated by using statistical significance testing. The authors propose two detection schemes based on the mobility of the authorized transmitter. The first scheme is based on the assumption that the Received Signal Strength (RSS) from a single transmitter1 decays approximately linearly with the logarithmic distance from the source, but it is no longer the case when the RSS is a sum of multiple transmitters at different locations. This scheme is used when the authorized user is mobile. The second scheme is based on the assumption that the transmitters at different locations will lead to different spatial distributions of the RSS and is used when the authorized transmitter is static. However, the authors assume that the unauthorized transmitters are non-colluding and that the authorized transmitters are always trustworthy. Khaledi et al. [80] propose a crowdsourced approach that simultaneously locates multiple transmitters using the received power measurement from mobile devices. Their proposed methodology is based on the following observations: a) receivers that are located near the transmitter observe higher power than the receivers that are distant from the transmitter, and b) the observed RSS at each receiver is primarily affected by the nearest transmitter. The local maxima of the RSSs observed by crowdsourced agents that are greater than a threshold are obtained and utilized to locate the unauthorized transmitters.

Sarkar et al. proposed a supervised deep learning-based spectrum sensing method known as RadYOLOLet. This approach is designed to detect low-power radar signals amidst interference and estimate the parameters of these radar signals [81]. This work is especially useful for detecting incumbent (in this case, radar) signals to prevent them from interference by secondary users in shared spectrum networks. Villa et al. build a CNN-based agent to detect incumbent radar transmissions and vacate the cellular bandwidth to avoid interference [82]. Shi and Sagduyu proposed a generative adversarial network (GAN) approach to sensing an incumbent presence in a spectrum band before a Next Generation (NextG) device (which is a secondary user) uses the frequency band for communication [83]. Although these approaches prevent interference when incumbents are operating by detecting their authorized transmissions, they have the potential to detect unauthorized transmissions for spectrum monitoring.

2.4.2 Intruder Identification and Punishment

Upon the occurrence of a potentially enforceable event, forensic analysis becomes essential to ascertain the viability of a claim for adjudication and to compile supporting evidence for the adjudication process. This includes the identification and punishment of spectrum access violators (also called intruders). In an ideal scenario, the intruder identification unit would opt for employing a PHY-layer authentication procedure for identification [84]. This procedure enables a receiver to swiftly discern between compliant and rogue transmitters, eliminating the need for unnecessary higher-layer processing. To make this method workable, it is essential for all Secondary User (SU) radios to include a mechanism for authenticating their waveforms and utilize tamper-resistant measures to prevent potential circumvention by malicious users [84]. PHY-layer authentication schemes can be intrinsic or extrinsic. Intrinsic schemes utilize the intrinsic characteristics of the transmitted signal as unique signatures to identify transmitters. These methods primarily include RF fingerprinting and electromagnetic signature identification [85–88]. However, such schemes are sensitive to environmental factors like interference, and temperature changes and are not as effective in "real-world scenarios" [84].

Extrinsic identification schemes, on the other hand, enable authorized transmitters to embed authentication signal intrinsically to their message signal (like machine authentication code, digital signatures, spectrum permits [89]), which can then be decoded by the receiver [84]. Such techniques include PHY-layer watermarking [90] and transmitter authentication [91, 92]. However, such schemes involve the superposition of the message signal over the authentication signal [90]. This means that the signal-to-noise ratio (SNR) of one signal is compromised by the other [84]. Another drawback of this scheme is that the SNR of the received signal should be high so that the received authentication signal is decoded and demodulated successfully by the receiver [84]. To address these challenges, a scheme called "blind transmitter identification" was developed to "enable a regulator to uniquely identify (or authenticate) a transmitter under low SNR and high multipath fading conditions while not requiring the regulator to have complete knowledge of the PHY-layer transmission parameters" [84, 93].

Punishment of intruders who are located and identified is usually of two types [43, 94]:

1. The malicious transmitter is not allowed to access the spectrum for a duration that is proportional to the severity of the spectrum access violation. This is usually achieved by revoking its license or operating rights [84]. Several works have focused on limiting the spectrum access of intruders. Sagduyu develops a game theoretic framework to study the interactions of intruders and defenders in shared spectrum networks [95]. Here, defenders of a shared spectrum network introduce controlled errors in intruders' spectrum access decisions to limit their spectrum access [95]. In [96], Sagduyu et al. investigate the impact of adversarial machine learning on wireless communications in the context of 5G in shared spectrum networks and develop a defense mechanism to address such attacks.
2. The malicious transmitter is financially penalized. The amount of penalty is again de-

pendent on the severity of harm caused and may include paying those who suffered due to the intruder [84].

2.4.3 Crowdsourced Approaches

A wide variety of research works has focused on the design of crowdsourcing frameworks, approaches, and methods to leverage the collective intellect of a crowd and engage experts and stakeholders to work collaboratively in sharing ideas, brainstorming solutions and advancing knowledge. While most of these crowdsourcing-driven research efforts share common buildings blocks, the features and context within which the crowdsourcing infrastructure is conceived differ to varying extents.

One important aspect of crowdsourcing, which has a significant impact on individuals' recruiting and participation in the event, is the incentive mechanism. To be effective, this mechanism must appeal to the targeted participants and motivate them to engage successfully in the planned crowdsourcing activities. A variety of methods have been proposed for creating effective engagement [97]. Some of these methods use mutual benefit or direct impact on the participants' individual interests to recruit participants [98]. Other methods use financial incentives or benefits to the general population to motivate participants' engagement in the crowdsourcing event [99–103].

Different typologies for crowdsourcing have been proposed, depending on the type of the crowdsourcing event, collaboration- or competition-based, the intended objectives, the participants' selection process, and the control of the tasks to achieve the targeted outcomes [104–106]. These topologies have been adopted, with varying degrees of success, in various activities and in different communities, including Wikipedia, Youtube, Healthcare, Disaster Management, Citizen Science, Transportation, City Management, Upwork, Gigster, TopCoder, Rainforest QA, Kaggle, GLG, HackerOne [107–111].

Numerous crowdsourced mechanisms for ex-post spectrum enforcement have been proposed. Jin et al. [89] propose the first crowdsourced spectrum misuse detection framework for DSA systems. This mechanism requires a legitimate transmitter to embed spectrum permit into its physical layer signals, which can be decoded and verified by ubiquitous crowdsourced, mobile users. This work primarily revolves around schemes for embedding spectrum permits in physical layer signals by the authorized transmitter. The authors do not discuss any mechanism to select the crowdsourced agents based on their capabilities. Dutta and Chiang [112] selects crowdsourced agents based on their spectrum misuse detection accuracy. The proposed mechanism shows improved performance over static enforcement and is also equipped to detect mobile violators. However, it is based on the assumption that all crowdsourced agents are trustworthy. Several incentive-based crowdsourced spectrum sensing works have been done over the past few years. Zhang et al. [113] proposes three online incentive mechanisms for smartphone sensing applications based on online reverse auctions. While the authors discuss auction mechanisms for both truthfulness and utility maximization, the value of a crowdsourced mobile detector to the platform is loosely defined. Yang et al. [114] examined two incentive-based crowdsourcing models: one in which a Stackelberg Equilibrium was calculated for the platform-centric model, and another where a truthful auction mechanism was proposed for the user-centric model. Zhu et al. [115] introduced an incentive-based auction mechanism aimed at enhancing bid fairness by addressing the impacts of malicious competition behavior and the "free-riding" phenomenon in crowdsourcing services. Lin et al. [116] take the Sybil attack into consideration for incentive-based crowdsourced spectrum sensing. Li et al. [117] propose a spectrum monitoring framework in multi-channel cognitive radio networks with limited monitoring resources. Spectrum misuse detectors make the best choice in choosing channels to monitor and in switching to a different set of channels by solving the problem as a Combinatorial Adversarial Multi-Armed Bandit Problem with Switching Costs. While the proposed methodology aims to balance strategy rewards and switching costs with exploration and exploitation, the authors assume that the spectrum monitors and malicious users are static. Salama et al. [118] proposed an optimal channel assignment framework for crowdsourced spectrum monitoring, where volunteers are assigned to monitor channels based on their availability patterns and are awarded incentives in return. The works [119] and [120] propose frameworks for crowdsourced spectrum sensing without violating the location privacy of mobile users.

This dissertation utilizes a hybrid crowdsourcing infrastructure that employs dedicated sentinels and recruits volunteers to improve monitoring effectiveness and reduce costs.

2.5 Summary

This chapter reviews the related work in the field of spectrum access enforcement in shared spectrum networks. It begins with an overview of access rights and their practices for their enforcement in the spectrum. This is followed by a discussion on the most common ex-ante enforcement practices. This is followed by a discussion on the related works of ex-post enforcement with primary focus on: 1) approaches for detection and localization of spectrum misuse, 2) schemes for intruder identification and punishment, and 3) existing crowdsourced approaches and their shortcomings.

3.0 System Model and Infrastructure Coverage

The primary challenge in the design of an effective spectrum enforcement infrastructure stems from the fact that it is not easy to determine where and how the spectrum monitoring resources are to be mobilized, given the non-deterministic nature of the mobile devices' behavior. It is equally difficult to determine how collaboration between spectrum monitoring devices must take place to ensure swift detection and response to spectrum misuse and access rights violations. To this end, a shared spectrum enforcement infrastructure is designed that utilizes crowdsourcing to detect spectrum access misuse effectively. This infrastructure consists of regions where devices can communicate in shared spectrum networks without compromising on the spectrum access rights of different stakeholders. Spectrum is allocated dynamically upon request and the enforcement of access rights is maintained by a hybrid architecture comprising a centralized cloud-based spectrum control system augmented by crowdsourced spectrum monitoring volunteers.

This chapter delves deep into the system model and the different components of the shared spectrum monitoring infrastructure. The strategy for ensuring effective coverage of the enforcement area is also discussed. In line with our research goal, a major portion of the chapter is dedicated to discussing the different components and functionalities of the infrastructure for effective management and deployment of crowdsourced spectrum monitoring volunteers in the enforcement area.

3.1 Infrastructure

The spectrum access rights enforcement area is partitioned into geographical sub-divisions called regions. By dividing the enforcement area into regions, enforcement efforts can be more targeted and responsive to local conditions and demands. An enforcement region is where spectrum access rights are enforced according to the underlying policies. The main functionalities that are supported in the enforcement area are 1) spectrum sharing, allowing for a flexible and dynamic allocation of frequencies to devices, 2) registration of crowdsourced volunteers and their selection to monitor the spectrum, and 3) automated adjudication for identifying intruders. To this end, the shared spectrum access enforcement infrastructure is built around a cloud-based centralized Spectrum Control System (SCS) and crowdsourced spectrum monitoring volunteers. The SCS consists of the following components:

1. *Portal*: The SCS portal includes the Portal UI and Portal API that manages the devices that are connected to SCS in the enforcement area. It serves as a platform for submitting requests for spectrum access, including the specific frequencies and geographic locations where the user intends to operate. It also provides real-time information about the availability of spectrum in different geographical regions with the help of visualization tools like heat maps.
2. *Spectrum Access Database*: This database stores information of all authorized devices registered with SCS in the enforcement area along with the information required to manage these devices. It is further responsible for dynamically allocating available spectrum to different users or services based on their frequency requirements while protecting the access rights of other authorized devices in the network by preventing interference.
3. *Access Enforcement Computational Infrastructure*: This infrastructure consists of computational servers and databases that are responsible for 1) registration, selection, and deployment of volunteers to monitor spectrum, 2) collecting spectrum monitoring reports from volunteers, and 3) automating adjudication for identifying intruders.

Authorized transmitters access available channels from the SCS through a local access point located in their current geographical region. The use of spectrum frequency by malicious users, known as *intruders*, represents a violation of spectrum access rights. Unauthorized spectrum access can be effectively prevented using a reliable spectrum monitoring infrastructure composed of trusted devices, referred to as *sentinels*. Sentinels are equipped with advanced authentication capabilities and are fully dedicated to monitoring spectrum access. Achieving high level of spectrum misuse detection over extended geographical coverage, however, is likely to come at a high cost of using a large number of sentinels. To minimize cost, this framework employs crowdsourcing to leverage the multitude of peer wireless devices, known as *volunteers*, thereby complementing the advanced capabilities of the trustworthy sentinels and enhancing the detection of unauthorized spectrum use. The volunteers are recruited to join the monitoring infrastructure, based on their software and hardware capabilities by the Access Enforcement Computational Infrastructure of SCS. The volunteers, however, may not be *honest* and as such may not always report spectrum access violations truthfully. To protect against volunteers' corruption, it is necessary to check and validate the reported spectrum channel access by the volunteer. To this end, the mobile *sentinels* are deployed to assess the behavior and performance of volunteers. This allows the SCS to make informed decisions on the selection of more trustworthy volunteers to monitor the spectrum. As shown in Figure 2, the shared spectrum enforcement infrastructure consists of the regions of enforcement with the centralized cloud-based spectrum control system that ensures enforcement of spectrum access rights with the help of crowdsourced spectrum monitoring volunteers and sentinels. In each region, the devices gain access to an available channel from the SCS through a local access point in the region [53, 54].

In the following sections, I discuss the algorithms used to divide the enforcement area into regions and delve deep into the Access Enforcement Computational Infrastructure of the SCS.

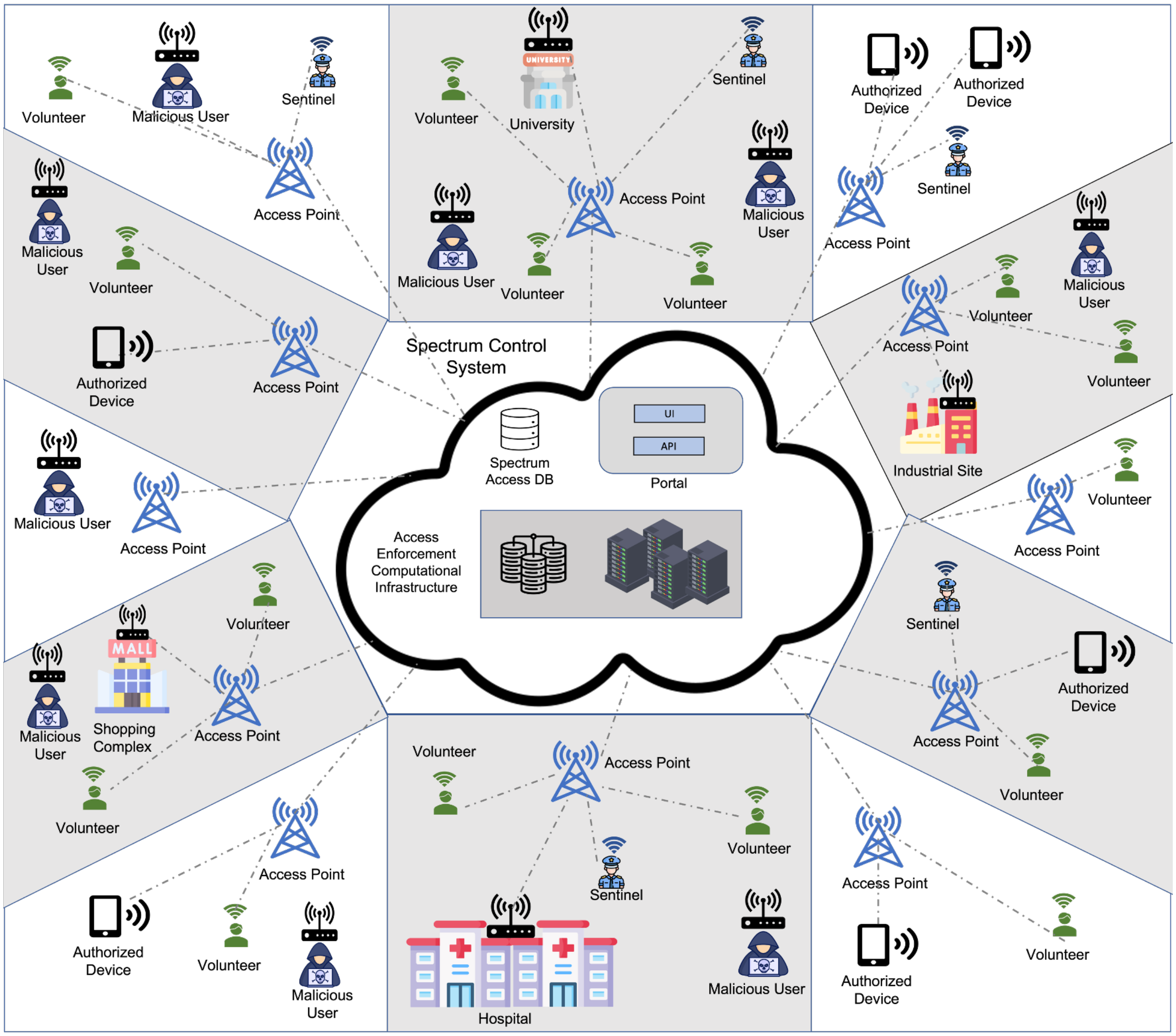


Figure 2: Shared Spectrum Enforcement Infrastructure

3.2 Enforcement Area Coverage

To ensure maximum coverage of the area for enforcement, a divide-and-conquer methodology is followed. We propose to divide the entire enforcement area, R, into smaller regions and then focus on solving the enforcement problem for a single region, $r \in R$. This, in turn, can be used for solving the enforcement problem for the whole area, R. For the division of R into regions, the employment of the Voronoi algorithm is proposed [47, 48, 53, 121]. The origin of Voronoi algorithm can be dated back to R. Descartes in his book on the principles of philosophy in the 17^{th} century [122]. He discussed the decomposition of space into convex regions, consisting of matter revolving around fixed stars [122]. The fundamental concept is that if there is a space, M, and a set, S, of sites, p, within M, then the region associated with site p includes all points, x, where the influence of p is the strongest [122]. This concept was formally introduced by Dirichlet [123] and Voronoi [124, 125], who applied it to the study of quadratic forms with sites as integer lattice points and influence defined by the Euclidean distance [122]. The resulting structure is known as the Dirichlet tessellation or Voronoi diagram.

It is assumed that the area of spectrum enforcement is a 2D Euclidean plane and there is a set of spectrum access points randomly distributed across this plane. For each spectrum access point, p, in the 2D plane, the corresponding Voronoi region, r, is the set of points in the plane that are at least as close to p as to any other access point in the enforcement area, R. [121].

Assuming that S is the set of at least three spectrum access points $p, q, r, ...$ in the 2D Euclidean plane that represents the spectrum enforcement area. For access point $p = (p_1, p_2)$ and point $z = (z_1, z_2)$, it is assumed that $d(p, z) = \sqrt{(p_1 - z_1)^2 + (p_2 - z_2)^2}$ denote the Euclidean distance, $\overline{pq}$ denote the line segment from p to q and $\overline{M}$ denote the closure of any set M [122].

Definition 3.2.1. For access points $p, q \in S$, let

$$B(p, q) = \{z | d(p, z) = d(q, z)\} \tag{1}$$

be the perpendicular bisector of $\overline{pq}$. It splits the halfplane

$$D(p, q) = \{z | d(p, z) < d(q, z)\} \tag{2}$$

containing access point p from the halfplane $D(q, p)$ containing access point q. The Voronoi region associated with access point p is formed by the intersection of the half-spaces defined by the perpendicular bisectors. Thus, the Voronoi region of p, with respect to the set of all access points, S, is defined below.

$$VR(p, S) = \bigcap_{q \in S, q \neq p} D(p, q) \tag{3}$$

Voronoi regions are thus convex polygons. Hence, the overall Voronoi diagram of S is defined by

$$V(S) = \bigcup_{p,q \in S, p \neq q} \overline{VR(p, S)} \bigcap \overline{VR(q, S)} \tag{4}$$

Thus, each Voronoi region $VR(p, S)$ can be defined as the intersection of $n - 1$ open halfplanes containing the access point p [122].

However, from an implementation perspective, this naive algorithm for the construction of Voronoi regions is not convenient. This is because:

- It may cause inconsistency due to precision problems.
- It fails to produce immediate neighborhood information.
- The time complexity of this algorithm is $O(n^2 logn)$ where n is the number of spectrum access points [126].

Therefore, the proposed algorithm for constructing Voronoi regions employs a divide-and-conquer approach. This approach can be outlined through the following steps [127].

1. Assuming the plane for computing the Voronoi diagram is rectangular, the spectrum access points nearest to each of the four corners of the rectangle are identified using Euclidean distance.
2. If all the corner points are nearest to the same access point, then all points within the rectangle defined by these corners are assigned to the region associated with that access point.
3. If all the corner points are not closest to the same access point, the rectangle defined by the four corners is subdivided into four smaller rectangles: the top-left, top-right, bottom-left, and bottom-right sections of the original plane. Each of these four fragments is then recursively processed starting from Step 1.

This recursive process continues until all subdivided rectangles cover a single seed closest to their four corner points. The algorithm may need to keep subdividing into smaller rectangles until a single discretized point is found, which is closest to only one seed. In the event of a tie, the first processed seed is considered closest [127]. The time complexity of this algorithm is $O(nlogn)$ where n is the number of spectrum access points in the geographical area of enforcement.

However, the Voronoi algorithm may yield regions of varying sizes, which can be a disadvantage because it may result in an uneven distribution of volunteers over time. This may result in possible loss in the detection of spectrum access violations. Thus, it is proposed to apply a relaxation to the Voronoi algorithm, called the Lloyd's Algorithm [52], which produces uniformly sized convex regions, and thus improves the probability of a fair distribution of volunteers over all regions [47, 48]. The Lloyd's algorithm can be explained by the following few steps:

1. The Voronoi regions for a set of spectrum access points, S, is generated.
2. The centroid of each Voronoi region is computed
3. The points in S are updated with the centroids of the corresponding region. This is continued until the set S has not converged.

Convergence of the above algorithm can simply be threshold-based where the process can be stopped when the cumulative distance between the points of two subsequent iterations is below a threshold [128]. It is necessary to observe that the spectrum access is fixed in the geographical regions. However, the size of each region can still be adjusted by using the centroids of these regions until a certain threshold is reached.

3.3 Access Enforcement Computational Infrastructure

The Access Enforcement Computational Infrastructure is a centralized platform that is responsible for ensuring effective monitoring and detection of spectrum access violations in the monitored region. The primary objectives of this infrastructure include:

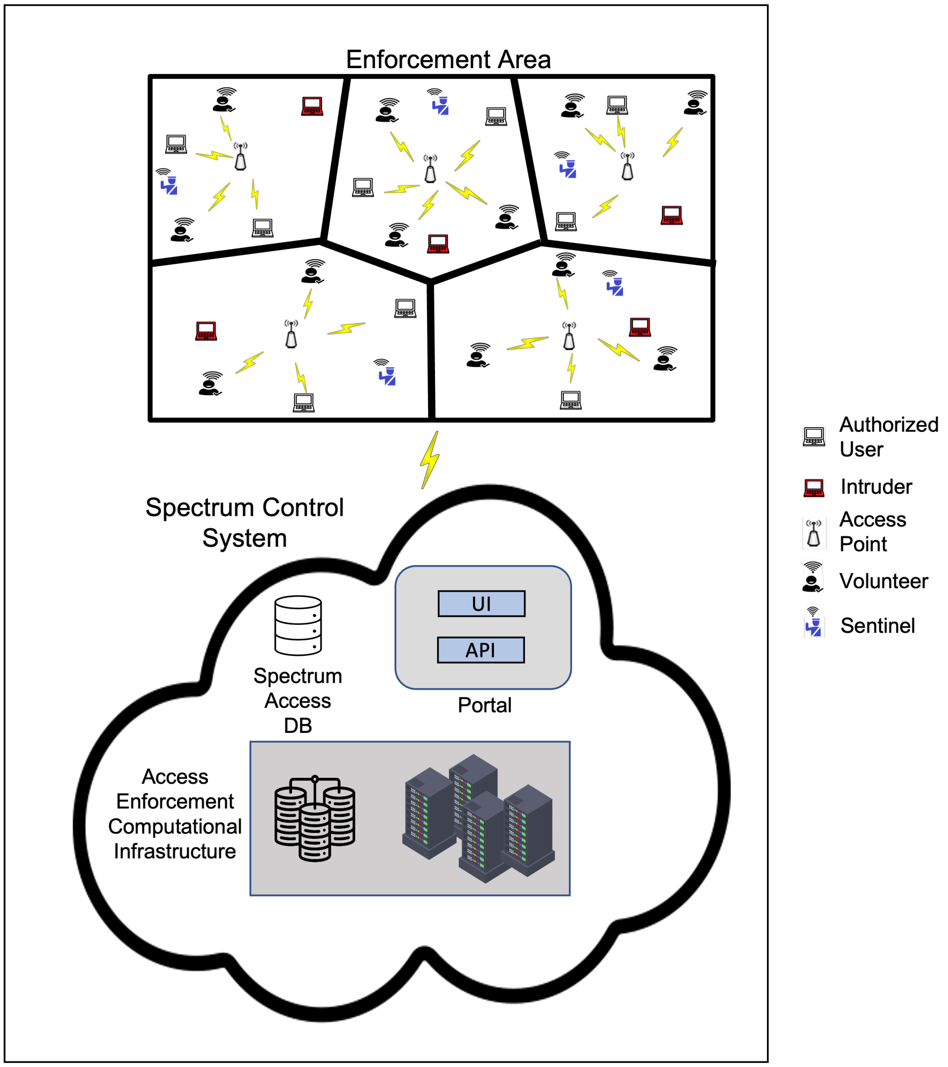


Figure 3: Division of Enforcement Area Into Regions using Voronoi and Lloyd's Algorithms

- Selection of crowdsourced volunteers to monitor channels in the enforcement area.
- Establishment of the trustworthiness of volunteers to monitor spectrum.
- Estimate of the performance of volunteers and their likelihood to be in a geographical region, so that only the most qualified volunteers can be selected to monitor the spectrum.

Attaining the above objectives is challenging because it is not easy to determine where, when, and how volunteers should be deployed. Additionally, it is necessary to make crucial decisions in near real-time regarding the eviction of volunteers when they perform poorly. Finally, it is challenging to ensure that the perpetrator of spectrum misuse is correctly identified every time and penalized after gathering irrefutable evidence in real-time. To support the above salient objectives, this infrastructure has the following functionalities:

- Support of registration of volunteers to the infrastructure.
- Selection of volunteers to monitor spectrum and their assignment to a geographical region.
- Maintenance of relevant volunteer information for estimating their performance and trustworthiness.
- Adjudication of the veracity of volunteer reports after monitoring spectrum, to identify and penalize the spectrum access intruders.

The Access Enforcement Computational Infrastructure consists of multiple components to support the required functionalities. The different components, sub-components, their functionalities, and characteristics are discussed in Section 3.3.1.

3.3.1 Components of the Access Enforcement Computational Infrastructure

The functionalities of the Access Enforcement Computational Infrastructure are supported by two primary components - the Volunteer Support Component and the Adjudication Component. A brief overview of these components and their sub-components is given below.

I Volunteer Support Component: This component is primarily responsible for addressing the registration of volunteers and their recruitment for spectrum monitoring. It consists of the following sub-components:

  a Volunteer Registration Database: This database stores the attributes related to the registration of volunteers.

  b Dynamic Attribute Database: This database stores volunteer attributes that are dynamically updated.

  c Volunteer Selection Unit: This unit is primarily responsible for selection of volunteers who are qualified to monitor spectrum in a geographical region.

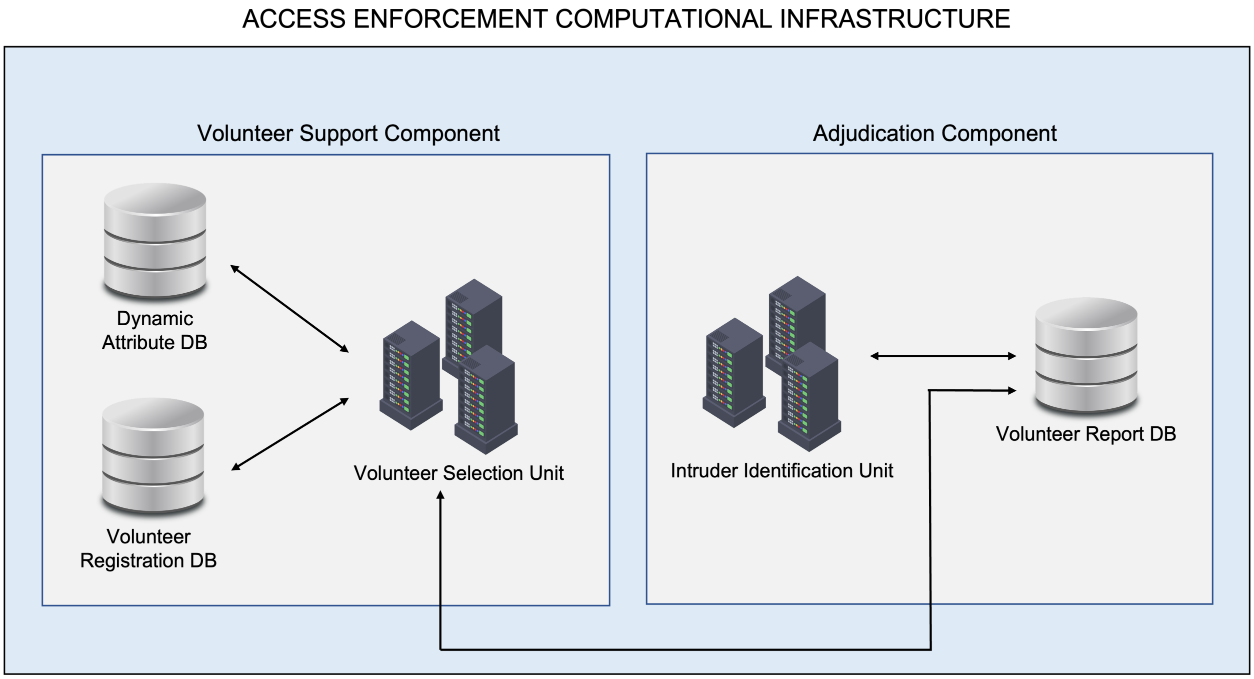

ACCESS ENFORCEMENT COMPUTATIONAL INFRASTRUCTURE
Volunteer Support Component
Adjudication Component
Dynamic Attribute DB
Volunteer Registration DB
Volunteer Selection Unit
Intruder Identification Unit
Volunteer Report DB

Figure 4: Components of the Access Enforcement Computational Infrastructure

II Adjudication Component: This component is mainly responsible for an automated adjudication for identifying and verifying spectrum access intruders. It consists of the following sub-components:

   a Volunteer Report Database: This database maintains the list of reports submitted by volunteers after monitoring spectrum.

   b Intruder Identification Unit: This unit is responsible for verifying the identity of spectrum intruders by utilizing the information in the Channel Occupancy Database and the Volunteer Report Database.

The different components of the Access Enforcement Computational Infrastructure are illustrated in Fig. 4. The detailed functionalities and characteristics of these components are discussed below.

I **Volunteer Support Component**: The Volunteer Support Component hosts sub-components for supporting the recruitment of volunteers and their selection to monitor

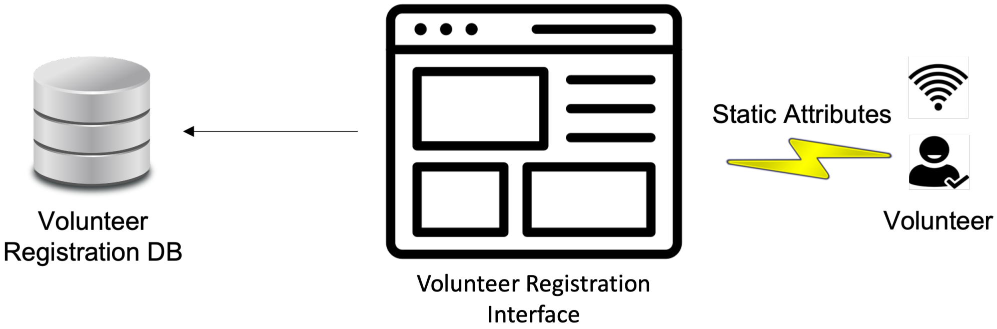


Figure 5: Volunteer registration methodology

spectrum. The recruitment of volunteers is facilitated by their registration through a registration interface. A database, referred to as the *Volunteer Registration Database*, is also maintained in the back-end for storing the *static* volunteer attributes that are collected during the registration process. In addition, there is a database, referred to as the *Dynamic Attribute Database*, that stores volunteer attributes which are collected and/or determined during run-time. Finally, there is a unit, referred to as the *Volunteer Selection Unit*, that utilizes the data in both the *Registration Database* and the *Dynamic Attribute Database* to perform the following major tasks:

- Assessment of the behavior and performance of volunteers.
- Implementation of an algorithm for selection of volunteers and their assignment to a geographical region for monitoring a channel.

a **Volunteer Registration Database**: Volunteer recruitment is facilitated by their registration to the Access Enforcement Computational Infrastructure of SCS. As shown in Fig. 5, Volunteer registration is assisted by a registration website which provides an interface for volunteers to provide the required information. This information is utilized to associate some *static* attributes with the volunteers. Such *static* attributes include the personal identification information, hardware device capability, the most likely ”home” and temporal availability of volunteers. These attributes are stored in the Volunteer Registration Database and are particularly useful for volunteer selection to monitor spectrum.

b **Dynamic Attribute Database**: This database is used to store the *Dynamic* attributes associated with volunteers. These attributes are updated during run-time and are utilized to make decisions on selection of volunteers and their assignment to channels in a geographical region. Dynamic attributes can be *situational* or *behavioral*. Situational attributes include attributes such as the current geographical location, average velocity, residual battery life of the sensing device, etc. Such situational attributes can be utilized to estimate the likelihood of volunteers to be in a geographical region and their capability to monitor channels. The performance of volunteers in monitoring spectrum also helps in assessment of their behavioral characteristics, like their trustworthiness in monitoring channels over prolonged intervals of time.

