QATAR UNIVERSITY

Graduate Studies

College of Engineering

Parameters Estimation to Achieve Distance-Based Security Breaching

A Thesis in

Computer Science and Engineering

By

Tara Thaer Salman

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

June 2015

Declaration

To the best of my knowledge, this thesis contains no material previously published or written by another person or institution, except where due reference is made in the text of the thesis. This thesis contains no material which has been accepted for the award of any other degree in any university or other institution.

Name Tara Thaer Salman

Signature _____ Date _____

Committee

The thesis of **Tara Thaer Salman** was reviewed and approved by the following:

We, the committee members listed below, accept and approve the Thesis/Dissertation of the student named above. To the best of this committees knowledge, the Thesis/Dissertation conforms the requirements of Qatar University, and we endorse this Thesis for examination.

Supervisor Tarek Mohammed Al-Fouly

Signature _____ Date _____

Abstract

Despite the great amount of research done on network security in the past decades, securing the communication from the physical layer perspective is still emerging and considerably a hot research area in wireless communication. Current researches involve the usage of physical layer characteristics as a secret key to be used between the communicating nodes. To that extent, as part of a physical layer security project, a security scheme based on the distance between communicating nodes was proposed. This thesis aims to propose and investigate some parameters estimation for the eavesdropper to achieve the breaching of distance based security scheme. Those parameters include the number of sources estimation, direction of arrivals estimation, distance estimation and the usage of such estimations to calculate the distance between the communicating nodes. By that, the distance in which the security key is based on can be estimated from an eavesdropper node and ideally the key should be generated correctly. The algorithms should handle accurate estimations with different circumstances, including low signal to noise ratio and an unknown number of applied signals which can be enforced by the two communicating nodes. Besides, it should be practically implementable so that it can be employed easily on any hardware.

Different algorithms for the number of sources estimation will be proposed to overcome the problem of unknown sources. Two directions of arrival algorithms will be investigated and one distance estimation algorithm will be explained to be later used in hardware implementation. The end results should clarify if the distance between communicating nodes can be estimated approximately, state the limitations of doing so, analyze the effect on the system and propose some ways to overcome such breaching schemes. It should be noted that the end objective is not to breach the security scheme, but to know its limitations in order to enhance it using further research.

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	**Description**
DoA	Direction of Arrival
SNR	Signal-to-Noise-Ratio
$MUSIC$	MUltiple SIgnal Classification
AIC	Akaikes Information Criterion
MDL	Minimum Description Length
EVD	Eigenvalues Decomposition
ULA	Uniform Linear Array
UCA	Uniform Circular Array
$MVDR$	Minimum Variance Distortion-less Response
$ESPRIT$	Estimation of Signal Parameter via Rotational Invariance Technique
GPS	Global Positioning Systems
ToA	Time of Arrival
$TDoA$	Time Difference of Arrival
$RSSI$	Received Signal Strength Indicator
DoS	Denial-of-Service
RF	Radio Frequency
CSI	channel state information
CR	Cogitative Radio
PUs	Primary Users
CUs	Secondary Users
CRB	Cramer-Rao Bounds
PNR	Peak to Noise Ratio
PAR	Peak to Average Ratio

EEE	Entropy Estimation of Eigenvalues
$CovM$	Covariance Matrix
$CorrM$	Correlation Matrix
MI	Moving increment
STD	Standard Deviation
MS	Moving STD
$OFDM$	Orthogonal Frequency Division Multiplexing
CGC	Channel Gain Component
$RFID$	Radio Frequency Identification
ICA	Independent Component Analysis
$AWGN$	Additive White Gaussian Noise
$QPSK$	Quadrature Phase Shift Keying
BMR	Bit-Mismatch Ratio
$NMSE$	Normalized Mean Square Error
$WARP$	Wireless open Access Research Platform
$FBGA$	Field-Programmable Gate Array

Acknowledgements

This thesis was made possible by the support of the NPRP grant 5-559-2-227 from the Qatar National Research Fund (QNRF). The statements made herein are solely the responsibility of the authors.

First and foremost I am grateful to my supervisor, Dr. Tarek El-Fouly, for his highest support, patience and guidance through this thesis and for allowing me to be part of the NPRP project mentioned earlier. I consider myself fortunate in being able to work with such a motivated and encouraging adviser over the past couple of years. Without his support, I would not be able to finish the work or even be interested in continuing my studies after B.Sc degree.

I am also greatly thankful to Dr. Amr Mohamed and Dr. Tamer Khattab for their continuous support and kind advice in both thesis and the project that I am part of. I received a great amount of help and motivation from both and this thesis would be missing a lot without their existence.

In addition, I would like to thank Mr. Ahmed Badawy who coauthored with me papers and had spent a great effort helping in this work. Ahmed is part of the same NPRP project referenced before and I offer my sincerest gratitude to him as a great helper and a second adviser in all the work done.

Finally, huge thanks go to Dr. Mohammed Samaka for his advice and encouragement over the past years. He had a great effect on me in continuing this work and motivating me to further study after earning this master's Degree. I wish to thank all the Computer Science and Engineering Department at Qatar University for giving us such a great opportunity to earn our degree and encouraging our hard work and achievements.

Dedication

I dedicate this thesis to my parents who have been greatly supporting me in every single step in my life. To my supervisor who has been a great pacemaker and actually raised research attitudes in me. To my best friend who has been there through all the good times and bad. And to the rest of the family and colleagues who supported me. I hope this work would make everyone proud of me as much as I am proud of them.

Chapter 1

Introduction

Signal processing is a critical technology that is involved in many applications including image, sound and video processing, wireless communication, control systems and source and channel coding. In computer networks, signal processing is highly involved in the physical layer and the presentation layer of the networking stack. In the context of the physical layer, signal processing techniques are used in cognitive radios, military communications and many other applications that are emerging in today's research. We are interested in the use of signal processing techniques in order to extract physical layer characteristics and then use them in security from the physical layer perspective, which is one of the newly emerging and highly critical fields in networking.

Traditionally, securing the communication system from the physical layer perspective is done using information theoretic approaches which are costly, impractical and rarely implementable on actual hardware. Cryptographic techniques that are applied to the upper layers in the networking stack sound promising; however, due to their key generation complexity and assumptions, it cannot be applied at the physical layer. On the other hand, physical layer characteristics such as channel state information, radio frequency and node localization information look helpful in key generation as they can act as a source of randomness and easy to get using signal processing techniques applied to the received signal in wireless channels.

Thus, as part of the research project, a physical layer key generation scheme was proposed based on the distance between the two authorized nodes. This scheme would use the distance between the nodes, which can be considered as a source of

randomness in mobile networks, to agree on a secret key that can be used by both nodes without enabling the eavesdropper, which is someone who is trying to breach the system, to know about it. In order for an eavesdropper to breach the system, she/he needs to listen to the communication while the key is exchanged, use multiple estimations to estimate the distance in which the key was based on and then generate the key based on those estimated distances.

To estimate the distance, the adversary needs to use multiple signal processing techniques that include a number of sources estimation, direction of arrival (DoA) estimations and distance estimation from both nodes. Since we are dealing with wireless channels, the possibility of receiving more than two sources is high, so the number of sources estimation is needed to know many sources are received and extract those of interest, i.e. extract the two communicating nodes. DoA estimation is needed to estimate the direction of signals coming from both communicating nodes while distance estimation is needed to estimate the distances from the eavesdropper to both nodes. Based on the estimated information, the eavesdropper can estimate the distance between the communicating nodes using triangle laws such as the law of cosine in which two sides and an angle of the triangle is known and the third side is required. The assumption over here is that the nodes are communicating back and forth so that the eavesdropper can receive a signal from both nodes.

The goal of this thesis is to estimate the required parameters and apply the previously stated procedure to estimate distance from an eavesdroppers point of view. The system can be viewed as in Fig. 1.1 where the node agrees on a distance-based security key and a passive eavesdropper that is just listening to their communication. The objective is to estimate the key and check the effect on the system in such scenarios. The importance of doing so is to determine how secure the communication is if the distance was exactly known by the eavesdropper and if she/he is able to breach the system or not. In addition, this work will help in gaining insight into distance-

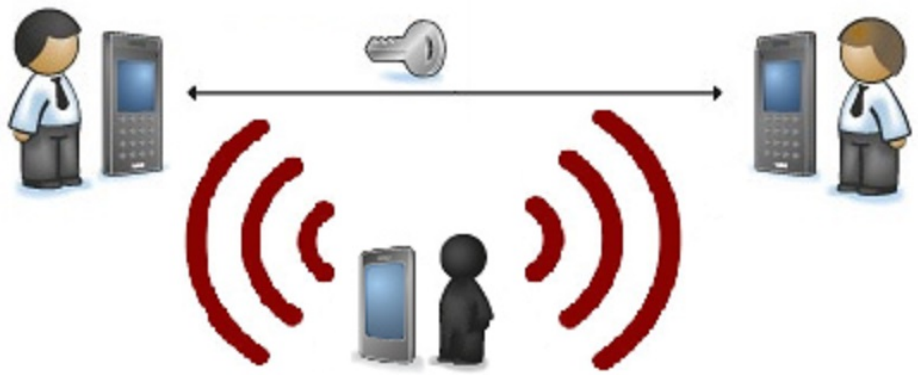


Figure 1.1: Overview

based security scheme and will act as a starting point for building stronger security algorithms that can secure the communication even if the distance is known.

Thus, the main objectives or points to achieve in this thesis are:

- Estimating the number of sources: As the eavesdropper is dealing with wireless channels, she/he might be receiving multiple signals from multiple sources. Therefore, the first step is to estimate how many signals she/he is receiving and then extract the ones of interest. Extracting the signals of interest would involve network protocols other than physical layer processing and hence such process is out of the scope of this thesis. However, this thesis estimates the number of sources and then assumes that the first two signals are the one of interest.
- Estimating DoA in an accurate way, taking into consideration low Signal-to-Noise-Ratio(SNR) and a small number of samples.
- Estimating the distances from the eavesdropper to the communicating nodes.
- Estimating the distance between the sender and the receiver, from the eavesdropper's point of view.
- Evaluating the estimated distance by its mean square error and normalized estimated distance.

- Generating a key based on the estimated distance and a simple key generation scheme.
- Evaluating the generated key by its bit error mismatch and compare it to communication nodes keys.

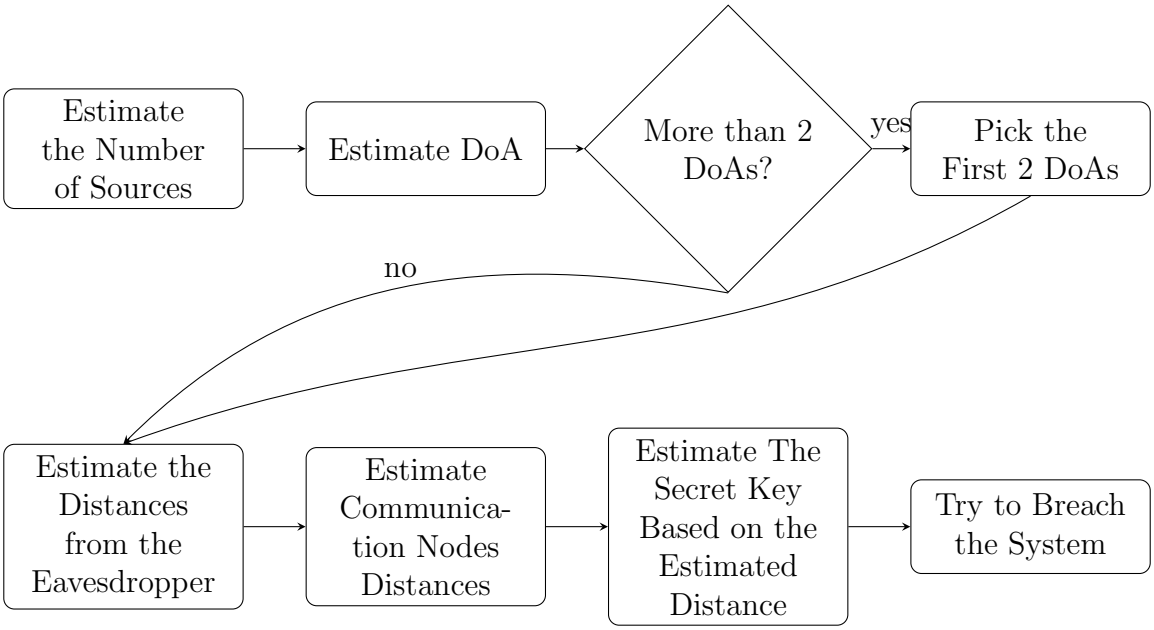


Figure 1.2: Breaching Scheme Processes

Fig. 1.2 illustrates the process of our breaching scheme. As shown in the figure, when the estimated DoAs exceed two, we assume that the first two estimated ones are our signals of interest, considering that these two will have the highest power to noise values. Extracting other DoAs can be considered as a future work and it is out of the scope of this thesis. It can also be seen that the last step investigates how close the estimated distance to the actual one and the possibility of being able to breach the system by generating an initial key which is compared to communicating nodes keys. The effect of such breaching is out of the scope of this thesis and can be considered as a future work in combination with the hardware implementation of this thesis.

Chapter 2

Background

In this chapter, a brief background about the Parameters Estimation used to reach thesis objectives will be presented followed by an introduction about physical layer security techniques.

2.1 Parameters Estimation

In this section, some fundamental concepts about the estimation techniques required in order to achieve the breaching goal are briefly discussed. This includes: number of sources estimation techniques, DoA estimation techniques and distance estimation techniques.

2.1.1 Number of Sources Estimation

Estimating the number of sources is considered as a critical parameter in DoA estimation besides its importance in other applications such as blind source separation [1] and channel order separation [2]. As an example, DoA subspace estimation algorithms such as MUltiple SIgnal Classification (MUSIC) involve eigenvalues decomposition which is done after knowing the number of sources received by the receiver array [3]. Thus, having the number of sources accurately estimated is critical in order for these algorithms to perform well and result in the correct directions. In this thesis, the number of sources estimation is critical in order to know how many signals are actually applied to the array, then exclude the signals of interest in order to do further processes on them.

In general, as in [4], estimating the number of sources can be classified into:

- **Information theoretic based estimation:** Algorithms such as Akaikes information criterion (AIC) and minimum description length (MDL) [5] use information theoretic criteria to estimate the number of sources. They search for a minimum value of the log like-hood function and an added penalty term, as will be explained later. Those algorithms are basically computationally complex due to the use of a complex minimization problems, as will be seen later, and the estimation of eigenvalues decomposition (EVD). Besides, they have some impractical assumptions such as sparse like noise, uncorrelated with the signal and the number of samples to be large enough. Hence, they fail in some practical scenarios such as underwater [6], and indoor offices [7].
- **Eigenvector based estimation:** The rank of the matrix composed of eigenvectors is used instead of the eigenvalues. In [8], the authors examine the rank of the appended eigenvector matrix. The rank increases with the number of subarray until the number of sources is reached where the rank stabilizes at that point and hence the number of sources can be estimated. The same assumptions of information theoretic approaches are applied; however, the advantage here is the use of non-coherent signals as well as fully coherent signals, which is a more practical scenario as signals are normally coherent. On the other hand, such algorithms involve finding the rank of a matrix besides EVD estimation, which increases the computational complexity, particularly for large matrices with large a number of elements. In addition, the sparse-like noise assumption is still used in algorithms, which makes such algorithms also fail in indoor and underwater environments. It is worth noting that [9] presented an algorithm to estimate both coherent and non coherent signals using the rank of different matrix formulations which could achieve better results. However, the complexity

of such an algorithm increases due to multiple matrix decompositions that are involved in estimation.

- **Threshold based estimation:** In these approaches, the number of sources is detected by setting up a threshold on the noise eigenvalues, i.e. if that threshold is exceeded, then it is a signal eigenvalue and the number of sources can be detected at that point. This threshold can be made on the upper bound of the noise eigenvalues [10], the difference between them [11] or any other mathematical relationship. The drawback in most of the techniques in this category is that the formula needed to estimate the threshold has an adjustment coefficient which needs to be set beforehand. This coefficient is not analytically found as finding such a thing through probability distributions and order statistics is very complex, as will be shown later. Instead, they estimate the coefficient through extensive computer simulation for each pair of antenna elements and the number of collected samples. In other words, if the number of antenna elements, the number of collected samples or both change, the coefficient has to be estimated accordingly, which adds a considerable burden to the system.

2.1.2 Direction of Arrival (DoA) Estimation

Estimating DoA from the received signal has drawn extensive attention in antenna array signal processing. The antenna array receives several signals from different directions and collects them at its elements with the added noise by the channel. Then, it processes this data to estimate the DoA of the received signals with different algorithms that were enhanced by the existence of smart antenna [12], [13].

Different types of arrays exist for signal processing, including uniform linear array (ULA), uniform circular array (UCA), phase array and many others [14]. For this thesis, UCA is used as it plays a significant role in DoA estimation due to its full 360 azimuth and 90 elevation angle coverage. Hence, some more details about UCA will

be found through this thesis; however, for more information about the other type of arrays, the reader can refer to [13].

DoA estimation algorithms can be classified into classical estimation and subspace estimation techniques. Below, some of the main estimation techniques that are involved in most DoA research are briefly discussed:

- **Bartlett algorithm:** A classical technique that steers the spectrum, finds the power of the received signal and determines the DoA by the location of the highest power in the spectrum. This is the simplest estimation algorithm; however, the main drawback is its low resolution as the angular separation between DoAs cannot exceed $2\pi/M$ in order for the DoAs to be estimated, where M is the number of array elements [15].
- **Minimum Variance Distortion-less Response (MVDR) algorithm:** A classical algorithm which is similar to Bartlett; however, it uses the inverse matrix of the received signal instead of the received signal itself. The DoA is estimated by scanning the spectrum and finding the highest power as well. This could solve the angular separation drawback relatively; however, it came with the complexity of finding the inverse of covariance matrix of the received signal [16].
- **MUltiple SIgnal Classification (MUSIC) algorithm:** A subspace estimation technique which uses eigenvalues decomposition to separate the noise and signal subspaces, relying on the fact that they are orthogonal. This provides the highest resolution at the expense of complexity and the need to know the exact number of sources. The complexity comes from eigenvalues decomposition on the covariance or the correlation matrix of the received signal which requires substantial computations. The number of sources is needed in order to separate the signal from noise subspaces [17].

- **Estimation of Signal Parameter via Rotational Invariance Technique (ESPRIT) algorithm:** A subspace estimation algorithm that exploits the rotational invariance of signal subspace which is created by eigenvalues decomposition of the received signal. This algorithm does not involve the exhaustive search of whole spectrum as in MUSIC and hence it is less complex; however, it has a lower resolution and less accurate estimation than MUSIC does [18].

2.1.3 Distance Estimation

In this section, some of the fundamental distance estimation algorithms will be stated along with their advantages and drawbacks. Distance estimation is needed to estimate the distance between the eavesdropper and both communicating nodes, which is needed with the direction of arrival estimation in order to estimate the distance between the two communicating nodes. Hence, estimating this distance is a critical task in hardware implementations of such a breaching scheme.

Estimating the distance at the receiver of two communicating nodes is a fundamental issue in many applications that involve wireless sensor network or wireless communication in general. Such applications include localization and positioning system such as Global Positioning Systems (GPS), patient localization systems, health care systems and intelligent transportation systems [19]. Thus, different techniques have been proposed to estimate the distance using different environment parameters and wireless characteristics. In general, as in [20], distance can be estimated using the following techniques:

- **Time of Arrival (ToA):** Distance between communicating nodes is directly proportional to the time that the signal takes to reach the other end. Hence, if the nodes were perfectly synchronized, the distance can be estimated using signal transmission time. i.e. if both sending and receiving times are known to the receiver, the distance can be estimated by $d = P_r(time_{arriving} - time_{sending})$

where P_r is the propagation speed [21]. This concept is applied in [22] to measure the distance in Zigbee networks and shows quite accurate results. However, the need for precise synchronization might be impractical in the hardware implementation, especially with different types of hardware.

- **Time Difference of Arrival (TDoA):** In such techniques, the distance is estimated based on the time difference that multiple signals will take to reach the destination. Hence, two types of signals, like RF and ultrasound, are needed to be sent to the same receiver at the same time and the distance can be estimated by the difference in their arriving times. i.e. the distance can be estimated by the formula $d = (P_r - P_u) * (time_1 - time_2)$ where P_r and P_u is the propagation speed for the RF and ultrasound signals, $time_1$ is the arriving time for the RF signal and $time_2$ is the arriving time for the ultrasound signal. This method is applied to WSN and Ad-Hoc networks in [23]. It is more precise than ToA; however, it needs extra hardware in order to send and receive two different types of signals and hence it adds more cost.

- **Received Signal Strength Indicator (RSSI):** RSSI can be used to estimate the distance of the receiver based on the received signal RSSI. This can be done using the log distance path loss exponent model, as will be discussed later in Section 4.4. Theoretically, there exists a direct radio propagation model to convert estimated RSSI at the receiver to an estimated distance. However, experimentally this method is inaccurate due to the environmental noise, obstacles, and antenna types. Thus, this method would need a system calibration in order to adjust some parameters and have a low error rate [24], [25].

  However, this method is the most widely used due to its simplicity and availability of RSSI readings with no added hardware. RSSI estimation circuitry is already deployed in most recent hardware and hence such estimation would be relatively easier than others.

2.2 Physical Layer Security

Physical layer security considers securing the communication system from the physical layer perspective. In general, security attacks can be classified into active and passive attacks. An active attack happens when an adversary tries to interfere with the communication by altering the exchanged messages, while a passive attack does not interrupt the network operation and the adversary just tries to get confidential information from the messages [26] [27]. Attacks at the physical layer can be classified into:

- **Denial-of-Service (DoS) attacks:** In such active attacks, the adversary attempts to overload the network by exhausting its resources. Jamming is a common way for DoS at the physical layer where the adversary utilizes the radio frequency by sending jamming signals and letting the nodes suffer from busy channels [28].
- **Masquerade attacks (or deauthentication attacks):** In these active attacks, the adversary pretends to be an authorized node and breaks the authentication system so it can illegally use the network resources. It usually involves other kinds of active or passive attacks such as authentication capturing in order to access the network [29].
- **Information disclosure and message modification:** It refers to active attacks that involve either modifying the message contents by adding or deleting based on adversaries' benefits or disclosing confidential information to unauthorized users in order to be used later [26], [27].
- **Eavesdropping and traffic analysis:** Eavesdropping is intercepting the authorized communication in order to gain confidential information without the authorized nodes knowledge. Traffic analysis is the usage of the communication between parties to determine the location and identities of the communication

parties. These kinds of passive attacks can be done even if the message is encrypted, and hence, they can help the eavesdropper in gaining information in order to perform other types of active attacks [26], [27].

In order to keep secured communication, one must either prevent eavesdropping or secure the communication even when eavesdropping happens, i.e. the eavesdropper cannot get helpful information by listening and analyzing the traffic. Preventing the network from eavesdropping in wireless is almost impossible due to the broadcasting nature and availability of the channels so it would be much easier if the security was done by preventing adversaries from gaining helpful information when eavesdropping.

Physical layer security algorithms can be classified into information theoretic analysis and cryptographic techniques. Below is a brief background about both techniques discussed with their current trends in securing the communication.

2.2.1 Information Theoretic Approaches

Information theoretic capacity is an average secrecy metric that indicates how much information has been leaked to the eavesdropper. The system here is designed and tuned to give a certain level of security, but it can never be guaranteed. The transmitter is required to have a partial or a full information about the channel since it will need such information to determine the security level or secrecy capacity [30], [31], [32]. Due to the inaccuracy of such information in practice and higher implementation cost, such systems are not available widely and only a few practical implementations were deployed to realize such systems.

2.2.2 Cryptographic Approaches

Cryptographic techniques are mostly based on encrypting the transmitted message by using either a public or a private key that is known to authorized nodes only. Traditional cryptographic techniques, which include public key infrastructure and

using one key all over the transmission, assume that the channel is perfect and there is no eavesdropping when transmitting the key or the random generator, i.e. the adverser does not have any information about how the key was generated. In addition, it assumes that the adversary has a limited computational capability such that it cannot estimate the key [33]. In the physical layer, the first assumption is weak since the wireless channel is open and eavesdroppers can access the channel between transmitters and receivers and extract the key. In addition, the dynamic changes in the wireless channel and mobility issues makee key distribution and key infrastructure a tedious task in such systems. More importantly, the security schemes at this layer should not assume anything about the computational capabilities of the adversary. Also, the premise that it is impossible for them to break the key is weak here and cannot be proven mathematically [34], [35]. Hence, the traditional cryptographic techniques do not suite the physical layer security, and new security schemes are needed in order to achieve the security needs or security capacities.

