2019

Secure State Estimation in the Presence of False Information Injection Attacks

Maitham ALSalman
Virginia Commonwealth University

SECURE STATE ESTIMATION IN THE PRESENCE OF FALSE INFORMATION INJECTION ATTACKS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

MAITHAM AL-SALMAN

B. Sc. in Electrical Engineering, AL-Mustansiriya University

Baghdad, 2008

M. Sc. in Electrical Engineering, AL-Mustansiriya University

Baghdad, 2011

Advisor: RUIXIN NIU

Associate Professor, Department of Electrical and Computer Engineering

Virginia Commonwealth University

Richmond, VA

December 2019

Acknowledgements

First, I would like to express my sincere gratitude to my advisor, Dr. Ruixin Niu, for his invaluable guidance and support throughout my graduate study. I am deeply indebted to his insightful suggestions on my research, which always bring my understanding of the research problem to a higher level. He spent endless time reviewing and proofreading my research documents and giving me excellent comments which always resulted in improved versions of research documents including this dissertation. This dissertation would not have been possible without his patience, encouragements, and immense knowledge.

In addition, I would like to thank the other members of my doctoral advisory committee: Dr. Alen Docef, Dr. Yuichi Motai, Dr. Wei Cheng, and Dr. Yanjun Qian, for their encouragement, assistance, insightful comments, and suggestions.

Further, I am thankful for the financial support from the Higher Committee for Education Development (HCED) in Iraq.

I would also like to thank my friends and colleagues for their friendship, and encouragement, throughout my period of study.

Finally, I would like to thank my loving parents and extended family members for their patience and support during these years.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

ABSTRACT

In this dissertation, we first investigate the problem of source location estimation in wireless sensor networks (WSNs) based on quantized data in the presence of false information attacks. Using a Gaussian mixture to model the possible attacks, we develop a maximum likelihood estimator (MLE) to locate the source with sensor data corrupted by injected false information, and call the approach quantized received signal strength with a Gaussian mixture model (Q-RSS-GM). The Cramer-Rao lower bound (CRLB) for this estimation problem is also derived to evaluate the estimation performance. It is shown that the proposed estimator is robust in various cases with different attack probabilities and parameter mismatch, and it significantly outperforms the approach that ignores the possible false information attacks.

Then, we investigate the case with the assumption that the fusion center does not have the knowledge of the attack probability and the attack noise power, which is a more realistic assumption for real life applications. We assume that the attack probability and power are random variables which follow certain uniform distributions and are independent and identically distributed among sensors. We derive the MLE for the localization problem based on quantized received signal strength which is corrupted by the false information injection. The CRLB for this estimation problem is also derived to evaluate the estimation performance. It is shown that the proposed new MLE is robust and provides excellent performance without knowing the attack parameters, such as attack probability and attack power.

The linear state estimation problem subjected to a spoofing attack (False Information Injection) is also considered in this dissertation. We investigate the problem of the Bayesian estimation in linear systems in the presence of false information injection attack. The relationship between the attacker and the defender is modeled from a minimax perspective, in which the attacker tries to maximize the cost function. On the other hand, the defender tries to optimize the detection threshold selection to minimize the attack effects on the system. We address the problem for two situations.

First, we consider that the attacker will attack with a deterministic bias injection. In this case, we derive the probabilities of detection and miss according to the non-central Chi squared distribution. The probability of false alarm is derived based on the Chi squared distribution. We investigate the minimax optimization problem for the cost function numerically. It is shown that for the attacker increasing the attack power will maximize the cost function in general. On the other hand, for the defender it is shown that there are three different regions in which the defender can work to design the defending strategy. The defender will be able to optimize the cost function if it has the prior information about the attacker power.