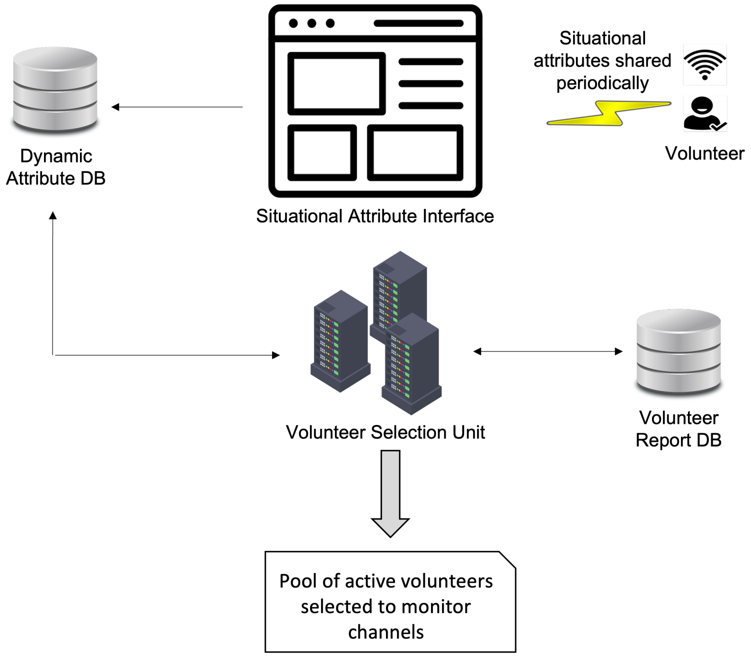

Situational attributes shared periodically
Volunteer
Dynamic Attribute DB
Situational Attribute Interface
Volunteer Selection Unit
Volunteer Report DB
Pool of active volunteers selected to monitor channels

Figure 6: Interaction between a volunteer and different components of the Access Enforcement Computational Infrastructure for updating the dynamic and behavioral attributes of the volunteer

c **Volunteer Selection Unit**: The Volunteer Selection Unit is responsible for selecting volunteers to monitor spectrum in a geographical area of enforcement over an epoch of enforcement. This is attained by implementing the following methodologies:

- Volunteer Selection Methodology: The selection of capable and qualified volunteers regularly, to monitor spectrum, is challenging. This is because an effective volunteer selection mechanism should ensure that there is
  - High accuracy in detection of enforceable events.
  - Adaptability to changes in the enforcement framework and environment.
  - Mutual satisfaction of the volunteers and the volunteer selection unit.

  Therefore, it is essential to implement an efficient algorithm for selecting volunteers at regular intervals. In addition, it is necessary to determine when and where to deploy volunteers for enforcement. For example, if a group of volunteers perform poorly in a monitoring epoch, then it is necessary to evict such volunteers from monitoring and select a new set of volunteers to replace them, even before the monitoring epoch ends. Finally, it is necessary to establish an efficient procedure for assignment of volunteers to monitor channels in a geographical region that they are most likely to reside in a monitoring interval. As shown in Fig. 6, the Volunteer Selection Unit utilizes the information in Volunteer Registration Database and the Dynamic Attribute Database to make decisions on selection and assignment of volunteers.

- Upgrade of Volunteer Attributes: The situational attributes of volunteers in the Dynamic Attribute Database are utilized to estimate, in run-time, their likelihood to reside in a geographical region during a spectrum monitoring epoch and estimate their performance in monitoring channels. In addition, the trustworthiness of volunteers is also regularly re-evaluated depending on their performance and behavior while monitoring spectrum. For this purpose, the Volunteer Selection Unit utilizes the information in the Volunteer Report Database of the Adjudication component and upgrades the trustworthiness of volunteers in the Dynamic Attribute database (as shown in Fig. 6).

II **Adjudication Component**:

When a potentially enforceable event has occurred, forensics must be performed to de-

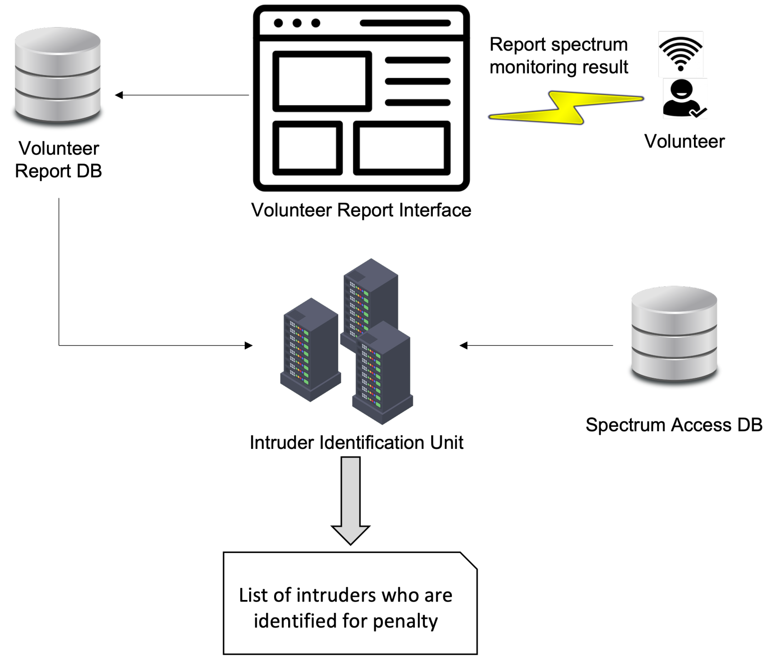

Volunteer Report DB
Volunteer Report Interface
Report spectrum monitoring result
Volunteer
Intruder Identification Unit
Spectrum Access DB
List of intruders who are identified for penalty

Figure 7: Functionality of the Adjudication Component in the Access Enforcement Computational Infrastructure

termine if a claim can be adjudicated and to build evidence in support of an adjudication process. Thus, one of the overarching objectives of the Access Enforcement Computational Infrastructure is building an information base for adjudication so that the event can be resolved and the costs appropriately internalized to the actors involved. Anderson et.al. write "spectrum forensic systems are designed to isolate an interference source, to provide evidence of interference and to do so rapidly, inexpensively and without burdensome complexity" [129]. The process of spectrum forensics consists of identifying the source of the interference, identifying the time and location of the interference, and assessing the impact of the event. Identifying the source of the interference may involve the individual radio or alternatively the operator/service provider. Other information that could prove valuable to an (automated) adjudication system may be the related actions, decisions, results, entries, etc. from the SCS as well as an after-the-fact statement of transmitter or receiver compliance. To this end, the Access Enforcement Computational Infrastructure maintains the Volunteer Report Database. In addition, there is an Intruder Identification Unit, which utilizes the information in this database along with information in the Spectrum Access database to identify and penalize the Spectrum Access intruders.

a **Volunteer Report Database**: This database maintains a record of all the reports submitted by volunteers while monitoring the spectrum. Volunteers are primarily responsible for detecting spectrum misuse. Numerous works have been proposed to detect misbehavior in wireless networks. Tang et al. [78] propose an adaptive approach for real-time selfish misbehavior detection in IEEE 802.11-based wireless networks. A basic misbehavior detector is designed based on the non-parametric cumulative sum (CUSUM) test. An adaptive version augments the performance of the basic detector by utilizing the Markov Decision Process for faster and more efficient misuse detection. However, this model is not suitable for multi-hop networks and is not scalable because of the sequential monitoring of nodes by a single detector. Liu et al. [79] investigate the problem of detecting unauthorized spectrum usage in DSA networks. The detection of anomalous spectrum usage is formulated by using statistical significance testing. The authors propose two detection schemes based on the mobility of the authorized transmitter. The first scheme is based on the assumption that the Received Signal Strength (RSS) from a single transmitter1 decays approximately linearly with the logarithmic distance from the source, but it is no longer the case when the RSS is a sum of multiple transmitters at different locations. This scheme is used when the authorized user is mobile. The second scheme is based on the assumption that the transmitters at different locations will lead to different spatial distributions of the RSS and is used when the authorized transmitter is static. However, the authors assume that the unauthorized transmitters are non-colluding and that the authorized transmitters are always trustworthy.

b **Intruder Identification Unit**: As shown in fig.7, the Intruder Identification Unit utilizes the information in the Volunteer Report Database and the Spectrum Access Database to identify the spectrum intruders. The first step to localizing and identifying a transmitter is to identify its signal. In an ideal scenario, the regulator would opt for employing a PHY-layer authentication procedure for identification [84]. This procedure enables a receiver to swiftly discern between compliant and rogue transmitters, eliminating the need for unnecessary higher-layer processing. To make this method workable, it is essential for all Secondary User (SU) radios to include a mechanism for authenticating their waveforms and utilize tamper-resistant measures to prevent potential circumvention by malicious users [84]. PHY-layer authentication schemes can be intrinsic or extrinsic. Intrinsic schemes utilize the intrinsic characteristics of the transmitted signal as unique signatures to identify transmitters. These methods primarily include RF fingerprinting and electromagnetic signature identification [85–88]. However, such schemes are sensitive to environmental factors like interference, and temperature changes and are not as effective in ”real-world scenarios” [84].

Extrinsic identification schemes, on the other hand, enable authorized transmitters to embed authentication signal intrinsically to their message signal (like machine authentication code, digital signatures, spectrum permits [89]), which can then be decoded by the receiver [84]. Such techniques include PHY-layer watermarking [90] and transmitter authentication [91, 92]. However, such schemes involve the superposition of the message signal over the authentication signal [90]. This means that the signal-to-noise ratio (SNR) of one signal is compromised by the other [84]. Another drawback of this scheme is that the SNR of the received signal should be high so that the received authentication signal is decoded and demodulated successfully by the receiver [84]. To address these challenges, a scheme called ”blind transmitter identification” was developed to ”enable a regulator to uniquely identify (or authenticate) a transmitter under low SNR and high multipath fading conditions while not requiring the regulator to have complete knowledge of the PHY-layer transmission parameters” [84, 93]. The Intruder Identification Unit utilizes the volunteer report, along with the information contained in the Spectrum Access database to determine the unauthorized spectrum intruders.

3.3.2 Volunteer Attributes

After the recruitment of volunteers, only the most *qualified* of the available volunteers should be selected to monitor spectrum in the area of enforcement. This is essential for ensuring an accurate detection of all enforceable events consistently over prolonged periods of time. However, determining the qualification of volunteers requires assessing their capability, availability, and performance in the past to predict their behavior and performance in the future. To this end, it is necessary to maintain multiple attributes that reflect the characteristics of a volunteer. As shown in Fig. 8, these volunteer attributes are primarily divided into two types - static and dynamic. Static attributes are established during the process of volunteer registration to the DSA Enforcement Infrastructure. On the contrary, dynamic volunteer attributes are periodically evaluated during run-time. Dynamic volunteer attributes are further divided into situational attributes and behavioral attributes. The remaining portion of this chapter is primarily dedicated to the discussion of these attributes that are associated with volunteers.

I **Static Attributes**: Static attributes are associated with volunteers during their registration to the DSA Enforcement Infrastructure. In the DSA Enforcement Infrastructure, these attributes are stored in the Volunteer Registration Database, which is a part of the Volunteer Support Component. These attributes help in comprehending the basic information, capability and availability of a volunteer who is freshly recruited. They help in the process of selection of freshly recruited volunteers, to monitor spectrum. These attributes usually remain unchanged during the course of spectrum monitoring. However, there may be circumstances in which these attributes may be optionally updated in the future, by the volunteers. An example of such occurrence is the change in sensing device that is used by volunteers over the course of spectrum enforcement. In such cases, volunteers need to register the new sensing device to the DSA Enforcement Infrastructure. The static attributes associated with a volunteer consists of their identification information, hardware device capability, most likely ”home” and availability to monitor spectrum. They are discussed in further details below.

a **Identification Information**: The identification information of volunteers primarily

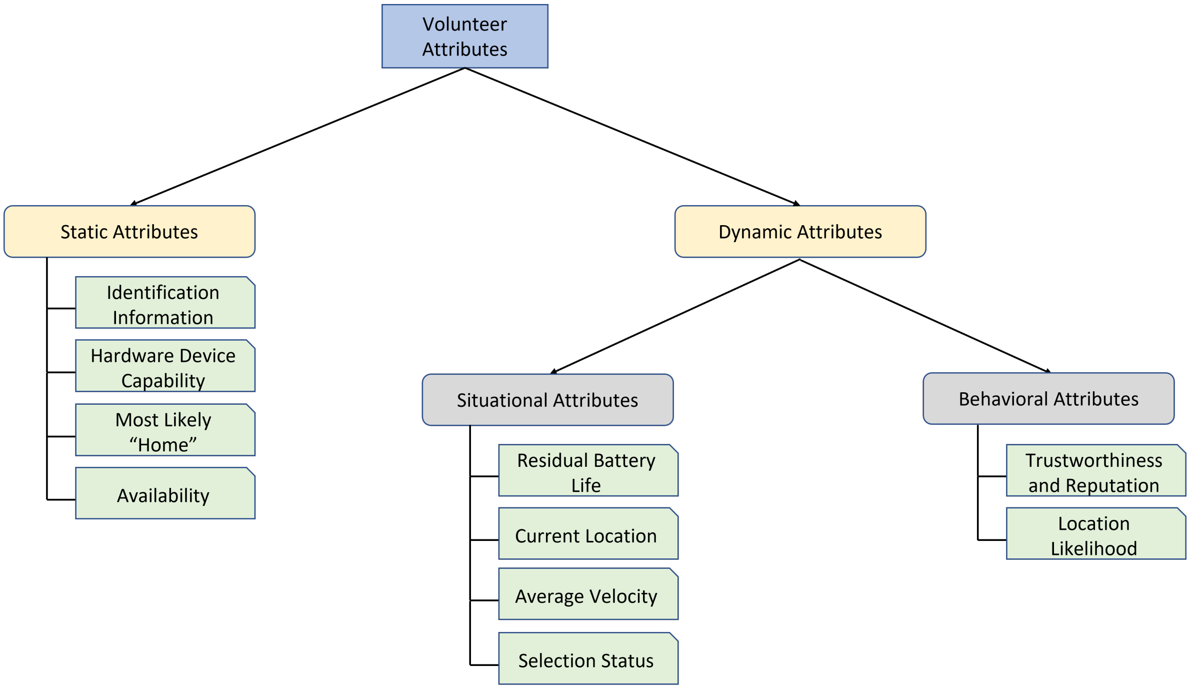


Figure 8: Classification of Volunteer Attributes

includes their full name and contact information, such as their email identity and phone number. It is, however, necessary to authenticate the identity of a volunteer to ensure security. Traditionally, in private and public networks, the process of authentication of user identity is usually Knowledge-based is completed through the use of a login user identification and a password or a personal identification number (PIN). Here the user's knowledge of the password or PIN is assumed to verify his/her identity. However, such systems are vulnerable, especially when the password/PIN are forgotten by a valid user or guessed by an impostor [130]. In some industries, like online banking, the use of user identification and password is sometimes sufficient because the users themselves would take extra precaution in not disclosing such information to others [131]. However, for registration of volunteers for spectrum enforcement, it is desirable to have additional levels of authentication to primarily prevent fraudulent spectrum users from performing the task of enforcement. Some of the alternative user authentication methodologies include:

- Token-based Approach: This approach utilizes the information in an item under a user's possession to verify the user identity. This includes information about the user's passport, driver's license, credit card, ID card, etc. However, such tokens may be lost, misplaced or stolen. Therefore, such approaches may suffer to ensure authentication of user identity [130].
- Biometric-based approach: User identification by using biometric utilizes the unique physiological and behavioral characteristics of a user for his/her identification. Such approaches include approaches such as face recognition, fingerprint recognition, Iris recognition, retina recognition, signature recognition, voice recognition, etc. This approach is usually more reliable than token-based and knowledge based approaches because many physiological and behavioral characteristics are usually uniquely associated with individuals [130, 132, 133]. Some of the major challenges that need to be addressed in a biometric recognition system include:
  - Biometric systems which utilize fingerprint recognition may suffer from being unable to successfully authenticate in the presence of dirt, cuts, physical wear and tear of the finger [134–136].
  - Numerous factors affect the performance of biometric systems which utilize face recognition for authentication. Such factors include light conditions, angles of face rotation, difference in facial expressions, etc [137, 138].
  - In systems that utilize signature recognition, an individual has the flexibility to change the signature when required. However, such authentication mechanisms may suffer from other drawbacks like evolution of signature with time and their likelihood of being influenced by physical and emotional conditions of the signatories. In addition, reproduction of signature by a professionl forger may fool a biometric system too [139].

  Some works have utilized deep learning for multimodal feature recognition to address these issues.

b **Hardware Device Capability**: Volunteers are required to register the devices that they will use for spectrum monitoring. The radio frequency (RF) sensor systems that are used to collect spectrum data share a common, general architecture. The basic architecture consists of an antenna, a preselector, a commercial-off-the-shelf (COTS) RF sensor and a measurement controller [140]. In such systems, the antenna type depends on several factors like the type of frequency band and service to be measured. The preselector connects the antenna to the RF sensor and is primarily responsible for reducing or rejecting the out-of-band and unwanted signals. The RF sensor is the primary component of the system and is primarily responsible for digitizing the RF signals received through the preselector to perform required "digital signal processing (DSP) functions to provide complex sampled data vs. time or frequency output data" [140]. The measurement controller is responsible for additional signal processing if required. Finally, the output data can be transmitted to a data staging server through the IP network [140].

The two usual types of RF sensors that are used for spectrum monitoring are swept-tuned and Discrete Fourier-Transform (DFT) based sensors. Swept tune sensors or the superheterodyne swept-based spectrum analyzers have the advantage of being able to perform very wide scan spans, extending several GHz. Laboratory quality swept tune sensors have been traditionally used for measuring short term spectrum usage and are known for accuracy in measuring radio spectrum. However, they are usually very expensive. On the other hand, DFT-based sensors are a special type of Software Defined Radio (SDR) receiver. This type of sensor digitizes the received signal and processes it using programmable software [140]. The National Telecommunications and Information Administration (NTIA) divides the types of spectrum monitoring sensors to four tiers — *high-tier*, *mid-tier*, *low-tier* and *very low-tier* [140]. The *high tier* sensors include the very expensive ($25,000 or more) laboratory grade spectrum monitoring devices like the spectrum analyzers and real-time analyzers. The *mid-tier* sensors are DFT-based sensors and are designed to be deployed outdoors, with cost in the range of $15,000 - $20,000. The low-tier sensors are development SDRs which are not waterproof and not meant for outdoor use. They usually cost in the range of $1000 to $5000. The very low-tier sensors include SDR receivers which are cheaper versions of the low-tier sensors [140]. A typical example of such sensors is a RTL-SDR receiver [45].

The high cost of at least the high-tier and mid-tier sensors makes it infeasible for the volunteers to use these devices for spectrum monitoring. Since we use crowdsourcing, we can take advantage of the preponderance of the crowsourced volunteers. This makes the use of inexpensive *very low-tier* sensors for spectrum monitoring a viable option too. In [45, 141], Nika et al. utilizes a prototype that connects smartphone with an external RTL-SDR dongle via USB port to monitor spectrum in the range of 52-2200 MHz. In this setup, the RTL-SDR sensor collects spectrum usage signals and the smartphone perform the task of data processing. The primary drawback of such systems is the limited sensitivity and bandwidth when compared to more expensive options. For example, the proposed RTL sensors usually have a bandwidth of 2.4 MHz when to 20 MHz of traditional USRP devices. This can, however, be addressed by taking more samples [45].

Therefore, while registering the devices to the DSA Enforcement Infrastructure, the volunteers should provide relevant information about the spectrum monitoring device that they will be using. Such information include:

- Device name and type.
- Frequency range that can be covered by the device.
- Maximum sample rate of the device.
- Serial number of the device used.

These relevant information regarding the spectrum monitoring device used by a volunteer is beneficial in making decisions regarding selection of volunteers and their assignment to monitor channels.

c **Most Likely ”Home”**: The most likely ”home” of a volunteer is usually the geographical location where the volunteer is likely to reside for the maximum amount of time. In real life, the most likely ”home” of a volunteer may be their house of residence or simply the location at which the volunteer expects to reside in for the majority of time. This information is useful because it can be used to assign a volunteer to monitor channels in the geographical area closer to their most likely ”home” location, when required.

d **Availability:** Availability of a volunteer includes the time slots in which the volunteer would be available to monitor spectrum. This is especially useful to know about the likely itinerary of a volunteer, which can be later utilized for selecting volunteers at time slots in which they are available to monitor spectrum.

II **Dynamic Attributes**: Dynamic attributes are associated with volunteers based on the information that is collected from volunteers during run-time. These attributes assist in comprehending the changing characteristics of a volunteer, like their location, sensing device's residual battery life, performance and behavior. Unlike static volunteer attributes, these attributes are updated regularly to estimate the most recent characteristics of a volunteer. Dynamic Attributes are primarily divided into two major types: situational attributes and behavioral attributes. Situational attributes are associated with volunteers on the basis of run-time information that is shared by them regularly, after recruitment. Such attributes include the current geographical location, residual battery life, average velocity, and the selection status of a volunteer. Behavioral attributes of volunteers are dynamically evaluated based on their behavior and performance. Such attributes include the likelihood of a volunteer to be in a geographical region, and the trustworthiness and reputation of volunteers.

i **Situational Attributes**: These attributes are associated with volunteers and are based on the information that they provide to the DSA Enforcement Infrastructure periodically. They primarily reflect the features that define a volunteer's dynamic state of affairs in a given time. Such attributes include:

(a) **Residual Battery Life**: The current State of Charge (SOC) of the battery in a spectrum sensing device used by a volunteer can be used to predict residual battery lifetime for a volunteer to monitor channels before the sensing device runs out of power. Recruiting a volunteer with a low estimated residual battery life is not desirable to the DSA Enforcement Infrastructure. This is because rejecting such volunteers will avoid the possibility of having their spectrum sensing device run out of power before the end of a monitoring epoch. To this end, volunteers must periodically report the SOC of their sensing device's battery, to the DSA Enforcement infrastructure.

Estimating the residual lifetime of a battery may depend on different factors. Some of these factors include the software used in the device, the user's usage pattern of the device, and the hardware specifications of the sensing device itself. Several works have been done to predict the residual battery life. The works in [142, 143] have focused on hardware and operating system level of battery life prediction. Such works primarily focus on utilizing the remaining battery capacity and the present rate of battery drain to predict the remaining battery life. Another area of related research is in model-based prediction of residual battery life. To this end, Doyle et al. have focused on a model-based approach for battery life prediction [144] by utilizing the discharge profile of the entire lifetime of the battery. The works in [145–147] have focused on the study of battery power consumption from the point of view of processor instructions. Finally, Kang et.al. have focused on predicting the residual battery life of a mobile device based on the usage patterns of individual users [148]. This work performs on-line analysis of the usage patterns of mobile devices, such as "the battery consumption rate when making voice calls, using data communication, or waiting for calls", to predict the residual battery life.

(b) **Current geographical location**: A volunteer needs to enable location sharing in their mobile devices to ensure that their location is collected periodically by the DSA enforcement infrastructure. Sharing of geographical location of a device is usually enabled by the GPS-capabilities in the device. GPS, or the Global Positioning System, is a satellite-based radio navigation system that provides users with Positioning, Navigation and Timing services [149]. Other such navigation systems around the world include the *NavIC* developed by India [150], *BeiDou* devloped by China [151], *GLONASS* developed by Russia and *Galileo* developed by the European Union [152]. However, such navigation systems may occasionally suffer failure due to jamming and interference [153]. To this end, Neinavaie et. al discuss about a substitute of GPS, which utilize the more abundant satellites on Earth's lower orbit, like the Starlink fleet. Hence, such systems ensure that the signals are received at a higher power, thereby making it more immune to jamming attacks and spoofs, than systems such as GPS and GLONASS [154]. However, continuous sharing of location may give rise to privacy concerns for the volunteers [155]. To this end, Michalevsky et al. utilize just the measured power profile of cellular devices as a time series data to determine the location and route taken by users [156]. Regular update about the geographical location of volunteers is essential because it helps in determining the likelihood of a volunteer to reside in a geographical region of enforcement.

(c) **Average Velocity**: It is necessary for the volunteers to submit their average velocity periodically because it allows the Volunteer Selection Unit to a) estimate the coverage that can be provided by a volunteer over a spectrum monitoring epoch and b) estimate the likelihood of a volunteer to be in a geographical region over a spectrum monitoring epoch. However, the average velocity of a person may not always be a trivial job to determine. In [157], Ji and Pachi investigate the stepping frequency and velocity of people walking by using statistical methods. Their results indicate that the velocity and stepping frequency for walking varies across genders and also across the venue. Hence, in such scenarios, where it is difficult to estimate the average velocity, volunteers can only report their activity and an estimated value will be used for the average velocity of volunteers in such instances.

(d) **Selection Status**: Volunteers are selected to monitor spectrum periodically at the beginning of every spectrum monitoring epoch (Chapter 4). However, not all volunteers are selected to monitor spectrum at every monitoring epoch. It is necessary to select volunteers who are estimated to do the job spectrum monitoring reliably and efficiently. Such volunteers who are selected to monitor spectrum at a monitoring epoch have an *Active* status. Otherwise, volunteers maintain a *Standby* status until they are selected again to monitor spectrum. Maintaining a status for volunteers is especially essential when a new pool of volunteers need to be ”patched” in to replace volunteers who are evicted for poor performance. This is done to efficiently continue with the process of monitoring spectrum before a spectrum monitoring epoch ends. A more detailed discussion of the volunteer selection mechanism is explained in Chapter 4.

ii **Behavioral Attribute**s: These attributes are associated with volunteers based on their behavior and performance in monitoring spectrum.

(a) **Trustworthiness and Reputation**: The effectiveness of crowdsourced ex-post spectrum monitoring relies on accurate and reliable data reported by volunteers after spectrum monitoring. If the volunteers are not trustworthy, the data they provide may be inaccurate or manipulated, leading to unreliable information about spectrum usage. Trustworthy volunteers facilitate effective monitoring of the spectrum and ensure that the collected data reflects the actual state of the spectrum. To this end, the Volunteer Selection Unit of the Access Enforcement Computational Infrastructure is aided by mobile sentinels in calculating the trustworthiness of volunteers based on their spectrum monitoring performance. This information is further used to build a reputation profile of volunteers, which in turn facilitates robust spectrum monitoring.

(b) **Location Likelihood**: Selecting volunteers to monitor the spectrum in an enforcement region where they are unlikely to visit is undesirable. Therefore, it is necessary to estimate the future location of volunteers before assigning them to monitor the spectrum in a region over a monitoring epoch. However, predicting future locations of volunteers is complex due to diverse mobility patterns and evolving intentions. Variations in mobility based on time add complexity, while uncertainty arises from individuals changing plans or deviating from routines too. Addressing these challenges is crucial for accurate location prediction of volunteers. To this end, a mobility model is developed in Chapter 6 using AI-based methodologies to predict the future location of volunteers.

3.4 Summary

In this chapter, the infrastructure and system model for effective ex-post enforcement of access rights in shared spectrum networks is discussed. The enforcement area is divided into regions where access rights are enforced according to the underlying principles. Enforcement of spectrum access rights is ensured by a shared spectrum access enforcement infrastructure which is built around a cloud-based centralized Spectrum Control System (SCS) and crowdsourced spectrum monitoring volunteers. The SCS consists of computational servers and databases that support volunteer registration and selection for spectrum monitoring. The following chapter builds upon this infrastructure and discusses the various strategies adopted to ensure effective volunteer selection.

4.0 Spectrum Monitoring Volunteer Selection

Volunteer selection is a basic component of the proposed ex-post spectrum access right enforcement framework. The main objective of the selection process is to determine, among a slate of registered candidates, the most *qualified* volunteers to perform reliably the shared-spectrum monitoring tasks over a shared-spectrum access region. To maximize coverage, the shared-spectrum access region is divided into non-overlapping coverage zones.

The spectrum monitoring process, within a coverage zone, takes place over successive time periods, referred to as *Monitoring Intervals (MI).* The selection must, therefore, assess candidates' monitoring *qualifications* to narrow down the pool of applicants until the required number of volunteers to cover a zone over an entire MI is reached. The required number of active volunteers is determined based on a greedy approach, which seeks to find the minimum number of *qualified* volunteers who, collectively, can carry out monitoring tasks over the targeted region. The selected volunteers are then assigned to actively engage in the monitoring of the spectrum, at the beginning of the MI, and report on the spectrum state in a timely manner. It is to be noted, however, that no matter how strategic the selection process, volunteer no-shows is likely to happen, as volunteers offer their assistance in exchange for no financial value. There can be multiple factors, some intentional and others unplanned, that prevent a volunteer from showing up. It is also the case that a *corrupt* volunteer, who does not adhere to the monitoring rules or fails to report truthfully on the state of the spectrum, is disqualified and immediately discharged. In order to overcome the occurrence of such events, the selection process must also identify a set of *standby* volunteers who may be called upon to replace a no-show or a disqualified volunteer.

In the first section of this chapter, the structure of the time interval is discussed. The proposed structure is designed to enable sustainable and robust assignment of volunteers to cover regions. It also facilitates accurate monitoring of the spectrum and reliable detection of spectrum access right violations, when they occur. From among all registered candidates, the most *qualified* volunteers are selected to monitor effectively the spectrum, in the geographical enforcement area under consideration. Given a target number of monitoring volunteers, an effective volunteer selection strategy must be designed to ensure reliable and sustainable spectrum monitoring over a MI. To this end, critical aspects of a criteria-based selection process include thoroughly assessing volunteers' *qualifications* and identifying specific determining factors. Among the different determinants, trust-based reputation and the likelihood of residing in the targeted region over the MI have been recognized as two of the most important factors. A weighted combination of these factors is used to assess the *qualification* of a volunteer to accomplish the monitoring task. The next section of the chapter discusses the basic properties an algorithm may have to achieve effective volunteer selection. Such an algorithm must be designed to ensure high geographical coverage, high spectrum access violation accuracy, and consistency in detecting spectrum access violations. It is also desirable that the algorithm accommodates volunteer's specific requests, when possible. To meet these requirements, combinations of variants of a Secretary-based algorithm, an online algorithm that attempts to optimize the probability of selecting the most *qualified* volunteers, and variants of the stable matching algorithms are used to develop two hybrid algorithms that ensure better coverage and higher rate of violation detection accuracy than what would have been achieved using only a variant of the Secretary-based algorithm or the stable matching algorithm [53]. These algorithms are described and discussed. The last section of the chapter discusses the experimental setup used to assess the performance of these algorithms and presents a comparative performance analysis of these algorithms. CSIM is used to simulate these algorithms. The results of the experiments indicate that a hybrid volunteer selection approach, which combines the secretary algorithm with a variant of the stable matching algorithm, achieves the most effective volunteer selection.

4.1 Spectrum Enforcement Time Interval Structure

As shown in Figure 9, shared-spectrum access monitoring takes place over consecutive MIs. Associated with each MI, for each zone, is a set of active volunteers and set of standby volunteers. The association of volunteers to monitored regions is carried out anew, at the beginning of each MI, by the Volunteer Selection Unit of the Access Enforcement Computational Infrastructure.

To enable effective and accurate monitoring of the spectrum, each MI is divided into sub-intervals, referred to as *Working Unit Intervals* (WUIs). A WUI represents the smallest time interval over which a spectrum user can accomplish useful, thus detectable, work [47, 48, 51, 53, 54]. A WUI is further divided into Access Slots, over which a user (authorized or unauthorized) can access a channel. In an access slot (AS), Sentinels and volunteers can also sense the spectrum to determine the type of ongoing access, legitimate or intrusive, and report any access violation, when it occurs.

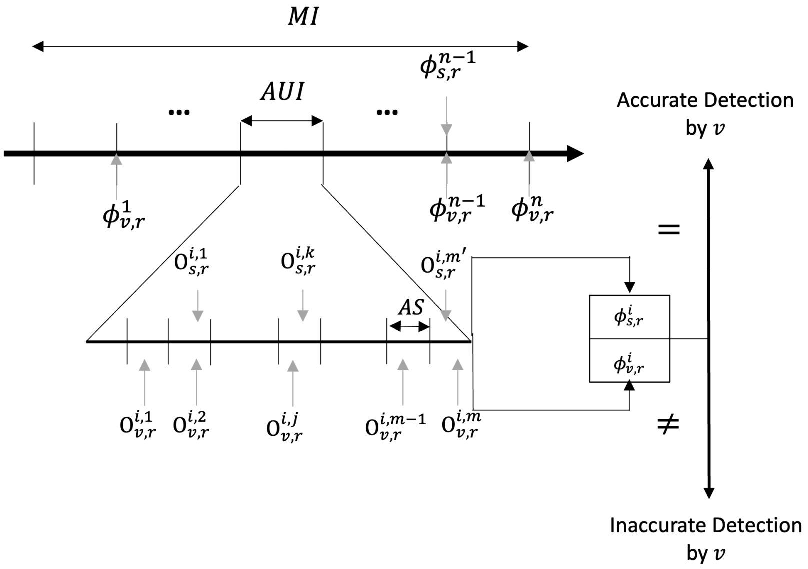


Figure 9: Decisions by volunteer v after every WUI and by sentinel s after random WUIs, for a given MI.

4.2 Volunteer Qualification

A central concern in the design of the shared-spectrum monitoring framework is the selection, throughout the monitoring process, of a group of highly qualified volunteers to reliably monitor spectrum activities, in a cooperative and timely manner. Candidates express their willingness to become a shared-spectrum monitor through the volunteer registration process. The process allows candidates to submit securely an application, which includes personal identification information and a volunteer profile, detailing the monitoring computational and communications capabilities of the applicant. Changes and updates to this information can be made anytime, throughout the monitoring process.

Given a slate of potential candidates, vital factors must be considered in the selection process to identify the candidates who are the best fit for monitoring, comprehensively and reliably, the shared spectrum in all regions of the monitoring area. The selection factors encompass aspects related to character traits, in terms of reliability, dependability, and integrity, ability traits, in terms of monitoring computing and communications capabilities, and spatiotemporal circumstances, in terms of the volunteer's physical location and temporal availability to monitor spectrum, for the prescribed time interval, in a given region of the coverage area.

Several disciplines in the area of social sciences have been interested, partially or wholly, in these factors, when studying collective behavior and human interactions and the interplay between cooperation, coordination, and delegation in human organizations [158–163]. Recently, social networking has emerged as a primary medium for people, across the globe, to share information and express opinions, interact virtually with each other, and collaborate towards accomplishing a myriad of tasks in various domains. The susceptibility of social networking to an increasingly larger number and more complex security threats, gave rise to an increased interest in the computing and communications communities to explore new approaches, that go beyond legacy security schemes, to ensure trustworthiness, integrity, and authenticity in social networking platforms [164–171].

To assess volunteers' qualifications to perform shared-spectrum monitoring, the following determinant factors are adopted in this thesis:

1. Trustworthiness, which is defined as the ability of a volunteer to be relied on in monitoring the shared spectrum, in a truthful manner. Trust is built over time, through repeated monitoring experiences and reliable reporting of the shared spectrum status.
2. Reputation, which is defined as the estimation in which a volunteer is held, based on their monitoring reporting record. Trustworthiness drives reputation. As such, volunteers use trust as the main determinant to build and sustain confidence in their reputations.
3. Spatiotemporal availability, which is defined as the likelihood of a volunteer performing shared-spectrum monitoring tasks, at the specified monitoring region over the prescribed monitoring time interval.

Combined, the above factors determine the dispositional qualification of a volunteer to carry out spectrum monitoring, in a reliable and timely manner. Such a qualification reflects the ability, capacity, and aptitude of a volunteer to monitor the shared spectrum in real time and report truthfully the status of the communication channel. It can be viewed as a measure of the predicted belief that the volunteer will be capable of performing the required monitoring task, in the specified time and the targeted region, reliably and timely. Volunteer qualification is calculated and updated at the end of every monitoring interval. Based on the updated qualification measure, a new set of volunteers is selected to monitor the spectrum in the upcoming Monitoring Interval.