Since the wireless channel characteristics are random by nature, they can be used as a source of randomness for the secret key generation between two nodes. Properties like radio frequency (RF) fingerprinting, channel state information (CSI), the distance between the nodes and multiple signal processing techniques can be used to generate a key to be used without message exchanging so that the eavesdropper cannot know about it. In addition, the employment of directional antennas and artificial noise, noise enforced in specific directions, can help in key generation and improvement of secret capacity by avoiding jamming and enhancing data rate. Coding techniques such as error correction codes or spread spectrum codes can also be used for key generation as they act as a common source of randomness. All these techniques can result in unique keys without message exchanging and hence in order for the eavesdropper to know the key, she/he needs specific signal processing and channel eavesdropping techniques, beside the need for knowing what characteristics are used to generate the key at that moment [36].

Chapter 3

Related Work

In this chapter, some recent applications of Parameters Estimation usage in wireless networks will be presented first. Then, recent work in number of sources estimation, physical layer security scheme and security breaching scheme will be discussed in brief.

3.1 Parameters Estimation in Other Wireless Context

Since this work tackles parameters estimation such as the number of sources, DoA and distance estimations, it is possible that such estimations are useful in other wireless communication applications. As an example, cogitative radio (CR) networks and military communication require localization and source number estimations in their processing. Hence, our estimation algorithms can be used in their context to localize or estimate parameters beside the use in physical layer security. Therefore, this section presents some recent use of our parameters estimation in the context of CR networks while the same thing can be applied to military applications.

CR network is a technology in which primary users (PUs) and secondary users (SUs) share the same spectrum to utilize resources and channels. SUs are only permitted to use the licensed bands and resources when they are not interfering PUs communication. In such networks, it is important for SUs to know the number of PUs and localize them, beside localizing other SUs in the communication range. Since location estimation is not available in practice, PUs will need to use signal process-

ing estimation techniques to estimate such parameters. As an example, in [37], the authors utilize DoA measurements along with RSSI reading to localize the PUs transmitters in fixed and uniform CR placements. They derive the Cramer-Rao Bounds (CRB) for joint RSSI/DoA localization in fixed CR placements, show a close form of CRB in uniform CR placements, practically implement their algorithm and compare the results to RSS and DoA-based location estimation. They use MUSIC for DoA estimation which relates to our number of sources and DoA estimations used in this thesis.

In [38], authors consider using MDL algorithm to estimate the number of PUs from SUs nodes without the need for any PU cooperation, assuming that all SUs are passive at that time and they are occupied with multiple antennas system. Their results show that PUs can have a close estimate of the number of PUs and provide additional information in case of estimation error. In that work, MDL can be replaced with our proposed algorithms which have a better performance and much less complexity, as will be seen later.

In addition, [39] uses distance estimation algorithm to estimate the distance between the base station and PUs, and they use this estimation to set maximum allowable transmit power in dynamic spectrum sharing environment. In [40], authors use a hybrid RSSI/DoA based algorithm to estimate the location and transmit power of an emitter in CR network. All those works and many others showed that our parameter estimation algorithms, especially the proposed ones for the number of sources estimation, can be used in CR networks to secure the communication beside its usage in localization or spectrum sensing purposes.

3.2 Number of Sources Estimation

One of the main contributions in this work is proposing some new algorithms for number of sources estimation. As will be seen later, MUSIC is used for DoA estima-

tion due to its high resolution and acceptable tolerance for low SNR values. However, one of its problems is the need for pre-knowledge of the number of sources which is unavailable in practice. Hence, number of sources estimation is needed to estimate how many signals are received by the antenna array at the receiver, who, in our case, is the eavesdropper.

In [41], the authors present an idea that assumes that the number of sources to be as maximum as the number of array elements, i.e., there exists a number of elements DoAs which include both virtual and actual ones. Then, the algorithm processes normal DoA estimation using any DoA estimation algorithm, [42], and finds the peak to noise ratio (PNR) of all DoA which contains both actual and virtual DoA. The PNR of the actual DoA can be distinguished from the virtual DoA PNR based on hypothesis testing, and hence the number of sources can be estimated using this testing. The authors use AIC heuristic to estimate the number by replacing the eigenvalues with the estimated PNR values. Results show a better detection probability, especially with low SNR and low number of samples. However, such a method cannot be applied to high resolution DoA estimation such as MUSIC as the algorithm would require the actual number of sources when applying eigenvalue decomposition (EVD), and hence, it would fail for an assumed number of DoA. Besides, such an algorithm is expected to be more computationally complex, especially with the fact that AIC was applied to estimate the actual number of sources.

Another work in [43] presents a threshold-based estimation algorithm that is based on the peak to average ratio (PAR) characteristics. The algorithm calculates the PAR values of the received data and calculates the differences between adjacent ones which is compared to a threshold. If the difference exceeds the threshold, then the number of sources is detected at the index of that difference. The drawback of this algorithm is in the threshold which is set based on the average gradient of the PAR values, the minimum PAR value and a coefficient that needs to be adjusted with different

number of samples, number of array elements and the minimum PAR value, which indicates that such algorithm would require a preconfigurable threshold.

A third work presented in [44] proposes an algorithm for the number of sources estimation in non-Gaussian noise channels. This algorithm compares the estimated eigenvalues of correlation matrix to an estimated noise variance. It first computes the noise correlation matrix then applies an EVD operation on it. The estimated noise eigenvalues are then used to calculate the noise variance. Generally, this algorithm shows a better estimation accuracy than MDL and entropy estimation of eigenvalues (EEE) in [45]. However, the noise variance estimation requires a large number of samples for accurate estimation. Moreover, the algorithm involves the estimation of two correlation matrices with their EVD, which is computationally inefficient especially with large number of array elements and large number of samples.

In [11], the authors follow a threshold based approach that depend on the increments of the eigenvalue of the covariance matrix. The number of signals in this approach can be detected when the increment exceeds a certain threshold which means that the eigenvalues have moved from noise subspace to signal subspace. The threshold is a single formula that depends on the minimum and maximum values of eigenvalues, number of array elements and the number of samples or data length of the received signal. The complexity of this algorithm is low; however, depending on the data length and the number of array elements, the coefficient that is included in the threshold formula is changed and that coefficient is found experimentally. Hence, the approach might be inflexible for different number of samples and elements in the array.

To address the complexity problem of traditional algorithms and the reconfigurability problem in the threshold-based algorithms, this work proposes some simple yet efficient solutions to estimate the number of sources. The proposed algorithms can be categorized into two main categories based on the matrix used in estimating the number of sources. Namely, the work proposes sample covariance matrix (CovM)-based

algorithms and auto correlation coefficient (CorrM)-based algorithms. It defines two decision statistics, namely moving increment (MI) and moving standard deviation (MS), which are used as the metrics to estimate the number of sources. In other words, in each category, it proposes two algorithms. First, the selected matrix is estimated, EVD is applied and then the decision statistics is estimated from the resulting eigenvalues. In the first category, the decision statistics are compared to a preset threshold, while in the second a simple maximization technique is proposed.

3.3 Physical Layer Secret Key Generation Techniques

Cryptographic techniques that utilize wireless channel characteristics have drawn a lot of research attention in the past few years. Some wireless characteristics were used as a source of randomness to share the key between communication parties. Those characteristics include location, RSSI value, Channel State Information (CSI), ambient audio, channel response and many others. In this section, some of the latest and most useful key generation schemes are briefly discussed. However, it is worth noting that many others are proposed in the literature; however, such details are out of the scope of this project.

In [46], the researchers exploit the channel response from Orthogonal Frequency Division Multiplexing (OFDM) sub-carrier to generate the key. They first use CSI information for the key generation, which should theoretically generate identical keys; however, practical implementation shows non-reciprocity, or mismatch, components in CSI measurements. These non-reciprocity components are due to the different electrical characteristic of wireless devices, especially for antenna gain and attenuation and resulted in a high bit mismatch in the generated keys. Thus, they have proposed a channel gain component (CGC) algorithm to mitigate these non-reciprocity

components by learning the channel response which can be done by sharing a small number of probe packets between the communication nodes. Results show that such a scheme achieves fast secret key generation and is resistant to attacks such as predictable channel attacks and stalking attacks, explained in the next section.

In [47], the authors analyze the information theoretical limits of key generation using nodes location. In specific, they have based the key generation scheme on the distance estimated between the nodes and found the secrecy capacity and achievable secret key rate. Analysis was done with and without global localization information. Results show that the secret key rate grows unbounded if the eavesdropper does not have the angle of arrival observation, hence she/he cannot estimate the distance. Otherwise, secret key generation cannot go beyond certain limits and if the eavesdropper had the ability to do some planned movements, then she/he could drastically reduce the secret key bit rate and hence breach the system.

Moreover, [48] and [35] discuss an RSSI-based security scheme for collaborative networks. Researchers in [49] propose a key generation scheme based on features from ambient audio. While [50] proposes a key generation scheme based on adaptive channel probing that is based on the proportional-integral-derivative controller, which is used to tune the probing rate.

The authors in [51] investigate an algorithm that hops around different physical layer parameters that are agreed between the communicating nodes to generate the key that is used through communication. In [52], the authors use Radio-frequency identification (RFID) and quantum key managements to generate the secret key. Authors in [53] consider establishing the key based on spatial and temporal correlation of the wireless channel coefficients which relied on CSI information secret key extraction. Finally, authors in [54] use an adaptive quantization mechanism for the secret key generation that derives the optimal quantization parameter to achieve a high secret key extraction rate.

3.4 Physical Layer Security Breaching

Since physical layer security schemes are newly emerging, breaching algorithms for such schemes are still limited and newly investigated. A limited research is done to tackle such breaching schemes, especially for the distance, DoA and CSI based information security schemes. In addition, most of the work done in the security algorithm itself assumes an adversary model for the eavesdropper, which decreases the possibility of breaching those security schemes. For example, in [47], the authors assume the eavesdropper has information about the relative locations of the communicating nodes and drove the theoretical bound of localization-based security. Hence, theoretically, even when the eavesdropper knows the location of legitimate nodes, the security can be reached with some limits. In [46], the authors consider their attack models under two types of attacks: predictable channel attacks[55] and stalking attacks [56]. Those attacks, discussed in the next paragraph, are harmful to RSSI and CSI-based secret key generation; however, it was proven that security can be achieved to some limits even under those attacks.

In predictable channel attacks, [55], the attacker uses some planned movement in a stationary environment to ensure that the communicating nodes build a secret key that can be predicted by him or her. For example, it can move between communicating nodes line of sight, causing a desired change in the channel and hence making the nodes agree on a key that can be predicted by him or her. In stalking attacks, [56], an adversary called stalker eavesdrops the trajectory of one communication node during the key establishment and hence can eavesdrop all the communication between the nodes.

In [57], the authors develop an algorithm that contradicts artificial noise security scheme and successfully eavesdrops the communication to access confidential information. The algorithm depends on independent component analysis (ICA) to cancel the noise added by the transmitter and break the security scheme. It utilizes an

existing ICA method, called fastICA [58], to cancel the artificial noise and achieve eavesdropping. FastICA is modified to handle complex value case which enables the breaching algorithm to decode the noisy received signal and extract the confidential data from the received signal.

In [59], the author tries to impose impersonation attacks to tackle RF-based security scheme. It studies the feasibility of such attacks in both modulation and transient based techniques for RF fingerprinting. Attacks are performed by either feature or signal replay from the eavesdropper. In feature replay attacks, which tackle the modulation based identification, the attacker modifies its own radio signal characteristics to match one of the targeted node. In signal replay attacks, the signal sent from the target node is saved by the eavesdropper to be used later without modifying the transient or modulation parts. This can help the eavesdropper to pass the authentication and the encryption if it was based on the RF fingerprint of the sender or receiver. Results, which were applied for authentication-based RF fingerprinting, show that replay attacks affected both modulation and transient RF fingerprinting while feature replay influenced only modulation based fingerprinting.

This work differs from others by proposing a practical, implementable breaching scheme without considering the security capacity limits and mathematical derivations which were considered in others work. To the best of our knowledge, this work is the first to tackle distance based security breaching which was proposed in [47] and [60]. The objective of this work is not to breach the distance based security scheme, but to check its strength against eavesdropping in order to further enhance it in a later work.

Chapter 4

Algorithms Design

In this chapter, we first present our system model that will be used with DoA and number of sources estimation techniques. We follow that by the number of sources estimation algorithms, in which we propose our new algorithms and present some others. Then, MVDR and MUSIC DoA algorithms will be explained to be compared later in the next chapter. After that, we present distance estimation techniques followed by our breaching scheme in which we will be discussing how breaching can happen by the eavesdropper.

4.1 System Model

In our system model, as shown in Fig. 4.1, we assume that the receiver is equipped with M-elements uniform circular array antenna, where M is 8 in the figure. Considering K signals are applied to the receiver's array, the received signal instant time t can be expressed as:

$$\mathbf{y(t)} = \sum_{\mathbf{k=1}}^{\mathbf{K}} \mathbf{a}(\phi_{\mathbf{k}})\ \mathbf{s_k(t)} + \mathbf{w(t)}, \tag{4.1}$$

where $\mathbf{a}(\phi_{\mathbf{k}})$ is the steering vector for the signal arriving at azimuth angle ϕ_k, $\mathbf{s_k(t)}$ is the applied signal from the k^{th} source at time t, and $\mathbf{w(t)}$ is the Additive White Gaussian Noise (AWGN). In matrix notations, (4.1) can be represented by:

$$\mathbf{Y} = \mathbf{AS} + \mathbf{W}, \tag{4.2}$$

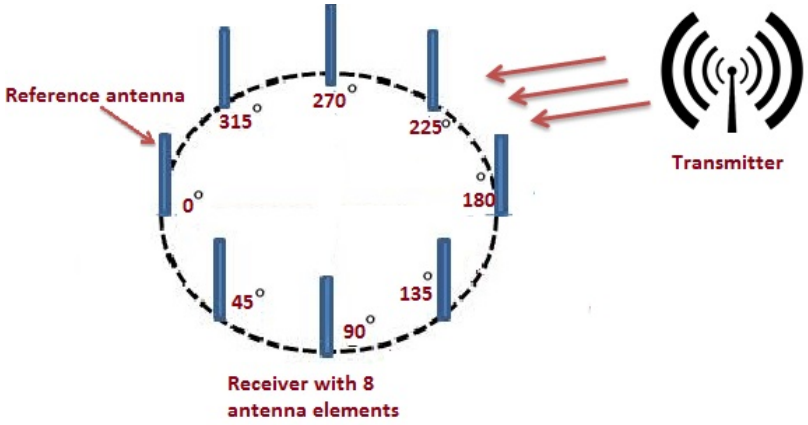


Figure 4.1: Circular array with one applied signal

where $Y \in \mathbb{C}^{M\times N}$, $\mathbf{A} \in \mathbb{C}^{M\times K}$, $\mathbf{S} \in \mathbb{C}^{K\times N}$, $\mathbf{W} \in \mathbb{C}^{M\times N}$, with N being the total number of collected samples and $\mathbb{C}$ is the set of complex numbers. The matrix of steering vectors is

$$\mathbf{A} = [\mathbf{a}(\phi_1), \mathbf{a}(\phi_2)...\mathbf{a}(\phi_k)]. \tag{4.3}$$

The steering vector $\mathbf{a}(\phi_k)$ for a uniform circular array (UCA) can be represented by:

$$\mathbf{a}(\phi_k) = e^{(2\pi/\eta r sin(\theta)(cos(\phi-\gamma)))}, \tag{4.4}$$

with waveform number η, radius r and γ is $360/N * (0 : N-1)$, θ is the Z-Plane angle which we assume to be orthogonal to the array and hence $sin(\theta)$ is 1 for the rest of the thesis.

4.1.1 Sample Covariance Matrix (CovM)

The covariance matrix of the received data can be expressed as:

$$\begin{aligned} \mathbf{R_{YY}} &= \mathbb{E}\left[\mathbf{YY^H}\right] \\ &= \mathbf{AR_{SS}A^H} + \mathbf{R_{WW}} \end{aligned} \tag{4.5}$$

where, $\mathbb{E}[.]$ denotes the expectation operation, H denotes the Hermitian operation, $\mathbf{R_{SS}}$ is the covariance matrix of the applied signal, $\mathbf{R_{WW}} = \sigma^2\mathbf{I}$ is the auto covariance matrix of the receivers AWGN with σ^2 is the noise variance and $\mathbf{I}$ is $M \times M$ unitary matrix.

In practice, CovM is estimated instead of the covariance matrix. We express $\mathbf{R_{YY}}$ as the CovM of N observation as:

$$\mathbf{R'_{YY}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Y}\mathbf{Y^H} \tag{4.6}$$

where $\mathbf{R'_{YY}}$ converge to $\mathbf{R_{YY}}$ for large number of samples.

CovM of the received signal from the M antenna array is typically estimated when estimating the DoA [16], [61]. For subspace-based techniques such as MUSIC [3], which is widely used and known for its superb performance particularly at low SNR levels, the EVD is applied on $\mathbf{R_{YY}}$ as a step to estimate the DoA. In other words, estimating the $\mathbf{R_{YY}}$ and its EVD is a conventional step in most of the DoA estimation algorithms.

Applying EVD on $\mathbf{R'_{YY}}$ leads to:

$$\begin{aligned}\mathbf{R'_{YY}} &= \mathbf{U_Y}\mathbf{\Lambda_Y}\mathbf{U_Y^H} \\ &= \mathbf{U_S}\mathbf{\Lambda_S}\mathbf{U_Y^H} + \mathbf{U_W}\mathbf{\Lambda_W}\mathbf{U_W^H},\end{aligned} \tag{4.7}$$

where $\mathbf{U_S}$ and $\mathbf{U_W}$ are the signal and noise subspaces unitary matrices respectively, and $\mathbf{\Lambda_S}$ and $\mathbf{\Lambda_W}$ are diagonal matrices of the eigenvalues of the signal and noise, respectively. Eq. (4.7) can be expressed as:

$$\mathbf{U_Y}\mathbf{\Lambda_Y}\mathbf{U_Y^H} = diag\left(\lambda_1, \lambda_2, ...\lambda_M\right) + \sigma^2\mathbf{I}. \tag{4.8}$$

The eigenvalues $(\lambda_1, \lambda_2, ...\lambda_M)$ with their corresponding eigenvectors $(e_1, e_2, ...e_M)$ define the signal and noise subspace as $\mathbf{U_S} = [e_1, ..., e_K]$ and $\mathbf{U_W} = [e_{K+1}, ..., e_M]$ respectively.

4.1.2 Auto Correlation Coefficient Matrix (CorrM)

In two of the proposed techniques, we exploit CorrM rather than CovM to estimate the number of applied sources. To define CorrM, we first redefine the covariance matrix in Eq. (4.5) as:

$$\begin{aligned} \mathbf{V_{YY}} &= \mathbb{E}\left[(\mathbf{Y} - \mu_\mathbf{Y})(\mathbf{Y} - \mu_\mathbf{Y})^H\right] \\ &= \mathbf{A}\mathbf{R_{SS}}\mathbf{A^H} + \mathbf{R_{WW}} - \mu_\mathbf{Y}\mu_\mathbf{Y}^\mathbf{H} \end{aligned} \tag{4.9}$$

where $\mu_\mathbf{Y} = \mathbb{E}[\mathbf{Y}]$. The elements in the diagonal of $\mathbf{V_{YY}}$ are the variances of $\mathbf{Y}$. CorrM is then given by:

$$\mathbf{C_{YY}} = (\mathbf{diag}(\mathbf{V_{YY}}))^{-\frac{1}{2}}\, \mathbf{V_{YY}}\, (\mathbf{diag}(\mathbf{V_{YY}}))^{-\frac{1}{2}}. \tag{4.10}$$

Then, we apply the EVD on $\mathbf{C_{YY}}$ which leads to

$$\mathbf{C_{YY}} = \mathbf{U_C}\mathbf{\Lambda_C}\mathbf{U_C^H}, \tag{4.11}$$

$$\begin{aligned} \mathbf{U_C}\mathbf{\Lambda_C}\mathbf{U_C^H} =& diag\left(\lambda_1^C, \lambda_2^C, ...\lambda_M^C\right) \\ &+ (diag(\mathbf{V_{YY}}))^{-\frac{1}{2}} \left(\sigma^2\mathbf{I} - \mu_\mathbf{Y}\mu_\mathbf{Y}^\mathbf{H}\right) \\ &(diag(\mathbf{V_{YY}}))^{-\frac{1}{2}}. \end{aligned} \tag{4.12}$$

The eigenvalues $(\lambda_1^C, \lambda_2^C, ...\lambda_M^C)$ with their corresponding eigenvectors $(e_1^C, e_2^C, ...e_M^C)$ define the signal and noise subspace as $\mathbf{U_S} = [e_1^C, ..., e_K^C]$ and $\mathbf{U_W} = [e_{K+1}^C, ..., e_M^C]$ respectively.

4.2 Number of Sources Estimation

In this section, we propose our simple yet efficient algorithms for the number of sources estimation. We start by explaining the traditional way of estimation, fallowed by the motivation behind our work, the decision statistics that we will be using, the principle of our proposed algorithms and finally our proposed algorithms.

4.2.1 Existing techniques

In this subsection, some of the traditional and most commonly used techniques will be discussed briefly. In specific, information theoretic approaches will be discussed to highlight their complexity and motivate the simplicity behind our work.

Akaikes information criterion (AIC) and minimum description length (MDL) are the most widely-used number of sources estimation techniques. They are ordered determination information theoretic models that use the eigenvalues of CovM to determine how many smallest eigenvalues are approximately equal. Those eigenvalues would lie in the noise subspace while others would lie in the signal subspace. Both algorithms consist of minimizing a criterion of log like-hood over the number of signals that are detectable. In this thesis, the derivation of these criteria is not stated, however the details of both of them can be found in [62]. When ordering CovM eigenvalues in a descending order, i.e., $\lambda_1 \geq \lambda_2 \geq ...\lambda_M$, AIC can be expressed as:

$$K_{AIC} = argmin_k \Bigg(-2\log \left(\frac{\Pi_{i=k+1}^{M} \lambda_i^{\frac{1}{M-k}}}{\frac{1}{M-k}\sum_{i=k+1}^{M}\lambda_i} \right)^{(M-k)N} + 2k(2M-k) \Bigg) \tag{4.13}$$

While MDL can be expressed as:

$$K_{MDL} = argmin_k \Bigg(-\log \left(\frac{\Pi_{i=k+1}^{M} \lambda_i^{\frac{1}{M-k}}}{\frac{1}{M-k}\sum_{i=k+1}^{M}\lambda_i} \right)^{(M-k)N} + \frac{1}{2}k(2M-k)\log(N) \Bigg) \tag{4.14}$$

where k is the index of the eigenvalues. In simulation results of this part, those two algorithms will be used as references for comparison with our proposed algorithms.

4.2.2 Motivation

The information theoretic approaches, AIC and MDL, are more computationally expensive than threshold-based approaches given that they need to solve the minimization problem given in Eq. (4.13) and Eq. (4.14) each time an estimation of the number of sources is needed. On the other hand, threshold-based approaches require extensive iterations in order to adjust coefficient parameters that change with several parameters such as N, M and SNR making its adjustment a tedious process.

This motivates us to propose new algorithms that strike a balance between complexity and extensive threshold adjustment. As will be shown later, the new algorithms are less complex than AIC and MDL, while they do not require threshold adjustment as in traditional threshold-based algorithms. They are a bit more complex than threshold based approach; however, that was done to overcome threshold adjusting problem which sound impractical in real life scenarios and hardware implementation.

The main contributions in this part as compared to existing include:

- Proposing four novel algorithms that use two different matrices to estimate the number of sources.
- Exploiting two different decision statistics to distinguish between noise and signal eigenvalues, i.e., Moving Increment (MI) and Moving STD (MS) .

- For the two CovM based algorithms, we define two non-reconfigurable formulas to estimate the threshold for each decision statistic. First, we find the distribution of the probability of false alarm of the MI case. We show it is a mathematically tedious process to estimate the threshold through this conventional process. We then derive the thresholds using regression analysis.
- For CorrM based algorithms, we redefine the problem as a simple maximization problem.
- We compare the performance of our proposed algorithms to the conventional high complexity AIC and MDL algorithms and show that our proposed algorithms have comparable performance at medium and high levels of SNR and better performance at low SNR values.

To the best of the our knowledge, estimating the number of sources through a maximization approach applied on the estimated eigenvalues of the CorrM, or by a single threshold formula without an adjusting coefficient applied on the eigenvalues of the CovM, have not been presented in the literature.

4.2.3 Simple yet Efficient Decision Statistics

In the proposed algorithms, two decision statistics will be used, namely MI and MS. First, we arrange the eigenvalues in an ascending order, rather than a descending order as in the case of AIC and MDL. Hence, the eigenvalues are arranged from the beginning as $(\lambda_1, \lambda_2, ...\lambda_M)$ where $\lambda_1 \leq \lambda_2 \leq ...\lambda_M$ and $(\lambda_1, \lambda_2, ...\lambda_{M-K-1})$ lay in the noise subspace while $(\lambda_{M-K}...\lambda_M)$ are in the signal subspace.