Secondly, we consider that the attacker will attack with a random bias injection. In this case, we calculate the probabilities of the detection and miss detection according to a numerical integration method. The probability of false alarm is obtained by using the Chi squared distribution. We also formulate the minmax optimization problem between the attacker and the defender. In this case, the attacker will be able to maximize its effect on the system by increasing the injected bias covarinace matrix, whereas if the defender has the prior information of the attack power, he/she will be able to select the optimum detection threshold to minimize the cost function.

We solve the minimax optimization problem numerically for both the cases with deterministic and random biases. Numerical results show that if the defender has prior knowledge of the attacker power, it can select the optimum detection threshold that minimizes the worst possible cost function accordingly. On the other hand, if the defender has no prior knowledge of the attacker's power, then the best strategy for the defender is to always reject the corrupted sensors' measurements.

CHAPTER 1

INTRODUCTION

1.1 Motivation and Background

The main goal of wireless sensor networks (WSNs) is to collect the sensors' measurements from a certain target or natural phenomena and send them to the fusion center (FC) which processes the data and makes the required decisions accordingly. Such characteristics can be employed in vast applications [1, 2, 3, 4]. The other important aspect that makes the WSNs applicable for a wide range of applications is the development of the wireless communication networks and devices which culminated with the Internet that gives the possibility of connecting multiple devices spread in a very wide area (around the globe). We can employ the devices connected to the Internet like mobile phones and PCs as sensor devices, which can be utilized to gather and transmit the data over the Internet. Such applications are called the internet of things (IoT) [5] .

Because of the breadth of the applications, in which WSNs can play a crucial role, there is increased interest in studying the challenges of the WSNs. There are two main challenges faced by sensors networks. The first challenge involves manufacturing of these critical devices and trying to equip them with the required parts that ensure their ability to operate in the field for long periods of time without human intervention. The second challenge involves assuring the reliability and the accuracy of the collected sensor measurements which can be affected by either an adversary, or natural noise interference, that might corrupt the data and mislead the fusion center (FC). In this dissertation, we are mainly concerned with the data reliability perspective and developing the appropriate estimators that have the capability to tolerate the corrupted data.

The developments in communication networks and devices did not change the basic structure of the WSN, which consists of a large number of devices (sensors) spread in a wide area also known as region of interest (ROI). The basic structure of the WSNs creates multiple threats to the security of the sensors' measurements which will affect the applications based on the received data. There are several limitations for the sensors devices like limited energy, communication and data processing power, making them susceptible to multiple kinds of attacks or natural noise interference, that may corrupt the collected data and cause the fusion center make inaccurate decisons about the monitored phenomena. Sensor data digitization and quantization have been adopted as a solution for the limited resources. For example, we can quantize the received signal strength at the sensors before its transmission. This is also known as quantized recived signal strength (Q-RSS) approach, which is used in many WSNs applications [6].

According to the previous illustration of the typical structure of the WSNs and the security issues that arise from such structure, the security of WSNs has become an important topic which has been studied recently [7, 8]. Several attacks that the WSNs might be facing, like Byzantine, man in the middle (MIMA), and spoofing attacks, have been studied and classified [3]. Intentional sensor destruction by an adversary could be also considered as a spoofing attack.

The spoofing attack changes the measurements from the sensors and it takes place either between the sensors and the source/phenomena (corrupt the data entering the sensors) or between the sensor and the fusion center (corrupt the data sent from the sensor to the fusion center). False information injection attacks are considered as spoofing attacks [9]. The spoofing attack in the wireless sensor networks and multiple sensors estimation systems has attracted increasing attention recently [3] . The spoofing attack can occur on the Global Positioning System (GPS), radar and sonar systems, localization system via WSNs [3, 10]. We will use these two terminologies i.e. false information injection and spoofing attack interchangeably in the dissertation. There are also Byzantine attacks which happen when the adversary takes control of a portion of the sensors and changes their quantized measurements [2, 11, 12].