In the following section, the concept of reputation, including the role trust plays in determining a volunteer's reputation, and the concept of spatiotemporal availability are discussed. The section concludes by showing how these factors are used as the dispositional qualification measure of a volunteer to perform shared spectrum monitoring, at the targeted region over the prescribed monitoring time interval, in a reliable and timely manner.

4.2.0.1 Reputation

The reputation of a volunteer is based on their ability to monitor the spectrum in a trustworthy manner. Volunteer reputation is built over time depending on how trustworthy a volunteer is in determining the state of a channel. The reputation of a volunteer relies on the trust associated with the volunteer. The trust of a volunteer, v, is determined by its past behavior. The behavior of a volunteer v is chiefly determined by its accuracy in the detection and reporting of spectrum access violations. As shown in Figure 7, a volunteer,v, in region r makes an observation $O^{i,j}_{v,r,c}$ of the access state of channel c in every AS j and a sentinel s makes an observation $O^{i,k}_{s,r,c}$ at a random AS k of a WUI i. Based on these observations, both v and s decide the spectrum access state over WUI i. We assume that a volunteer v's decision $\theta^{i}_{v,r,c}$ is accurate if it is the same as the decision $\theta^{i}_{v,r,c}$ of sentinel s. The trustworthiness of a volunteer is determined by their accuracy in the detection of spectrum access violation as given by (5).

$$T_{v,r,c} = \sum_{MI} \frac{s_{v,r,c}(MI)}{s_{v,r,c}(MI) + f_{v,r,c}(MI)} \tag{5}$$

In equation (5) $s_{v,r,c}(MI)$ and $f_{v,r,c}(MI)$ are the number of times that the decisions of v and s matched and did not match for channel c in r respectively, over a given MI [47, 48, 51]. A sentinel s decides to monitor c at random WUIs to verify the decisions made by the volunteers. The minimum number of observations η required by a sentinel in a WUI to determine the ground truth of spectrum access state with a margin of error δ at β% confidence level and critical value z^* is given by (6).

$$\eta \geq 0.25.(z^*/\delta)^2 \tag{6}$$

The reputation $\Gamma_{v,r,c}$ of a volunteer v in r for channel c is established based on the volunteer's trustworthiness over an extended duration. The tenet of our approach is to increase the reputation slowly after success and penalize the reputation rapidly after it falls below a threshold. The reputation $\Gamma^{z,i+1}_{v,r,c}$ of a volunteer v at the beginning of WUI $i+1$ of MI z for monitoring channel c in r is given by (7).

$$\Gamma^{z,i+1}_{v,r,c} = \begin{cases} \Gamma^{z,i}_{v,r,c} + f(T^{z,i}_{v,r,c}), & \text{if accurate} \\ \Gamma^{z,i}_{v,r,c} - g(T^{z,i}_{v,r,c}), & \text{otherwise} \end{cases} \tag{7}$$

where $T^{z,i}_{v,r,c}$ is the trustworthiness of v for monitoring channel c in r after it makes a decision in WUI i of MI z, $f(T^{z,i}_{v,r,c}) = \kappa.T^{z,i}_{v,r,c}$ (where κ is a system parameter) and $g(T^{z,i}_{v,r,c}) = e^{\lambda.(1-T^{z,i}_{v,r,c})}$, such that λ increases if $\Gamma^{z,i}_{v,r,c} < \zeta$, where ζ is the threshold below which we decrease reputation more rapidly. So, the reputation is increased linearly when an accurate decision is made by v and decreased exponentially otherwise [51].

4.2.0.2 Proportion of Residence Time

A second critical attribute in determining volunteer qualification is their likelihood to reside in a region for which they are candidates to be selected to monitor the spectrum. This is achieved by evaluating the estimated duration of a volunteer to reach a target region, r, the proportion of time that a volunteer resides in r compared to other regions, and the estimated sojourn time of a volunteer in r to determine the likelihood of a volunteer to reside in r over a Monitoring Interval. In the following, the different parameters for measuring volunteer qualification are discussed in detail. Volunteers who have a high likelihood of residing in a region r are preferred to monitor r.

The proportion of residence time $\rho_v(r)$ of a volunteer v in r is given by (8).

$$\rho_v(r) = \frac{\tau_v(r)}{\sum_{r \in R} \tau_v(r)} \tag{8}$$

where $\tau_v(r)$ is the total time spent by v in r.

4.2.0.3 Duration to Destination

Volunteers who are likely to reach a region r in the shortest duration are preferred to monitor r. At any given time t, the location L^t_v of volunteer v enables us to estimate the shortest duration $\Upsilon^t_v(r)$ needed by v to reach a region r, as shown in (9).

$$\Upsilon_v^t(r) = \frac{\gamma.d(L_v^t, \mathcal{O}_r)}{\tilde{\mu_v}} \tag{9}$$

where $\gamma > 0$ is a system parameter, $\tilde{\mu_v}$ is v's average velocity, $\mathcal{O}_r$ is the centroid of region r and $d(L_v^t, \mathcal{O}_r)$ is the shortest distance between L_v^t and $\mathcal{O}_r$.

4.2.0.4 Sojourn Time

Volunteers who are most likely to reside a major proportion of time in r after a visit to r are preferred. To this end, we estimate the sojourn time of a volunteer $v \in V$ in $r \in R$ after every visit of v to r. After the j^{th} visit of v to r, we measure its $(j-1)^{th}$ sojourn time, $S_v^{j-1}(r)$ as the difference between its $(j-1)^{th}$ departure time, $dep_v^{(j-1)}(r)$ from r and its $(j-1)^{th}$ arrival time, $arr_v^{(j-1)}(r)$ in r. Based on this information, the proportion of time that v is likely to stay in r before its j^{th} departure from r is estimated as an exponentially smoothed average, given by (10).

$$\tilde{S}_v^j(r) = \alpha.S_v^{j-1}(r) + (1-\alpha).\tilde{S}_v^{j-1}(r) \tag{10}$$

$$\alpha = h.(E_v^{j-1}(r))^2/\sigma_v^j(r) \tag{11}$$

where $0 < h < 1$, $E_v^{j-1}(r) = S_v^{j-1}(r) - \tilde{S}_v^{j-1}(r)$ is the prediction error on visit j, and $\sigma_v^j(r)$ is the average of the past square prediction errors, as shown in (12).

$$\sigma_v^j(r) = h.(E_v^{j-1}(r))^2 + (1-h).\sigma_v^{j-1}(r) \tag{12}$$

4.2.1 Formulation of Volunteer Qualification

The Volunteer Selection Unit selects up to k *qualified* volunteers to monitor R at the beginning of every MI. This is determined by the estimated Qualification $Q_{v,r,c}(MI)$ of a volunteer v to monitor a channel c in $r \in R$ over the next MI, given by (13), defined below.

$$Q_{v,r,c}(MI) = f(\Gamma_{v,r,c}, \tilde{S}_v^j(r), \Upsilon_v^t(r), \rho_v(r)) \tag{13}$$

We normalize $\Gamma_{v,r,c}, \tilde{S}_v^j(r), \Upsilon_v^t(r), \rho_v(r)$ by using the min-max normalization technique [172] such that $0 \leq \Gamma_{v,r,c}, \tilde{S}_v^j(r), \Upsilon_v^t(r), \rho_v(r) \leq 1$. Clearly, reputation is the most significant component of the selection metric. Untrustworthy volunteers, regardless of their location or likelihood of being in a region, must be eliminated. Furthermore, since a volunteer can only monitor a channel in a region that it resides in over a WUI, the proportion of residence time $\rho_v(r)$ of v in r is next in priority. With this assumption, we explore ways to aggregate the above four parameters in f in order to assess their impact in measuring the qualification of a volunteer as shown in f1f4.

$$f_1 = p_0.(\log(1+p_1).\log(1+p_2)) \tag{14}$$

$$f_2 = p_0.(\log(1+p_1)+\log(1+p_2)) \tag{15}$$

$$f_2 = p_0.(\frac{w_1}{w_1+w_2}p_1 + \frac{w_2}{w_1+w_2}p_2) \tag{16}$$

$$f_2 = p_0.(\frac{w_1}{w_1+w_2}p_1.\frac{w_2}{w_1+w_2}p_2) \tag{17}$$

In the above equations, we assume that $p_0 = \rho_v(r).e^{\beta.\Gamma_{v,r,c}}$ where $\beta > 0$, $p_1 = 1 - \Upsilon_v^t(r)$ and $p_2 = \tilde{S}_v^j(r)$, w_1 and w_2 are the weights associated with p_1 and p_2 respectively. We define p_1 as such because lower duration to destination $\Upsilon_v^t(r)$ is preferred, unlike the other three parameters. We observe that reputation $\Gamma_{v,r,c}$, being the dominating factor, exponentially impacts the qualification $Q_{v,r,c}(MI)$, while the parameter $\rho_v(r)$ is multiplied linearly to it. In (14) and (15), parameters $\tilde{S}_v^j(r)$ and $\Upsilon_v^t(r)$ are used logarithmically and thus have a sub-linear impact. $\tilde{S}_v^j(r)$ and $\Upsilon_v^t(r)$ are aggregated by addition and multiplication in (14) and (15) respectively. On the contrary, in (16) and (17), parameters $\tilde{S}_v^j(r)$ and $\Upsilon_v^t(r)$ are combined in a weighted linear manner and aggregated by addition and multiplication respectively.

4.3 Volunteer Selection Methodology

An efficient methodology for selection of mobile volunteers to monitor spectrum is crucial for more efficient ex post spectrum enforcement. In order to design the selection methodology, we define the following metrics/criteria.

1. *Coverage*: Volunteers should be selected such that they provide spatial coverage of the enforcement area and coverage of channels in the area. *Good* spatial coverage ensures that the volunteers selected are not clustered in a small area. This ensures that the malicious users can be localized after they violate the terms of spectrum access. Good channel coverage ensures that a high number of vulnerable channels are monitored consistently to detect spectrum misuse.
2. *Accuracy of Detection*: The accuracy of detection of a volunteer v depends on several factors like the reliability of v, the Received Signal Strength of v, the transmit power of v and the device characteristics of v. Higher accuracy of detection by the Volunteers who are selected for spectrum enforcement is preferred.
3. *Consistency*: It is required to ensure that there is consistent coverage of the enforcement area over any duration of enforcement. To this, we ensure that a new set of volunteers is selected after a certain epoch based on the renewed calculation of volunteer qualification.
4. *Mutual Satisfaction*: Apart from ensuring that there is consistent and efficient enforcement of spectrum access, we propose to develop a volunteer selection methodology that ensures high volunteer satisfaction along with the satisfaction of the enforcement platform.

The above criteria are considered when designing a selection methodology to achieve the objectives of the proposed framework. It is unlikely that a single volunteer selection algorithm can achieve all the requirements. Thus, the feasible strategy is to strike a balance that approximates as closely as possible to my stated objectives.

4.4 Overall Volunteer Selection Strategy

As discussed in Chapter 3, the recruitment of volunteers is addressed by the Volunteer Recruitment Infrastructure. In this infrastructure, the Volunteer Selection Unit primarily deals with selection of volunteers for monitoring spectrum in every Monitoring Interval. To this end, the Volunteer Selection Unit utilizes the information contained in the Volunteer Registration Database for selecting the first batch of volunteers to monitor spectrum at the commencement of the first Monitoring interval. Thereafter, in every other Monitoring Intervals, it utilizes the information in both the Volunteer Registration Database and the Behavioral Attribute Database to select volunteers. A Monitoring Interval may be large. This is because it costs time to deploy volunteers and sentinels, monitoring the spectrum channels and establishing the state of the channels, before an actual decision can be made about the spectrum access state of all the channels, the behavior and performance of volunteers, and the overall effectiveness of the spectrum enforcement scheme. However, there may arise a circumstance in which some of the deployed volunteers perform poorly in a consistent manner. This may be because of environmental factors like low signal strength due to attenuation or behavioral factors like collusion by a volunteer with spectrum access intruders. In such cases, the poor performing volunteers need to be replaced before the Monitoring Interval ends. To this end, two basic strategies are employed as follows:

- **Maintaining *Active* and *Standby* Volunteers**: Volunteers are selected to monitor spectrum periodically at the beginning of every spectrum monitoring epoch (Chapter 4). However, not all volunteers are selected to monitor spectrum at every monitoring interval. It is necessary to select volunteers who are estimated to do the job of spectrum monitoring reliably and efficiently. Such volunteers who are selected to monitor the spectrum at the beginning of a monitoring interval have an *Active* status. Otherwise, volunteers maintain a *Standby* status until they are selected again to monitor spectrum.

Maintaining a status for volunteers is especially essential when a new pool of volunteers need to be "patched" in to replace volunteers who are evicted for poor performance. This is done to efficiently continue with the process of monitoring spectrum before a spectrum monitoring epoch ends.

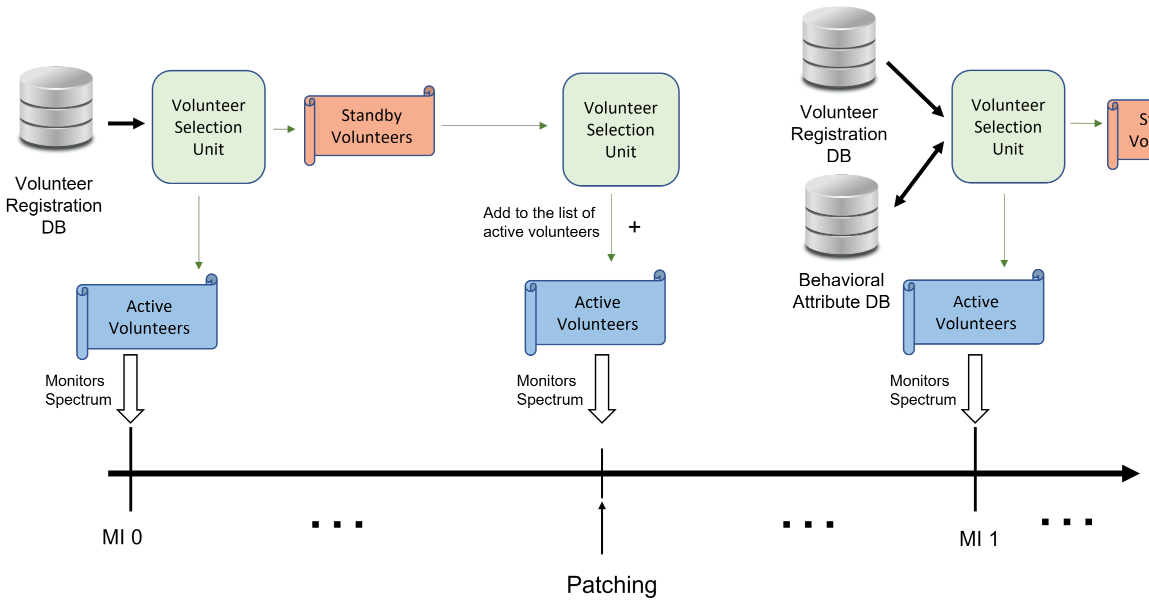


Figure 10: Sample of the volunteer selection strategy for the initial two Monitoring Intervals

- **Patching**: Patching refers to the strategy of making amends to the list of active volunteers in a monitoring interval. If the performance of some active volunteers falls below a threshold, then they are evicted from monitoring spectrum and a new set of volunteers is selected from the set of standby volunteer for monitoring spectrum in rest of the monitoring interval. This ensures that the poor enforcement in the elapsed portion of the Monitoring Interval can be compensated by the spectrum monitoring performed by these newly selected volunteers in rest of the monitoring interval.

4.5 Number of Volunteers Required For Selection To Monitor Spectrum In an Enforcement Region

A shared spectrum network that is composed of several wireless devices and access points services a geographical area. Spectrum is monitored by crowdsourced volunteers in this geographical area to enforce the rights to access spectrum. As discussed in chapter 3, a geographical area of spectrum access enforcement is divided into regions using an approach based on Voronoi and Lloyd's algorithms. To guarantee robust spectrum monitoring, an "optimal" coverage of each region is required. Achieving an "optimal" coverage depends on the volunteers' ability to monitor a specific portion of the geographical area. Based on the capability of the spectrum sensing device used by a volunteer, a coverage zone is a sub-area within a region and is considered to be the unit of coverage by a volunteer in the geographical area of enforcement.

Assuming a free space path [173] and isotropic antennas of the spectrum sensing devices, the average distance, $\overline{d}$, before which the received power reduces below threshold ϕ is given by (18)

$$\overline{d} = \sqrt{\frac{\overline{P_t}}{\phi} \cdot \left(\frac{c^2}{(4\pi f)^2} \right)} \tag{18}$$

In (18),

- $\overline{P_t}$: Average Transmitted power of spectrum sensing device of all volunteers.
- ϕ: Specified threshold power level (received power should not fall below this level).
- c: Speed of light in vacuum.
- f: Frequency of the transmitted signal.

Based on $\overline{d}$, a region can be divided into square-shaped zones, such that the circumscribed circle of the square has radius of length $\overline{d}$. As shown in figure 11, each region is partitioned into coverage zones. While the area of coverage is represented by a circle with radius $\overline{d}$ (which is calculated in (18)), the zone of coverage is its circumscribing square.

Considering that the antennas are isotropic, the average area of coverage, A, of a spectrum sensing device can be given by (19).

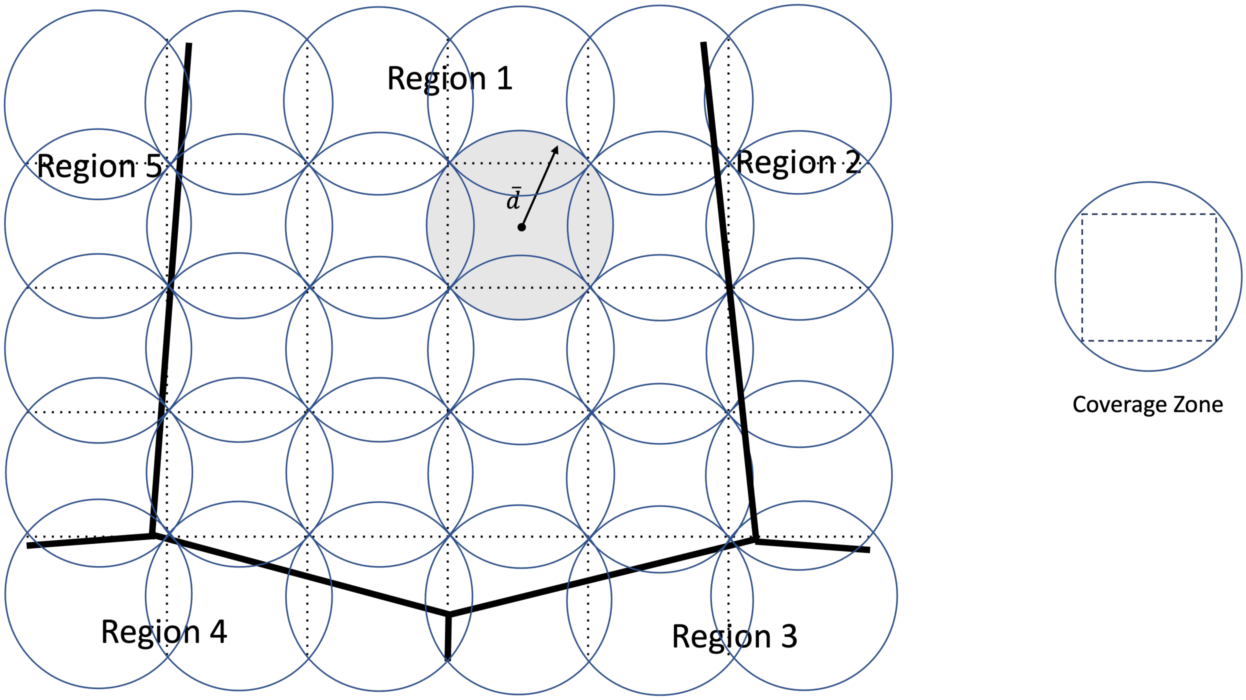


Figure 11: Coverage Zones in A Region of Enforcement

$$A = \pi \cdot (\overline{d})^2 \tag{19}$$

Therefore, a side of the circumscribing square (which is the coverage zone), is of length $\sqrt{2}.\overline{d}$, and its area is $2.\overline{d}^2$. Therefore, the number of coverage zones, $|Z|$, in an enforcement region, r, of surface area $\mathscr{R}$ is be given by (20)

$$|Z| = \lceil \frac{\mathscr{R}}{2.(\overline{d})^2} \rceil \tag{20}$$

It is highly unlikely for a volunteer to cover an entire region during a spectrum monitoring epoch. This stems from two primary reasons — 1) Limitation of the capability of the spectrum sensing device used by a volunteer, 2) A volunteer is mobile and may likely move out of a zone within an epoch.

To this end, it is necessary to select eligible volunteers for monitoring spectrum in a coverage zone. Given a coverage zone z, let $V_z = \{v_i | 1 \leq i \leq K_z\}$ be the set of eligible volunteers to cover zone z. A volunteer, v_i is eligible to cover zone z iff

1. Volunteer, v_i, is present in z.
2. The coverage radius, d_i, provided by v_i is greater than or equal to $\overline{d}$.

As shown in (18), it is known that the geographical coverage of a volunteer, v, depends on the transmitted power of the spectrum sensing device used by the volunteer. Therefore, the radius, d_v of the area covered by the sensing device of a volunteer is given by (21).

$$d_v = \sqrt{\frac{P_t^v}{\phi} \cdot \left(\frac{c^2}{(4\pi f^v)^2}\right)} \tag{21}$$

In (21), P_t^v represents the transmitted power of the sensing device used by volunteer v, f^v represents the frequency of the transmitted signal and c is the speed of light. Therefore, the area of geographical coverage that can be provided by a volunteer, v, is $\pi \cdot (d_v)^2$ or the area of a circle with radius d_v. Is the coverage provided by v enough for a coverage zone z? This depends on the value of d_v when compared to $\overline{d}$. In the worst case, if a volunteer, v, is on one of the corners of a square zone z, then d_v has to be at least $2.\overline{d}$ to cover z. Therefore, the probability that v is able to cover z is given by C_v^z as defined in (22).

$$C_v^z = Pr(d_v >= 2.\overline{d}) \tag{22}$$

In addition to the geographical coverage that can be provided by a volunteer, it is necessary to estimate the likelihood of a volunteer being present in a coverage zone, z. This is because if a volunteer is not present in a coverage zone over an epoch, then they cannot geographically cover it to monitor the spectrum. The volunteer mobility model predicts that a set of volunteers, $V_z = \{v_1, v_2, ..., v_k\}$, lands in zone z (as discussed later in Chapter 6). For every volunteer $v \in V_z$, the probability, ρ_v^z, of v to arrive in zone z is equal to the accuracy of the location prediction of v by v's mobility model. To ensure high geographical coverage of zone z, the expected coverage that can be provided by volunteers in V_z should be greater than or equal to a threshold Θ and is given by (23).

$$\sum_{v \in V_z} \rho_v^z \cdot C_v^z \geq \Theta \tag{23}$$

To this end, selecting all the volunteers in set V_z will ensure "optimal" coverage of the area of spectrum enforcement. However, there is a tradeoff primarily because selecting a large number of volunteers is expensive and has extra overheads. This is because maintaining the quality of the selected volunteers requires consistent analysis of their performance which can be both resource and time-intensive. Moreover, managing a large pool of volunteers can incur excess administrative costs in tasks such as communication and coordination with the volunteers.

Therefore, the objective is to minimize the number of volunteers selected from V_z to cover zone z while ensuring that the expected coverage provided by the volunteers is greater than or equal to the threshold, Θ. To this end, it is assumed that $\hat{V_z}$ is the set of minimum number of volunteers selected out of V_z to ensure that the coverage provided by the volunteers in zone z is at least Θ and that the value provided by each volunteer, v, to zone z is $\mu_v^z = \rho_v^z \cdot C_v^z$. Therefore, the minimum number of volunteers, $|\hat{V_z}|$, that ensure net value, $\sum_{v \in \hat{V_z}} \mu_v^z \geq \Theta$, needs to be determined. In the optimal solution to selecting volunteers for ensuring a net value of at least Θ, there must be a first volunteer, v_i, such that its value, $\mu_{v_i}^z \leq \Theta$. Let the minimum number of volunteers selected, $|\hat{V_z}|$ be represented by $T_z[\Theta]$, which represents the minimum number of volunteers whose total value is Θ. Hence, if $\mu_{v_i}^z$ is the value provided by the first volunteer, v_i, to zone z, in the optimal solution to getting a total value of Θ, then $T_z[\Theta] = 1 + T_z[\Theta - \mu_{v_i}^z]$ i.e., value $\mu_{v_i}^z$ added to $T_z[\Theta - \mu_{v_i}^z]$ to optimally get value of Θ. It is not known who is the first volunteer, v_i, that gives the optimal solution but it can be determined by checking the value provided by all the volunteers in $\hat{V_z}$, and the value of the optimal solution must correspond to the minimum value of $1 + T_z[\Theta - \mu_{v_i}^z]$, by definition. Moreover, if $\Theta = 0$, then the number of volunteers required is 0, as shown in (24).

$$T^z[\Theta] = \begin{cases} 0 & \text{if } \Theta = 0 \\ \min_{i:\mu_{v_i} \leq \Theta}\{1 + T[\Theta - \mu_{v_i}]\} & \text{if } \Theta > 0 \end{cases} \tag{24}$$

Algorithm 1 gives a bottom-up solution to determine the minimum number of volunteers in set V_z that can provide expected coverage of value, Θ, to zone z.

Algorithm 1 Minimum Number of Volunteers in Zone $z(V_z, k, \Theta)$

```
Data: Array $V_z$ of volunteers in zone $z$, integer $k$ for number of volunteers in $z$, threshold $\Theta$
Result: Minimum number of volunteers that can provide value of $\Theta$, $T_z[\Theta]$
Initialize array $T_z$ of size $\Theta + 1$ with all elements set to $\infty$
$T_z[0] \leftarrow 0$
for $p \leftarrow 0$ to $\Theta$ do
    $\min \leftarrow \infty$
    for $i \leftarrow 1$ to $|V_z|$ do
        if $\mu_{v_i}^z \leq p$ then
            if $1 + T_z[p - \mu_{v_i}^z] < \min$ then
                $\min \leftarrow 1 + T_z[p - \mu_{v_i}^z]$
            end if
        end if
    end for
    $T_z[p] \leftarrow \min$
end for
return $T_z[\Theta]$
```

The above algorithm recursively solves the equation in (24) and returns $T_z[\Theta]$ which is the minimum number of volunteers that can achieve the expected geographical coverage of Θ. Therefore, the target number of volunteers to be selected, V_T, in a geographical area of spectrum enforcement is the sum of the minimum number of volunteers in zone z to achieve expected coverage over Θ across all zones in every region, r, of the enforcement area, R, as shown in (25).

$$V_T = \sum_{r \in R} \sum_{z} T_z^r[\Theta] \tag{25}$$

Volunteers who are selected to monitor spectrum over a geographical area may not always be honest in reporting the spectrum access events. Corrupt volunteers present the risk of disturbing the spectrum monitoring process by reporting falsely on the spectrum access. To overcome the expulsion of corrupt volunteers upon detection, it is necessary to select the number of volunteers such that the overall corruption is mitigated.

In addition to satisfying the geographical constraints, the target number of volunteers to be selected, $V_T(z)$, must take into consideration the likelihood of the presence of corrupt volunteers in the pool of selected volunteers, V_z. Assuming that in a given zone, z, the probability of a volunteer being corrupted, P_C, is uniform, the probability that the number of corrupt volunteers, $V_T^C(z)$, among a total number of volunteers, $V_T(z)$, is equal to k, can be expressed in (26).

$$\Pr(V_T^C(z) = k) = \binom{V_T(z)}{k} P_C^k (1 - P_C)^{V_T(z)-k} \tag{26}$$

The expected number of corrupt volunteers, $\overline{V_T^C}$, of the binomial distribution is given by (27).

$$\overline{V_T^C}(z) = \sum_{k=1}^{V_T(z)} \binom{V_T(z)}{k} P_C^k (1 - P_C)^{V_T(z)-k} = P_C \cdot V_T(z) \tag{27}$$

To mitigate the impact of volunteer corruption, it is necessary to increase the size of selected volunteers, $V_T(z)$, by $\overline{V_T^C}(z)$ standby volunteers from the set of volunteers, V_z. The standby volunteers are engaged in the monitoring process any time an active volunteer who has been identified as corrupt is evicted. It is to be noted, however, that the size of the remaining volunteers in V_z may be less than $\overline{V_T^C}(z)$. Taking this into consideration, the number of standby volunteers, S_z, is the minimum of the expected number of corrupt volunteers among $V_T(z)$, and the remaining number of volunteers in V_z, as shown in (28).

$$S_z = minimum(||V_z|| - V_T, \overline{V_T^C}) \tag{28}$$

4.6 Volunteer Selection Algorithms

In this section, the algorithms for effective volunteer selection are discussed. The overall selection algorithm is based on a combination of a threshold-based Multiple-Choice Secretary algorithm and variations of the stable matching algorithms.

4.6.1 Multiple-Choice Secretary Algorithm

The Multiple-Choice Secretary algorithm (*MC-Secretary*) used here is a threshold-based volunteer selection algorithm that attempts to optimize the probability of selecting the most qualified volunteers [47, 48, 51, 174, 175]. This algorithm (executed by function *Multiple_Choice_Secretary()*), takes as input the set of all volunteers V and region r for which volunteers are to be selected to monitor spectrum. Using this algorithm, we select up to k_r volunteers such that $k_r = k/||R||$, where k is the maximum number of volunteers to be selected for the entire area of enforcement R. A random sample m_r is drawn from a binomial distribution $Binomial(||Q_r||, 1/2)$ (where Q_r is the number of volunteers available for selection in r), from which we select up to $k_r/2$ volunteers recursively [47, 48, 51, 174]. The initial m_r observed qualification values of volunteers are used to determine a threshold so that among the remaining volunteers, only those whose qualification values exceed this threshold are selected [47, 48, 51, 174]. This algorithm returns the set of selected volunteers $V_{S,r}$ in r. However, *MC-Secretary* does not assign channels to volunteers for monitoring. As discussed in [47, 51], we devise a modified Round Robin channel assignment scheme for assigning channels to volunteers based on their qualification (13) to monitor a channel. This scheme is executed by the function *Assign_Channels()*, where a hash table $H_{c,V_{S,r}}(MI)$ is maintained in which a channel c is mapped to the list $\Lambda_{c,V_{S,r}}$ of all volunteers in $V_{S,r}$ ordered in descending order by their qualification values to monitor c in r over a MI. For every region $r \in R$, a channel c is assigned in a round robin manner to the topmost v in $\Lambda_{c,V_{S,r}}$. After this assignment of c to v in r, v is deleted from the list $\Lambda_{c,V_{S,r}}$ of every channel in $H_{c,V_{S,r}}(MI)$, which ensures that a volunteer monitors only one channel in r over a given MI [47,51]. Algorithms *VM*, *RVM* and *MC-Secretary* (combined with *Assign_Channels()*) have polynomial time complexity in terms of the total number of volunteers, channels and regions.

Algorithm 2 Multi_Choice_Secretary(V, r)

```
Maintain matrix $\Psi_{V,R}$ that stores qualification values $\forall v \in V, \forall r \in R$, list of selected volunteers $V_{S,r}, \forall r \in R$
for all $r \in R$ do
    if $t = MI$ then
        $Q_r \leftarrow \Psi_{V,R}[r]$
        if $k_r = 1$ then
            Run Classic Secretary Algorithm
        else
            $m_r \leftarrow Binom(||Q_r||, 1/2)$
            if $m_r > k_r/2$ then
                $l_r \leftarrow k_r/2$
            else
                $l_r \leftarrow m_r$
            end if
            Recursively select upto $l_r$ volunteers
            $B_r \leftarrow descending\_sort(Q_r[1], ..., Q_r[m_r])$
            $threshold \leftarrow B_r[l_r]$
            for $i \leftarrow m_r + 1, ..., ||Q_r||$ do
                if $Q_r[i] > threshold$ and $||V_{S,r}|| < k_r$ then
                    $V_{S,r} \leftarrow V_{S,r} \cup v$
                else
                    Reject $v$
                end if
            end for
        end if
    end if
end for
return $V_{S,r}$
```

4.6.2 Stable Matching Algorithms

Stable Matching was first proposed by Gale and Shapley as the Deferred Acceptance algorithm in their seminal work [176] primarily for matching a set of men to a set of women (such that the size of both sets is equal) for marriage based on the preferences of both men and women. The authors further extended the algorithm to solve the college admission problem wherein college applicants are matched to colleges based on the selection criteria of the colleges and the preferences of the applicants. "An assignment of applicants to colleges is called unstable if there are two applicants α and β who are assigned to colleges A and B, respectively, although β prefers B to A and A prefers β to α" [176]. Several online auction mechanisms [113, 116, 177] have been proposed for crowdsourced spectrum sensing. While online auction mechanisms are able to match volunteers to channels over certain utility functions, they do not ensure stable matching [178, 179]. Stable matching is desirable because it considers both the preferences of volunteers and channel attributes for assignment of channels to volunteers.

In Computer Science, stable matching has been used primarily for resource allocation. Xu and Li [180] use stable matching algorithms to efficiently match virtual machines in the cloud (with heterogeneous resource needs) to servers. Saad et al. [181] use the concept of stable matching for college admissions to match users (students) to small and macro cell stations (colleges) to satisfy the utility of users and the coverage of cell stations. Gu et al. [182] match device-to-device users to cellular users for the purpose of efficient sharing of resources. Bayat et al. [183, 184] use a distributed algorithm for stable matching friendly of jammers to transmission pairs to protect from eavesdropping during transmission. The works [178, 179] focus on stable matching of spectrum buyers (Secondary Users) and sellers(Incumbents) to achieve Dynamic Spectrum Access in a distributed manner. However, to the best of our knowledge, there has been no work that applies stable matching for efficient spectrum enforcement and misuse detection.

Stable matching is desirable because it considers both the preferences of volunteers and channel attributes for assignment of channels to volunteers. Since factors like monitoring device characteristics can cause a volunteer to prefer one channel over the other, having a stable match ensures lesser incentives for a volunteer to deviate from the current match and thereby lesser overheads and better *volunteer happiness* [178,179]. The proposed approach is similar to the methodology for stable matching of college applicants (volunteers) to colleges (channels).

4.6.2.1 Volunteer Matching

The Volunteer Matching algorithm (VM) selects volunteers by using a variation of the Gale Shapley algorithm that is proposed for college admissions [176]. To this end, every volunteer maintains a Priority list P_v^r which contains the list of channels that a volunteer v can monitor in region r. The channels in P_v^r are ordered by v's preferences. We assume that every volunteer v prefers to monitor channels in the region r that it is currently residing in. Furthermore, in P_v^r, the channels are ordered by the potential quality of spectrum misuse detection $\psi_v^{c,r}$ (s.t. $\psi_1 \leq \psi_v^{c,r} \leq \psi_2$) by volunteer v in channel c of region r. It is assumed that $\psi_v^{c,r}$ depends on the characteristics of the spectrum sensing device used by v. Similarly, the Volunteer Service unit Ω_r of every region $r \in R$ maintains a Candidate list χ_c^r for every $c \in C$ such that it contains a list of volunteers (sorted in descending order by their qualification) who apply to monitor c in r.

