Studies on Radio Environment Map Construction Methods under Threatening Environment

GAO YING

Department of Communication Engineering and Informatics
The University of Electro-Communications

This dissertation is submitted for the degree of
Doctor of Philosophy

March 2024

Supervisory Committee

Chairperson: Professor Takeo Fujii

1. Member: Professor Koji Ishibashi

2. Member: Associate Professor Koichi Adachi

3. Member: Associate Professor Katsuya Suto

4. Member: Associate Professor Hideki Yagi

Day of the Pre-defense: May 15, 2023.

Day of the Final Defense: Feb. 06, 2024.

I would like to dedicate this thesis to my loving parents . . .

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

GAO YING
March 2024

Acknowledgements

I would like to express my sincere gratitude to everyone who has contributed to my journey in completing my Ph.D. study.

First and foremost, I would like to thank my supervisor Prof. Takeo FUJII, for his invaluable guidance, support, and encouragement throughout my study. His expertise and insight were instrumental in shaping the direction of my research, and I am grateful for his unwavering commitment to my success. Beyond this, Prof. FUJII provided me with a welcoming environment and a sense of belonging, making my transition to a new country and culture much smoother. He has always been approachable, understanding, and caring, always making time to listen and offer advice, not only about my research but also life in Japan.

I would also like to extend my thanks to the members of my dissertation committee, Prof. Koji ISHIBASHI, Assoc. Prof. Koichi ADACHI, Assoc. Prof. Katsuya SUTO, and Assoc. Prof. Hideki YAGI for their insightful feedback and constructive criticism throughout this process. Their expertise in their respective fields provided me with valuable perspectives on my research, and their input has been instrumental in shaping the final outcome of this dissertation.

Furthermore, I would like to thank all the staff and colleagues at Advanced Wireless and Communication research Center (AWCC), who provided me with support, inspiration, and encouragement throughout my research journey. Their camaraderie and friendship made this experience both meaningful and enjoyable.

Additionally, I would like to express my sincere gratitude for the MEXT scholarship support throughout my academic journey. This generous support has allowed me to pursue my academic and professional goals without financial burden, and has given me the opportunity to engage in research and extracurricular activities that have enriched my educational experience.

Finally, I would like to express my deep gratitude to my family, for their unwavering support, encouragement, and understanding throughout this journey. Their love and encouragement gave me the strength and motivation to persevere through the challenges of completing a Ph.D., and I am forever grateful for their presence in my life.

Thank you to everyone who has contributed to my personal and professional growth. Your support and encouragement have meant a lot to me, and I will always be grateful for your kindness and generosity.

和文概要

コグニティブ無線は、利用周波数をダイナミックに切り替えるダイナミックスペクトラムアクセス(DSA: Dynamic Spectrum Access)を用いることで、無線周波数逼迫の課題を解決する新しい無線技術として期待されている。コグニティブ無線によるDSAでは、免許無線局が利用していないとき免許帯域を免許不要局が適応的に利用する検討が行われている。このようなコグニティブ無線ネットワークの実現には、スペクトラム環境を認識し、空き帯域を発見したうえで、適切にスペクトラムを利用するアルゴリズムが必要となる。この時、効率的なスペクトラム管理のために、正確な無線環境の把握が重要となる。その実現の一手法として、無線環境を表現する無線環境マップ(REM: Radio Environment Map)の導入が考えられている。REMを活用することで、コグニティブ無線機はどの周波数を使うことで効率的かつ大容量の通信が可能かを把握し、環境に適応した運用を行うことが可能となる。

REMは多数のユーザから観測情報を集めることで効率的に構築することが可能となるが、悪意のあるユーザが偽のセンサ情報を送るなど多様な攻撃を行うことで、その正確性が低下する課題がある。不正確なREMは、通信効率の低下や、無線システム相互干渉の増加、ネットワーク性能の低下につながる可能性を持つため、攻撃にロバストなREM構築および管理手法が重要となる。

そこで、本論文では、悪意のあるユーザが存在する環境でも、信頼性が高く正確なREMを構築する手法を提案する。

一つ目の手法では、悪意のある端末と誠実な端末を効果的に区別することができるDouble Layer Monitor (DLM) アルゴリズムの概念を提案している。提案するDLMモデルでは、同じメッシュからのすべてのレポートとの類似度を計算する。単一メッシュでは、正常な情報よりも悪意のある情報が多い可能性があるため、類似度だけをチェックすると、悪意のある情報に目を奪われてしまう可能性があり、それを避けるため過去の評価結果も考慮する。さらに、空間情報アルゴリズムに基づき、悪意のあるノードを正確に特定することで、ネットワーク性能の向上を図る。また、最適な攻撃戦略を検討し、最適な攻撃における最大誤差を求めることで、強いネットワークセキュリティを達成した。

二つ目の手法はデータセットが不十分な状態も考慮する。このような場合、悪意のある端末の影響を避けるために、少数のトラストノードを設定するKriging-based Trust Nodes Aided (KTNA) REM構築アルゴリズムを提案する。ノードのリアルタイム評価と累積評価を行うことで、REMの精度を向上させ、各センシングスロットにおいて各位置で利用できるデータ量が少ないことに起因する精度制限の問題を解決する。本手法は、悪意のあるノードの影響を効果的に回避し、ネットワーク全体のパフォーマンスを向上させることができた。

加えて、KTNAアルゴリズムに基づき、信頼セット内の信頼できる情報を自動的に収集し、無効な情報を削除する再起動フェーズを追加した改良型KTNAアルゴリズムを提案し、周辺環境が変化した際にシステムを初期化する手法を検討する。KTNAアルゴリズムに基づき、精密なREMを構築し、選択されたデータに基づいて、平均パスロスやシャドーイングの影響を含むチャネル状態を推定できるKTNA+システムの構造を提示する。様々なシミュレーションにより、KTNA+システムの性能を検証し、ネットワーク性能と向上とセキュリティ耐性の向上に有効であることを実証した。

従来、ユーザデータを用いて協調センシングを行う際には、悪意のある端末からの干渉を回避することは困難である。多くの研究は、情報提供端末が信頼できることを前提にしている。しかし、様々な携帯電話から情報を収集するような実世界のアプリケーションにクラウドセンシングの関連技術を適用する場合、すべてのユーザの絶対的な正確さを保証することはできず、利己的なユーザが周波数リソースを独占しようとしたり、主要ユーザの通信を妨害したりすることを防ぐことは困難であった。今回の研究はその解決法の提示を行っている。

本研究は、悪意のある情報を根絶することで、今後のREMを用いたカバレッジ最適化、リソース割り当て、干渉解析、位置推定などの研究において、より安全な環境を提供できると考える。今後、より安全・快適・安心な通信環境を提供するためには、脅威環境下でのREM構築やチャネル推定システムに関する研究を幅広く行うことが必要となる。

Abstract

Cognitive radio (CR) is a wireless communication technology that has emerged as a promising solution to address the spectrum scarcity problem by enabling dynamic spectrum access (DSA). It allows unlicensed users to opportunistically access the licensed spectrum when it is not in use by licensed users. Cognitive radio networks (CRNs) rely on intelligent algorithms to sense the spectrum environment, detect available frequencies, and optimize the use of the spectrum. However, accurate knowledge of the radio environment is crucial for efficient spectrum management. This has led to the development of the Radio Environment Map (REM), which provides a comprehensive representation of the radio environment. The REM enables cognitive radios to make informed decisions on spectrum access, leading to efficient spectrum utilization and increased network capacity.

However, the accuracy of the Radio Environment Map (REM) can be threatened by various malicious attacks, in particular, data falsification attacks have a serious impact on the accuracy of REM construction. REM plays a critical role in CRNs, inaccurate or outdated REM information can lead to inefficient spectrum usage, interference, and reduced network performance. Therefore, it is essential to develop robust REM construction and maintenance techniques to ensure the accuracy and reliability of the map.

In our study, we propose two REM construction methods to address the challenge of developing a reliable and accurate REM under malicious attack. Firstly, when there is sufficient data available, we propose a Double-Layer Monitor (DLM) based on spatial-information approach, including IDW and spatial correlation, to deal with various data falsification attacks in the network. By analyzing the real-time similarities and historical performance, to reject the malicious information from the database.

However, in scenarios where there is limited data available, such as in a new or rapidly changing radio environment, the proposed DLM-based approach may not be effective. Therefore, we also propose a Kriging-based Trust Nodes Aided (KTNA) algorithm to construct the REM, by adding a small amount of trust nodes, effectively guaranteeing the accuracy of the REM. Additionally, based on the KTNA algorithm, we add the reboot part, so the system can lead to an accurate REM under the dynamic environment. We also

give a solution for channel estimation including the path-loss and shadowing effect under a threatening environment.

Overall, our study contributes to the development of reliable and accurate REM construction methods for CRNs under threatening environments, which is essential for the successful deployment of cognitive radio technology in the future.

Contents

List of Figures

List of Tables

Nomenclature

Roman Symbols

a	reward index in Chapter 5
$ATRe_k(i)$	the accumulative total reputation of the i-th terminal at k-th slot
b	penalty index in Chapter 5
C	covariance
D	number of meshes
d	distance
d_0	reference distance
d_{cor}	correlation distance
H_0	Null hypothesis
H_1	Alternative hypothesis
$HisRe_i^H$	historical reliability of the i-th terminal at the step H
$\mathcal{HT}$	honest terminal set
$I_{\max}$	maximum interference
L	propagation loss
m	sill variance
$\mathcal{MT}$	malicious terminal set
N	number of the sensing terminals

P	received signal power
p	IDW power parameter
P'	reported signal power
$\hat{P}$	estimated signal power
P_a	malicious terminals attacking probability
$para$	a reward-penalty parameter
P_d	detection probability
P_f	false alarm probability
P_m	miss detection probability
P_{T_x}	primary transmission power
$\hat{P}_{T_x}$	estimated transmission power
Q	spatial random process
Q'	detrend spatial random process
r	range
Re	a reward-penalty function
Rel	normalized supported level in DLM
Sim	similarity degree in DLM
Sup	supported level in DLM
W	shadowing loss
$\hat{W}$	estimated shadowing loss

Greek Symbols

α	similarity threshold in DLM
β	historical reliability threshold in DLM
Δd_{ij}	distance between the i-th and j-th terminals

δ	data falsification attacking index
ϵ	adjusted error
η	path-loss coefficient
$\hat{\eta}$	estimated path-loss coefficient
Γ	SIR
γ	Seimivariogram
κ	data falsification attack attacking threshold
λ	wavelength of the signal
μ	spatial process drift
ω	weight factor
$\omega\_HisRe_i^H$	the reputation weight of $HisRe_i^H$
$\rho_{i,j}$	correlation index
υ	Lagrange multiplier
ξ	spatial random fluctuation exponent
ζ	accumulative total reputation system threshold

Subscripts

cor	correlation
max	maximum value
min	minimum value
Tx	transmitter
Rx	receiver
sum	summation

Other Symbols

ABMT	All But Malicious Terminals

AE	Adjusted Error
arg	argument
AST	All Sensed Terminals
AT	Accumulative Total
CR	Cognitive Radio
CRNs	Cognitive Radio Networks
CSS	Collaborative Spectrum Sensing
DLM	Double-Layer Monitor
FCC	Federal Communications Commission
FC	Fusion Center
GP	Gaussian Process
GPS	Globla Positioning System
histo	same with DLM in Chapter 3
IDW	Inverse Distance Weighted
IoT and IoV	Internet of Things/ Vehicles
JIS	Japanese Industrial Standard
KTNA	Kriging-Based Trust Nodes Aided
M2M	Machine-to-Machine
MAE	Mean Absolute Error
NSED	Normalized Sensed-and-Estimated Difference
PUE	Primary User Emulation
PUs	Primary Users
REM	Radio Environment Map
RF	Radio Frequency

RMSE	Root Mean Squared Error
RSSI	Received Signal Strength Indicator
RT	Real Time
SDR	Software-Defined Radio
sim	same with DLM first layer in Chapter 3
SIR	Signal-to-Interference power Ratio
SNR	Signal-to-Noise Ratio
SPTF	Spectrum Policy Task Force
SSDF	Spectrum Sensing Data Falsification
SS	Spectrum Sensing
SUs	Secondary Users
SVM	Support Vector Machines
TVWS	Television White Space
WLAN	Wireless Local Area Network
CDF	Cumulative Distribution Function

Chapter 1

Introduction

1.1 History and Background

The rapid expansion of wireless communication technology in recent years has created an extraordinary need for radio spectrum. Nevertheless, the increase in demand has faced a constraint due to the limited availability of resources. The spectrum, which is crucial for many wireless communication applications like terrestrial and satellite communications, Wi-Fi, and bluetooth, is currently encountering constraints that impede its capacity to keep up with the increasing demand. This predicament has led to a critical scarcity of available frequencies, attracting the focus of governments, regulatory bodies, and industry participants worldwide.

The limited availability of spectrum resources presents a barrier, which is caused by the allocation of licensed frequencies for various applications. As depicted in Fig.1.1 for the United States, the frequencies that are most sought after have already been allocated, which presents a difficult situation for wireless operators as they endeavour to get an adequate amount of spectrum. The limited availability of this resource directly adds to the deterioration of network performance, exacerbating the problem of spectrum shortage.

As a solution to this difficulty, the adoption of intelligent spectrum sharing and reuse has become a prominent concept, with Cognitive Radio (CR) technology leading the way. Recognised for its capacity to improve the utilisation of spectrum and mitigate the shortage of spectrum resources, CR technology is considered a vital element in meeting the increasing demands for both new and existing services. Nevertheless, the successful functioning of CR systems is highly dependent on dependable spectrum awareness. This specific aspect has captured the attention of scholars, functioning as a strategy for detecting signals with adaptable uses in radio surveillance, spectrum allocation, and diverse fields.

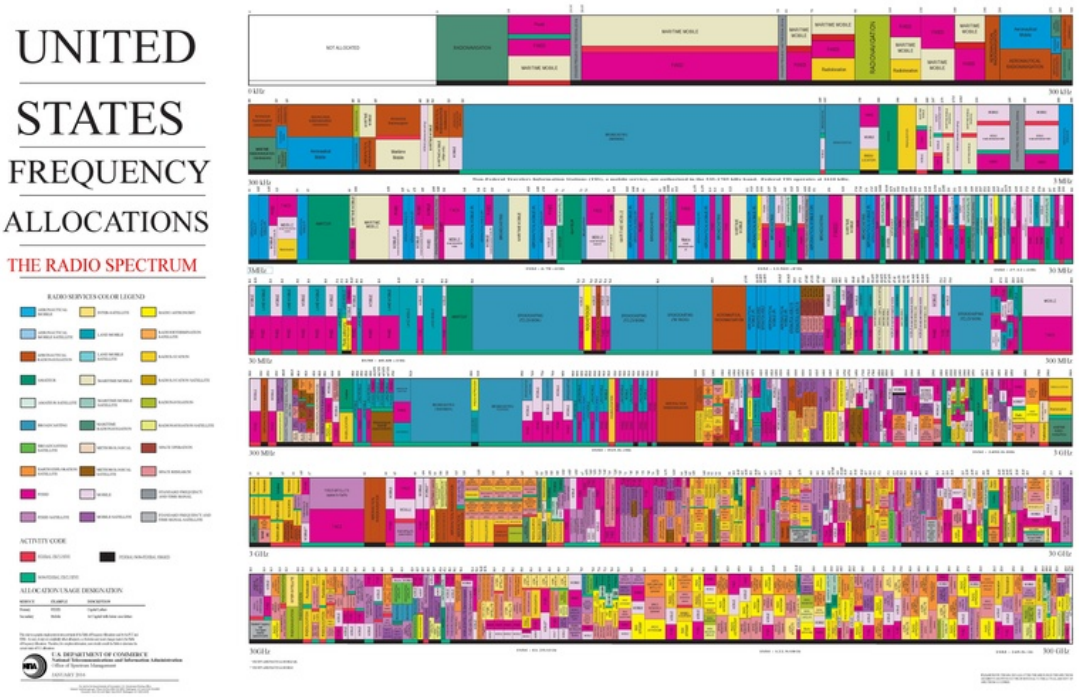


Figure 1.1 United States frequency allocation chart 2016 [1].

1.1.1 The first proposal of cognitive radio (CR)

Enhancing user perception and service security in the context of the Internet of Everything (IoE) and the rapid expansion of services is a significant research focus. Wireless network resource control and optimising service security performance are key areas of interest. The expected rise in the number of connected mobile devices is predicted to create a mismatch between the growing need for spectrum resources and the limited availability of spectrum. The pursuit of higher communication frequency bands is accompanied by the urgent adoption of dynamic spectrum sharing schemes and flexible allocation of spectrum resources. [2][3].

The conventional techniques used for allocating spectrum, as emphasised by the US Spectrum Policy Task Force (SPTF) and the Federal Communications Commission (FCC), have shown inefficient utilisation of the spectrum. The spectrum management philosophy of exclusivity, which is considered overly restrictive, impedes the availability of spectrum access [4]. According to spectrum usage surveys, the utilisation of available spectrum in a given region is estimated to be between 2% and 6% [5][6]. The analysis conducted by the Berkely US Wireless Research Centre highlights the fact that the assigned spectrum resources are not being fully utilised. This is clear from the presence of unoccupied bands and severe under-utilization. Fig.1.2 and Fig.1.3 depict conceptual representations of spectrum waste in the spatial and time-frequency domains, respectively.

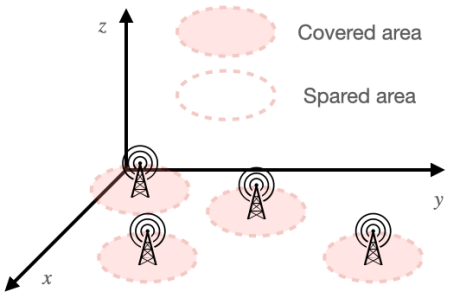


Figure 1.2 Spatial hole.

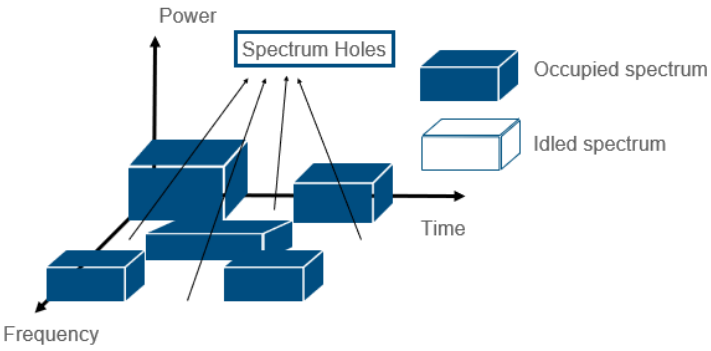


Figure 1.3 Time-frequency hole.

In response to these challenges, experiments monitoring radio spectrum use across common frequency bands reveal wide variations in spectrum utilization across time, frequency, and spatial domains [7, 8]. Fixed spectrum allocation methods result in low spectrum utilization and potential wasted spectrum, denying unlicensed users access to vacant bands. The advent of Cognitive Radio (CR) offers a viable solution.

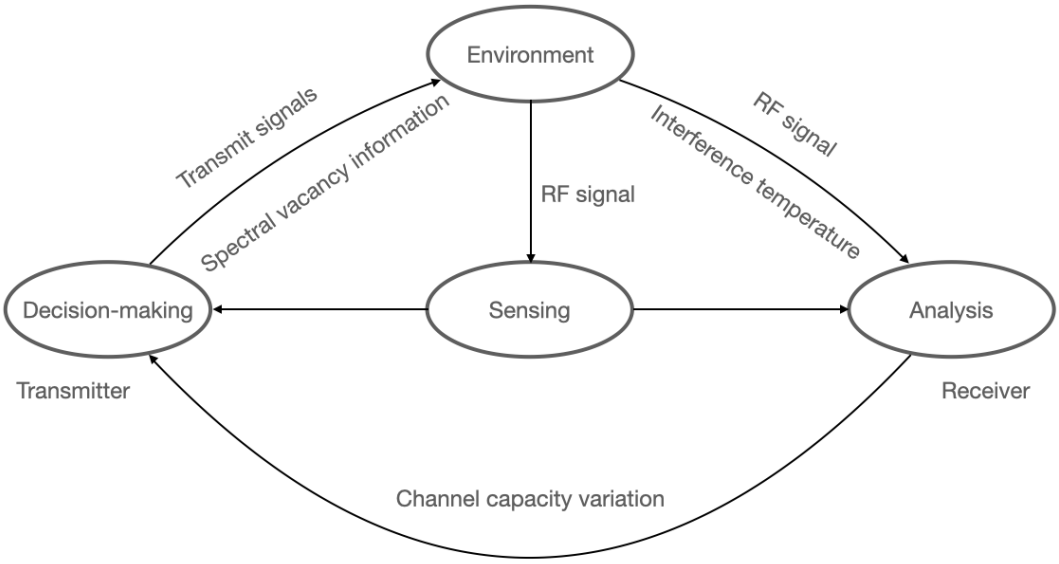


Figure 1.4 Architecture model of a CR loop.

The idea of CR dates back to the early 1990s when Joseph Mitola III, a researcher at the Swedish Defence Research Agency, proposed the concept of software-defined radio (SDR). Mitola's vision was to create a radio system that could reconfigure itself to operate on any frequency band, modulate any waveform, and adapt to any wireless environment. The concept

of SDR later evolved into CR, which added the ability to sense the wireless environment and make decisions based on the available information.

In the late 1990s, the FCC in the United States recognized the potential of CR to address the spectrum scarcity problem and issued a notice of proposed rule making to explore the use of CR in the wireless industry. This led to the formation of a group of industry experts who worked to develop standards and regulations for CR.

Since then, CR has been the subject of extensive research and development efforts, and it has been successfully deployed in various applications, including public safety networks, and wireless sensor networks. In recent years, CR has also gained attention in the field of 5G and beyond wireless communication systems, where it can play a crucial role in enabling efficient spectrum utilization and enhancing network performance.

CR is a revolutionary technology that has the potential to significantly improve the efficiency and reliability of wireless communication systems. It is an intelligent radio technology that uses machine learning algorithms and signal processing techniques to dynamically adapt to the wireless environment and optimize the use of available spectrum. CR technique allow unlicensed terminals, which is also called Secondary Users (SUs), opportunistic access to authorised spectrum allocated to the licensed terminals, which is also called Primary Users (PUs), for spectrum sharing.

In cognitive radio networks (CRNs), one key technique is the Spectrum Sensing (SS), the main function of SS is to detect the usage status of an authorized frequency band. Once an available spectrum hole is detected, the SU can dynamically access it by adjusting its own operating parameters. Due to the significant advantages of CRNs in terms of spectrum resource reallocation, these years, it has been widely used for integration with IoT [9–11], Internet of Vehicles (IoV) [12–14], wireless sensor networks [15–17] and so on.

1.1.2 The emergence of radio environment map (REM)

Radio Environment Map (REM) is a tool that provides statistical information on the radio frequency (RF) environment in a given area [18–20]. It is a critical component of CRNs and can be used to optimize the use of available spectrum resources. The idea of REMs dates back to the early 2000s when CR was emerging as a new technology for improving spectrum utilization [21]. REM technology is based on the concept of using sensors to monitor the RF environment in a given area and then creating a map of the spectrum use in that area. The map can then be used by CR devices to make decisions on which frequency bands to use and how to optimize their transmissions.

REM assists in making better use of geolocation information, such as the efficient use of spectral resources in wireless communication systems. It is also defined as a database with

geo-located information applied to dynamic spectrum access in CR systems. REM can be used to make decisions in a variety of applications, including coverage optimization [19], resource allocation [22], interference analysis [23], location estimation [24][25], and so on. The development of REM technology has played a significant role in advancing the field of CR and improving the efficiency and reliability of wireless communication systems.

Crowdsourcing has emerged as a popular approach to data collection and analysis, leveraging the power of the crowd to complete complex tasks efficiently and accurately [26]. In wireless communication systems, crowdsourcing can be applied to construct REMs, which provide information about the propagation of radio signals in a given area. Crowdsourcing REM construction involves the participation of multiple terminals in the area, which collaboratively gather data and transmit it to a central server for processing. By pooling data from multiple sources, the resulting REM can provide a more accurate representation of the wireless environment. However, the open nature of crowdsourcing introduces security threats, such as the intentional or unintentional falsification of data by malicious terminals [27]. To address this issue, secure crowdsourcing REM construction methods need to be studied.

1.1.3 Security issues faced by CRNs

SS is used for detecting the current band usage accurately in real-time, which is an important aspect and a key prerequisite for CR implementation. While a single SU adopts a specific method to detect the operating status of a particular PU in a certain frequency band is called local sensing or independent sensing. The process of detection by multiple SUs working and sharing the information together is called Collaborative Spectrum Sensing (CSS), in which perceived limitations of individual users' sensing can be effectively avoided.

However, the open nature of the wireless channel exposes CR networks to a number of serious security threats. Besides facing some traditional security threats (like denial-of-service attacks), the cognitive and intelligent nature for dynamic spectrum access also make CRNs face some specific safety hazards and challenges. For example, Primary User Emulation (PUE) and Spectrum Sensing Data Falsification (SSDF) attack [28]. Among them, SSDF can have a serious impact on the CRNs through malicious users tampering with sensing information, thereby blinding the fusion center and causing the database to make incorrect global decisions. With their low-cost and high-reward attack modes, as well as their flexibility and diversity of attack methods, the SSDF attacks have become a major component of cognitive radio network attacks and be widely used and discussed. By simply tampering with sensing information, global decisions can be influenced, which in turn can have a deleterious

effect on the cognitive network. Hence, it is a big challenge for the effective defense against SSDF attacks.[29]

1.2 Related Work

1.2.1 Spectrum sensing

CR is a wireless communication technology that enables dynamic spectrum access for wireless devices to improve spectrum utilization efficiency. CR utilizes SS techniques to detect and utilize the available spectrum in real-time. There are two main types of sensing methods used in CR: local sensing and collaborative sensing.

1.2.1.1 Local sensing

Local sensing involves a single device that senses the spectrum locally and makes a decision based on its own observations. This approach is simple and efficient, but may not provide accurate results in dynamic environments or in the presence of hidden terminals.

Local sensing can be performed using different techniques, such as energy detection, matched filtering, and cyclostationary feature detection [30–33]. Among them, energy detection is a simple and widely used technique that involves measuring the energy level in a particular frequency band and comparing it with a threshold to determine the presence of a primary user signal [30][31]. Matched filtering, on the other hand, is a technique that correlates the received signal with a known template or reference signal to detect the presence of a signal with a specific waveform or modulation scheme [32]. Cyclostationary feature detection is a more advanced technique that exploits the statistical properties of the primary user signal, which exhibit periodicity in their auto-correlation and power spectral density functions. Cyclostationary feature detection involves extracting specific features, such as cyclic frequencies or correlation coefficients, from the signal and comparing them with a threshold to detect the presence of a primary user signal [33].

Local sensing is efficient and does not require coordination with other devices, but it may not provide accurate results in dynamic environments or in the presence of hidden terminals.

1.2.1.2 Collaborative sensing

Collaborative sensing (also cooperative sensing) is a technique in which multiple cognitive terminals work together to sense the radio frequency spectrum. In collaborative sensing, each terminal shares its sensing results with other terminals in the network, allowing the network

to combine and aggregate the information from multiple terminals to make a more accurate decision about the presence or absence of primary user signals in the spectrum. Collaborative sensing can be classified into four categories: centralized, distributed, relay-assisted, and clustering. [34–37]

Over them, the centralized and distributed networks are the widely used model. Centralized collaborative sensing involves a Fusion Center (FC) that collects sensing reports from all the terminals in the network and makes a final decision based on the aggregated data. In this approach, the FC has complete control over the sensing process and can optimize the sensing parameters and algorithms to improve the sensing performance. However, centralized collaborative sensing has some disadvantages, such as high communication overhead, and potential privacy concerns. Distributed collaborative sensing, on the other hand, involves each sensing terminal making its own decision based on its local sensing results and sharing its decision with neighboring terminals. In this approach, there is no FC controlling the sensing process, and each sensing terminal can adapt its sensing parameters and algorithms based on its own local environment. Distributed collaborative sensing has some advantages over centralized collaborative sensing, such as low communication overhead, robustness against node failures, and better privacy protection. However, distributed collaborative sensing may suffer from inconsistency and lack of synchronization among the different terminals, which can affect the overall sensing performance.

Additionally, decision fusion and data fusion are two different approaches to combining the sensing data from multiple cognitive terminals to detect the presence or absence of primary users in the spectrum. [38][39]

Decision fusion also called hard decisions, is a technique in which each cognitive radio makes a binary decision (e.g., "signal present" or "signal absent") based on its local sensing results, and a FC combines these binary decisions using a OR/ AND/ K out of N/ Majority rule. The FC then makes a final decision about the presence or absence of a primary terminal's signal. Decision fusion can be more efficient in terms of communication overhead and delay than data fusion, as it requires each cognitive terminal to transmit less information. However, decision fusion may suffer from the limitations of individual sensing, as it does not take into account the diversity and redundancy of the sensing results from multiple terminals.

Data fusion also called soft decisions, is a technique in which each cognitive terminal reports its raw sensing data, such as the received signal strength or the spectrum occupancy, to an FC. The FC then combines the data from all the cognitive terminals using techniques such as maximum likelihood estimation or Bayesian inference to make a decision about

the presence or absence of a primary terminal's signal. Data fusion can provide a more accurate and reliable decision than individual sensing, as it takes into account the diversity and redundancy of the sensing results from multiple terminals. However, data fusion requires each cognitive terminal to transmit more information, which may cause higher communication overhead and delay.

The choice between decision fusion and data fusion depends on various factors such as the network topology, the sensing technologies, and the computational capabilities of the CRNs. Data combination can provide more accurate and reliable results, but it requires more communication overhead and computational resources. Decision combination, on the other hand, is simpler and more efficient, but it may suffer from low accuracy and reliability due to the diversity of the sensing results from different cognitive terminals.

1.2.2 Spectrum sensing data falsification attack

When several secondary terminals are involved in CSS, detection accuracy is improved; nevertheless, when terminals accidentally or maliciously broadcast false sensing data to the database during collaboration, it may lead to bad global decision-making [28]. This CSS assault is also known as the SSDF attack, and it has a devastating effect on the accuracy of collaborative detection [29].

Consequently, it is crucial to implement sufficient security measures in a wireless environment plagued by malicious terminals. In most cases, there are three broad categories into which current solutions might be placed. In the first category, false spectrum measurements are isolated using statistical anomaly detection and removed. The evidence theory-based CSS approach proposed by Han et al. as a defense against the SSDF assault in reference [40]. This technique assesses the evidence's credibility by its degree of resemblance and filters out any pieces of evidence that have a low similarity degree. In [41], a Bayesian-based approach was proposed for determining the credibility of spectrum sensing reports and filtering out those that could be inaccurate. As the second category, references like [42] and [43], as well as reference [44], which proposed a partitioning around medoid algorithm to cluster and reputation adjust the terminals to improve sensing. In this category, the database monitors long-term behaviors using a reputation system to distinguish the malicious terminals. FastDtec, described in [45], is a trust assessment technique developed to counter both collaborative and independent attacks. The third category employs machine learning techniques to identify valid measures from malicious ones. Spectrum readings were classified as valid or invalid using a Support Vector Machines (SVM) based approach, as cited in reference [46]. A Joint spectrum sensing and resource allocation (JSSRA) strategy, which is

itself a learning process using the trust degree combination approach to resist assault, was presented in the [47].

1.2.3 Radio environment map

In order to enhance the accuracy of the spectrum estimation, one simple method is using the REM construction to refine a primary terminal's coverage map. REM is a map that describes the radio activities of primary users, the REM includes the received signal strength (RSS) of primary users at different locations of interest, which can be directly measured through spectrum sensing or estimated using appropriate statistical spatial interpolation methods.

REMs have seen a recent uptick in popularity for use in wireless communication systems [18–20]. Effective use of spectrum resources in wireless communication networks is one example of how REMs improve geo-location data use. They are also understood as geo-spatial databases used for cognitive radio systems' dynamic spectrum access. The accumulated measurement data may be used to generate an REM that displays the typical received power at each location. Having access to this information allows us to make accurate predictions about the local radio environment, such as cellular network coverage and Wireless Local Area Network (WLAN) communication quality [48][49]. Decision-making using REMs is possible in many contexts, such as optimization of coverage [19], allocation of resource [22], analysis of interference [23], and estimation of location [24, 25, 50].

One of the key benefits of REM is that it allows cognitive terminals to operate more efficiently and reliably in dynamic and uncertain radio environments. By providing accurate information about the radio environment, REM enables cognitive terminals to adapt their behavior, avoiding interference with other users and maximizing spectral efficiency. REM is also important for ensuring spectrum sharing and coordination between different users and applications, as it provides a common reference point for all parties involved.

The well-known method of spatial interpolation known as Ordinary Kriging is capable of exactly obtaining a REM value. Ordinary Kriging was used by Phillips et al. [51], in order to figure out how well a WiMAX network operating at 2.5 GHz would cover a university campus in the United States. Additionally, a study in [52] revealed how the accuracy of the Television White Space (TVWS) geo-location database could be improved by using Ordinary Kriging to anticipate the primary user's signal intensity from a very small number of samples. This research was released to show how the accuracy of the database might be raised. Another piece of measurement study conducted in Seattle, Washington in [53] reinforces the benefits of Ordinary Kriging over model-based prediction approaches such as the Longley-Rice model, FCC F-Curves, and k closest neighbor.