MI decision statistic

The first decision statistic is the moving increment, δ, which is simply the difference between two consecutive eigenvalues. It can be expressed as:

$$\delta_i = \lambda_i - \lambda_{i-1} \quad \text{for} \quad i = 2, 3.., M. \tag{4.15}$$

MS decision statistic

The second proposed decision statistic used as a metric to decide on the number of sources is the moving standard deviation of the estimated eigenvalues, α. The sample standard deviation, in general, is a measure of variance or difference of the sample from the mean, it can be calculated by:

$$s_M = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(x_i - u)^2}, \tag{4.16}$$

where u is the mean and M is the size of the sample or, in our case, the size of the eigenvalues involved in standard deviation calculation.

The biased standard deviation of two consecutive eigenvalues, can be calculated as:

$$STD(i) = \sqrt{(\lambda_i - u)^2 + (\lambda_{i-1} - u)^2}, \tag{4.17}$$

where u is the mean of the two eigenvalues involved which is given by:

$$u = \frac{\lambda_i + \lambda_{i-1}}{2}. \tag{4.18}$$

Then, the second decision statistics, which is the MS (α), is defined as the difference between two consecutive STDs as follows

$$\alpha_i = STD(i) - STD(i-1), \qquad \text{for } i = 3, 4, ..., M. \tag{4.19}$$

Proposition 1 *α_i can be rewritten as:*

$$\alpha_i = \frac{1}{\sqrt{2}} \left(\lambda_i - 2\lambda_{i-1} + \lambda_{i-2} \right). \tag{4.20}$$

Proof. The proof for *Proposition 1* is provided in Appendix A ∎

4.2.4 Principle of the Proposed Algorithms

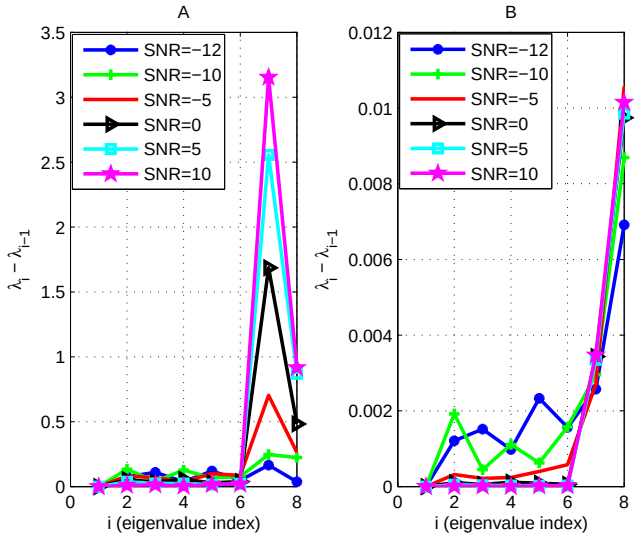


Figure 4.2: Change in Eigenvalues of MI with different SNR for (A) CorrM (B) CovM

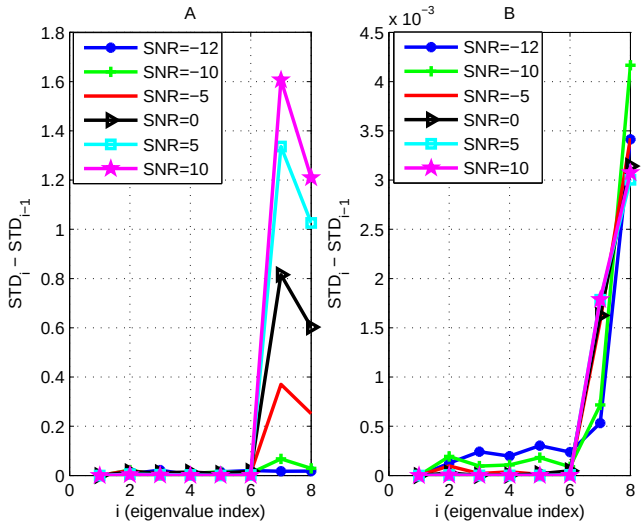


Figure 4.3: Change in Eigenvalues of MS with different SNR for (A) CorrM (B) CovM

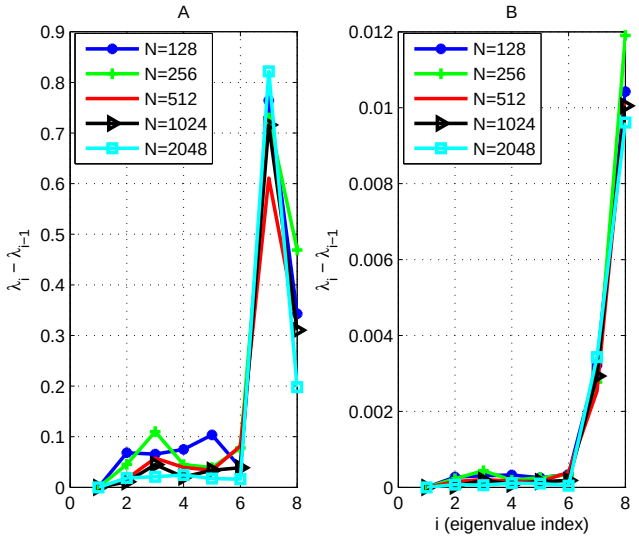


Figure 4.4: Change in Eigenvalues of MI with different number of samples for (A) CorrM (B) CovM

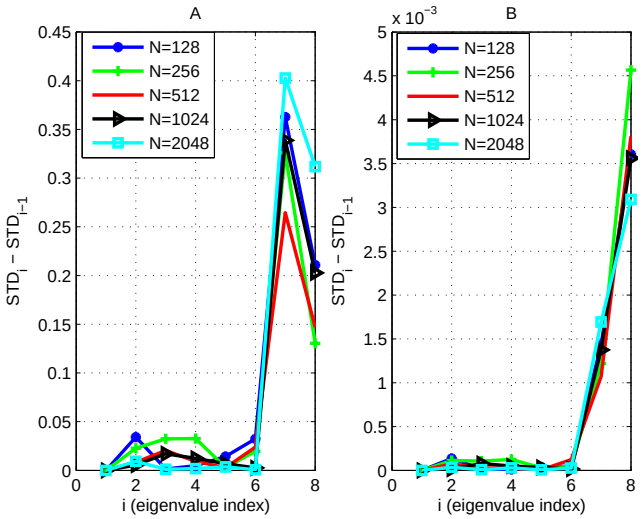


Figure 4.5: Change in Eigenvalues of MS with different number of samples for (A) CorrM (B) CovM

It can be inferred from Eq. (4.8) and Eq. (4.12) that since the eigenvalues of the signal subspace contain both signal and noise powers, the values of sources' signal eigenvalues are expected to be higher than noise eigenvalues at moderate and high SNR values. At the same time, the noise eigenvalues are expected to be comparable to one another.

The main advantage of using EVD of CorrM in Eq. (4.12) rather than the EVD of CovM in Eq. (4.8) is that the difference between signal eigenvalues and the noise eigenvalues is more accentuated, which leads to an easier and more efficient estimation

of the number of sources, particularly at low SNR values. Moreover, the mathematical operation applied to estimate a decision statistic, which is then used to decide on the number of sources, can be as simple as our proposed MI or MS rather than the complicated decision statistic for the AIC and MDL given in Eq. (4.13) and Eq. (4.14).

To illustrate the concepts applied in the previous paragraph, we plot the MI and MS of the estimated eigenvalues of CorrM and CovM for different SNR values in Fig. 4.2 and Fig. 4.3, respectively, and different number of collected samples in Fig. 4.4 and Fig. 4.5, respectively. The simulation parameters for the first two figures are 8 element antenna array, 2 applied signals, 1024 samples and different SNR values. The simulation parameters for the next two figures are the same except that the SNR is kept fixed at -7 dB and the number of samples changed from 128 to 2048.

From those figures, one can see that for our first category, CorrM based eigenvalues, the jump in the decision statistic when moving from the noise subspace to the signal subspace is always the highest. The decision statistics then starts to decrease. In other words, the highest increment in the decision statistic always happens when moving from noise subspace to signal subspace. On the contrary, when using the same two decision statistics with the eigenvalues of CovM, a threshold needs to be set at the first jump between the noise and signal subspaces as this jump is not necessarily the highest. This implies that when using the decision statistics of CorrM, the problem is transformed into a simple maximization problem, where the index at which the highest jump occurs is searched for, while for the case of using the decision statistic of CovM, the decision statistics should be compared to a threshold to decide on the number of sources. This threshold is either found by extensive simulations, which was used traditionally, or by an equation that handles all parameters that affect the threshold.

It should be noted here that Appendix B presents some simple examples that will explain the proposed algorithm principle in two working examples and how different algorithms detect the number of sources.

4.2.5 Proposed Algorithms

In this section, we present our four proposed algorithms based on two categories: CovM and CorrM-based approaches. In each category, we propose to use MI and MS decision statistic which will end up with four proposed different algorithms divided into two subsections: one for each category. In other words, each category will be represented in a subsection and each one will propose two algorithms. The pseudo code for the algorithms is summarized in Algorithm 1. At the beginning, we generate two, or more, signals from different sources, the signal will be transmitter using the wireless channel and received by the antenna array elements at the receiver. In simulation, such process is done by generating any signal, multiplying it by the steering vector, expressed in Eq. (4.4), and adding the noise and channel effect resulting from transmitting the signal. At the received, the signals are collected from the antenna array and processed to estimate the number of sources and DoA. First, the correlation or covariance matrix is calculated and EVD is applied to estimate the eigenvalues and eigenvector for number of sources and DoA estimation, respectively. After that, the eigenvalues are arranged in ascending order, the decision statistic is estimated and each algorithm starts its process separately, as will be seen in the following subsections.

CorrM Based Algorithms

In the first category, CorrM eigenvalues are used to estimate the number of sources by the two decision statistics, MI and MS. As shown in section 4.2.4, a simple maximization problem can be found to estimate the number of sources using both of our

Algorithm 1 Number of Sources Estimation Algorithm

Assuming

M: number of elements % *array elements of the circular array*

N: number of samples % *samples to be collected at each antenna*

step 0:Collect The Received Signal

1: Generate a QPSK signal *%QPSK signal to be sent by the sources*
%To apply the effect of receiving those sources at antenna elements

2: Multiply by the circular array steering vector

3: Add a noise *%Channel and Noise Effect*

The received signal is now

$$\mathbf{Y} = \mathbf{AS} + \mathbf{W},$$

where $Y \in \mathbb{C}^{M\times N}$, $\mathbf{A} \in \mathbb{C}^{M\times K}$,$\mathbf{S} \in \mathbb{C}^{K\times N}$, $\mathbf{W} \in \mathbb{C}^{M\times N}$, with N being the total number of collected samples and $\mathbb{C}$ is the set of complex numbers.

step 1:Get The Eigenvalues

1: Calculate CovM or CorrM matrices by (4.7) or (4.10) *%Matrices Calculation*
%Eigenvalue Decomposition Operation

2: Apply EVD to the matrices (4.8) and (4.12)

3: Extract The Eigenvalues:

Eigenvalues $(\lambda_1, \lambda_2, ...\lambda_M)$ Eigenvalues $(\lambda_1^C, \lambda_2^C, ...\lambda_M^C)$

4: Arrange Eigenvalues in ascending order:

$\lambda_1 \leq \lambda_2 \leq ...\lambda_M$ $\lambda_1^C \leq \lambda_2^C \leq ...\lambda_M^C$

step 2: Estimate the Number of Sources

1: Apply one of decision statistics Eq. (4.20) Eq. (4.15) *%Applying MS and MI decision statistics*

2: Find the location of the first signal eigenvalue *%Find either the highest increment in decision statistics or the first increment that exceed a threshold*

1: **for** $i = 2\ i < M$ **do**
Calculate δ_i or α_i
2: **if** $(\delta_i - \delta_{i-1} > \xi_n)$ **then**
j=i
Break ;
3: **end if**
4: **end for**

1: **for** $i = 2\ i < M$ **do**
Calculate δ_i or α_i
2: **end for**

$$j = \arg\max_i \delta_i.$$

3: Find The estimated number of sources: *%number of left eigenvalues*
K=M-i+1

metrics, MS and MI. Hence, the highest increment would then imply the shift between the noise eigenvalues to the signal eigenvalues.

For MI, expressed in Eq. (4.15), the index at which this shift happens can be estimated as:

$$j = \arg\max_i \delta_i.$$

In this case, the number of sources can be given by $K = M - j + 1$.

Similarly for MS, expressed in Eq. (4.20), the highest index at which the shift between the noise eigenvalues and the signal eigenvalues can be estimated as:

$$j = \arg\max_i \alpha_i$$

Consequently, the number of sources can be given by $K = M - j + 1$.

CovM Based Algorithms

In the second category, CovM eigenvalues are used and a threshold is found to distinguish between noise and signal eigenvalues.

First, we use the MI of the eigenvalues, δ_i, and compare it to a threshold. Hence, the number of sources is estimated when:

$$\delta_i \geq \xi_n. \tag{4.21}$$

where δ_i is expressed in (4.15) and ξ_n is a threshold that depends on the noise power, SNR value, number of samples, number of elements and received power. The number of sources can be determined when δ_i accedes ξ_n and at that point $K = M - i + 1$.

To estimate the threshold ξ_n, the probability distribution of δ_i has to be derived first. Thus, for a given probability of false alarm (P_f), ξ_n can be estimated as:

$$P_f = Pr\left(\delta_i \geq \xi_n | i \leq M - K\right). \tag{4.22}$$

In case of noise only, i.e., $K = 0$, the received samples follow $\mathcal{N}(0, \sigma^2)$, and therefore $\mathbf{R_{YY}}$ follows a Wishart distribution with N degrees of freedom and variance Σ, i.e., $\mathbf{R_{YY}}$ follows $\mathcal{W}(N, \Sigma)$ [63]. The empirical distribution function of the noise eigenvalues can be expressed by [63], [64]

$$F^R(\lambda) = \frac{\text{number of eigenvalues of } \mathbf{R_{YY}} \leq \lambda}{M} \tag{4.23}$$

according to [64], F^R converges to f^W with a high probability when the number of samples $N \to \infty$. f^W follows a Marcenko-Pastur density function [65], which can be expressed as:

$$f_W(\lambda) = dF^W(\lambda) = max(0, (1 - G))\delta(\lambda) + \frac{\sqrt{(\lambda - a_-)}\sqrt{(a_+ - \lambda)}}{2\pi\sigma\lambda(1/G)}\Pi_{[a_-, a_+]}(\lambda) \tag{4.24}$$

where G is the samples to elements ratio (N/M), $a_\pm = \sigma(1 \pm 1/\sqrt{G})^2$, $\delta(\lambda)$ is the delta function, and the function $\Pi_{[a,b]}(\lambda)$ equals 1 for $a \leq \lambda \leq b$ and 0 otherwise. We derive the probability distribution of the noise eigenvalues of δ_i in *Proposition 2*.

Proposition 2 *The probability distribution of δ_i can be given by (4.25).*

Proof. The proof for *Proposition 2* is provided in Appendix C. ∎

Consequently, (4.22) can be rewritten as:

$$\begin{aligned} P_f =& Pr\left(\delta_i \geq \xi_n | i \leq M - K\right) \\ =& 1 - F_{\delta_i}(\xi_n). \end{aligned} \tag{4.26}$$

$$
\begin{aligned}
F_{\delta_i}(\xi_n) = & \left(\frac{G}{2\pi\sigma}\right)^i \int_{-\infty}^{\xi_n} \frac{(M-K)!}{(i-2)!(M-K-i)!} \int_{a_-}^{a_+} \frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda} \\
& \left\{ \frac{1}{4\sqrt{a_-}\sqrt{a_+}} \left[2\arcsin\left(\frac{-2\lambda+a_-+a_+}{a_--a_+}\right) a_-^{3/2}\sqrt{a_+} + 2\arcsin\left(\frac{-2\lambda+a_-+a_+}{a_--a_+}\right) a_+^{3/2}\sqrt{a_-} + 4a_+a_- \right.\right. \\
& \arctan\left(\frac{1}{2}\frac{2a_-a_+ - a_+\lambda - a_-\lambda}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+-\lambda)}\sqrt{(\lambda-a_-)}}\right) + \pi a_-^{3/2}\sqrt{a_+} + \pi\sqrt{a_-}a_+^{3/2} - 2a_-a_+\pi \\
& \left.\left. + 4\sqrt{a_-}\sqrt{a_+}\sqrt{\lambda-a_-}\sqrt{a_+-\lambda} \right] \right\}^{i-2} \left(\frac{\sqrt{(\lambda+\delta)-a_-}\sqrt{a_+-(\lambda+\delta)}}{\lambda+\delta}\right) \\
& \left(1 - \frac{G}{2\pi\sigma}\left\{ \frac{1}{4\sqrt{a_-}\sqrt{a_+}} \left[2\arcsin\left(\frac{-2(\lambda+\delta)+a_-+a_+}{a_--a_+}\right) a_-^{3/2}\sqrt{a_+} \right.\right.\right. \\
& + 2\arcsin\left(\frac{-2(\lambda+\delta)+a_-+a_+}{a_--a_+}\right) a_+^{3/2}\sqrt{a_-} + 4a_+a_- \arctan\left(\frac{1}{2}\frac{2a_-a_+ - a_+(\lambda+\delta) - a_-(\lambda+\delta)}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+-(\lambda+\delta))}\sqrt{((\lambda+\delta)-a_-)}}\right) \\
& \left.\left.\left. + \pi a_-^{3/2}\sqrt{a_+} + \pi\sqrt{a_-}a_+^{3/2} - 2a_-a_+\pi + 4\sqrt{a_-}\sqrt{a_+}\sqrt{(\lambda+\delta)-a_-}\sqrt{a_+-(\lambda+\delta)} \right] \right\} \right)^{M-K-i} d\lambda d\delta \quad (4.25)
\end{aligned}
$$

Conventionally, in order to estimate the threshold ξ_n, Eq. (4.26) has to be solved for a desired P_f. This requires calculating the integral in (4.25). Which is highly prohibitive due to the following:

1. For a given P_f, it requires the total number of noise eigenvalues $(M-K)$ to be known a prior in order to solve for ξ_n. This is not feasible since the total number of sources (K) is our unknown to be estimated.

2. Another approach is to try to minimize P_f by differentiating it twice, once with respect to K and another with respect to ξ_n. Then equating the output to zero to solve for the two unknown. This, as can be seen from (4.25), is a mathematically tedious process and will have a high computational complexity.

Alternatively, we estimate the threshold in Eq. (4.22) through multiple linear regression and least square fitting approaches. A regression model is a statistical process that is used to estimate the relationship between multiple explanatory variables and one dependent variable that depends on them. That is, giving some data sets, or samples, that relate multiple variables to one dependent variable, a linear function can be found to estimate the dependent variable from those multiple variables. This function tries to fit all data points in one equation using linear least squares fittings or any other fitting approach [66], [67], [68]. In our case, the dependent variable is

the threshold that we need to estimate having multiple other variables, such as SNR value, number of samples, number of elements and the received power, i.e., we find a function for the threshold depending on all those variables.

Using the previous concept in Section 4.2.4, depicted in Fig. 4.2.(B) and Fig. 4.4.(B), the threshold can be defined in multiple regions as follows:

$$\xi_n = P_s \times \begin{cases} \rho, & \text{for} \quad SNR > 2\ dB \\ \rho/8, & \text{for} \quad N > 10000 \\ \rho/2, & \text{for} \quad SNR > -2\ dB \quad N < 100 \\ \rho/6, & \text{for} \quad SNR < -6\ dB \quad N > 1000 \\ \rho/4, & \text{for} \quad elsewhere \end{cases} \tag{4.27}$$

Taking some samples from those regions and applying the concept of least square fitting [66] to estimate the threshold function, the threshold can be defined by:

$$\xi_n = P_s\ (-7.75.\ 10^{-3}\ SNR - 3.77.\ 10^{-5}\ N + 1.05) \tag{4.28}$$

This threshold takes into consideration specific regions of interest that include: SNR values $SNR \in [-20dB : 40dB]$, received power $P_s \in [0dBm : -100dBm]$, number of samples $N \in [2^6 : 2^{14}]$, and number of antenna array element of less than 30. Outside these regions the threshold might fail to estimate the number of sources correctly and hence requires further analysis. However, it is a very rare case that the threshold goes beyond these regions, especially in practical implementations.

Another technique is to use MS instead of MI and set a threshold to distinguish between noise and signal eigenvalues. Hence, the decision is taken on the number of sources

$$\alpha_i \geq \gamma_n. \tag{4.29}$$

where α_i is expressed in (4.20) and γ_n is a threshold that depends on the number of samples, the number of elements and the received power. The number of sources can be detected when α_i accedes γ_n and, at that point, $K = M - i + 1$.

The probability of false alarm for MS can be given by:

$$P_f = Pr\left(\alpha_i \geq \gamma_n | i \leq M - K\right). \tag{4.30}$$

As in the MI case, in order to estimate the threshold γ_n, the probability distribution function $F_{\alpha_i}(\gamma_n)$ has to be expressed first, then solve (4.30) for a given P_f. As a matter of fact, $F_{\alpha_i}(\gamma_n)$ is expected to be more complicated to solve than $F_{\delta_i}(\xi_n)$. Again, we find the threshold in (4.30) through a least squares fitting approach [66]. Using the previous concept in Sec 4.2.4, depicted in Fig. =4.3.(B) and Fig. 4.5.(B), the threshold can be defined with multiple regions using:

$$\gamma_n = \begin{cases} 0.45P_s, & \text{for} \quad 500 < N < 1000 \quad M \leq 8 \\ P_s, & \text{for} \quad 100 < N < 500 \quad M > 8 \\ 0.75P_s, & elsewhere \end{cases} \tag{4.31}$$

Taking some samples from those regions and using a linear least squares regression to estimate the threshold function, the threshold can be defined by:

$$\gamma_n = P_s \left(2.5.\ 10^{-5}.\ N - 0.12.\ 10^{-4}.\ M + 0.66\right) \tag{4.32}$$

4.3 Direction of Arrival (DoA) Estimation

One of the main targets of this work is estimating DoA which will be used later in the breaching scheme. Hence, two methods are considered for DoA estimation, namely MVDR and MUSIC. MVDR was chosen for its simplicity and better resolution that Capon which lies in the simple traditional algorithms [16]. MUSIC was selected for its

high resolution and popularity among subspace algorithms. Algorithm 2 highlights both MVDR and MUSIC algorithm using CovM and the following sections explain both algorithms in details. The same steps are applied to CorrM by replacing CovM calculations with CorrM.

As with number of sources estimation, Section 4.2.5, the first step is to collect data from the antenna elements either by simulation or from actual hardware. Then, data is organized in a matrix and the covariance matrix or the correlation matrix is calculated. After that, EVD is applied as MUSIC algorithms will rely on it and the noise and subspace are separated; on the assumption that the number of sources is known or estimated. MUSIC will use the orthogonality between the noise subspace and the signal to search for DoA, while MVDR will use the orthogonality fact between the inverse covariance matrix and the signal to search for DoA. The details of such estimation procedure will be explained in the following sections for each algorithm.

4.3.1 MUSIC

Since 1986, when it was first proposed in [17], MUSIC has been the most well known and widely used algorithm for DoA estimation. The reason why such a method is popular goes back to its high resolution in estimating DoA (the angles can be as close as 5° in good SNR conditions) and its high tolerance to low SNR values, which is better than other DoA estimation algorithms. However, this comes with the cost of high computational complexity that resulted from EVD which is a complex operation and requires high computation resources. Hence, its complexity is compromised by its high performance. Since a lot of attempts have been found in the literature to reduce this complexity [69], [70], MUSIC complexity will not be considered when choosing between algorithms.