The algorithm VM (shown by Algorithm 3) takes the set of all volunteers, V, set of all channels, C, and set of regions, R as input and returns the set of selected volunteers V_S as output. Every $v \in V$ is added to the Candidate list χ_c^r associated with c in the region r such that c is the first choice of v in its Priority list P_v^r (lines 6-7). From the candidate list, χ_c^r, of channel c in r, only the top q_c^r (s.t. $q_c^r = k/(||R||.||C_r||)$), where k is the maximum number of volunteers to be selected in R, $||R||$ is the number of regions and $||C_r||$ is the number of channels in region r) candidates ranked by their qualification are stored in the waiting list W_c^r and the remaining candidates are rejected and stored in a reject list Θ_c^r. This is repeated for all the channels in C (lines 11-19). Volunteers rejected by a channel c will apply to their next choice in P_v^r. This process continues until every $v \in V$ is rejected by all channels or placed on the waiting list, W_c^r, of a channel c in r. Finally, volunteers in W_c^r for each channel c in r are selected to monitor channel c in r (for every $c \in C$ and $r \in R$) (lines 21-24).

Algorithm 3 Volunteer_Match(V, R, C)

```
Maintain Channel Priority list $P_v^r$, Candidate list $\chi_c^r$, Waiting List $W_c^r$, Reject list $\Theta_c^r$, list of selected volunteers $V_{S,r}$, $\forall v \in V, \forall r \in R, \forall c \in C$.
$W \leftarrow \cup_{c,r} W_c^r$
while $\exists v, r, c : v \notin W$ and $v \notin \Theta_c^r$ do
    $W \leftarrow \cup_{c,r} W_c^r$
    for all $v \in V$ do
        if $\exists r, c : v \notin W$ and $v \notin \Theta_c^r$ then
            $top\_channel \leftarrow P_v^r[1]$
            $\chi_{top\_channel}^r \leftarrow \chi_{top\_channel}^r \cup v$
            $Delete\ the\ entry\ P_v^r[1] from P_v^r$
        end if
    end for
    for all $c \in C$ do
        for all volunteers $v \in \chi_c^r$ do
            if $v \in top\_q_r(\chi_c^r)$ then
                $W_c^r \leftarrow W_c^r \cup v$
            else
                $\Theta_c^r \leftarrow \Theta_c^r \cup v$
            end if
        end for
    end for
end while
for all $c \in C$ in all $r \in R$ do
    $Assign\ channel\ c\ to\ all\ v \in W_c^r$
    $V_{S,r} \leftarrow V_{S,r} \cup W_c^r$
end for
$V_S \leftarrow \cup_r V_{S,r}$
return $V_S$
```

4.6.2.2 Reverse Volunteer Matching

Similar to Algorithm 3, we assume that for the Reverse Volunteer Matching algorithm (RVM), every volunteer $v \in V$ maintains a Priority list P_v^r of channels ordered by v's preferences to monitor channels in R. In contrast to Algorithm 3, where volunteers propose to the Volunteer Selection Unit to be matched to a channel of their choice and get matched accordingly, in RVM (shown by Algorithm 4), the DSA infrastructure initially accumulates all the proposals from volunteers in V and then proposes back to volunteers (based on their qualification) with offers to match them to *suitable* channels [176].

In Algorithm 4, the Volunteer Service unit Ω_r associated with r constructs a Candidate list χ_c^r for every channel c in r such that it contains all the volunteers who applied to monitor c in r (lines 1-3). This candidate list χ_c^r is sorted in descending order by the qualification of volunteers (line 7) and handed to the Volunteer Selection Unit. For every $c \in C$, the Volunteer Selection Unit proposes an offer to monitor c in r to the first volunteer v_{top} in χ_c^r and if volunteer v_{top} is not assigned to monitor any other channel (or is free), then v_{top} and c are matched and the matched pair is stored in the match list M_c^r maintained by Ω_r for every c in r (lines 8-11). On the contrary, if volunteer v has already been assigned a channel c' to monitor in region r' (r' may or may not be the same as r), then the rank of preference of v to monitor c is compared to that of monitoring c' (lines 12-13). If v prefers to monitor c in r to its current match c' in region r', then v is matched to c and removed from the match list $M_{c'}^{r'}$ (lines 13-15). This continues till every channel $c \in C$ in every $r \in R$ has at most q_c^r matched volunteers or there are no more volunteer left to propose to in χ_c^r. Finally, all the volunteers in M_c^r of c in r are selected and assigned to monitor c in r (for every $c \in C$ in $r \in R$) (lines 20-23).

Algorithm 4 Reverse_Volunteer_Match(V, R, C)

```
Maintain Channel Priority list $P_v^r$, Candidate list $\chi_c^r$, Channel Match list $M_c^r$, list of selected volunteers $V_{S,r}$, $\forall r \in R, \forall c \in C$.
for all $c \in C$ do
    $\chi_c^r \leftarrow \chi_c^r \cup v, \forall v$ applying to monitor $c \in C$
end for
while $\exists r, c : ||\chi_c^r|| > 0$ and $||M_c^r|| < q$ do
    for all $c \in C$ do
        $n \leftarrow ||\chi_c^r||$
        $\chi_c^r \leftarrow descending\_sort(\chi_c^r[1], ..., \chi_c^r[n])$
        $v_{top} \leftarrow \chi_c^r[1]$
        Delete the entry $\chi_c^r[1]$ from $\chi_c^r$
        if $v_{top}$ is free then
            $M_c^r \leftarrow M_c^r \cup v_{top}$
        else if $v_{top} \in M_{c'}^r$ then
            if $rank(v_{top}, c) > rank(v_{top}, c')$ then
                $M_c^r \leftarrow M_c^r \cup v_{top}$
                $M_{c'}^{r'} \leftarrow M_{c'}^{r'} - v_{top}$
            end if
        end if
    end for
end while
for all $c \in C$ in all $r \in R$ do
    Assign channel c to all $v \in M_c^r$
    $V_{S,r} \leftarrow V_{S,r} \cup M_c^r$
end for
$V_S \leftarrow \cup_r V_{S,r}$
return $V_S$
```

4.6.3 A Composite approach for volunteer selection

The Multiple - Choice Secretary algorithm succeeds in selecting volunteers with high qualification values and when combined with the $Assign\_Channels()$ function, succeeds in giving high performance in terms of accuracy of misuse detection and coverage when compared to an algorithm that selects volunteers randomly. However, this algorithm does not guarantee a stable matching of volunteers and channels and hence does not ensure mutual satisfaction of volunteers and spectrum monitoring platform. In addition, the Secretary-based algorithm is not adaptable to the changing behavior of volunteers as it may result in starvation of volunteers who behave/perform poorly in the initial rounds of selection. A stable matching algorithm for volunteer selection ensures that there is stable matching of volunteers to channels and hence high mutual satisfaction of volunteers and the spectrum monitoring platform. However, this methodology does not ensure high performance (in terms of coverage and accuracy). Therefore, we propose to combine the benefits of the two algorithms to achieve the requirements for an efficient selection of volunteers. Figure 12 demonstrates the usage of a composite algorithm that combines the benefits of the two algorithms (MC Secretary and Stable Matching) to select volunteers from V. The selected volunteers are represented by the set V_S which are then assigned for monitoring channels in regions $r_1, r_2, ..., r_R$.

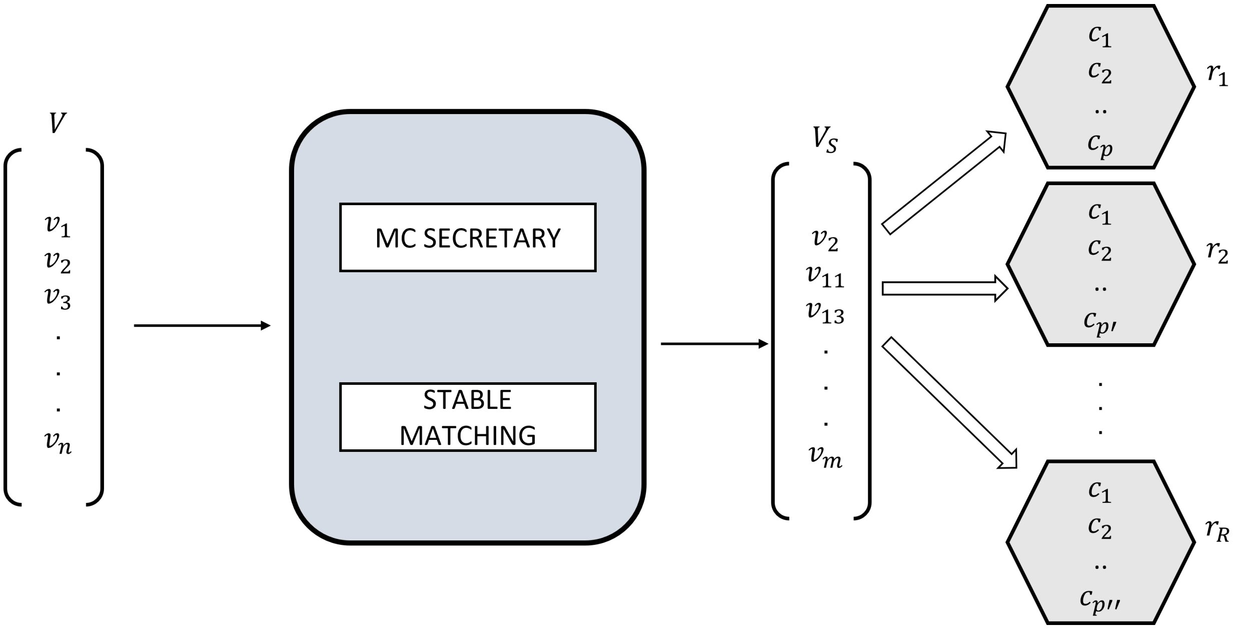


Figure 12: A composite approach for volunteer selection

4.6.4 Hybrid Algorithms

We develop two hybrid algorithms, *HYBRID-VM* and *HYBRID-RVM* as shown by Algorithms 5 and 6 by blending *MC-Secretary* with *VM* and *RVM* respectively.

Algorithm 5 Hybrid_Volunteer_Match(V, R, C)

```
Maintain list of selected volunteers $V_{S,r}$, $\forall v \in V, \forall r \in R$.
for all $r \in R$ do
    $V'_{S,r} \leftarrow$ Multiple_Choice_Secretary($V$, $r$)
    $V_{S,r} \leftarrow$ Volunteer_Match($V'_{S,r}$, $R$, $C$)
end for
$V_S \leftarrow \cup_r V_{S,r}$
return $V_S$
```

Algorithm 6 Hybrid_Reverse_Volunteer_Match(V, R, C)

```
Maintain list of selected volunteers $V_{S,r}$, $\forall v \in V, \forall r \in R$.
for all $r \in R$ do
    $V'_{S,r} \leftarrow$ Multiple_Choice_Secretary($V$, $r$)
    $V_{S,r} \leftarrow$ Reverse_Volunteer_Match($V'_{S,r}$, $R$, $C$)
end for
$V_S \leftarrow \cup_r V_{S,r}$
return $V_S$
```

In both *HYBRID-VM* and *HYBRID-RVM*, we feed the volunteers who are selected by *Multiple_Choice_Secretary()* to functions *Volunteer_Match()* and *Reverse_Volunteer_Match()* respectively. This is done because it establishes a threshold above which volunteers are chosen for matching and thereby improves on the individual performance of the Secretary-based algorithm and the vanilla matching algorithms.

Table 1: Simulation Parameters

Parameter	**Value**
Area of Enforcement	$500m \times 1000m$
Population	1070
Number of Volunteers	183
Number of channels per region	5
Number of Regions	2
Mobility Model	Random Waypoint
Maximum Battery Capacity of Volunteer	$7Wh$
System Parameter h	0.03
System Parameter κ	1
Reputation Threshold ζ	-10
Number of WUIs	5580

4.6.5 Random Algorithm

This algorithm is the baseline for other algorithms to compare against. Here, volunteers are selected randomly irrespective of their qualification values. The selected volunteers are assigned channels to monitor, in a round-robin fashion, irrespective of their qualifications or preferences.

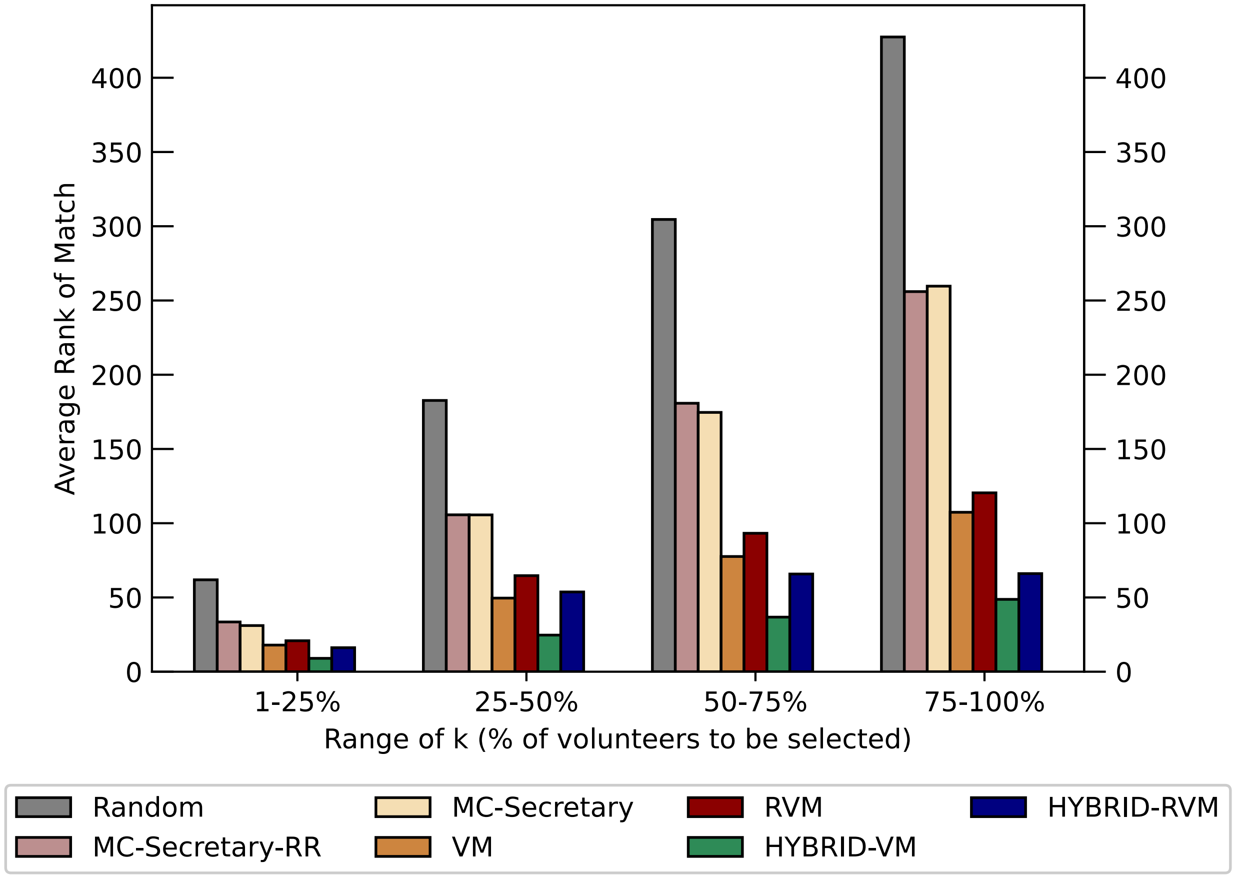


Figure 13: Comparison of Average Rank of Match with change in k.

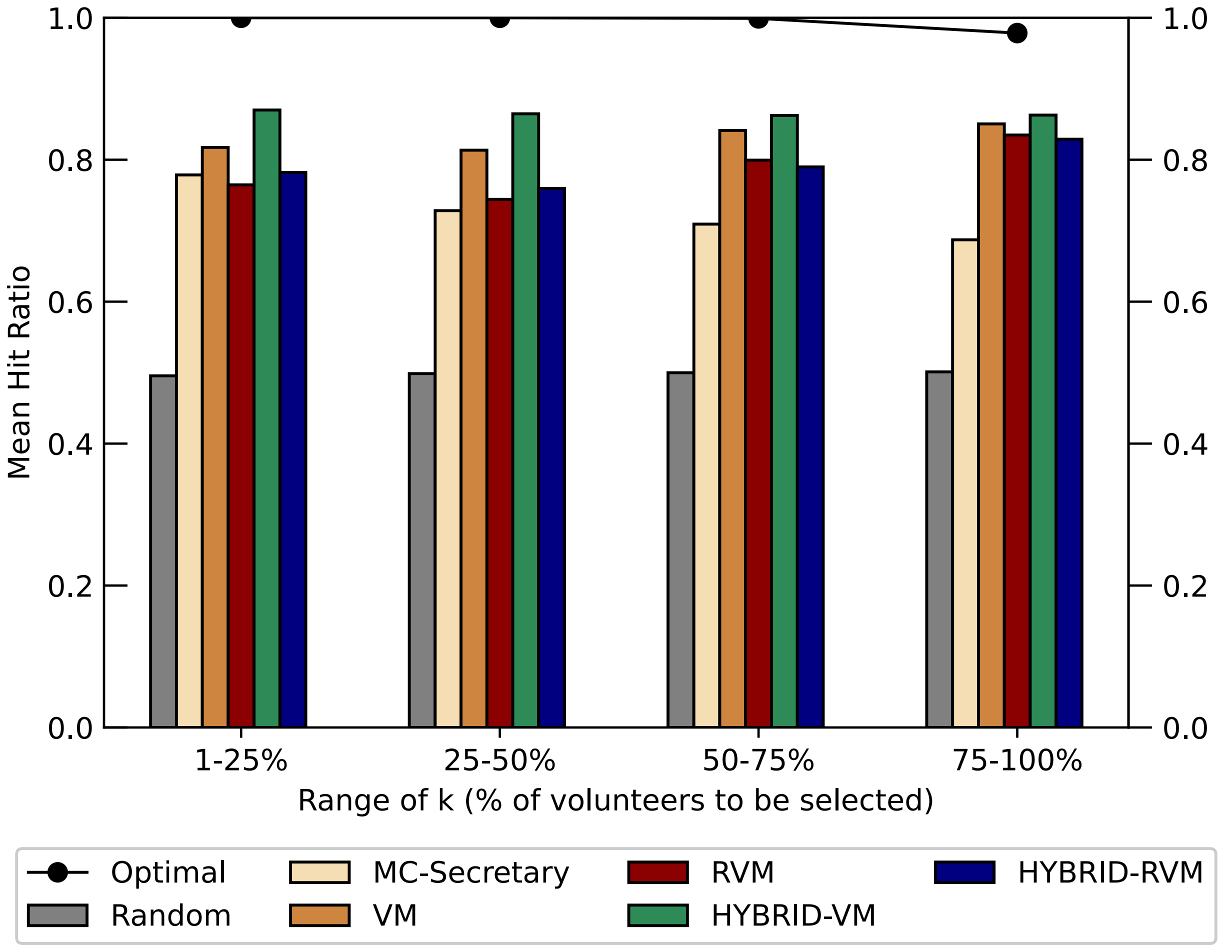


Figure 14: Comparison of Mean Hit Ratio with change in k.

4.7 Experiments and Results

We assume that one MI consists of five WUIs and that volunteers are selected at the beginning of every MI (starting from the second MI). Values chosen for the Simulation parameters are shown in Table 5.1. The performance of the volunteer selection algorithms is measured using the following metrics:

1. *Average Rank of Match*: Average rank of the channel (that a selected volunteer v is

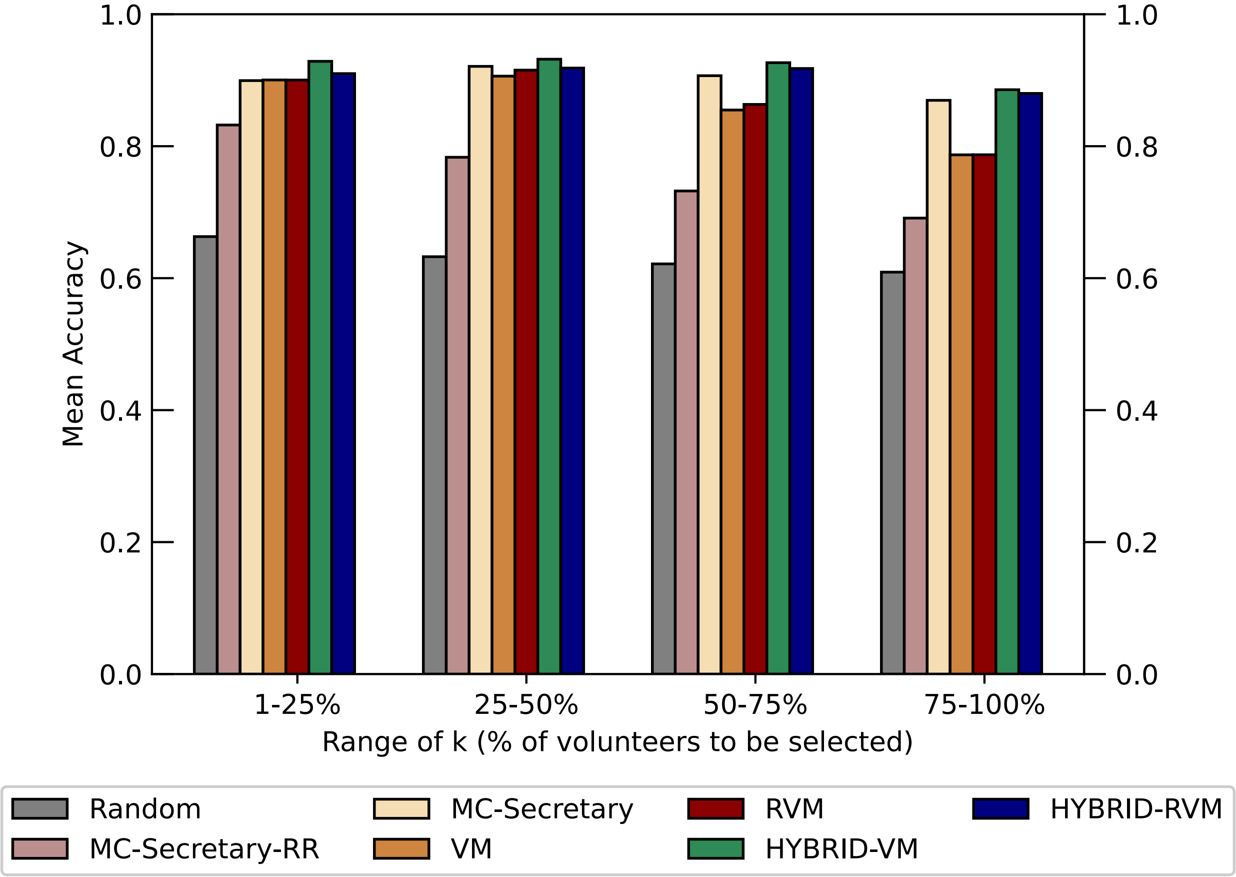


Figure 15: Comparison of Mean Accuracy of detection with change in k.

assigned to monitor) in v's priority list P_v^r, for all $v \in V$ who are selected for spectrum monitoring over the duration of the simulation. A lower value indicates higher *volunteer happiness*.

2. *Mean Hit Ratio*: Ratio of the number of hits to the total number of hits and misses. If a volunteer v selected for monitoring region r is in r at the beginning of a WUI of a MI that v is selected for, then it is a *hit*, otherwise, it is a *miss*. A higher hit ratio indicates higher coverage of R by volunteers over the period of enforcement [47, 48, 53].

3. *Mean Accuracy of Detection*: All volunteers detect misuse with probability δ (where $\delta = 0.5$ for corrupt volunteers and $\delta = 1$ for honest volunteers) times the potential

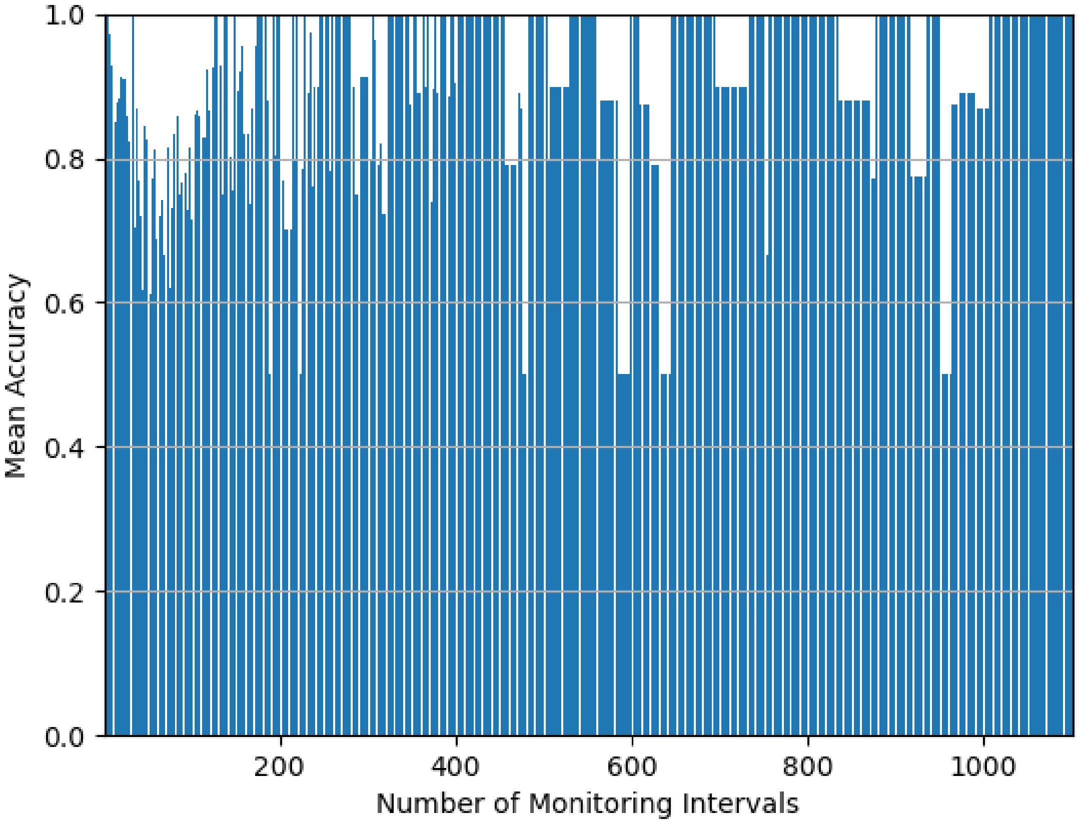


Figure 16: Change in mean accuracy with time for volunteer selection using HYBRID-VM with $k = 0 - 25\%$.

quality of spectrum misuse detection $\psi_v^{c,r}$ (Section 4.6.2.1). A misuse detection result by a volunteer is accurate if it is above a threshold (i.e., matches that of a sentinel s in r at a WUI in which s monitors).

In Fig. 13 and 15, *MC-Secretary-RR* refers to *MC-Secretary* with simple Round Robin channel assignment (irrespective of volunteer qualification) unlike *MC-Secretary* which uses *Assign_Channels()*. As expected, the baseline Random algorithm performs the worst in all experiments. In Fig. 13, we observe that all the matching algorithms (*VM*, *RVM*, *HYBRID-VM* and *HYBRID-RVM*) have lower Average Rank of Match (i.e., higher *Volun-*

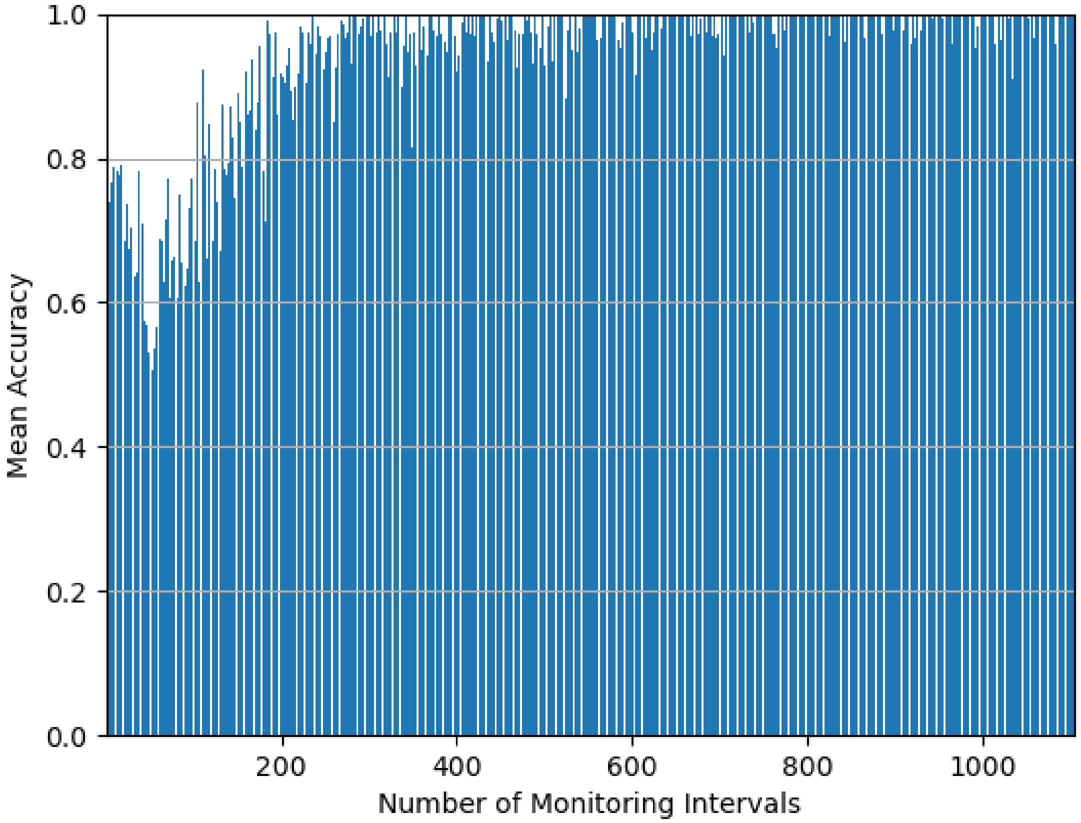


Figure 17: Change in mean accuracy with time for volunteer selection using HYBRID-VM with $k = 25 - 50\%$.

teer Happiness) than the remaining algorithms because they utilize volunteer preferences for selection and channel assignment. Also, *VM* (or *HYBRID-VM*) performs better than *RVM*(or *HYBRID-RVM*) because stable matching algorithms are biased towards the ones who propose [176]. In *VM* and *HYBRID-VM*, volunteers propose to the DSA infrastructure to get matched, unlike in *RVM* and *HYBRID-RVM* where the DSA infrastructure makes the final proposal. In Fig. 14, we include a clairvoyant *Optimal* algorithm which calculates in hindsight the optimal mean hit ratio after selecting k_r volunteers for every $r \in R$. Its mean hit ratio goes below 1 when k increases because the proportion of k_r volunteers staying in r

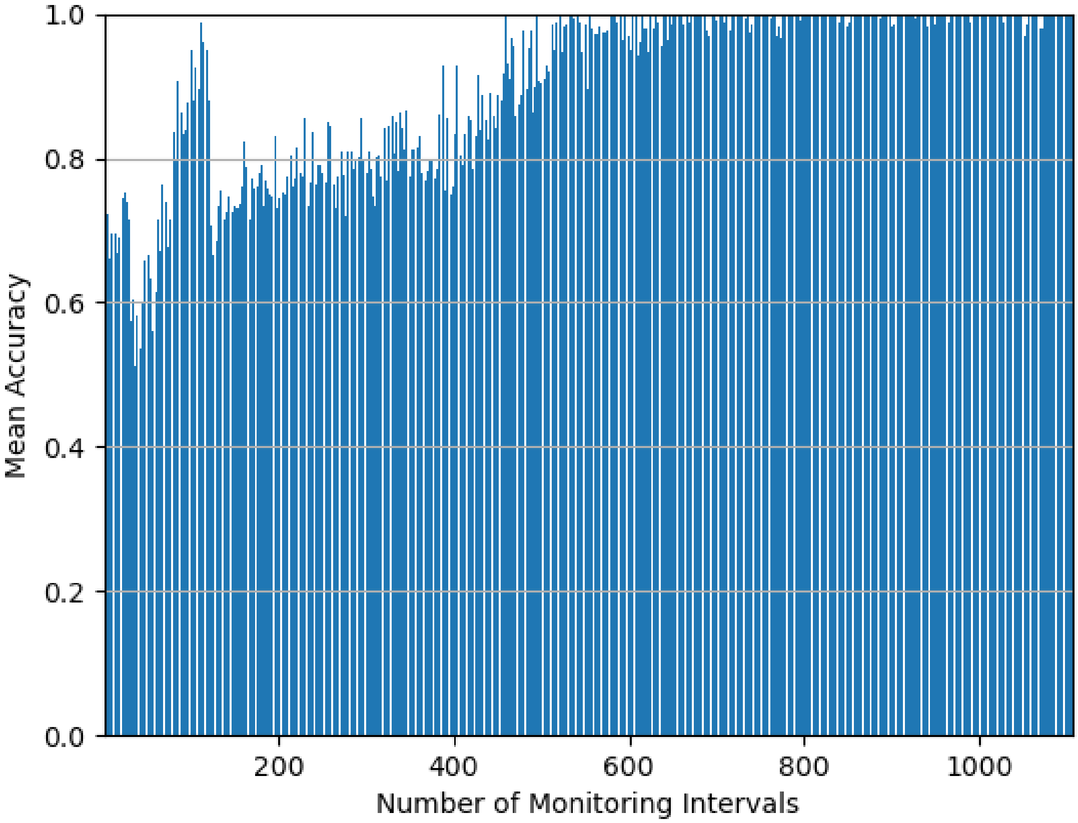


Figure 18: Change in mean accuracy with time for volunteer selection using HYBRID-VM with $k = 50 - 75\%$.

decreases. Interestingly, *VM* and *RVM* give better hit ratio than *MC-Secretary* as k increases because unlike *MC-Secretary* they consider volunteer preference to monitor channels in their region of residence. The hybrid algorithms utilize this advantage of the vanilla matching algorithms to give higher hit ratio than MC-Secretary as k increases. *HYBRID-VM* gives the best mean hit ratio over all ranges of k and performs better on average than *VM* (by 4.1%), *MC-Secretary*(by 19.2%), *RVM* (by 10.1%) and *HYBRID-RVM* (by 9.5%). In Fig. 15, we observe that *MC-Secretary* gives higher detection accuracy than *MC-Secretary-RR* because it uses *Assign_Channels()*. Interestingly, *MC-Secretary* performs better than *VM*

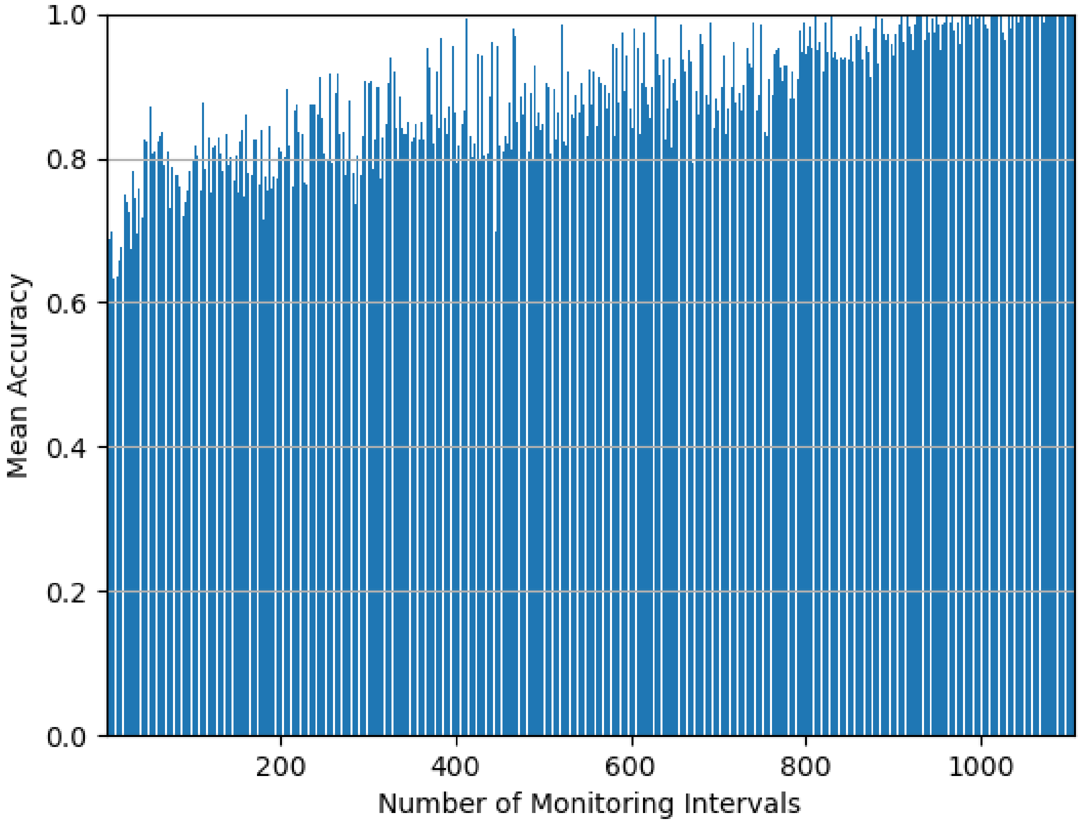


Figure 19: Change in mean accuracy with time for volunteer selection using HYBRID-VM with $k = 75 - 100\%$.

and *RVM* in terms of detection accuracy than in terms of mean hit ratio because it selects volunteers based only on their qualification (13), in which volunteer reputation (being combined exponentially) dominates. Also, *VM* and *RVM* perform poorly as range of k increases because volunteer preferences of channels do not necessarily align with the best interests of the DSA infrastructure. The hybrid algorithms utilize this advantage of MC-Secretary and give better detection accuracy than *VM* and *RVM* (with *HYBRID-VM* giving the best mean accuracy for all ranges of k).