Since the accuracy of the REM can improve the efficiency of the spectrum directly, the accuracy of the measurement data is significantly important.

1.2.4 Challenges

It is widely known that statistical spatial interpolation techniques, such as Ordinary Kriging, are sensitive to outliers, which may be induced by masking or cheating. All existing works in the REM-related studies use the premise that all data can be trusted, but none of these works really prove this claim. The evidence shown in [54], that the precision of the prediction may be impacted even by a small quantity of inaccurate data.

Additionally, the goal of REM construction is to estimate the Received Signal Strength Indicator (RSSI) of the primary terminal at the area of interest, so the method to estimate the RSSI at a certain location under a threatening environment needs to be studied. The secure mechanisms in CRNs intend to detect whether or not the primary terminal exists at the location of interest, which is not suited for REM construction problems.

Accordingly, improving communication quality by constructing an accurate REM under attacks, is a problem that needs to be addressed.

1.3 Contributions

The ever-growing demand for high-speed and reliable wireless communication has led to an increasing number of wireless networks in recent years. However, wireless networks are easy to be affected by various security threats, such as a wide variety of data falsification attacks. Therefore, it is crucial to develop effective security mechanisms to ensure the authenticity, stability, and availability of wireless networks.

To address the above issues, we propose novel algorithms and systems to improve the performance and security of wireless networks. Specifically, we consider the scenarios when the data size is sufficient and insufficient conditions separately, and give different solutions to them. We introduce the Double-Layer Monitor (DLM) algorithm under the sufficient data condition. Additionally, the Kriging-based Trust Nodes Aided Radio Environment Map construction (KTNA-REM) algorithm to an insufficient condition, which can effectively identify malicious nodes, and avoid their impact. Also, the improved KTNA system, which is called KTNA+ system can estimate channel conditions accurately. Our contributions can provide valuable insights into enhancing the performance and security of wireless networks and benefit their development.

1.3.1 An improvement of security scheme for REM construction under massive attacking

In Chapter 3, during the process of creating the REM, we offer a technique for anti-malicious terminals that utilizes information depending on its physical location. The following are some of the contributions that this chapter makes:

- By using similarity comparison and sustainable monitor, we propose a DLM algorithm to identify malicious terminals. Also, the reward-penalty function of DLM algorithm is optimized in order to make it more efficient.
- In order to improve the network functionality, we suggested to use a DLM that is based on spatial information methods. These algorithms include inverse distance weighting (IDW) and spatial correlation.
- In order to thoroughly assess the performance of our algorithms, their error performance under the optimal attack strategy (strongest attack) is investigated.

1.3.2 Kriging-based trust nodes aided REM construction method

In Chapter 4, in order to create a radio map with high accuracy using a small number of terminals in a threatening environment, we present a KTNA-REM algorithm. KTNA-REM is a real-time constructing method that removes the less trustworthy data from the dataset and keeps the most trustworthy data after each sensing slot. The simulation results demonstrate that the KTNA-REM maintains consistent performance in the face of a variety of threats. This chapter makes the following contributions:

- To mitigate the effects of malicious nodes, we present the KTNA-REM technique, in which we set a small amount of the trust nodes in the interested communication area.
- The accuracy of REM has been greatly enhanced by establishing the trustset, which was enhanced by analyzing the accumulative total reputation. The problem of low precision caused by a lack of data for sensing has been addressed.

1.3.3 An improved Kriging-based radio environment map construction and channel estimation system in threatening environments

In chapter 5, to get closer to a high-precision radio map while employing a limited number of terminals under attacking, we improved our KTNA-REM generation technique into a REM

construct and channel estimate system named as KTNA+ [55]. The KTNA+ is a channel estimate system and REM construction system based on the KTNA algorithm. The map is dynamically updated by the system. The simulation results demonstrate the system's consistent operation in the face of a variety of threats. This chapter makes the following contributions:

- This is a novel and challenging study designed to estimate channel information based on REM construction data in threatening environments.
- We present an improved version of the KTNA algorithm. The influence of malicious terminals may be mitigated by automatically collecting a limited group of trust nodes. In addition, the reboot phase allows it to efficiently remove faulty data and restart the system at the appropriate moment.
- We discuss the architecture of the KTNA+ system, which, using the KTNA algorithm as a foundation, can accurately generate a REM based on the estimation of the channel state, including the average path-loss and the influence of shadowing. The exhaustive simulation results verified the effectiveness.

REM is essential for assessing the wireless environment and enhancing the quality of wireless communication to fulfill the ever-increasing demands. Nevertheless, when implementing the technology in practical scenarios, like gathering data from different mobile devices, it is challenging to ensure complete honesty from all users. Additionally, preventing self-interested users from monopolizing spectrum resources or interfering with the communication of primary users becomes a complex task. Our research specifically addresses this aspect of the problem and significantly contributes to the extensive variety of practical applications in the future.

1.4 Organization

This dissertation provides a study about the REM construction under attacking methods. We introduce the related background and knowledge in Chapters 1 and 2. Next, Chapter 3 discusses how to improve the accuracy of constructing a REM with the interference of malicious terminals affect with massive data. After we propose a trust nodes-aided method to deal with the condition when the collected data is insufficient in Chapter 4. In Chapter 5, we extend our algorithm to a comprehensive REM construction system, that can effectively counter malicious terminals attacks and changes in the surrounding environment. Finally, the conclusion and future works are list in Chapter 6.

Chapter 1: Introduction In this chapter, we give an introduction about the background, related works, and challenges. We summarize the conducted research contributions from chapter 3 to 5 and the organization of this dissertation is also described in this chapter.

Chapter 2: System model In this chapter, we explain the specific operating system models of spectrum sensing. Additionally, we introduce the basic database model and REM construction model, which will be used in the later chapters. Moreover, we introduce sensing security issues, which is the main concern in our research, especially, the data falsification attack.

Chapter 3: An improvement of security scheme for REM construction under massive attacking In this chapter, we introduce the DLM algorithm, based on the sufficient collected data, by checking the similarity degree of the real-step information as well as the historical performance, also the sustainable monitored reliability to distinguish the malicious information, by flagging the malicious terminals to remove their information from the database, thus effectively reduces the interference caused by the malicious terminals. Additionally, we improve the DLM by combining with spatial information, including DLM based on spatial correlation and DLM based on IDW. Finally, an optimal attack (strongest attack) under the given secure algorithm is presented.

Chapter 4: Kriging-based trust nodes aided REM construction method The above chapter mainly considers when the collected data is sufficient, in this chapter, we introduce a new scheme, to construct the REM based on the insufficient dataset. By adding a small amount of trust nodes, against the malicious terminals attack, the real-time and the accumulative total reputation need to be checked in our algorithm. Based on the kriging interpolation, the accuracy of REM construction increased a lot. By adapting our method, it is able to guarantee REM accuracy and network security when the data is insufficient.

Chapter 5: A kriging-based REM construction and channel estimation system in threatening environments The above chapters mainly construct the REM under a stable surrounding environment. In this chapter, we expand our algorithm to a REM construct system, which considers the changes in the environment to reboot the system. Also, we give a channel estimation solution including the estimation of the average path-loss and shadowing impact under the threatening environment. Based on this, we constructed the REM and adjusted the REM with the adjusted error support, to achieve a high-quality REM under the threatening environment.

Chapter 6: Conclusions and future scopes In this chapter, we give the conclusions of this dissertation. Additionally, we discuss the future scopes based on our study in this chapter.

Chapter 2

System Model

To enhance communication quality, REM proves invaluable by effectively estimating the surrounding environment through the aggregation of sensing reports obtained from various terminals. This data is systematically collected and stored in a comprehensive database, enabling the construction of an accurate REM. By leveraging this resource, communication networks can adapt and optimize their performance, ensuring a more reliable and efficient exchange of information in response to the ever-changing radio conditions.

In this chapter, first introduces the sensing model, database model and REM construction model are explained next, and finally discusses the sensing security issues in the CRNs.

2.1 Sensing Model

Spectrum sensing is the cornerstone of cognitive wireless networks towards practical applications, while sensing performance is the basic indicator of sensing effectiveness. How to design an effective and reliable sensing network has become a hot research topic and focus in academia.

2.1.1 Spectrum Sensing

Accurate detection of spectrum holes is the primary task for CR system implementation. In order to accurately detect weak signals emitted by the PU, SUs must have high local sensing sensitivity, and the binary assumption problem in spectrum sensing is modeled as follows,

$$x(t) = \begin{cases} n(t), & \mathrm{H}_0 \\ h \cdot s(t) + n(t), & \mathrm{H}_1 \end{cases}, \tag{2.1}$$

where $x(t)$ is the received signal of the secondary terminal, $s(t)$ is the signal of the primary terminals, $n(t)$ is the noise, and the h is the channel gain, hypotheses H_0 and H_1 indicate whether the sensing frequency band is idle or not, respectively.

Additionally, the detection probability P_d, the false alarm probability P_f, and the miss detection probability P_m are important indices to evaluate the sensing performance. These indices can be indicated as follows,

$$\begin{cases} P_d = \Pr(H_1|H_1) \\ P_f = \Pr(H_1|H_0) \\ P_m = 1 - P_d = \Pr(H_0|H_1) \end{cases}, \tag{2.2}$$

where P_d indicates the probability that the secondary terminal correctly detects when the primary terminal is working, P_f indicates the probability that the secondary terminal falsely detects that the primary terminal exists when the primary terminal is not working, and P_m indicates the probability that when the primary terminal is working, however, the secondary terminal judges the channel condition as idle.

2.1.2 Detection methods

As a basic element of a cooperative sensing system, local sensing by a single node can be divided into two types from the perspective of signal detection technology: coherent sensing and non-coherent sensing. The classification is based on whether prior knowledge of PU signals is required. The frequency bandwidth of the detection object is divided into narrowband and wideband. The overall classification diagram is presented in Fig.2.1.

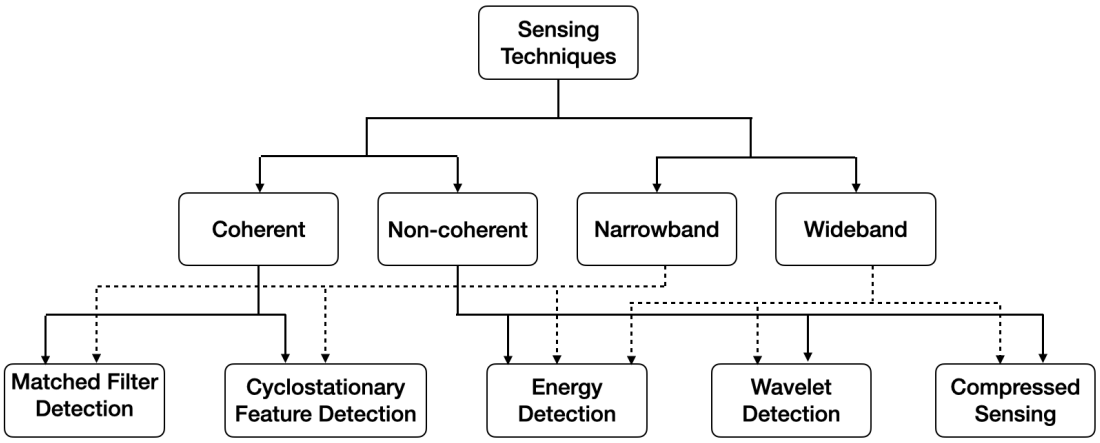


Figure 2.1 Sensing classification diagram.

Since the accuracy of spectrum sensing is very critical to the cognitive network, accurate sensing results are the prerequisite for ensuring the stability of the cognitive network. High-precision sensing results can improve the capacity and stability of the cognitive network. Conversely, low-precision sensing results will cause great waste, interference and even damage to the network. In this dissertation, we mainly used Energy detection.

Energy detection is a commonly used technique for spectrum sensing in CRNs. It involves measuring the energy of the received signal in a particular frequency band and comparing it to a threshold value to determine the presence or absence of a primary user signal.

Given $x(t)$ is the signal to be detected, $n(t)$ is the white noise, for each sensing slot T and the detection bandwidth W, the detection statistics Y can be demoted as follows,

$$Y = \sum_{k=1}^{2TW} |x(k)|^2, \tag{2.3}$$

where $2TW$ is the time-bandwidth product. Then we compare the detection statistics Y with the given threshold λ, if $Y > \lambda$, then we judge the primary terminal is working, otherwise, we consider the certain band is idled. The energy detection principle block diagram is like Fig.2.2 shows.

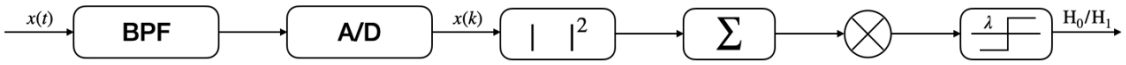


Figure 2.2 Diagram of energy detection.

The energy detection technique is simple and easy to implement, and it can be used with a wide range of modulation schemes and signal types. However, it has some limitations, such as sensitivity to noise and interference, and the requirement for accurate estimation of the noise power. These shortages can be improved by using collaborative sensing or using adaptive thresholding.

Indeed, energy detection is a valuable technique for spectrum sensing in CRNs, and its performance can be improved by combining it with other techniques or by using adaptive and intelligent sensing algorithms. Because it is simple and easy to implement, most collaborative spectrum sensing uses energy detection for local spectrum sensing.

2.1.3 Network architectures

According to the differences between the sensing nodes interacting with the information in CRNs, the network architectures usually can be classified as centralized, distributed,

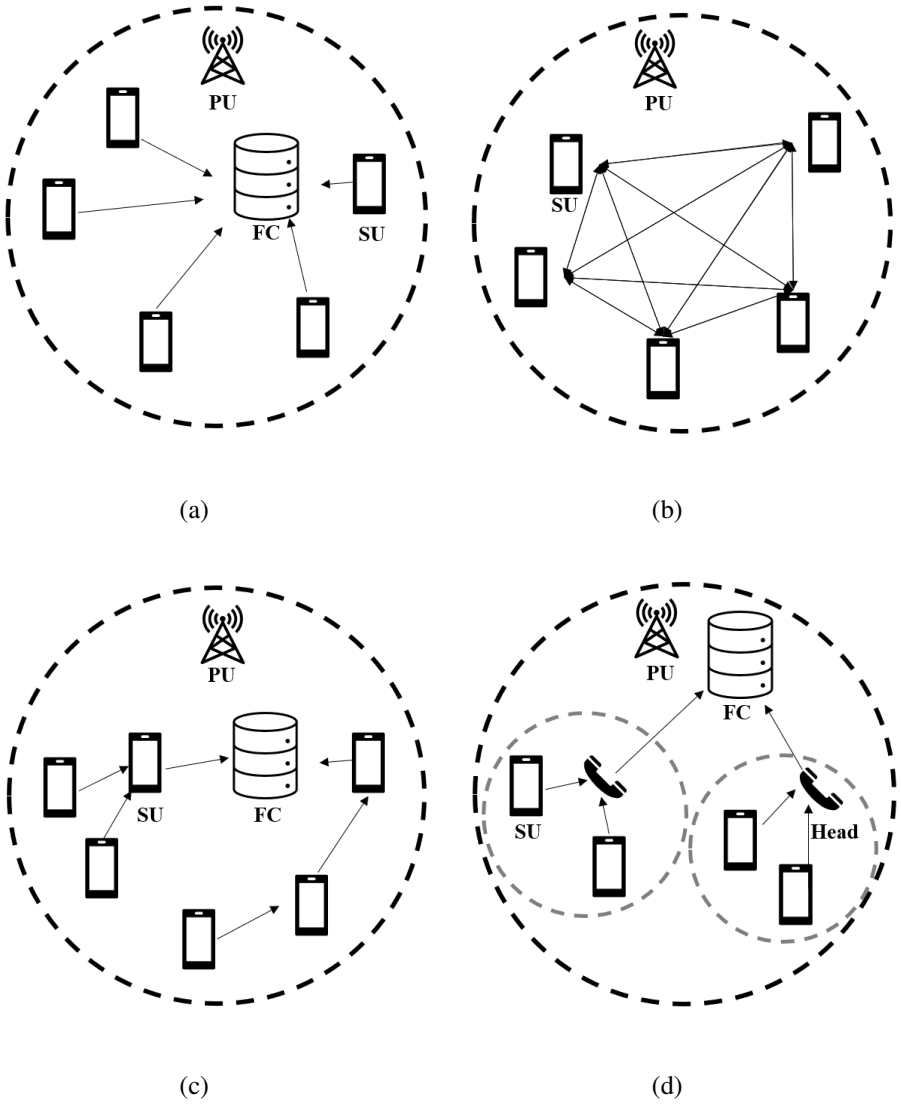


Figure 2.3 Collaborative spectrum sensing mechanisms: (a) Centralized. (b) Distributed. (c) Relay-assisted. (d) Clustering.

relay-assisted, and clustering networks as Fig.2.3 shows. The centralized networks represent the cognitive radio terminals that upload all sensed information to the database, and the database collects and merges all the received information to make the final global decision, the schematic diagram is as Fig.2.3(a). For the distributed networks as Fig.2.3(b), they do not have a control center or coordination process, all the sensing nodes have the same weight in the networks, after sensing they exchange information with each other, and make a global decision together. In a relay-assisted CRN as Fig.2.3(c), the sensing nodes communicate indirectly, through one or more relay nodes, instead of communicating with the database

directly. The relay nodes can be either dedicated or opportunistic, depending on whether they are specifically deployed for relaying or opportunistically take advantage of their location and sensing capabilities to assist in the transmission. Considering a large number of sensing nodes in the CRN, the clustering network like Fig.2.3(d), divides all the sensing nodes into several clusters, and each cluster has a head node that collects and processes the data reported from other nodes in the same cluster (the process of combining data within each cluster is equivalent to centralized network), and then reports the processed results to the center for a final decision. In this dissertation, we choose centralized network architecture, which means combining sensing results from multiple cognitive radio nodes in one fusion center (database), with the key objective of making accurate global decisions.

2.2 Database Model

In this section, we introduce a novel concept of a spectrum database that can be used to achieve fully efficient spectrum sharing. The database for a REM can be used by cognitive radio devices, for example, mobile phones or vehicles, to make decisions about how to adjust their transmission parameters to optimize their performance while avoiding interference to other devices.

The REM database can be created by combining the sensing and modeling techniques together. The database can be generated by collecting data about the RF environment from the cognitive radio devices and the estimated performance of the un-sensed locations.

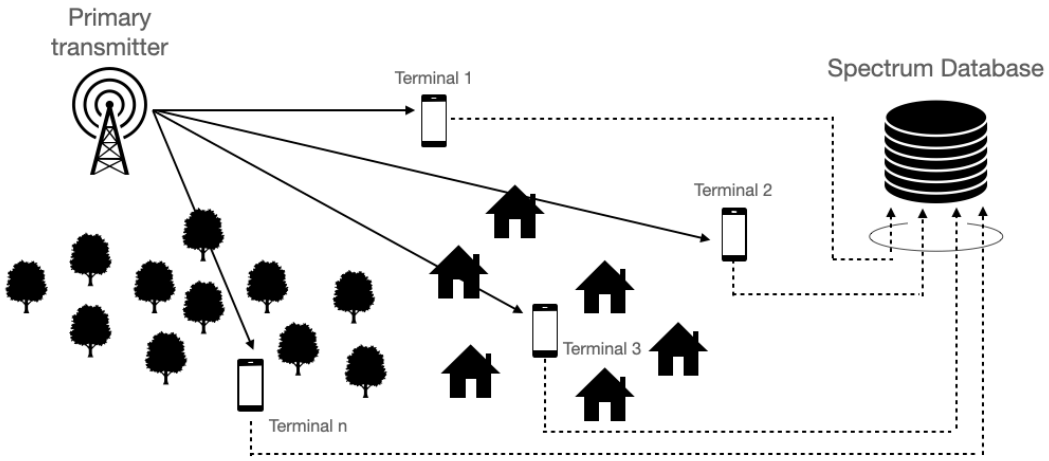


Figure 2.4 A concept of the centralized spectrum database.

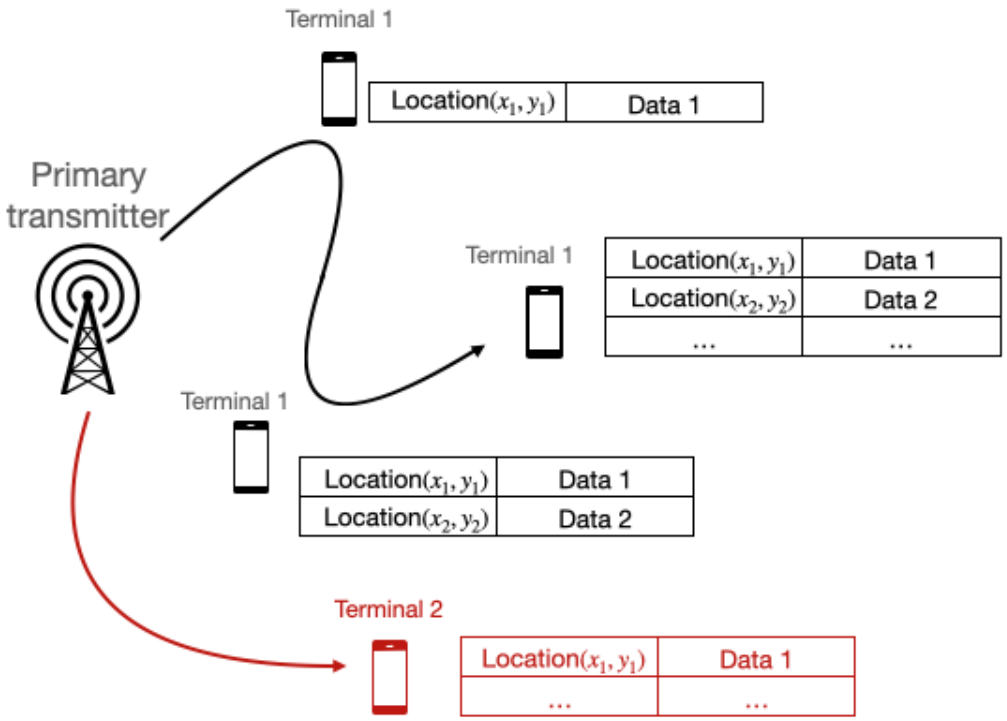


Figure 2.5 A concept of the distributed spectrum database.

The database can be either centralized or distributed, depending on the particular implementation. In a centralized REM database, all cognitive radio devices connect to a central server to report and access the data, as Fig.2.4 shows. In contrast, each device in a distributed database keeps a copy of the database and occasionally updates it using data from its own local sensors as Fig.2.5 shows, each terminal has the information from the database, and then will update its own datasets based on their own movements and measurements.

In this research, we use the centralized spectrum database in Fig.2.4. The terminals collect data by measuring the received signals from primary users. The collected data is location-specific and reported to the database, the reported dataset can be Table 2.1. Once enough data is gathered, the database estimates the radio environment characteristics of the PUs through statistical processing using the large dataset. And then the terminals can connect to the database via a wireless access point, such as a cellular network, to learn the radio environment around them and adjust their communication parameters, in order to avoid disturbing the primary terminals' communications.

Table 2.1 Dataset information.

Item	Type	Size(Byte)	Remarks
Measurement date	datetime	8	YYYY/MM/DD hh:mm:ss
Transmission latitude	double	8	Transmission latitude (°)
Transmission longitude	double	8	Transmission longitude (°)
Reception latitude	double	8	Reception latitude (°)
Reception longitude	double	8	Reception longitude (°)
Center frequency	double	8	Center frequency (Hz)
Received Signal Strength Indicator (RSSI)	double	8	RSSI (mW)
Packet ID	integer	4	-
Transmitter ID	char	17	-
Receiver ID	char	17	-
[1]Transmitter mesh code (10m)	char	2	XXXX-XX-XX-XX-XX
Transmitter mesh code (5m)	char	2	XXXX-XX-XX-XX-XX-XX
Receiver mesh code (10m)	char	2	XXXX-XX-XX-XX-XX
Receiver mesh code (5m)	char	2	XXXX-XX-XX-XX-XX-XX
Saved date	datetime	8	YYYY/MM/DD hh:mm:ss

[1] mesh code will be introduced in Section 2.3.3

2.3 REM Model

REM is a digital representation method which can provide the information including the RF information at different locations and so on. It can be used in many different fields including resource allocation, interference analysis, location estimation, optimization, etc.

2.3.1 Crowdsourcing

To determine the average values throughout the creation of the radio map, instantaneous received signal power samples are necessary. The strong observation capability of the spectrum analyzer makes it possible to precisely observe the instantaneous received signal power. However, the analyzer always has a high cost, which means that it cannot be common used in reality. In order to collect the received signal power samples at a cheap production cost, a new measurement tool is needed.

In order to create a REM, one method known as "crowdsourcing" involves gathering wireless signal data from a significant number of mobile terminals or other wireless sensors placed throughout a certain area. For example, the smartphones. Due to their lower cost and wide used, many researchers suggested to use them for measurement. This method makes use of the common mobile devices to gather data and offers an affordable solution to

get a complete understanding of the radio environment. The gathered information is then combined and examined to create a REM that can be applied to various wireless applications. Crowdsourcing enables the effective gathering of data from numerous sources across a big geographic area. [26][56]

2.3.2 Data collecting

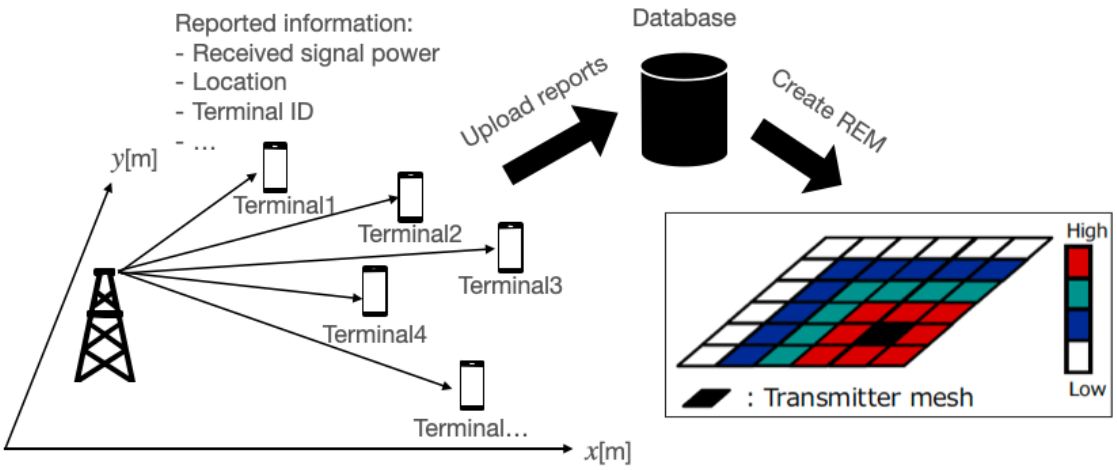


Figure 2.6 A concept of the conventional REM.

As crowdsourcing method we mentioned above, when the fixed transmitter sends a signal in the interested area, the mobile terminals, like smartphones, collect data and store these data at their own memory first. After a certain time, like one day, each terminal reports the stored dataset to the database. The database can receive massive data in a short time and then merge all the reported dataset on the server, the database statistically processed these datasets, and generate the REM. In this dissertation, we also consider when the mobile terminals do not have enough memory, they report their received dataset more often, we discuss the method for database construct REM based on such insufficient dataset in Chapter 4 and 5.

A concept of the conventional REM is shown in Fig.2.6. Several terminals are used to collect the spectrum information in the communication area. The terminals' location can be obtained from Global Positioning System (GPS). Terminals can collect the information including the terminal's ID, location, time, frequency and power send to the database. The database can be installed in the cloud or a base station which can store massive data and then base on these datasets generate the REM.

2.3.3 Mesh structure

The average received signal strength might not be precisely understood because the instantaneous received signal power includes a multipath fading component. An evenly split geographic area is referred to as a mesh, or some researchers also call it a grid. To counteract the fading effects, the database first computes a mesh based on latitude and longitude. The average received signal power is then calculated from each mesh's instantaneous received signal power samples.

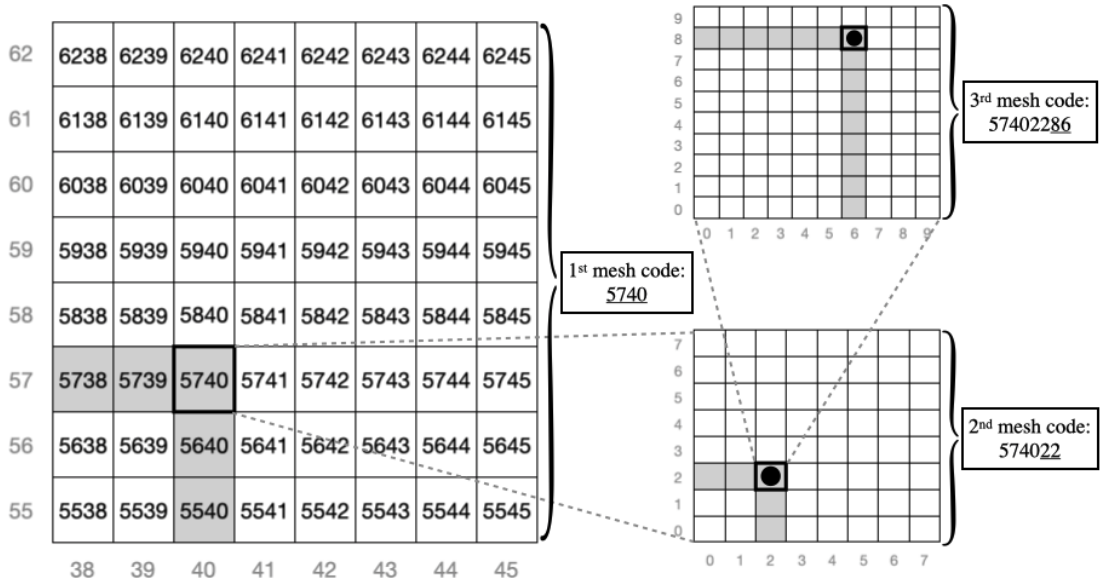


Figure 2.7 Schematic representation of mesh codes based on JIS X0410.

We can use the codes to distinguish each mesh square as a distinct location based on latitude and longitude. The codes are also helpful for mesh square statistics data analysis. Mesh codes are depicted schematically in Fig.2.7. Grid square codes based on JIS X0410 [57], often known as Japanese Industrial Standard (JIS), are utilized. Grid square statistics generation processes for both government of Japan statistics and industry applications are defined by JIS X0410. In Japan, some statistical surveys, such as censuses, economic surveys, and censuses, national land numeric information, facilities, natural environment and land usage, etc. are provided by the Statistics Bureau, the Ministry of Internal Affairs and Communications, the Ministry of Land, Infrastructure, Transport and Tourism [58].

As the JIS X0410, the codes are represented by numerical digits whose lengths are determined by the geographic resolution, and their notation is "puqvrw". The "1st mesh code" for "pu" is four numerical digits and an 80 km mesh square. Additionally, "puqv" is known as "2nd mesh code," which divides "1st mesh code" into 8 equal parts of length and

breadth (10km mesh square code) and is expressed by six numerical digits. Additionally, "puqvrw" is known as "3rd mesh code," which divides "2nd mesh code" by 10 equally (1km mesh square code) and is represented by eight numeric digits. The "Standard Area Grid" is the common name for the 3rd mesh code [59]. Additionally, the conversion process is indicated as follows,

$$p = \lfloor \text{latitude} \times 60/40 \rfloor (\text{p is two digits}), \tag{2.4}$$

$$u = \lfloor \text{longtitude} - 100 \rfloor, \tag{2.5}$$

$$\begin{cases} a = (\text{latitude} \times 60/40 - \text{p}) \times 40 \\ q = \lfloor a/5 \rfloor (q \text{ is a one digit}) \end{cases}, \tag{2.6}$$

$$\begin{cases} f = \text{longitude} - 100 - \text{u} \\ v = \lfloor f \times 8 \rfloor (v \text{ is a one digit}) \end{cases}, \tag{2.7}$$

$$\begin{cases} b = (a/5 - q) \times 5 \\ r = \lfloor b \times 2 \rfloor (r \text{ is a one digit}) \end{cases}, \tag{2.8}$$

$$\begin{cases} g = (f \times 8 - v) \\ w = \lfloor g \times 80 \rfloor (w \text{ is a one digit}) \end{cases}, \tag{2.9}$$

where, $\lfloor \rfloor$ denotes the floor function. The mesh size can be set as 10m, 5m, 2m, or 1m depends on the required accuracy [60].

2.3.4 Data processing

The database initially gives a mesh code to the received dataset based on the location information. The database manager determined the mesh size throughout this procedure. The average received signal power is then calculated for each divided mesh using the formula below,

$$\bar{P}_m = \frac{1}{N_m} \sum_{i=0}^{N_m - 1} P_{m,i} \quad [\text{mW}], \tag{2.10}$$

where, $\bar{P}_m$ denotes the average received power in the m-th mesh, $P_{m,i}$ [mW] indicates the i-th data information in the m-th mesh. N_m represents the number of received information in the m-th mesh.