As shown in Algorithm 2, MUSIC depends on EVD operation which is done after knowing the number of sources and results in eigenvalues and eigenvectors. After

Algorithm 2 MUSIC & MVDR Estimation Algorithms

Assuming

M: number of elements % *array elements of the circular array*

N: number of samples % *samples to be collected at each antenna*

K: number of applied signal with ϕ_i angle % *either known or estimated*

step 0:Collect The Received Signal

1: Generate a QPSK signal %*QPSK signal to be sent by the sources*
%*To apply the effect of receiving those sources at antenna elements*

2: Multiply by the circular array steering vector

3: Add a noise %*Channel and Noise Effect*

The received signal is now

$$\mathbf{Y} = \mathbf{AS} + \mathbf{W},$$

where $Y \in \mathbb{C}^{M\times N}$, $\mathbf{A} \in \mathbb{C}^{M\times K}$,$\mathbf{S} \in \mathbb{C}^{K\times N}$, $\mathbf{W} \in \mathbb{C}^{M\times N}$, with N being the total number of collected samples and $\mathbb{C}$ is the set of complex numbers.

step 1:Get The Eigenvectors

1: Calculate CovM or CorrM matrices by (4.7) or (4.10) %*Matrices Calculation*
%*Eigenvalue Decomposition Operation*

2: Apply EVD to the matrices (4.8) and (4.12)

3: Extract The Eigenvector:
$\mathbf{U_S} = [e_1, ..., e_K]$ is the signal subspace and $\mathbf{U_W} = [e_{K+1}, ..., e_M]$ is the noise subspace

step 2: Estimate DoA

1: $\phi \in [0 : 360]$

2: **for** $i = 1 \rightarrow length(\phi)$ **do**
Find the steering vector as in Eq. (4.4) % *To search the full circular spectrum*
Apply MUSIC or MVDR spectrum where $a(\phi)$ is the steering vector at the current ϕ_i %*Based on the orthogonality fact*

$$P_{MUSIC_i}(\phi_i) = \frac{1}{a(\phi_i)^T U_W U_W^T a(\phi_i)}$$

$$P_{MVRD_i}(\phi_i) = \frac{1}{a(\phi_i)^T R_{YY}^{-1} a(\phi_i)}$$

3: **end for**

4: The location of first K peaks of the resulting P_{MUSIC} or P_{MVDR} are the estimated DoA

knowing the number of sources either in prior or by estimation, the algorithm applies EVD on CorrM or CovM to decompose into signal and noise subspace. The signal subspace is $U_S = [e_K, ..., e_M]$, while the noise subspace is $U_W = [e_1, ..., e_{N-K}]$, where K is the number of sources estimated in the previous algorithm, explained earlier in Section 4.2.

The algorithm depends on the fact that the noise and signal subspace are orthogonal to each other. Hence, using the Euclidean distance equation, $||a(\phi)U_W||^2$ should be 0 at the angles of arrivals. So, MUSIC attempts to find the steering vectors which are as orthogonal to noise subspace as possible. It searches the spectrum to find all steering vectors and apply

$$\begin{aligned} P_{MUSIC}(\phi) =& \frac{1}{||a(\phi)U_W||^2} \\ =& \frac{1}{a(\phi)^T U_W U_W^T a(\phi)} \end{aligned} \tag{4.33}$$

for all ϕ in the spectrum which is $[1 : 360]$ in UCA where $a(\phi)$ is defined in Eq. (4.4). Then, the algorithm searches for the location of maximum K peaks in the spectrum where K is the number of sources as estimated before. The variance in this estimation approaches the Cramer-Rao lower bound when SNR approach infinity [71].

4.3.2 MVDR

MVDR is a traditional estimation algorithm that does not depend on subspace decomposition as MUSIC does and, assumingly, it is less complex than MUSIC. It was presented for the first time in [16] in 1969 on a modification of Capon algorithm. It has a better resolution than Capon (reach about 20° in good SNR conditions), however its SNR tolerance is less than other subspace algorithms.

As shown in Algorithm 2, the idea behind MVDR is to use the inverse of CovM or CorrM matrices of the received signal. Multiplying that inverse by the steering vector should result in close to 0 value at the directions of arrivals. Therefore, MVDR

attempts to multiply the steering vector of the spectrum by the inverse CovM and find the maximum peaks which will represent DoAs. It searches the spectrum to find all steering vectors and apply

$$P_{MVRD}(\phi) = \frac{1}{a(\phi)^T R_{YY}^{-1} a(\phi)} \tag{4.34}$$

For all ϕ in the spectrum which is $[1:360]$ in UCA where $a(\phi)$ is defined in Eq. 4.4. Then, the algorithm searches for the location of maximum K peaks in the spectrum where K is the number of sources as estimated before. This can approach MUSIC performance when SNR tends to infinity.

4.4 Distance Estimation

Due to its easy implementation and high availability, RSSI reading is the most common technique for distance estimation in wireless channels. RSSI is a measure of the power of the received signal which can be estimated easily in most of the recent transceivers. These transceivers are already equipped with circuits to estimate RSSI value; and hence, this can be the simplest approach to estimate the distance from the received signal.

In order to estimate the distance, models like free space propagation and the two-ray ground reflection were presented in the literature, however the log distance path loss model is the most commonly used model for such estimation. Unlike other models, the log distance path loss model can be used for both indoor and outdoor environments. The model maps RSSI readings to distances using:

$$P_r = P_r(d_0) - 10\ n_p\ log_{10}\frac{d}{d_0}\ +\ X_\sigma \tag{4.35}$$

Where P_r is the average received power, measured in dBW, and that is usually RSSI value. $P_r(d_0)$ is the RSSI value at a reference point d_0, n_p is the path loss exponent,

Table 4.1: Derived Path Loss Model for Different Environments

Environment	**Average Power Equation**
Line of Sight Indoor	$P_r = -16.21log_{10}(d) - 40.412$
Non Line of Sight Indoor	$P_r = -23.41log_{10}(d) - 48.67$
Outdoor	$P_r = -13.196log_{10}(d) - 32.6600$

d is the required distance and X_σ is a normally distributed variable with 0 means and σ standard deviation. Using a reference distance of 1 meter, Eq. (4.35) can be expressed as:

$$P_r = -10\ n_p\ log_{10}(d)\ +\ C \tag{4.36}$$

Where C is $P_r(1) + X_\sigma$. Hence, the distance can be estimated by:

$$d = 10^{-\frac{RSSI-C}{10n_p}} \tag{4.37}$$

In general, n_p and C change in different environments and circumstances. To have an accurate estimation of this distance, those variables need adjusting from time to time in order to accommodate environmental changes. In [72], the researchers characterized the path-loss exponent and the constant C in 3 categories: line of sight indoor environment, non-line of sight indoor environment and outdoor environment. Table. 4.1 summarizes derived path loss exponent model for the three environments where they used up to 10 m distance.

However, for accurate estimation of the distance, this equation still needs adjusting of the included parameters as those parameters change with different environmental factors. This can be done prior to the breaching scheme and that will be the only parameters in this thesis that need prior adjusting. To adjust those parameters, different distances needed to be considered and multiple readings of the corresponding RSSI values are measured. Then, the average RSSI value is calculated for each distance and the equation is found using least square fitting linear regression model.

Thus, for accurate estimations, Eq. (4.35) will need adjusting when implementing the breaching scheme in hardware at a later stage of this thesis. Simulations of these algorithms can be done only in a specific environment where the coefficients are known, RSSI can be estimated from the received signal and the distance can be calculated directly from the path-loss model equation. In the context of this thesis, we will not be estimating the distance in the presented way for two reasons: first, as we are working with simulation, the coefficients for the environment cannot be found except if we implement a specific one, i.e. we will need to assume some existing environments and use their coefficients, and then, we want to generalize our breaching scheme for every environment. Hence, the simulation will not include distance estimated by RSSI; however, that is assumed to be easily implemented with the previously stated procedure when applying the scheme to our hardware environments. Instead, in this thesis we surveyed distance estimation algorithms' accuracy, assumed some exact distances and then added random errors based on algorithms' accuracy. In particular, we surveyed the most well-known algorithms for distance estimation which are the RSSI-based and ToA based algorithms. RSSI-based algorithms error can reach 20% of the actual distance, even with good coefficients estimation, [73] [74], while ToA based algorithms can reach 10% error with good clock synchronization between the communication nodes [75]. We assume some actual distances and add random error that does not exceed 20% in case of RSSI and 10% in case of ToA. Then, we used the resulting distance for breaching scheme and key generation as will be seen later in simulation results.

4.5 Key Generation Scheme

After estimating the distance, the secret key should be generated based on a specific scheme, which is agreed between Alice and Bob and we assume Eve to know the scheme in prior. In this work, we follow a simple key generation scheme presented

in [60] in the same project that this thesis is part of. We excluded information reconciliation and privacy amplification to be tackled at a later stage of the project; however, we will stop at the generated bits and evaluate them by the bit mismatch ratio as will be seen later. Algorithm 3 shows the modified scheme for our case where we excluded the last three steps and the channel measurement estimation.

Algorithm 3 Key Generation Scheme

step 0: Initialization

Bob, Alice, and Eve collect RSSI values from multiple beacons
They average their collected values
Estimate the distance Based on the averaged value.
They repeat that for a large number of iteration (100)

step 1: Uniform Quantization

Alice, Bob and Eve Quantize their bits based on uniform quantization

$$Y = Q(X) \quad \text{for} X \in (d_i, d_{i+1})$$

step 2: Encoding

They encode each quantized value separately and generate their keys accordingly

In this scheme, RSSI values are collected and the distance is estimated for averaged RSSI values. Then, we have three common sources of randomness, i.e. the estimated distances for each node, and the first step is to quantize them into bit streams to be used as a secret key for Alice and Bob and estimated key for Eve. Uniform quantization, [76], was used in which:

$$Y = Q(X) \quad \text{for} X \in (d_i, d_{i+1}) \tag{4.38}$$

where d is an interval and X is the input which is the estimated distance in our case. In this quantization, the spaces along the x-axis (time) and the y-axis (distance) is uniformly distributed. Then, we encode the bits as in [77] where each quantized value is encoded with multiple values so that we do not end in a high bit error mismatch (BMR). In our case, we choose 7 quantization bits and 2 encoded bits. After

estimating the distance, we divide it by the maximum estimated one and multiply by a $2^{Quantization\ Bits} - 1$. Then this quantized bit is transferred to a bit stream and the encoding bits are attached to the end of each quantized bits.

In an actual key generation, removing the last two stages will result in a high BMR that might be unacceptable in practice; however, we are concerned with the parameters estimation and how close the generated keys are to each other. The last two stages will be considered in future extensions; however, our objective currently is just to compare between the generated keys based on the distance only.

4.6 Breaching Scheme

In this section, we discuss how the breaching of distance based security schemes can happen. To start, we first define our communication nodes as Alice and Bob and our eavesdropper as Eve. To estimate the direction of arrival, we used a circular array with 8 antenna elements. Hence, Eve will need to be equipped with a circular array with 8 elements, as was shown in Fig. 4.1. However, both Alice and Bob can be equipped with only one antenna element to estimate the distance between them. Fig. 4.6.(A) demonstrates a simple overview of our system without showing the equipped antennas. In this figure, node 1 and 2 are Alice and Bob respectively while node e is Eve.

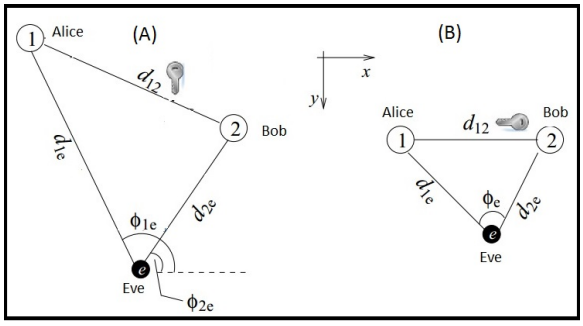


Figure 4.6: Breaching Scheme Overview

As can be seen in the figure, node 1 and 2 communicate using their secret key that is based on the distance between them, i.e. based on d_{12}. Hence, the distance between them is first estimated, as in Section 4.4 or any other distance estimation algorithm, before they start their secret communication. Node e is listening to the communication and trying to breach the system by estimating the key. Since it is equipped with the circular array, it can estimate the direction of arrival of both Alice and Bob. We denote the direction of arrival as ϕ_{1e} and ϕ_{2e}. Then Eve can estimate its distances from both Alice and Bob which are donated by d_{1e} and d_{2e}, respectively, using the estimation algorithm in section 4.4.

As a result, Eve can form a triangle as in Fig. 4.6.(B) where ϕ_e is the difference between ϕ_{2e} and ϕ_{1e} and d_{1e} and d_{2e} are estimated. Using that triangle, Eve can use the cosine law to estimate d_{12}, by having an angle and 2 sides and calculating the third side. In other words, at an instant of time, t, Eve can estimate the distance by:

$$d_{12} = d_{1e}^2 + d_{2e}^2 - 2 \times d_{1e} \times d_{2e} cos\left(\phi_e\right) \tag{4.39}$$

Now ideally, Alice, Bob and Eve need to estimate d_{12} exactly the same. In this way, the generated key between Alice and Bob and the breaching key by Eve are identical and, they do not need further processing. However, practically as the distances are estimated, they can never be exact and always have some errors. As a result, Alice and Bob will have some bit mismatch between their generated keys and they need further processing to agree on a final key. Further processes include information reconciliation and privacy amplification which are out of the scope of this thesis.

In the breaching scheme, for the number of sources estimation, two of the proposed algorithms, MMS_{CORR} and MMI_{COV}, will be used as they showed the best performance. MUSIC will be used for DoA estimation as it showed a superior performance compared to MVDR. For distance estimation, two scenarios will be used:

- **Scenario 1: Distance estimation Based on RSSI**

In this scenario, the distance is assumed to have a fixed value and a random error of less 20% is added, as RSSI-based distance algorithms can reach 20% error even with good configuration. Results are expected not to be good when DoA and the number of sources are not estimated correctly as the distance will be far from the actual distance. However, when the number of sources and DoA are estimated correctly, the estimated distances will be so close to the actual one and results are highly comparable to Alice and Bob distance estimation.

- **Scenario 2: Distance estimation Based on ToA**

  In this scenario, the distance is assumed to have a fixed value and a random error of less 10% is added, as ToA based distance algorithm can reach 10% error even with perfect clock synchronization. Results are expected to be better than the first scenario as the error is less in this case. Even though it was stated earlier that ToA will not be used in the context of this project, it was simulated to check on its performance, compared RSSI-based algorithm. ToA distance estimation might be used between the communicating nodes, as they can synchronize their clocks based on their communication, and in the case of powerful adversary where the eavesdropper can synchronize his close even with passive eavesdropping.

After that, Alice_Bob distance is estimated and the key can be generated based on section 4.5 where we assume the distance as an input to the key generation scheme. As will be seen later in simulation results, the estimated distances will be evaluated by comparing it to the actual one, as this will help in analyzing the scheme and test its limitations. Then the generated key will be compared to a reference one in terms of the bit mismatch in order to compare Eve's keys to Alice and Bob ones. After this stage, if Alice and Bob are using the public channel to communicate for information reconciliation and privacy amplification, Eve can listen to their communication and fix its key accordingly; however, as stated earlier this is out of the scope of this work.

It is worth noting here that all the breaching scenarios in this thesis take only two sources into consideration. Having more than two sources can end up estimating more than one distance and the eavesdropper needs to consider them all. This case will be discussed in the future direction section.

Chapter 5

Implementation and Results

In this chapter, we start by explaining how the implementation for our algorithms was done, going through the simulation results for the number of sources and DoA algorithms and finally we end by simulation results for the breaching and key generation schemes.

5.1 Implementation

The implementation and simulation of the number of sources and DoA algorithm were done using MATLAB. The reason why MATLAB is chosen is for its easy implementation when dealing with matrix operation and signal processing, which is involved almost everywhere in our algorithms. Besides, MATLAB is considerably the best when dealing with lower layer communication algorithms due to its embedded functions such as Communication toolbox. Finally, MATLAB is supported by WARP hardware, which will be used later for hardware implementation of the project. Thus, dealing with MATLAB in the simulation would ease the process of applying with algorithms to hardware and fast the hardware and software implementation.

For number of sources estimation, as in Algorithm 1, the implementation starts by generating a set of random binary numbers and modulating them by Quadrature Phase Shift Keying (QPSK) modulation using the Communication toolbox in MATLAB. Then, the white Gaussian noise is added by the SNR value of interest and the new matrix is in the form of the received signal as in Eq. (4.1) which will be processed to estimate the number of sources estimation.

Based on the previous estimate and the eigenvector that is gotten from the covariance matrix, the new signal and noise subspaces are created. The noise subspace is the first $M - K$ columns of the eigenvector where K is the estimated number of sources and the signal subspace are the last K columns of the eigenvector. Then, DoA is estimated for both MVDR and MUSIC as was presented in Algorithm 2 where K peaks of the circular array spectrum are searched for.

The distance estimation algorithm depends on the environment and the measured RSSI by the hardware. Hence, this algorithm will not be simulated; however, it will be assumed in order to simulate the breaching algorithm. For the distance estimation, the traditional method that is based on RSSI was used in indoor and outdoor environments and this method has been implemented previously in the same lab so simulating it in software will not add a value. Besides, the coefficient in the algorithm is environment dependent and needs to be found in the lab with the hardware which is another reason for not simulating the distance based algorithm.

Hardware implementation is an extension of this thesis that was already started. Section 6.2 will handle some initial setup of the hardware used and some initial results of DoA estimation with their problems and proposed solutions.

5.2 Simulation Results

Simulation results are done to test DoA and the number of sources estimation algorithms performance in different scenarios. Our objective is to compare algorithms and choose the best to be used in the breaching scheme. For DoA estimation, MUSIC and MVDR are simulated and the end results show that even though MUSIC is more complex, it had a better performance in terms of resolution and low SNR tolerance. Hence, MUSIC will be used in the breaching algorithm simulation results. For the number of sources estimation, MDL, AIC, and the four proposed algorithms are included in the simulation to choose the best to be used in the actual breaching

algorithm. Two of the proposed algorithms, $\mathbf{MS_{CORR}}$ and $\mathbf{MI_{COV}}$, show the best performance among others, even better than AIC and MDL at some cases, while they are much less complex. Hence, they will be used in simulations of the breaching scheme.

Performance metric to be used for comparison is the percentage error rate which can be expressed by the number of successful estimations divided by the total number of estimations. It is expressed as:

$$error\ rate = \left(1 - \frac{number\ of\ successes}{number\ of\ runs}\right) \times 100 \tag{5.1}$$

Another performance metric that will be shown in some simulations is the confidence interval. This will help evaluating the population parameter, or the estimated values for each algorithm, the range of estimates and how close they are to the actual values. In this metric, upper and lower bounds are found and the confidence interval is the range between them. Hence, it can be expressed as:

$$CI \in [\bar{X} - n\frac{\sigma}{L}, \bar{X} + n\frac{\sigma}{L}] \tag{5.2}$$

where $\bar{X}$ is the mean of the estimated values, σ is the standard deviation, L is the number of sample, and n is a population parameter which is set to 1.96 in case of 95% confidence interval.

The simulation was running for 10000 and the average error was gotten and used in all fallowing figures unless otherwise stated.

5.2.1 Number of Sources Estimation

The simulation results for number of sources estimation are done to test the algorithms in different scenarios that include different number of samples, SNR values, the number of applied signals and the number of elements that construct the array.

Table 5.1: Figures Legend Abbreviations for Number of Sources Estimation

Abbreviation	Algorithm
AIC	AIC Algorithm
MDL	MDL Algorithm
$\mathbf{MS_{CORR}}$	Proposed MS for CorrM Algorithm
$\mathbf{MI_{CORR}}$	Proposed MI for CorrM Algorithm
$\mathbf{MS_{COV}}$	Proposed MS for CovM Algorithm
$\mathbf{MI_{COV}}$	Proposed MI for CovM Algorithm

Except with different number of elements, the array configuration used is a uniform circular array with 8 elements. The original signal is a QPSK signal and the noise added is a white Gaussian noise with the different SNR values.

For simplicity and arrangement of figures, Table 5.1 represents the notation of the legends that will be used later in all of our simulation figures.

Error rate vs. SNR

The first simulation is done to test AIC, MDL and the proposed algorithms performance with different SNR values. SNR values range from -20 to 15 dB, the number of samples is fixed to 1024 and the actual number of sources is 2.

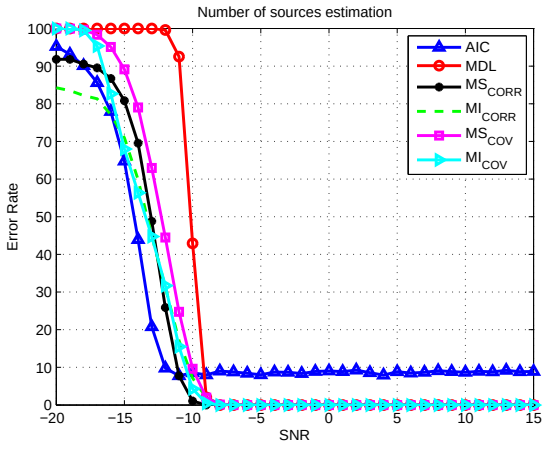


Figure 5.1: Effect Of Different SNR at N=1024 sample

As shown in Fig. 5.1, the proposed algorithms behave better than MDL at low SNR values and better than AIC at high SNR values. For less than -10 dB, the proposed algorithms have a comparable performance to AIC and better than MDL

with $\mathbf{MI_{Corr}}$ having the least error in less than -15 dB. Between -15 and -12 dB, all algorithms have a very comparable performance; however, the estimation error rate for all algorithms is high for less than -12 dB SNR values. This error is due to the inconsistent change in eigenvalues that results from high noise. At SNR values higher than -10 dB, the performance of MDL and the proposed algorithms come to be the same with a minimum error rate that is almost 0% while AIC keeps its error rate of about 10%. The reason why AIC is not giving lower error rate is the overestimation of the number of sources which happens with relatively high SNR values. This overestimation is probably due to AIC added penalty term as was proven by [78].

Another test is done to check the performance with low number of samples where it is set to 100 samples and the same other configurations are applied. As shown in Fig. 5.2, CorrM based algorithms outperform others with low SNR levels that are less than -10 dB. This goes back to the better contrast between the signal and noise eigenvalues that is introduced by exploiting CorrM and hence better number of sources estimation. After -5 dB, the performance of MDL and the proposed algorithms come to be the same with minimum error rate that is almost 0% while AIC kept its error rate of 10% which happened due to overestimation of the number of sources.

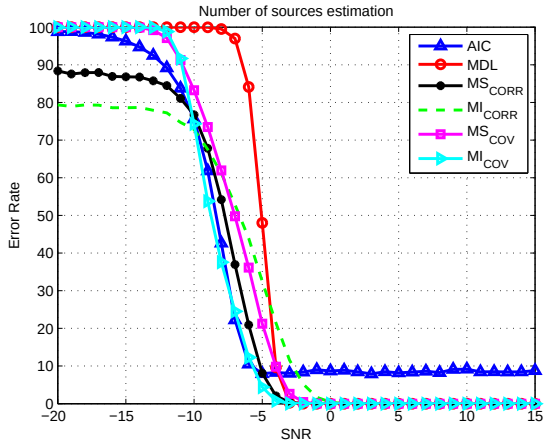


Figure 5.2: Effect Of Different SNR at N=100 sample

Error rate vs. number of samples

One of the important parameters to consider in any algorithm design is the number of samples needed by the algorithm to estimate correctly. This is important for algorithm practical implementation as the number of samples needs to be minimized in such scenarios. Simulation parameters in this test are: SNR value of -5 dB, 2 applied signals and different number of samples.

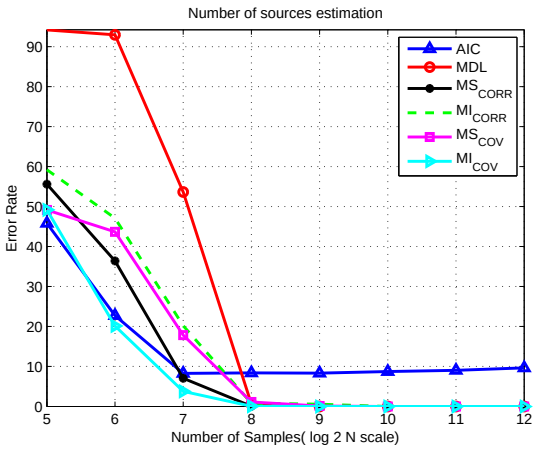


Figure 5.3: Effect Of Different Number of samples at SNR=-5

As can be seen in Fig. 5.3, the proposed algorithms have a better performance than MDL and similar performance to AIC for low number of samples. MDL algorithm underestimates the number of sources with low number of samples as eigenvalues are not well distrusted in a way that can be detected by the algorithm. MDL and the proposed algorithms have the same performance for more than 2^8 samples which is almost 0% error rate. AIC, on the other hand, overestimates the number of samples and hence has its 10% error rate, which is found in almost all test cases that are conducted in this thesis.

Error rate vs. different number of applied sources

Different algorithms might have different sensitivities in terms of the number of sources they can estimate. Hence, in this simulation, algorithms are tested against a

different number of applied sources at SNR value of -5 dB and the number of samples equal to 1024 samples.

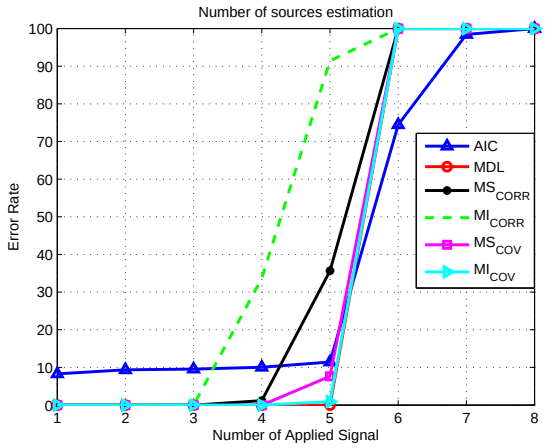


Figure 5.4: Effect Of Different Number of applied Signals at SNR=-5, N =1024 with 8 elements UCA array

As can be seen in Fig. 5.4, AIC outperforms all other algorithms in the maximum number of sources it can estimate. MDL, MS for both matrices and $\mathbf{MI_{Corr}}$ can estimate up to 5 elements with less than 20% error and can not estimate more, while AIC can estimate 6 and 7 signals, but with a very high error rate. By looking at some other results with different angular separation between the applied signal, it is noted that the separation between DoA angles have a great impact on the total number of sources the algorithms can estimate. The higher the separation is, the easier it is for the algorithm to detect more sources.