In the final set of experiments, the spectrum access violation detection accuracy is compared for different values of k using the HYBRID-VM algorithm for volunteer selection, as shown in Figs. 16, 17, 18 and 19. In these experiments, corrupt volunteers, whose normalized reputations fall below a threshold ($\nu = 0.01$), are expelled from the set of active volunteers and are never selected again for spectrum monitoring. In Figs. 17, 18 and 19, it is observed that the detection accuracy increases and converges to 1 as the number of monitoring intervals increases. This is because the number of corrupt volunteers who get selected keeps decreasing with the increase in monitoring intervals and hence the detection accuracy achieved by the remaining set of honest volunteers reaches 1. It is also observed that as the value of k increases, the number of Monitoring Intervals required for the detection accuracy to converge to 1 increases. This is because, with the increase in the number of selected volunteers (i.e., with the increase in the value of k), the proportion of corrupt volunteers selected also increases. Therefore, as k increases, it requires more Monitoring Intervals before the pool of selected volunteers is free from corrupt volunteers, and hence requires more Monitoring Intervals for the detection accuracy to converge to 1. Finally, as observed in Fig. 16, for $k = 0 - 25\%$, the detection accuracy does not increase gradually to converge to 1 and instead shows a variance in detection accuracy values throughout the plot. This is because when fewer volunteers are selected, it takes longer time to build the reputation profile of corrupt volunteers as the average proportion of corrupt volunteers selected in a Monitoring Interval is lesser in this case. Hence, in this scenario, when there are corrupt volunteers in the pool of selected volunteers, the detection accuracy falls below 1, and otherwise the detection accuracy remains 1 when honest volunteers are selected.

4.8 Summary

In this chapter, we formulated the volunteer selection algorithm for spectrum enforcement as two variants of the stable matching algorithm [176], *VM*, and *RVM*. While the matching algorithms take volunteer preferences into consideration, they do not always produce the best performance for the DSA infrastructure (in terms of coverage and accuracy). So, we combined a variant of the Multiple-choice Secretary algorithm [174] with the matching algorithms to develop two hybrid algorithms, *HYBRID-VM* and *HYBRID-RVM*. We observed that *HYBRID-VM* gave the best performance in terms of accuracy of detection, region coverage and *volunteer happiness* when compared to the other algorithms.

5.0 Spectrum Monitoring Approaches For Robust Enforcement of Access Rights

An effective spectrum monitoring scheme should be robust [54]. This will ensure consistency in detecting spectrum access violations and catching intruders over prolonged intervals. An effective spectrum monitoring scheme must be able to detect spectrum access violations successfully against

1. Errors caused due to hardware failure or environmental factors like poor received signal strength due to location constraints, etc.
2. Intruders, who exhibit different behavior while accessing spectrum illegally.
3. Volunteers, who exhibit different behavior while reporting about spectrum misuse.

In this chapter, I address the challenges of robust and reliable detection of unauthorized spectrum access over prolonged time intervals. For this purpose, a framework is proposed that is driven by efficient spectrum sampling and reputation management of crowdsourced spectrum monitoring agents. This ensures robustness and success in detecting spectrum access violations and catching intruders consistently. Simulation results indicate that the proposed scheme successfully catches intruders across all spectrum monitoring intervals, despite varying intruder profiles, sensing failure rates, and the trustworthiness of volunteers.

This chapter is organized as follows. The motivation for this work is first discussed. Then, the spectrum monitoring scheme and the key players in our framework—volunteers, sentinels, and intruders—are examined. The spectrum sampling strategy of sentinels and volunteers is then addressed. Next, the reputation management algorithm is explained. Finally, the experimental setup is described, and the obtained results are presented.

5.1 Motivation

In this chapter, I address the challenge of maintaining the robustness of the spectrum enforcement scheme, to detect unauthorized access to spectrum consistently over prolonged intervals of time. This is desirable to ensure consistency in performance and persistence of the system against varying physical and environmental adversities. I propose a framework, driven by a spectrum sampling scheme and reputation management of recruited crowdsourced volunteers. This facilitates the detection and identification of spectrum intruders against different intruder profiles, types of volunteers, and sensing failure rates, with minimal harm and consistent success. In other words, the main purpose of this chapter is to ensure an effective Quality of Sensing, which is persistently robust and successful against unauthorized access in a shared spectrum network. This work builds on [47,48,51,53], with the following key contributions [54]:

1. An efficient sampling strategy for spectrum monitoring by crowdsourced spectrum monitoring agents, is proposed.
2. An algorithm to manage the reputation of crowdsourced monitoring agents, is proposed, which in turn ensures a reliable and robust detection of spectrum misuse consistently.

While section 5.2 discusses the spectrum sampling scheme, section 5.3 discusses the reputation management algorithm and section 5.4 discusses the simulation settings and experiment results.

5.2 Spectrum Sampling Scheme

As discussed in Chapter 4, the total spectrum monitoring period is divided into Monitoring Intervals, *MIs*, at the beginning of which a new set of volunteers is selected *MI* by the *Volunteer Selection* unit of the Access Enforcement Computational Infrastructure, based on their *qualification*. Each *MI* is divided into sub-intervals called Working Unit Intervals (*WUIs*), which is the smallest time interval over which a user (authorized or unauthorized) can accomplish useful work [47,48,51,53]. A *WUI* is further divided into access slots, over which a user (authorized or unauthorized) can access a channel. A sentinel and a volunteer can also sense a channel over an access slot to determine the access type of the channel. The access type of a channel is determined by whether it is empty or if it is accessed by an authorized user and/or an intruder. Both the volunteers and sentinels primarily attempt to detect the spectrum misuse by an intruder. The following sub-sections discuss spectrum access by three key players in the spectrum monitoring scheme - Volunteer, Sentinel, and Intruder.

5.2.1 Volunteer Spectrum Sampling Scheme

It is assumed that a volunteer is either corrupt or honest. A corrupt volunteer is assumed to collude with intruders to gain access to the spectrum illegally. To this end, a corrupt volunteer deterministically fails to report an intruder's illegal use of spectrum over an access slot in a *WUI* but reports the ground truth when the spectrum is not in use by an intruder. In other words, a corrupt volunteer chooses to always lie about the usage of spectrum by intruders. On the contrary, an honest volunteer always reports the ground truth. However, both the honest and corrupt volunteers report the access state of a channel with a probability of sensing failure, P_f^v. A sensing failure may occur due to physical and environmental constraints, like poor signal strength received by the sensing device due to the location of a volunteer or simply due to device hardware failure.

A trivial approach for sampling spectrum to detect and catch an intruder is to monitor spectrum in every access slot of a *MI.* However, considering the energy constraints of a volunteer, I propose a more efficient scheme for sampling the spectrum. A volunteer needs to monitor a channel at the same time that an intruder uses it, to be able to detect and report the intruder. However, the timing and pattern of spectrum usage by an intruder are unknown to the volunteer. If it is assumed that a *WUI* consists of τ access slots, then an intruder would need to use at least τ access slots consecutively, to do any useful work. In order to avoid unnecessary overhead, a volunteer would need to observe a user accessing the channel consecutively for more than one access slot before communicating to the *Spectrum*

Access Database to verify the identity of the user. To this end, a volunteer decides to verify the legitimacy of a user u only after it consecutively observes that the channel is being used by u for at least $\tau/2 + 1$ access slots. Assuming that it takes $\tau/2 - 1$ access slots to verify the legitimacy of the user u from the *Spectrum Access Database* and to subsequently jam the channel for preventing further use by the intruder, the Cost incurred for verifying and catching an intruder should be at most 1 *WUI*, in cases where the spectrum sensing is perfect and without any failure by the volunteer. Here, *Cost* incurred is the maximum number of *WUIs* for which a channel is used by an intruder before being caught. For this purpose, a volunteer needs to monitor only the first $\tau/2+1$ slots in every *WUI*. In fig. 20, the volunteer monitoring scheme is shown over two *WUIs*. It is observed that no matter which access slot the intruder starts intruding, a volunteer can consecutively catch the intruder stealing spectrum for at least $\tau/2+1$ slots, which is enough to initiate verification of intruder identity by using the Spectrum Access Database.

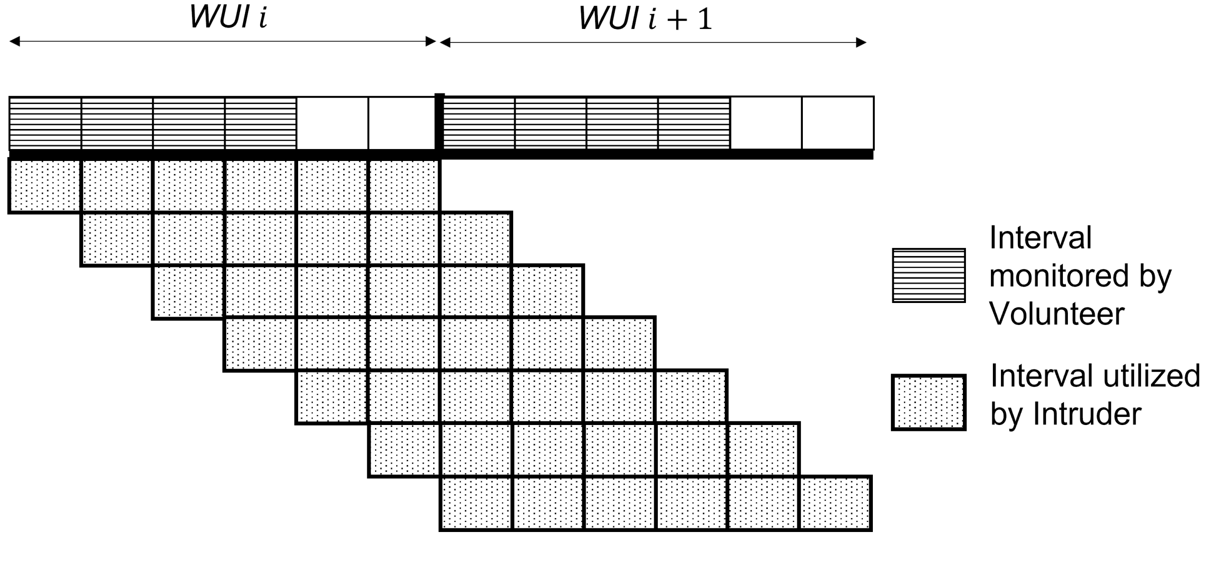

WUI i
WUI i + 1
Interval monitored by Volunteer
Interval utilized by Intruder

Figure 20: Volunteer Spectrum Monitoring Scheme Over Two *WUIs*.

5.2.2 Sentinel Spectrum Sampling Scheme

A sentinel uses a random process to monitor the spectrum. In a given *WUI*, k, a sentinel chooses to randomly the channel and use the outcome to assess the trustworthiness of a volunteer, v. The challenge is to compute the probability of monitoring a channel during a WUI, k, to ensure that a sufficient number of samples, n, is collected to accurately assess a volunteer v's trustworthiness and to minimize the energy required by the sentinel to monitor a channel. If I consider S to be the number of *WUIs* to be sampled, the goal is to determine the proportion, n, of *WUIs* to be sampled in order to accurately assess the trustworthiness of a volunteer. If it is assumed that the number of sampled *WUIs*, S, is much smaller than the total number of *WUIs*, N, in a Monitoring Interval, S can be statistically modeled as a normal distribution. Assuming a margin of error, δ, in estimating the proportion of sampled *WUIs* at $\beta\%$ confidence level, the following equation (29) holds.

$$\begin{aligned} Confidence\_Interval &= \hat{p} + z^{*}.\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ \Longrightarrow n &= \frac{z^{*(2)} \cdot p \cdot (1-p)}{\delta^{2}} \\ \Longrightarrow n &\geq 0.25.(\frac{z^{*}}{\delta})^{2} \end{aligned} \tag{29}$$

where $\hat{p}$ is the sample proportion, z^{*} is the critical value and n is the minimum sample size required. Since the total number of *WUIs* in a Monitoring Intervals, N, is finite, the minimum number of *WUIs* required to be sampled can be reduced further by using (30) [185].

$$n_{min} = \frac{n}{1 + \frac{n-1}{N}} \tag{30}$$

Therefore, it is needed to ensure that the expected number of samples, S, taken is at least n_{min}. If σ is the probability of sampling a *WUI*, the probability to sample n_{min} *WUIs* in a *MI* is given by (31).

$$\begin{aligned} \sum_{k=1}^{N} k.\binom{n_{min}}{k}\sigma^{k}.(1-\sigma)^{n-k} \geq n_{min} \\ \Longrightarrow N.\sigma \geq n_{min} \\ \Longrightarrow \sigma \geq \frac{n_{min}}{N} \end{aligned} \tag{31}$$

where N is the total number of *WUIs* in a *MI*.

5.2.3 Intruder Spectrum Access

An intruder attempts to utilize the spectrum illegitimately in an access slot. To define the behavior of an intruder, its profile is constructed by associating two main parameters with it. The first parameter is the probability of an intrusion attempt in an access slot, ρ. The value of ρ varies with the type of intruder. An intruder which attempts to intrude with higher frequency will have a higher value of ρ. The second parameter that is used to define the profile of an intruder is the number of intervals that are stolen by the intruder consecutively. It is known that an intruder will have to utilize at least τ access slots consecutively, to do some useful work (Section 5.2.1). Thus, the number of intervals that are to be stolen by an intruder is given by $\eta.\tau$ where η is varied based on the audacity of an intruder. The two parameters, ρ and τ, in combination, measure the propensity to risk of an intruder. The intruders' inclination to take risks can, thus, be measured by how persistent the intruder is to start using the channel and how long it will stay in the channel after acquiring access to it.

5.3 Volunteer Reputation Management

At the beginning of a *MI*, new volunteers are selected to monitor the channels in a geographical region. Additionally, sentinels monitor the spectrum to assess the trustworthiness of the selected volunteers, so that volunteers who serially exhibit untrustworthy behavior are relegated to the point of no selection for monitoring the spectrum. A volunteer is deemed untrustworthy when the spectrum access state reported by it does not match the spectrum access state reported by a sentinel. However, volunteer observations can also be inaccurate due to technical fallacy. To this end, the reputation of volunteers is built over a period of time, based on the foundation of their trustworthiness exhibited over a period of time.

An access slot can be used to verify the behavior of a volunteer, v, if and only if sentinel, s, senses without failure, and a channel, c, is sensed simultaneously by both v and s. As shown in algorithm 7, if both s and v monitor spectrum over an access slot, the number of verified intervals is incremented by 1 (lines 9-10). In a verified interval, if the observation (about the spectrum access state) made by the volunteer, $O^{i,k}_{v,c}$, matches with that of the sentinel, $O^{i,k}_{s,c}$, then the success score of v for WUI i is incremented by 1, otherwise the failure score of v for WUI i is incremented by 1 (lines 12-15). At the end of WUI, i, I verify if the failure score of v in WUI i is less than ψ (also referred to as the Discrepancy Tolerance) times the total number of verified intervals in the AUI (line 20). In such cases, I update the volunteer failure score, F^v_m, that is maintained for v in MI m. Furthermore, I penalize v by decreasing its reputation exponentially in terms of its F^v_m score(line 22). If the new reputation of v lies below the reputation threshold, θ, I decide to expel v, calculate the *Cost* incurred (discussed in more detail in Section 5.4) and select a new volunteer, if available (lines 23-26). If, however, the new reputation of v does not fall below θ, then the volunteer failure score, F^v_m, is saved so that it can be used for a more severe penalty, if the failure score is above ψ times the total number of verified intervals, next time. On the other hand, if the failure score of v is less than or equal to ψ times the total number of verified intervals in the WUI i, then I increase the volunteer reputation linearly and do not update the volunteer failure score, F^v_m, either (line 29). Hence, the basic principle of our approach for reputation management is to increase the reputation slowly after success and penalize the reputation progressively more rapidly with an increase in the number of failures.

Table 5.1: Simulation Parameters

Parameter	**Value**
Number of channels	1
Number of Access Slots in a *WUI*	20
Number of *WUIs*, N, in a *MI*	5000
Number of *MIs*	50
Confidence level, β	95%
Margin of Error, δ	2
Reputation Threshold, θ	0
System Parameter, α	1
System Parameter, χ	1
Probability of Sensing Failure of a Sentinel	0.01

5.4 Experiments and Results

Several experiments are conducted to assess the performance of the proposed scheme to detect intrusion in a channel. The quality of spectrum sensing by utilizing the proposed spectrum monitoring scheme is determined by its robustness and success against sensing failures, varying types of intruders, and volunteers. The values that are chosen for the simulation parameters are shown in Table 5.1. Since the channels operate independently, analyzing the performance of the scheme over a single channel, given the intruder profile and volunteer type, will be sufficient to determine the performance over multiple Monitoring Intervals. I assess the performance in terms of the average *Cost* incurred to assert *Success* or *Failure* of the proposed scheme. The *Cost* incurred measures the average number of useful *WUIs* used by an intruder successfully before it is caught in a Monitoring Interval. The

Algorithm 7 Volunteer Reputation Management

```
for all $m \in MI$ do
  Let $v$ be the volunteer and $s$ be the sentinel monitoring channel $c$.
  for all $i \in WUI$ do
    Maintain volunteer failure score, $F_m^v$, of $v$ for $MI$ $m$.
    Success score of $WUI$ $i$, $S_i \leftarrow 0$
    Failure score of $WUI$ $i$, $F_i \leftarrow 0$
    Number of verified intervals, $T \leftarrow 0$
    for each access slot $k$ do
      if $s$ does not have a sensing failure then
        if both $v$ and $s$ monitor channel $c$ then
          $T \leftarrow T + 1$
          if $O_{v,c}^{i,k} = O_{s,c}^{i,k}$ then
            $S_i \leftarrow S_i + 1$
          else
            $F_i \leftarrow F_i + 1$
          end if
        end if
      end if
    end for
```

```
if $F_i > \psi.T$ then
    $F_m^v \leftarrow F_m^v + F_i$
    Reputation, $\Gamma_{v,c}^i = \Gamma_{v,c}^{i-1} - e^{\alpha.F_m^v}$
    if $\Gamma_{v,c}^i < \theta$ then
        1) Save the volunteer failure score, $F_m^v$ and expel volunteer $v$ from monitoring $c$.
        2) Calculate $Cost$ incurred
        3) Select a new volunteer, if available. Otherwise, quit.
    end if
else
    Reputation, $\Gamma_{v,c}^i = \Gamma_{v,c}^{i-1} + \chi.S_i$
end if

=0
```

average of the *Cost* incurred is calculated over all the Monitoring Intervals that are tested in the simulation. If an intruder is caught at every Monitoring Interval, then it highlights a robust monitoring scheme and is considered a success.

In the first experiment, an honest volunteer attempts to catch an intruder in every *MI*, wherein the intruder attempts intrusion in each MI until it is caught in that MI. Here, I assess the performance of the monitoring scheme with one honest volunteer who monitors spectrum (using the methodology in Section 5.2.1) for different intruder profiles, across multiple sensing failure probabilities of the volunteer. The results of these experiments are shown in Tables 2.1-2.3. Here, the probability of sensing failure of v, P_f^v, is varied by selecting values in the range given by $P_f^v = \{0.1, 0.2, 0.3\}$. The type of intruder is also varied by having values in the range of $\eta = \{1, 3, 5\}$ and $\rho = \{0.3, 0.5, 0.8\}$. I observe that the average *Cost* increases as the probability of sensing failure of a volunteer increases because the average number of *WUIs* utilized by an intruder, before being caught by an honest volunteer increases with the increase in sensing failures. As shown in Tables 2.1-2.3, the proposed scheme is successful over all *MIs* for all the variations in sensing failure probability and types of intruders in the experiments. This establishes the robustness and success of our scheme when an honest

Table 2.1: Cost, Success with Honest Volunteer, $P_f^v = 0.1$

Intruder Profile	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.8$
$\eta = 1$	1.40, Yes	1.44, Yes	1.40, Yes
$\eta = 3$	1.24, Yes	1.46, Yes	1.58, Yes
$\eta = 5$	1.54, Yes	1.56, Yes	1.68, Yes
Average Cost = 1.48, Success = Yes			

Table 2.2: Cost, Success with Honest Volunteer, $P_f^v = 0.2$

Intruder Profile	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.8$
$\eta = 1$	1.78, Yes	1.58, Yes	1.62, Yes
$\eta = 3$	1.56, Yes	1.68, Yes	1.78, Yes
$\eta = 5$	1.72, Yes	1.70, Yes	1.86, Yes
Average Cost = 1.70, Success = Yes			

Table 2.3: Cost, Success with Honest Volunteer, $P_f^v = 0.3$

Intruder Profile	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.8$
$\eta = 1$	2.20, Yes	1.68, Yes	1.66, Yes
$\eta = 3$	1.58, Yes	1.76, Yes	1.80, Yes
$\eta = 5$	1.88, Yes	1.90, Yes	1.94, Yes
Average Cost = 1.82, Success = Yes			

volunteer monitors spectrum.

Table 3: Decision on Choosing Discrepancy Tolerance (ψ) to Avoid Eviction of Honest Volunteer

ψ	0.65	0.7	0.75	0.8	0.85	0.9
Eviction	Yes	Yes	Yes	Yes	No	No

Next, I assess the performance of our scheme against corrupt volunteers. For this purpose, I first determine the value of the Discrepancy Tolerance, ψ, such that honest volunteers are not expelled by mistake. An honest volunteer may also be unable to report the correct spectrum access state due to sensing failures. However, it is undesirable to expel an honest volunteer from the list of selected volunteers to monitor the spectrum, for the errors due to sensing failures. For this purpose, I empirically determine the value of ψ, with an honest volunteer having $P_f^v = 0.3$ against all the intruder profiles, As shown in Table 3, I avoid evicting honest volunteers in all cases when ψ is greater than or equal to 0.85. Hence, I utilize this value of ψ for the rest of our experiments. For the next set of experiments (whose results are shown in Tables 4.1-4.3), I measure the average *Cost* incurred and determine the *Success* of the spectrum monitoring scheme over different sensing failure rates of a corrupt volunteer across different intruder profiles. In these experiments, a corrupt volunteer monitors the spectrum at the beginning of each *MI*, and the sentinel attempts to expel the corrupt volunteer in every *MI*. Here, the $Cost$ incurred is calculated until the corrupt volunteer is expelled by the sentinel by using Algorithm 7. This is because after the expulsion of a corrupt volunteer, the sentinel itself can report and catch the intruder, without incurring any additional *Cost*. As expected, the average *Cost* incurred with corrupt volunteers (in Tables 4.1-4.3) is greater than the average *Cost* incurred with honest volunteers (in Tables 2.1-2.3) for all the sensing failure probabilities, P_f^v, of volunteer v. This is because a corrupt volunteer colludes with the intruder and thereby helps the intruder to use the channel for more *WUIs*, before it is expelled and the intruder is caught by the sentinel. Also, I observe

Table 4.1: Cost, Success with Corrupt Volunteer, $P_f^v = 0.1$

Intruder Profile	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.8$
$\eta = 1$	8.72, Yes	6.62, Yes	7.18, Yes
$\eta = 3$	7.44, Yes	6.86, Yes	6.88, Yes
$\eta = 5$	6.84, Yes	7.44, Yes	6.32, Yes
Average Cost = 7.14, Success = Yes			

Table 4.2: Cost, Success with Corrupt Volunteer, $P_f^v = 0.2$

Intruder Profile	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.8$
$\eta = 1$	8.62, Yes	6.34, Yes	7.06, Yes
$\eta = 3$	7.34, Yes	6.86, Yes	6.88, Yes
$\eta = 5$	6.84, Yes	7.44, Yes	6.32, Yes
Average Cost = 7.08, Success = Yes			

Table 4.3: Cost, Success with Corrupt Volunteer, $P_f^v = 0.3$

Intruder Profile	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.8$
$\eta = 1$	7.84, Yes	6.32, Yes	6.92, Yes
$\eta = 3$	7.34, Yes	6.86, Yes	6.88, Yes
$\eta = 5$	6.82, Yes	7.44, Yes	6.30, Yes
Average Cost = 6.97, Success = Yes			

that the corrupt volunteer is successfully evicted in all *MIs*. Another interesting observation that I make in Tables 4.1-4.3 is that the average cost incurred over all the *MIs* and against all intruder types, decreases with the increase in P_f^v. This is unlike what I see in the case of an honest volunteer (in Tables 2.1-2.3). This is because more sensing failures are likely to result in more discrepancy between the volunteer's observation and sentinel's observation of spectrum access state and hence, would result in faster expulsion of the corrupt volunteer by using Algorithm 7.

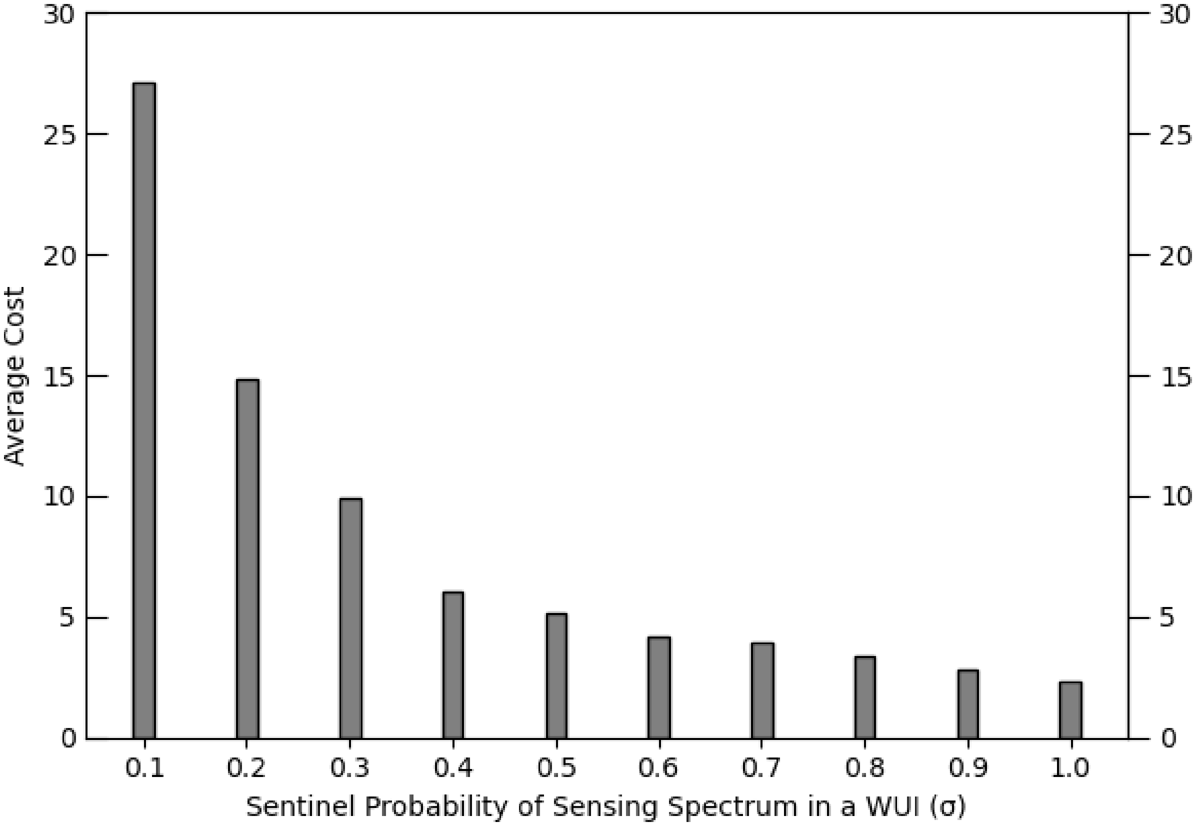


Figure 21: Average Cost Incurred With Change in Sentinel's Probability of Sensing (σ) when a Corrupt Volunteer Monitors Spectrum .

Finally, as shown in fig. 21, I assess the impact of varying the sentinel's Probability of monitoring spectrum (also given by σ) in a *WUI*, when a corrupt volunteer monitors spectrum. It is observed that the average *Cost* incurred reduces with the increase in σ. This is because if a sentinel monitors spectrum more frequently, then the corrupt volunteer can be expelled earlier and thereby incurring lesser average *Cost*. However, due to the advanced technological capabilities and limited availability of sentinels, monitoring spectrum too frequently would incur high overheads and expenses. Hence, there is a need to balance the sentinel's frequency of monitoring spectrum and the average *Cost* incurred. In our experiment, it is observed that even with the lowest value of σ that I tested (i.e., for $\sigma = 0.1$), the average *Cost* incurred is still small enough to expel the corrupt volunteer and catch the intruder before a *MI* ends.

5.5 Summary

In this chapter, I focused on methodologies to ensure robust detection of unauthorized spectrum access in shared spectrum networks, consistently over prolonged intervals of time. To this end, I proposed a framework that is driven by efficient spectrum sampling by crowdsourced agents. This is augmented by a methodology for efficient volunteer reputation management to ensure robustness against corrupt volunteers and intruders, irrespective of their propensity to take risk for illegally using spectrum. It was observed that the proposed framework worked successfully in establishing robustness over different intruder profiles, volunteer types and sensing failure rates.

6.0 AI-Based Methodologies For Estimating Location Likelihood of Volunteers

Ensuring effective volunteer selection for spectrum monitoring involves the selection of individuals qualified to consistently monitor the spectrum over extended periods. In Chapter 4, the process of selecting qualified volunteers is outlined through the utilization of a combination of Multi-Choice Secretary and Stable matching-based algorithms. Volunteers considered for monitoring a specific geographical region, r, over a monitoring interval should demonstrate both trustworthy behavior and a high likelihood of remaining within r during the monitoring interval. A reliable reputation management scheme, introduced in Chapter 5, is employed to assert the trustworthiness of the active volunteers. While the prediction of a volunteer's location likelihood can be achieved using metrics such as sojourn time, proportion of residence time, and duration to destination (as discussed in Chapter 4), forecasting volunteer mobility is challenging due to its dynamic and unpredictable nature. To address this challenge, AI-based methodologies are explored in this chapter to develop a volunteer mobility model capable of accurately predicting future volunteer locations.

This chapter addresses the challenges associated with learning volunteer mobility patterns and predicting their future locations. A mobility model is introduced, employing ML-based techniques to learn volunteers' mobility patterns and forecast their future locations. The performance of these ML-based techniques is primarily assessed through a comparative analysis of their accuracy in predicting future volunteer regions and trajectories.

The chapter is organized as follows: the motivation behind using ML models to predict volunteers' mobility is discussed, followed by a detailed description of the volunteer mobility model. The experimental setup to assess the performance of the proposed ML algorithms is then described, followed by insights derived from their analysis. Finally, the experimental results are presented and statistically validated.

6.1 Motivation

As discussed in Chapter 4, volunteers are selected to monitor the spectrum in a specific geographical region during a monitoring interval. Therefore, for a volunteer to be considered a candidate for selection to monitor the spectrum in the region r during the next monitoring interval (MI), m, the volunteer must reside in r over m.