2.4 Sensing Security Issues

Compared to traditional wireless networks, cognitive wireless networks are different in their mode of operation and network architecture, therefore face unique security issues. Due to the difference in priority between primary terminals and secondary terminals for spectrum resource utilization, malicious attack nodes are able to gain access to launch attacks on dynamic spectrum access links during the spectrum sensing part.

2.4.1 Attacking classification

Although CRNs hold the potential to enhance spectrum utilization, increase network capacity, and reduce interference, they are not immune to security threats and attacks that can compromise their functionality and performance. Such attacks can target various aspects of CRNs, including spectrum sensing, allocation, access, sharing, and mobility. These attacks can significantly impact the network performance and pose substantial challenges to the design and deployment of secure CRNs. Figure 2.8 illustrates the different classes of attacks in CRNs. This dissertation specifically focuses on data falsification attacks on spectrum sensing, aiming to find effective ways to counter their impact on cognitive radio accuracy.

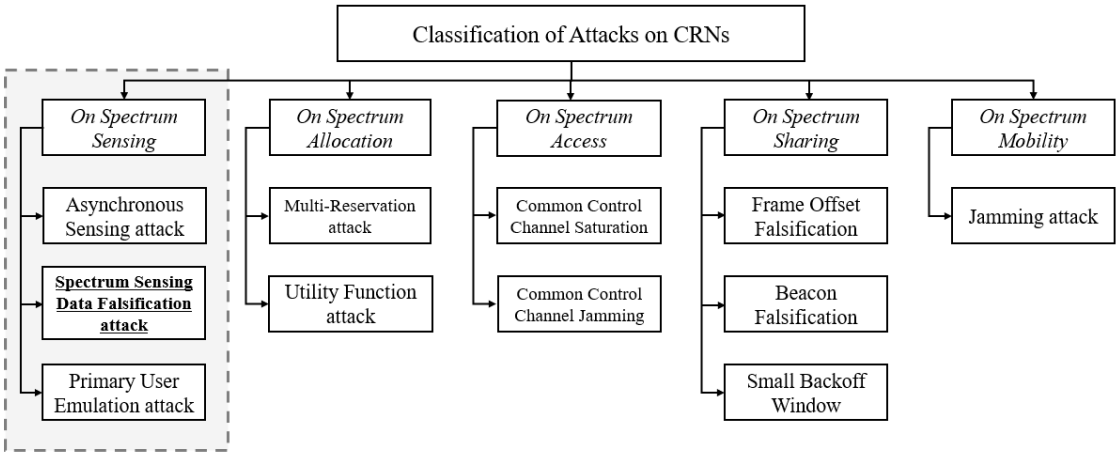


Figure 2.8 Diagram of attacks in CRNs.

Above all attacking methods, SSDF can have a serious impact on the CRNs through malicious users tampering with sensing information, thereby blinding the fusion center and causing the database to make incorrect global decisions. With their low-cost and high-reward attacking modes, as well as their flexibility and diversity of attacking methods, the SSDF attacks have become a major component of cognitive radio network attacks. By simply tampering with sensing information, global decisions can be influenced, which in turn can have a deleterious effect on the cognitive network. Hence, it is a big challenge for the effective defense against SSDF attacks. In our research, we investigate different algorithms to resist SSDF attacks in order to maximize the accuracy of the global results.

2.4.2 Time and spatial domain attack

Time domain approaches primarily aim to comprehend the temporal attributes of signals, encompassing their duration, timing, and temporal fluctuations. Time-domain spectrum sensing methods analyze the amplitude and timing characteristics of signals to ascertain their existence or absence within a specific frequency range. Energy detection, cyclostationary feature detection, and matched filter detection all belong to the time domain sensing.

A time domain attack in spectrum sensing involves manipulating the temporal characteristics of signals to deceive or disrupt the spectrum sensing process. Attackers may use techniques such as jamming, altering signal timings, or introducing false temporal patterns to mislead cognitive radios in their perception of spectrum occupancy. SSDF attack which we will introduce in Section 4.2.3 falls into the time domain attack.

In contrast, the spatial domain in spectrum sensing refers to the use of multiple antennas or spatial information to analyze signals. Spatial domain techniques involve examining the spatial characteristics of signals, including their direction of arrival, and interference patterns.

A spatial domain attack in spectrum sensing could involve manipulating the spatial characteristics of signals to deceive or disrupt the sensing process. For example, an attacker might deploy directional jammers to interfere with signals in specific spatial directions or attempt to create false spatial patterns to mislead the sensing system.

2.4.3 Data falsification attack

Cognitive radio is a technology that allows wireless communication devices to access the idle radio spectrum dynamically, which is an effective way to improve spectrum utilization. However, like any wireless communication system, cognitive radio networks are easy to be attacked.

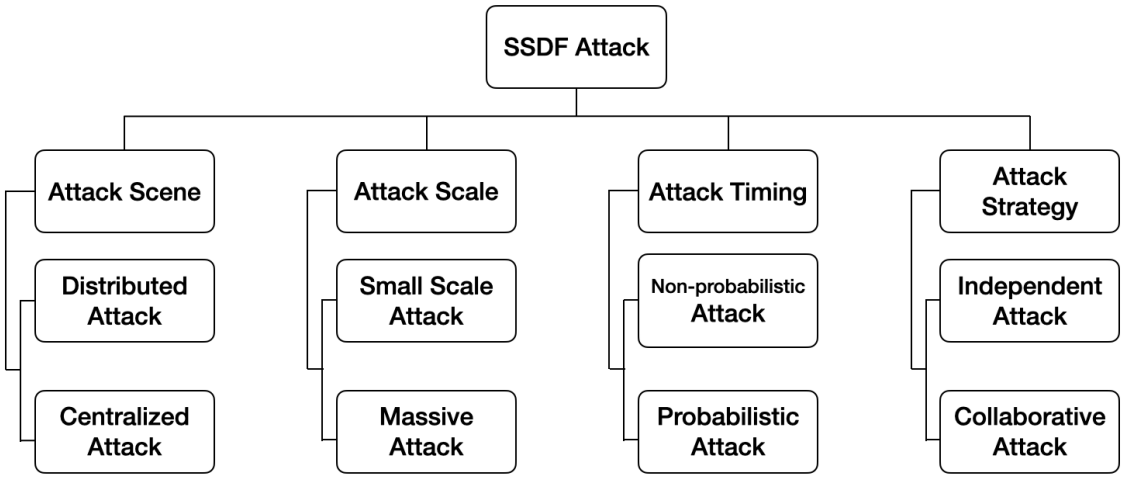


Figure 2.9 Diagram of SSDF attack.

A data falsification attack is also called a Byzantine attack, which can cause massive disruption to cognitive networks, based on the fact that CRNs rely on accurate sensing of the radio spectrum to identify the spectrum conditions, to avoid interference with the licensed users. SSDF attack can be carried out by one or a set of attackers who has the ability to transmit signals on the same frequency bands as the cognitive radio network, they launch false information to the database and can cause the database to make the wrong decisions about the spectrum conditions. These attackers intentionally send error-aware reports to the database in order to fulfill some of their selfish requirements, for example, to disturb the licensed users' communication or to occupy the idle band by themselves. The attackers, also called malicious terminals, have some strategies to send the wrong information.

- Always Idle: The malicious terminals always send a lower report than the real received signal power in a soft decision network. The same can be known, they always send Idle condition to the database if they are in a hard decision network. By using this attacking strategy, the malicious terminal even do not need to sense the channel, and their goal to attack network is to lead a big interference to the primary terminals, as long as they blind the database, they do not care about the real channel conditions.

- Always Busy: In contrast, the malicious terminals always send a higher report than the real sensed power in order to blind the database as the channel is busy in every sensing slot if they are in a soft decision network. Or they always send 1 to the database if they are in a hard decision network. Different with the Always Idle attack, they should sense the band although the malicious terminals always report the high power to the database. Their goal to launch a Always Busy attack is to blind the database to make a wrong

decision, and reduce the usage of the channel, then they can use the idle channel by themselves.

- Always Opposite: In this attacking strategy, the malicious terminals need to sense the channel and based on their real decision, rewrite their information and send to the database. In this case, the reports from them are always wrong. Such strategy is highly disruptive, but the malicious terminals are also relatively easy to be detected by the system due to their constant error reporting.

- Hybrid attacking: Hybrid attacking is the attack method terminals combined different strategies together to protect themselves and attack the network. In this strategy, some slots the malicious terminals send the correct information and some slots they choose to attack. When they choose to attack, the attacking strength also can be different. Since attacks by malicious users are variable, it is difficult for the system to detect them.

In our research, we assume that there are malicious terminals present in the communication environment. These terminals can compromise the integrity of the REM by manipulating the sensing power of the spectrum and submitting false power information to the database. Such erroneous data can result in significant errors in the REM, which could potentially serve the selfish interests of the malicious terminals.

Here, we present the hybrid attack model as follows: malicious terminals alter their data by comparing it to the κ power threshold. If the sensing power of the malicious terminal exceeds the threshold in each sensing position, the sensing power is rewritten by multiplying the attack index with the probability P_a and the incorrect dataset is reported to the database. If not, malicious terminals will transmit the correct dataset to the database. Following is the reported strength of the malicious terminal:

$$P'(s_i) = \begin{cases} P(s_i) \cdot \delta, & \text{if } P(s_i) > \kappa \text{ with } P_a \\ P(s_i) \cdot \delta, & \text{if } P(s_i) < \kappa \text{ with } P_b \\ P(s_i), & \text{otherwise} \end{cases}, \tag{2.11}$$

where δ represents the attacking index, and P_a and P_b represents the attack probability over or lower than the threshold, respectively. If $\delta > 1$, the attacking strength increases with the attacking index, if the attacking index $0 < \delta < 1$, the attacking strength decreases as the attacking index increases.

Fig. 2.10 shows the normalized Mean of Absolute value of Errors (MAE) and the Root of the Mean of the Square of Errors (RMSE) under different attacking index $\delta \in [0.3, 1.6]$,

also we set the number of malicious terminals increased from 0 to 90 out of 100 in total. In the figure, the legend means the MAE or RMSE under different attacking indexes, for example, the first legend means the MAE curve when the attacking index is equal to 0.3. The figure shows that when the attacking index δ is approached to 1, the MAE and RMSE decreased, otherwise the error increased. Additionally, when the number of malicious terminals increases, the error also increases.

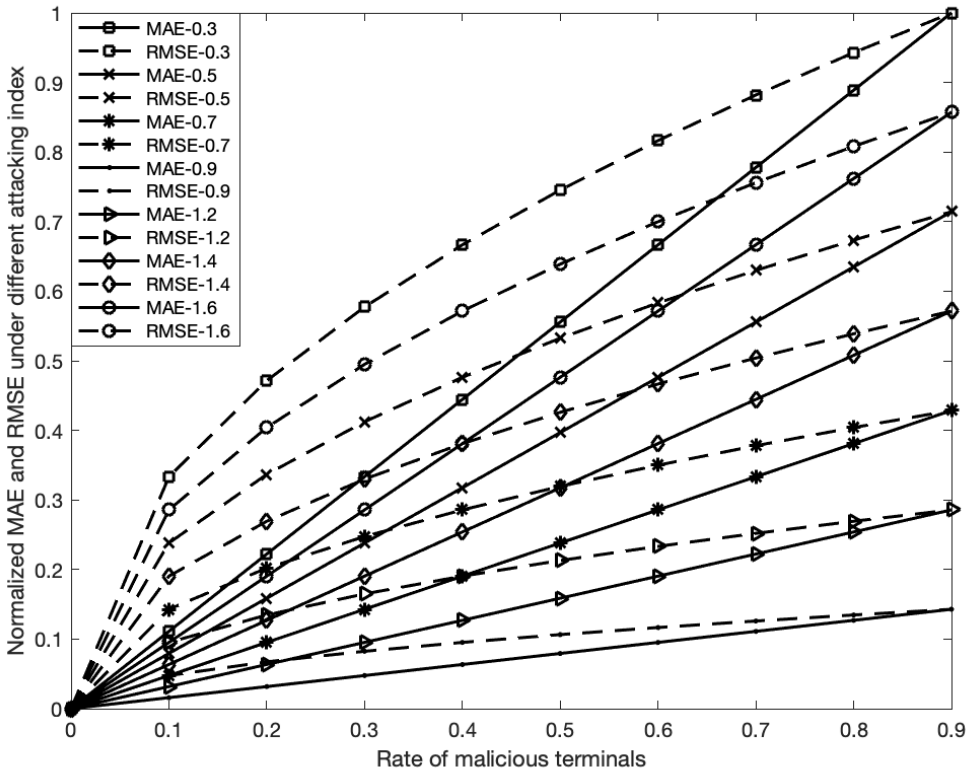


Figure 2.10 Effect on data falsification attack.

2.5 Chapter Summary

In this chapter, we introduced various models. The chapter begins with the introduction of spectrum sensing models and techniques, which involve network architectures usually used. Additionally, the database model, which is used to manage spectrum usage information, is introduced. Lastly, the REM model, which is used to simulate the radio propagation and interference effects in the cognitive radio network, is discussed. Finally, the chapter also

addresses the sensing security issues that arise due to malicious terminals attempting to disrupt the sensing process or occupy the idle channel by themselves.

Chapter 3

An Improvement of Security Scheme for REM Construction under Massive Attacking

Radio Environment Map is a popular method to estimate the communication area. However, the open characteristic of the wireless communication network always faces some threats. To construct the radio map, usually we split the map into several small meshes, the terminals which are located in the same mesh make the contribution on estimate the average power of the mesh. Since it is very difficult to get the ground truth information, in this chapter, we use the average power as the ground truth. However, although several terminals work together to construct the part of the mesh, the database is still easy to be cheated by the malicious terminals cheating.

The arrangement of chapters as follows: Section 3.1 introduces the background of our study, section 3.2 explains the system models, and the implementation framework of our proposed method is discussed in section 3.3. The attacking strategies in this chapter is shown as section 3.4. Finally, the simulation results are shown at section 3.5, and finally, the chapter summary is in section 3.6.

3.1 Background

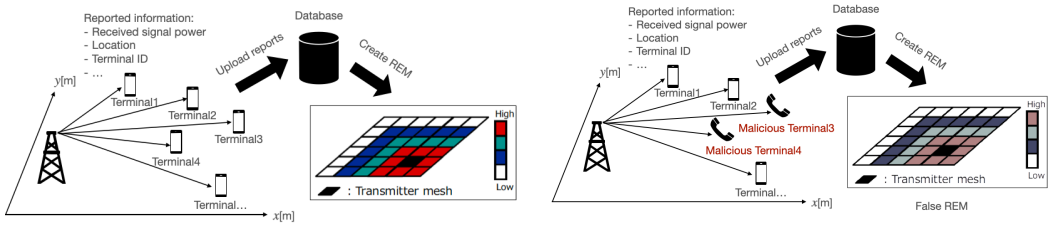
The REM can be used to manage inter-transmitter interference by storing and analyzing extensive information, such as the mean received signal power within the communication area. By analyzing the measurement datasets, an estimate of the average received signal power can also be obtained. In the case of TVWS systems, the REM, which is stored in

the spectrum database, is used by SUs to identify available white spaces and allowable interference power. According to [20], an effective spectrum-sharing system can be achieved through the utilization of REM.

The REM generation process necessitates the utilization of environmental information sourced from a database. Terminals are positioned at random within the communication area and record several details including received signal power, terminal ID, and location, among others. To construct REM, a sufficient dataset from the communication area is necessary. Hence, the precision of the terminal-reported information is vital for REM, and the accuracy of REM construction serves as a crucial metric that directly impacts spectral efficiency.

In [23], REM was created without taking into account its shadowing impact. Reference [61] modeled spatial spectrum sharing over log-normal channels, and the results indicated that the Kriging-aided method can simplify the model. In [62], a neural network was utilized to enhance accuracy for both fixed and distributed transmitter systems. REM can be generated using experimental measurement datasets. References [63][64] constructed a model for a distributed transmitter system, analyzed the frequency correlation of shadowing, and modeled the V2V communication environment using measurements.

One approach for producing REMs involves utilizing data from terminals within the communication area. The accuracy of these REMs relies heavily on the quality of the received data. However, if there are malicious terminals present in the area, they can manipulate the data and upload incorrect information to the database, attempting to deceive the system for their own benefit. As reported in [45, 65, 66], known as data falsification attacks, these types of attacks are common in the spectrum sensing field, and there are various examples of them, including always present, always absent, and always opposite attacks. Malicious terminals may also work independently or collaboratively to perform attacks [45]. Various methods have been proposed to address these attacks, such as using optimal likelihood ratio tests [67] or a penalty-based Dempster-Shafer theory of evidence to deal with uncertainty representation [68–70]. However, when the number of datasets is limited, especially after removing malicious terminals, the REMs may not be accurate enough. To enhance the precision of the REMs, researchers have demonstrated that shadowing in the communication area exhibits spatial correlation, and the interpolation method can be employed to estimate unknown points' data [71–73].



(a) Without malicious terminals (True Map) (b) With malicious terminals (Fake Map)

Figure 3.1 A concept of the conventional REM.

3.2 System Description

3.2.1 REM model

To create a REM for the communication area, we assume the utilization of multiple mobile terminals for gathering spectrum information. These terminals are randomly placed based on their movements. As stated in [61], they can gather information and transmit it to a database, which may be located in the cloud or base station, capable of storing substantial amounts of data. Once the database obtains adequate information from the terminals, it can construct a REM. Our research only considers information such as the terminal ID, location, and received signal power. The received signal power is used to construct the REM since mobile terminals can quickly obtain it. For example, a smartphone can use the API of the Android OS to obtain the received signal power of the cellular system and WLAN. Hence, the measurement cost of received power is low. Fig.3.1 depicts the conventional REM concept.

In order to mitigate the effects of small-scale fading, we divide the communication area into several two-dimensional meshes. We compute the average power of each mesh using data collected from within it. When the mesh size is small enough, shadowing effects can be disregarded and the accuracy of the REM can be enhanced. Nonetheless, in cases where there are malicious terminals present in the communication area, the REM's accuracy can be compromised by the transmission of malicious information, which means the constructed map may have big differences between the real condition due to the effect caused by the malicious terminals. To address this issue, it is necessary to develop an algorithm that can discern between valid and malicious information from such terminals, as stated in [74].

3.2.2 Attacking model

In the communication environment, we take into account the presence of malicious terminals that may try to attack the database. Their goal is to satisfy their own interests, such as interfering with primary terminal frequency bands or monopolizing free frequency bands. To construct REM, it is necessary to collect the received signal power from mobile terminals distributed in space and calculate the average power of each mesh. However, the accuracy of REM can be compromised by an efficiency attack known as data falsification, which can cause damage to the system.

Data falsification attacks, also known as Byzantine attacks, involve terminals that maliciously alter the data they sense from the spectrum in order to deceive the database. Previous research, such as [65], has presented examples of this type of attack. In this study, we assume malicious terminals modify their data by comparing it with a power threshold κ. During each sensing slot, if the sensing power of the malicious terminal surpasses the threshold, they rewrite the sensing power by multiplying a specific factor δ (known as the attacking index) with the probability P_a, and report the false data to the database. Conversely, if the sensing power is below the threshold, the malicious terminals report the correct dataset to the database. The reported power of a malicious terminal is calculated as follows,

$$P'(i) = \begin{cases} P(i) \cdot \delta, & \text{if } P(i) > \kappa \text{ with } P_a \\ P(i), & \text{otherwise} \end{cases}, \tag{3.1}$$

If $\delta > 1$, the attacking strength increases with the attacking index, if the attacking index is $0 < \delta < 1$, the attacking strength decreases as the attacking index increases. The concept of REM damage from a data falsification attack is shown in Fig.3.1(b).

3.3 Sensing Scheme Against Data Falsification Attack

This section considers the scenario where multiple terminals move randomly within a communication area, and they transmit information directly to the database, or after rewriting it. The database accumulates data over multiple time slots, resulting in a large amount of stored data. To improve the performance of the REM, a technique called Double-Layer Monitor mechanism is proposed, which identifies and eliminates malicious terminals and their corresponding data from the database.

3.3.1 Double-layer monitor (DLM)

Numerous security issues plague wireless communication networks with an open characteristic, the Double-Layer Monitor technique is proposed as a way to improve the estimation accuracy of REM. We configure the weight allocation component to enhance the performance of malicious terminal evaluations. The DLM consists of two layers: the similarity comparison method and the sustainable monitor.

3.3.1.1 Similarity comparison (first layer)

In order to detect malicious datasets within the database, we examine the possibility of determining their similarity in the initial layer. This involves computing the similarity degree between every pair of datasets located within the same mesh. If the mesh is of a small enough size, the shadowing index can be regarded as uniform throughout the mesh. This means that terminals within the same mesh will only experience differences in path loss and fading. If the mesh size is small, the path loss values are also similar, therefore, it is reasonable to use the similarity degree as an indicator to identify malicious datasets.

During their movements, the malicious terminals carry out attacks on the data by rewriting the information they receive and sending false information to the database. As a result, even when located in the same place, the data from the malicious terminals differ from those of the honest terminals, which would normally differ due to fading. We determine the similarity between pairs of data and compare them. If the terminals exhibit a high degree of similarity, we conclude that they are likely to be honest terminals. Conversely, if the similarity degrees are low, we consider them to be malicious.

The equation for calculating the similarity degree is calculated as follows,

$$sim(i,j) = \frac{max(P'(i), P'(j))}{\frac{1}{2}(P'(i) + P'(j))}, \tag{3.2}$$

where i and j indicate different reports in one mesh. The following similarity matrix can be constructed for the received signal strength across the entire mesh.

$$Sim = \begin{bmatrix} 1 & \cdots & sim(1,j) & \cdots & sim(1,n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ sim(i,1) & \cdots & 1 & \cdots & sim(i,n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ sim(n,1) & \cdots & sim(n,j) & \cdots & 1 \end{bmatrix}, \tag{3.3}$$

then, the support level of each data can be calculated as follows,

$$Sup(i) = \sum_{j=1}^{n} sim(i,j) \quad j \neq i \quad i,j = 1 \cdots n, \tag{3.4}$$

consequently, the terminal i's reliability can be achieved by normalizing the support as follows:

$$Rel(i) = \frac{Sup(i)}{max(Sup(1), Sup(2), \cdots, Sup(n))}. \tag{3.5}$$

The similarity comparison method is like Algorithm 1 represented below, where $m = 1, 2, 3, ..., D$ represents the mesh number, and we presume there are a total of D meshes separated from the communication area. $i = 1, 2, 3, ..., N$ represents the number of reports, and we presume that each mesh contains N reports. $\mathcal{HT}$ and $\mathcal{MT}$ denote, respectively, the set names of honest terminals and malicious terminals. α represents the similarity threshold; if $Rel(i)$ is greater than α, the terminal is deemed trustworthy; otherwise, it is malicious.

In cases where malicious datasets make up a significant portion of the datasets within a mesh, relying solely on similarity degree to determine reliability may not be effective in removing them. Additionally, if malicious terminals engage in dynamic attacks by changing their attack patterns, constantly resetting the threshold during their movement may not be a practical solution. To tackle this issue, a sustainable monitoring approach is being considered.

3.3.1.2 Sustainable monitor (second layer)

As noted previously, depending on only a similarity comparison is insufficient to eliminate malicious data from the power fusion. In order to address this issue, we explored the possibility of utilizing a sustainable monitor to improve performance. A sustainable monitor employs terminal ID information and monitors the terminal's behavior as it transitions to another mesh network. This enables ongoing evaluation of reliability, and if a terminal's behavior is consistently flagged as malicious, all data associated with that terminal can be removed from the database.

Real-step confidence When comparing the reported power to the average power after power fusion in each mesh, the difference is known as the real-step confidence. We regard the terminal to be confident if the difference is small; otherwise, we recommend lowering confidence. The formula for the $Bias$ confidence index is as follows:

Algorithm 1 Similarity Comparison

Require:

The parameters related to the Database, such as $P'(i)$, $\mathcal{HT}$, $\mathcal{MT}$, α, etc.

Ensure:

for Each mesh $m \in D$ **do**
  Initialize $\mathcal{MT} = \phi$,$\mathcal{HT} = N$,$P_{all} = 0$
  for Each reported power in each mesh $i \in N$ **do**
    Calculate the similarity degree
    Generate the similarity matrix
    Calculate the $Sup(i)$ for each dataset
    Normalization get $Rel(i)$
    if $Rel(i) \geq \alpha$ **then**
      $P_{all} \leftarrow P_{all} + P'(i)$
    else
      $\mathcal{MT} \leftarrow \mathcal{MT} + \{i\}$
      $\mathcal{HT} \leftarrow \mathcal{HT} - \{i\}$
    end if
  end for
  Calculate the average power $\overline{P}_m = \frac{P_{all}}{|\mathcal{HT}|}$
end for
Generate the REM

$$Bias(i) = |P'(i) - \overline{P}|, \tag{3.6}$$

where $Bias(i)$ represents the absolute value of the difference between the reported power from i-th terminal and the mesh average power.

Historical reliability Historical reliability records are gradually updated. When the terminal's historical reliability is less than the threshold, it is considered malicious, and all data associated with that ID are eliminated from the power fusion.

$$HisRe_i^H = l * HisRe_i^{H-1} + (-1)^{para} * \mathrm{Re}(\cdot), \tag{3.7}$$

here, $HisRe_i^H$ denotes the i-th terminal's historical record of reliability at the $H(H = 2, 3, 4, ...)$ step.The influence of historical reliability may be thought of as impact factor l pair, with a bigger value of l having a greater effect. The reward-penalty parameter, $para = 0, 1$, indicates that the historical reliability should increase or decrease. When $Bias < \zeta$, $para = 0$, the historical reliability increases. And decrease when $para = 1$. Here, ζ is the real-step confidence threshold. By resetting the reward-penalty function $\mathrm{Re}(\cdot)$, we are able to get a

more accurate historical computation compared to our earlier study [74]. The $Bias$-related fitting function is denoted by Re$(\cdot)$. It is expected that as $Bias$ climbs, $Re(\cdot)$ will decline slowly, but when $Bias \geq \zeta$, Re$(\cdot)$ will increase quickly.

Weight allocation Once we have computed the historical reliability, we modify the weight assigned to each terminal accordingly. Terminals that exhibit greater reliability are given higher weights, while those that do not are given lower weights. If a terminal's weight falls to zero, its reliability is also set to zero.

$$w\_HisRe_i^H = \frac{HisRe_i^H}{\max(HisRe^H)}, \tag{3.8}$$

here w_HisRe represents the weight, and $\max(HisRe^H)$ represents the maximum $HisRe$ for all the terminals in the same mesh.

The DLM is like Algorithm 2 illustrated below. The parameter β denotes the threshold for historical reliability, and H represents the step number of the terminal. By utilizing the terminal ID data, we can continuously monitor the performance of the terminals as they move, enabling us to evaluate their reliability constantly. This allows us to detect malicious terminals even in cases of dynamic attacks (when malicious terminals alter their attacking indexes while moving) or small-scale attacks. We can identify such malicious terminals step-by-step based on their historical reliability and exclude them from the database.

3.3.2 DLM with spatial information

As previously mentioned, a DLM has the ability to detect and remove datasets transmitted by malicious terminals from the power fusion. This can result in varying numbers of reports for each mesh, not only due to the presence of malicious terminals, but also because of the random movements of the terminals. The REM is generated by the database based on the reports received from the terminals. In cases where there are insufficient datasets from a particular mesh, the REM may deviate significantly from reality. By obtaining more reliable datasets in the database, the accuracy of the REM can be improved in this collaborative sensing network. Interpolation may be considered as a solution to address the issue of uneven distribution of information across the meshes.

3.3.2.1 DLM based on spatial correlation

The exponential decay model is a widely recognized representation of the spatial correlation of shadowing, and the correlation index can be defined as follows [75],

Algorithm 2 Double Layer Monitor

Require:

The parameters related to the Database, such as $P'(i)$, $\mathcal{HT}$, $\mathcal{MT}$, α, β, terminal ID, etc.

Ensure:

for Each reported power in each mesh $i \in N$ **do**
  Initialize $\mathcal{MT} = \phi$,$\mathcal{HT} = N$,$P_{all} = 0$,$HisRe_i^H = 0.1$
  for Each terminal's step H **do**
    Calculate the Similarity comparison get the $Rel(i)$
    if $Rel(i) > \alpha$ && $w\_HisRe_i^H \geq \beta$ **then**
      $P_{all} \leftarrow P_{all} + P'(i)$
    else
      $\mathcal{MT} \leftarrow \mathcal{MT} + \{i\}$
      $\mathcal{HT} \leftarrow \mathcal{HT} - \{i\}$
    end if
    Calculate the average power $\overline{P}$
    Calculate the Real-step confidence $Bias(i)$ and $para$
    Calculate the Historical reliability $HisRe_i^H$
    Do the weight allocation
  end for
  Update the $HisRe_i^H$ and w_HisRe_iH
  Move to next step $H \leftarrow H + 1$
end for
Generate the REM

$$\rho_{ij} = \frac{E[W(s_i)W(s_j)]}{\sigma^2} = \exp(-\frac{\Delta d_{ij}}{d_{\text{cor}}}\ln 2), \tag{3.9}$$

where Δd_{ij}[m] represents the distance between two different terminals i and j, and d_{cor}[m] is the correlation distance, defined as the location where $\rho_{ij} = 0.5$. In an urban area, the correlation distance was approximately 20 [m] according to an experiment [76].

When the correlation distance is held constant, as in Fig.3.2, a smaller distance between two points indicates a greater correlation. The cumulative distribution function (CDF) curves at various thresholds are shown in Fig.3.3. The significance level for the ρ index is denoted by θ. The observation reports may be utilized to estimate the information at the test locations only when $\rho > \theta$. The values shown in Fig.3.3 are tabulated in reference Table 3.1.

In order to address the issue of unequal information distribution across the meshes and enhance the accuracy of the REM, we utilize a DLM based on a spatial correlation algorithm, illustrated in Algorithm 3. By performing appropriate interpolation, we can gather adequate data for estimating the environment, thereby compensating for the lack of information caused

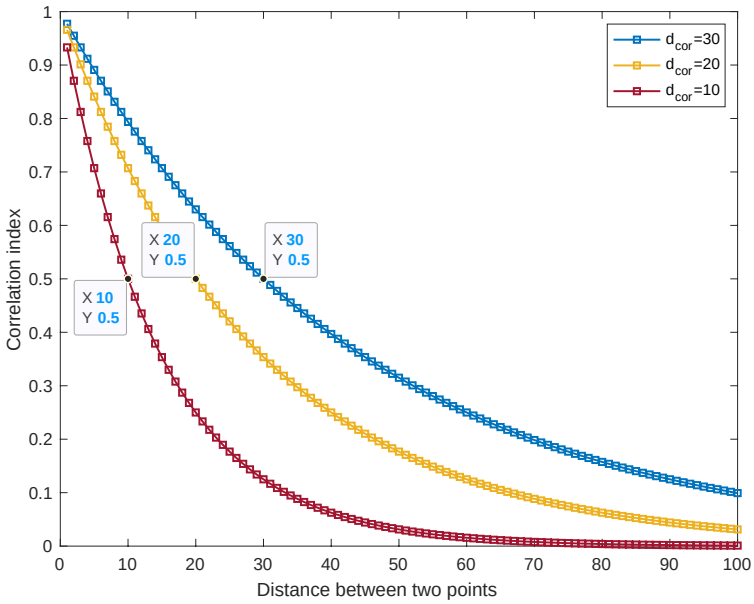


Figure 3.2 Relationship between distance [m] and the correlation index.

by random movements of the terminals and resolving the problem of information loss caused by malicious attacks on the terminals.

Table 3.1 Detail value of spatial correlation.

	MAE [dB]	μ	σ
No Algorithms	3.4799	-0.1443	0.8231
$\theta = 0.75$	3.4090	-0.1147	0.8122
$\theta = 0.85$	2.8055	-0.0260	0.7078
$\theta = 0.95$	2.3808	0.1338	0.5947

3.3.2.2 DLM based on inverse distance weighting (IDW)

When we can do interpolation based on a large number of known observation points, IDW is a deterministic approach may be used. Each value is derived using a weighted average of nearby known observation locations, with each observation point's weight determined by the inverse of the distance between itself and the unknown location.