Algorithm performance with different array elements

We then examine the effect of increasing the number of elements that construct the array at SNR value of -5 dB; 100 samples and 2 sources are applied to the array. The number of samples is chosen to be low so that some algorithm would fail at this stage while others do not. If we choose the number of samples to be high, all algorithms will estimate correctly at this SNR even with 6 antenna elements.

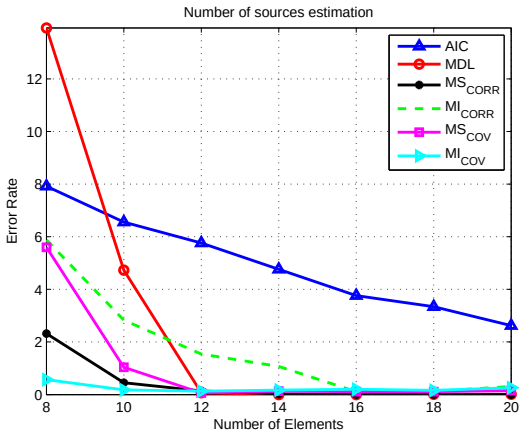


Figure 5.5: Effect Of increasing the number of elements that construct the array at N =100

As shown in Fig. 5.5, $\mathbf{MS_{Corr}}$ and $\mathbf{MI_{Cov}}$ have the best performance among others for the number of elements higher than or equal to 8. As the number of elements increases, the error decreases till it comes to almost 0% error rate for MDL and proposed algorithms, which happens after 12 elements.

Confidence interval with different SNR values

In this simulation, the confidence interval of each algorithm is tested against different SNR values, as SNR is one of the most important metric that we test against. Simulation parameters in this subsection are: number of samples of 2^10, two signals located at 100° and 140° and different SNR values that range from -20 to 15 dB. The y-axis represents the upper and lower bounds of the confidence interval, while the confidence is the range between them. As shown in Fig. 5.6, all algorithms had a very small confidence intervals which indicates that the estimated values are almost the same. The proposed algorithms based on CorrM had a slight variation at very low SNR values; however, they are the closest to the actual value, two sources. Even though it is not so obvious in the figure, AIC algorithm has its estimates between 2 and 3 for high SNR values and that is why it has an error rate. However, it is not shown as the average is closer to 2 and only 10% of the estimates are 3.

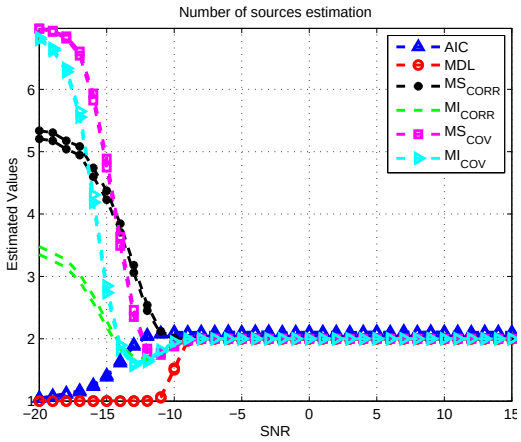


Figure 5.6: Confidence interval with different SNR values

Complexity comparison

As the proposed algorithms are claimed to be simple, their complexity should be much less than AIC and MDL. To prove that, we use the simulation run-time as a metric to compare the complexity of the 6 presented algorithms. In Fig. 5.7, we plot the simulation run-time for all presented algorithms versus the number of collected samples: (a) actual simulation, run-time in seconds and (b) the simulation run-time normalized to the AIC run-time. In Fig. 5.8, we plot the simulation run-time for all proposed algorithms versus the number of antenna elements: (a) actual simulation run-time in seconds and (b) the simulation run-time normalized to the AIC run-time.

As can be seen in Fig. 5.7, our proposed algorithms have a much less simulation run-time, which is the translation of a much less complexity. AIC and MDL have a comparable complexity with MDL achieving a slightly less run-time. At a lower number of samples, MDL is taking 90% of the time AIC takes while it is taking 80% at a higher number of samples. As expected, as the number of samples increases, the simulation run-time for all algorithms increases. Our CorrM algorithms achieve simulation run-time that is less than 2% of that achieved using AIC while our CovM algorithms are achieving simulation run-time that is less than 25% of that achieved using AIC.

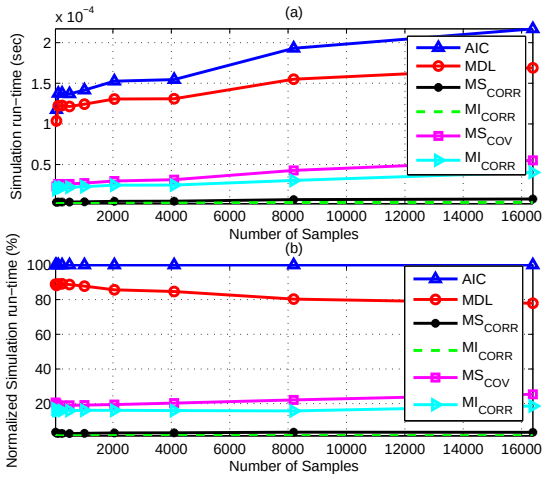


Figure 5.7: Simulation run-time versus number of samples: (a) Actual run-time in seconds (b) Run-time normalized to AIC run-time

Similarly, from Fig. 5.8, our proposed algorithms have drastically improved the simulation run-time while AIC and MDL are having a comparable run-time. Our CorrM algorithms achieve simulation run-time that is less than 5% of that achieved using AIC at low number of antenna elements and less than 2% at higher number of antenna elements, while our CovM algorithms are achieving simulation run-time that is less than 25% of that achieved using AIC at low number of antenna elements and less than 10% at higher number of antenna elements.

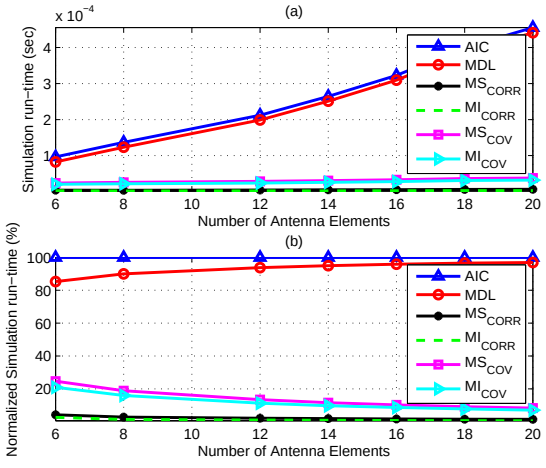


Figure 5.8: Simulation run-time versus number of antenna elements: (a) Actual run-time in seconds (b) Run-time normalized to AIC run-time

5.2.2 Direction of Arrival (DoA) Estimation

Simulation results for DoA estimation were conducted to check the performance of MVDR and MUSIC in different scenarios. This includes testing with different SNR values, different angle separation (resolution), different number of samples, different number of elements that construct the array and different number of applied signals. Unless otherwise stated, an 8 element UCA was used to simulate the results since only 8 elements array will be used for later hardware implementation.

Error rate versus SNR

To test algorithms tolerance to SNR values, SNR values range from -20 to 15 dB, the number of samples is fixed to 1024 and two signals are applied with an angular separation of 40°. Results, as shown in Fig. 5.9, prove that MUSIC outperforms MVDR in terms of error rates and SNR tolerance. MUSIC can estimate correct DoAs at SNR values of -5 dB while MVDR can estimate correct DoAs only after 0 dB. The reason why the error rate increases as SNR decreases is due to high noise effect which is involved in both algorithms calculations and leads to less accurate estimation and hence higher error rate.

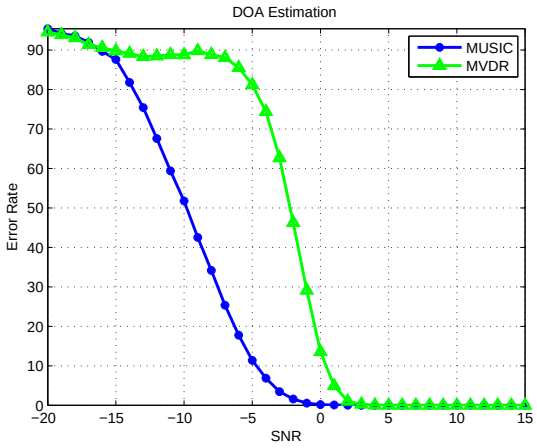


Figure 5.9: Error rate of MVDR and MUSIC with Different SNR values

Error rate versus angular resolution

Another simulation is done to test the error rates of both algorithms with different angular resolution at SNR of 0 dB, number of samples of 1024 and 8 elements UCA with 2 applied signals. Results, as can be seen in Fig. 5.10, show that MUSIC is able to estimate the angles correctly if they are separated by 20° with about 20% error rate. Even though it can detect that there were 2 signals for less than this separation, but it could not estimate the angles correctly and hence the error rate was high. MVDR can detect the two signals at 30° separation; however, it is not accurate estimation until 40° separation. The reason why MUSIC could have better estimation with small separations goes back to EVD effect which separates the noise from signal subspaces and search for the noise orthogonality instead of the full CovM orthogonality in case of MVDR.

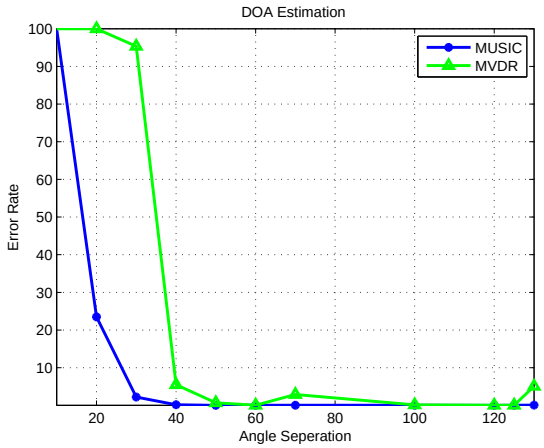


Figure 5.10: Error rate of MVDR and MUSIC with Different angular resolution

Error rate versus different applied signals

A third simulation is done to test algorithms error rates with different number of applied signals that are received by the array. SNR is fixed to -5 dB, the number of samples is 1024 and the angular separations vary, but they are sufficient enough for both algorithms not to fail. As can be seen in Fig. 5.11, MUSIC performs much

better than MVDR when the number of sources increases. Even though the number of elements was 8, MUSIC can estimate 5 signals with an error rate of less than 20%. MVDR, on the other hand, is able to estimate only 2 signals with an error rate of less than 20%. After 2 signals, the error rate starts increasing and the algorithm fails in estimating DoA which can be a main drawback in our case. This simulation shows the most noticeable advantage of MUSIC over MVDR, beside the high resolution, and it is one of the most important reasons to pick MUSIC over MVDR for the breaching scheme.

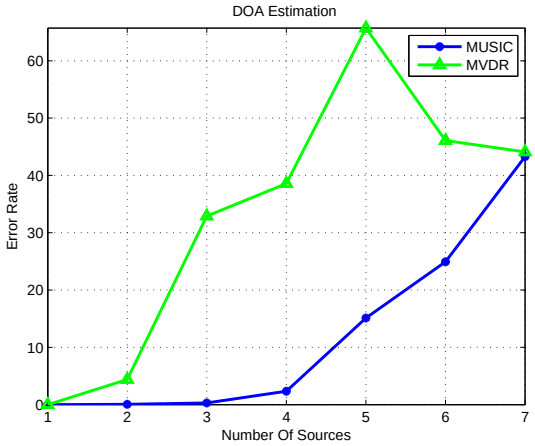
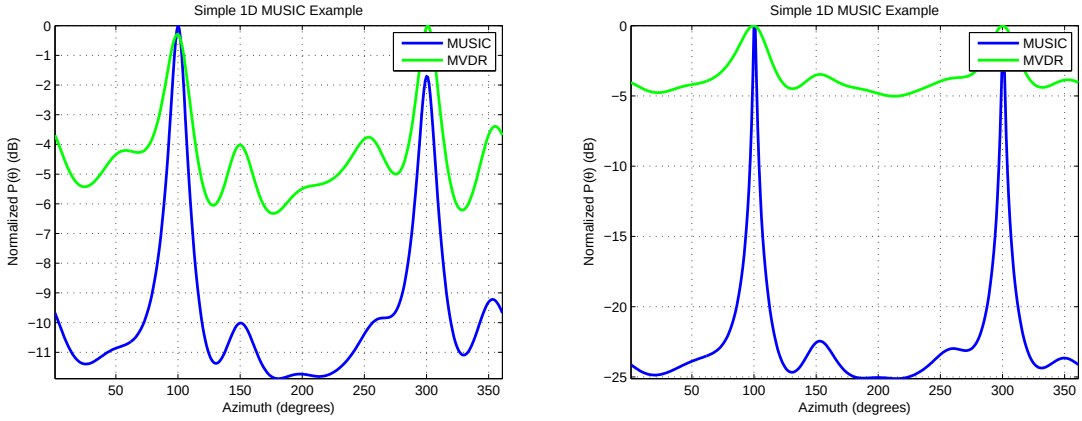


Figure 5.11: Error rate of MVDR and MUSIC with Different number of applied signals

Error rate at various number of samples

Another simulation is done to test the error rates with different number of samples that are used to estimate DoA. As stated before, this is particularly important in any hardware implementation as the number of samples is limited to hardware capabilities which is needed to be minimized as much as possible. Hence, both MUSIC and MVDR are tested against different number of samples while fixing SNR value to -5 and 2 applied signals are generated with 100° separation.

Even though the algorithms can detect the signals at low number of samples, as shown in Fig. 5.12, the accuracy of both algorithms is not good and the estimated DoAs are out of tolerance range, i.e. there was more than 1° error compared to actual



(a) L=64 (b) L=1024

Figure 5.12: Comparison between MVDR and MUSIC with different number of samples

values. Hence, in terms of the error rates, as can be seen in Fig. 5.13, both algorithms have a very high error rate with 2^7 samples or less. After that, both error rates are much enhanced until they reach almost 0 at 2^{10} samples or more. This can prove that even if the algorithms can detect DoAs, they will not able to estimate them accurately and hence they fail at low number of samples due to incorrect estimation of the CovM. However, in our case, that will not be a problem as we assume enough number of samples already, taking about 2^{14} samples in hardware implementation.

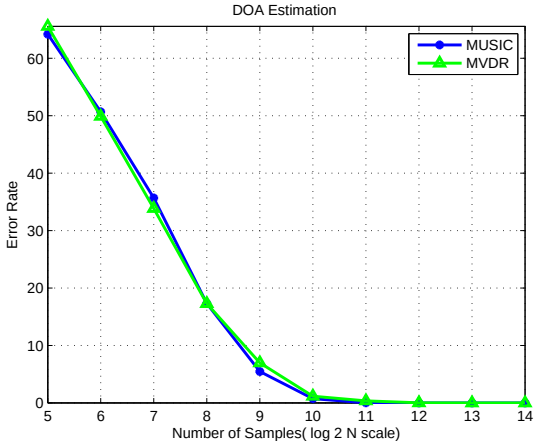


Figure 5.13: Error rate of MVDR and MUSIC with Different number of snapshots

Error rate versus different array elements

A final simulation is done to test algorithms error rate with different number of elements that construct the array. Theoretically, as the number of elements increases, the estimation is enhanced, but with a penalty of complexity increase. Of course, as we are concerned with hardware implementation, a compromise between accuracy and complexity should be done in order to benefit from both factors. To reach this compromise, this simulation is done to test how the performance gets affected when increasing the number of array elements. SNR value was fixed to -5, two signals are applied with 100° separation and the number of samples is 1024.

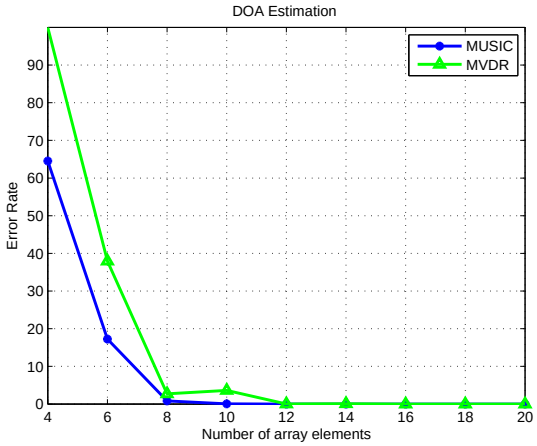


Figure 5.14: Error rate of MVDR and MUSIC with Different number of array elements

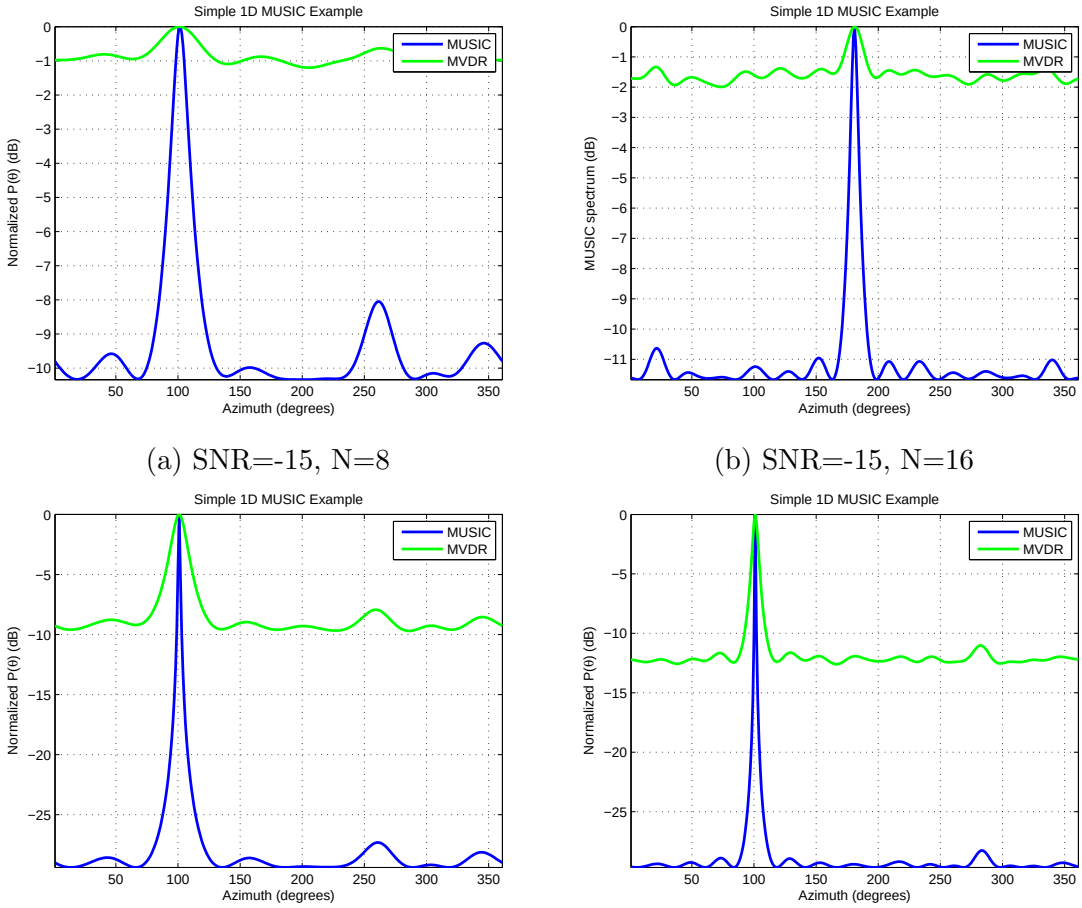
As shown in Fig. 5.14, MVDR cannot estimate with 6 elements, but it has a correct estimation at 8 elements while MUSIC can estimate even with 6 elements. It's worth noting here that increasing the array elements affected the estimation in both algorithms. The resolutions are much enhanced for both algorithms as the number of elements increases; however, from 8 to 16 elements the change is not that worthy, especially when considering the cost added. Hence, using 8 elements would be sufficient enough for good estimation in our application and will be used later in breaching simulation and hardware implementation.

MVDR & MUSIC PNR values

This simulation is done to check the peak to noise ratio (PNR) of both algorithms at different SNR values and different number of elements that construct UCA, where only one signal is applied to the array. Fig. 5.15 shows PNR values of both algorithms with different SNR values and different number of elements. As can be seen, MUSIC outperforms MVDR by achieving a very high PNR, y-axis, even with low SNR values, achieving about 10 dB for SNR of -15 dB while it is only 1 dB in case of MVDR. Besides, as SNR values and number of elements increases, PNR value will be getting higher due to less noise effect. Higher PNR values would mean better estimation as signal peaks will be more distinguishable from noise peaks, even with relatively high noise peaks.

Confidence interval with different SNR values

In this simulation, the confidence interval of MVDR and MUSIC algorithms are tested with different SNR values. Simulation parameters in this subsection are: number of samples of 2^10, two signals located at 100° and 140° and different SNR values that range from -20 to 15 dB. The confidence interval of one angle, 100° is calculated for both algorithms, and the y-axis represents the upper and lower bounds, while the interval is the range between them. As shown in Fig. 5.16, MUSIC has a much better performance than MVDR. The estimation is wrong for both algorithm when SNR values are less than -10 dB. However, after that MUSIC estimation gets to the actual value with minimum confidence interval, while MVDR doesn't get to the actual value until 0 dB. These SNR boundaries are slightly different than the one in the error rate simulation which is due to the search of one angle instead of two in the case of error rate.



(a) SNR=-15, N=8

(b) SNR=-15, N=16

(c) SNR=-5, N=8

(d) SNR=-5, N=16

Figure 5.15: Comparison between MVDR and MUSIC PNR values at different SNR and different number of elements

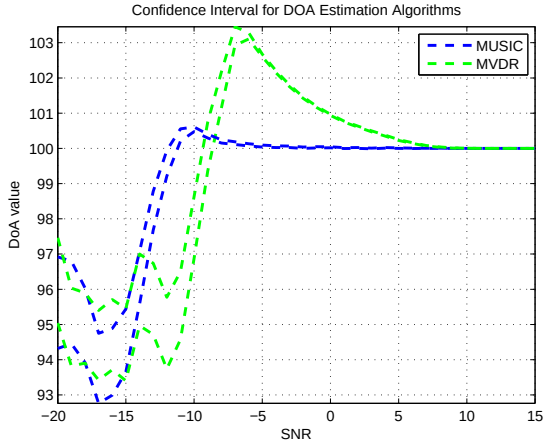


Figure 5.16: DoA confidence interval with different SNR values

5.3 Breaching Scheme Simulation Results

In this section, we test the performance of the estimated Alice-Bob distance in terms of Normalized Mean Square Error (NMSE), normalized distance and bit-mismatch ratio (BMR). NMSE and the normalized distance would show how the estimated distances differ from the actual one while BMR would show the bit mismatch between the generated keys and Alice's key which we consider as a reference key. The normalized distance estimation is given by:

$$Normalized\ Distance = \frac{Average\ Estimated\ Distance}{Actual\ Distance} \times 100 \tag{5.3}$$

while NMSE can be given by:

$$NMSE = \frac{\frac{1}{T}\sum(\rho' - \rho)^2}{\rho'^2} \tag{5.4}$$

where T is the number of trials, ρ' is the actual distance, ρ is the estimated distance. BMR is the number of mismatched bits divided by the total number of bits in the generated keys, it can be represented by:

$$BMR = \frac{Number\ of\ Different\ Bits}{Total\ Number\ of\ Bits} \tag{5.5}$$

Distances will be assumed with some error of less than 20% and 10% for RSSI and ToA distance estimation algorithms, respectively, as stated earlier in Section 4.6. Each of Alice, Bob and Eve will be estimating distances and generating their own keys with Eve using the two number of sources estimation algorithms, MMS_{CORR} and MMI_{COV}. As a result, 8 distances will be estimated and 8 keys will be generated. Distances are compared to the actual one in terms of NMSE and normalized distance estimation while the keys are compared to Alice key, as a reference key, in terms of BMR. Simulation results consider different numbers of samples, different SNR values,

Table 5.2: Figures Legend Abbreviations for Breaching Scheme Algorithms

Abbreviation	Algorithm
Alice_Bob	Estimated distance by Alice
Alice_Bob_MMS$_{\mathbf{Corr}}$	Estimated Distance by Eve Using **MMS$_{\mathbf{Corr}}$** Algorithm
Alice_Bob_MMI$_{\mathbf{COV}}$	Estimated Distance by Eve **MMI$_{\mathbf{COV}}$** Algorithm
Bob_Alice	Estimated distance by Bob

and different angular separation as parameters to change and see their effect on the presented algorithms.