This necessitates predicting the future locations of volunteers. Estimating the future locations of mobile volunteers depends upon various factors, including their historical location data, mobility patterns, probable places of residence and work, itinerary, and overall profile. The task of predicting the future locations of volunteers is challenging due to several factors:

1. Different individuals exhibit unique mobility patterns and preferences. Creating a generalized model for location prediction that accommodates diverse volunteer behaviors and preferences is challenging, as one-size-fits-all approaches may not effectively capture individual variances.
2. Volunteer intentions and contexts may evolve. Predicting locations requires considering the dynamic nature of user contexts, such as work, home, or recreational activities, and adapting models to changing preferences and habits.
3. Trace location data may be affected by inaccuracies, noise, or interference from various sources. Filtering out noise and ensuring the quality of input data is crucial for building reliable location prediction models.
4. Volunteer mobility can have varying patterns based on time of day, day of week, and even seasons. Predicting future locations requires accounting for these temporal variations.
5. Volunteer mobility is uncertain, as individuals may change their plans or deviate from usual routines.

To address the above challenges, a volunteer mobility model is built to capture the mobility patterns of volunteers, which will be discussed in the following sections.

6.2 Volunteer Mobility Model

In order to develop a volunteer mobility model, I analyzed datasets that contain GPS trajectories of taxi drivers [186, 187] and common people [188–190] in Bejing that were collected over a period of 1 week and 5 years, respectively. The analysis of these datasets guided the development of the mobility model used to determine the likelihood of a volunteer being present in a given region over a specific monitoring interval. In the following, the sequential nature of the data is discussed first, followed by an elaboration on the basic components of the model.

Sequential data refers to a type of data where the order and arrangement of elements are important. Here, each data point is positioned in a specific sequence, and the relationships or patterns between these elements are often crucial for understanding the data's underlying structure or behavior [191]. Sequential data is valuable because understanding a sequence of information is crucial for gaining insights and making future predictions. Examples of sequential data include time series, where observations are recorded over time, and sequences of events or symbols, such as sentences in natural language, DNA sequences [192, 193], or musical notes in a composition [194]. Sequential data helps to address the challenges of location prediction in the following ways.

1. It helps to capture the temporal dynamics of volunteer mobility. It facilitates understanding how location location at one point in time helps to estimate the volunteer location in subsequent time points.
2. It helps in recognizing recurrent patterns in the volunteer movement. Daily routines, periodic trips, or other cyclic behaviors become evident when analyzing data sequentially, aiding in the prediction of future locations based on historical trends.
3. It facilitates the detection of irregularities in a volunteer's movement patterns. Sudden deviations from regular behavior can be indicative of changes in plans or unexpected events. This contributes to more robust prediction models.
4. Such type of data also enables the personalization of location prediction models. By understanding an individual's unique mobility patterns, models can tailor predictions

to the specific profile of that volunteer, improving the relevance and accuracy of future location prediction.

Based on the sequential nature of the data, it is possible to predict the trajectory a volunteer is likely to follow on the way to their intended destination. To determine this trajectory, taking into consideration the contextual dependencies discussed above, a methodology based on machine learning is adopted.

The trajectory of a volunteer is a sequence of location time steps. To capture the dependencies, trends, or patterns that evolve over the ordered sequence of data points, a machine learning-based algorithm should carry out the analysis of the sequential data to predict the next location based on previous location time steps. To carry out the analysis, Recurrent Neural Network (RNN) and Transformer-based methodologies are adopted.

Recurrent Neural Networks (RNNs) are commonly employed in machine learning to handle and extract meaningful information from sequential data. The architecture of RNNs and their variants are designed to capture dependencies in sequential data and make predictions based on historical context. Therefore, I utilize the two variants of RNNs, Long Short Term Memory (LSTM) networks and Gated Recurrent Units (GRU), to model sequential mobility data for predicting the future location of volunteers. In addition, the recent popularity of Transformers being able to capture long-term dependencies motivates the comparison of a Transformer-based mobility model with the LSTM and GRU-based models.

6.3 Recurrent Neural Networks

Rumelhart et al. introduced the backpropagation algorithm, a foundational supervised learning technique enabling efficient computation of gradients for neural network weights, thus facilitating error minimization during training [191]. This work demonstrated the applicability of backpropagation to sequential data and thus laid the groundwork for the development of recurrent neural networks (RNNs). It emphasized the learning of distributed representations, showcasing the network's ability to capture intricate features and relationships within the data. By addressing the credit assignment problem, it established neural networks as a potent model for machine learning, leading to subsequent advancements in deep learning and artificial intelligence [191]. Although the authors did not explicitly introduce the term "recurrent neural network", their work has significantly influenced the direction of neural network research, especially in the field of modeling and learning from sequential dependencies. [191].

RNNs handle sequential data by maintaining a hidden state that captures temporal dependencies and adapts to variable-length sequences. Their ability to retain information from previous time steps enables context-aware processing, making them ideal for tasks where order matters [195]. RNNs, unlike traditional feedforward neural networks, are characterized by their directed cycles. This unique structure allows them to capture dependencies in sequential information and display dynamic temporal behavior. The primary feature of RNNs is their ability to maintain hidden states that store information about previous inputs in the sequence, which enables them to consider context and learn patterns over time. The hidden state h_t of an RNN at time t is computed based on the current input x_t and the previous hidden state h_{t-1} through a set of learned parameters. Mathematically, this can be expressed as $h_t = f(W_h h_{t-1} + W_x x_t)$, where f is an activation function, W_h and W_x are weight matrices, and h_t represents the hidden state at time t [195, 196].

A critical issue in training RNNs is the problem of vanishing gradients [195]. It arises during the backpropagation process, where the gradient of the loss function is calculated and propagated backward through the network to update the weights. This issue becomes pronounced, especially when dealing with long sequences or dependencies. The vanishing gradient problem occurs because as the gradients are backpropagated through time steps, they can diminish exponentially and approach zero. This phenomenon is problematic for RNNs attempting to capture long-term dependencies, as the influence of early time steps becomes negligible during the process of updating weights. Consequently, the network struggles to learn and remember information from distant past inputs, hindering its ability to capture long-range dependencies in sequential data [195, 196]. In RNNs, the same set of weights is reused across all time steps, and during backpropagation, the gradients are multiplied by these weights at each step. If these weights are small, the gradients shrink exponentially with each time step, leading to vanishing gradients [195]. This issue impacts the network's capacity to learn and represent sequential patterns effectively, especially in tasks where understanding long-term dependencies is crucial. To address the vanishing gradient problem, more advanced architectures, such as LSTM networks and GRUs, have been developed. These architectures incorporate mechanisms like memory cells and gating units that enable the network to selectively retain and update information over extended sequences, mitigating the vanishing gradient problem [195, 196].

6.3.1 Long Short Term Memory

The LSTM architecture was devised to address the vanishing gradient problem inherent in training RNNs on long sequences of data [197]. The key innovation is in the incorporation of a memory cell and three gating mechanisms: *Input Gate*, *Forget Gate*, and *Output Gate*. The *Input Gate* modulates the influx of new information into the Memory Cell and makes decisions about which components of the input should be stored for future reference [197]. Meanwhile, the *Forget Gate* is in charge of determining what information in the Memory Cell is to be discarded, facilitating the selective removal of irrelevant data. The *Output Gate* assumes the role of a decision-making entity. It combines details from the current input and the Memory Cell contents. This process helps decide which part of the information should be shared as the final output from the LSTM cell. This mirrors the gate's capacity to modulate the flow of information, contributing to the network's adaptability and its ability to selectively propagate relevant information. The Memory Cell, serving as a dedicated storage unit, is adept at retaining information over extended sequences, thus mitigating the vanishing gradient problem. In other words, this is achieved by using gating mechanisms that let the network adjust information flow, deciding what to remember or forget [195,197,198]. The use of sigmoid and tanh activation functions in the gating mechanisms ensures that information flow is controlled continuously. Importantly, the LSTM architecture incorporates the concept of ”constant error carousel”, which aids in preserving error signals across multiple time steps, facilitating the propagation of gradients through time without degradation [197, 198]. The weighted connections associated with the gates are learned during training through backpropagation, enabling the network to adapt and optimize its performance on the given task. This comprehensive architecture, with its nuanced control over information flow, has proven highly effective in capturing both short-term and long-term dependencies in sequential data, making LSTMs particularly well-suited for a diverse range of applications, including natural language processing and time series prediction [197, 198].

6.3.2 Gated Recurrent Unit

Cho et al. introduced the Gated Recurrent Unit (GRU), which is a type of RNN architecture. Both GRU and LSTM are designed to address the vanishing gradient problem and enable RNNs to capture long-range dependencies in sequential data [199]. The key innovation in both architectures is the incorporation of gating mechanisms that regulate the flow of information within the network [200]. The GRU simplifies the architecture of the LSTM by combining the memory cell and hidden state into a single state, and it uses two gates: an update gate and a reset gate.

The GRU architecture consists of the following [195, 199]:

1. Candidate Hidden State: The candidate hidden state is a new hidden state that could be updated to become the next hidden state. It is calculated based on the input, the reset gate, and the previous hidden state.
2. Hidden State: The hidden state is updated by blending the candidate hidden state and the previous hidden state using the update gate. It represents the current memory or information retained by the network.
3. Reset Gate: The reset gate controls the amount of past information to be forgotten. It influences the information that is discarded from the previous hidden state in calculating the new candidate hidden state.
4. Update Gate: The update gate controls the trade-off between the new candidate hidden state and the previous hidden state. It determines how much of the past information should be carried forward to the future.

In the following, the tradeoffs between LSTM and GRU networks are discussed.

6.3.3 Tradeoffs between LSTM and GRU

The choice between GRU and LSTM networks involves tradeoffs based on the requirements of a task. The following are some key tradeoffs between GRU and LSTM [195,199,201]:

1. The architecture of LSTMs is more complex compared to GRUs. They consist of an input gate, an output gate, a forget gate, and a memory cell. On the other hand, GRUs are simpler, primarily because they combine the memory cell and hidden state into a single state and utilize only two gates - the update gate and the reset gate.
2. LSTMs may require more careful tuning of hyperparameters to achieve optimal performance. GRUs are often more straightforward to use and may require less fine-tuning of hyperparameters.
3. Since LSTMs usually have more parameters compared to GRUs, they can potentially be more prone to overfitting, especially with limited data. GRUs have fewer parameters, which can be helpful when there are smaller datasets or when there are limited computational resources.
4. LSTMs can capture long-term dependencies in sequential data effectively. This is mainly because they have a dedicated memory cell and mechanisms to control the flow of information over extended time steps. On the other hand, GRUs may be less effective than LSTMs in modeling long-term dependencies, as they lack a separate memory cell.
5. The complexity of LSTMs can result in slower training speeds, especially when dealing with large datasets. GRUs often train faster due to their simpler architecture. This can be helpful, especially in applications where training efficiency is a critical factor.

Oftentimes, the performance of GRU and LSTM can vary based on the nature of the task. In some cases, GRUs may outperform LSTMs, while in others, LSTMs perform better. The choice often depends on empirical evaluation and experimentation. Therefore, experiments are conducted to do a comparative analysis of the effectiveness of LSTM and GRU to predict the future location of volunteers.

6.4 Transformer

The Transformer is a neural network architecture that was introduced by Vaswani et al. in 2017 [202]. The transformer architecture's primary innovation is its self-attention mechanism, which allows it to more efficiently capture dependencies between different positions in a sequence than traditional RNNs. "Unlike RNNs, the transformer does not process data in sequence (i.e. in order), which allows for more parallelization and reduces training time" by taking advantage of multiple GPU cores [203]. This model has proven to be highly successful in different machine-learning tasks, especially in natural language processing. Some of the major tradeoffs between transformers and RNN-based models (like LSTM and GRU) are [202, 204, 205]:

1. The entire sequence is processed in parallel by a transformer, making them more computationally efficient than RNN-based models, especially for long sequences, when trained on GPUs. However, RNNs process data sequentially. While this is less computationally efficient, it is, however, beneficial if the order of the sequence is important for the task [202, 204, 205].
2. Transformers do not inherently understand the sequential order of input data. Positional encoding is used to provide some information about the order of elements, but it may not always be as effective in capturing temporal dependencies as the sequential processing of RNNs [204, 205].
3. While transformers have a self-attention mechanism to capture relationships between different positions in a sequence, they may not explicitly maintain temporal memory over time. For certain time series tasks with long-range dependencies, models like LSTMs with dedicated memory cells could perform better [204, 205].
4. LSTMs and GRUs share parameters across time steps. This is beneficial for tasks requiring historical context. Transformers' position-wise attention allows them to attend to different positions in the sequence. This enables transformers to capture dependencies without using shared parameters [204, 205].

The choice between RNN-based models and transformers ultimately depends on the characteristics of the task and the nature of the data. Transformers might excel in scenarios where parallelized processing and attention mechanisms are crucial, while RNN-based models may be preferable for tasks with a clear sequential structure. Due to the recent popularity of Transformers and their advantages in parallelization to capture global contexts, its performance in predicting volunteer location is compared against LSTM and GRU.

6.5 Experimental Setup

The primary objective is to predict the future location of volunteers over a given monitoring interval. For this purpose, the LSTM, GRU, and Transformer models are evaluated on real-world mobility datasets of users. In the following sub-sections, I discuss the dataset and data preprocessing techniques, the experimental methodology, the metrics, the results of my experiments and their statistical validation.

6.5.1 Dataset and Data Preprocessing

The volunteer mobility models are assessed using the Geolife dataset which is a collection of GPS trajectories [188–190]. This dataset was gathered by Microsoft Research Asia over five years (from April 2007 to August 2012) from 182 users as part of the Geolife project. It consists of time-stamped points with latitude, longitude, and altitude information. The dataset comprises 17,621 trajectories covering a total distance of 1,292,951 kilometers and a total duration of 50,176 hours. These trajectories were recorded by various GPS loggers and GPS-enabled phones, with different sampling rates. Approximately 91.5% of the trajectories are densely represented, with points logged every 1-5 seconds or every 5-10 meters [188–190]. The dataset captures diverse outdoor movements, encompassing daily routines like commuting to work or going home, as well as recreational activities such as shopping, sightseeing, dining, hiking, and cycling. This trajectory dataset can potentially be applied in various fields, including mining mobility patterns, recognizing user activities, studying location-based social networks, addressing location privacy concerns, and making location-based recommendations. While the dataset is distributed across 30 cities in China, as well as some cities in the USA and Europe, the majority of the data originates from Beijing, China [188–190].

This dataset is cleaned and preprocessed for training and testing the AI-based models by undertaking the following steps:

1. All the rows with null or void columns are dropped.
2. Since RNN and Transformer models are usually proficient in learning long-term dependencies in large datasets, these models are trained and tested with the data of volunteers that have at least 5000 timestamped data points.
3. Since the majority of data points originate in Beijing, China, only the location of volunteers over Beijing, specifically in the region bounded by latitudes 39.6 and 40.2 and longitudes 116.0 and 116.8 are predicted. To this end, only the data points that are within this geographical area are utilized.

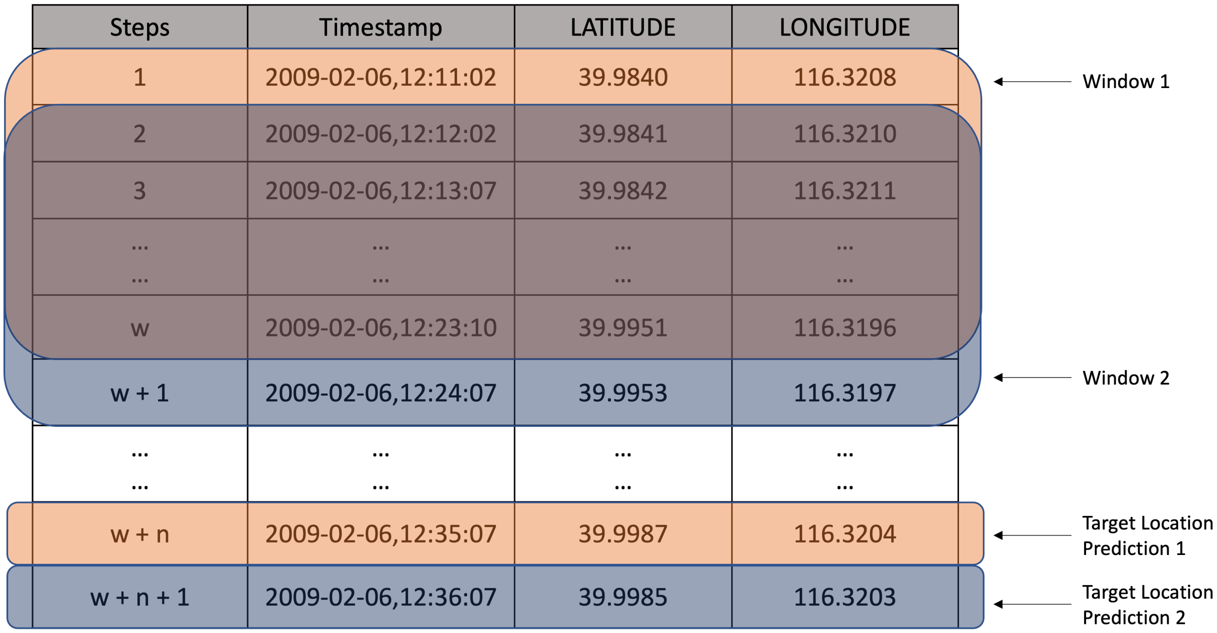
The following discusses the experimental methodology to do a comparative analysis of the performance of the models.

6.5.2 Methodology

In this work, the performance of the AI-based models is compared. In addition, a weighted random approach is also used to predict the future location of volunteers. This section highlights the methodologies utilized for both these approaches.

6.5.2.1 AI-Based Location Prediction Approaches

The LSTM, GRU, and Transformer models are trained by using a sliding window technique, such that a series of overlapping sequences or *windows* is created from the original sequential mobility data. As shown in Figure 22, the sequential data of latitudes and longitudes is divided into fixed-size *windows*. Each window of size w contains a number of consecutive location time steps. There is a corresponding target location, which is predicted based on w over n steps in the future. In order to capture the continuous patterns and dependencies in the sequential data, the *windows* overlap. This means that each time step is part of multiple *windows*. For example, with a *window* of size w, the window starting at time step i would be $i, i+1, i+2, ..., i+w$ and the target location to estimate would be in time step $i+w+n$. Similarly, the next window would be $i+1, i+2, i+3, ..., i+w+1$ and the target location to estimate would be in time step $i+w+n+1$. These overlapping *windows* are used to train the mobility models using a supervised learning approach. The input sequences are fed into the network, and the model is adjusted to minimize the difference between its predictions and the actual target outputs. This training process using sliding windows enables the mobility models to learn and capture temporal patterns, dependencies, and trends present in the time series mobility data. After training, the mobility models can be used to make predictions on new, unseen data by sliding the window similarly.



Steps	Timestamp	LATITUDE	LONGITUDE
1	2009-02-06,12:11:02	39.9840	116.3208
2	2009-02-06,12:12:02	39.9841	116.3210
3	2009-02-06,12:13:07	39.9842	116.3211
… …	… …	… …	… …
w	2009-02-06,12:23:10	39.9951	116.3196
w + 1	2009-02-06,12:24:07	39.9953	116.3197
… …	… …	… …	… …
w + n	2009-02-06,12:35:07	39.9987	116.3204
w + n + 1	2009-02-06,12:36:07	39.9985	116.3203

Figure 22: Sliding Windows for Training and Testing LSTM and GRU models

As discussed in chapter 4, a set of volunteers is selected to monitor the spectrum over an MI in a geographical region. Volunteer selection is facilitated by a hybrid algorithm built upon Secretary and Stable Matching algorithms. Volunteer selection, however, depends on the qualification of a volunteer to monitor the spectrum in a geographical region. One factor

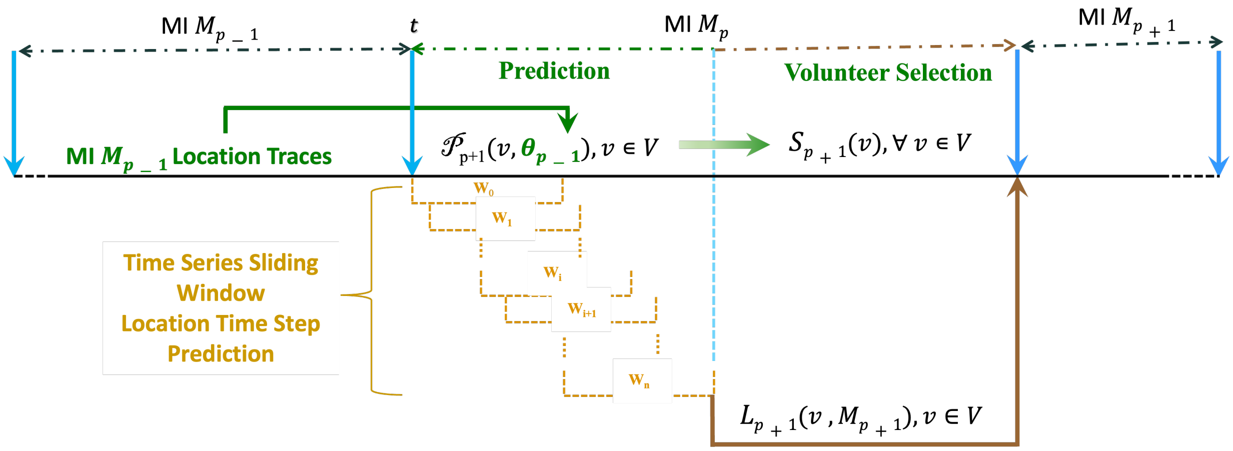

MI M_{p-1}
t
MI M_p
MI M_{p+1}
Prediction
Volunteer Selection
MI M_{p-1} Location Traces
$\mathscr{P}_{p+1}(v, \theta_{p-1}), v \in V$
$S_{p+1}(v), \forall v \in V$
W_0
W_1
W_i
W_{i+1}
W_n
Time Series Sliding Window Location Time Step Prediction
$L_{p+1}(v, M_{p+1}), v \in V$

Figure 23: Volunteer Region Prediction for Selection In a Monitoring Interval

that qualifies a volunteer to monitor the spectrum in a region r is their likelihood to be in r over an MI. For this purpose, the LSTM, GRU, and Transformer-based mobility models are utilized to predict the location of a volunteer at the start of an MI. This predicted future volunteer location is then fed to the Volunteer Selection algorithms to select and assign volunteers to all the monitoring regions in the spectrum enforcement area. As shown in figure 23, in a MI, M_p, the mobility models are used for location prediction, $\mathscr{P}_{p+1}(v, \theta_{p-1})$, of a volunteer v over MI M_{p+1} by learning v's mobility pattern from past location traces, θ_{p-1}, in MI, M_{p-1}. The LSTM, GRU, and Transformer-based mobility models utilize the sliding windows $w_0, w_1, w_2, ..., w_n$ to learn the temporal dynamics of the mobility pattern of volunteer v and predict the location of v at a future timestamp. The window, w_n, is used to predict the location, $L_{p+1}(v, M_{p+1})$, of v at the start of MI M_{p+1}. The predicted location using $\mathscr{P}_{p+1}(v, \theta_{p-1})$ is fed to the Volunteer Selection algorithm, $S_{p+1}(v)$, to aid in volunteer selection and assignment over the MI, M_{p+1}.

6.5.2.2 Weighted Random Location Prediction Approach

The current region of residence, r, of a volunteer, v, is derived from their current location, at every time point in the Geolife dataset. This information is used to identify the

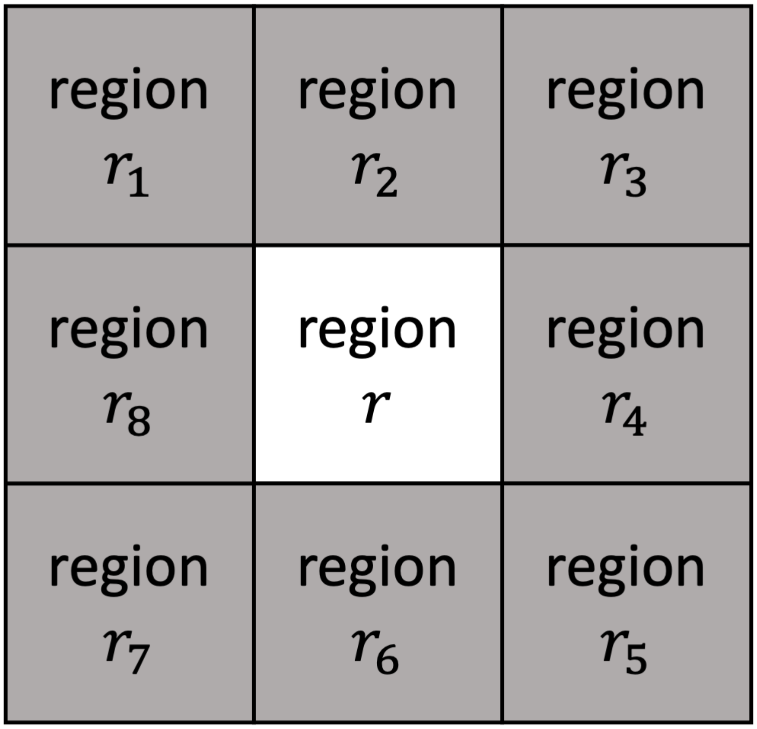


Figure 24: Neighboring Regions of the Current Region of Residence, r, of a Volunteer, v.

neighboring regions adjacent to r. As shown in Fig. 24, regions $r_1, r_2, r_3, r_4, r_5, r_6, r_7$ and r_8 are the neighbors of the current region, r. It is assumed that a volunteer will either stay in the current region, r, or move to one of the neighboring regions, at the next time point. To estimate the future region of residence of a volunteer, the distances between the current location, L_v, of a volunteer to the centroid of every neighboring region, including the current region, are calculated, as shown in the expression depicted in (32).

$$w_r^v = \frac{\frac{1}{d_{L_v,c_r}}}{\sum_{r_i \in R_N} \frac{1}{d_{L_v,c_{r_i}}}} \tag{32}$$

In the above expression, w_r^v is the weight associated with region r, d_{L_v,c_r} is the distance between the current location, L_v, of volunteer v and the centroid, c_r of the region r, R_N is the set of regions that includes the current region of residence, r, of v and all of its neighboring regions. The region of residence of v, over the next time point, is estimated in a weighted random manner, using weight, w_r^v, for region $r \in R_N$. Consequently, the probability of selecting region r as the region where a volunteer, v, resides in the next time point is given by $P(r) = w_r^v$. Therefore, the process of choosing the next region of residence, $r_i \in R_N$ involves calculating these weights and then selecting r_i according to the resulting probabilities, $P(r_i), \forall r_i \in R_N$.

6.5.3 Performance Metrics

The following metrics are used for evaluating the performance of the volunteer location prediction approaches.

- *Accuracy*: As discussed in Chapter 3, effective spectrum monitoring is ensured by dividing the geographical area into smaller regions. For this experiment, the geographical area is divided into 49 equal-sized square regions, as shown in Figure 25. The latitude and longitude values that are predicted by the deep learning models for a volunteer are mapped to one of these regions. To determine the accuracy of prediction of the LSTM and GRU models, their predicted region for a volunteer is compared to the actual region in which the volunteer resides at a given time in the future.

  Assuming that,

  - γ_v is the number of predictions made for a volunteer, v.
  - X_v^i is an indicator function such that $X_v^i = 1$ if the predicted region is the same as the actual region for the i^{th} location prediction of a volunteer, v, and $X_v^i = 0$ otherwise.

  The accuracy, A_v, in predicting the future regions of a volunteer, v, is defined as the sum of the indicator functions divided by the total number of predictions.

$$A_v = \frac{1}{\gamma_v} \sum_{i=1}^{\gamma_v} X_v^i \tag{33}$$

  In other words, accuracy is the ratio of the frequency of correct predictions to the total number of predictions.

- *Root Mean Square Error (RMSE)*: RMSE provides a measure of the average deviation between predicted and actual values of latitude and longitude, with lower RMSE values indicating better predictive accuracy. This is given by (34) as shown below.

$$\text{RMSE}_v = \sqrt{\frac{1}{\gamma_v}\sum_{i=1}^{\gamma_v}(y_v^i - \hat{y}_v^i)^2} \tag{34}$$

  In (34), γ_v represents the total number of predictions made for a volunteer v, y_v^i represents the observed latitude and longitude values for the i^{th} data point, and $\hat{y}_v^i$ represents the predicted latitude and longitude values for the i^{th} data point.

- *Geodesic Distance*: The geodesic distance, d_g^v, between the predicted location and the actual location of a volunteer, v, is the next performance metric that is used. It is a measure of the shortest path between the predicted and actual location of v. It takes into account the curvature of the sphere and provides a more accurate distance calculation for points specified by latitude and longitude than a simple Euclidean distance. It is often used in geography, navigation, and mapping applications where precise measurements of distances over the Earth's surface are required [206, 207]. A lower geodesic distance between the predicted and actual location of a volunteer indicates better performance of the volunteer mobility model.

  One common method to calculate the geodesic distance is using the Haversine formula, which considers the spherical geometry of the Earth [206, 207]. As shown in (35), it involves trigonometric functions to determine the great-circle distance between two points on the Earth's surface, given their latitude and longitude coordinates.

$$\begin{aligned} d_g^v = R \cdot 2 \cdot \arcsin\bigg(& \\ & \sqrt{\text{hav}(\Delta\text{lat}) + \cos(\text{lat}_1)\cdot\cos(\text{lat}_2)\cdot\text{hav}(\Delta\text{lon})}\bigg) \end{aligned} \tag{35}$$

  where,

$$\text{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right)$$

In (35), R is Earth's radius (mean radius = 3,958.8 miles), Δlat is the difference between predicted and actual latitudes, Δlon is the difference between predicted and actual longitudes, $\cos(\text{lat}_1)$ and $\cos(\text{lat}_2)$ are the cosines of the predicted and actual latitudes, $\text{hav}(\theta)$ is the haversine function which is equal to the sine of half of the central angle between the points.

- *Execution Time*: The execution time is a metric that gauges the average duration required to train and test a deep-learning model for predicting a volunteer's future location. A lower execution time indicates a better performance. .

6.5.4 Results

After the *Geolife* dataset is cleaned and preprocessed using the steps underlined in Section 6.5.1, the number of volunteers whose location traces are used to train and test the deep learning models is 116. In both the LSTM and GRU-based volunteer mobility models, there are four LSTM and GRU layers, respectively. The first three layers consist of 64 units each and return the entire sequence of outputs, while the fourth layer, with 32 units, only returns the final output in the sequence. The Rectified Linear Unit (ReLU) activation function is used in all the layers to promote non-linearity in the model. The architecture of both the mobility models concludes with a final Dense layer featuring two units, indicative of its application to regression tasks aiming to predict the future latitude and longitude values of a volunteer.

The transformer model is designed for time series regression. With an input feature dimensionality of 2, representing latitude and longitude, and a sequence length of 12 timestamps, the model's architecture is tailored for the specific characteristics of the data. It comprises an initial linear embedding layer that transforms the input features into a hidden space of size 64. The transformer encoder stack, consisting of two layers with four attention heads each, is used to capture temporal dependencies in the data. A dropout layer is added with a dropout rate of 0.2 to avoid overfitting the data to the model. The final linear layer maps the flattened output from the transformer to a 2-dimensional output, to predict the target latitude and longitude coordinates. The model utilizes the mean squared error loss

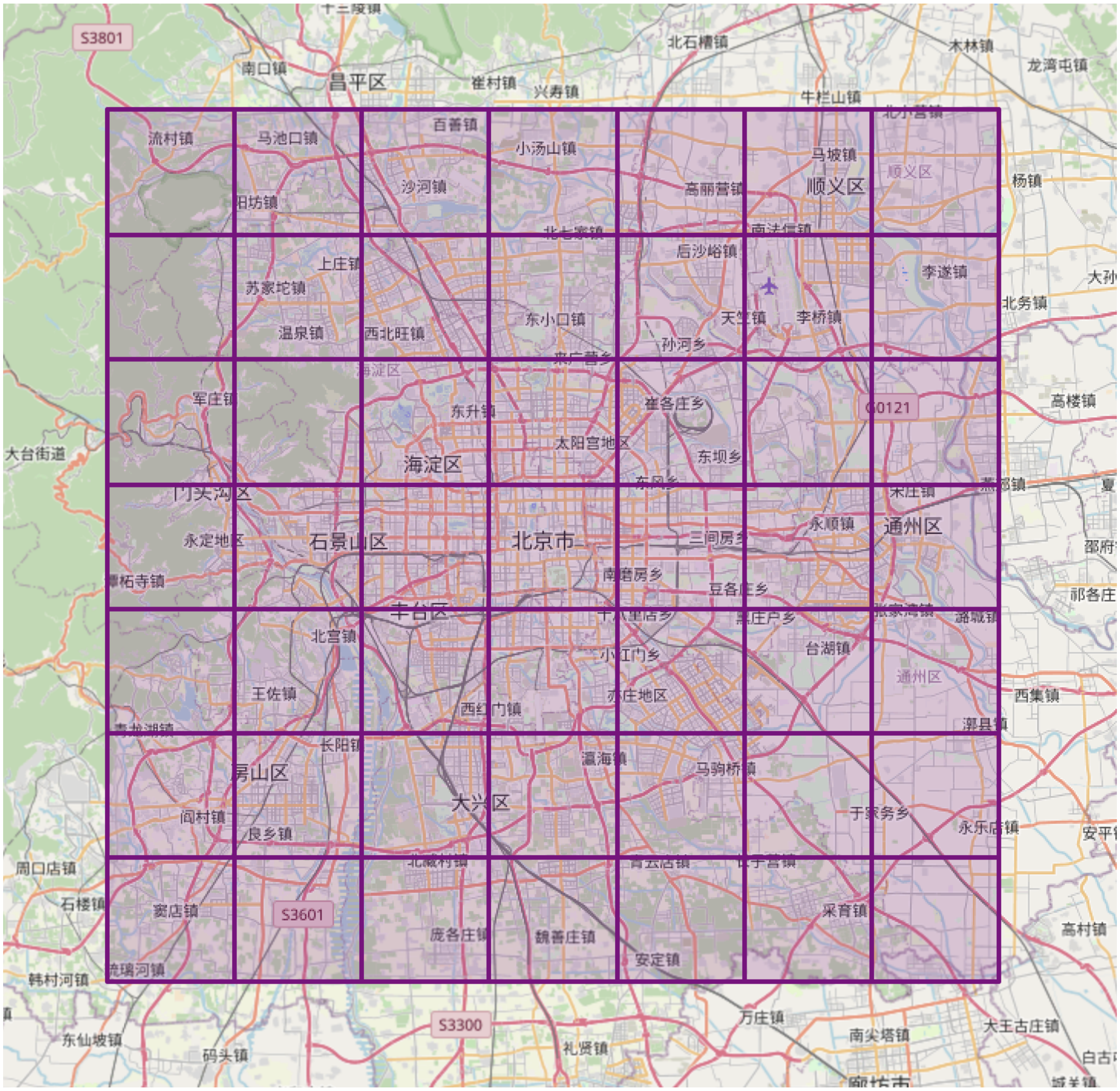


Figure 25: Regions in the monitoring area

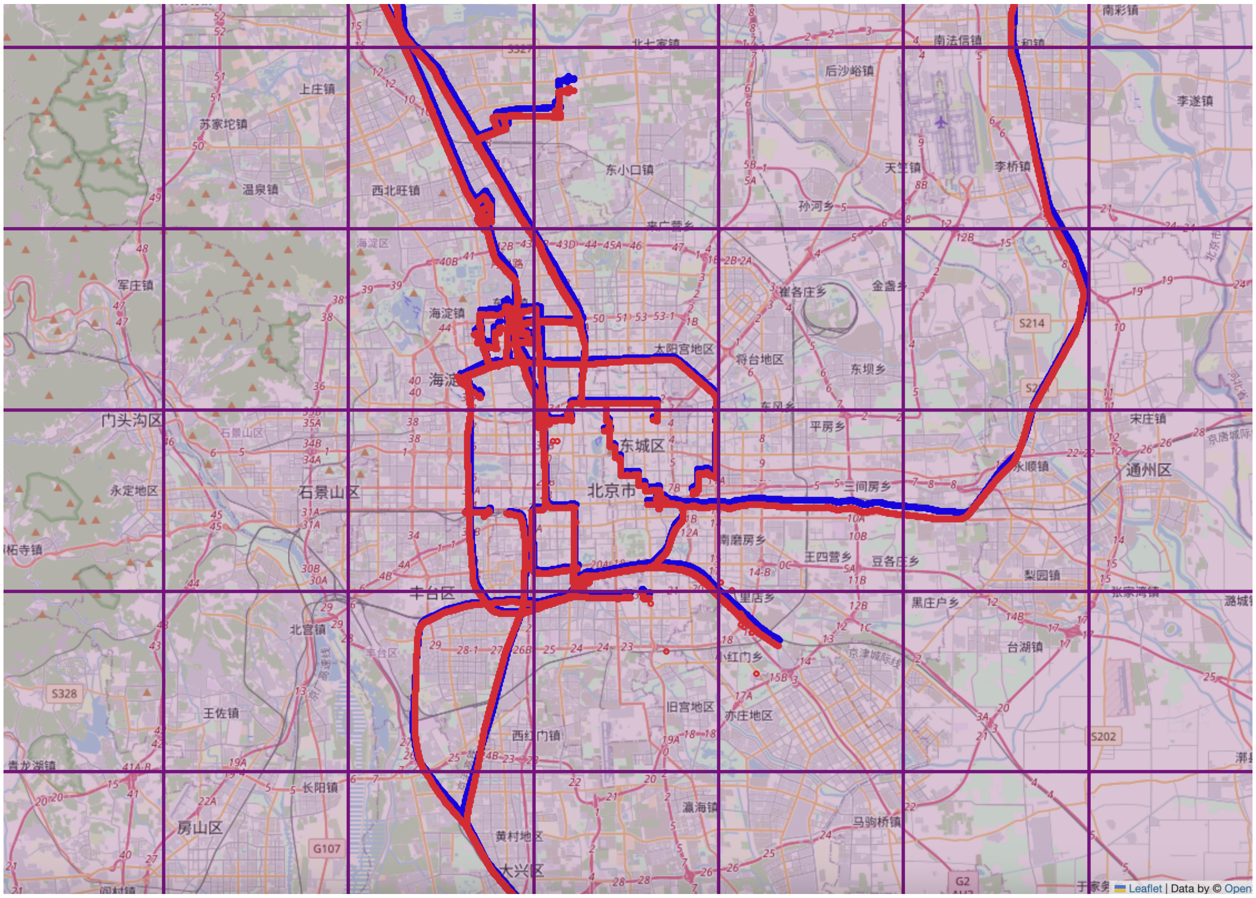


Figure 26: Predicted Trajectory(in red) versus Actual Trajectory (in blue) of a Volunteer Using GRU

and Adam optimizer with a learning rate of 0.0005 to optimize the prediction accuracy over 10 training epochs.