The problem of general parameter estimation in the presence of Byzantine or spoofing attacks has been investigated in [1, 2, 12, 13, 14, 3]. In [1], Byzantine attacks on sensor networks estimating unknown parameters based on binary data have been studied. Two attack strategies depending on the amount of information available to the adversary, namely full information and information free attacks, have been investigated. Full information attacks proved to be more destructive. In [2, 12], Byzantine attacks on sensor networks were studied and a method for attack detection and classification was proposed, which has been employed to improve the estimation performance. Asymptotic results were provided as the number of measurements and the number of sensors increase. In [13], in the presence of false information injection (spoofing) attacks, several detection-estimation strategies have been proposed to minimize the average system estimation mean squared error (MSE). In [14], a heuristic method was proposed to jointly identify the attacked sensors and estimate the desired parameter. The attacks might be caused by intentionally injected false data or by natural phenomena.

The problem of localization in sensor networks under Byzantine or spoofing attacks was studied in [11, 15, 9, 16, 10]. In [11], the optimal Byzantine attack strategies to compromise a localization system based on quantized sensor data were proposed and the posterior Cramer-Rao lower bound (PCRLB) for performance evaluation was derived. An approach based on error correcting codes was developed in [15] to mitigate Byzantine attacks and estimate the target location. In [10], the identification and detection of location spoofing attacks on sensor networks based on RSS and beamforming estimates have been examined. In [9], a class of approaches was developed to detect the attacks on a localization sensor network based on quantized data, where the attack is a combination of man-in-the-middle, hacking, and spoofing attacks. In [16], median based robust methods were developed to tolerate the bad sensor data attacked by the adversary, rather than detecting and eliminating them.

In [7], we addressed the issue of false information injection attacks using the Gaussian mixture model (GMM). The GMM has been used for localization in sensor networks [17, 18, 19, 20, 21, 22]. For example, in [17], centralized and distributed maximum likelihood based algorithms for location estimation using RSS were proposed. In [22], a nonlinear and nonconvex source localization problem based on RSS has been relaxed and solved using a semidefinite Gaussian mixture algorithm. However, in all the work in [17, 18, 19, 20, 21, 22], analog sensor data have been assumed.

Inspired by [6] and the aforementioned work, which has used the GMM noise model, we proposed a Maximum Likelihood estimator (MLE) based approach to deal with quantized received signal strength corrupted by spoofing attacks with a Gaussian mixture model (Q-RSS-GM). To the best of our knowledge, we are the first to apply GMM to deal with quantized RSS data. In the dissertation, we also derive the CRLB for the proposed estimator to evaluate its performance.

It is worth mentioning here that our work on localization based WSNs is composed of two parts. In both parts, we have proposed a source location estimation approach based on MLE - QRSS assuming a false information injection attack which can also be called more generally a spoofing attack. In the first part, we assume that the random attack is *i.i.d* at different sensors with deterministic parameters for both the attack probability p_a and the false information injection signal power σ_2^2. In contrast to the first part, in the second part we assume that both of the p_a and σ_2^2 are random variables. We assume that the attack probability p_a and the attacker signal power σ_2^2 are following two different uniform distributions. Based on this assumption, we derive the probability mass function for the quantize received signal strength (QRSS). For both parts, we derive the maximum likelihood estimators based on quantized received signal strength (MLE - QRSS), and in each case we have derived the corresponding CRLB for the proposed estimator for performance evaluation.

The problem of false information injection attack for linear system has been studied in [4, 13, 23] and references therein. In [4], the problem of optimal false information injection attack in multi-sensor system was investigated. In [4], the attacker tries to maximize Kalman filter's estimation error. It was shown that to maximize the estimation error, the attacker can optimize the power allocation among the sensors, and can optimally design the injected noise error covariance matrix $\boldsymbol{P}_{bb}$ from the adversary perspective.