We assume N observation sites, with coordinates (x_i, y_i), where $i = 1, 2, 3, ..., N$, are spread out uniformly throughout the region of interest. Since we are only concerned with a two-dimensional scenario here, we may interpret the x_i, y_i as representing the horizontal and

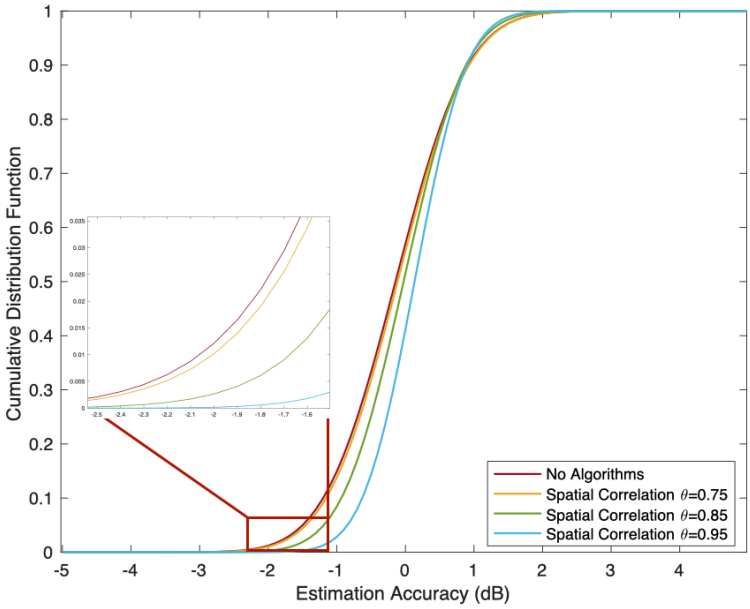


Figure 3.3 Cumulative distribution function of spatial correlation.

vertical distances, respectively, from the point of observation. Power data associated with coordinates is denoted by the notation $P(x_i, y_i)$. The ordinary IDW interpolation function can be expressed as follows,

$$\begin{cases} \widetilde{P}(x_0, y_0) = \sum_{i=1}^{N} w_i P(x_i, y_i), & \text{if } d_i \neq 0 \text{ for all } i \\ \widetilde{P}(x_0, y_0) = P(x_i, y_i), & \text{if } d_i = 0 \text{ for some } i \end{cases}, \tag{3.10}$$

where x_0, y_0 is the position of the interpolated point, $\widetilde{P}(x_0, y_0)$ is the power interpolated at the location, w_i is the given weight from each observation point, which can be represented as follows,

$$w_i = \frac{d_i^{-p}}{\sum_{i=1}^{N} d_i^{-p}}, \tag{3.11}$$

here, the distance between the i-th observation point and the interpolated point is denoted by d_i, and is defined as $d_i = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2}$. For a bigger value of the IDW power parameter, p (a positive real number), the nearest points have a greater impact on the interpolated point. Typically, [0.5,3] is the widely used range of p. [77] [78]

Because a larger value of p results in a greater degree of related on the points next to the one being interpolated, the performance of the interpolation improves as demonstrated in Fig.3.4. Sensing power is impacted by path loss, shadowing, and fading; path loss and

Algorithm 3 DLM based on Spatial Correlation

Require:

The parameters related to the Database, such as $P'(i)$, $\mathcal{HT}$, $\mathcal{MT}$, α, β, terminal ID, etc.

Ensure:

1: **for** Each reported power in each mesh $i \in N$ **do**
2: Do the Double Layer Monitor
3: Get $\mathcal{MT}$ and $\mathcal{HT}$
4: Calculate the amount of data which need to be interpolation
5: Generate the random location in the mesh
6: Interpolation by Spatial Correlation
7: **end for**
8: Generate the REM

Table 3.2 Detail value of IDW.

	MAE [dB]	μ	σ
No Algorithms	3.4799	-0.1443	0.8231
$p = 1$	2.3127	-0.0694	0.5760
$p = 2$	1.4599	-0.0155	0.4013
$p = 3$	1.1435	0.0077	0.3219

shadowing, in turn, are influenced by the placements of the terminals. Fading does not have an effect on the location. Therefore, if there is a shorter distance between two terminals, then the sensing power values of those terminals will be more comparable to one another, and the mistake caused by interpolation should be lower. The values that are shown in Fig.3.4 are presented in further detail in Table 3.2. As can be seen, an increased value for the IDW power parameter is associated with a decreased value for the inaccuracy in the interpolated point.

Interpolation is one method that may be used to address the issue of uneven information distribution throughout the meshes, as was discussed before. IDW interpolation is a technique that, like Algorithm 3, which makes use of knowledge of the geographical information, has the potential to address this issue and provide adequate data in order to create the REM. In addition, the estimated data that are created by IDW are more connected to the points that are closer, as a result of the addition of the weight that varies according to distance. This should result in improved performance when compared with the spatial correlation. Algorithm 4 provides a description of the DLM that is based on IDW.

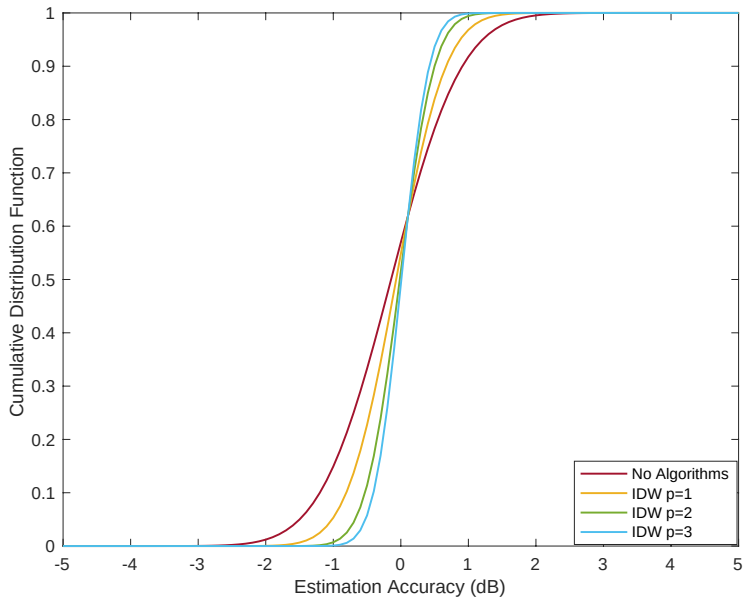


Figure 3.4 Cumulative distribution function of IDW.

3.4 Different Attack Strategies

The construction of REMs using regular algorithms is always at risk of malicious terminal attacks. These attacks can impact the REM in various ways. To provide a comprehensive evaluation of performance, this section introduces different attack strategies, categorized as black-box and white-box.

3.4.1 Black-box attack

A black-box attack refers to a scenario where the malicious terminals do not possess knowledge about the database. Although they may have a head of the attacker to make the decision of the attack strategy, they lack information about the algorithms and parameters used in the database setup. This limits their ability to plan the attack strategy.

3.4.1.1 Static attack and dynamic attack

Without loss of generality, we examine both static and dynamic attacks in our research. A static attack involves all the malicious terminals employing the same attack strategy. These terminals utilize a constant attacking index throughout the sensing process, without the ability to pause or modify the index. Conversely, a dynamic attack allows the malicious terminals to alter their attack strategy during the sensing process. They can choose to change the attacking

Algorithm 4 DLM based on IDW

Require:
The parameters related to the Database, such as $P'(i)$, $\mathcal{HT}$, $\mathcal{MT}$, α, β, terminal ID, etc.
Ensure:
1: **for** Each reported power in each mesh $i \in N$ **do**
2: Do the Double Layer Monitor
3: Get $\mathcal{MT}$ and $\mathcal{HT}$
4: Calculate the amount of data which need to be interpolation
5: Generate the random location in the mesh
6: Interpolation with Inverse Distance Weighting
7: **end for**
8: Generate the REM

index at any time, and may even abstain from attacking in certain sensing slots to evade detection by the database.

3.4.1.2 Independent attack and collaborative attack

An independent attack allows malicious terminals to operate autonomously, choosing the attacking index and attacking slot based on their individual preferences. This approach offers greater subjective initiative to the attackers and can increase the complexity of the attack system. Since attackers make decisions independently, the risk of being detected is significantly reduced. In contrast, a collaborative attack requires all malicious terminals to select the same attack strategy. While this type of attack may have a greater impact on the sensing system, it also has a higher risk of being detected compared to independent attacks.

3.4.2 White-box attack

In contrast to a black-box attack, a white-box attack is carried out by malicious terminals that have comprehensive knowledge of the security system of the whole network. Although this information may be used to perform the functions of a database, the terminals' primary objective is to raise the mistake rate of the REM in order to reap the advantages of using the incorrect map.

Malicious terminals may perform an attack that is most effective by utilizing the information provided by the security system. According to Equ.(3.1), as the attacking index $\delta(\delta \geq 1)$ grows, the attacking strength increases, the reliability $Rel(i)$ of malicious terminals falls, and the detection likelihood of malicious terminals increases. All of these results are reflected by the fact that the attacking index is increasing. On the other hand, as the attacking index $\delta(0 < \delta < 1)$ decreases, the attacking strength rises, the reliability $Rel(i)$ of malicious

terminals declines, and the detection likelihood of malicious terminals increases. This is because the attacking index is inversely proportional to the reliability of malicious terminals. When $Rel(i) \leq \alpha$, even though $Bias(i)$ is small, the data can be judged as malicious terminals at the first layer similarity comparison step; therefore, there is a trade-off problem that is δ need to be strong and do not be distinguished at the same time. This is why it is important to consider $Rel(i)$. δ has to be specified in the following manner in order to get the ideal attacking index:

$$\arg\max_{\delta} \sum_{i=1}^{M} [Rel(i) > \alpha] * Bias(i). \tag{3.12}$$

where $[A \cdot B]$ represents a "SPECIAL IF FUNCTION", select all values A that satisfy $A \cdot B$. δ is the attacking index, when δ increases, $Rel(i)$ decreases, and $Bias(i)$ increases.

Equ.(3.12) represents, changing attacking index δ and select all $Rel(i)$ satisfy $Rel(i) > \alpha$, and multiply the selected values by $Bias(i)$. Find the value of δ when the following formula reaches the maximum value.

We make the assumption that the malicious terminals have access to an attack center, similar to a database. We also assume that the attack center may hack the database and acquire confidential information. As a result, the attack center has complete knowledge of the DLM operation method and its associated parameters. This allows the malicious terminals to generate their optimal attack strategy within our security network. Here, the optimal attack strategy means the strongest attack strategy to our system.

The optimal approach is to ensure that the malicious datasets can participate in the power fusion step and create as many errors as possible. In this scenario, the $HisRe$ metric is monitored, and as long as it is lower than the threshold, the attacker should set the attack model to silent and transmit accurate data to increase $HisRe$. Once $HisRe$ is deemed safe, the attacker can resume the attack, resulting in errors in the REM.

3.5 Results and Discussion

In this part, we will demonstrate the simulation results that are used to validate the performance of the algorithms discussed before. Additionally, "sim" is the abbreviated name for similarity comparison, and "corr" is the abbreviated name for spatial correlation. The term "histo" in this context indicates that the algorithms take the historical reliability into consideration; hence, it is the shortened version for DLM. In the first column of Table 3.5, the following methods are listed: DLM based on IDW, DLM based on spatial correlation, DLM alone, similarity comparison (first layer) based on IDW, similarity comparison based

on spatial correlation, similarity comparison alone, bi_weight from reference [79], average combination from reference [80], IDW alone, spatial correlation alone, no algorithms and only normal terminals. Besides, MAE is used to represent accuracy in the CDF figures in this chapter.

For the 'average combination' [80] is using the sample mean and sample standard deviation of the energy values of all users to calculate the outlier factors,

$$o_n = \frac{e_n - \mu}{\sigma}, \tag{3.13}$$

where, μ and σ are the mean and standard deviation of the dataset, respectively. The users whose outliers values have a magnitude above the threshold are considered as malicious.

The 'bi_weight' [79] can be calculated as follows,

$$\hat{\mu} = \frac{\sum w_n e_n}{\sum w_n}, \tag{3.14}$$

where,

$$w_n = \begin{cases} (1 - (\frac{e_n - \hat{\mu}}{c_1 S}^2)^2, & \frac{e_n - \hat{\mu}}{c_1 S}^2 < 1 \\ 0, \quad \text{otherwise} \end{cases}, \tag{3.15}$$

and,

$$S = \text{median}|\text{e}_\text{n} - \hat{\mu}|. \tag{3.16}$$

The median absolute deviation (MAD) of the 'bi_weight' is given by,

$$\hat{\sigma} = \sqrt{\frac{N \sum_{u_n{}^2} (e_n - \hat{\mu})^2 (1 - u_n^2)^4}{s(-1+s)}}, \tag{3.17}$$

where,

$$s = \sum_{u_n{}^2 < 1} (1 - u_n^2)(1 - 5u_n^2), \tag{3.18}$$

and,

$$u_n = \frac{e_n - \hat{\mu}}{c_2 \cdot \text{median}|\text{e}_\text{n} - \hat{\mu}|}. \tag{3.19}$$

Overall, the comparison methods are shown in Table 3.3.

Table 3.3 Comparison methods.

Names	Algorithms
histo+IDW	Double layer monitor based on IDW
histo+corr	Double layer monitor based on spatial correlation
histo only	Double layer monitor only
sim+IDW	Similarity comparison (1st layer) based on IDW
sim+corr	Similarity comparison (1st layer) based on spatial correlation
sim only	Similarity comparison (1st layer) only
bi_weight	Reference [79]
average combination	Reference [80]
IDW only	inverse distance weighting
corr only	spatial correlation
with all terminals	No any secure or spatial algorithms
ON	Only normal terminals' information is selected

3.5.1 Simulation setup

We form a square that was 10 meters on each side inside the communication area. For purposes of simplicity, we will assume that each terminal takes a random path (a random walk from one mesh to the next mesh), that they all travel the same distance of 20 steps from one side of the communication area to the other, and that they only upload a single set of data at each step. Because the movement of each terminal is unpredictable, the total number of reports generated by each mesh is unique. A mesh that is traversed by a greater number of terminals may produce more reports. Overall, 900 routes are generated by random, and from those, 100 are selected by random to operate as malicious terminals. The parameters of the simulation can be found in Table 3.4.

3.5.2 Radio propagation model

Let $\boldsymbol{s}_{T_x}$ denote the primary user's transmitter location, therefore, we assume that the received signal power of the terminal which is located at $\boldsymbol{s} = (x, y)$ is given as follows,

$$P(\boldsymbol{s}) = P_{T_x} - L(d_0) - 10\eta \log_{10}(\frac{d_m}{d_0}) + W + F, \tag{3.20}$$

where P_{T_x} represents the primary transmission power in the dBm domain and $d_m = ||\boldsymbol{s}_{T_x} - \boldsymbol{s}||$ represents the distance [m] between the transmitter and the sensing terminal which located at $\boldsymbol{s}$, d_0 represents the reference distance [m], η represents the path loss index, W represents

Table 3.4 Simulation parameters.

Parameter	Value
Mesh size [m^2]	10×10
Mesh amount	20×20
Center frequency [GHz]	3.5
Transmission power [dBm]	29
Reference distance [m]	10
path-loss index η	3.5
Standard deviation of W	6
Similarity threshold α	0.95
Bias threshold ζ	0.8
Historical reliability threshold β	0.1
Correlation threshold	0.75
IDW power parameter	3
The number of routes	900
Percentage of malicious routes	11.11
Steps for each route	20
attacking index	0.5-1.5

the shadowing loss [dB] at the location $\boldsymbol{s}$ and W follows a log-normal distribution with a standard deviation of σ [dB]. F represents the small-scale fading [dB]. $L(d_0)$ represents the free-space path loss [dB], which is calculated as follows,

$$L(d_0) = 10\log_{10}(\frac{4\pi d_0}{\lambda})^2, \tag{3.21}$$

where λ represents the wavelength [m].

3.5.3 Fixed-terminal condition

In this subsection, we check the simple condition, in which we do not consider the random routes of the terminals. The environment is simple and ideal. We set the terminals on the map at fixed locations.

First, we would like to check when the number of malicious-terminal reports increases, the error of the REM, where the error is the distance between the REM constructed power strength with the real power strength. We set 200 reports in each mesh, and the ratio of malicious reports ranges from 5% to 20%. The attacking index is set to $\delta = 0.7$ all the time, the Mean Absolute Error (MAE) which can be denoted as,

$$MAE = \frac{1}{D}\sum_{i=1}^{D}(|\overline{P}_{\text{true}}(i) - \overline{P'}(i)|). \tag{3.22}$$

where, $\overline{P}_{\text{true}}(i)$ is the ground-truth power strength of the i-th interested mesh, $\overline{P'}(i)$ is the estimated power strength of the i-th interested mesh under attacking. MAE of the fixed-terminal condition is shown in Fig.3.5.

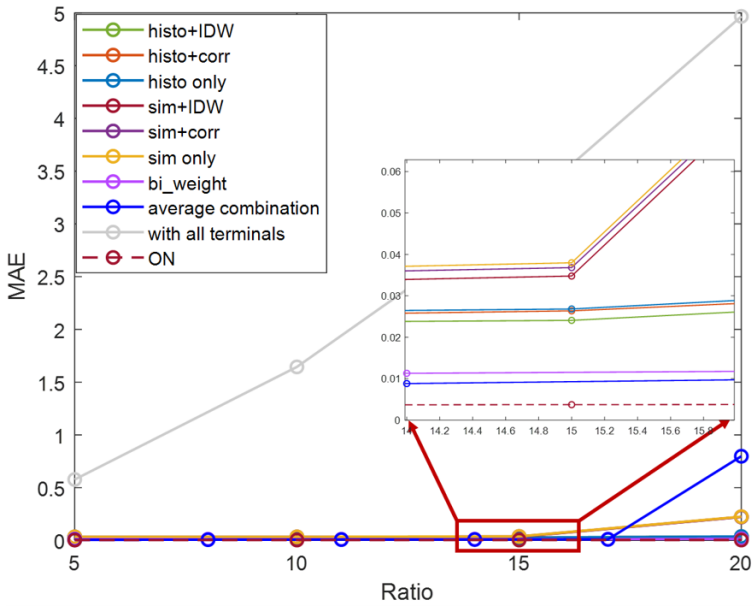


Figure 3.5 MAE versus of fixed reports.

We can find the MAE increases fast when the ratio of malicious-terminal reports increases with all terminals and, obviously, the error comes from the effects of the malicious terminals attack. When the amount of malicious terminals increases, the database receives more malicious reports, and the rate of honest reports decreases, the accuracy of the performance decreases. After adding the histo part, the methods have a similar performance when the ratio of malicious terminals increases that is because by using the historical method with the reputation, the database can judge the terminals continuously, so the malicious terminals are easy to be found. By using the similarity method, when the amount of malicious terminals increases, the performance has a slight increase, since the amount of honest terminals is still over the amount of malicious, by comparison, the database can make the correct decision. bi_weight and average combination worse than our proposed method, since they only consider the reports of power, they didn't use any geo-information to do the interpolation.

In addition, we examine the performance when the ratio of malicious reports is fixed but the number of reports received in each mesh varies. The ratio of malicious-terminal reports

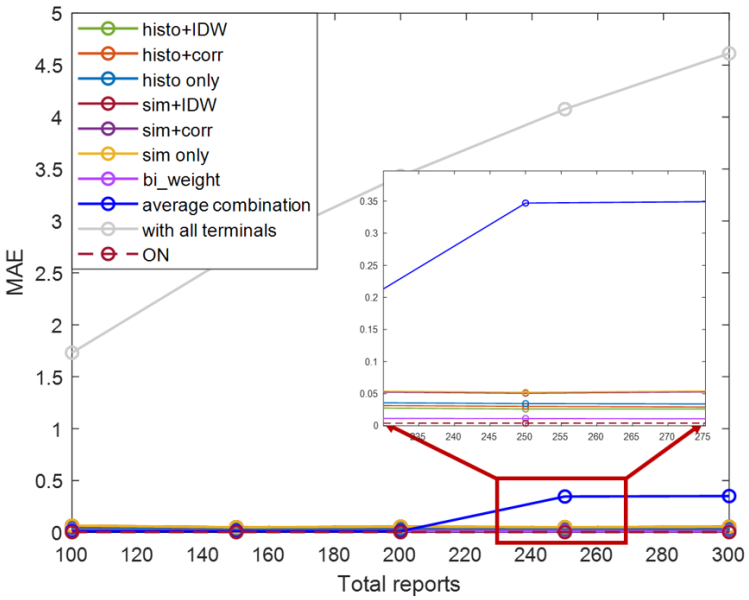


Figure 3.6 MAE versus of fixed ratio.

is set to 15%, and the number of reports in each mesh ranges between 100 and 300. As demonstrated in Fig.3.6, the error rises as the total number of reports increases.

But in fact, the situation is more complicated since the terminals' movements are unpredictable, the quantity of reports received and the malicious reports in each mesh are distinct. We test how well it works under a variety of random terminal movements and attack scenarios in the following sections.

3.5.4 Under static attack

As previously stated, in a static attack, malicious terminals execute the same attack at all times, they do not alter their attack strategy (including the attacking index and attacking position) during their movements. They are unable to stop once they start acting in an attacking manner. We categorize static attacks as either independent or cooperative. The malicious terminals rate for each mesh in our interested area is shown as Fig.3.7. In mesh **1-20**, the malicious terminals' rate is 0.03448, and more normal terminals passed by than the malicious ones, however, in mesh **13-13**, the malicious terminals' rate is 0.6429, which means, malicious terminals' information is over half. The performance of these two meshes also detailed in the following.

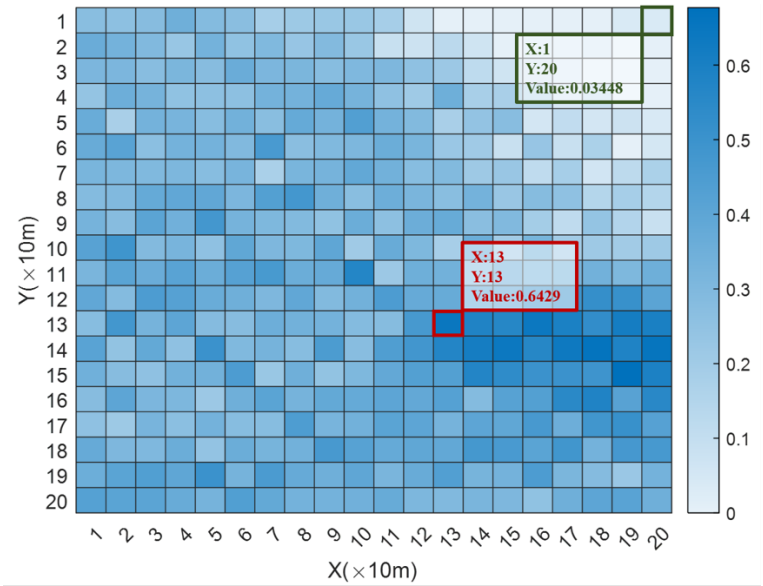


Figure 3.7 Malicious terminals' rate.

3.5.4.1 Under independent strategy

During a static attack using the independent approach, malicious terminals choose their attacking indexes randomly and launch attacks at any point in their trips. After settling on a set of attacking indices, they don't alter them mid-move. Assume for the moment that malicious nodes choose their attacking indices at random between 0.5 and 1.5. The attacking-index information of each malicious route is shown in Fig.3.8, and the CDF curves after applying the security method are displayed in Fig.3.9.

As shown in Fig.3.8, we set 100 malicious terminals in total and the horizontal axis indicates the route number of the malicious terminal, and the vertical axis indicates the attacking index δ, from the figure can find, the first malicious route take $\delta_1 = 1.31$ as the attacking strength and do this attack during its movement entirely. Additionally, as the definition of the falsification attack, when the attack index is approached to 1, the attacking strength becomes weaker.

DLM based on IDW outperforms DLM based on spatial correlation when malicious terminals use a static attack using the independent method, as illustrated in Fig.3.9. Algorithms that take into account past reliability do better than those that focus just on similarities in the present. Algorithms that take into account geographical information also have an advantage over their non-spatial counterparts.

Table 3.5 displays the results of these estimate measurements, including the average accuracy and the individual accuracy in mesh **1-20** and **13-13**. The accuracy here is checked by calculating MAE which is denoted as Equ.(3.22).

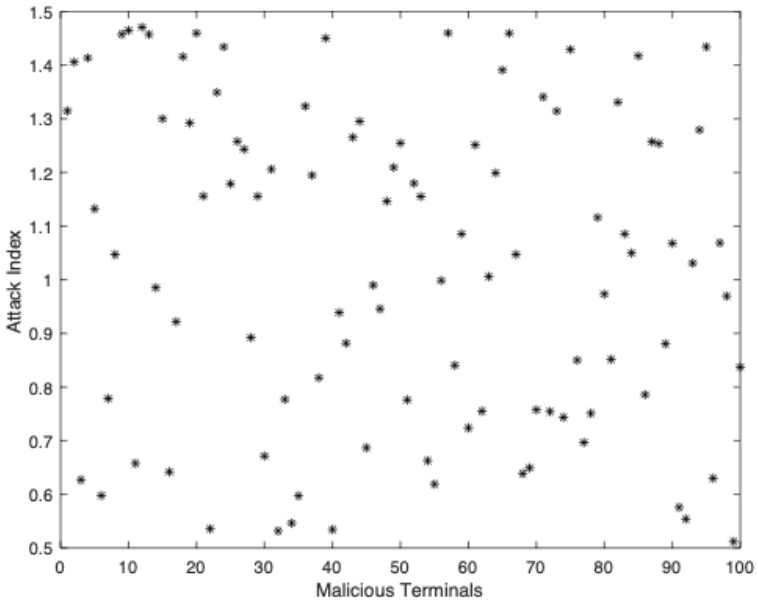


Figure 3.8 Attacking index of malicious terminals.

Table 3.5 MAE [dB] under independent static attack.

	Total avg.	1-20	13-13
histo+IDW	0.0422	0.0115	0.0078
histo+corr	0.0478	0.0143	0.0031
histo only	0.0579	0.0299	0.0190
sim+IDW	0.7400	0.0115	12.9669
sim+corr	0.7474	0.0230	12.8949
sim only	0.7515	0.0299	13.4674
bi_weight	3.0371	0.0141	24.5666
average combination	4.7192	0.1249	26.5893
IDW only	10.7568	0.8229	22.3174
corr only	10.7733	0.9963	20.9597
with all terminals	10.7779	0.9978	21.5496
ON	0.0579	0.0299	0.0190

The MAEs are minimized in DLM based on IDW, and DLM based on spatial correlation is superior to employing only the DLM as Table 3.5 shows. Using geographical information and including historical data both boost performance over relying just on similarity comparison. Without the security method, employing simply geographical information might actually result in a bigger inaccuracy rather than an improvement in REM precision. This is due to the fact that malicious information might be used for estimating purposes in the interpolated

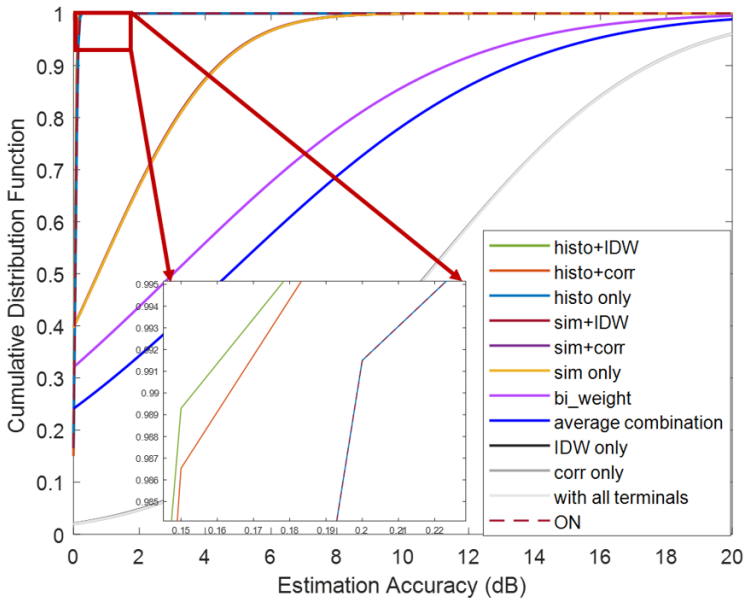


Figure 3.9 CDF of static attack under independent.

locations, which could result in much more mistake than would occur without interpolation. When the security technique is implemented at the lowest possible level, however, malicious information may be purged from the database and interpolation can improve precision.

3.5.4.2 Under collaborative strategy

When malicious terminals work together, they are unable to independently make the choice to launch an attack, which is not the case with the independent method. To submit false data to the database, the malicious terminals need to choose the same attacking index and follow the same procedure. Due to this subsection's emphasis on the static attack method, all malicious terminals use the same attacking index and continuously transmit inaccurate data to the database regardless of where they are located.

Fig.3.10 shows the performance under collaborative attack. Based on the definition of the SSDF attack, when $\delta < 1$, the attacking strength decreases when the attacking index increases, and the detection of the malicious terminals becomes harder if only using the first layer. Conversely, when $\delta > 1$, the attacking strength increases when the attacking index increases.

From Table 3.6 shown, except for our proposed methods, all other methods lead to a big error when the malicious information in one mesh is over half, although some of them are able to have a good performance in the total average, they can make a big interference in some certain mesh. By using our DLM based algorithm, this problem is solved, which means, we have a good performance when malicious information is localized over 50%.

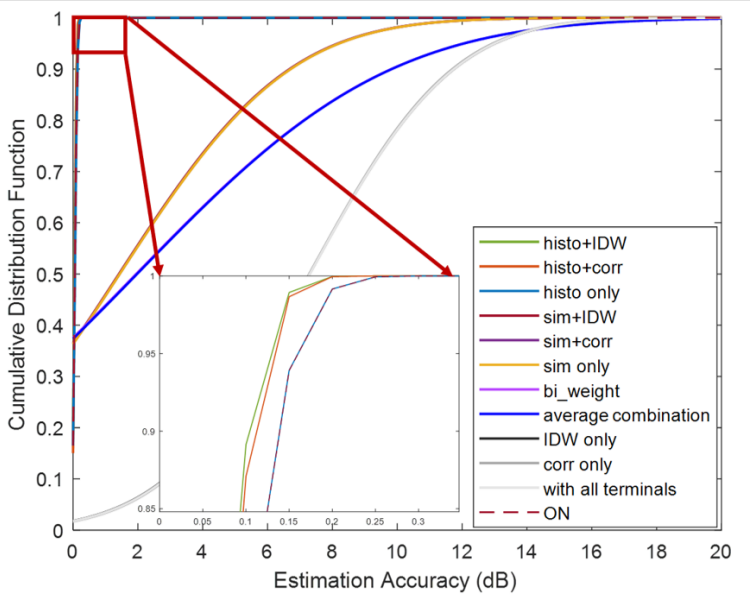


Figure 3.10 CDF of static attack under collaborative.

Table 3.6 MAE [dB] under collaborative static attack.

	Total avg.	1-20	13-13
histo+IDW	0.0472	0.0115	0.0078
histo+corr	0.0513	0.0143	0.0031
histo only	0.0625	0.0299	0.0190
sim+IDW	1.4157	0.0115	24.0892
sim+corr	1.4352	0.0230	24.0883
sim only	1.4391	0.0299	24.0963
bi_weight	1.9437	0.0141	24.0097
average combination	1.9737	0.1249	24.0097
IDW only	7.2928	0.6721	15.8268
corr only	7.2915	0.8127	15.1970
with all terminals	7.3014	0.8128	15.4836
ON	0.0579	0.0299	0.0190

3.5.5 Under dynamic attack

As we noted before, a dynamic attack is an intelligent attack approach in which malicious terminals may vary their attacking parameter as they are moving. This allows the malicious terminals to more effectively carry out their attacks. They are able to adjust the attacking indexes they use and even determine when those attacks launch; as a result, they are able to choose given steps which to undertake an attack and select other steps during which they will

not attack in order to defend the detection from the database. In this scenario, the attacking behavior of the malicious terminals is not fixed; rather, they are able to modify it whenever they deem it necessary. Dynamic attacks are divided into independent and collaborative categories.

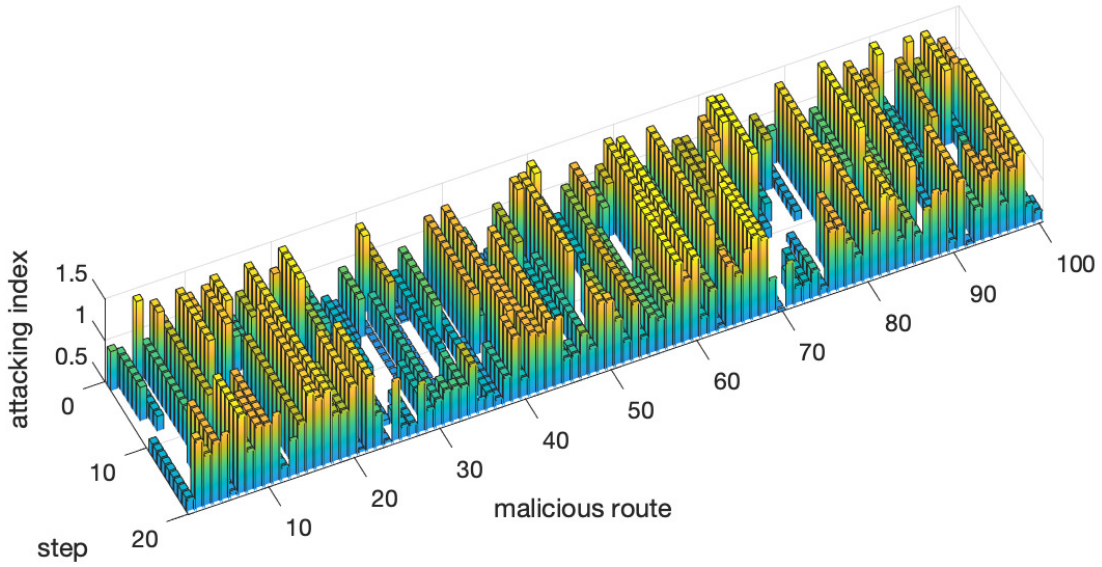


Figure 3.11 Attacking parameter condition.