The first 70% of the location traces of each volunteer is used for training and the remaining 30% is used for testing. While training and testing, a sliding window size is 12 and the 12^{th} location sequence from the end of the window is predicted. During training, the model's weights are iteratively updated to minimize the specified loss function over several epochs. The training concludes after 10 epochs.

Figures 26, 27 and 28 exhibit the predicted trajectory versus the actual trajectory of a volunteer picked randomly from the Geolife dataset. The performance of the three mobility

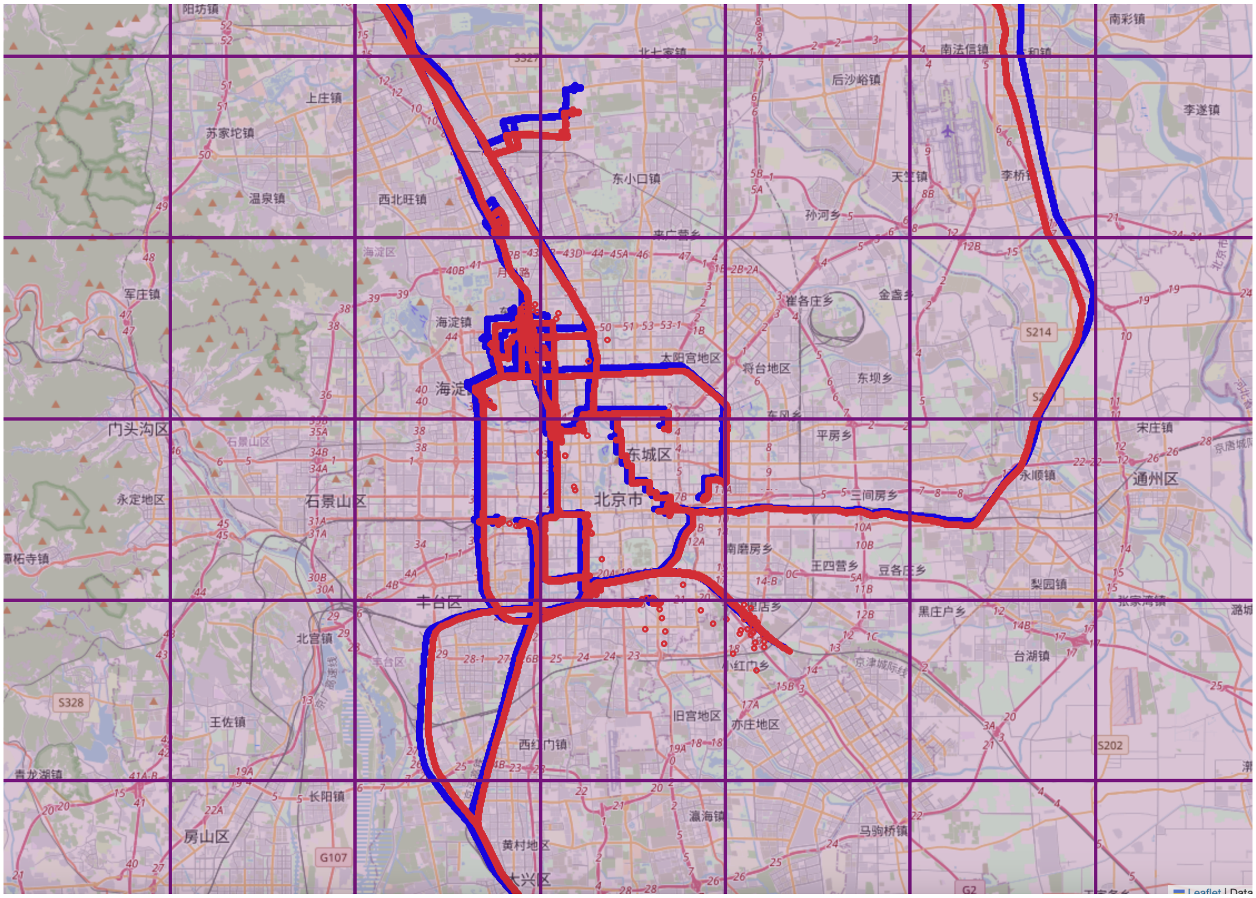


Figure 27: Predicted Trajectory(in red) versus Actual Trajectory (in blue) of a Volunteer Using LSTM

models is evaluated by using the four performance metrics discussed in Section 6.5.3. As shown in Table 6.1, the mean accuracy in predicting the future region of all the volunteers by the GRU-based mobility model is 0.9234, which is higher than the accuracy of prediction achieved by the LSTM and Tranformer-based mobility models. Similarly, as shown in tables 6.2 and 6.3, the GRU-based model outperforms the LSTM and Transformer in both mean RMSE and mean geodesic distance (in miles). Finally, as shown in Table 6.4, the GRU-based model takes an average of 299.428 seconds to train and test on the location trace of each user, compared to LSTM, which takes an average of 311.521 seconds, and the Transformer, which takes an average of 320.636 seconds. This is expected as GRUs train faster due to their

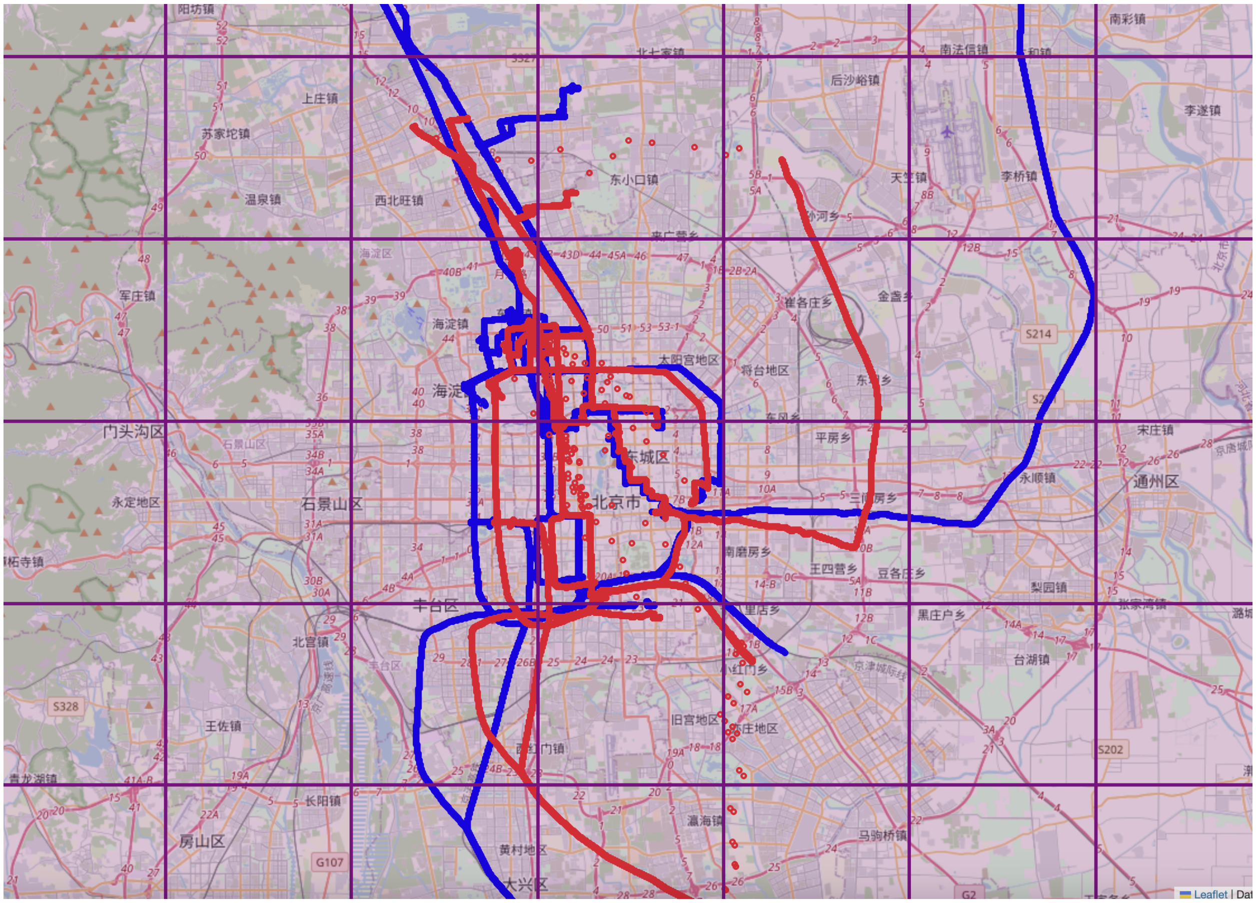


Figure 28: Predicted Trajectory(in red) versus Actual Trajectory (in blue) of a Volunteer Using Transformer

simpler architecture. Therefore, based on these performance metrics, a GRU-based mobility model is preferable to use compared to LSTM and Transformer. However, it is necessary to determine if the difference in performance between the three models is statistically significant.

In addition to the AI-based approaches, volunteer location prediction is also done using a weighted random approach as discussed in section 6.5.2.2. For this approach, all the non-null data points of all the users in the Geolife dataset are used for prediction. This is because, unlike the AI-based approaches, using this approach does not require training on the volunteer mobility data, to predict future locations. To be consistent with the sliding window size of 12 used in the AI-based approaches, given a location time step, l, I predict the

Table 6.1: Accuracy

Model	**Accuracy**		
	Mean	**Std. Deviation**	**95% Confidence Interval**
LSTM	0.9179	0.1018	[0.8949, 0.9375]
GRU	**0.9234**	0.1078	**[0.8977, 0.9449]**
Transformer	0.8151	0.1737	[0.7879, 0.8416]
Weighted Random	0.0207	0.0313	[0.0154, 0.0257]

Table 6.2: RMSE

Model	**RMSE**		
	Mean	**Std. Deviation**	**95% Confidence Interval**
LSTM	0.0063	0.0067	[0.0047, 0.0081]
GRU	**0.0053**	0.0053	**[0.0043, 0.0065]**
Transformer	0.0135	0.0102	[0.0107, 0.0162]

region of residence of a volunteer after every location time step and compare the predicted region with the actual region only after the 12^{th} location time step from l. As shown in Table 6.1, the accuracy of predicting the future region of a volunteer using this approach is 0.0207, which is significantly lower than all the AI-based approaches. This is expected because of the random nature of prediction.

Table 6.3: Geodesic Distance

Model	Geodesic Distance (miles)		
	Mean	**Std. Deviation**	**95% Confidence Interval**
LSTM	0.5298	0.55213	[0.3992, 0.6819]
GRU	**0.4448**	0.44495	**[0.3609, 0.5437]**
Transformer	1.1440	0.8467	[0.9025, 1.3732]

Table 6.4: Execution Time

Model	Execution Time (seconds)		
	Mean	**Std. Deviation**	**95% Confidence Interval**
LSTM	311.521	448.359	[220.0406, 404.1181]
GRU	**299.428**	411.566	**[211.3795, 390.4299]**
Transformer	320.636	453.153	[223.1013, 394.0880]

6.5.5 Statistical Validation of The Experimental Results

The experimental results have been statistically validated using two methods: non-overlapping batch means and the paired t-test. The following provides a brief overview of these methods and discusses the results obtained through their application.

6.5.5.1 Non-Overlapping Batch Means

Nonoverlapping Batch Means (NOBM) is a statistical method commonly used in simulation output analysis, especially for estimating steady-state means and constructing confidence intervals [208]. The objective of NOBM in my experiments is to estimate the mean of a performance metric value μ_O from a sequence of observations $\{O_i : i = 1, \ldots, n\}$ generated by testing the LSTM and GRU-based volunteer mobility models on location traces of 116 volunteers in the dataset.

The sequence of observations is divided into B batches, each of size z. This results in $n = Bz$ observations for each performance metric. This is done without overlapping, which ensures that each data point belongs to only one batch [208].

For each batch j (where $j = 1, 2, \ldots, B$), the batch mean $Y_j(z)$ is calculated by using (36)

$$Y_j(z) = \frac{1}{z} \sum_{i=z(j-1)+1}^{zj} X_i \tag{36}$$

The grand mean, Y, which is the average of all batch means, is calculated using (37).

$$Y = \frac{1}{B} \sum_{j=1}^{B} Y_j(z) \tag{37}$$

For each batch j, the batch variance $S^2_{z,B}$ is obtained from (38).

$$S^2_{z,B} = \frac{1}{B-1} \sum_{j=1}^{B} [Y_j(z) - Y]^2 \tag{38}$$

The $100(1-\alpha)\%$ confidence interval for μ_O is given by (39).

$$Y(z, B) \pm t_{1-\alpha/2,B-1} \frac{S_{z,B}}{\sqrt{B}} \tag{39}$$

where $t_{1-\alpha/2,B-1}$ is the critical value from the t-distribution with $B-1$ degrees of freedom. As $z \to \infty$ with B fixed (so that $n \to \infty$), the NOBM confidence interval is asymptotically valid. The tables 6.1, 6.2, 6.3 and 6.4 show the 95% confidence interval of the four performance metrics with batch size, $z = 10$.

6.5.5.2 Paired t-tests

To statistically validate whether the results obtained by the GRU and LSTM models are significantly different, the paired t-test is utilized. In this methodology, the performance metric values of GRU and LSTM-based models in predicting the future location of all 116 volunteers are compared. For this purpose, paired t-tests are performed for the values of all the performance metrics that are discussed in Section 6.5.3 [209]. The null and alternative hypotheses for each paired t-test are defined as follows:

Null Hypothesis (H_0): Metric value of one model = Metric value of a different model

Alternative Hypothesis (H_1): Metric value of one model $\neq$ Metric value of a different model

A significance level (α) of 0.05 is chosen. The paired t-tests that were performed to assess whether there is a significant difference between the performance metric values of the two models are given by (40) [209].

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{40}$$

In (40), s_d is the standard deviation of the differences, $\bar{d}$ is the mean of the differences between paired observations, and n is the number of paired observations. The paired t-test yielded a t-statistic of t and a p-value of p. The decision rule was based on comparing the p-value to the chosen significance level:

$$\text{If } p < \alpha, \text{ reject the null hypothesis.}$$

$$\text{If } p \geq \alpha, \text{ fail to reject the null hypothesis.}$$

As shown in Tables 6.6 and 6.7, the p-values of mean accuracy, RMSE, and Geodesic Distance are lesser than the significance value of 0.05. This implies that there is a significant difference in the performance of the Transformer-based mobility model when compared to LSTM and GRU-based mobility models. Since the mean accuracy is lesser and the RMSE

Table 6.5: Paired t-test Results of LSTM versus GRU

Accuracy		**RMSE**		**Geodesic Distance**		**Execution Time**	
t	**p-value**	**t**	**p-value**	**t**	**p-value**	**t**	**p-value**
-0.559	0.577	2.953	**0.004**	3.107	**0.002**	1.419	0.159

Table 6.6: Paired t-test Results of LSTM versus Transformer

Accuracy		**RMSE**		**Geodesic Distance**		**Execution Time**	
t	**p-value**	**t**	**p-value**	**t**	**p-value**	**t**	**p-value**
6.443	$\mathbf{2.84 \times 10^{-9}}$	-8.88	$\mathbf{1.05 \times 10^{-14}}$	-9.0	$\mathbf{5.39 \times 10^{-15}}$	-0.705	0.482

Table 6.7: Paired t-test Results of GRU versus Transformer

Accuracy		**RMSE**		**Geodesic Distance**		**Execution Time**	
t	**p-value**	**t**	**p-value**	**t**	**p-value**	**t**	**p-value**
8.078	$\mathbf{7.26 \times 10^{-13}}$	-10.2	$\mathbf{8.69 \times 10^{-18}}$	-10.523	$\mathbf{1.52 \times 10^{-18}}$	-1.594	0.114

and Geodesic Distance of the Transformer-based model are higher than both LSTM and GRU, it can be concluded that Transformers perform poorly on these three metrics when compared to LSTM and GRU. As shown in Table 6.5, the p-values of mean accuracy and execution time are both greater than the significance value of 0.05. This implies that the null hypothesis cannot be rejected and thus there is no significant difference in the performance of the LSTM and GRU models in terms of mean accuracy and execution time. However, for mean RMSE and geodesic distance, the p-values are 0.004 and 0.002 respectively, which are both less than the significance value of 0.05. This implies that the performances of GRU and LSTM are significantly different in terms of mean RMSE and geodesic distance. Since the average RMSE and geodesic distance of GRU is less than that of LSTM (as shown in tables 6.2, 6.3), it is concluded that the GRU-based model performs significantly better than LSTM in terms of RMSE and geodesic distance. Lower RMSE and geodesic distance indicate a more accurate trajectory prediction of volunteers. For spectrum monitoring, it is more important to predict the future region of a volunteer in low execution time than to predict the exact future trajectory of a volunteer. Therefore, both LSTM and GRU can be interchangeably used to model volunteer mobility for spectrum monitoring.

6.6 Summary

This chapter addresses the effectiveness of spectrum monitoring by employing a volunteer mobility model which is built using deep learning algorithms. This model predicts the future locations of volunteers based on their historical location traces. The performance of the model is evaluated using four performance metrics, including Accuracy, Root Mean Square Error, Geodesic Distance, and Execution Time. Results from the Geolife dataset reveal that the GRU-based mobility model consistently outperforms the LSTM and Transformer-based models across all the metrics. Statistical validation confirms that Transformer performs the worst in terms of accuracy, RMSE and Geodesic distance when compared to LSTM and GRU. Between LSTM and GRU, there is a significant difference in performance in terms of RMSE and geodesic distance. This suggests that the GRU model is preferable for accurate trajectory prediction in the context of spectrum monitoring. However, for predicting the future region of volunteers, there is no significant difference in the performance of the LSTM and GRU. Hence, either LSTM or GRU can be used to predict the future region of volunteers for monitoring spectrum in a specific monitoring interval.

7.0 Conclusion and Future Directions

The chapter concludes the dissertation, highlighting the major contributions of the dissertation and discussing the future scope of this research.

7.1 Dissertation Contributions

With the exponential increase in the use of wireless services, the demand for additional spectrum is steadily on the rise. To address the spectrum scarcity problem, there is a need for spectrum sharing. To this end, the FCC proposed Dynamic Spectrum Access (DSA), wherein licensed frequency bands when idle, are utilized by unlicensed users. As spectrum sharing becomes more intense with more stakeholders, we can expect an increase in the number of potentially enforceable events [47]. Hence, the success of shared spectrum networks depends on the ability to effectively monitor the spectrum and enforce spectrum access policies. The primary focus of this dissertation is on efficient ex-post (punitive) spectrum enforcement. Traditional deployment of dedicated physical monitoring infrastructure [71] neither ensures high coverage of channels in the area of enforcement nor is cost-effective. Hence, a crowdsourced approach for spectrum enforcement is utilized [47, 48, 51, 53].

A shared spectrum enforcement infrastructure is developed for efficient coverage. The proposed approach divides the area of enforcement by utilizing a clustering algorithm and utilizes a divide-and-conquer approach for achieving efficient coverage of the area of enforcement. It is believed that spectrum misuse and access right violations can be effectively prevented using a trusted infrastructure, composed of a minimal number of dedicated devices with advanced trust and authentication capabilities, augmented with an opportunistic infrastructure of peer wireless devices with various software and hardware capabilities.

A spectrum sampling strategy by crowdsourced spectrum monitoring agents is developed such that it ensures effectiveness and accuracy in detecting spectrum access violations. This strategy, when coupled with a reputation management methodology enables us to develop a robust system to detect spectrum infractions against varying types of intruders, crowdsourced spectrum monitoring agents, and environmental factors.

The success of a crowdsourced infrastructure depends primarily on the performance and trustworthiness of the recruited crowdsourced agents. A crowdsourced monitoring agent should have a high likelihood of residing in the geographical region for which the agent is recruited. To this end, Machine learning-based approaches like LSTM, GRU, and Transformers are utilized to predict their likelihood of being in a geographical region of enforcement over epochs of time. Another criterion that needs to be considered is the trustworthiness of the crowdsourced agents. Failure to report spectrum misuse consistently should be penalized. However, such failures can also occur due to 'acts of God' like sensing device failure and signal reception. Therefore, effective assessment of agents' monitoring performance should be immune to network failure occurrences. To meet these requirements, I developed a methodology to determine the qualification of a crowdsourced monitoring agent, based on their monitoring device capabilities, their likelihood to reside in a region for an extended period of time, and their reputation in reporting accurate infraction events. The recruitment scheme also strives to take into consideration the personal preferences of the crowdsourced agents. In my work, the monitoring agent selection algorithm for spectrum access enforcement is formulated as two variants of the stable matching algorithm, VM and RVM. While the matching algorithms consider volunteer preferences, they do not always produce the best performance for the DSA infrastructure (in terms of coverage and accuracy). To this end, a variant of the Multiple-choice Secretary algorithm is combined with stable matching algorithms to develop two hybrid algorithms, HYBRID-VM and HYBRID-RVM. The analysis shows that HYBRID-VM resulted in the best performance, in terms of detection accuracy, region coverage, and volunteer happiness when compared to the other algorithms [51, 53].

7.2 Future Research

In the following, I discuss potential directions of future research in next-generation shared spectrum networks.

7.2.1 Enhanced Volunteer Infrastructure

In the proposed framework, volunteers may register with information that is not genuine. For example, if the location that a volunteer shares is not accurate, then it can hinder the recruitment of qualified volunteers, thereby hampering effective spectrum enforcement. Another example of a lapse in security can occur when a volunteer registers with a fake identity. Information related to the crowdsourced volunteers must be verified to avoid adverse issues arising from security lapses. This is an area that I plan to address in future research.

Volunteers who are recruited to monitor the spectrum may collude with the spectrum intruders and utilize the spectrum illegally themselves. Such collaboration can cause additional overheads that hinder the overall detection of spectrum misuse in a geographical region of enforcement. Collusions can additionally be of varying types – volunteer-volunteer collusion, volunteer-intruder collusion, and collusion among multiple volunteers and intruders. This is an area of research that I haven't explored yet and plan to address in the future.

7.2.2 Cognitive Architecture For Spectrum Enforcement

Wireless networks have ushered in an era of pervasive connectivity, transcending boundaries and enabling effortless communications, anytime and anywhere. Technology continues to advance rapidly, and next-generation wireless networks are already being envisioned. The continued advances in computing and wireless technologies, however, will bring together a range of emerging technologies, including augmented and virtual reality, digital twins and yet-to-be-invented smart infrastructure, paving the way for innovative services and applications across various industries and critical sectors of the economy. The ability of future wireless networks to meet the expected exponential increase in demand for high-bandwidth and sustained connectivity is paramount for the sustained deployment of immersive and persistent network environments to support future applications. It is expected that the number of smart interconnected devices will reach 40 billion by 2050, paving the way for large-scale deployment of immersive virtual environments that blur the lines between physical and digital worlds. New paradigms of spectrum sharing and coexistence must be developed for the wireless infrastructure to perform its functionalities, effectively and efficiently. Perceptive and reasoning capabilities will be required to enable cognitive radios to resolve complex challenges autonomously and self-optimize connectivity among a massive number of users. Embedding intelligence and cognitive capabilities in future software-defined radios raises several challenging concerns, which may lead to network breaches, disturb connectivity, and hinder the successful deployment and operation of future wireless networks. Unintended and malicious disturbances caused by smart rogue devices can degrade the quality of radio communication and, possibly, disrupt it entirely. In this emerging wireless ecosystem, new approaches to monitoring and access control are bound to play a key role in ensuring sustained high-quality service and uninterrupted operation. Cognitive access rights management and enforcement will continue to be an integral part of future spectrum-shared networks. Without proper enforcement, it would be impossible to protect access rights in large-scale, high-bandwidth wireless environments. Current access right enforcement frameworks, however, lack the capability to respond swiftly and effectively to the expected large number of access violations staged by a massive number of heterogeneous, cognitive rogue devices.

Future access right enforcement systems are, therefore, expected to embed in their design cognitive and perceptive capabilities, along with automation and self-management, to operate autonomously and monitor spectrum access effectively and reliability. An architectural transformation is, required to augment spectrum access enforcement with an intelligent monitoring layer with inbuilt cognitive and machine learning capabilities. The derived knowledge is used for recommendation and automation across different regions of the wireless network in a coordinated manner. Generative artificial intelligence provides a unique opportunity to bring deep intelligence into the spectrum access rights enforcement. Future research will focus on exploring fundamentals along with innovations to bring Generative AI advances toward the design and commercial deployment of cognitive spectrum access right enforcement, in large-scale, heterogeneous wireless networks. Current spectrum access systems are mainly focused on resolving interference, with minimal consideration of spectrum misuse and access violations. The proposed method will take an integrative approach to holistically address mobility management, sensing and localization, smart signaling, interference management, and spectrum access misuse, in a coherent manner. A hybrid monitoring architecture will be investigated, whereby knowledge is maintained both within the core and the extreme edges of the network to allow the central controller and the embedded mobile spectrum access monitors to collect information, share knowledge, and take coordinated, context-aware actions in response to spectrum disturbances. The ability to distribute spectrum monitoring between the central controller and smart trusted devices is crucial to achieve scale, and adequate responsiveness to spectrum misuse, while reducing infrastructure cost and energy consumption and optimizing monitoring performance. Machine learning algorithms will be explored to harness the combined power of generative AI and wireless spectrum enforcement expertise in acquiring knowledge and making effective and reliable decisions.

7.2.3 Cognitive Protocols For Spectrum Enforcement

AI is no longer just a "futuristic trop", it has become a key enabler of technological innovation and will continue to fuel new discoveries across the scientific, engineering, humanistic, and artistic fields. The confluence of computing, communications, and machine learning (ML) is having a transformative impact on the way future networks will be built, deployed, and managed. In the not-so-distant future, I envision our networks to be powered by cognitive robots and intelligent communication devices capable of autonomously planning and solving complex networking problems for optimized operability and enhanced user experience. A major focus of my future research is to explore new and innovative ways to harness the potential of Generative AI to build networking protocols and technology that empower users in their everyday interactions with other people and with their sounding environments. My current work demonstrated the feasibility of building a crowdsourced, spectrum monitoring framework, empowered by ML to regiment shared spectrum access and prevent intrusion and violation of access rights. The critical insights gained through this work will provide me with the basis for the development of new networking paradigms, that seamlessly integrate predictive and generative ML approaches, to enable secure, scalable real-time network management and monitoring. One focus of my future research will be on new cognitive ways to efficiently monitor network behavior, accelerate network failure resolution, and proactively engineer guidance and remedial actions to prevent future undesirable behaviors before they occur. Another focus of my future research will be on automating routine network tasks and using generative AI to create real-time, context-aware responses to optimize spectrum usage and enable personalized user interactions with people and the surrounding environment. These approaches will have a significant impact on the way network engineers develop, deploy, and operate networks and the variety of ways users will interact with each other and the physical world, leading to new kinds of knowledge artifacts and productive and meaningful experiences.

Bibliography

[1] S. Shavell, “The optimal structure of law enforcement,” *The Journal of Law and Economics*, vol. 36, no. 1, Part 2, pp. 255–287, 1993.

[2] A. Malki and M. B. Weiss, “Ex-post enforcement in spectrum sharing,” in *2014 TPRC Conference Paper*, 2014.

[3] P. Steenkiste, D. Sicker, G. Minden, and D. Raychaudhuri, “Future directions in cognitive radio network research,” in *NSF workshop report*, vol. 4, pp. 1–2, 2009.

[4] FCC, “Measuring Broadband America — fcc.gov.” `https://www.fcc.gov/general/measuring-broadband-america`. [Accessed 02-08-2024].

[5] R. Zhang, F. Gao, and Y.-C. Liang, “Cognitive beamforming made practical: Effective interference channel and learning-throughput tradeoff,” *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 706–718, 2010.

[6] X. Li, N. Zhao, Y. Sun, and F. R. Yu, “Interference alignment based on antenna selection with imperfect channel state information in cognitive radio networks,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5497–5511, 2015.

[7] Q. Zhao and B. M. Sadler, “A survey of dynamic spectrum access,” *IEEE signal processing magazine*, vol. 24, no. 3, pp. 79–89, 2007.

[8] D. N. Hatfield and P. J. Weiser, “Property rights in spectrum: Taking the next step,” in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, pp. 43–55, IEEE, 2005.

[9] L. Xu, R. Tonjes, T. Paila, W. Hansmann, M. Frank, and M. Albrecht, “Drive-ing to the internet: Dynamic radio for ip services in vehicular environments,” in *Proceedings 25th Annual IEEE Conference on Local Computer Networks. LCN 2000*, pp. 281–289, IEEE, 2000.

[10] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE journal on selected areas in communications*, vol. 23, no. 2, pp. 201–220, 2005.

[11] B. Wang and K. R. Liu, "Advances in cognitive radio networks: A survey," *IEEE Journal of selected topics in signal processing*, vol. 5, no. 1, pp. 5–23, 2010.

[12] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, 2009.

[13] M. Song, C. Xin, Y. Zhao, and X. Cheng, "Dynamic spectrum access: from cognitive radio to network radio," *IEEE Wireless Communications*, vol. 19, no. 1, pp. 23–29, 2012.

[14] X. Yuan, C. Jiang, Y. Shi, Y. T. Hou, W. Lou, and S. Kompella, "Beyond interference avoidance: On transparent coexistence for multi-hop secondary cr networks," in *2013 IEEE International Conference on Sensing, Communications and Networking (SECON)*, pp. 398–405, IEEE, 2013.

[15] X. Yuan, Y. Shi, Y. T. Hou, W. Lou, and S. Kompella, "Ups: A united cooperative paradigm for primary and secondary networks," in *2013 IEEE 10th International Conference on Mobile Ad-Hoc and Sensor Systems*, pp. 78–85, IEEE, 2013.

[16] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh, "White space networking with wi-fi like connectivity," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 4, pp. 27–38, 2009.

[17] D. Hlavacek and J. M. Chang, "A layered approach to cognitive radio network security: A survey," *Computer Networks*, vol. 75, pp. 414–436, 2014.

[18] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *IEEE Journal on selected areas in communications*, vol. 25, no. 3, pp. 517–528, 2007.

[19] J. Huang, R. A. Berry, and M. L. Honig, "Spectrum sharing with distributed interference compensation," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, pp. 88–93, IEEE, 2005.

[20] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *1999 IEEE International Workshop on Mobile Multimedia Communications (MoMuC'99)(Cat. No. 99EX384)*, pp. 3–10, IEEE, 1999.

[21] Q. Zhao, Y. Chen, and A. Swami, “Cognitive mac protocols for dynamic spectrum access,” in *Cognitive Wireless Communication Networks*, pp. 271–301, Springer, 2007.

[22] P. S. Rosenbloom, “Towards a new cognitive hourglass: Uniform implementation of cognitive architecture via factor graphs,” in *Proceedings of the 9th international conference on cognitive modeling*, pp. 116–121, Citeseer, 2009.

[23] P. Rosenbloom, “Speculations on leveraging graphical models for architectural integration of visual representation and reasoning,” in *Workshops at the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[24] X. Foukas, M. K. Marina, and K. Kontovasilis, “Iris: Deep reinforcement learning driven shared spectrum access architecture for indoor neutral-host small cells,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 8, pp. 1820–1837, 2019.