All the presented scenarios in this section will follow Fig. 5.17 for the actual assumed distances between Eve_Bob and Eve_Alice where the key is based on Alice_Bob distance, d_{12}, which depends on the value of ϕ_e. ϕ_e is the difference between two actual DoAs and changes depending on the simulation.

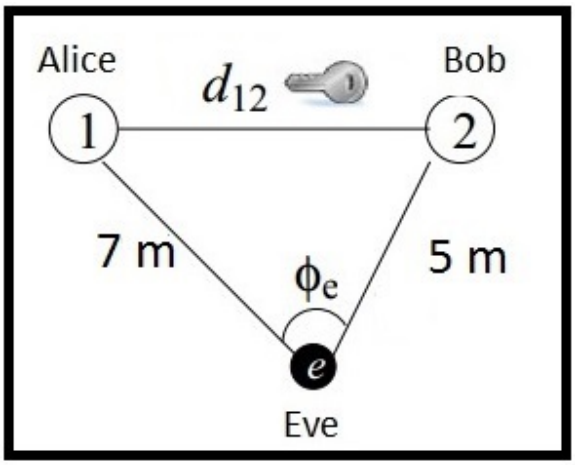


Figure 5.17: Simulation Setup for the Breaching Algorithm

5.3.1 Distance Estimation versus Different Parameters

For simplicity and arrangement of figures in this subsection, table 5.2 represents the notation of the legends that will be used for figures and within discussions.

Number of samples

In this simulation, SNR value is fixed to -5 dB and the angles are set to 100° and 200°, i.e. the angular separation is 100°. The number of samples ranges from 2^5 to 2^{14}

to test the distance estimation with different samples. As can be seen in Fig. 5.18, Eve cannot estimate the correct distance for less than 2^8 samples. The estimated distances are far away from the actual one and this is due to incorrect estimation of DoA and the number of sources. The reason why the average estimated distances are always less than the actual one goes back to some incorrect estimation of the number of sources which put the estimated distance to its default value, 0 meter. As a result, when averaging the distances, those 0 estimated distances will lead to an average distance of less than the actual one. Bob and Alice can have approximate distances, even with a low number of samples as they do not need DoA estimation and they just estimate the distance which is close to the actual one. It can be seen as well that ToA and RSSI-based estimation has a close performance in terms of the estimated distance tolerance as the tolerance was based on DoA not on distance estimation. Fig. 5.19 shows NMSE for the 4 algorithms which indicates that the error gets close to 0 after 2^8 samples. It should be noted here that Eve estimated distances are sometimes overlapping with each others while Alice and Bob distances are overlapped always, which is resulting in two curves instead of four at some points of the figures.

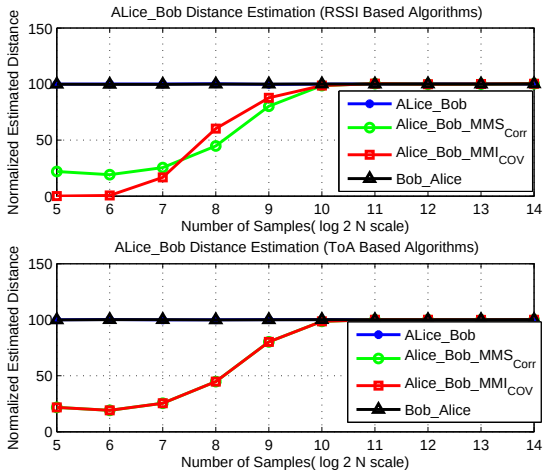


Figure 5.18: Normalized distance estimation with different number of samples at SNR=-7, Angular separation=100

To show NMSE when DoA and the number of sources are estimated correctly, the same simulation is regenerated for number of samples greater than 2^9. Fig 5.20

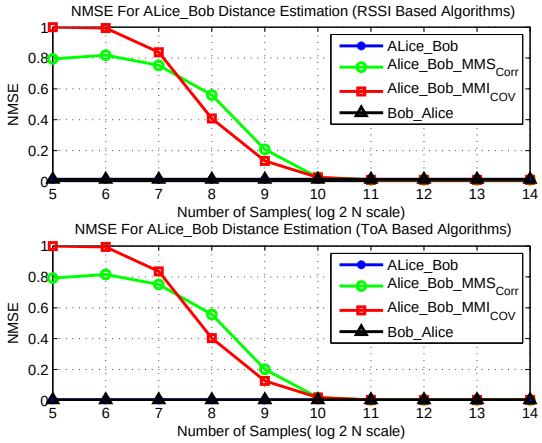


Figure 5.19: NMSE with different number of samples at SNR=-7, Angular separation=100

shows the NMSE for this simulation in order to compare between Alice, Bob and Eve estimated distance in good number of samples conditions. As can be seen, Eve NMSE are relatively close to Alice and Bob NMSE when the number of samples is more than 2^{10}, due to correct estimation of DoA and number of sources algorithms. It can be seen as well that when DoA and the number of sources were estimated correctly, ToA based algorithm has a less NMSE than RSSI-based algorithm, which is due to less estimation error in ToA based algorithms, i.e. 10% instead of 20% in case of RSSI-based algorithms. Also, Eve curves might overlap while Alice and bob curves are overlapping.

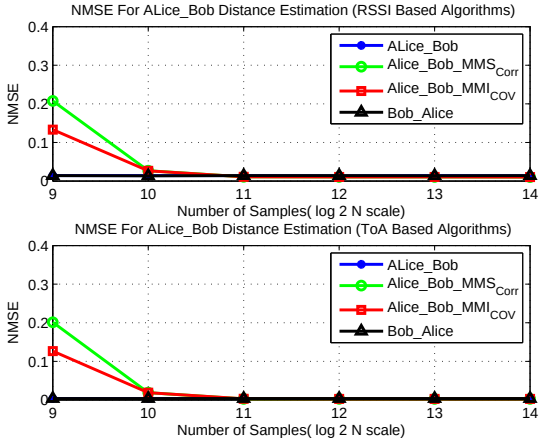


Figure 5.20: NMSE with different good condition number of samples at SNR=-7, Angular separation = 100

SNR

In this simulation, the number of samples is fixed to 2^{14} and the angular separation is fixed to 100°. SNR values range from -20 to 20 dB and the distances are estimated by Alice, Bob and Eve. As shown in Fig. 5.21, Eve can not estimate the distance for less than -15 dB, but it can estimate approximate distances after that. The reason why the estimation fails for less -15 dB is due to failure in DoA estimation which is expected at such low SNR values. After that, all algorithms come to almost the actual distance estimation and have almost 0 NMSE as shown in Fig. 5.22.

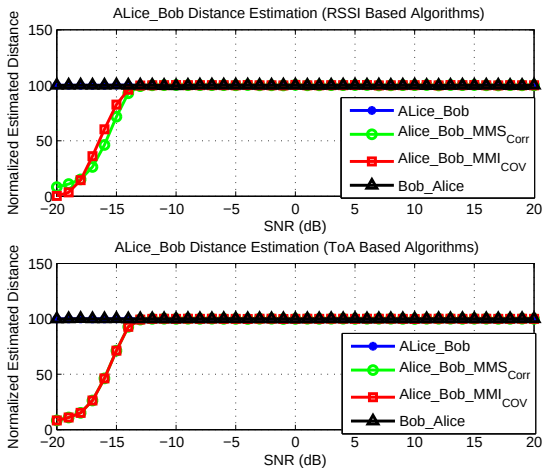


Figure 5.21: Normalized distance with SNR values at N=2^{14}, Angular separation=100

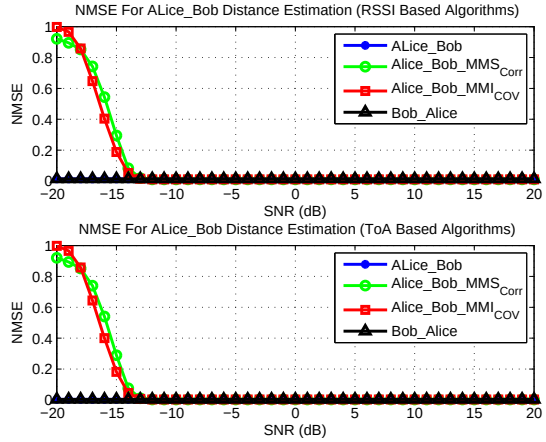


Figure 5.22: NMSE with SNR values at N=2^{14}, Angular separation = 100

To check the performance when DoA and the number of samples are estimated correctly, the same simulation is regenerated with SNR values of -10 to 20 dB. As shown in Fig. 5.23, Alice, Bob, and Eve have a comparable performance in terms of NMSE with Eve having a slightly less value than Bob and Alice do. This is due to the fact that Eve has its error in Alice−Eve and Bob−Eve distances rather than Alice−Bob distance which cause a less error in Alice−Bob distance as some errors cancel each other. It is obvious as well that ToA based algorithms have less NMSE than RSSI-based due to less estimation error in ToA based algorithms. Moreover, Eve curves might overlap while Alice and bob curves are overlapping.

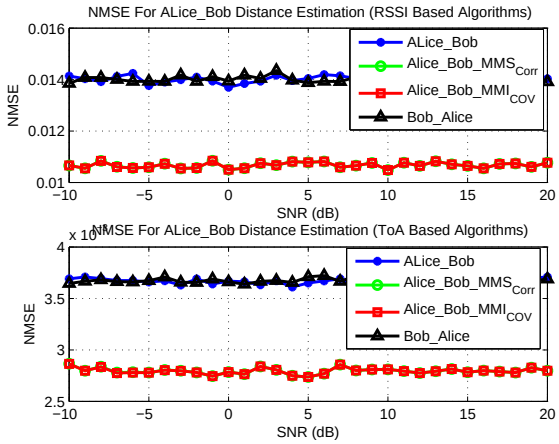


Figure 5.23: NMSE with good condition SNR values at N=2^{14}, Angular separation = 100

Angular separation

In this simulation, SNR value is fixed at 0 dB, the number of samples is fixed to 2^{14}, and the angular separations change from 10 to 140°. As shown in Fig. 5.24, $\mathbf{MMI_{COV}}$ algorithm estimates correctly when the separation is 15° while $\mathbf{MMS_{CORR}}$ algorithm estimates correctly when the separation is 30°. The reason behind that is the resolution of CorrM based algorithm which is less than CovM algorithms and causes the failure of number of sources estimation for less than 30° separation.

Table 5.3: Figures Legend Abbreviations for Breaching Scheme Algorithms

Abbreviation	Algorithm
Alice_Bob_Key	Key generated by Bob
Alice_Bob_MMS$_{CORR}$	Key generated by Eve Using MMS_{CORR} Algorithm
Alice_Bob_MMI$_{COV}$	Key generated by Eve Using MMI_{COV} Algorithm

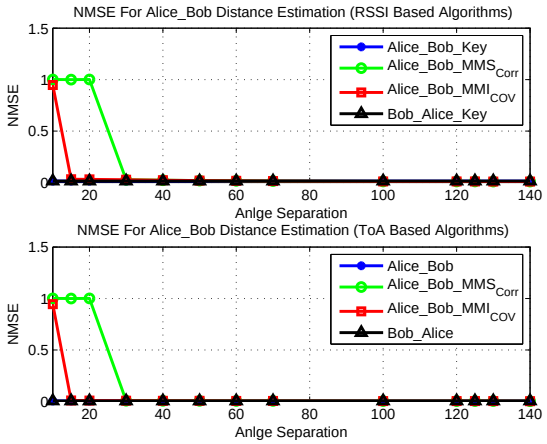


Figure 5.24: NMSE with different angles separation at N=2^{14}, SNR =0 dB

5.3.2 Keys BMR versus Different Parameters

In this subsection, secret keys are generated using the estimated distances by Alice, Bob and Eve with Eve using 2 algorithms. Then, Alice's key is chosen as a reference key and other keys are compared to it in terms of bit mismatch. Simulations are done with different SNR and different number of samples conditions; however, it is not tested with different angular separations as such results are already expected, the key will not work for less than 30°, but it has a comparable result for more than that.

For simplicity and arrangement of figures in this subsection, Table 5.3 represents the notation of the legends that will be used for figures in this section.

Number of Samples

In the first simulation, SNR value is fixed to -5 dB, the angular separation is 100° and the number of samples ranges from 2^5 to 2^{14}. Results, as shown in Fig. 5.25, prove that Eve will have a higher BMR than Bob at low number of samples. Hence,

even if Eve listens to information reconciliation and privacy amplification, she will not be able to estimate the key at that stage. The reason for such high BMR at that stage is the incorrect estimations of the number of sources or DoA which happen at relatively low number of samples and result in wrong or far distance estimation.

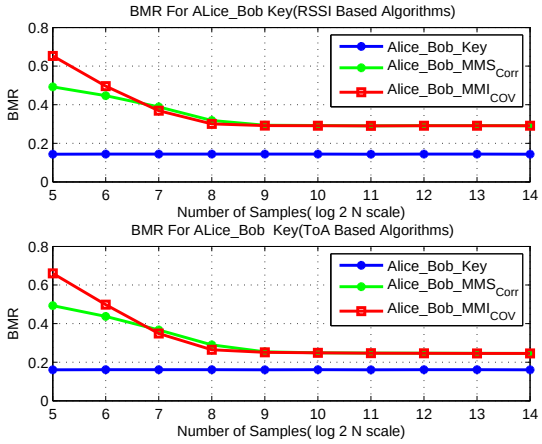


Figure 5.25: Key BMR with different number of samples at SNR=-5, Angular separation = 100

After that, Eve estimates DoA approximately and could have a lower BMR; however, it is still higher than Bob BMR, especially for RSSI-based distance estimation scenario. As shown in Fig. 5.26, Bob BMR is less 20% which indicates that there is a small mismatch between Alice and Bob generated keys. Eve, on the other hand, has about 25% mismatch in case of ToA based scenario and 30% mismatch in case of RSSI-based scenario. With further information reconciliation and privacy amplification stages Bob will be able to fix his mismatch as they are less than 20-25%. If Eve listens to those stages then she might be able to fix the mismatch in case of ToA based algorithm; however, it is still difficult to fix the mismatch in case of RSSI-based algorithm.

SNR

In the second simulation, the number of samples is fixed to 2^{14} and SNR values range from -10 to 20 dB while the angular separation is 100. As shown in Fig. 5.27, Eve BMR is higher than Bob one's even when DoA and the number of samples are

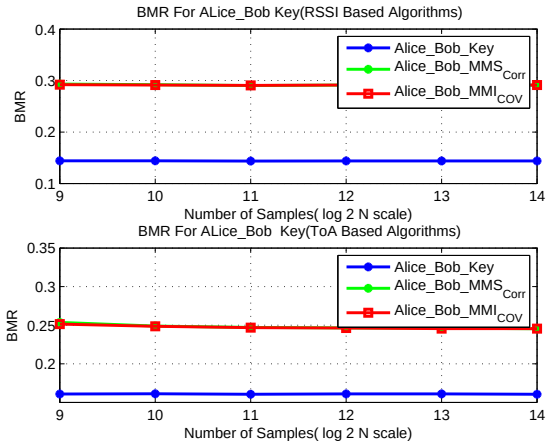


Figure 5.26: Key BMR with different number of samples at SNR=-7, Angular separation = 100

estimated correctly. The reason behind that is the error added by ToA or RSSI-based algorithm which is random and uncorrelated from Bob and Alice errors. This resulted in different quantization bits which will be resulting in high BMR, especially in the case of RSSI-based algorithm. As with different number of samples, Eve might be able to breach the system with ToA based algorithms; however, it is still difficult to breach in case of RSSI-based algorithm.

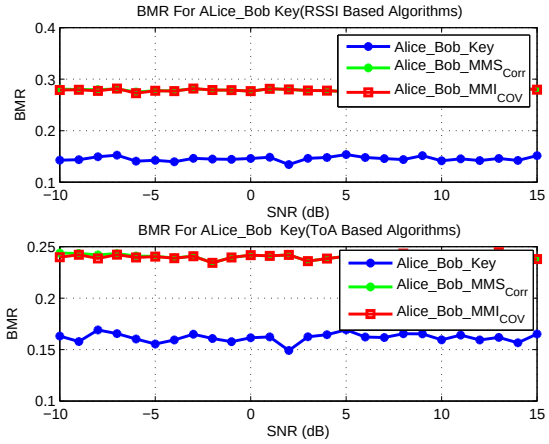


Figure 5.27: Key BMR with SNR values at N=2^{14}, Angular separation = 100

Chapter 6

Discussion

In this chapter, we start by discussing some of the breaching scheme weaknesses in order for further enhancement in the future. Then we discuss an extension of the project which already started, namely hardware implementation. Later on, we discuss some challenges that were faced through the hardware implementation and the number of sources estimation. Finally, we discuss some future directions that can be done for the project in short and long terms.

6.1 Breaching Scheme Weaknesses

In this section, weaknesses of the breaching scheme will be discussed in brief. This aims to introduce some ways to enhance the security and the breaching schemes which will be considered as a future direction of this thesis. The main weak points that will be discussed here are the assumption of two sources in the breaching scheme, complexity of the algorithm and the usage of empirical thresholds with the number of sources CovM based algorithms.

Complexity problem with the breaching scheme comes from its eigenvalues decomposition which is needed for both the number of sources and DoA estimation. This process is traditionally complex and requires M^3 iterations, hence, its complexity is approximated by $O(M^3)$. Besides, the complexity of covariance matrix finding is about $O(M^2N)$ and the complexity of the spectral search in DoA estimation is $O(M^2A)$ when N is the number of samples, M is the number of array elements and A is the angles in which DoAs are search for, i.e. 360^circ in case of circular array [79]

[80]. The number of sources estimation techniques are relatively simple, as shown in simulation results, and their complexity with distance estimation complexity can be negligible in comparison to DoA complexity. Hence, the big problem is with DoA estimation complexity as this is the most complex process in our work. There exist many attempts in the literature to solve MUSIC complexity; an example can be found in [81] [82]; however, we consider solving the complexity problem as a future direction and we just use the traditional MUSIC algorithm in our context.

The second weakness lies in the threshold definition of CovM-based number of sources estimation algorithms. This threshold is found empirically for both MS and MI decision statistics, even though we show a closure form the distribution of threshold. However, there is a possibility that when implementing those algorithms in hardware, this threshold will need to be redefined from the experimental data rather than empirical simulation data. At that stage, linear regression can be used as well to define the threshold equation for both decision statistics.

A final weak point to be considered is the way we handle more than two sources applied to the array. As stated earlier, receiving more than two sources is a valid case especially in wireless channels. We handle this case by choosing the two highest peaks to noise ratio DoAs and consider them as our target nodes, or communicating nodes. Of course, that might not a practical assumption; however, this can be a valid assumption with a simple simulation and hardware implementation as we have the control of the signal powers. A further discussion on this point will be considered later in the future direction section.

6.2 Hardware Implementation as an Extension

Implementing the breaching scheme on actual hardware is an important part of the project that this thesis was part of. This can prove the feasibility of both the security scheme and the breaching scheme, and it suggests an initial model for hardware im-

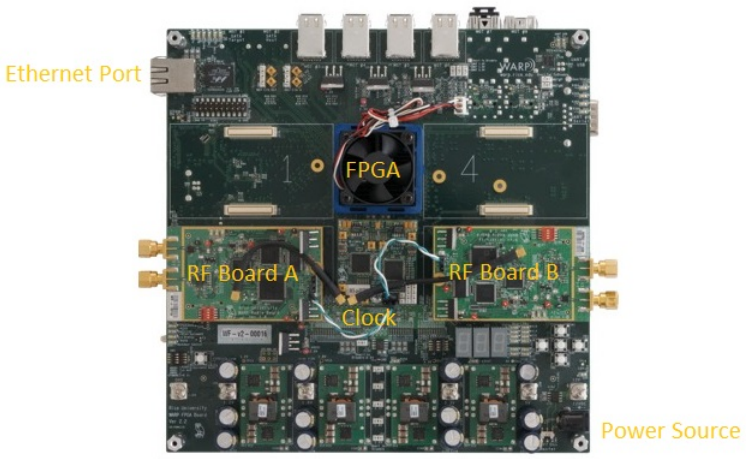


Figure 6.1: WARP Hardware

plementation of physical layer security schemes. Traditionally, physical layer security was mostly considered with software implementation and information theoretic approaches which are not feasible in actual implementations; however, the current trend in this field is for hardware implementation of newly raised security schemes. Channel information, RSSI or the distance is available in hardware modules now and hence estimating such values and building the security based on them can be considered as an easy task compared to traditional cryptographic techniques.

To that extent, some work already started in employing our scheme to hardware using Wireless open Access Research Platform (WARP) hardware. In this subsection, we introduce the hardware first, then the configuration that will be used in implementation, some initial results, the problem introduced and finally the planned solutions and work to be done in that part.

Wireless Open Access Research Platform (WARP) hardware

WARP is a scalable, extendable and programmable wireless platform built by Rice University to be used by academics and industrial projects in MAC and PHY layer protocol implementation. As in Fig. 6.1, the hardware consists of 4 RF daughter boards, where only A and B are placed while C and D can be placed in 1 and 4

slot respectively, Xilinx Virtex-II Pro FPGA, sampling and RF clocks and Ethernet port to be connected to a switch. RF boards operate at 2.4 or 5 GHz to provide the waveform access to the transceivers without fallowing any wireless standards, i.e. any standard can be used, and requiring the waveform to be handled at the FPGA level [83].

WARPLAB framework allows many physical layer algorithms to be constructed and tested on MATLAB offline bases. It combines MATLAB and FPGA implementation to allow the customization and flexibility of the implemented algorithms. The FPGA is programmed externally to allow programming of RF board connections and clock synchronization while the reference design uses MATLAB to control the RF board and perform signal processing which is the core of this work.

Hardware configuration

In order to implement DoA algorithm, we use a circular array with 8 elements and 0.06 meter radius. Hence, a circular array of 8 elements was designed externally and we connected it to our 8 WARP RF boards coming from 2 WARP kits to the array elements. As shown in Fig 6.2, the array elements are connected to the WARP hardware through RF cables. Synchronization between WARP nodes was done by connecting clocks through MMCX-Cables and connecting the triggers of both nodes to one reference node. Fig. 6.3 shows the connection between the WARP nodes clocks and triggers where node 1 is the reference node and node 2 is the synchronized node.

After that, the FPGA is programmed to its default state where 4 RF boards will be working and one node will be receiving its clock from the other, for clock synchronization. The 4 RF boards are set to receive mode and a signal is sent from a signal generator with -20 dB power. The signal is received by the 8 antennas, calibrated to match their phase shifts and attenuation and finally saved as a matrix of 8×2^{14} where 2^{14} is the number of samples. Then an offline process starts by applying our estimation algorithms and estimates the communicating nodes distance

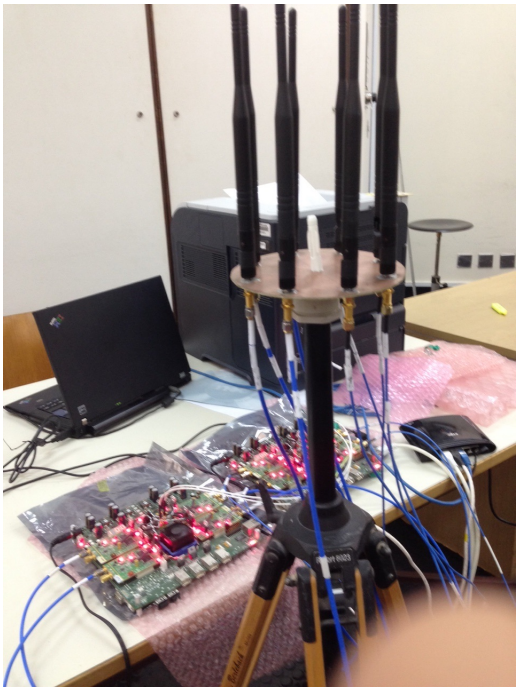


Figure 6.2: UCA connection to RF boards

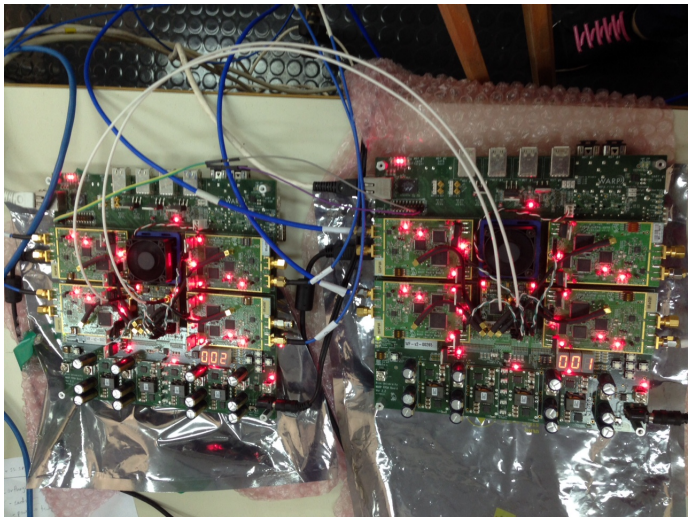


Figure 6.3: RF board connections for synchronization

by: estimating the number of sources, DoA, distances from the eavesdropper. Finally, we estimate the key based on the estimated distance and compare it to Alice and Bob keys.