3.5.5.1 Under independent strategy

In a dynamic attack utilizing an independent strategy, malicious terminals can independently make the attacking decision and establish the attacking parameters, including selecting and modifying their attacking indexes. At each phase, they can also determine whether or not to attack. In this paper, steps in which malicious terminals do not attack are referred to as "Silent Mode," and steps in which they do attack are referred to as "Active Mode." The Silent Mode refers to the positions in which malicious terminals appear as trustworthy terminals. To protect themselves and avoid detection, they should send the correct information to the

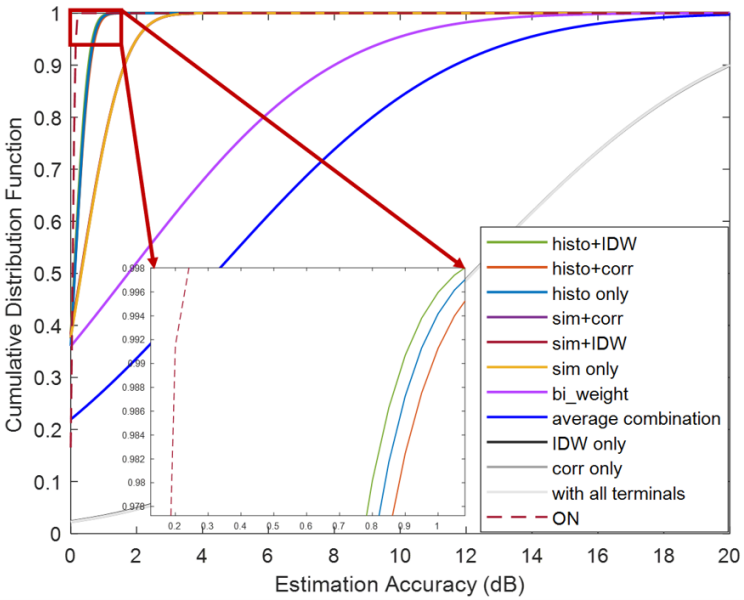


Figure 3.12 CDF of dynamic attack under independent.

database in this mode. Thus, $\delta = 1$ in the Silent Mode. For this section, we also presume that the attacking index ranges from 0.5 to 1.5 in the Active Mode, because, according to the preceding analysis, it is simpler to identify malicious terminals when the attacking strength is too high. In addition, each malicious terminal chooses its transition steps randomly throughout its entire journey. The attacking parameter conditions are depicted in Fig.3.11, while the CDF curves following the implementation of the security algorithm are depicted in Fig.3.12.

Fig.3.11 shows the attacking parameter of the dynamic attack. There are 100 malicious terminals in the interested communication area in total, and each terminal has 20 reports during their moving. The horizontal axis indicates the ID of the malicious terminals' route, the vertical axis indicates the different steps during their moving, and the bar height indicates the attacking index. The attacking index during the Silent Mode period has been set to 0 in this figure so that the attacking slot may be seen more clearly, however, in reality, it should be set to 1. Each terminal has three states during their moving, during the first and the third states, the malicious terminals do attack, and during the second state, the malicious terminals act as honest terminals. The time point for them to change their states and the attacking strength during their first and third states are chosen randomly by themselves. Consequently, in Fig.3.11, for the steps that lack the data, the malicious terminals behave as honest terminals, transmitting the correct information and not launching an attack. Each terminal randomly chooses the attacking index and determines the Silent Mode and Active Mode. Fig.3.12 depicts the performance of each algorithm under this attack strategy. The estimation precision

Table 3.7 MAE [dB] under independent dynamic attack.

	Total avg.	1-20	13-13
histo+IDW	0.1044	0.0115	0.0078
histo+corr	0.1184	0.0143	0.0031
histo only	0.1242	0.0299	0.0190
sim+IDW	0.3077	0.0115	9.3373
sim+corr	0.3155	0.0230	9.6441
sim only	0.3206	0.0299	10.0708
bi_weight	1.7590	0.0141	22.3640
average combination	4.3876	0.1249	22.3640
IDW only	12.1743	0.6199	30.0273
corr only	12.1908	0.7493	27.3028
with all terminals	12.1905	0.7489	28.0240
ON	0.0579	0.0299	0.0190

is shown in Table 3.7. The algorithms that use historical data outperform those that use only similarity data, and both outperform those that do not use security data. The security algorithms that take spatial information into account outperform those that do not. IDW-based DLM obtains the highest accuracy among the algorithms.

3.5.5.2 Under collaborative strategy

In contrast to the independent attack strategy, the collaborative strategy is a group attack in which malicious terminals attack synchronously, they share the same attacking parameters, and can be considered to have the same brain. After the leader of the malicious terminals chooses the attacking index and steps, the other terminals adopt the same strategy.

In this subsection, the independent attack plan is implemented by the first malicious terminal, which acts as the network's de facto leader. The attacking parameter condition is set to the same as the 9^{th} malicious route depicted in Fig.3.11, the Silent Mode lasts from step 9 to step 11, and the attacking indexes in Active Mode are 0.979523385210219 and 1.49949162009770, respectively. The other terminals all attack in the same way that the leader does. Fig.3.13 displays the CDF curves, whereas Table 3.8 displays the estimate accuracy. Results from simulations show that DLM based on IDW provides the highest performance, and that using historical data may boost estimate precision.

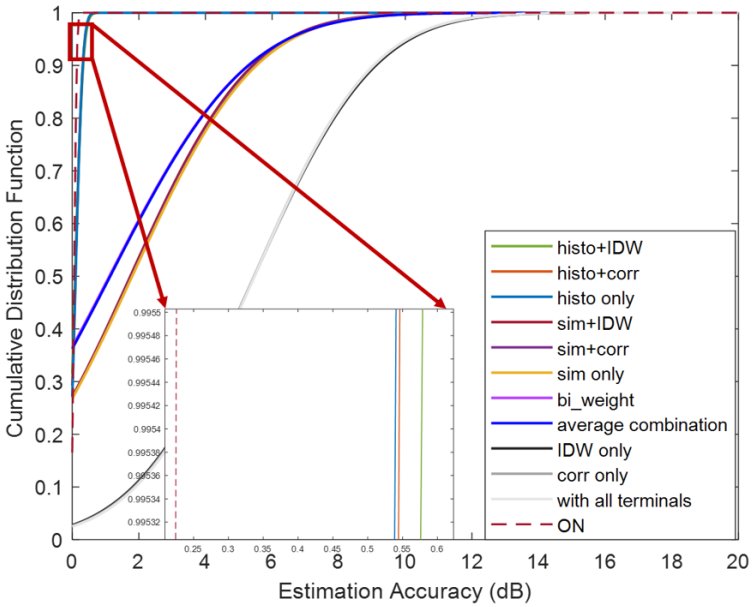


Figure 3.13 CDF of dynamic attack under collaborative.

Table 3.8 MAE [dB] under collaborative dynamic attack.

	Total avg.	1-20	13-13
histo+IDW	0.0892	0.0115	0.7754
histo+corr	0.0942	0.0143	0.5829
histo only	0.1013	0.0299	0.7628
sim+IDW	1.7445	0.0115	7.9138
sim+corr	1.7746	0.0230	7.6009
sim only	1.8013	0.0299	7.7489
bi_weight	1.1218	0.0141	11.9594
average combination	1.1504	0.1249	11.9594
IDW only	5.4660	0.3316	7.9138
corr only	5.4589	0.3984	7.6009
with all terminals	5.4736	0.3954	7.7489
ON	0.0579	0.0299	0.0190

3.5.6 Under optimal attack

To comprehensively evaluate the performance, we also generated the optimal (strongest) white-box attack strategy condition. As stated in section 3.4.2, if the malicious terminals have an attack center that can acquire the complete information of the security network, including algorithm operation and parameter setting, they can function as a database and emulate the reliability of each malicious terminal. In this scenario, the types of attacks become complex

and diverse, and the network's security can be severely compromised. The attacking index δ must satisfy Equ.(3.12) due to a trade-off problem between the detection probability and the attacking intensity. In addition, the attack strategy must take into account $HisRe$ to ensure that malicious terminals can join the power fusion.

The results of attacks using various attacking indices are shown in Fig.3.14. The usual attacking is shown by the colored dashed lines, while the malicious terminals constantly launch attacks. Different colored solid lines represent optimal attacking, where past reliability is tracked. As long as the $HisRe$ is lower than the threshold, the malicious terminals should switch their mode to Silent Mode in order to improve the system's reliability. This will guarantee that the incorrect information may join the power fusion, which will result in an error in the REM. They should switch back to Active Mode whenever the $HisRe$ is high enough to join into power fusion. This structure needs to be used repeatedly.

The proportion of malicious terminals that are not discovered after the first layer (similarity comparison) is shown by the color green in Fig.3.14. If the percentage of the malicious terminals is higher, it demonstrates that the attacking performance was better. The findings of the simulation indicate that the optimal attack approach is superior to the standard method in terms of effectiveness. Similarly, the color yellow represents the proportion of malicious terminals that remain undetected after the second layer of protection (DLM). The ideal attack performs better, once more. In addition, when the attacking index is less than 0.8, malicious terminals are simple to identify even when using similarity comparison only. This is due to the fact that malicious information is noticeably distinct from original information. When the attacking index is greater than 0.9, the malicious terminals are not easily distinguishable even when using DLM. This is due to the fact that the difference between the malicious information and the original information is tiny. Additionally, when taking into consideration the fact that different terminals suffer from different channel conditions, such as path loss and fading, the algorithm is unable to determine whether the difference was caused by the malicious action or by the channel difference. The lines of light-blue depict the typical degree of similarity degree across malicious terminals; a greater value implies that it will be more challenging to determine whether or not the terminal is malicious. Both the weight allocation and the historical reliability of malicious terminals are represented by the red and grey colors, respectively. The historical reliability is similar with the similarity degree; in both cases, a greater value suggests a lesser possibility of being found. The RMSE for various attacking indices is shown in a dark blue color. When the attacking index is close to 0.9, attacking effectiveness is at its peak. When the attacking index is raised, the attacking strength is weakened, resulting in a smaller RMSE. As an example, when the attacking index is extremely near to 1, the REM error is minor even do not use a security method.

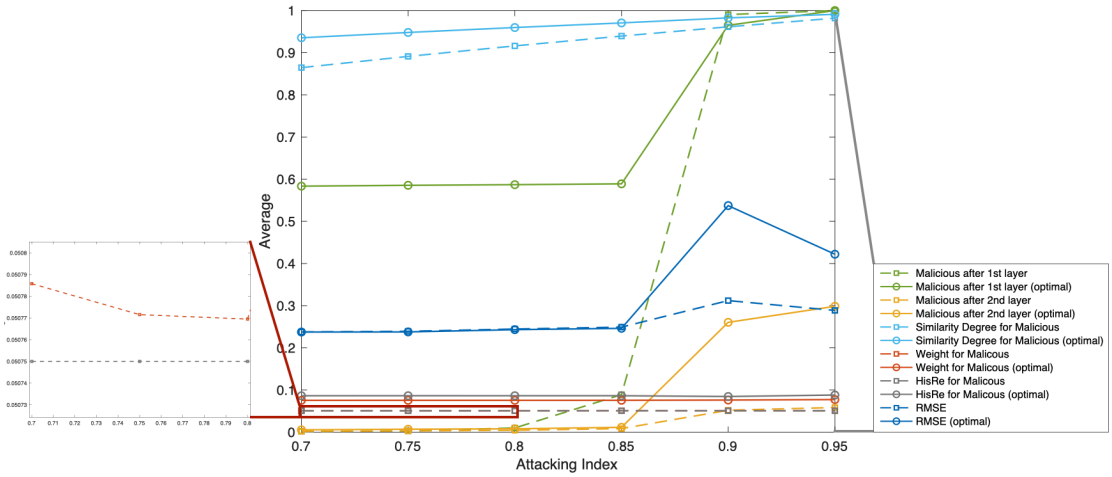


Figure 3.14 Attacking performance under different attacking index.

Figs.3.15 and 3.16 show the historical reliability under the optimal attack with an attacking index $\delta = 0.9$. The former shows the total map for all the malicious routes, there are 100 malicious terminals in the interested communication area, and each terminal has 20 reports during their moving. The horizontal axis indicates the ID of the malicious terminals' route, the vertical axis indicates the different steps during their moving, the bar height indicates the historical reliability of each terminal at different steps of its movement.

The first six malicious routes are shown on the map in Fig.3.16. The grey bar represents an attack being carried out by the malicious terminal. When there is no gray area, the malicious terminal is behaving like a trustworthy terminal throughout that process step. The orange bar shows how reliable every step has been in the past. Since the results of the historical reliability calculation for each terminal are collected at the end of the step and influence the next step, we can identify the moment the malicious terminal switches to Silent Mode. After this step, the historical reliability gradually improves. Additionally, the reliability of past events can quickly decline when the malicious terminal launches an attack.

The CDF curves obtained with the DLM under various attacking indices are shown in Fig.3.17, and the errors obtained with various techniques are displayed in Table 3.9. The error that occurs when an attacker performs the optimal attack using the DLM algorithm is shown in the "DLM(optimal)" column, the error that occurs when an attacker uses the normal attack using the DLM algorithm is shown in the "DLM" column, and the error that occurs when no security methods are employed is shown in the rightmost column, named as "with all terminals". The figure and table show that the REM suffers the most errors when challenged with an optimal attack technique, when attackers launched small strength attacks, $\delta = 0.9$ reached the optimal behavior for the malicious terminals, which means the worst condition

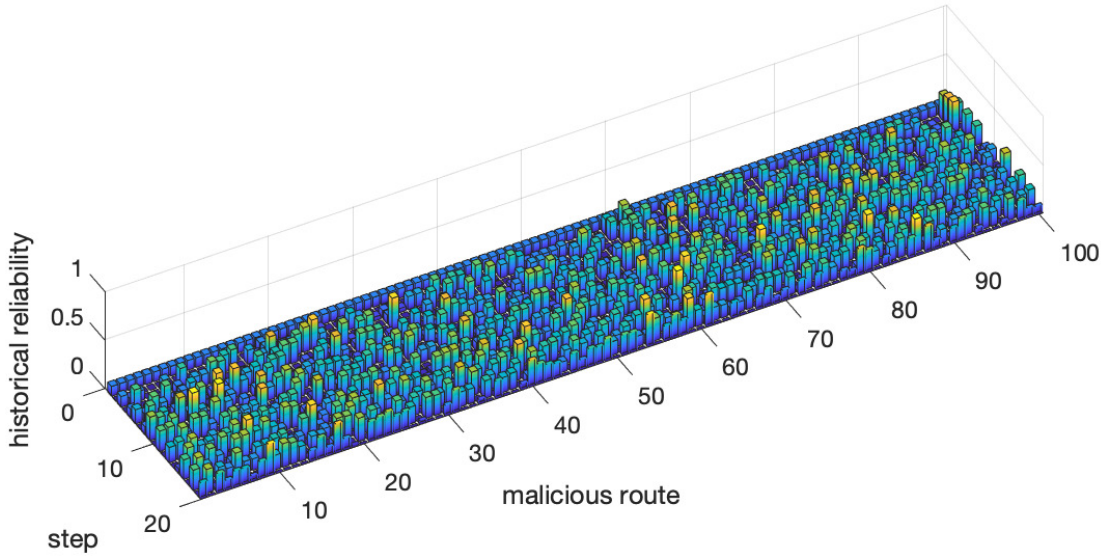


Figure 3.15 Historical reliability under optimal attack.

for the normal fusion database, but that DLM errors remain less than 0.5dB regardless of the severity of the attack.

3.5.7 Discussion

From the multiple simulation results, we found that the DLM has significant advantages for data falsification attacks. After historical information is added, the database can continually monitor the behavior of malicious terminals, and the performance improves significantly after the historical reliability is calculated. In addition, we considered using spatial information to improve the algorithm. The simulation results indicated that the DLM based on IDW is better than the DLM based on spatial correlation, and both have better performance than the DLM alone. Finally, we used the full knowledge of the network and launched the optimal attack (strongest attack), and the results indicated that the optimal attack (strongest attack) has advantages (the error of the REM increases), nonetheless, our security algorithm can ensure

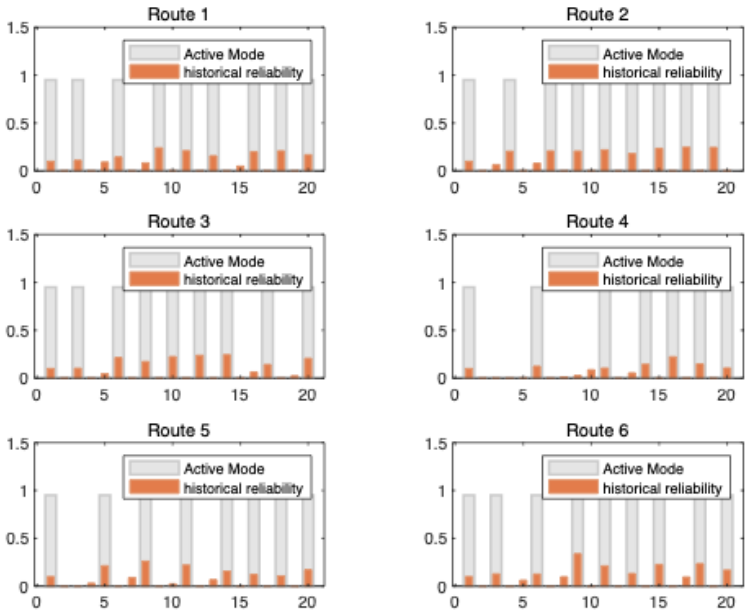


Figure 3.16 Attacking condition of first 6 malicious routes.

Table 3.9 MAE [dB] under optimal attack.

δ	Total avg.		
	DLM(optimal)	DLM	with all terminals
0.85	0.0684	0.0999	**1.2877**
0.87	0.1529	0.1076	1.1160
0.90	**0.3878**	**0.2240**	0.8585
0.93	0.2846	0.1600	0.6009
0.95	0.2104	0.1241	0.4292

the accuracy of the network. The complexity of algorithms we mentioned in this chapter are as follows:

histo: $O(n^2)$

sim: $O(n^2)$

bi_weight: $O(mn)$

average combination: $O(n)$

Combined with the simulation results above, it is easy to find that the average combination method has the lowest complexity, however, the accuracy is the worst. Although bi_weight can have a good performance under the condition that the number of malicious terminals' reports is less than half, when the malicious reports' number is over half, the performance of this method decreases significantly. However, our methods are relatively complex compared

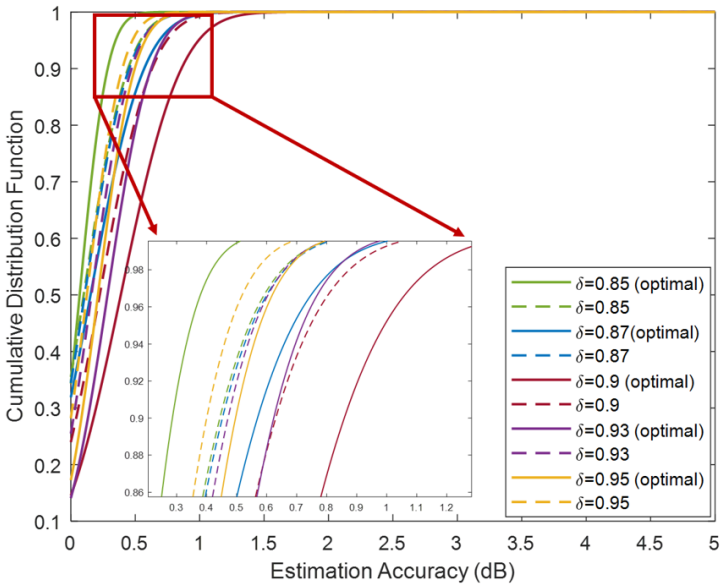


Figure 3.17 CDF under different attacking index.

with others, by using our proposed method, even when the malicious reports are over half in one mesh, still can maintain the high accuracy of the REM.

Additionally, we stored data in large-capacity and high-computing places, such as a database, to complete the elimination of malicious nodes. Our algorithm indeed has a high complexity, but it is very effective for sensing environments that require higher precision.

3.6 Chapter Summary

We proposed DLM based on spatial information algorithms, including IDW and spatial correlation, to deal with various data falsification attacks in the network. To ensure that the historical reliability decreases rapidly for malicious terminals and increases slowly for honest ones, we introduced the reward-penalty function of the DLM algorithm. The algorithm was evaluated under different attack scenarios such as static, dynamic, independent, and collaborative attacks, and the simulation results demonstrated its effectiveness in eliminating malicious information and improving the REM's accuracy. Furthermore, we defined the optimal attack (strongest attack) model based on our algorithm and attempted to check the network's robustness, and the simulation results revealed the superiority of the proposed attack strategy over the normal one. Importantly, our security algorithm remained effective even under the optimal attack scenario.

Chapter 4

Kriging-based Trust Nodes Aided REM Construction Method

The utilization of crowdsourcing for spectrum sensing offers a promising solution for generating the REM in large communication areas. In this chapter, we introduce the KTNA-REM system, which is based on Kriging and aided by trust nodes. By incorporating a small number of trust nodes, the database can continuously evaluate the reputation of terminals and generate the REM based on highly reliable data, even when the number of malicious terminals is significant. Simulation results demonstrate the effectiveness and accuracy of the proposed approach.

The chapter is organized as follows: Section 4.1 introduces the background of our study, section 4.2 explains the system models, and the implementation framework of our proposed method is discussed in section 4.3. The simulation results are shown at section 4.4 and finally, the chapter summary is in section 4.5.

4.1 Background

Radio Environment Map, or REM, is a tool that may be used to control the interference caused by several transmitters. For example, secondary users in TVWS systems often identify the white spaces and the permitted interference level based on the REM that is recorded in the spectrum database. The information included in the REM may be used to offer an efficient spectrum sharing system, as shown by the reference [20], which illustrates that employing this information can give an efficient spectrum sharing system.

The accuracy of REM construction is critical as it directly impacts spectral efficiency. The REM is generated by analyzing information from multiple terminals that sense the

environment conditions and send their data to the database. This information includes details such as received signal power, terminal ID, and location. Based on this data, the database can generate the REM for a given communication area. Therefore, the accuracy of the information reported by the terminals to the database plays a crucial role in ensuring the accuracy of the REM construction.

Kriging-based methods are widely used to improve the accuracy of REM construction. In Kriging-based REM construction, the database first collects a limited number of datasets and estimates the signal strength at unsampled locations using a mathematical model. Kriging interpolation is a popular approach that utilizes the spatial correlation between different locations to interpolate the signal strength at unsampled locations. By minimizing the variance of the estimation error under the unbiased estimation constraint, Kriging-based methods can generate a high accuracy REM with a limited amount of data. Moreover, the addition of trust nodes, which can provide reliable information and help to evaluate the reputation of the terminals, can further improve the accuracy of the REM. By applying the Kriging interpolation method, the database can use less data to estimate and interpolate a high accuracy REM [81].

It is important to note that the reliability of the reporting data can be compromised by various security threats in the open environment of crowdsourced REM construction. One such threat is the SSDF attack, also known as Byzantine attacks, where malicious or dishonest terminals intentionally or unintentionally provide false information to the database. This can significantly degrade the accuracy of the REM and affect the spectral efficiency of the system. Reputation-based algorithms have been proposed as a solution to deal with malicious terminals by updating the reputation of each terminal based on rewarding or punishing mechanisms [82]. Another approach is to use classifiers to detect and filter out the malicious terminals during sensing [47]. However, there is still limited research on addressing these challenges specifically in the context of REM construction, where the location and transmission activity of primary terminals are known, but the signal strength needs to be estimated at each location of interest. Chapter 3 has focused on collecting a large amount of data to remove malicious information, which can lead to significant time overheads in the initial phase of the system [83].

The KTNA-REM algorithm, introduced in this chapter, addresses the challenge of constructing a high-precision radio map using only a small number of terminals in a threatening environment. Unlike previous methods, the KTNA-REM algorithm does not require storing a large amount of data to begin detecting and is a real-time construction method that only stores high-reliability data, discarding lower reliability data in every sensing slot. The algorithm achieves map updates through data-driven methods.

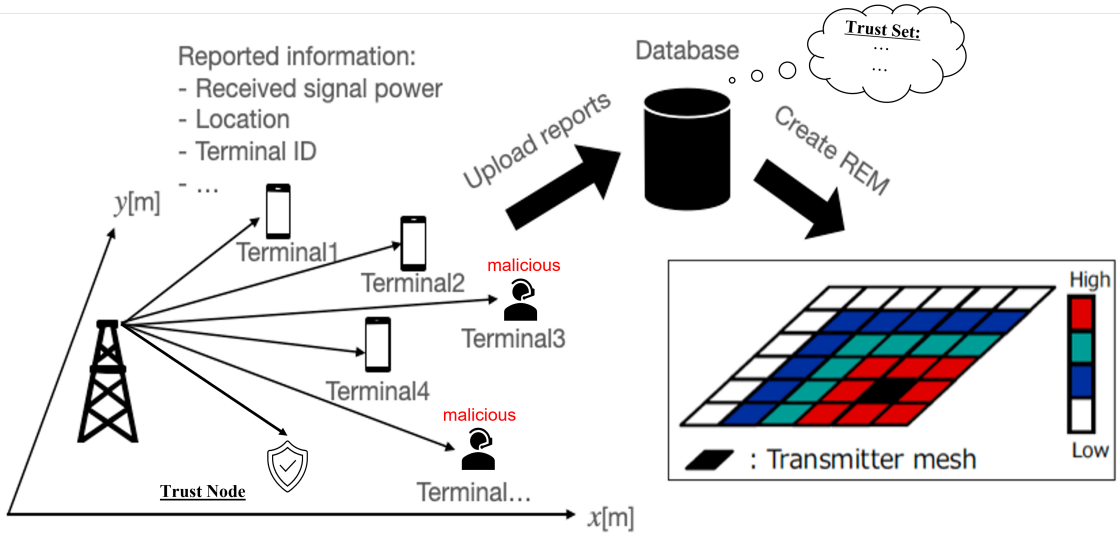


Figure 4.1 A concept of the conventional REM based on KTNA.

4.2 System Description

4.2.1 REM model

To create the REM for a communication area, multiple terminals are utilized to collect spectrum information within the area. The locations of these terminals are determined based on random movements, and both normal and malicious terminals are assumed to be present. Additionally, a small number of trust nodes are placed at fixed locations within the communication area to provide reliable information. The information collected by the terminals includes their ID, location, time, frequency, and power, which is then sent to the database for storage. The database, which can be located in the cloud or a base station with ample storage capacity, uses these datasets to generate the REM. A representation of a conventional REM is provided in Figure 4.1.

To mitigate the impact of small-scale fading, we divide the communication area into two-dimensional grids or meshes. This allows us to group the datasets from the same mesh together and calculate the average power for that particular mesh. When the mesh size is small enough, the impact of shadowing can be neglected, and the accuracy of the REM can be improved.

4.2.2 Radio propagation model

In this section, we define the dataset vector $\boldsymbol{S} = (P(\boldsymbol{s_1}), P(\boldsymbol{s_2}), ..., P(\boldsymbol{s_N}))^T$, where $\boldsymbol{s_i} = (x_i, y_i)$ is the measurement location of each terminal and $P(\boldsymbol{s_i})$ is the received signal

power. Let $\boldsymbol{s_{T_x}} = (x_{T_x}, y_{T_x})$ denote the primary user's transmitter location, therefore, the received signal power of the terminal which is located at $\boldsymbol{s}$ is given as follows,

$$P(\boldsymbol{s}) = P_{T_x} - L(\boldsymbol{s}) + W + F, \tag{4.1}$$

where P_{T_x} represents the primary transmission power in the dBm domain and $L(\boldsymbol{s})$ is the path loss between $\boldsymbol{s}$ and $\boldsymbol{s_{T_x}}$. W represents the shadowing loss [dB] at the location $\boldsymbol{s}$ and W follows a log-normal distribution with a standard deviation of σ [dB]. As both L and W are location-dependent scalars, we express them as functions of the location vector $\boldsymbol{s}$. F represents the small-scale fading [dB]. Here, we assume the spatial correlation in shadowing follows a typical model [75],

$$\begin{aligned} \rho_{i,j} &= \frac{\mathsf{E}[(W(\boldsymbol{s_i}) - \mathsf{E}[W(\boldsymbol{s_i})])((W(\boldsymbol{s_j}) - \mathsf{E}[W(\boldsymbol{s_j})])]}{\sigma_i \sigma_j} \\ &\approx \exp\left\{\left(\frac{-||\boldsymbol{s_i} - \boldsymbol{s_j}|| \ln 2}{d_{cor}}\right)\right\} \end{aligned}, \tag{4.2}$$

where $||\cdot||$ is Euclidean distance, σ_i represents the standard deviation of $W(\boldsymbol{s_i})$. d_{cor} is the correlation distance, defined as a point on $\rho_{i,j} = 0.5$.

4.2.3 Attacking model

We assume that there are malicious terminals in the communication environment, and these terminals have the ability to blind the database by rewriting the sensing power of the spectrum and reporting incorrect information about the power to the database. As a result, these incorrect datasets can cause a significant amount of error in the REM. This is done in order to fulfill the malicious terminals' own self-serving requirements. Data falsification attack is the name given to the technique of attack that involves rewriting the sensing power.

In a data falsification attack, also known as a Byzantine attack, malicious terminals alter the data they collect using their sensing capabilities in the spectral domain in order to deceive the database. This kind of attack model was shown by reference [65]. One such traditional method of data fabrication attack is outlined here. In order to alter their data, malicious terminals check it against a power threshold, denoted by κ. If the malicious terminal's sensing power is over the threshold for any given sensing slot, it will rewrite the sensing power multiplied by an attack index κ with a probability P_a to submit an incorrect dataset to the database. If not, malicious terminals will upload the right dataset. Here is the reported strength of the malicious terminal,

$$P'(i) = \begin{cases} P(i) \cdot \delta, & \text{if } P(i) > \kappa \text{ with } P_a \\ P(i), & \text{otherwise} \end{cases}, \tag{4.3}$$

where, the probability of an attack, denoted by P_a, and the attacking index, denoted by δ. If the attacking index is greater than one, the attacking strength rises with the attacking index; otherwise, when attacking index is less than one, the attacking strength falls as the attacking index rises.

4.2.4 Ordinary Kriging

Ordinary Kriging [1] is one of the most popular way to do the spatial interpolation for the REM. The method to interpolate the unknown value is like follows,

$$\hat{P}(\boldsymbol{s_j}) = \sum_{i=1}^{N} \omega(i) P(\boldsymbol{s_i}), \tag{4.4}$$

where, $\omega(i)$ is the weight factor of the i-th terminal multiplied by the known value $P(i)$, and $\hat{P}(\boldsymbol{s_j})$ is the interpolated value of $P(\boldsymbol{s_j})$. Kriging minimizes the estimation error by optimizing the $\omega(i)$. Additionally, ordinary Kriging assumes that $\mathsf{E}[P(\boldsymbol{s_i})] = const.$ over any i, besides, the spatial correlation property of $P(\boldsymbol{s_i})$ is static over the entire measurement area. Usually, ordinary Kriging can be applied for the estimate the shadowing index $W(\boldsymbol{s_j})$, and the interpolation result at j can be given by,

$$\begin{aligned} \hat{P}(\boldsymbol{s_j}) = \hat{P}_{T_x} - 10\hat{\eta}\log_{10}||\boldsymbol{s_j} - \boldsymbol{s_{T_x}}|| \\ + \sum_{i=1}^{N} \omega(\boldsymbol{s_i})(P(\boldsymbol{s_i}) - (\hat{P}_{T_x} - 10\hat{\eta}\log_{10}||\boldsymbol{s_i} - \boldsymbol{s_{T_x}}||)) \end{aligned}, \tag{4.5}$$

where, $\hat{P}_{T_x}$ and $\hat{\eta}$ are the estimated P_{T_x} and η.