[25] S. Hu, F. Li, H. Guo, P. Wang, G. Bi, Y. Gao, Z. Liu, and B. Yu, “Tdcs-idma system for cognitive radio networks with cloud,” *IEEE Access*, vol. 6, pp. 20520–20530, 2018.

[26] J. Mitola and G. Q. Maguire, “Cognitive radio: making software radios more personal,” *IEEE personal communications*, vol. 6, no. 4, pp. 13–18, 1999.

[27] D. B. Rawat, S. Shetty, and K. Raza, “Game theoretic dynamic spectrum access in cloud-based cognitive radio networks,” in *2014 IEEE International Conference on Cloud Engineering*, pp. 586–591, IEEE, 2014.

[28] I. Foster, Y. Zhao, I. Raicu, and S. Lu, “Cloud computing and grid computing 360-degree compared,” in *2008 grid computing environments workshop*, pp. 1–10, Ieee, 2008.

[29] S.-H. Wu, H.-L. Chao, C.-H. Ko, S.-R. Mo, C.-T. Jiang, T.-L. Li, C.-C. Cheng, and C.-F. Liang, “A cloud model and concept prototype for cognitive radio networks,” *IEEE Wireless Communications*, vol. 19, no. 4, pp. 49–58, 2012.

[30] W.-Y. Lee and I. F. Akyildiz, “Optimal spectrum sensing framework for cognitive radio networks,” *IEEE Transactions on wireless communications*, vol. 7, no. 10, pp. 3845–3857, 2008.

[31] P. S. Upadhyaya, V. K. Shah, and J. H. Reed, “Cross-layer band selection and routing design for diverse band-aware dsa networks,” in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2020.

[32] M. Sherman, A. N. Mody, R. Martinez, C. Rodriguez, and R. Reddy, “Ieee standards supporting cognitive radio and networks, dynamic spectrum access, and coexistence,” *IEEE Communications Magazine*, vol. 46, no. 7, pp. 72–79, 2008.

[33] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, “Ieee 802.22: The first cognitive radio wireless regional area network standard,” *IEEE communications magazine*, vol. 47, no. 1, pp. 130–138, 2009.

[34] A. B. Flores, R. E. Guerra, E. W. Knightly, P. Ecclesine, and S. Pandey, “Ieee 802.11 af: A standard for tv white space spectrum sharing,” *IEEE Communications Magazine*, vol. 51, no. 10, pp. 92–100, 2013.

[35] “Spectrum Sharing — nist.gov.” `https://www.nist.gov/advanced-communications/spectrum-sharing`. [Accessed 03-08-2024].

[36] J. M. Peha, “Sharing spectrum through spectrum policy reform and cognitive radio,” *Proceedings of the IEEE*, vol. 97, no. 4, pp. 708–719, 2009.

[37] NIST, “NextG: 5G-and-Beyond Technology — nist.gov.” `https://www.nist.gov/advanced-communications/nextg-5g-and-beyond-technology`. [Accessed 03-08-2024].

[38] “3.5 GHz Band Overview — fcc.gov.” `https://www.fcc.gov/wireless/bureau-divisions/mobility-division/35-ghz-band/35-ghz-band-overview`. [Accessed 03-08-2024].

[39] “Enabling abundant wireless connectivity for everyone — google.com.” `https://www.google.com/get/spectrumdatabase/#!#cbrs`. [Accessed 03-08-2024].

[40] M. R. Souryal, T. T. Nguyen, and N. J. LaSorte, “3.5 ghz federal incumbent protection algorithms,” in *2018 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, pp. 1–5, IEEE, 2018.

[41] 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," *Proceedings of the IEEE*, vol. 100, no. 12, pp. 3172–3186, 2012.

[42] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxylakis, "A survey on security threats and detection techniques in cognitive radio networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 428–445, 2012.

[43] K. A. Woyach, A. Sahai, G. Atia, and V. Saligrama, "Crime and punishment for cognitive radios," in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 236–243, IEEE, 2008.

[44] G. Atia, A. Sahai, and V. Saligrama, "Spectrum enforcement and liability assignment in cognitive radio systems," in *2008 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 1–12, IEEE, 2008.

[45] A. Nika, Z. Li, Y. Zhu, Y. Zhu, B. Y. Zhao, X. Zhou, and H. Zheng, "Empirical validation of commodity spectrum monitoring," in *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pp. 96–108, 2016.

[46] E. Schlager and E. Ostrom, "Property-rights regimes and natural resources: a conceptual analysis," *Land economics*, pp. 249–262, 1992.

[47] D. Das, P. Bustamante, T. Znati, M. M. Gomez, M. Weiss, and J. S. Rose, "Misuse detection in dynamic spectrum sharing wireless networks across multiple channels," *International Journal on Advances in Networks and Services Volume 12, Number 3 & 4, 2019*, 2019.

[48] D. Das, T. Znati, M. B. Weiss, P. Bustamante, M. M. Gomez, and J. S. Rose, "Crowdsourced misuse detection in dynamic spectrum sharing wireless networks," in *International Conference on Networks (ICN)*, 2019.

[49] A. Gopinathan, Z. Li, and C. Wu, "Strategyproof auctions for balancing social welfare and fairness in secondary spectrum markets," in *2011 Proceedings IEEE INFOCOM*, pp. 3020–3028, IEEE, 2011.

[50] J. Howe, "The Rise of Crowdsourcing — wired.com." `https://www.wired.com/2006/06/crowds/`, 2006. [Accessed 04-08-2024].

[51] D. Das, J. S. Rose, T. Znati, P. Bustamante, M. Weiss, and M. M. Gomez, "Spectrum misuse detection in cooperative wireless networks," in *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*, pp. 1–6, IEEE, 2020.

[52] Q. Du, M. Emelianenko, and L. Ju, "Convergence of the lloyd algorithm for computing centroidal voronoi tessellations," *SIAM journal on numerical analysis*, vol. 44, no. 1, pp. 102–119, 2006.

[53] D. Das, T. Znati, M. B. Weiss, M. M. Gomez, P. Bustamante, and J. S. Rose, "Matchmaking of volunteers and channels for dynamic spectrum access enforcement," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2020.

[54] D. Das, T. Znati, and M. B. Weiss, "Efficient monitoring of dynamic spectrum access for robust and reliable detection of unauthorized access," in *MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM)*, pp. 729–734, IEEE, 2022.

[55] N. N. Krishnan, R. Kumbhkar, N. B. Mandayam, I. Seskar, and S. Kompella, "Coexistence of radar and communication systems in cbrs bands through downlink power control," in *MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM)*, pp. 713–718, IEEE, 2017.

[56] H.-Y. Kuo, S.-Y. Liu, C.-Y. Huang, Y.-C. Chen, and M.-H. Xie, "Reliable data transmission through private cbrs networks," *arXiv preprint arXiv:2310.14171*, 2023.

[57] X. Ying, M. M. Buddhikot, and S. Roy, "Sas-assisted coexistence-aware dynamic channel assignment in cbrs band," *IEEE Transactions on Wireless Communications*, vol. 17, no. 9, pp. 6307–6320, 2018.

[58] Y. Xiao, S. Shi, W. Lou, C. Wang, X. Li, N. Zhang, Y. T. Hou, and J. H. Reed, "Decentralized spectrum access system: Vision, challenges, and a blockchain solution," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 220–228, 2022.

[59] M. Grissa, A. A. Yavuz, and B. Hamdaoui, "Trustsas: A trustworthy spectrum access system for the 3.5 ghz cbrs band," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pp. 1495–1503, IEEE, 2019.

[60] M. I. Rochman, V. Sathya, B. Payne, M. Yavuz, and M. Ghosh, "A measurement study of the impact of adjacent channel interference between c-band and cbrs," in

2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1–7, IEEE, 2023.

[61] D. N. Hatfield, L. Claudy, M. Gorenberg, D. Gurney, G. Lapin, B. Markwalter, G. Mendenhall, P. de Vries, and D. Roberson, "Introduction to interference resolution, enforcement and radio noise," *FCC Technological Advisory Council, Washington, DC, Tech. Rep. FCC*, pp. 14–31, 2014.

[62] M. W. Toffel and J. L. Short, "Coming clean and cleaning up: Does voluntary self-reporting indicate effective self-policing?," *The Journal of Law and Economics*, vol. 54, no. 3, pp. 609–649, 2011.

[63] R. H. Coase, "The federal communications commission," *The Journal of Law and Economics*, vol. 2, pp. 1–40, 1959.

[64] B. Bahrak, A. Deshpande, *et al.*, "Spectrum access policy reasoning for policy-based cognitive radios," *Computer Networks*, vol. 56, no. 11, pp. 2649–2663, 2012.

[65] E. Ostrom, "Beyond markets and states: polycentric governance of complex economic systems," *American economic review*, vol. 100, no. 3, pp. 641–672, 2010.

[66] E. Schlager and E. Ostrom, "Property-rights regimes and natural resources: a conceptual analysis," *Land economics*, pp. 249–262, 1992.

[67] B. Danev and S. Capkun, "Transient-based identification of wireless sensor nodes," in *2009 International Conference on Information Processing in Sensor Networks*, pp. 25–36, IEEE, 2009.

[68] H. Demsetz, "The exchange and enforcement of property rights," *The Journal of Law and Economics*, vol. 7, pp. 11–26, 1964.

[69] E. H. Dinan and B. Jabbari, "Spreading codes for direct sequence cdma and wideband cdma cellular networks," *IEEE communications magazine*, vol. 36, no. 9, pp. 48–54, 1998.

[70] F. Perich and M. McHenry, "Policy-based spectrum access control for dynamic spectrum access network radios," *Journal of Web Semantics*, vol. 7, no. 1, pp. 21–27, 2009.

[71] M. B. Weiss, M. Altamimi, and M. McHenry, “Enforcement and spectrum sharing: A case study of the 1695–1710 mhz band,” in *8th International Conference on Cognitive Radio Oriented Wireless Networks*, pp. 7–12, IEEE, 2013.

[72] P. Bustamante, D. Das, J. S. Rose, M. Gomez, M. B. Weiss, J.-M. Park, and T. Znati, “Toward automated enforcement of radio interference,” in *TPRC48: The 48th Research Conference on Communication, Information and Internet Policy*, 2020.

[73] R. Murty, R. Chandra, T. Moscibroda, and P. Bahl, “Senseless: A database-driven white spaces network,” *IEEE Transactions on Mobile Computing*, vol. 11, no. 2, pp. 189–203, 2011.

[74] R. A. O’Connor, “Understanding television’s grade a and grade b service contours,” *IEEE transactions on broadcasting*, vol. 47, no. 3, pp. 309–314, 2001.

[75] M. Vu, N. Devroye, and V. Tarokh, “On the primary exclusive region of cognitive networks,” *IEEE transactions on wireless communications*, vol. 8, no. 7, pp. 3380–3385, 2009.

[76] S. Kusaladharma and C. Tellambura, “Aggregate interference analysis for underlay cognitive radio networks,” *IEEE Wireless communications letters*, vol. 1, no. 6, pp. 641–644, 2012.

[77] S. Bhattarai, J.-M. Park, and W. Lehr, “Dynamic exclusion zones for protecting primary users in database-driven spectrum sharing,” *IEEE/ACM Transactions on Networking*, vol. 28, no. 4, pp. 1506–1519, 2020.

[78] J. Tang and Y. Cheng, “Selfish misbehavior detection in 802.11 based wireless networks: An adaptive approach based on markov decision process,” in *2013 Proceedings IEEE INFOCOM*, pp. 1357–1365, IEEE, 2013.

[79] S. Liu, L. J. Greenstein, W. Trappe, and Y. Chen, “Detecting anomalous spectrum usage in dynamic spectrum access networks,” *Ad Hoc Networks*, vol. 10, no. 5, pp. 831–844, 2012.

[80] M. Khaledi, M. Khaledi, S. Sarkar, S. Kasera, N. Patwari, K. Derr, and S. Ramirez, “Simultaneous power-based localization of transmitters for crowdsourced spectrum monitoring,” in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pp. 235–247, 2017.

[81] S. Sarkar, D. Guo, and D. Cabric, “Radyololet: Radar detection and parameter estimation using yolo and wavelet,” *arXiv preprint arXiv:2309.12094*, 2023.

[82] D. Villa, D. Uvaydov, L. Bonati, P. Johari, J. M. Jornet, and T. Melodia, “Intelligent radar detection in cbrs band in the colosseum wireless network emulator,” *arXiv preprint arXiv:2309.08861*, 2023.

[83] Y. Shi and Y. E. Sagduyu, “Sensing-throughput tradeoffs with generative adversarial networks for nextg spectrum sharing,” in *MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM)*, pp. 13–18, IEEE, 2022.

[84] J.-M. Park, J. H. Reed, A. Beex, T. C. Clancy, V. Kumar, and B. Bahrak, “Security and enforcement in spectrum sharing,” *Proceedings of the IEEE*, vol. 102, no. 3, pp. 270–281, 2014.

[85] M. Barbeau, J. Hall, and E. Kranakis, “Detection of rogue devices in bluetooth networks using radio frequency fingerprinting,” in *proceedings of the 3rd IASTED International Conference on Communications and Computer Networks, CCN*, pp. 4–6, 2006.

[86] K. Kim, C. M. Spooner, I. Akbar, and J. H. Reed, “Specific emitter identification for cognitive radio with application to ieee 802.11,” in *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*, pp. 1–5, IEEE, 2008.

[87] K. Remley, C. A. Grosvenor, R. T. Johnk, D. R. Novotny, P. D. Hale, M. McKinley, A. Karygiannis, and E. Antonakakis, “Electromagnetic signatures of wlan cards and network security,” in *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology, 2005.*, pp. 484–488, IEEE, 2005.

[88] B. Danev, H. Luecken, S. Capkun, and K. El Defrawy, “Attacks on physical-layer identification,” in *Proceedings of the third ACM conference on Wireless network security*, pp. 89–98, 2010.

[89] X. Jin, J. Sun, R. Zhang, Y. Zhang, and C. Zhang, “Specguard: Spectrum misuse detection in dynamic spectrum access systems,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 12, pp. 2925–2938, 2018.

[90] N. Goergen, T. C. Clancy, and T. R. Newman, “Physical layer authentication watermarks through synthetic channel emulation,” in *2010 IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*, pp. 1–7, IEEE, 2010.

[91] X. Wang, Y. Wu, and B. Caron, “Transmitter identification using embedded pseudo random sequences,” *IEEE Transactions on broadcasting*, vol. 50, no. 3, pp. 244–252, 2004.

[92] X. Tan, K. Borle, W. Du, and B. Chen, “Cryptographic link signatures for spectrum usage authentication in cognitive radio,” in *Proceedings of the fourth ACM conference on Wireless network security*, pp. 79–90, 2011.

[93] V. Kumar, H. Li, J.-M. J. Park, and K. Bian, “Crowd-sourced authentication for enforcement in dynamic spectrum sharing,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 625–636, 2019.

[94] Y. Hou and M. Li, “Enforcing spectrum access rules in cognitive radio networks through cooperative jamming,” in *International Conference on Wireless Algorithms, Systems, and Applications*, pp. 440–453, Springer, 2013.

[95] Y. E. Sagduyu, “Adversarial machine learning and defense game for nextg signal classification with deep learning,” in *MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM)*, pp. 1076–1081, IEEE, 2022.

[96] Y. E. Sagduyu, T. Erpek, and Y. Shi, “Adversarial machine learning for 5g communications security,” *Game Theory and Machine Learning for Cyber Security*, pp. 270–288, 2021.

[97] J. Siegel, “Innovative methods in stakeholder engagement: An environmental scan,” 2012.

[98] J. M. Leimeister, M. Huber, U. Bretschneider, and H. Krcmar, “Leveraging crowdsourcing: activation-supporting components for it-based ideas competition,” *Journal of management information systems*, vol. 26, no. 1, pp. 197–224, 2009.

[99] S. O. Alexander Hars, “Working for free? motivations for participating in open-source projects,” *International journal of electronic commerce*, vol. 6, no. 3, pp. 25–39, 2002.

[100] G. Hertel, S. Niedner, and S. Herrmann, “Motivation of software developers in open source projects: an internet-based survey of contributors to the linux kernel,” *Research policy*, vol. 32, no. 7, pp. 1159–1177, 2003.

[101] K. R. Lakhani and R. G. Wolf, "Why hackers do what they do: Understanding motivation and effort in free/open source software projects," 2005.

[102] J. Lerner and J. Tirole, "Some simple economics of open source," *The journal of industrial economics*, vol. 50, no. 2, pp. 197–234, 2002.

[103] M. I. Rochman, V. Sathya, N. Nunez, D. Fernandez, M. Ghosh, A. S. Ibrahim, and W. Payne, "A comparison study of cellular deployments in chicago and miami using apps on smartphones," in *Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & CHaracterization*, pp. 61–68, 2022.

[104] A. Afuah and C. L. Tucci, "Reflections on the 2022 amr decade award: crowdsourcing as a solution to distant search," *Academy of Management Review*, vol. 48, no. 4, pp. 597–610, 2023.

[105] J. Pohlisch, "Internal crowdsourcing at sap," in *European Conference on Innovation and Entrepreneurship*, pp. 1201–XXIV, Academic Conferences International Limited, 2019.

[106] J. Pohlisch, "An introduction to internal crowdsourcing," *Internal Crowdsourcing in Companies*, p. 15, 2021.

[107] X. Xerandy, F. Ai, T. Znati, L. K. Comfort, and F. A. Ismail, "Device-to-device communication: A scalable, socially aware, land-based infrastructure to support community resilience in disaster events," *Hazardous Seas: A Sociotechnical Framework for Early Tsunami Detection and Warning*, p. 91, 2023.

[108] T. Nelson, A. Roy, C. Ferster, J. Fischer, V. Brum-Bastos, K. Laberee, H. Yu, and M. Winters, "Generalized model for mapping bicycle ridership with crowdsourced data," *Transportation Research Part C: Emerging Technologies*, vol. 125, p. 102981, 2021.

[109] E. Glińska, H. Kiryluk, and K. Ilczuk, "Crowdsourcing initiatives in city management: the perspective of polish local governments," *Ekonomia i Środowisko*, no. 3, pp. 287–311, 2022.

[110] C. Muller, L. Chapman, S. Johnston, C. Kidd, S. Illingworth, G. Foody, A. Overeem, and R. Leigh, "Crowdsourcing for climate and atmospheric sciences: current status and future potential," *International Journal of Climatology*, vol. 35, no. 11, pp. 3185–3203, 2015.

[111] D. Palacios-Marqués, J. F. Gallego-Nicholls, and M. Guijarro-García, “A recipe for success: Crowdsourcing, online social networks, and their impact on organizational performance,” *Technological Forecasting and Social Change*, vol. 165, p. 120566, 2021.

[112] A. Dutta and M. Chiang, ““See Something, Say Something” Crowdsourced Enforcement of Spectrum Policies,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 67–80, 2016.

[113] X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, and X. Li, “Free market of crowdsourcing: Incentive mechanism design for mobile sensing,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 12, pp. 3190–3200, 2014.

[114] D. Yang, G. Xue, X. Fang, and J. Tang, “Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing,” in *Proceedings of the 18th annual international conference on Mobile computing and networking*, pp. 173–184, 2012.

[115] X. Zhu, J. An, M. Yang, L. Xiang, Q. Yang, and X. Gui, “A fair incentive mechanism for crowdsourcing in crowd sensing,” *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1364–1372, 2016.

[116] J. Lin, M. Li, D. Yang, G. Xue, and J. Tang, “Sybil-proof incentive mechanisms for crowdsensing,” in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pp. 1–9, IEEE, 2017.

[117] M. Li, D. Yang, J. Lin, and J. Tang, “Specwatch: A framework for adversarial spectrum monitoring with unknown statistics,” *Computer Networks*, vol. 143, pp. 176–190, 2018.

[118] A. M. Salama, M. Li, and D. Yang, “Optimal crowdsourced channel monitoring in cognitive radio networks,” in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2017.

[119] X. Jin, R. Zhang, Y. Chen, T. Li, and Y. Zhang, “Dpsense: Differentially private crowdsourced spectrum sensing,” in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 296–307, 2016.

[120] X. Jin and Y. Zhang, “Privacy-preserving crowdsourced spectrum sensing,” *IEEE/ACM Transactions on Networking*, vol. 26, no. 3, pp. 1236–1249, 2018.

[121] F. Aurenhammer, “Voronoi diagrams—a survey of a fundamental geometric data structure,” *ACM Computing Surveys (CSUR)*, vol. 23, no. 3, pp. 345–405, 1991.

[122] F. Aurenhammer and R. Klein, “Voronoi diagrams.,” *Handbook of computational geometry*, vol. 5, no. 10, pp. 201–290, 2000.

[123] G. Lejeune Dirichlet, “Über die reduction der positiven quadratischen formen mit drei unbestimmten ganzen zahlen.,” *Journal für die reine und angewandte Mathematik (Crelles Journal)*, vol. 1850, no. 40, pp. 209–227, 1850.

[124] G. F. Voronoı, “Deuxieme mémoire: recherches sur les paralléloedres primitifs,” *J. reine angew. Math*, vol. 136, pp. 67–181, 1909.

[125] G. Voronoi, “Nouvelles applications des paramètres continus à la théorie des formes quadratiques. deuxième mémoire. recherches sur les parallélloèdres primitifs.,” *Journal für die reine und angewandte Mathematik*, vol. 134, pp. 198–287, 1908.

[126] V. Sacristán, “Algorithms for constructing voronoi diagrams.” `https://www.ic.unicamp.br/~rezende/ensino/mo619/Sacristan,%20Voronoi%20Diagrams.pdf`. [PowerPoint slides].

[127] E. Smith, C. Trefftz, and B. DeVries, “A divide-and-conquer algorithm for computing voronoi diagrams,” in *2020 IEEE International Conference on Electro Information Technology (EIT)*, pp. 495–499, IEEE, 2020.

[128] S. Lloyd, “Least squares quantization in pcm,” *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[129] A. Anderson, X. Wang, K. R. Baker, and D. Grunwald, “Systems for spectrum forensics,” in *Proceedings of the 2nd International Workshop on Hot Topics in Wireless*, pp. 26–30, 2015.

[130] A. Jain, L. Hong, and S. Pankanti, “Biometric identification,” *Communications of the ACM*, vol. 43, no. 2, pp. 90–98, 2000.

[131] J. L. Bailie and M. A. Jortberg, “Online learner authentication: Verifying the identity of online users,” *Journal of Online Learning and Teaching*, vol. 5, no. 2, pp. 197–207, 2009.

[132] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 4–20, 2004.

[133] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki, "An introduction evaluating biometric systems," *Computer*, vol. 33, no. 2, pp. 56–63, 2000.

[134] W. Meng, D. S. Wong, S. Furnell, and J. Zhou, "Surveying the development of biometric user authentication on mobile phones," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1268–1293, 2014.

[135] K. Dharavath, F. A. Talukdar, and R. H. Laskar, "Study on biometric authentication systems, challenges and future trends: A review," in *2013 IEEE international conference on computational intelligence and computing research*, pp. 1–7, IEEE, 2013.

[136] I. M. Alsaadi, "Physiological biometric authentication systems, advantages, disadvantages and future development: A review," *International Journal of Scientific & Technology Research*, vol. 4, no. 12, pp. 285–289, 2015.

[137] A. Eng and L. A. Wahsheh, "Look into my eyes: A survey of biometric security," in *2013 10th International Conference on Information Technology: New Generations*, pp. 422–427, IEEE, 2013.

[138] A. N. Kataria, D. M. Adhyaru, A. K. Sharma, and T. H. Zaveri, "A survey of automated biometric authentication techniques," in *2013 Nirma university international conference on engineering (NUiCONE)*, pp. 1–6, IEEE, 2013.

[139] M. Faundez-Zanuy, "Signature recognition state-of-the-art," *IEEE aerospace and electronic systems magazine*, vol. 20, no. 7, pp. 28–32, 2005.

[140] J. A. Wepman, B. L. Bedford, H. E. Ottke, and M. G. Cotton, "Rf sensors for spectrum monitoring applications: Fundamentals and rf performance test plan," *Technical Report NTIA TR-15-519*, August 2015.

[141] A. Nika, Z. Zhang, X. Zhou, B. Y. Zhao, and H. Zheng, "Towards commoditized real-time spectrum monitoring," in *Proceedings of the 1st ACM Workshop on Hot Topics in Wireless*, pp. 25–30, 2014.

[142] Compaq, Intel, Microsoft, Phoenix, and Toshiba, “Advanced configuration and power interface specification,” 2002.

[143] Intel and Microsoft, “Advanced power management(apm) bios interface specification,” 1996.

[144] M. Doyle, T. F. Fuller, and J. Newman, “Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell,” *Journal of the Electrochemical society*, vol. 140, no. 6, p. 1526, 1993.

[145] V. Tiwari, S. Malik, A. Wolfe, and M. T.-C. Lee, “Instruction level power analysis and optimization of software,” *Technologies for wireless computing*, pp. 139–154, 1996.

[146] H. Saputra, M. Kandemir, N. Vijaykrishnan, M. J. Irwin, J. S. Hu, C.-H. Hsu, and U. Kremer, “Energy-conscious compilation based on voltage scaling,” in *Proceedings of the joint conference on languages, compilers and tools for embedded systems: software and compilers for embedded systems*, pp. 2–11, 2002.

[147] V. Tiwari, S. Malik, and A. Wolfe, “Power analysis of embedded software: A first step towards software power minimization,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 2, no. 4, pp. 437–445, 1994.

[148] J.-M. Kang, S.-s. Seo, and J. W.-K. Hong, “Personalized battery lifetime prediction for mobile devices based on usage patterns,” *Journal of Computing Science and Engineering*, vol. 5, no. 4, pp. 338–345, 2011.

[149] C. J. Hegarty, “The global positioning system (gps),” *Springer handbook of global navigation satellite systems*, pp. 197–218, 2017.

[150] S. Dan, A. Santra, S. Mahato, and A. Bose, “Navic performance over the service region: Availability and solution quality,” *Sādhanā*, vol. 45, pp. 1–7, 2020.

[151] Y. Yang, W. Gao, S. Guo, Y. Mao, and Y. Yang, “Introduction to beidou-3 navigation satellite system,” *Navigation*, vol. 66, no. 1, pp. 7–18, 2019.

[152] X. Li, M. Ge, X. Dai, X. Ren, M. Fritsche, J. Wickert, and H. Schuh, “Accuracy and reliability of multi-gnss real-time precise positioning: Gps, glonass, beidou, and galileo,” *Journal of geodesy*, vol. 89, no. 6, pp. 607–635, 2015.

[153] M. Warren, M. Greeff, B. Patel, J. Collier, A. P. Schoellig, and T. D. Barfoot, "There's no place like home: visual teach and repeat for emergency return of multirotor uavs during gps failure," *IEEE Robotics and automation letters*, vol. 4, no. 1, pp. 161–168, 2018.

[154] M. Neinavaie, J. Khalife, and Z. M. Kassas, "Exploiting starlink signals for navigation: first results," in *Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021)*, pp. 2766–2773, 2021.

[155] J. Y. Tsai, P. G. Kelley, L. F. Cranor, and N. Sadeh, "Location-sharing technologies: Privacy risks and controls," *Isjlp*, vol. 6, p. 119, 2010.

[156] Y. Michalevsky, A. Schulman, G. A. Veerapandian, D. Boneh, and G. Nakibly, "{PowerSpy}: Location tracking using mobile device power analysis," in *24th USENIX Security Symposium (USENIX Security 15)*, pp. 785–800, 2015.

[157] T. Ji and A. Pachi, "Frequency and velocity of people walking," *Struct. Eng*, vol. 84, no. 3, pp. 36–40, 2005.

[158] K. J. Arrow, *The limits of organization.* WW Norton & Company, 1974.

[159] P. R. Cohen and H. J. Levesque, "Intention is choice with commitment," *Artificial intelligence*, vol. 42, no. 2-3, pp. 213–261, 1990.

[160] J. S. Coleman, *Foundations of social theory.* Harvard university press, 1994.

[161] R. M. Kramer, "Trust and distrust in organizations: Emerging perspectives, enduring questions," *Annual review of psychology*, vol. 50, no. 1, pp. 569–598, 1999.

[162] H. C. Triandis, "Organizations," 1960.

[163] A. Herzig, E. Lorini, J. F. Hübner, and L. Vercouter, "A logic of trust and reputation," *Logic Journal of the IGPL*, vol. 18, no. 1, pp. 214–244, 2010.

[164] S. Ali, N. Islam, A. Rauf, I. U. Din, M. Guizani, and J. J. Rodrigues, "Privacy and security issues in online social networks," *Future Internet*, vol. 10, no. 12, p. 114, 2018.

[165] M. Taddicken, “The ‘privacy paradox’in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure,” *Journal of computer-mediated communication*, vol. 19, no. 2, pp. 248–273, 2014.

[166] M. Fire, G. Katz, and Y. Elovici, “Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies,” *Human journal*, vol. 1, no. 1, pp. 26–39, 2012.

[167] A. Chaabane, Y. Ding, R. Dey, M. A. Kaafar, and K. W. Ross, “A closer look at third-party osn applications: are they leaking your personal information?,” in *Passive and Active Measurement: 15th International Conference, PAM 2014, Los Angeles, CA, USA, March 10-11, 2014, Proceedings 15*, pp. 235–246, Springer, 2014.

[168] M. Li, S. Yu, N. Cao, and W. Lou, “Privacy-preserving distributed profile matching in proximity-based mobile social networks,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2024–2033, 2013.

[169] C. Lv, X. Xiao, L. Zhang, and T. Yu, “Publishing common neighbors histograms of social networks under edge differential privacy,” in *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security*, pp. 1099–1113, 2024.

[170] J. Shen, J. Tian, and Z. Wang, “Privacy-preserving algorithm based on vulnerable nodes for social relationships,” *The Journal of Supercomputing*, pp. 1–28, 2024.

[171] Y. Mahajan, Z. Guo, J.-H. Cho, and I.-R. Chen, “Privacy-preserving and diversity-aware trust-based team formation in online social networks,” 2023.

[172] B. Talukder, K. W. Hipel, and G. W. vanLoon, “Developing composite indicators for agricultural sustainability assessment: Effect of normalization and aggregation techniques,” *Resources*, vol. 6, no. 4, p. 66, 2017.

[173] H. T. Friis, “A note on a simple transmission formula,” *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, 1946.

[174] R. Kleinberg, “A multiple-choice secretary algorithm with applications to online auctions,” in *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 630–631, 2005.

[175] M. Babaioff, N. Immorlica, D. Kempe, and R. Kleinberg, "Online auctions and generalized secretary problems," *ACM SIGecom Exchanges*, vol. 7, no. 2, pp. 1–11, 2008.

[176] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.

[177] D. Yang, X. Zhang, and G. Xue, "Promise: A framework for truthful and profit maximizing spectrum double auctions," in *IEEE INFOCOM 2014-IEEE conference on computer communications*, pp. 109–117, IEEE, 2014.

[178] Y. Chen, Y. Xiong, Q. Wang, X. Yin, and B. Li, "Stable matching for spectrum market with guaranteed minimum requirement," in *Proceedings of the 18th ACM international symposium on mobile Ad hoc networking and computing*, pp. 1–10, 2017.

[179] Y. Chen, L. Jiang, H. Cai, J. Zhang, and B. Li, "Spectrum matching," in *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, pp. 590–599, IEEE, 2016.

[180] H. Xu and B. Li, "Anchor: A versatile and efficient framework for resource management in the cloud," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 6, pp. 1066–1076, 2012.

[181] W. Saad, Z. Han, R. Zheng, M. Debbah, and H. V. Poor, "A college admissions game for uplink user association in wireless small cell networks," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pp. 1096–1104, IEEE, 2014.

[182] Y. Gu, Y. Zhang, M. Pan, and Z. Han, "Cheating in matching of device to device pairs in cellular networks," in *2014 IEEE global communications conference*, pp. 4910–4915, IEEE, 2014.

[183] S. Bayat, R. H. Louie, Z. Han, Y. Li, and B. Vucetic, "Distributed stable matching algorithm for physical layer security with multiple source-destination pairs and jammer nodes," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2688–2693, IEEE, 2012.

[184] S. Bayat, R. H. Louie, Z. Han, B. Vucetic, and Y. Li, "Physical-layer security in distributed wireless networks using matching theory," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 5, pp. 717–732, 2013.

[185] G. D. Israel, "Determining sample size (pp. 1-5)," *Gainesville: University of Florida Cooperative Extension Service, Institute of Food and Agriculture Sciences, EDIS*, 1992.

[186] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 316–324, 2011.

[187] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*, pp. 99–108, 2010.

[188] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th international conference on World wide web*, pp. 791–800, 2009.

[189] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on gps data," in *Proceedings of the 10th international conference on Ubiquitous computing*, pp. 312–321, 2008.

[190] Y. Zheng, X. Xie, W.-Y. Ma, *et al.*, "Geolife: A collaborative social networking service among user, location and trajectory.," *IEEE Data Eng. Bull.*, vol. 33, no. 2, pp. 32–39, 2010.

[191] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[192] J. Besemer and M. Borodovsky, "Genemark: web software for gene finding in prokaryotes, eukaryotes and viruses," *Nucleic acids research*, vol. 33, no. suppl_2, pp. W451–W454, 2005.

[193] J. Zhou and O. G. Troyanskaya, "Predicting effects of noncoding variants with deep learning–based sequence model," *Nature methods*, vol. 12, no. 10, pp. 931–934, 2015.

[194] A. Dubreuil, *Hands-on music generation with magenta: Explore the role of deep learning in music generation and assisted music composition.* Packt Publishing Ltd, 2020.

[195] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning.* MIT press, 2016.

[196] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.

[197] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[198] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[199] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[200] W. Forecasting, M. Diqi, J. L. A. KM, and C. D. Sleman, "Harnessing the power of stacked gru for accurate weather predictions," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, vol. 6, no. 2, pp. 208–219, 2023.

[201] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[202] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[203] K. Elhariri, "The Transformer Model — towardsdatascience.com." `https://towardsdatascience.com/attention-is-all-you-need-e498378552f9`, March 2022. [Accessed 04-08-2024].

[204] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into deep learning.* Cambridge University Press, 2023.

[205] U. Kamath, K. Graham, and W. Emara, *Transformers for machine learning: a deep dive.* Chapman and Hall/CRC, 2022.

[206] C. C. Robusto, "The cosine-haversine formula," *The American Mathematical Monthly*, vol. 64, no. 1, pp. 38–40, 1957.

[207] P. Dauni, M. Firdaus, R. Asfariani, M. Saputra, A. Hidayat, and W. Zulfikar, "Implementation of haversine formula for school location tracking," in *Journal of Physics: Conference Series*, vol. 1402, p. 077028, IOP Publishing, 2019.

[208] N. M. Steiger and J. R. Wilson, "Improved batching for confidence interval construction in steady-state simulation," in *Proceedings of the 31st conference on Winter simulation: Simulation—a bridge to the future-Volume 1*, pp. 442–451, 1999.

[209] Student, "The probable error of a mean," *Biometrika*, pp. 1–25, 1908.