Table 6.1: Mean , Variance and STD of the Hardware Estimated DoA

Expected DoA Range	**Estimated One**		
	Mean	**Variance**	**STD**
65-75	70	0	0
115-125	118.4300	0.2678	0.5175
190-200	190.7143	0.2420	0.4919
250-260	256.8400	0.1358	0.3685
340-350	346.7500	0.1894	0.4352

Initial results

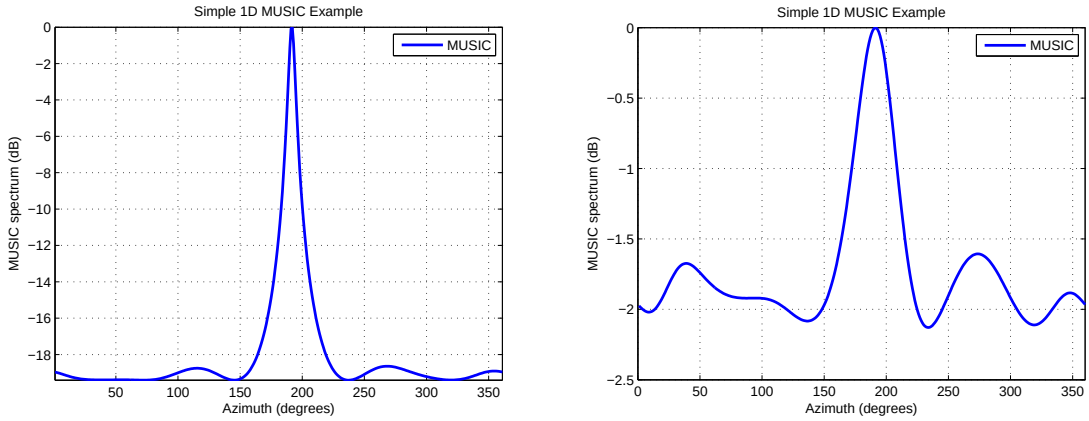
DoA algorithm was the first stage in hardware implementation as it is the hardest and it is the core of the breaching scheme. A matrix of 8×2^{14} is received and needed to be processed in order to estimate the direction from which the signal is coming. One signal with -20 dB power is applied and the process follows Algorithm 2 where the received signal is already saved in the matrix so that we need step 1 and 2 of the algorithm only. The circular array is fixed while the source is moving for 5 different runs at locations. The actual values of the directions are not known exactly, but they are approximated to 70, 120, 190, 250 and 330 which is close to the estimated ones. At each location, DoA is estimated and saved where the variation in the estimated angles is small and can reach a maximum of 2° only. Table 6.1 shows the range of the actual DoA, mean, variance and standard deviation of the hardware estimated ones. It should be noted however those results were done for 100 runs only; and hence, the results might not be that accurate.

Problem raised and proposed solutions

As shown in initial results, hardware implementation of MUSIC gave correct estimated DoA up to some limits; however, many problems were introduced at that stage and it took a lot of time to be tracked and solved. Some of these problems will be mentioned later in the next section; however, even after all the calibration and setup phases, there are two major problems which could not be solved till now. These

problems are basically the peak to noise ratio (PNR) values and the change in phase offsets among nodes, which is changing the reference antenna for each run.

The first problem in which the PNR values are wrong is a more serious one in our context and would fail the breaching scheme as the algorithm would not estimate two sources. This problem can be visualized in Fig. 6.4 where the PNR value of the hardware is compared to the simulation one, at low SNR condition. As can be seen, PNR in the case of hardware is much less than the one for simulation, even though the SNR value was -10 dB in simulation case. Obviously, this PNR is wrong for MUSIC algorithm as this algorithm should have much higher values even with low SNR conditions, which could conclude that the problem is not with low SNR. Employing two sources based on these results does not work as the signals will be lost with all this added noise, i.e. the signal will not be detected.



(a) Simulation (b) Hardware

Figure 6.4: DoA spectrum power of 1 signal at 190°

It is possible that the problem is caused by the environment in our labs, which might introduce noise into the environment in a way that failed DoA implementation especially for MUSIC algorithm. This high noise could be non orthogonal to the signal subspace which is resulted from wireless interferences and reflections and would fail MUSIC algorithm as it is based on the orthogonality fact. The second problem is the change in phase offsets among RF nodes, which is a feature in the hardware and

it is resulting in the reference node being changed each time the code is run. As a result, DoA estimation is changing each time that the code is run. This is due to the RF reference, which is the location of 0°, changes each time. This problem has not been tackled as the actual value of the DoA wouldn't effect our scheme as soon as the difference between DoAs are the same, which conceptually should happen.

6.3 Challenges

In this section, some of the challenges that were faced during the project are discussed as well as the ways that were used to overcome these challenges.

6.3.1 Number of Sources Estimation

In number of sources estimation, the big challenge was to set a threshold for CovM based algorithms. Initially, it was clear that a threshold can be set to differentiate between noise and signal eigenvalues; however, that threshold changes with different parameters and hence a mathematical way was needed to find the threshold distribution using eigenvalues distribution. To achieve that, we started with the distribution of the eigenvalues given in Eq. (4.24) and from that we found the distribution of ordered eigenvalues. Then, we found the distribution of the moving increments, which is the difference between consecutive eigenvalues, given in Eq. (C.7). This turned out to be too complex to solve so we just stated the distribution of the threshold in Eq. (4.25) without solving for its integration and showed the detailed proof in appendix C to prove its mathematical tediousness. Then, we moved to the linear regression analysis in order to find the threshold which would count as the easiest way to deal with such cases empirically and experimentally if the mathematical derivations were not available. When applying this work to hardware, there might be a need to reset the threshold equation based on the environment; however, we expect the change to

be minimal and handle different SNR values and the number of samples conditions, not as what was achieved in previous work.

6.3.2 Hardware Implementation

When proposing this thesis, we thought of hardware implementation as a stage in it; however, this was not achieved due to a group of problems that were faced while implementing the project of hardware. First, the available WARP kits at Qatar University were not enough for our implementation purposes and hence it was needed to be ordered from outside Qatar and that needed a lot of time and cost to be finished. Later on, another purchase was needed for some extra hardwares to connect between nodes, and those hardwares were not taken into account at the beginning.

After that, the hardware needed to be configured carefully to set the RF boards and test their connectivity which took a significant amount of time due to the lack of experience in hardware work and the introduction of new problems that were taken into consideration. In general, dealing with WARP hardware was a challenging task due to its sensitivity, unavailability of resources and lack of use at Qatar university. The sensitivity comes from the fact that there are too many sensors on the board and can be easily damaged; however, extensive force is needed in order to remove or install any RF board or any wire. Unavailability of resources come from the fact that the hardware does not come with manuals or help and they depend totally on online resources and forums. Finally, the lack of use comes from the fact that only a few people have used the hardware before and the problems that were faced in this project were not faced by others.

A third problem that was faced in hardware was the calibration phase, which was needed to calibrate the antenna's phase shift and attenuation and normalize them to one reference point. This problem was raised with hardware implementation as the received signals can come at different times and attenuated with different factors

due to distances and channel effects. Hence, the calibration process was important in order to synchronize the received signals with the same phase shift and attenuation. Using the network analyzer can be a way to do the calibration; however, it was just arriving at the university and needs a calibration phase of the device itself and hence this process took a long time to be completed especially with the cables that were used in the project. By the end, the calibration phase could not be guaranteed 100% due to continuous changes in wires and environmental factors which could affect the phase and attenuation.

6.4 Future Directions

In this section, some of the future directions of this thesis will be discussed in brief. As this thesis is part of a bigger project, some of those directions have already started while some others will be considered in the near future. Of course, much more directions can be considered if we consider physical layer security techniques; however, this section will pick up the most important ones.

6.4.1 Key Generation Scheme

One of the key future directions of this thesis is to continue on the complete key regeneration scheme that involves information reconciliation and privacy amplification and the effect on Eve's estimated key on the system. In information reconciliation, Alice permutes its bits and divides them into blocks to be sent to Bob. When Bob receives these blocks, he checks his permutations with the received ones and corrects his generated bits accordingly. In privacy amplification, Alice and Bob agree on the number of hash functions, probably one of them sends it to the other, and they apply those hash functions on the bits which will, theoretically, result in identical bits. From Eve's point of view, she needs to know the arrangement of blocks sent, i.e. how many bits per block and if there are consequence bits. Also, she needs to know the

hash functions that were used in order to apply them to its generated bits. In other words, Eve needs to be as capable as Bob in order to end up with identical bits. This might be done in practice if Eve is powerful enough and listens to Bob−Alice communication when they exchange such information. For time limitations, this was not considered in the context of this thesis; however, it can be considered as a near future direction in the context of the project that the thesis is part of.

6.4.2 Hardware Implementation

As stated earlier, hardware implementation of the scheme is an essential part of the project, which already started at a later stage of the thesis. A near future direction is to solve the existing problems that were discussed before and continue with the implementation of the rest of algorithms to finalize a hardware testbed for the security scheme, its breaching scheme and the enhanced security scheme.

6.4.3 Enhancing Security Scheme

As stated earlier, the aim of this thesis is to check the feasibility of breaching the security system by an eavesdropper and hence help in enhancing the security scheme. As the current results showed that Eve will have about 25% error in case of ToA based algorithm and about 30% error in case of RSSI based, Eve might be able to breach the system in some cases if she is capable enough. Hence, privacy amplification and information reconciliation have to be either hidden from Eve or done in a way that Eve cant understand.

Distance estimation results showed that the distances can be estimated approximately at relatively low SNR, and hence the system can be breached if the eavesdropper can listen to further communication on the key agreement. Then, security scheme should be enhanced and done by:

- **Forcing low SNR condition for eavesdroppers:** The communicating node can force low SNR values for eavesdroppers by employing a null signal to the directions other than the intended receiver direction. This can be done using directed antennas and special signal processing techniques. By doing that, the eavesdropper will not be able to estimate the number of sources or DoA and hence the breaching scheme would fail to estimate the correct distance.

- **Using Another Source of Randomness:** Security can be enhanced by adding another source of randomness that is used to generate the key. In [60], the authors exploit the use of both channel measurements and distance to generate the key which would fail this breaching scheme with minimal changes in the key generation scheme. There exists some ways for the eavesdropper to breach such a system by combining the channel measurements estimates and distance estimates; however, it will need more computations and hence the eavesdropper needs to be more powerful.

6.4.4 Considering More Than Two Sources

As wireless channels are open and can be accessed by anyone, the possibility of receiving more than two sources at the same time is high and should be considered when implementing such schemes. As shown in the results, the number of sources and DoA can work fine with more than two sources, however the problem comes when distinguishing communicating nodes from other external nodes, i.e. when estimating the communicating nodes distance which is the last stage. In order to solve this, one needs a cross-layer design to access the MAC address in the MAC layer frame or any address that can distinguish between nodes of interest and other nodes. This direction will not, mostly, be considered in the near future as it involves some work on the MAC layer which is not easy with WARP hardware; however, it can be considered in simulations of some of those future directions.

6.4.5 Breaching Scheme for Channel Based Security Scheme

Channel measurement security schemes are widely used in physical layer security rather than distance based. These schemes are based on channel gain and amplitude which are available and can be considered as a source of randomness as well. A further work can be done to breach the channel-based security scheme as it is more common and by that we can breach distance-based, channel-based and their combination. In order to breach the channel based scheme, the eavesdropper needs to be in the line of sights between the communicating nodes, and she/he utilizes the signal it receives to estimate the channel measurements, which will have close values to Bob's and Alice' channel measurements. Of course, being in the line of sight and forcing the communicating nodes to communicate on this path is the most difficult part, but there exists some ways to do so. It should be noted that in order to breach the security scheme that is based on both distance and channel, as in [60], the eavesdropper will need to employ both distance and channel estimation techniques and hence one can imagine the added complexity by adding only one other source of randomness.

Chapter 7

Conclusion

This work presents a breaching scheme that tackles distance based physical layer security scheme. In such a security scheme, the communicating nodes use their distances as a source of randomness in mobile networks and they base their key on it. The breaching scheme tries to estimate that distance by multiple algorithms that include number of sources estimation, DoA estimation, distance estimation and the use of the cosine law of a triangle to find the required distance.

We propose four different algorithms for the number of sources estimation that utilize both CovM and CorrM to get the eigenvalues and find the number of sources based on simple decision statistics. CovM- based algorithms use a threshold, which is derived using least squares linear regression fitting, while CorrM-based algorithms depended on a simple search for the maximum value of the decision statistics. We compared our proposed algorithms to traditional information theoretic ones in terms of efficiency and complexity, and the results show that our algorithms are more robust and much simpler than others. In addition, we presented a comparison of two DoA estimation techniques, namely MUSIC and MVDR, and results show that MUSIC outperformed MVDR in many ways. We showed how to estimate the distance in a hardware environment; however, in our simulation, we just assume some values for the exact distance and add a random error to them.

We evaluate our breaching scheme in terms of estimated normalized distance, NMSE for the estimated distance and a simple BMR calculation for the estimated key which needs to be extended for information reconciliation and privacy amplifications stages. Our results show that Eve might be able to breach the system as soon as

she has a good number of sources and DoA estimations. She will have a well estimated distance at relatively low SNR of about -10 dB if we consider 2^{14} samples and good angular separation. Eve's BMR accedes Bob's one especially with RSSI-based distance estimation algorithms. Hence, Eve might be able to breach the system in case of ToA-based algorithm; however, it is still difficult to breach RSSI-based algorithms due to high BMR results.

Future directions of this thesis include further key comparison results, hardware implementation, improving the security scheme and key generation scheme and finally considering breaching schemes for channel measurement security schemes.

Bibliography

[1] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[2] A. Liavas, P. Regalia, and J.-P. Delmas, "Blind channel approximation: effective channel order determination," *Signal Processing, IEEE Transactions on*, vol. 47, no. 12, pp. 3336–3344, Dec 1999.

[3] R. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*. Stanford University, 1981. [Online]. Available: http://books.google.com.qa/books?id=mLKUnQEACAAJ

[4] J.-S. Jiang and M.-A. Ingram, "Robust detection of number of sources using the transformed rotational matrix," in *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, vol. 1, March 2004, pp. 501–506 Vol.1.

[5] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 33, no. 2, pp. 387–392, Apr 1985.

[6] O. Hu, F. Zheng, and M. Faulkner, "Detecting the number of signals using antenna array: a single threshold solution," in *Signal Processing and Its Applications, 1999. ISSPA '99. Proceedings of the Fifth International Symposium on*, vol. 2, 1999, pp. 905–908 vol.2.

[7] J.-S. Jiang and M. A. Ingram, "Path models and mimo capacity for measured indoor channels at 5.8 ghz," *ANTEM*, pp. 603–609, 2002.

[8] A. Di and L. Tian, "Matrix decomposition and multiple source location," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, vol. 9, Mar 1984, pp. 722–725.

[9] K. Chen, J. Xin, H. Tao, N. Zheng, and A. Sano, "Detection of the number of non-coherent and coherent signals with a simple planar array," in *Signal Processing Advances in Wireless Communications (SPAWC), 2014 IEEE 15th International Workshop on*, June 2014, pp. 204–208.

[10] W. Chen, K. M. Wong, and J. Reilly, "Detection of the number of signals: a predicted eigen-threshold approach," *Signal Processing, IEEE Transactions on*, vol. 39, no. 5, pp. 1088–1098, May 1991.

[11] O. Hu, F. Zheng, and M. Faulkner, "Detecting the number of signals using antenna array: a single threshold solution," in *Signal Processing and Its Applications, 1999. ISSPA '99. Proceedings of the Fifth International Symposium on*, vol. 2, 1999, pp. 905–908 vol.2.

[12] H. Xiong, "Antenna array geometries and algorithms for direction of arrival estimation," Ph.D. dissertation, March 2013. [Online]. Available: http://eprints.nottingham.ac.uk/13016/

[13] T. Tuncer and B. Friedlander, *Classical and Modern Direction-of-Arrival Estimation*. Elsevier Science, 2009. [Online]. Available: https://books.google.com.qa/books?id=1aQbxKJI2CsC

[14] C. Balanis, *Antenna theory: analysis and design*, ser. Harper & Row series in electrical engineering. Wiley, 1982. [Online]. Available: http://books.google.com.qa/books?id=wARTAAAAMAAJ

[15] M. S. Bartlett, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, no. 1/2, pp. pp. 1–16, 1950. [Online]. Available: http://www.jstor.org/stable/2332141

[16] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug 1969.

[17] R. Schmidt, "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions on*, vol. 34, no. 3, pp. 276–280, Mar 1986.

[18] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984–995, Jul 1989.

[19] C. Chen and K. Feng, "Enhanced distance and location estimation for broadband wireless networks," *Mobile Computing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[20] J. Jung and C. Bae, "M2m distance estimation in indoor wireless network," in *Advanced Communication Technology (ICACT), 2013 15th International Conference on*, Jan 2013, pp. 287–290.

[21] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice*. Springer Vienna, 2012. [Online]. Available: https://books.google.com.qa/books?id=tLavBQAAQBAJ

[22] S. Schwarzer, M. Vossiek, M. Pichler, and A. Stelzer, "Precise distance measurement with ieee 802.15.4 (zigbee) devices," in *Radio and Wireless Symposium, 2008 IEEE*, Jan 2008, pp. 779–782.

[23] A. Savvides, C.-C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '01. New York, NY, USA: ACM, 2001, pp. 166–179. [Online]. Available: http://doi.acm.org/10.1145/381677.381693

[24] N. Patwari and A. O. Hero, III, "Using proximity and quantized rss for sensor localization in wireless networks," in *Proceedings of the 2Nd ACM International Conference on Wireless Sensor Networks and Applications*, ser. WSNA '03. New York, NY, USA: ACM, 2003, pp. 20–29. [Online]. Available: http://doi.acm.org/10.1145/941350.941354

[25] K. Whitehouse and D. Culler, "Calibration as parameter estimation in sensor networks," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, ser. WSNA '02. New York, NY, USA: ACM, 2002, pp. 59–67. [Online]. Available: http://doi.acm.org/10.1145/570738.570747

[26] Y.-S. Shiu, S. Y. Chang, H.-C. Wu, S.-H. Huang, and H.-H. Chen, "Physical layer security in wireless networks: a tutorial," *Wireless Communications, IEEE*, vol. 18, no. 2, pp. 66–74, April 2011.

[27] W. Stallings, *Cryptography and Network Security: Principles and Practice*, 3rd ed. Pearson Education, 2002.

[28] C. Murthy and B. Manoj, *Ad Hoc wireless networks: architectures and protocols*, ser. Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR, 2004. [Online]. Available: http://books.google.com.qa/books?id=fvVSAAAAMAAJ

[29] K. Bicakci and B. Tavli, "Denial-of-service attacks and countermeasures in ieee 802.11 wireless networks," *Comput. Stand. Interfaces*, vol. 31, no. 5, pp. 931–941, Sep. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.csi.2008.09.038

[30] M. Bloch, J. Barros, M. Rodrigues, and S. McLaughlin, "Wireless information-theoretic security," *Information Theory, IEEE Transactions on*, vol. 54, no. 6, pp. 2515–2534, June 2008.

[31] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas i: The misome wiretap channel," *Information Theory, IEEE Transactions on*, vol. 56, no. 7, pp. 3088–3104, July 2010.

[32] E. Tekin and A. Yener, "The gaussian multiple access wire-tap channel," *Information Theory, IEEE Transactions on*, vol. 54, no. 12, pp. 5747–5755, Dec 2008.

[33] J. Massey, "An introduction to contemporary cryptology," *Proceedings of the IEEE*, vol. 76, no. 5, pp. 533–549, May 1988.

[34] A. Mukherjee, S. Fakoorian, J. Huang, and A. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *Communications Surveys Tutorials, IEEE*, vol. 16, no. 3, pp. 1550–1573, Third 2014.

[35] H. Liu, J. Yang, Y. Wang, and Y. Chen, "Collaborative secret key extraction leveraging received signal strength in mobile wireless networks," in *INFOCOM, 2012 Proceedings IEEE*, March 2012, pp. 927–935.

[36] R. Liu and W. Trappe, *Securing wireless communications at the physical layer*. Springer, 2010.

[37] J. Wang, J. Chen, and D. Cabric, "Cramer-rao bounds for joint rss/doa-based primary-user localization in cognitive radio networks," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 3, pp. 1363–1375, March 2013.

[38] S. Tiiro, K. Umebayashi, and Y. Suzuki, "Cooperative source number estimation for cognitive radio networks," in *Information Networking (ICOIN), 2014 International Conference on*, Feb 2014, pp. 401–405.

[39] K. Umebayashi, J. Lehtomaki, and Y. Suzuki, "Study on spectrum sharing method based on distance estimation for cognitive radio networks," in *Global Communications Conference (GLOBECOM), 2013 IEEE*, Dec 2013, pp. 890–895.

[40] J. Werner, A. Hakkarainen, and M. Valkama, "Cramer-rao bounds for hybrid rss-doa based emitter location and transmit power estimation in cognitive radio systems," in *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*, Sept 2013, pp. 1–7.

[41] J. Jiang, Y. Du, and P. Wei, "Detection of the number of sources based on power estimation," in *Communications, Circuits and Systems (ICCCAS), 2013 International Conference on*, vol. 2, Nov 2013, pp. 278–282.

[42] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984–995, Jul 1989.

[43] W. Jing, H. Jianguo, H. Chengbing, H. Hai, and W. Huihui, "Detection of the number of signals by a threshold based on peak-to-average ratio," in *Signal Processing, Communications and Computing (ICSPCC), 2011 IEEE International Conference on*, Sept 2011, pp. 1–4.

[44] G. Anand and P. Nagesha, "Source number estimation in non-gaussian noise," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, Sept 2014, pp. 1711–1715.

[45] H. Asadi and B. Seyfe, "Source number estimation via entropy estimation of eigenvalues (eee) in gaussian and non-gaussian noise," *arXiv preprint arXiv:1311.6051*, 2013.

[46] H. Liu, Y. Wang, J. Yang, and Y. Chen, "Fast and practical secret key extraction by exploiting channel response," in *INFOCOM, 2013 Proceedings IEEE*, April 2013, pp. 3048–3056.

[47] O. Gungor, F. Chen, and C. Koksal, "Secret key generation via localization and mobility," *Vehicular Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[48] H. Liu, J. Yang, Y. Wang, Y. Chen, and C. Koksal, "Group secret key generation via received signal strength: Protocols, achievable rates, and implementation," *Mobile Computing, IEEE Transactions on*, vol. 13, no. 12, pp. 2820–2835, Dec 2014.

[49] D. Schurmann and S. Sigg, "Secure communication based on ambient audio," *Mobile Computing, IEEE Transactions on*, vol. 12, no. 2, pp. 358–370, Feb 2013.

[50] Y. Wei, K. Zeng, and P. Mohapatra, "Adaptive wireless channel probing for shared key generation based on pid controller," *Mobile Computing, IEEE Transactions on*, vol. 12, no. 9, pp. 1842–1852, Sept 2013.

[51] Y. Qiao, K. Srinivasan, and A. Arora, "Configuration hopping: A secure communication protocol without explicit key exchange," in *Stabilization, Safety, and Security of Distributed Systems*, ser. Lecture Notes in Computer Science, P. Felber and V. Garg, Eds. Springer International Publishing, 2014, vol. 8756, pp. 268–282. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11764-5_19

[52] V. Thayananthan, A. Alzahrani, and M. S. Qureshi, "Efficient techniques of key management and quantum cryptography in rfid networks," *Security and Communication Networks*, vol. 8, no. 4, pp. 589–597, 2015. [Online]. Available: http://dx.doi.org/10.1002/sec.1005

[53] C. Chen and M. Jensen, "Secret key establishment using temporally and spatially correlated wireless channel coefficients," *Mobile Computing, IEEE Transactions on*, vol. 10, no. 2, pp. 205–215, Feb 2011.

[54] Y. El Hajj Shehadeh, O. Alfandi, and D. Hogrefe, "Towards robust key extraction from multipath wireless channels," *Communications and Networks, Journal of*, vol. 14, no. 4, pp. 385–395, Aug 2012.