Ordinary Kriging determines the optimal weight which can minimize the variance of estimation error $\sigma_L^2 = \mathsf{Var}[\hat{W}_{\boldsymbol{s_j}} - W_{\boldsymbol{s_j}}]$, where $\mathsf{Var}[\cdot]$ is the variance of the random variable. The wights need to satisfy: $\sum_i^N \omega(i) = 1$. By using Lagrange multiplier, the objective function

[1] Gaussian Process (GP) is a common non-parametric model in the field of machine learning, similar to kriging. If the covariance function is identical, the output of simple and ordinary kriging is identical to the mathematical expectation of the output of GP under normal likelihood. The difference between Kriging and GP is that the former assumes the random field is an inherently smooth process and provides its optimal unbiased estimate across the test sample, whereas the latter assumes the random field is a Gaussian process and provides its entire distribution over the test sample's posterior.[84–86]

can be written as,

$$\phi(\omega(i), \upsilon) = \sigma_L^2 - 2\upsilon(\sum_{i=1}^{N} \omega(i) - 1), \tag{4.6}$$

where, υ is the Lagrange multiplier. Here, σ_L^2 can be written as follows,

$$\sigma_L^2 = -\gamma(d_{j,j}) - \sum_{i=1}^{N}\sum_{j=1}^{N} \omega(i)\omega(j)\gamma(d_{i,j}) + 2\sum_{i=1}^{2} \omega(i)\gamma(d_{i,j}), \tag{4.7}$$

where, $d_{i,j} \triangleq ||\boldsymbol{s_i} - \boldsymbol{s_j}||$. In addition, γ is the semivariogram defined as,

$$\gamma(d_{i,j}) = \frac{1}{2}\mathsf{Var}[\hat{W}(\boldsymbol{s_i}) - \hat{W}(\boldsymbol{s_j})]. \tag{4.8}$$

The solution of the optimal $\omega(i)$ is given by,

$$\begin{pmatrix} \omega(1) \\ \omega(2) \\ \vdots \\ \omega(N) \\ \upsilon \end{pmatrix} = \begin{pmatrix} \gamma(d_{1,1}) & \cdots & \gamma(d_{1,N}) & 1 \\ \gamma(d_{2,1}) & \cdots & \gamma(d_{2,N}) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(d_{N,1}) & \cdots & \gamma(d_{N,N}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(d_{1,j}) \\ \gamma(d_{2,j}) \\ \vdots \\ \gamma(d_{N,j}) \\ 1 \end{pmatrix}. \tag{4.9}$$

4.3 Proposed Method

This section presents the KTNA-REM algorithm, which involves incorporating a limited number of trust nodes to counteract attacks. The trust nodes have predetermined, fixed locations, and to minimize overhead, they are only active during the initial sensing slot. Following the first sensing round, the trust nodes remain inactive, and their sensing data is added directly to the trust set.

4.3.1 Real-time comparison

To identify malicious datasets in the database, we compare the received power of each terminal with its estimated power using the interpolated map. We use the sensed information from the trust set and a Kriging-based interpolation method to compute the real-time REM, where the interpolated data at a particular point represents the terminals' estimated power. Due to the fact that the Kriging-based REM is highly accurate even with limited data, the trust set initially stores only data from the trust nodes during sensing. We compute the difference

between the received and estimated power as follows,

$$Bias_k(i) = |P'_k(\boldsymbol{s_i}) - \hat{P}_k(\boldsymbol{s_i})|, \tag{4.10}$$

where $\hat{P}_k(\boldsymbol{s_i})$ is the power estimated at the k-th sensing slot using Kriging interpolation. We should get rid of the really big outliers since they may have high $Bias$ and influence the reputation evaluation of other terminals. The formula for determining an outlier is as follows,

$$o_k(i) = \frac{Bias_k(i) - \mu_k}{\sigma_k}, \tag{4.11}$$

where, the sample mean and standard deviation of the differential value $Bias$ at the k-th sensing slot are denoted by μ_k and σ_k, respectively. The outlier of the i-th terminal in the k-th sensing slot is denoted by $o_k(i)$. The max-min scaling is as follows, excluding the estimated values for the i-th terminal if $o_k(i) > \theta$,

$$Bias^*_k(i) = \frac{Bias_k(i) - \mathsf{min}}{\mathsf{max} - \mathsf{min}}, \tag{4.12}$$

where, $Bias^*_k(i)$ is the normalized sensed-and-estimated difference (NSED), max and min are the maximum and minimum data after remove the extremely large outliers at the k-th sensing slot.

4.3.2 Accumulative-total reputation

When malicious terminals conduct data falsification attacks, they change the received information and send incorrect data to the database, even if they are at the same location as honest terminals. As a result, the datasets of malicious terminals are different from those of honest terminals. After calculating the NSED of each dataset, we compare it with the system threshold ζ and update the real-time comparison accordingly.

$$\begin{cases} ATRe_{k+1}(i) = ATRe_k(i) + a, & \text{if } Bias^*_k(i) < \zeta \\ ATRe_{k+1}(i) = ATRe_k(i) - b, & \mathsf{otherwise} \end{cases}, \tag{4.13}$$

where, $ATRe_k(i)$ is the accumulative-total reputation of i-th terminals at k-th sensing slot. a and b are the reward and penalty index, respectively. Note that, to effectively remove malicious terminals, it is important to ensure that the rate of decrease is faster than the rate of increase. If the calculated value of $ATRe_k(i)$ is high enough, it is considered that the

terminal has a high probability of being an honest terminal, and the sensing data from that terminal can be added to the trust set.

4.4 Simulation Results

We will compare the performance of the proposed KTNA-REM algorithm with four other strategies in this section. The communication area contains a total of 100 sensing terminals, including some malicious ones, and the trust nodes are evenly distributed across the area. The four strategies that we will compare are,

- *All Sensed Terminals under Real-Time (AST/RT)*: the database utilizes all information from the terminals, including malicious data, during the real-time sensing slot.
- *All But Malicious Terminals under Real-Time (ABMT/RT)*: the database constructs the REM by using only the honest information in real-time.
- *All Sensed Terminals under Accumulative-Total (AST/AT)*: the database collects and stores all information from the terminals, including the malicious ones, in real-time and from previous sensing slots. This means that all information is used equally without any distinction in each sensing slot.
- *All But Malicious Terminals under Accumulative-Total (ABMT/AT)*: the database constructs the REM using both real-time and previous rounds' information, but removes the malicious information before constructing the REM. Note that since it is impossible for the database to know which terminals are malicious, the accuracy of this strategy can be considered as the upper bound of any mechanism that can be achieved.

The parameters in this part is shown as TABLE 4.1.

4.4.1 Impact of different attacking strength

In this section, we evaluate the performance of our algorithm under different levels of attack after convergence. Specifically, we consider a scenario where 40 out of 100 terminals are malicious and there are 10 trust nodes. To assess the performance, MAE is an essential measurement to check the performance which can be calculated as Equ.(3.22).

As defined in section 4.2.3, the attacking index δ determines the attack's intensity. When δ approaches 1, the attacking power decreases; otherwise, the attacking power increases. MAEs under AST/RT, ABMT/RT, AST/AT, ABMT/AT, and our proposed method with increased attacking strength ($\delta = 0.5 \sim 0.9$) are depicted in Fig.4.2. Note that the ABMT/AT

Table 4.1 Simulation parameters.

Parameter	Value
Communication area [m^2]	100×100
Mesh size [m^2]	5×5
The number of meshes	400
The number of terminals	100
Center frequency [GHz]	3.5
Transmission power [dBm]	29
Reference distance [m]	10
Path-loss index η	3.5
Standard deviation of W	6

curve is unaffected by the change in attacking intensity, as it represents the sum of all the honest terminals. In this case, this curve is depicted as a reference for the algorithm's upper bound. As shown in Fig.4.2, the MAE of AST/AT is the best performance when only Ordinary Kriging is used to generate the REM, the real-time generated REM is is worse than the accumulative total since less information can be analyzed, and the MAE of both decreases as the attacking strength decreases. After convergence, the performance of our proposed method is comparable to that of the upper bound (ABMT/AT). The performance of the ABMT under real-time sensing is also stable; the error in this results from the lesser quantity of information that can be used in a given sensing period.

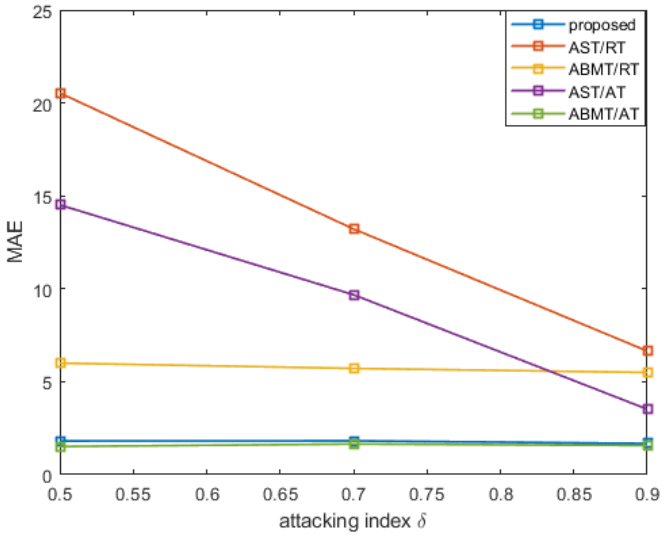
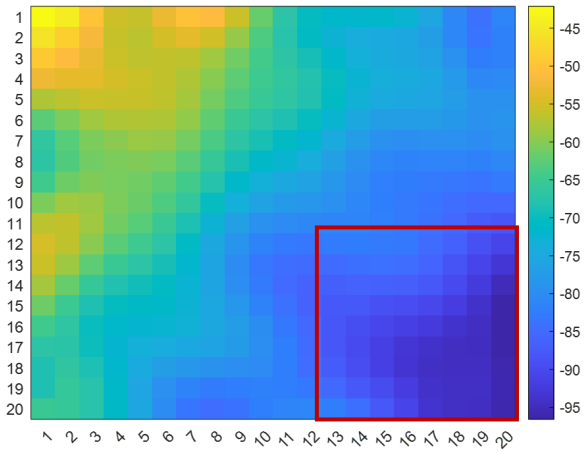
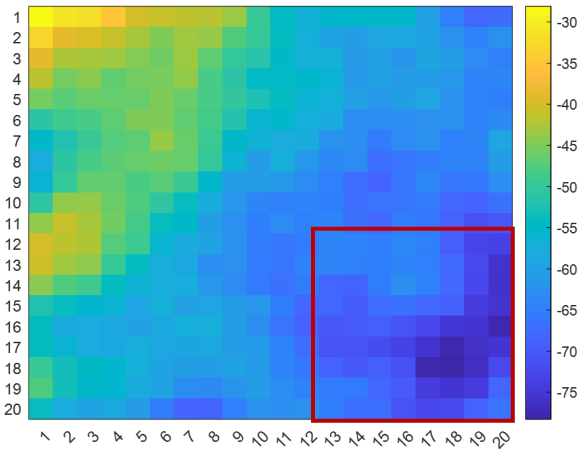


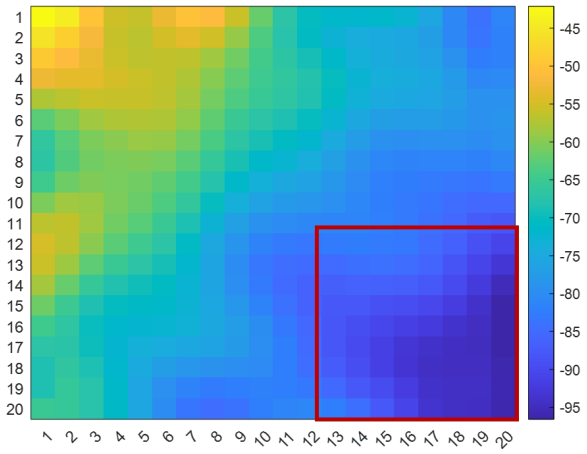
Figure 4.2 MAE under different attacking strength.



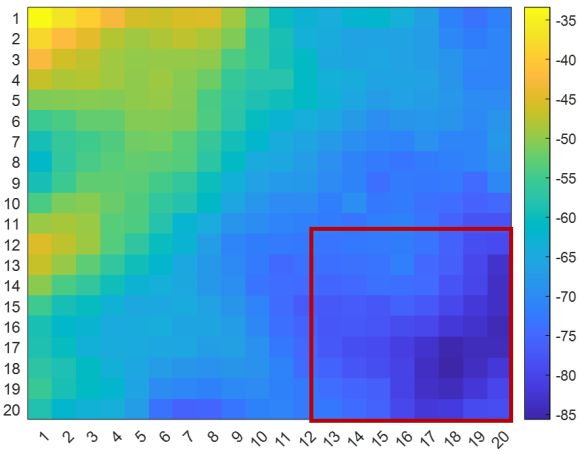
(a) proposed ($\delta = 0.5$)



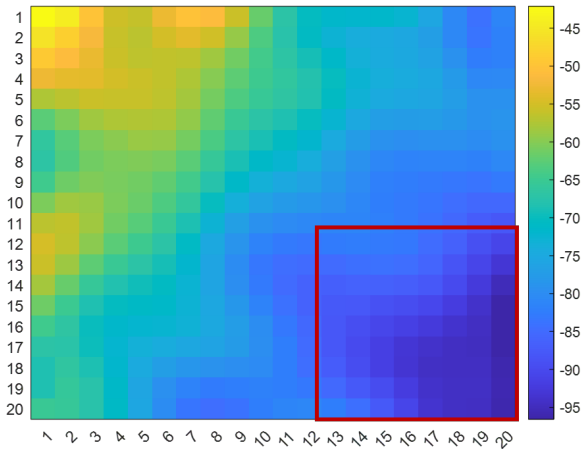
(b) AST ($\delta = 0.5$)



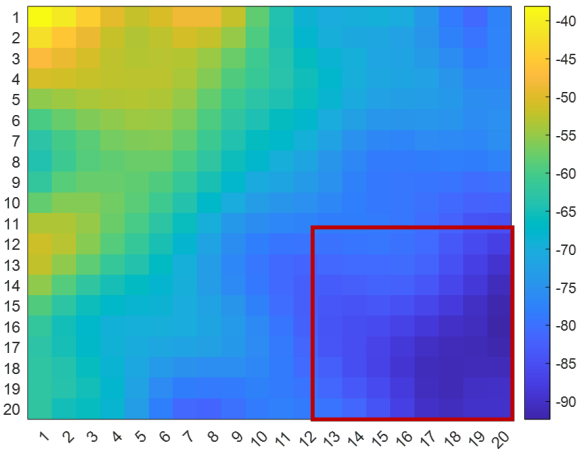
(c) proposed ($\delta = 0.7$)



(d) AST ($\delta = 0.7$)

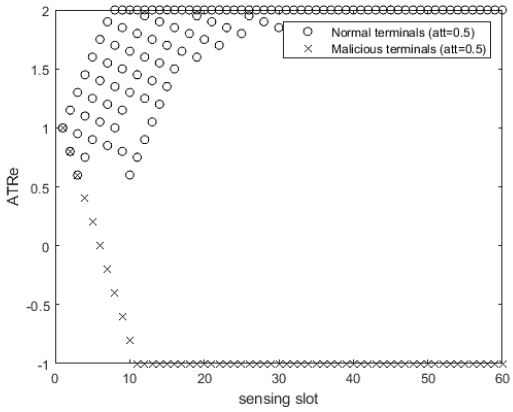


(e) proposed ($\delta = 0.9$)

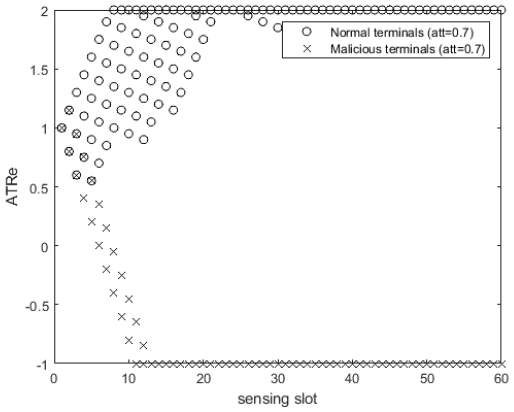


(f) AST ($\delta = 0.9$)

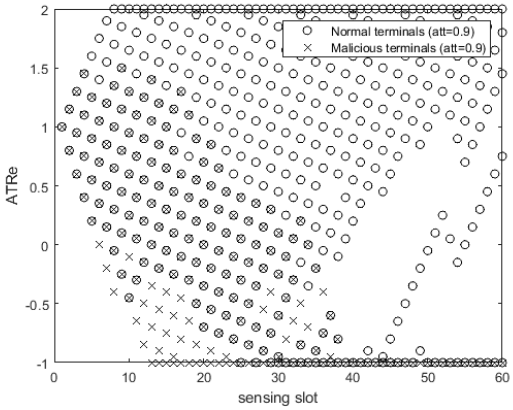
Figure 4.3 REM [dBm] under different attacking strength



(a) $\delta = 0.5$



(b) $\delta = 0.7$



(c) $\delta = 0.9$

Figure 4.4 $ATRe$ under different attacking strength.

The maps generated under different attacking strengths are shown in Fig.4.3. The figures in the left column show the maps generated by our proposed method, while those in the right column show the maps generated by AST/AT. The figures in the left column are similar because even with different levels of attack, the algorithm generates a REM that excludes terminals with low accumulative total reputations. This iterative process effectively removes the malicious terminals and selects the reliable ones. On the other hand, the figures in the right column have clearer differences due to the interference from the malicious terminals. As the attacking strength weakens from top to bottom, the differences between the proposed method and AST also decrease.

We can check the general trend of $ATRe$ under different attacking strengths for each sensing slot in Fig.4.4. The approach is more efficient at detecting malicious terminals with $\delta = 0.5$ and $\delta = 0.7$ than with $\delta = 0.9$ since the former have a stronger attacking strength, a

bigger *Bias*, and are thus easier to identify. As can be shown in Fig.5.6(c), the system is able to identify malicious reports after 40 iterations, even though the malicious terminals only make little changes.

4.4.2 Impact of the amount of malicious terminals

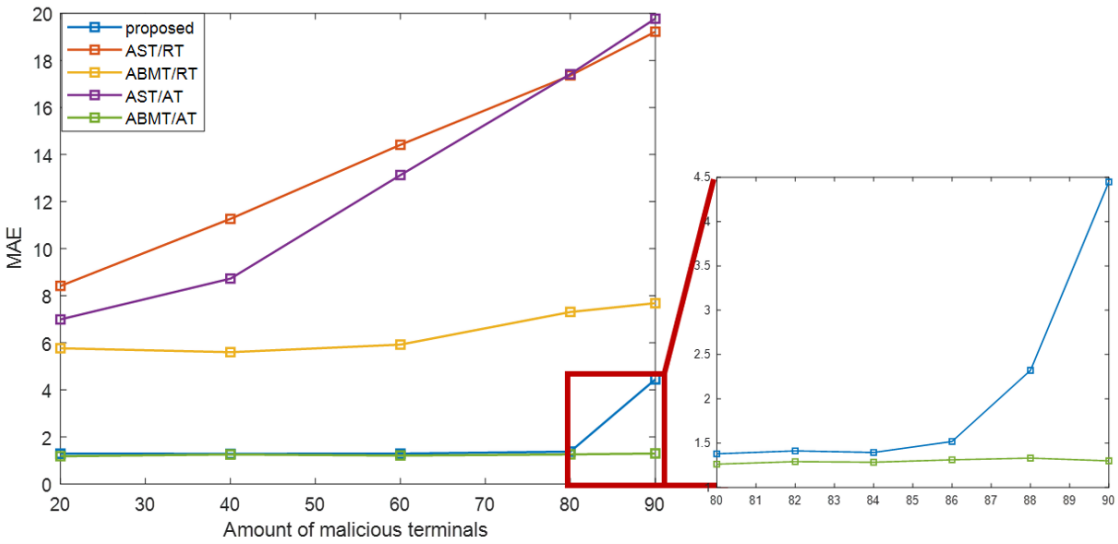
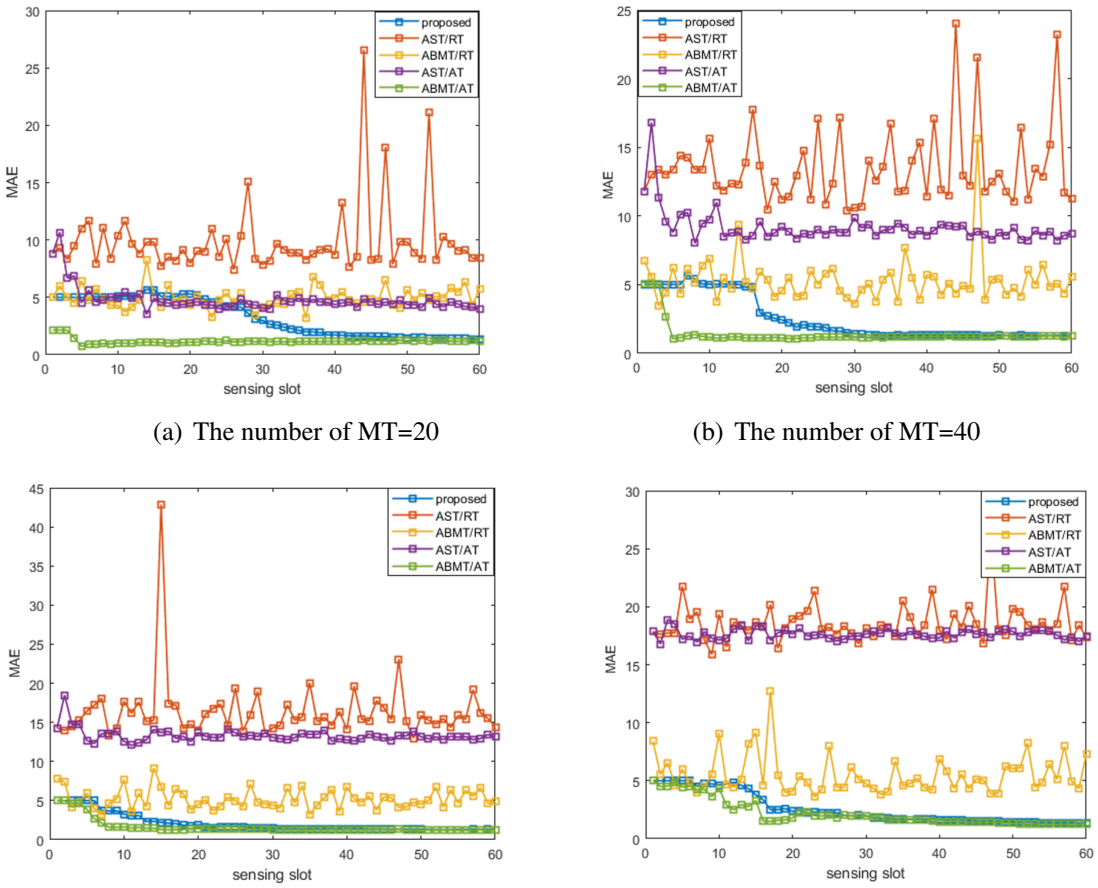


Figure 4.5 MAE under different amount of malicious terminals.

In this section, we examine the effect of increasing the number of malicious terminals from 20 to 90 out of 100. When the number of malicious terminals over 50, it indicates that the malicious terminals occupied more than half of the total terminals. Fig.4.5 depicts MAEs with varying amounts of maliciousness. As the number of malicious terminals increases, AST errors increase rapidly. Using accumulative information still results in a significant error, especially when more than half of terminals are malicious. In contrast, as the number of malicious terminals increases, the proposed algorithm's error slightly increases. When the number of malicious terminals is less than 84 out of 100, the MAE of our proposed method is very similar to the bound. The performance improved significantly after adding the trust nodes aided system, and by examining the performance as depicted in Fig.4.5, we can conclude that the system functions properly. In addition, Fig.4.6 displays the mean errors for each sensing slot. The convergence pace varied barely based on the variations in malicious terminals, and the algorithm can converge under various conditions within 50 rounds.

Fig.4.7 illustrates the trend of $ATRe$ under varying numbers of malicious terminals per sensing slot. In comparison to 20 malicious terminals, when the number of malicious terminals increases, convergence speed increases as well. This is the reason why Fig.4.6(a)



(a) The number of MT=20

(b) The number of MT=40

(c) The number of MT=60

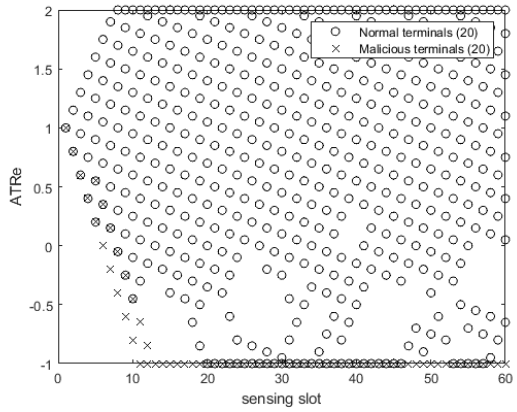
(d) The number of MT=80

Figure 4.6 MAE [dB] under different amount of malicious terminals.

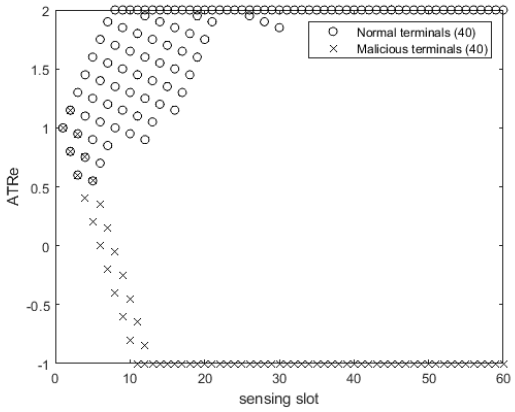
converges more slowly than in other situations: when the number of malicious terminals is small, the algorithm is able to eliminate them effectively, but with a higher probability of false alarms (mistakenly identifying normal terminals as malicious).

4.4.3 Impact of different amount of trust nodes

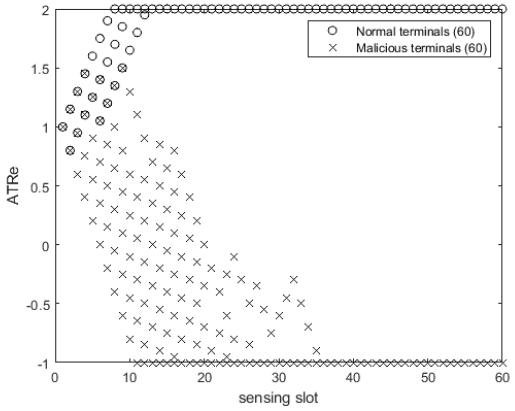
This section demonstrates the impact of increasing the number of trust nodes from 1 to 20. In this section, we set the attacking index to $\delta = 0.7$ and the number of malicious terminals to 40. Since trust nodes are distinct from normal terminals, estimation precision is fixed when using AST and ABMT. Similarly to what was stated previously, the ABMT utilizing accumulative total information only with honest terminals provides the highest achievable performance. Fig.4.8 demonstrates that by using a system aided by trust nodes, performance was significantly enhanced. And the more trust nodes we employ, the higher the system's



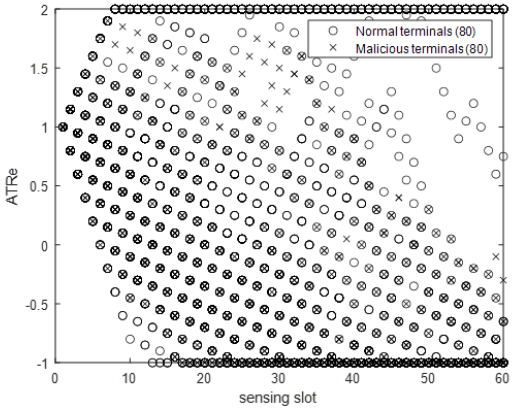
(a) Malicious terminals=20



(b) Malicious terminals=40



(c) Malicious terminals=60



(d) Malicious terminals=80

Figure 4.7 $ATRe$ under different amount of malicious terminals.

performance. However, when the number of trust nodes in the system increased from 10 to 20, the performance did not change significantly. Regard to the fact that adding trust nodes also increases the overheads of the network, the simulation results indicate that ten trust nodes are preferable. Fig.4.9 illustrates the trend of $ATRe$ for varying numbers of trust nodes. A round icon indicates a normal terminal's reputation and a cross indicates a malicious terminal's reputation. After adding trust nodes, the system can reliably separate terminals based on their status.

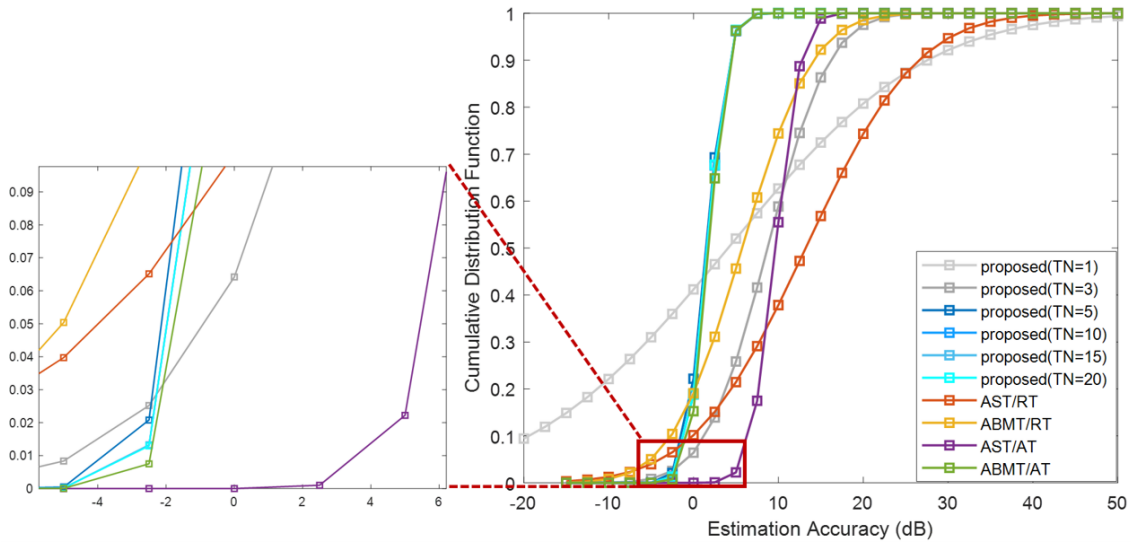
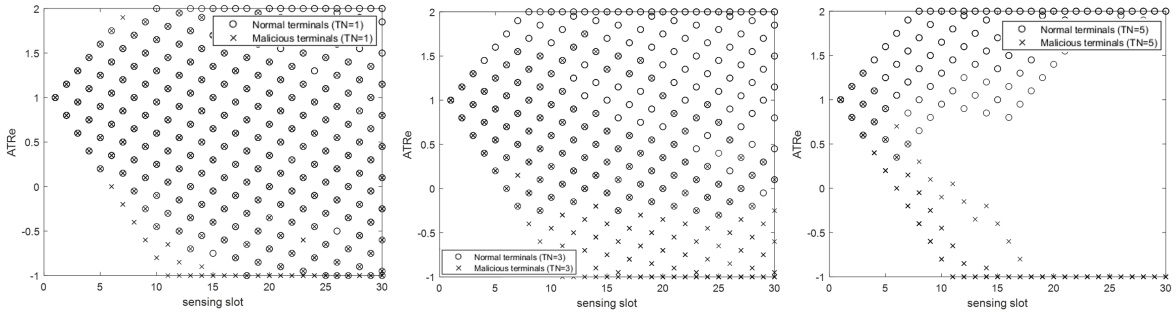
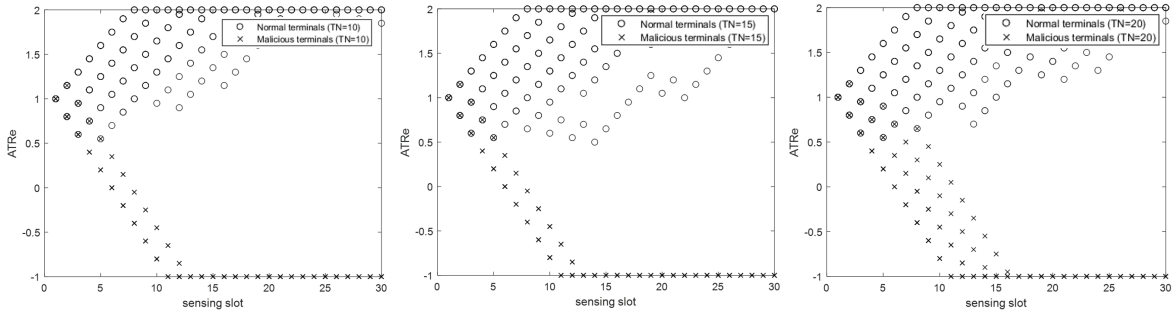


Figure 4.8 CDF under different amount of trust nodes.



(a) Trust nodes=1 (b) Trust nodes=3 (c) Trust nodes=5



(d) Trust nodes=10 (e) Trust nodes=15 (f) Trust nodes=20

Figure 4.9 $ATRe$ under different amount of trust nodes.

4.5 Chapter Summary

In this chapter, we introduced the design of the Kriging-based Trust Nodes aided REM construction method. By adding a small number of trust nodes, the system evaluates real-time comparisons and accumulative total reputations to select reliable datasets for constructing the REM in a threatening environment. The proposed method achieves high map accuracy by monitoring the cumulative behavior of even a small number of participants. Furthermore, the system maintains stable performance even when more than half of the terminals are malicious. Simulation results demonstrate the robustness and stability of the proposed method under varying conditions.

Chapter 5

A REM Construction and Channel Estimation System in Threatening Environments

An accurate REM can bring the CRN with lots of benefits. This chapter aims to propose a solution to the security threats in REM construction while ensuring accurate channel estimation. To achieve this goal, we introduce the Kriging-based trust nodes aided (KTNA) REM construction method, complemented with a channel estimation (KTNA+) system. By incorporating a small number of trust nodes, the proposed method mitigates the effects of malicious attacks and enhances the precision of channel estimation. Simulation results demonstrate the effectiveness of the proposed method in accurately estimating the average path-loss and shadowing impacts in threatening environments.