[55] S. Jana, S. N. Premnath, M. Clark, S. K. Kasera, N. Patwari, and S. V. Krishnamurthy, "On the effectiveness of secret key extraction from wireless signal strength in real environments," in *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '09. New York, NY, USA: ACM, 2009, pp. 321–332. [Online]. Available: http://doi.acm.org/10.1145/1614320.1614356

[56] H. Liu, J. Yang, Y. Wang, and Y. Chen, "Collaborative secret key extraction leveraging received signal strength in mobile wireless networks," in *INFOCOM, 2012 Proceedings IEEE*, March 2012, pp. 927–935.

[57] F. Wu, W. Wang, B. Yao, and Q. Yin, "Effective eavesdropping in the artificial noise aided security scheme," in *Communications in China (ICCC), 2013 IEEE/CIC International Conference on*, Aug 2013, pp. 214–218.

[58] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 10, no. 3, pp. 626–634, May 1999.

[59] B. Danev, H. Luecken, S. Capkun, and K. El Defrawy, "Attacks on physical-layer identification," in *Proceedings of the Third ACM Conference on Wireless Network Security*, ser. WiSec '10. New York, NY, USA: ACM, 2010, pp. 89–98. [Online]. Available: http://doi.acm.org/10.1145/1741866.1741882

[60] A. Badawy, T. Khattab, T. ElFouly, A. Mohamed, and D. Trinchero, "Secret key generation based on channel and distance measurements," in *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2014 6th International Congress on*, Oct 2014, pp. 136–142.

[61] M. S. BARTLETT, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, no. 1-2, pp. 1–16, 1950. [Online]. Available: http://biomet.oxfordjournals.org/content/37/1-2/1.short

[62] H. Akaike, "A new look at the statistical model identification," *Automatic Control, IEEE Transactions on*, vol. 19, no. 6, pp. 716–723, Dec 1974.

[63] M. Xiao, P. Wei, and H.-M. Tai, "Estimation of the number of sources based on hypothesis testing," *Communications and Networks, Journal of*, vol. 14, no. 5, pp. 481–486, Oct 2012.

[64] R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 2625–2638, July 2008.

[65] V. A. Marenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, p. 457, 1967. [Online]. Available: http://stacks.iop.org/0025-5734/1/i=4/a=A01

[66] D. Freedman, *Statistical Models: Theory and Practice*. Cambridge University Press, 2009. [Online]. Available: https://books.google.com.qa/books?id=fW_9BV5Wpf8C

[67] P. Hansen, V. Pereyra, and G. Scherer, *Least Squares Data Fitting with Applications*, ser. Least Squares Data Fitting with Applications. Johns Hopkins University Press, 2012. [Online]. Available: https://books.google.com.qa/books?id=ynMi1ND6YjYC

[68] J. Myers, A. Well, and R. Lorch, *Research Design and Statistical Analysis: Third Edition*. Taylor & Francis, 2013. [Online]. Available: https://books.google.com.qa/books?id=7ep8FjCcloYC

[69] F. Yan, M. Jin, and X. Qiao, "Low-complexity doa estimation based on compressed music and its performance analysis," *Signal Processing, IEEE Transactions on*, vol. 61, no. 8, pp. 1915–1930, April 2013.

[70] T. Ferreira, S. Netto, and P. Diniz, "Direction-of-arrival estimation using a low-complexity covariance-based approach," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 3, pp. 1924–1934, JULY 2012.

[71] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the music algorithm," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 4, pp. 532–544, Apr 1988.

[72] R. Al Alawi, "Rssi based location estimation in wireless sensors networks," in *Networks (ICON), 2011 17th IEEE International Conference on*, Dec 2011, pp. 118–122.

[73] D. Gualda, J. Urena, J. Garcia, E. Garcia, and D. Ruiz, "Rssi distance estimation based on genetic programming," in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*, Oct 2013, pp. 1–8.

[74] M. Pricone and A. Caracas, "A heterogeneous rssi-based localization system for indoor and outdoor sports activities," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International*, Aug 2014, pp. 274–280.

[75] G. Shen, R. Zetik, and R. Thoma, "Performance comparison of toa and tdoa based location estimation algorithms in los environment," in *Positioning, Navigation and Communication, 2008. WPNC 2008. 5th Workshop on*, March 2008, pp. 71–78.

[76] L. Tan and J. Jiang, *Digital Signal Processing: Fundamentals and Applications*. Elsevier Science, 2013. [Online]. Available: https://books.google.com.qa/books?id=M9-OhaJSwAEC

[77] J. Zhang, S. Kasera, and N. Patwari, "Mobility assisted secret key generation using wireless link signatures," in *INFOCOM, 2010 Proceedings IEEE*, March 2010, pp. 1–5.

[78] E. Fishler, M. Grosmann, and H. Messer, "Detection of signals by information theoretic criteria: general asymptotic performance analysis," *Signal Processing, IEEE Transactions on*, vol. 50, no. 5, pp. 1027–1036, May 2002.

[79] M. Rubsamen and A. Gershman, "Direction-of-arrival estimation for nonuniform sensor arrays: From manifold separation to fourier domain music methods," *Signal Processing, IEEE Transactions on*, vol. 57, no. 2, pp. 588–599, Feb 2009.

[80] K. Mizutani, T. Ito, M. Sugimoto, and H. Hashizume, "Tsat-music: a novel algorithm for rapid and accurate ultrasonic 3d localization," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, 2011. [Online]. Available: http://dx.doi.org/10.1186/1687-6180-2011-101

[81] F. Yan, M. Jin, and X. Qiao, "Low-complexity doa estimation based on compressed music and its performance analysis," *Signal Processing, IEEE Transactions on*, vol. 61, no. 8, pp. 1915–1930, April 2013.

[82] Y. Jing, N. Feng, and Y. Shen, "Modified music estimation for correlated signals with compressive sampling arrays," *Systems Engineering and Electronics, Journal of*, vol. 25, no. 5, pp. 755–759, Oct 2014.

[83] "Warp project." [Online]. Available: http://warpproject.org

[84] H. David and H. Nagaraja, *Order Statistics*, ser. Wiley Series in Probability and Statistics. Wiley, 2004. [Online]. Available: https://books.google.com.qa/books?id=bdhzFXg6xFkC

Appendix A

Proof of Proposition 1

By substituting (4.18) into (4.17):

$$STD(i) = \sqrt{\left(\lambda_i - \frac{\lambda_i + \lambda_{i-1}}{2}\right)^2 + \left(\lambda_{i-1} - \frac{\lambda_i + \lambda_{i-1}}{2}\right)^2}. \tag{A.1}$$

$STD(i)$ with some manipulation can be written as:

$$STD(i) = \sqrt{\left(\frac{\lambda_i - \lambda_{i-1}}{2}\right)^2 + \left(\frac{\lambda_{i-1} - \lambda_i}{2}\right)^2}. \tag{A.2}$$

Because of the $(.)^2$ operation, $\left(\frac{\lambda_{i-1}-\lambda_i}{2}\right)^2$ can be written as $\left(\frac{\lambda_i-\lambda_{i-1}}{2}\right)^2$. This leads to:

$$STD(i) = \sqrt{2\left(\frac{\lambda_i - \lambda_{i-1}}{2}\right)^2}, \tag{A.3}$$

which can be further reduced to:

$$STD(i) = \frac{\lambda_i - \lambda_{i-1}}{\sqrt{2}}. \tag{A.4}$$

Applying the same for $STD(i-1)$ leads to:

$$STD(i-1) = \frac{\lambda_{i-1} - \lambda_{i-2}}{\sqrt{2}}. \tag{A.5}$$

Then substituting (A.4) and (A.5) into (4.19), α_i can be written as:

$$\alpha_i = \frac{(\lambda_i - \lambda_{i-1}) - (\lambda_{i-1} - \lambda_{i-2})}{\sqrt{2}}, \tag{A.6}$$

which results directly to (4.20).

Appendix B

Sample Examples for Number of Sources Estimation

In this appendix, we show two simple examples of number of sources estimation to illustrate the process highlighted in 4.2.4. In the first example, the number of sources estimation works fine for all algorithms while it fails for the second one. In each example, we show the eigenvalues, MI values and MS values and plot them with respect to eigenvalues index.

The first toy problem parameters are: SNR value of -5 dB, number of samples of 2^{10} and 2 sources separated by 100°. With such parameters, all the proposed algorithms had a correct estimation of 2 number of sources. As shown in Table B.1 and Fig. B.1, CorrM MS and MI have a maximum value when the eigenvalues move from the noise to signal subspace and hence the number of sources is estimated at that index. For CovM based MS and MI, there is a clear jump from noise to signal subspace' however, it was not the maximum jump and hence a threshold is needed. The estimated thresholds are 0.0014 and 0.00065 for MI and MS respectively. Algorithms works perfectly in this case and the difference between the noise eigenvalues does not exceed the threshold.

CovM Based			CorrM Based		
Eigenvalue	**MI**	**MS**	**Eigenvalue**	**MI**	**MS**
0.0029	0	0	0.6881	0	0
0.0030	0.0001	0.0000	0.7147	0.0267	0.0267
0.0031	0.0002	0.0000	0.7501	0.0354	0.0354
0.0032	0.0000	-0.0001	0.7608	0.0107	0.0107
0.0033	0.0002	0.0001	0.7959	0.0351	0.0351
0.0034	0.0001	-0.0000	0.8180	0.0220	0.0220
0.0066	0.0032	0.0016	1.5884	0.7704	0.7704
0.0170	0.0103	0.0036	1.8840	0.2957	0.2957

Table B.1: First Examples Eigenvalues, MI and MS values

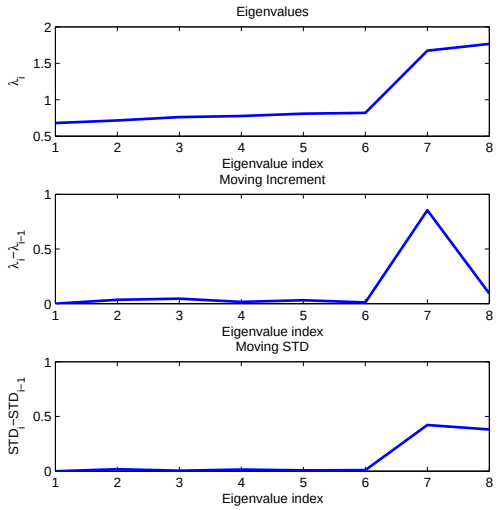


Figure B.1: Eigenvalues, Moving increment and Moving STD for CorrM Based Algorithm (1^{st} example)

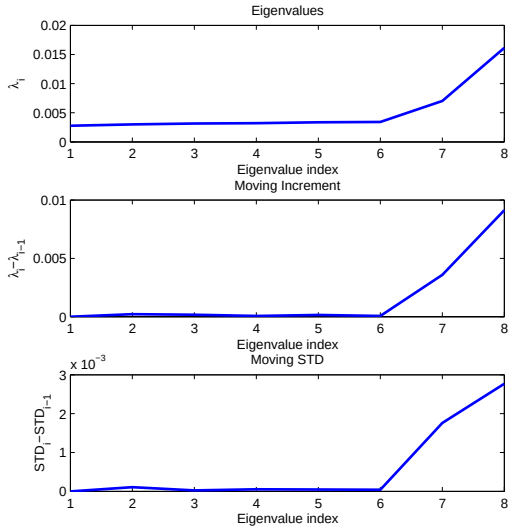


Figure B.2: Eigenvalues, Moving increment and Moving STD for CovM Based Algorithm (1^{st} example)

In the second example, low SNR and number of samples conditions are applied in order to fail the proposed algorithms. SNR value is set to -10 dB and the number of samples is 100 sample only and 2 sources are applied separated by 100°. CorrM based algorithms and $\mathbf{MMS_{Cov}}$ underestimate the number of sources as one, mostly while $\mathbf{MMS_{Cov}}$ differs in its estimation as the difference in noise eigenvalues exceed the calculated threshold. Table B.2 and Figs. B.3 and B.4 show the results of this simulation and where it fails. As can be seen in Fig. B.3, the maximization point is located at the last index, hence the number of sources is underestimated to 1 instead of 2. In Fig. B.4, MS algorithm exceed the threshold at the last index while the MI algorithm exceed the threshold at eigenvalues index 4. Hence, MS underestimates the number of sources to 1 while MI overestimates to 5 where the thresholds are 0.00065 & 0.0018 for MS and MI respectively.

CovM Based			CorrM Based		
Eigenvalue	**MI**	**MS**	**Eigenvalue**	**MI**	**MS**
0.0064	0	0	0.6010	0	0
0.0072	0.0007	0.0004	0.6754	0.0744	0.0372
0.0086	0.0014	0.0003	0.7796	0.1042	0.0149
0.0107	0.0021	0.0003	0.9256	0.1460	0.0209
0.0115	0.0009	-0.0006	0.9996	0.0740	0.0360
0.0131	0.0015	0.0003	1.1315	0.1319	0.0290
0.0145	0.0015	0	1.2888	0.1573	0.0127
0.0263	0.0118	0.0052	1.5984	0.3096	0.0762

Table B.2: Second Examples Eigenvalues, MI and MS values

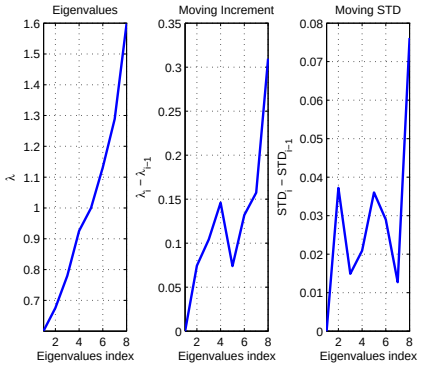


Figure B.3: Eigenvalues, Moving increment and Moving STD for CorrM Based Algorithm (2^{nd} example)

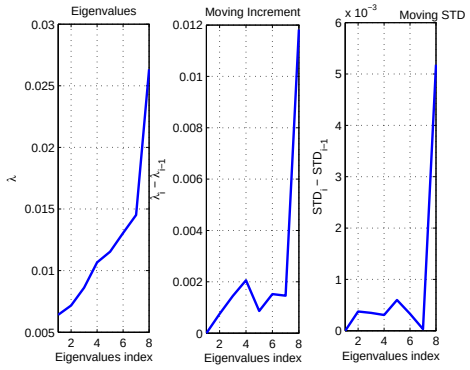


Figure B.4: Eigenvalues, Moving increment and Moving STD for CovM Based Algorithm (2^{nd} example)

Appendix C

Proof for Proposition 2

The joint probability density function of the difference between two ordered independent random variables, X_r and X_S with $1 \leq r < s < n$, $W_{rs} = X_s - X_r$ is [84]:

$$f_{W_{rs}}(w_{rs}) = D_{rs} \int_{-\infty}^{\infty} F(x)^{r-1} f(x) \Big(F(x + w_{rs}) - F(x) \Big)^{s-r-1} f(x + w_{rs}) \Big(1 - F(x + w_{rs}) \Big)^{n-s} dx. \tag{C.1}$$

In our case, we need to find the joint the probability density function of $\delta_i = \lambda_i - \lambda_{i-1}$, hence, we let $r = i - 1$, $s = i$ and $\delta_i = W_{rs}$ and let $D = D_{rs}$. In this way, (C.1) for $\delta_i = \delta$ and $\lambda_i = \lambda$ can be redefined as :

$$f_{\delta_i}(\delta) = D \int_{-\infty}^{\infty} F(\lambda)^{i-2} f(\lambda) \Big(F(\lambda + \delta) - F(\lambda) \Big)^{i-i+1-1} f(\lambda + \delta) \Big(1 - F(\lambda + \delta) \Big)^{M-K-i} d\lambda. \tag{C.2}$$

Since $i - i + 1 - 1 = 0$, the term $\Big(F(\lambda + \delta) - F(x) \Big)^{i-i+1-1}$ will be 1 and (C.2) can be rewritten as :

$$f_{\delta_i}(\delta) = D \int_{-\infty}^{\infty} F(\lambda)^{i-2} f(\lambda) f(\lambda + \delta) \Big(1 - F(\lambda + \delta) \Big)^{M-K-i} d\lambda,$$

where D is a constant defined as:

$$\begin{aligned} D &= \frac{(M-K)!}{(r-1)!(s-r-1)!(M-K-s)!} = \frac{(M-K)!}{(i-2)!(i-i+1-1)!(M-K-i)!} \\ &= \frac{(M-K)!}{(i-2)!(M-K-i)!} \end{aligned} \tag{C.3}$$

$f(\lambda)$ is defined in (4.24). On the assumption that G is greater than 1, i.e. $N > M$, the first term in (4.24) will be canceled and (4.24) will be represented by its second term. $F(\lambda)$ can be defined as:

$$\begin{aligned} F(\lambda) &= \int_{-\infty}^{\lambda} f(\lambda) d\lambda = \int_{-\infty}^{\lambda} \frac{\sqrt{(\lambda - a_-)(a_+ - \lambda)}}{2\pi\sigma\lambda(1/G)} \Pi_{[a_-, a_+]}(\lambda) d\lambda = \int_{-\infty}^{\lambda} \frac{G}{2\pi\sigma} \frac{\sqrt{(\lambda - a_-)(a_+ - \lambda)}}{\lambda} \Pi_{[a_-, a_+]}(\lambda) d\lambda \\ &= \frac{G}{2\pi\sigma} \int_{a_-}^{\lambda} \frac{\sqrt{(\lambda - a_-)}\sqrt{(a_+ - \lambda)}}{\lambda} d\lambda \end{aligned} \tag{C.4}$$

Assuming that $0 \leq a_- \leq \lambda \leq a_+$:

$$
\begin{aligned}
F(\lambda) &= \frac{G}{2\pi\sigma}\frac{1}{4}\frac{1}{\sqrt{a_-}\sqrt{a_+}}\Bigg(2\arcsin\big(\frac{-2\lambda+a_-+a_+}{a_--a_+}\big)a_-^{3/2}\sqrt{a_+}+2\arcsin\big(\frac{-2\lambda+a_-+a_+}{a_--a_+}\big)a_+^{3/2}\sqrt{a_-}+4a_+a_- \\
&\arctan\Big(\frac{1}{2}\frac{2a_-a_+-a_+\lambda-a_-\lambda}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+-\lambda)(\lambda-a_-)}}\Big)+\pi a_-^{3/2}\sqrt{a_+}+\pi\sqrt{a_-}a_+^{3/2}-2a_-a_+\pi+4\sqrt{a_-}\sqrt{a_+}\sqrt{\lambda-a_-}\sqrt{a_+-\lambda}\Bigg) \\
&= \frac{G}{4*2\pi\sigma}\Bigg\{\frac{1}{\sqrt{a_-}\sqrt{a_+}}\Bigg(2\arcsin\big(\frac{-2\lambda+a_-+a_+}{a_--a_+}\big)a_-^{3/2}\sqrt{a_+}+2\arcsin\big(\frac{-2\lambda+a_-+a_+}{a_--a_+}\big)a_+^{3/2}\sqrt{a_-}+4a_+a_- \\
&\arctan\Big(\frac{1}{2}\frac{2a_-a_+-a_+\lambda-a_-\lambda}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+-\lambda)(\lambda-a_-)}}\Big)+\pi a_-^{3/2}\sqrt{a_+}+\pi\sqrt{a_-}a_+^{3/2}-2a_-a_+\pi+4\sqrt{a_-}\sqrt{a_+}\sqrt{\lambda-a_-}\sqrt{a_+-\lambda}\Bigg)\Bigg\}
\end{aligned}
\tag{C.5}
$$

By that, (C.2) can be rewritten as:

$$
\begin{aligned}
f_{\delta_i}(\delta) &= \frac{(M-K)!}{(i-2)!(M-K-i)!}\int_{-\infty}^{\infty}\Bigg(\frac{G}{2\pi\sigma}\int_{a_-}^{\lambda}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda}d\lambda\Bigg)^{i-2}\Bigg(\frac{G}{2\pi\sigma}\frac{\sqrt{(\lambda-a_-)(a_+-\lambda)}}{\lambda}\Pi_{[a_-,a_+]}(\lambda)\Bigg) \\
&\Bigg(\frac{G}{2\pi\sigma}\frac{\sqrt{(\lambda+\delta-a_-)(a_+-\lambda-\delta)}}{\lambda}\Pi_{[a_-,a_+]}(\lambda+\delta)\Bigg)\Bigg(1-\frac{G}{2\pi\sigma}\int_{a_-}^{\lambda+\delta}\frac{\sqrt{(\lambda+\delta-a_-)}\sqrt{(a_+-\lambda-\delta)}}{\lambda+\delta}d\lambda\Bigg)^{M-K-i}d\lambda \\
&= \frac{(M-K)!}{(i-2)!(M-K-i)!}\Bigg(\frac{G}{2\pi\sigma}\Bigg)^{1+i-2+1}\int_{a_-}^{a_+}\frac{\sqrt{(\lambda-a_-)(a_+-\lambda)}}{\lambda}\Bigg(\int_{a_-}^{\lambda}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda}d\lambda\Bigg)^{i-2} \\
&\Bigg(\frac{\sqrt{(\lambda+\delta-a_-)(a_+-\lambda-\delta)}}{\lambda}\Bigg)\Bigg(1-\frac{G}{2\pi\sigma}\int_{a_-}^{\lambda+\delta}\frac{\sqrt{(\lambda+\delta-a_-)}\sqrt{(a_+-\lambda-\delta)}}{\lambda+\delta}d\lambda\Bigg)^{M-K-i}d\lambda \\
&= \frac{(M-K)!}{(i-2)!(M-K-i)!}\Bigg(\frac{G}{2\pi\sigma}\Bigg)^{i}\int_{a_-}^{a_+}\frac{\sqrt{(\lambda-a_-)(a_+-\lambda)}}{\lambda}\Bigg(\int_{a_-}^{\lambda}\frac{\sqrt{(\lambda-a_-)}\sqrt{(a_+-\lambda)}}{\lambda}d\lambda\Bigg)^{i-2} \\
&\Bigg(\frac{\sqrt{(\lambda+\delta-a_-)(a_+-\lambda-\delta)}}{\lambda}\Bigg)\Bigg(1-\frac{G}{2\pi\sigma}\int_{a_-}^{\lambda+\delta}\frac{\sqrt{(\lambda+\delta-a_-)}\sqrt{(a_+-\lambda-\delta)}}{\lambda+\delta}d\lambda\Bigg)^{M-K-i}d\lambda,
\end{aligned}
\tag{C.6}
$$

Substituting (C.5) in (C.6) will lead to:

$$
\begin{aligned}
f_{\delta_i}(\delta) &= \frac{(M-K)!}{(i-2)!(M-K-i)!}\Bigg(\frac{G}{2\pi\sigma}\Bigg)^{i}\int_{a_-}^{a_+}\frac{\sqrt{(\lambda-a_-)(a_+-\lambda)}}{\lambda} \\
&\Bigg\{\frac{1}{4\sqrt{a_-}\sqrt{a_+}}\Bigg[2\arcsin\big(\frac{-2\lambda+a_-+a_+}{a_--a_+}\big)a_-^{3/2}\sqrt{a_+}+2\arcsin\big(\frac{-2\lambda+a_-+a_+}{a_--a_+}\big)a_+^{3/2}\sqrt{a_-}+4a_+a_- \\
&\arctan\Big(\frac{1}{2}\frac{2a_-a_+-a_+\lambda-a_-\lambda}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+-\lambda)(\lambda-a_-)}}\Big)+\pi a_-^{3/2}\sqrt{a_+}+\pi\sqrt{a_-}a_+^{3/2}-2a_-a_+\pi+4\sqrt{a_-}\sqrt{a_+}\sqrt{\lambda-a_-}\sqrt{a_+-\lambda}\Bigg]\Bigg\}^{i-2} \\
&\Bigg(\frac{\sqrt{(\lambda+\delta-a_-)(a_+-\lambda-\delta)}}{\lambda}\Bigg)\Bigg(1-\frac{G}{2\pi\sigma}\Bigg\{\frac{1}{4\sqrt{a_-}\sqrt{a_+}}\Bigg[2\arcsin\big(\frac{-2(\lambda+\delta)+a_-+a_+}{a_--a_+}\big)a_-^{3/2}\sqrt{a_+} \\
&+2\arcsin\big(\frac{-2(\lambda+\delta)+a_-+a_+}{a_--a_+}\big)a_+^{3/2}\sqrt{a_-}+4a_+a_-\arctan\Big(\frac{1}{2}\frac{2a_-a_+-a_+(\lambda+\delta)-a_-(\lambda+\delta)}{\sqrt{a_-}\sqrt{a_+}\sqrt{(a_+-(\lambda+\delta))((\lambda+\delta)-a_-)}}\Big) \\
&+\pi a_-^{3/2}\sqrt{a_+}+\pi\sqrt{a_-}a_+^{3/2}-2a_-a_+\pi+4\sqrt{a_-}\sqrt{a_+}\sqrt{(\lambda+\delta)-a_-}\sqrt{a_+-(\lambda+\delta)}\Bigg]\Bigg\}\Bigg)^{M-K-i}d\lambda
\end{aligned}
\tag{C.7}
$$

which leads directly to (4.25) for the probability distribution function.