The arrangements of this chapter are as follows: the background of our research is represented in section 5.1, the system related model is shown in section 5.2, and our channel estimated model is as section 5.3 shown. Some simulation results are shown in section 5.4, finally, the summary of this chapter is in section 5.5.

5.1 Background

It is essential to comprehend how each application associated with the REM concept is determined by the use of location information in the system model and what type of quantity comprises the REM concept. In this study, we regard the REM to represent a physical quantity of the environment, namely the signal intensity, or radio signal power. By analyzing data from multiple terminals that perceive the environment and transmit it to a database, the REM

can be generated. The database then generates the REM for a particular communication area based on information such as the received signal strength, terminal ID, and terminal location, among other things. As the precision of REM construction has a direct impact on spectral efficiency, it is crucial that terminals report accurate data to the database.

Due to the open nature of networks, a number of security risks put the accuracy of data reporting at risk. For instance, even though several terminals worked together to generate a portion of the map, the database is still susceptible to being compromised by information sent by egotistical or dishonest terminals to meet their needs (occupying the channel or causing a significant amount of interference from licensed terminals). The accuracy of REM is considerably decreased by this kind of attack, also called as a Byzantine attack or a SSDF attack.[29, 40–47].

Reputation-based algorithms are key strategies for dealing with threatening terminals, and anti-attack studies have recently attracted a lot of interest. By rewarding or punishing the terminals to differentiate between their various statuses, they update each terminal's reputation [42–45]. Using a classifier to identify malicious terminals is another prevalent trend; researchers have trained classifiers to identify and filter malicious terminals during sensing using SVM [46]. Although the primary terminal's location and transmission activity are known, these studies do not apply to REM construction because it is necessary to assess the signal intensity at each important point.

Additionally, using a crowdsourcing-assisted radio map, which has been the subject of research in recent years [56][64], is a typical way to obtain high-precision maps. The average power is then calculated using the location-based data that the database initially gathers in an adequate amount. By figuring out the average receiving power at each place, the REM may be made to be very near to the actual situation. Large-scale sensor networks are expensive to operate and maintain in practice, and a lack of environmental knowledge can result in critical mistakes while building REMs.

The spatial statistical characterisation of the radio environment, which is challenging due to the numerous wireless channel propagation mechanisms, is directly related to the applicability of spatial predictions to wireless communication systems [87][88]. These mechanisms have an impact on the signals that are received, which has an impact on the REM generating process. As a result, processing geo-located observations to accurately produce coverage maps is challenging.

We enhanced our Kriging-based trust nodes aided (KTNA) REM generation method into a channel estimation (KTNA+) system to address these issues. This system generates a high-precision radio map utilizing a minimal number of terminals in such dangerous areas.

5.2 System Description

5.2.1 REM model

To illustrate the REM for the communication space, we suppose that the spectrum data is gathered in the communication area using a number of terminals. The locations of the terminals, including both harmless and malicious terminals, are determined at random based on their random movements. Additionally, we place a few trust nodes in the communication area whose positions are fixed and whose information is always trustworthy. The terminal's ID, location, time, frequency, and power can all be collected by the terminal and sent to the database. The database can be setup in the cloud or at a base station, where it can store vast amounts of data. The REM is then produced based on these datasets. Fig.4.1 depicts a concept of the conventional REM.

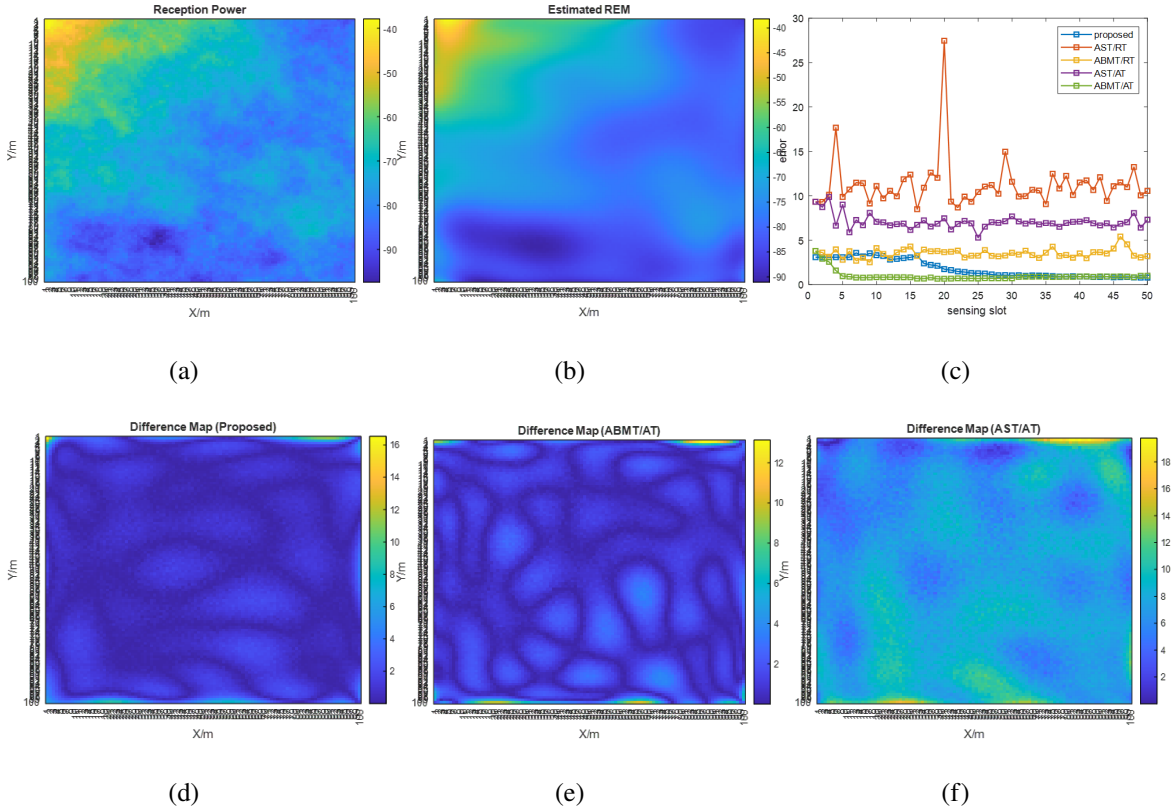
5.2.2 Reputation model

The KTNA-REM construction approach, which augments anti-attacking with a few trust nodes, is described in section 4.3. Keep in mind that the trust node's location is fixed, and in order to reduce overhead, trust nodes only need to work during the initial slot for sensing. After that, they are free to remain silent. The reports from trust nodes are stored directly in the trustset.

In order to ensure safety, we also think about the reboot phase. After sorting the accumulative reliability, if the top reliable nodes have a large $Bias$ value, this indicates that the environment may change while sensing (for example, it may rain or snow). In this situation, we must reboot the system, which entails clearing the trust set and launching the trust nodes for sensing again.

Fig.5.1(a) and 5.1(b) illustrate the KTNA-based realization and estimation maps, respectively. Fig.5.1(c) illustrates how the estimated mean error of the KTNA method varies with sensing slot. We deploy 100 sensing terminals in the communication zone, of which 30 are malicious. Ten nodes of trust are equitably spread throughout the entire communication area. We evaluate the performance against the next four strategies which we introduced in section 4.4.

The results of the simulations demonstrate the effectiveness of our KTNA method in removing malicious terminals and estimating REM. After the algorithm's convergence, the performance infinitely approaches the upper bound. The difference map between the real reception power and the estimated map produced by our KTNA method is shown in Fig.5.1(d).



(a) (b) (c)

(d) (e) (f)

Figure 5.1 Simulation results of reputation model [dB]: (a) Real Reception Power. (b) Estimated REM with the KTNA method. (c) Average error vary with sensing slot. (d) Difference Map: difference between the real reception power and the estimated map. (e) Difference REM: difference between the real reception power and the ABMT/AT based estimated map. (f) Difference REM: difference between the real reception power and the AST/AT based estimated map.

Fig.5.1(e) and 5.1(f) are the difference maps based on the two comparison methods, with the former being the upper bound of what can be achieved.

It is crucial to emphasize that these conclusions only apply to a single realization of the fix-attacking procedure. More attacking techniques and performance analyses are provided in section 5.4. With the help of our KTNA algorithm, we can quickly determine how the malicious terminals are affecting the overall communication network in this subsection. We introduce the channel estimation approach in the part after that to evaluate the effects of path loss and shadowing in a safe environment.

5.3 Adjusted Channel Estimation

In this section, we discussed the radio environment's interpretation as a spatial random process denoted by Q. The process is composed of two fundamental components, and it is defined for all spatial locations $\boldsymbol{s} = (x, y)$. The representation of $Q(\boldsymbol{s})$ can be expressed as follows,

$$Q(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \xi(\boldsymbol{s}), \text{with } \boldsymbol{s} \in D, \tag{5.1}$$

where $\mu(\boldsymbol{s})$ denotes the spatial process drift, which is often known as the trend in geo-statistical terminology, $\xi(\boldsymbol{s})$ consists of the zero-mean spatial random fluctuations of the spatial model, and the two-dimensional spatial domain D indicates the coverage region of interest.

In this research, $\xi(\boldsymbol{s})$ represents the shadowing effects of the wireless channel, whereas $\mu(\boldsymbol{s})$ represents path-loss in the wireless communication environment. The random variables $Q(\boldsymbol{s_i})$ collected to form the spatial random process $Q(\boldsymbol{s})$ are measurements of reception power, whose statistical properties depend on $\mu(\boldsymbol{s})$ and $\xi(\boldsymbol{s})$. Therefore, the aim of spatial estimation is to estimate the impact of these components from the relevant measurement data.

5.3.1 Path-loss estimation

The mean component $\mu(\boldsymbol{s})$ in trend modeling is based on the average path-loss of the wireless channel. It is determined by the average received signal power and can be expressed as follows,

$$\mu(\boldsymbol{s}) = P_{T_x} - \sum_{k=1}^{p} a_k f_k(\boldsymbol{s}), \tag{5.2}$$

where P_{T_x} [dBm] represents the transmission power, $f_k(\boldsymbol{s})$ represents the model's base function, and a_k represents the unknown constant trend coefficients. This chapter considers the well-known log-distance model for the path-loss model, which can be represented as [89],

$$L(d_{\boldsymbol{s_{T_x}},\boldsymbol{s}}) = L_0(d_0) + 10\log_{10}(\frac{d_{\boldsymbol{s_{T_x}},\boldsymbol{s}}}{d_0})^{\eta}, \tag{5.3}$$

d_0 [m] is the reference distance, also known as the critical distance, η is the path-loss index, and L_0 is the free-space path-loss, and it can be calculated as follows,

$$L_0(d_0) = 10\log_{10}(\frac{4\pi d_0}{\lambda})^2, \tag{5.4}$$

where λ [m] is the signal's wavelength. Using the model's base function, the average path-loss can be expressed as,

$$\mu(\boldsymbol{s}) = P_{T_x} - \eta f(\boldsymbol{s}) = P_{T_x} - 10\eta \log_{10}(d_{\boldsymbol{s_{T_x}},\boldsymbol{s}}). \tag{5.5}$$

5.3.2 Spatial estimation

The spatial random fluctuation exponent $\xi(\boldsymbol{s})$ in the wireless communication system affects the received signal power, as shown in the model presented in Equ.(5.1). This is due to the presence of obstacles such as buildings, mountains, and other obstructions that surround the terminals. These obstacles can cause random changes in the signal broadcast system. As the location, size, and dielectric characteristics of these obstructions are not known beforehand, statistical models that can characterize such fluctuations become crucial.

In our study, the random fluctuation is characterized by log-normal shadowing, and the model can be stated as [90],

$$Q(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \xi(\boldsymbol{s}) = P_{T_x} - 10\eta \log_{10}(\boldsymbol{s_{T_x}}, \boldsymbol{s}) + \xi(\boldsymbol{s}), \tag{5.6}$$

where $\xi(\boldsymbol{s})$ represents the shadowing at location $s(x_i, y_i)$, which is a zero-mean Gaussian spatial random process with standard deviation σ. Additionally, the covariance between two different random variables can be denoted as,

$$C(\boldsymbol{s_i}, \boldsymbol{s_j}) = \mathbb{E}\{[Q(\boldsymbol{s_i}) - \mu(\boldsymbol{s_i})][Q(\boldsymbol{s_j}) - \mu(\boldsymbol{s_j})]\}, \tag{5.7}$$

where $\mathbb{E}\{\cdot\}$ represents the expectation operator. The formula for the semivariogram function is,

$$\begin{aligned} \gamma(\boldsymbol{s_i}, \boldsymbol{s_j}) &= \frac{1}{2}\mathsf{Var}\{Q(\boldsymbol{s_i}) - Q(\boldsymbol{s_j})\} \\ &= \frac{1}{2}\mathbb{E}\{[Q(\boldsymbol{s_i}) - Q(\boldsymbol{s_j}) - \mathbb{E}\{Q\boldsymbol{s_i} - Q(\boldsymbol{s_j})\}]^2\} \end{aligned}. \tag{5.8}$$

Note that the received signal power is influenced by the location of the receiving terminals and the path-loss trend affects the stationarity of the process. Therefore, it is essential to estimate the trend component first. Once the trend component is estimated, the covariance and semivariance between any two points can be computed based on their separation vector h where $h \equiv \boldsymbol{s_i} - \boldsymbol{s_j}$. As a result, the covariance and semivariogram can be expressed as functions of the separation vector h only. Thus the covariance and semivariogram can be re-written as,

$$\begin{cases} C(h) = \mathbb{E}\{[Q(\boldsymbol{s_i}+h)-\mu][Q(\boldsymbol{s_i})-\mu]\}, & \forall \boldsymbol{s_i}, \boldsymbol{s_i}+h \in D, \\ \gamma(h) = \frac{1}{2}\mathsf{Var}\{Q(\boldsymbol{s_i}+h)-Q(\boldsymbol{s_i})\}, & \forall \boldsymbol{s_i}, \boldsymbol{s_i}+h \in D. \end{cases} \tag{5.9}$$

The Gudmundson model is commonly used to characterize the correlation of shadowing and belongs to the class of correlation models that consider the separated distance h between received data sets. Empirical studies and measurement campaigns have demonstrated that the spatial correlation of the log-normal shadowing in wireless communication channels decays exponentially with distance. Therefore, in this study, an isotropic-exponential structured covariance model is selected for constructing the spatial random process, which is given by [90],

$$C(|h|) = me^{-\frac{|h|}{r}}, \tag{5.10}$$

where m indicates the variance of the spatial random process which is called *sill variance*, and r here is called as *range* and reflects the exponential decay of the covariance function. Since the covariance function and the semivariogram function have a direct relationship under the wide sense stationarity condition, which is $C(\boldsymbol{s_i}, \boldsymbol{s_j}) = C(0) - \gamma(\boldsymbol{s_i}, \boldsymbol{s_j})$. So the semivariogram model can be written as,

$$\gamma(|h|) = m\left\{1 - e^{-\frac{|h|}{r}}\right\}. \tag{5.11}$$

Since the estimation of the power at the unknown location needs to depend on the ability of the environment, the semivariogram evaluation is important for the REM construction.

5.3.3 Adjusted REM construction

To clarify, as previously mentioned, the estimation approach presented in Equ.(5.1) involves estimating the path-loss and shadowing maps separately before combining them to generate the overall map. The reason for estimating the path-loss map first is due to its impact on the stationarity of the spatial random process, which in turn can affect the spatial estimation and semivariogram, leading to biased results. Therefore, the path-loss is estimated first, and based on the estimated model, the measurement dataset is detrended. This means that the data used in the spatial estimation and semivariogram are given by,

$$Q'(\boldsymbol{s}) = Q(\boldsymbol{s}) - \mu(\boldsymbol{s}). \tag{5.12}$$

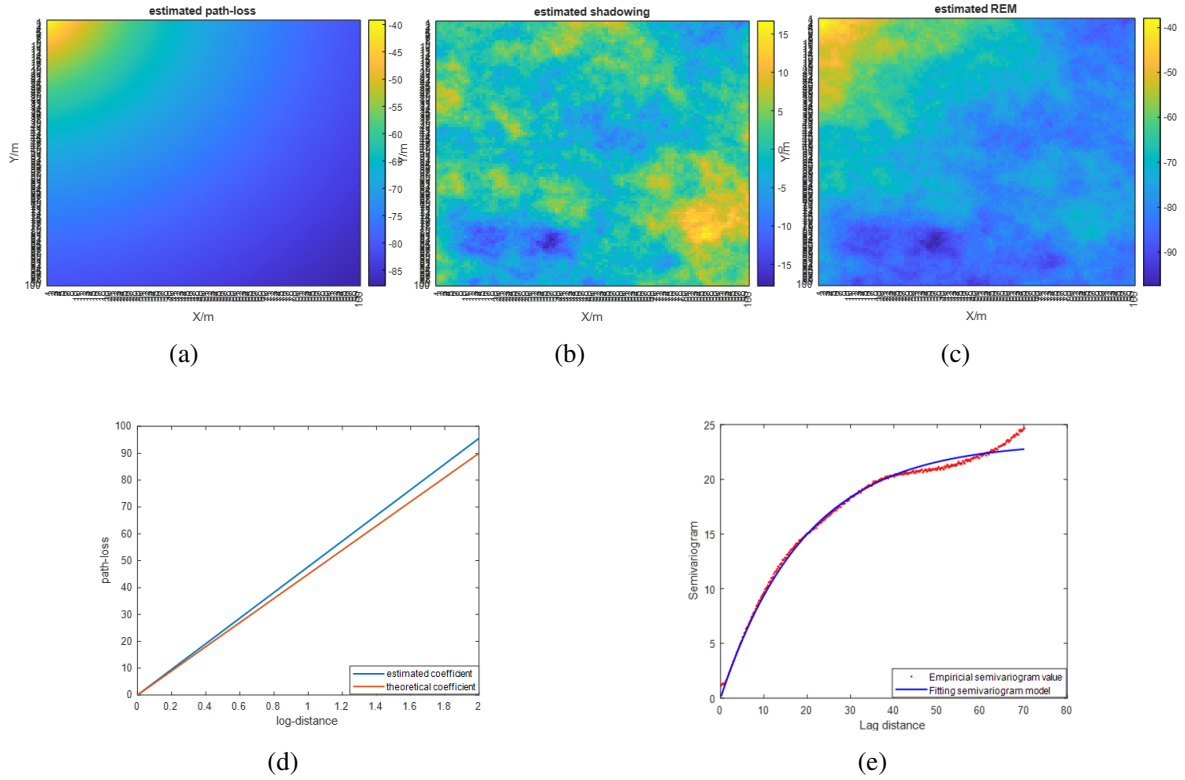


Figure 5.2 Simulation results of channel estimation [dB]: (a) Estimated path-loss map. (b) Estimated shadowing map. (c) Estimated REM. (d) Estimated path-loss index. (e) Experimental semi-variogram.

The experimental semivariogram is produced from the regionalized measurements using the moments of Matheron approach, which is given by [91],

$$\hat{\gamma}(h) = \frac{1}{2N_h} \sum_{\boldsymbol{s_i}-\boldsymbol{s_j}=h} [Q'(\boldsymbol{s_i}) - Q'(\boldsymbol{s_j})]^2, \forall \boldsymbol{s_i} \in D, i = 1,2,...,N, \tag{5.13}$$

where N_h is the number of data pairs whose distance from the location is h. In reality, N_h must be suitably large because measurements must be taken for each h, however, the set of measurements is finite. To address this issue, the semivariogram can be estimated via pre-calculated separate distances, known as lag distance.

The spatial estimation model's simulation results are displayed in Fig.5.2. The average path-loss impact is depicted in Fig.5.2(a) and the estimated shadowing impact is depicted in Fig.5.2(b), The combination of these two maps results in the spatial random process depicted in Fig.5.2(c). Fig.5.2(d) displays the estimated path-loss index, the estimated $\eta = 4.77824654048242$, which the reference index $\eta_0 = 3.5$, the semivariogram is shown as Fig.5.2(e), the upper bound parameter $m = 23.5223582293268$ is referred to as *sill*, and

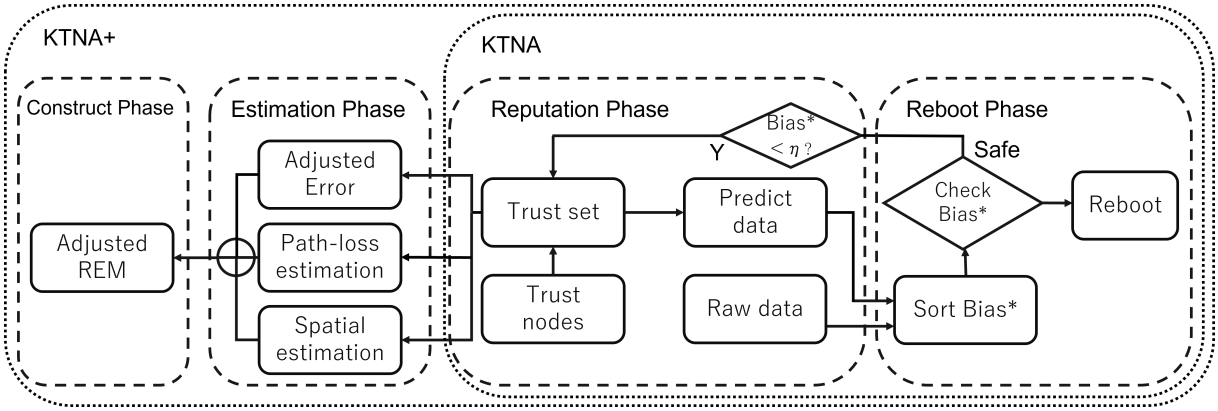


Figure 5.3 The structure of KTNA+ system.

can be used to measure shadowing intensity. $r = 19.7715335842846$ is referred to as the *range* parameter. In particular, the spatial correlation is stronger for data separated by smaller distances than r. *Range* is proportional to the semivariogram's slope and can be interpreted as the correlation distance $d_{cor} = 20$ [m] of the shadowing. Parameter $c_0 = 0$ indicates a sudden change in the spatial variability of the random process at small separation distances, which may be influenced by measurement errors, local variability, or superimposed noise.

As shown in Fig.5.2, although the algorithm can estimate the channel condition well, it still has some errors. In order to proceed with a more accurate REM, we calculate the combination map of the Adjusted Error (AE) based on the dataset as follows,

$$\begin{aligned}\epsilon &= \frac{\sum_{i=1}^{N}(Q(\boldsymbol{s}) - \hat{Q}(\boldsymbol{s}))}{N} = \frac{\sum_{i=1}^{N}(Q(\boldsymbol{s}) - (\hat{\mu}(\boldsymbol{s}) + \hat{\xi}(\boldsymbol{s}))}{N}) \\ &= \frac{\sum_{i=1}^{N}(Q(\boldsymbol{s}) - (P_{T_x} - 10\hat{\eta}\log_{10}(d_{\boldsymbol{s_{T_x}},\boldsymbol{s},\boldsymbol{s}}) + \hat{\xi}(\boldsymbol{s})))}{N}.\end{aligned} \tag{5.14}$$

Here, $\boldsymbol{s} = (x_i, y_i)$ is the measurement location, $\hat{\mu}$ and $\hat{\xi}$ is the estimated path-loss impact and the shadowing impact, respectively. Then we rewrite the Equ.(5.1) as,

$$\hat{Q}(\boldsymbol{s}) = \hat{\mu}(\boldsymbol{s}) + \hat{\xi}(\boldsymbol{s}) + \epsilon, \tag{5.15}$$

where, $\hat{Q}(\boldsymbol{s})$ is the adjusted REM, also we named the whole system based on the reputation model and adjusted channel estimation model as KTNA+ system, and the structure is like Fig.5.3. Note that, once the KTNA convergences, the KTNA+ system can store the spatial estimated impact, as long as we get the random spatial information, the secure system can be simplified as a linear regression problem.

Table 5.1 Simulation parameters.

Parameter	Value
Communication area [m^2]	500×500
Mesh size [m^2]	5×5
The number of meshes	10000
The number of terminals	100
Center frequency [GHz]	3.5
Transmission power [dBm]	29
Reference distance [m]	10
Path-loss index η	3.5
Standard deviation of W	6

5.4 Simulation Results

In this section, we present the results of several simulations conducted using the proposed method. The simulations were carried out using MATLAB R2022a. We considered an area of interest for communication and placed 100 terminals in this area. All terminals were assumed to have the same moving speed and a random direction within the area. The parameters in this part are shown as TABLE 5.1.

5.4.1 Different amount of malicious terminals

The number of malicious terminals increases from 20 to 40 out of 100 in this subsection. We examine the CDF of the MAE as depicted in Fig.5.4, which is a crucial performance measurement. MAE can be calculated as Equ.(3.22).

In the reputation phase, we compared our proposed method with two secure methods, the similarity degree method (SimD) and Histo based reputation method, which are extensively used in the spectrum sensing field to eliminate the impact of malicious terminals. Typically, these methods use the power similarity of neighboring nodes for malicious node exclusion, and the average power of normal terminals is used to update historical information [40]. In addition, because the output of the first two methods is based on the average power of the mesh, in order to obtain a higher quality REM, we examined two widely used interpolation methods, IDW and Kriging.

As Fig.5.4 shows, the proposed method has a better performance than the traditional average power output methods. The performance decreases when the amount of malicious

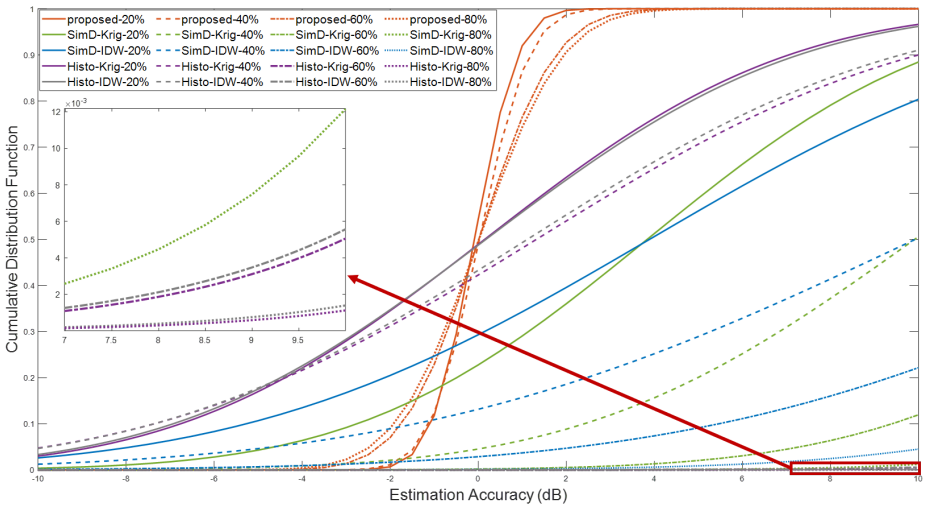


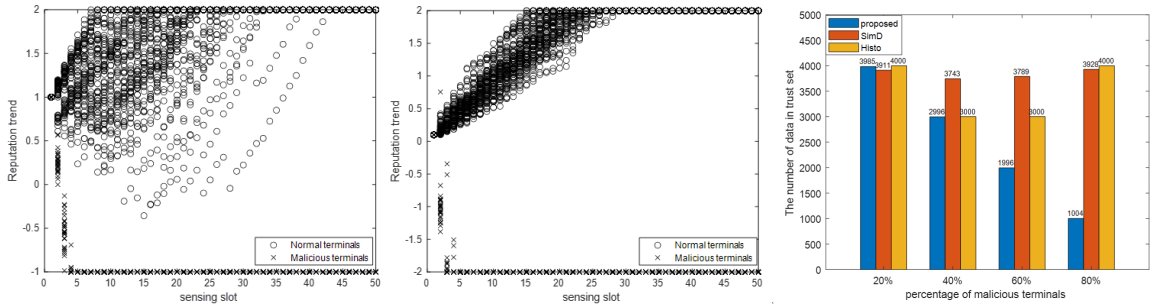
Figure 5.4 CDF under different amount of malicious terminals.

terminals decreases. After doing the interpolation and adding the spatial information, the performance improved for the Histo and SimD methods.

Additionally, Fig.5.5(a) and Fig.5.5(b) show the reputation trend of each terminal when the amount of malicious is 30 out of 100 under the proposed method and Histo based method, respectively. The reputation can be calculated as Equ.(4.13). Both of these methods are effective at identifying malicious terminals and rejecting them from the database. Additionally, the Histo-based method has a stronger reward performance than the proposed method, but the estimated map is not as accurate as ours. The efficacy of the proposed method is superior to that of traditional REM construction techniques due to the elimination of the need to calculate the average power of each mesh. Instead, interpolation is performed using the raw data from the normal terminals. SimD and Histo-based methods must compute the average power during their procedure; despite having a decent reputation for removing malicious terminals, they still result in varying degrees of information loss when constructing the REM.

Truly, it is possible to eliminate the loss of information by ignoring the outputted average power and using the raw information defined as the normal terminals to interpolation, however, the computational requirements can be quite significant. Fig.5.5(c) depicts the quantity of data stored by the trust set. As indicated by the blue color bar, the trust nodes assisted phase determines that the database can store fewer records for the intended purpose.

The performance of the estimation phase is presented in Table 5.2, and lines 2 through 7 indicate the MAE of the estimated RSSI when using various methodologies. $\hat{\eta}$ represents the estimated path-loss index, whereas $\hat{d}_{cor}$ [m] represents the estimated correlation distance.



(a) (b) (c)

Figure 5.5 The Comparison of different methods: (a) The reputation trend under proposed method. (b) The reputation trend under Histo-based method. (c) The selected number of different method.

Table 5.2 Performance under different amount of malicious terminals.

Methods	Parameters	20%	40%	60%	80%
AST	MAE_RSSI	5.262156	10.53644	15.88108	21.19951
SimD-Krig	MAE_RSSI	5.179132	10.10587	16.79927	22.42956
SimD-IDW	MAE_RSSI	6.144702	10.79592	17.00903	22.47131
Histo-Krig	MAE_RSSI	4.116176	4.825126	25.69812	26.22508
Histo-IDW	MAE_RSSI	4.233108	4.785248	25.83612	26.31063
KTNA+	MAE_RSSI	**0.593164**	**0.678816**	**1.064683**	**1.114734**
	MAE_W	0.80597	0.893773	1.238783	1.274118
	$\hat{\eta}$	3.44829	3.44859	3.44543	3.44437
	$\hat{d}_{cor}$	20.7715	18.7074	17.6088	16.3715

According to the table, the traditional methodologies (SimD and Histo) for estimating RSSIs have large discrepancies with the ground truth. In contrast, our proposed method KTNA+ system performs well in channel estimation and REM construction.

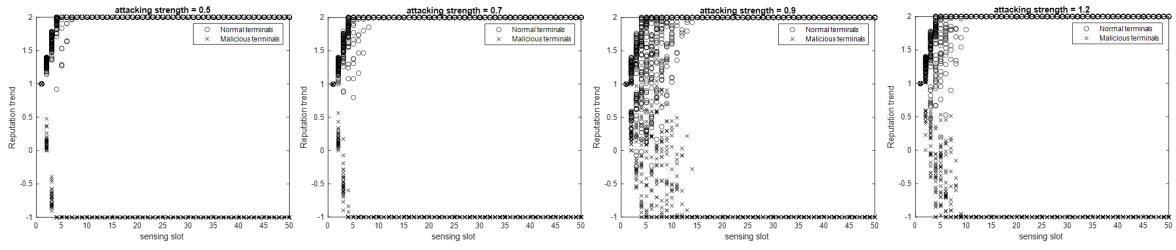
5.4.2 Different attacking strength

Impact on various levels of attacking strength is evaluated here. The percentage of malicious terminals is set at 30 while the percentage of trust nodes is set at 10. According to the definition in section 4.2.3, the attack's power is determined by the attacking index δ. It's simple math: if δ is less than 1, the attacker's strength drops when δ increases; if it's more than 1, the attacker's strength rises when δ increases.

Table 5.3 Performance under different attacking strength.

Methods	**Parameters**	$\delta = 0.5$	$\delta = 0.7$	$\delta = 0.9$	$\delta = 1.2$
AST	MAE_RSSI	13.13383	7.87430	2.656462	5.291361
SimD-Krig	MAE_RSSI	11.51336	7.359291	4.610962	5.863282
SimD-IDW	MAE_RSSI	12.42125	8.348438	4.829906	6.491345
Histo-Krig	MAE_RSSI	4.248715	4.083502	4.63966	4.090303
Histo-IDW	MAE_RSSI	4.4088	4.261046	4.952128	4.19464
KTNA+	MAE_RSSI	**0.674949**	**0.674776**	**0.675088**	**0.674051**
	MAE_W	0.88657	0.886485	0.892108	0.880136
	$\hat{\eta}$	3.4496	3.45004	3.45048	3.44994
	$\hat{d}_{cor}$	19.3911	19.5169	19.4004	19.2658

Table 5.3 displays the MAE as well as certain specific performances. The theoretical average path loss in the environment is $\eta = 3.5$ and the correlation distance is $d_{cor} = 20$ [m] [92]. When the attacking strength is high (δ is away from one), KTNA is better able to identify malicious terminals, leading to more precise estimates. Aside from that, the malicious information is too close to the actual information for KTNA to make a decision without further data. Our suggested technique outperforms competing algorithms in two key areas: minimizing the impact of malicious attacks and generating an accurate environment estimate over a range of attack intensities. Even if malicious information is removed, the massive fusion of information causes the output data to lose the texture structure contained in the information itself, which has a significant effect on the estimation of the channel environment, leading to a larger error for the SimD and Histo methods.



(a) (b) (c) (d)

Figure 5.6 The performance under different attacking strength: (a)$\delta = 0.5$. (b) $\delta = 0.7$. (c) $\delta = 0.9$. (d) $\delta = 1.2$.

Fig.5.6 depicts the progression of the reputation. Equ.(4.13) determines whether a node's reputation should improve or decrease based on whether or not its normalized $Bias_k^*$ is below the threshold value, at which point it is assumed to have a high likelihood of being a

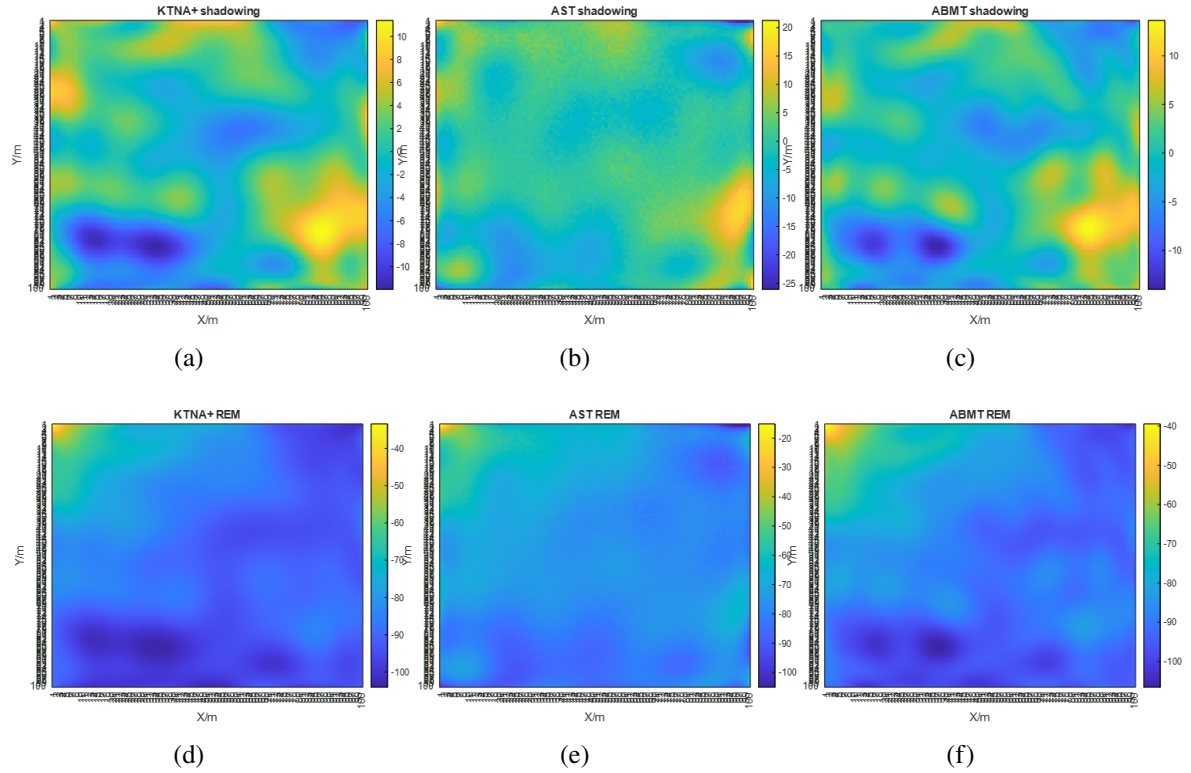


Figure 5.7 The performance of the different maps [dB]: (a)Estimated shadowing by KTNA+. (b) Estimated shadowing by AST. (c) Estimated shadowing by ABMT. (d) Estimated REM by KTNA+. (e) Estimated REM by AST. (f) Estimated REM by ABMT.

normal terminal. Here, we adjust the reward index to $a = \zeta - Bias_k^*$, where ζ is the threshold. Furthermore, $b = Bias_k^*$ is the penalty index. This means that the standing of any terminal might rise or fall rather quickly. As can be seen in Fig.5.6, when the attacking index $\delta < 1$, as in Fig.5.6(a)-5.6(c), the attacking strength is weaker when δ approaches to one, the difference between the estimated power and the reported power is less, and the database needs more rounds to distinguish the status of the terminals, so the reputation of Fig.5.6(c) rises or falls more slowly. When the attacking index is greater than one, the attacking strength increases as δ far away from one. That's why Fig.5.6(d)'s reputation shifts are picking up speed once again.

Note that the ABMT approach may be thought of as the upper limit that the algorithm can reach; the difference between the AST maps and the ABMT maps can be thought of as the effect of the malicious terminals. The estimated map under $\delta = 0.5$ is shown in Fig.5.7. The predicted shadowing effect is shown in Fig.5.7(a)-Fig.5.7(c). Fig.5.7(c) is the ideal way for estimating the effects of shadowing since it takes into account only trustworthy pieces of data while immediately discarding anything malicious. Since Fig.5.7(b) incorporates all

available data, including malicious terminal information, into its shadowing effect estimate, it stands in contrast to Fig.5.7(c). Our proposed method's projected shadowing effect is shown in Fig.5.7(a). We discover that our method's performance is comparable to that of the Fig.5.7(c), which indicates that it is able to identify malicious terminals and provide an accurate assessment of the shadowing effect.

The simulated results of the estimated REM are shown in Fig.5.7(d)-Fig.5.7(f). The significance of the malicious terminals' impact is shown by contrasting Fig.5.7(e) with Fig.5.7(f). Analyzing the differences between Fig.5.7(d) and Fig.5.7(f), the simulation results indicate our method reduces the impact of malicious terminals on the environment effectively.

5.5 Chapter Summary

In this chapter, we introduced the design of the KTNA+ system, which aims to achieve a secure REM construction and channel estimation in a threatening wireless communication environment. Our approach involves the incorporation of trust nodes in the system to ensure the rejection of malicious data from the database by awarding or punishing the terminals in the interest area. We also developed a method for estimating the channel conditions, taking into account the path-loss and shadowing effects in the wireless communication area. Through extensive simulations, we validated the effectiveness of our proposed system in constructing the REM and accurately estimating the channel under various attack scenarios.

Chapter 6

Conclusions and Future Scopes

This chapter marks the conclusion of our research on highly efficient and secure REM construction. To begin, we provide a summary of the contributions and findings from each preceding chapter. Following this, potential future research directions are explored and discussed.

6.1 Conclusions

In order to improve the quality of wireless communication and meet the infinite growth of wireless communication demand, REM plays a crucial role in evaluating the wireless environment. However, when the relevant technology is applied to real-world applications, such as collecting information from various mobile phones, the absolute honesty of all users cannot be guaranteed, and it is difficult to prevent selfish users from attempting to monopolize spectrum resources or disrupt the primary user's communication. Motivated by the fact that the efficiency of the dynamic spectrum access strongly depends on the accuracy of the constructed REM, we have comprehensively investigated the highly secure REM construction methods under threatening environments satisfied with 1) distinguishing the malicious terminals and 2) improving the REM accuracy. As the comprehensive conclusion of the analysis, the individual methods appearing in this dissertation have shown the following accomplishments in meeting the two targets:

Our proposed DLM can deal with the condition in which the malicious terminals amount is less than normal terminals. Our method and bi_weight can easily distinguish malicious terminals and construct the REM when the amount of malicious terminals is less than 50%, however, the performance of similarity degree, and average_combination depends on the attacking index δ a lot, where, when δ is approach to one, attacking strength is weak, it is difficult for them to distinguish the difference between malicious terminals and normal

Table 6.1 Performance under different algorithms.

Algorithms	Target 1		Target 2		Complexity	CE
	$< 50\%$	$> 50\%$	$< 50\%$	$> 50\%$		
DLM	∘	△	∘	△	$O(n^2)$	−−
similarity degree	△	×	∘	×	$O(n^2)$	−−
bi_weight	∘	×	∘	×	$O(mn)$	−−
average combination	△	×	∘	×	$O(n)$	−−
AST/RT	−−	−−	×	×	$O(n^3)$	−−
ABMT/RT	−−	−−	∘	∘	$O(n^3)$	−−
AST/AT	−−	−−	×	×	$O(n^3)$	−−
ABMT/AT	−−	−−	∘	∘	$O(n^3)$	−−
KTNA	∘	∘	∘	∘	$O(n^3)$	−−
KTNA+	∘	∘	∘	∘	$O(n^3)$	∘

[1] ∘ : good performance
[2] × : bad performance
[3] △ : performance discusses by case
[4] −− : no performance
[5] CE : channel estimation capability

terminals, although they can meet target 2, it comes at the cost of a higher probability of false alarms.

When the amount of malicious terminals is over 50%, the comparison methods all lose their functions, since they do not have the memory to trace the past performance, when the malicious terminal occupies more than normal, they will mis-detect the normal as the malicious. Our proposed DLM has the historical reliability layer, we can target the malicious behavior and mark them, finally, in some mesh in which more malicious terminals passed by, still can maintain the performance. However, our method only works in the case where fewer meshes have more malicious information, if the total amount of malicious terminals is larger than normal, we also lose the precision.

AST/RT, ABMT/RT, AST/AT, and ABMT/AT do not have the performance about distinguishing malicious terminals, since they are references. All AST methods used all the information from the database, and all ABMT methods used all the normal terminals' information from the database. KTNA and KTNA+ can face all the targets since they add trust nodes in the surrounding environment, based on the sensing information from the trust nodes, we distinguish the malicious or normals. Even when the number of malicious terminals is over half, KTNA and KTNA+ still can distinguish the malicious and construct the REM precisely.

By the way, KTNA is the algorithm by compares with the trust nodes, selects reliable information, and uses reliable information to do interpretation and generate the REM directly. KTNA+ is the system, which uses reliable information to estimate the channel condition and then construct the REM. Both of them have good performances.

The common outlier and abnormal detection methods are mainly focused on eliminating values with large differences from others. The datasets are considered as independent, which means not time-related. Outlier detection is more suited for the condition in which the outliers are caused by unstable sensing, not malicious nature.

Different from the common outlier and abnormal, malicious terminals have the nature to destroy networks on purpose instead of unintentionally, and the performance can be monitored continuously. Especially, when the malicious reports in one mesh are larger than the normal reports, by using the outlier detection, it will judge the normal reports as outliers falsely. Also, when an extremely large outlier exists, it will affect other outlier detection. By using our proposed method, since we judge the error continuously, this could be solved.

As we mentioned in Section 1.1.2, REM can be used to make decisions in a variety of applications, including coverage optimization [19], resource allocation [22], interference analysis [23], location estimation [24][25], and so on. The accuracy of the REM is indeed required to be different for different applications, but certainly, an estimation error higher than 5 dB is definitely not allowed. Regardless of the intended use, most REM errors need to be kept below 3dB error or even less, however, as the results shown in our dissertation, without our methods, the error of the REM is quite high on average or in some certain mesh, especially when the rate of malicious terminals is high. It is shown that, as the REM error increases, interference is significantly affected in reference [23], and for 2dB error in REM, caused around 10% capacity decrease for the small cell users in reference [93].

Most studies presume that information provided by terminals is trustworthy, and our research precisely bridges this part of the gap and makes an important contribution to the comprehensive and wide range of practical applications in the future.

6.2 Future Scopes

The primary objective of this dissertation is to enhance the wireless communication quality by constructing the REM. We found the precision of the REM could decrease significantly when facing data falsification attacks. Additionally, there are several research areas that offer opportunities for future investigation. In the final section of this dissertation, we provide a brief overview of the outstanding issues and potential avenues for further research.

Toward a more flexible system for REM construction In this dissertation, we mainly concentrated on against the data falsification attack, due to it is easy to launch, and can have an important impact on the entire network system, and can even completely destroy the communication system, so in recent years, it has been the focus of scholars' research. However, as we mentioned in section 1.1.3, SSDF is not the only attack method that can be launched in the CRNs. More flexible attacks could happen in the communication network. Therefore, a system that can effectively identify attack methods and respond quickly to ensure the stability and security of the communication system needs to be studied.

Toward the completion of distributed networks In this dissertation, the fixed primary terminal operating within the centralized network was the main focus. However, given the prevalence of interconnected systems and devices today, studying distributed networks has become increasingly important. Distributed networks have the potential to increase extensibility, lower costs, and increase system efficiency. Distributed network research can also aid the ever-increasing need for connectivity as a result of the expanding number of devices.

Toward five-dimensional REM construction Although we assumed a two-dimensional radio environment in this research, the transmitter and receiver heights have a substantial impact on radio propagation characteristics. Additionally, we must take into account the activities in the time domain and various frequency bands. For example, the efficiency of sharing the spectrum can be increased through statistical analysis in the time domain, such as occupancy rate and transition rate. As a result, it is important to study the five-dimensional (longitude-latitude-altitude-time-frequency) model in the future.

Bibliography

[1] U. S. G. P. Office, "Secure cooperative sensing techniques for cognitive radio systems," in *Commerce Dept., National Telecommunications and Information Administration, Office of Spectrum Management*, May 2016.

[2] S. Bhattarai, J.-M. J. Park, B. Gao, K. Bian, and W. Lehr, "An overview of dynamic spectrum sharing: Ongoing initiatives, challenges, and a roadmap for future research," *IEEE Transactions on Cognitive Communications and Networking*, vol. 2, no. 2, pp. 110–128, June 2016.

[3] J. H. Reed, J. T. Bernhard, and J.-M. Park, "Spectrum access technologies: The past, the present, and the future," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1676–1684, May 2012.

[4] F. C. Commission, "Spectrum policy task force report," *Report of the Spectrum Efficiency Working Group*, 2002.

[5] M. A. McHenry, "Nsf spectrum occupancy measurements project summary shared spectrum co. report," 2005.

[6] F. C. Commission, "Spectrum policy task force report. technical report 02-135," *Spectrum Policy Task Force Report*, 2002.

[7] S. Barnes, P. Botha, and B. Maharaj, "Spectral occupation of tv broadcast bands: Measurement and analysis," *Measurement*, vol. 93, pp. 272–277, Nov. 2016.

[8] S. Yin, D. Chen, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: A large-scale spectrum measurement study," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 1033–1046, June 2012.

[9] D. S. Gurjar, H. H. Nguyen, and H. D. Tuan, "Wireless information and power transfer for iot applications in overlay cognitive radio networks," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3257–3270, April 2019.

[10] A. Das, N. Das, A. Das Barman, and S. Dhar, "Energy incentive for packet relay using cognitive radio in iot networks," *IEEE Communications Letters*, vol. 23, no. 9, pp. 1581–1585, Sep. 2019.

[11] L. Xu, W. Yin, X. Zhang, and Y. Yang, "Fairness-aware throughput maximization over cognitive heterogeneous noma networks for industrial cognitive iot," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4723–4733, Aug 2020.

[12] A. Paul, A. Daniel, A. Ahmad, and S. Rho, "Cooperative cognitive intelligence for internet of vehicles," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1249–1258, Sep. 2017.

[13] G. Rathee, F. Ahmad, F. Kurugollu, M. A. Azad, R. Iqbal, and M. Imran, "Crt-biov: A cognitive radio technique for blockchain-enabled internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4005–4015, July 2021.

[14] M. A. Hossain, R. M. Noor, K.-L. A. Yau, S. R. Azzuhri, M. R. Z'aba, and I. Ahmedy, "Comprehensive survey of machine learning approaches in cognitive radio-based vehicular ad hoc networks," *IEEE Access*, vol. 8, pp. 78 054–78 108, Apr. 2020.

[15] Z. Zhang, X. Wen, H. Xu, and L. Yuan, "Sensing nodes selective fusion scheme of spectrum sensing in spectrum-heterogeneous cognitive wireless sensor networks," *IEEE Sensors Journal*, vol. 18, no. 1, pp. 436–445, Jan 2018.

[16] I. Kakalou and K. E. Psannis, "Sustainable and efficient data collection in cognitive radio sensor networks," *IEEE Transactions on Sustainable Computing*, vol. 4, no. 1, pp. 29–38, Jan 2019.

[17] A. Sarthi, D. S. Gurjar, C. Sai, P. Pattanayak, and A. Bhardwaj, "Performance impact of hardware impairments on wireless powered cognitive radio sensor networks," *IEEE Sensors Letters*, vol. 4, no. 6, pp. 1–4, June 2020.

[18] H. B. Yilmaz, T. Tugcu, F. Alagöz, and S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," *IEEE Communications Magazine*, vol. 51, no. 12, pp. 162–169, December 2013.

[19] H. Braham, S. B. Jemaa, G. Fort, E. Moulines, and B. Sayrac, "Fixed rank kriging for cellular coverage analysis," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4212–4222, May 2017.

[20] J. Perez-Romero, A. Zalonis, L. Boukhatem, A. Kliks, K. Koutlia, N. Dimitriou, and R. Kurda, "On the use of radio environment maps for interference management in heterogeneous networks," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 184–191, August 2015.

[21] Y. Zhao, J. H. Reed, S. Mao, and K. K. Bae, "Overhead analysis for radio environment mapenabled cognitive radio networks," in *2006 1st IEEE Workshop on Networking Technologies for Software Defined Radio Networks*, Sep. 2006, pp. 18–25.

[22] E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed kriging," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1205–1211, March 2011.

[23] K. Sato and T. Fujii, "Kriging-based interference power constraint: Integrated design of the radio environment map and transmission power," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 1, pp. 13–25, March 2017.

[24] F. Gu, J. Niu, and L. Duan, "Waipo: A fusion-based collaborative indoor localization system on smartphones," *IEEE/ACM Transactions on Networking*, vol. 25, no. 4, pp. 2267–2280, Aug 2017.

[25] C. Wu, Z. Yang, and Y. Liu, "Smartphones based crowdsourcing for indoor localization," *IEEE Transactions on Mobile Computing*, vol. 14, no. 2, pp. 444–457, Feb 2015.

[26] B. A. Huberman, "Crowdsourcing and attention," *Computer*, vol. 41, no. 11, pp. 103–105, Nov 2008.

[27] Y. Hu and R. Zhang, "A spatiotemporal approach for secure crowdsourced radio environment map construction," *IEEE/ACM Transactions on Networking*, vol. 28, no. 4, pp. 1790–1803, Aug 2020.

[28] J. Li, Z. Feng, Z. Feng, and P. Zhang, "A survey of security issues in cognitive radio networks," *China Communications*, vol. 12, no. 3, pp. 132–150, Mar 2015.

[29] F. Ye, X. Zhang, and Y. Li, "Comprehensive reputation-based security mechanism against dynamic ssdf attack in cognitive radio networks," *Symmetry*, vol. 8, no. 12, 2016.

[30] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, April 1967.

[31] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 21–24, Jan 2007.

[32] G. Bharathy, V. Rajendran, T. Tamilselvi, and M. Meena, "A study and simulation of spectrum sensing schemes for cognitive radio networks," in *2020 7th International Conference on Smart Structures and Systems (ICSSS)*, July 2020, pp. 1–11.

[33] M. Derakhshani, M. Nasiri-Kenari, and T. Le-Ngoc, "Cooperative cyclostationary spectrum sensing in cognitive radios at low snr regimes," in *2010 IEEE International Conference on Communications*, May 2010, pp. 1–5.

[34] T. Xiong, Z. Li, Y.-D. Yao, and P. Qi, "Random, persistent, and adaptive spectrum sensing strategies for multiband spectrum sensing in cognitive radio networks with secondary user hardware limitation," *IEEE Access*, vol. 5, pp. 14 854–14 866, August 2017.

[35] A. Azarfar, C.-H. Liu, J.-F. Frigon, B. Sansò, and D. Cabric, "Joint transmission and cooperative spectrum sensing scheduling optimization in multi-channel dynamic spectrum access networks," in *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, March 2017, pp. 1–10.

[36] B. Wang, K. R. Liu, and T. C. Clancy, "Evolutionary cooperative spectrum sensing game: how to collaborate?" *IEEE Transactions on Communications*, vol. 58, no. 3, pp. 890–900, March 2010.

[37] W. Zhang, Y. Yang, C. K. Yeo, and L. Deng, "Cluster-based cooperative spectrum sensing assignment strategy in cognitive radio networks," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sep. 2014, pp. 1–5.

[38] W. Ejaz, G. Hattab, N. Cherif, M. Ibnkahla, F. Abdelkefi, and M. Siala, "Cooperative spectrum sensing with heterogeneous devices: Hard combining versus soft combining," *IEEE Systems Journal*, vol. 12, no. 1, pp. 981–992, March 2018.

[39] S. Chaudhari, J. Lunden, V. Koivunen, and H. V. Poor, "Cooperative sensing with imperfect reporting channels: Hard decisions or soft decisions?" *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 18–28, Jan 2012.

[40] Y. Han, Q. Chen, and J.-X. Wang, "An enhanced d-s theory cooperative spectrum sensing algorithm against ssdf attack," in *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, May 2012, pp. 1–5.

[41] W. Wang, H. Li, Y. Sun, and Z. Han, "Catchit: Detect malicious nodes in collaborative spectrum sensing," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, Nov 2009, pp. 1–6.

[42] L. Ma, Y. Xiang, Q. Pei, Y. Xiang, and H. Zhu, "Robust reputation-based cooperative spectrum sensing via imperfect common control channel," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 3950–3963, May 2018.

[43] J. Feng, S. Li, S. Lv, H. Wang, and A. Fu, "Securing cooperative spectrum sensing against collusive false feedback attack in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8276–8287, Sep. 2018.

[44] C. S. Hyder, B. Grebur, L. Xiao, and M. Ellison, "Arc: Adaptive reputation based clustering against spectrum sensing data falsification attacks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1707–1719, Aug 2014.

[45] J. Ren, Y. Zhang, Q. Ye, K. Yang, K. Zhang, and X. S. Shen, "Exploiting secure and energy-efficient collaborative spectrum sensing for cognitive radio sensor networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 6813–6827, Oct 2016.

[46] H. Zhu, T. Song, J. Wu, X. Li, and J. Hu, "Cooperative spectrum sensing algorithm based on support vector machine against ssdf attack," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[47] H. Chen, M. Zhou, L. Xie, K. Wang, and J. Li, "Joint spectrum sensing and resource allocation scheme in cognitive radio networks with spectrum sensing data falsification attack," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 9181–9191, Nov 2016.

[48] S. Bi, J. Lyu, Z. Ding, and R. Zhang, "Engineering radio maps for wireless resource management," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 133–141, April 2019.

[49] Z. El-friakh, A. M. Voicu, S. Shabani, L. Simić, and P. Mähönen, "Crowdsourced indoor wi-fi rems: Does the spatial interpolation method matter?" in *2018 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Oct 2018, pp. 1–10.

[50] L. Yang, N. Wu, B. Li, W. Yuan, and L. Hanzo, "Indoor localization based on factor graphs: A unified framework," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 4353–4366, March 2023.

[51] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," in *2012 IEEE International Symposium on Dynamic Spectrum Access Networks*, Oct 2012, pp. 422–433.

[52] A. Achtzehn, J. Riihijärvi, and P. Mähönen, "Improving accuracy for tvws geolocation databases: Results from measurement-driven estimation approaches," in *2014 IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, April 2014, pp. 392–403.

[53] X. Ying, C. W. Kim, and S. Roy, "Revisiting tv coverage estimation with measurement-based statistical interpolation," in *2015 7th International Conference on Communication Systems and Networks (COMSNETS)*, Jan 2015, pp. 1–8.

[54] X. Liu, F. Chen, and C.-T. Lu, "Robust prediction and outlier detection for spatial datasets," in *2012 IEEE 12th International Conference on Data Mining*, Dec 2012, pp. 469–478.

[55] Y. Gao and T. Fujii, "Kriging-based trust nodes aided rem construction under threatening environment," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, Sep. 2022, pp. 1–7.

[56] S.-H. Jung and D. Han, "Automated construction and maintenance of wi-fi radio maps for crowdsourcing-based indoor positioning systems," *IEEE Access*, vol. 6, pp. 1764–1777, December 2018.

[57] S. B. of Japan, "Method of demarcation for grid square," [Online]. available:http://www.stat.go.jp/english/data/mesh/05.html.

[58] A.-H. Sato, S. Nishimura, and H. Tsubaki, "World grid square codes: Definition and an example of world grid square data," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec 2017, pp. 4238–4247.

[59] J. Tada and K. Sato, "An implementation of a grid square codes generator on a risc-v processor," *International Journal of Networking and Computing*, vol. 12, no. 1, pp. 204–217, Jan 2022.

[60] K. Katagiri, K. Sato, and T. Fujii, "Crowdsourcing-assisted radio environment maps for v2v communication systems," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–5.

[61] K. Sato, K. Inage, and T. Fujii, "Modeling the kriging-aided spatial spectrum sharing over log-normal channels," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 749–752, June 2019.

[62] ——, "On the performance of neural network residual kriging in radio environment mapping," *IEEE Access*, vol. 7, pp. 94 557–94 568, July 2019.

[63] ——, "Frequency correlation of shadowing over tv bands in suburban area," *Electron. Lett.*, vol. 54, no. 1, pp. 6–8, Jan. 2018.

[64] K. Katagiri, K. Sato, and T. Fujii, "Crowdsourcing-assisted radio environment database for v2v communications," *Sensors*, vol. 18, no. 4, p. 1183, April 2018.

[65] S. Men, P. Chargé, and S. Pillement, "A robust cooperative spectrum sensing method against faulty nodes in cwsns," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 334–339.

[66] Y. Gao, M. Diao, and T. Fujii, "Sensor selection based on dempster-shafer evidence theory under collaborative spectrum sensing in cognitive radio sensor networks," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Sep. 2019, pp. 1–7.

[67] B. Chen and P. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 693–699, Feb 2005.

[68] P. Qihang, Z. Kun, W. Jun, and L. Shaoqian, "A distributed spectrum sensing scheme based on credibility and evidence theory in cognitive radio context," in *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2006, pp. 1–5.

[69] N. Nguyen-Thanh and I. Koo, "An enhanced cooperative spectrum sensing scheme based on evidence theory and reliability source evaluation in cognitive radio context," *IEEE Communications Letters*, vol. 13, no. 7, pp. 492–494, July 2009.

[70] ——, "Evidence-theory-based cooperative spectrum sensing with efficient quantization method in cognitive radio," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 185–195, Jan 2011.

[71] A. Ghasemi and E. S. Sousa, "Asymptotic performance of collaborative spectrum sensing under correlated log-normal shadowing," *IEEE Communications Letters*, vol. 11, no. 1, pp. 34–36, Jan 2007.

[72] R. Zhang, J. Wei, D. G. Michelson, and V. C. M. Leung, "Outage probability of mrc diversity over correlated shadowed fading channels," *IEEE Wireless Communications Letters*, vol. 1, no. 5, pp. 516–519, October 2012.

[73] R. Wan, M. Wu, L. Hu, and H. Wang, "Energy-efficient cooperative spectrum sensing scheme based on spatial correlation for cognitive internet of things," *IEEE Access*, vol. 8, pp. 139 501–139 511, August 2020.

[74] Y. Gao and T. Fujii, "Improvement of radio environment map under data falsification attack," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, Sep. 2021, pp. 1–6.

[75] D. Giancristofaro, "Correlation model for shadow fading in mobile radio channels," *Electronics Letters*, vol. 32, pp. 958 – 959, 06 1996.

[76] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol. 27, pp. 2145–2146(1), November 1991.

[77] A. Sree Dhevi, "Imputing missing values using inverse distance weighted interpolation for time series data," in *2014 Sixth International Conference on Advanced Computing (ICoAC)*, Dec 2014, pp. 255–259.

[78] E. Oktavia, Widyawan, and I. W. Mustika, "Inverse distance weighting and kriging spatial interpolation for data center thermal monitoring," in *2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Aug 2016, pp. 69–74.

[79] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Malicious user detection in a cognitive radio cooperative sensing system," *IEEE Transactions on Wireless Communications*, vol. 9, no. 8, pp. 2488–2497, August 2010.

[80] ——, "Secure cooperative sensing techniques for cognitive radio systems," in *2008 IEEE International Conference on Communications*, May 2008, pp. 3406–3410.

[81] S. R. X. Ying and R. Poovendran, "Incentivizing crowdsourcing for radio environment mapping with statistical interpolation," in *IEEE DySPAN*, Sep. 2015, p. 365–374.

[82] K. Zeng, P. Pawelczak, and D. Cabric, "Reputation-based cooperative spectrum sensing with trusted nodes assistance," *IEEE Communications Letters*, vol. 14, no. 3, pp. 226–228, March 2010.

[83] Y. Gao and T. Fujii, "An improvement of security scheme for radio environment map under massive attacking," *IEEE Access*, vol. 10, pp. 45 508–45 521, April 2022.

[84] P. Zhen, B. Zhang, Y.-Q. Xu, Z. Chen, H. Wang, and D. Guo, "Radio environment map construction based on gaussian process with positional uncertainty," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1639–1643, Aug 2022.

[85] X. Wang, X. Wang, S. Mao, J. Zhang, S. C. G. Periaswamy, and J. Patton, "Indoor radio map construction and localization with deep gaussian processes," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11 238–11 249, Nov 2020.

[86] B. Ferris, D. Fox, and N. Lawrence, "Wifi-slam using gaussian process latent variable models," in *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, ser. IJCAI'07, San Francisco, CA, USA, 2007, p. 2480–2485.

[87] M. Malmirchegini and Y. Mostofi, "On the spatial predictability of communication channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 964–978, March 2012.

[88] S. S. Szyszkowicz, H. Yanikomeroglu, and J. S. Thompson, "On the feasibility of wireless shadowing correlation models," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4222–4236, Nov 2010.

[89] M. Iskander and Z. Yun, "Propagation prediction models for wireless communication systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 662–673, March 2002.

[90] G. Andrea, *Wireless Communications*, 2005.

[91] A. O. Margaret and W. Richard, *Basic Steps in Geostatistics: The Variogram and Kriging*, 2015.

[92] K. Katagiri, K. Sato, K. Inage, and T. Fujii, "Dynamic radio map using statistical hypothesis testing," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 3, pp. 752–766, Sep. 2021.

[93] J. Perez-Romero, A. Zalonis, L. Boukhatem, A. Kliks, K. Koutlia, N. Dimitriou, and R. Kurda, "On the use of radio environment maps for interference management in heterogeneous networks," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 184–191, August 2015.

Publications

List of Publications Directly Related to The Dissertation

Journal Papers

1. Y. Gao and T. Fujii, "An Improvement of Security Scheme for Radio Environment Map Under Massive Attacking," in *IEEE Access*, vol. 10, pp. 45508-45521, May 2022, doi: 10.1109/ACCESS.2022.3170478. (Related to Chapter 3)

2. Y. Gao and T. Fujii, "A Kriging-based Radio Environment Map Construction and Channel Estimation System in Threatening Environments," in *IEEE Access*, vol. 11, pp. 38136-38148, April 2023, doi: 10.1109/ACCESS.2023.3267973. (Related to Chapter 5)

Refereed International Conference Papers

1. Y. Gao and T. Fujii, "Kriging-based Trust Nodes Aided REM Construction under Threatening Environment," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, London, United Kingdom, Sep. 2022, pp. 1-7, doi: 10.1109/VTC2022-Fall57202.2022.10012983. (Related to Chapter 4)

List of Referenced Publications

Refereed International Conference Papers

1. Y. Gao and T. Fujii, "Terminal Selection Based on Multi-armed Bandit under Threatening Environment for Radio Environment Map Construction," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, Helsinki, Finland, June 2022, pp. 1-6, doi: 10.1109/VTC2022-Spring54318.2022.9861006.

2. Y. Gao and T. Fujii, "Improvement of Radio Environment Map under Data Falsification Attack," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, Norman, OK, USA, Sep. 2021, pp. 1-6, doi: 10.1109/VTC2021-Fall52928.2021.9625446.

3. Y. Gao, M. Diao and T. Fujii, "Sensor Selection Based on Dempster-Shafer Evidence Theory under Collaborative Spectrum Sensing in Cognitive Radio Sensor Networks," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Honolulu, HI, USA, Sep. 2019, pp. 1-7, doi: 10.1109/VTCFall.2019.8891189.

Domestic Conference Papers

1. Y. Gao and T. Fujii, " A Trustworthy Approach for Radio Environment Map Construction in CRNs," in *IEICE Communication Society RISING2022*, Kyoto, Japan, Oct. 2022.

2. Y. Gao, M. Diao and T. Fujii, " Reputation-Based Spectrum Sensing Strategy in Cognitive Radio Sensor Networks," in *Technical Report of IEICE, SR2019*, Osaka, Japan, July 2019.