In [13], the problem of false information attack detection strategies was investigated for Bayesian estimator. It was assumed that the attacker will use the optimal attack strategies to maximize the estimation error in case that no detection strategy was available at the estimation system. The defender aims to minmize the cost function, which is the trace of the MSE matrix for the estimation error. The proposed defending strategies depend on two principles. The first is the detection and discard strategy which will discard the attacked sensor after detection. The second strategy is detection and incorporation, which will utilize the information of the detected sensor and improve the MSE based on the available information. The detection and discard strategy is robust to the mismatch between the detector design assumption and the real attack parameters.

Inspired by the previous work [4, 13], we are investigating the solution to the minimax optimization problem. First, we assumed that the attacker will attack with a deterministic constant bias, and the attacker is trying to maximize the cost. On the other hand, the defender will work on minimizing the cost function by trying to select an optimum detection threshold, under hypothesis H_1, when there is an attack, the distribution of the detector's test statistic has been derived, which is non-central Chi squared distribution.

Next, we assume that the attacker will attack with a random noise and in this case the attacker will try to optimize the attack noise covariance matrix $\boldsymbol{P}_{bb}$. On the other hand, the defender will work on minimizing the cost by optimizing the detection threshold selection. In this case, we use numerical integration to calculate the detection and miss detection probabilities.

1.2 Contributions

In the first problem of this dissertation, the problem of false information injection is addressed using Gaussian mixture model in [7]. We have derived the MLE and its corresponding CRLB. The simulation results show that our proposed estimator provides robust performance under the spoofing attacks. They also show that the proposed estimator provides an acceptable performance in the case of mismatch between the nominal (estimated) and the true attack parameters.

The second problem is a natural extension to the first problem. The assumption of random attack probability and random noise power adopted by the estimator will make the estimator more robust to a general false information attack. The proposed estimator is also based on Q-RSS data, and we also derive the CRLB under this more general assumption for estimator performance evaluation.

In the second problem, we have derived the MLE and its corresponding CRLB.

In this case the situation is more general and realistic. The simulation results show that our proposed estimator provides robust performance under the spoofing attack assumption. They also show that the proposed estimator provides an excellent performance in the case of mismatch between the nominal (estimated) and the true attack parameters.

Note that the work in our dissertation is different from all the work mentioned earlier in [11, 15, 9, 16, 10]. First, it is different from [11, 15] in that we assume the sensor network is under spoofing attacks rather than Byzantine attacks. It is different from [10, 9], since our proposed approach does not try to detect the attacked senors. Instead, it is a localization approach that is robust under false information injection attacks, by assuming a certain attack probability. In this sense, our approach adopts a similar philosophy as that in [16], i.e. developing a robust approach that tolerates the bad sensor data. However, our work is quite different from [16]. In [16], the sensor data are assumed to be analog, the sensing modalities include distance (range) measurements and RF fingerprinting, and the robust approach is based on the median. In contrast, in this dissertation we propose a maximum likelihood estimator (MLE) based on quantized received signal strength data subject to false information injection attacks, by using a Gaussian mixture (GM) model, which can be used to address a general spoofing attack.

The third problem is a minimax optimization problem from both the attacker and defender perspectives. The optimal false information attack has been studied in [4], and the attack detection strategies was investigated in [13]. However, in both of the previous works and to the best of our knowledge, we are the first to investigate the minimax problem between the attacker and the defender from both the adversary and the defender perspectives. We have formulated the relationship between the attacker and the defender as a minimax optimization problem, where the attacker tries to maximize the cost function of the estimation error for the Bayesian estimator by controlling the attack bias vector. Whereas, the defender tries to optimize its defending strategy by selecting the optimum detection threshold that minimizes the cost function, which is trace of the average mean squared error of the estimator. We solved the minimax optimization problem numerically. Numerical results show that when the defender has no knowledge of the attacker power, the optimum strategy is to always reject the corrupted sensors' measurements by setting the threshold to zero. If the defender has a prior knowledge of the attacker power, then the defender will be able to minimize the attack effect by selecting the optimum threshold which will depend on the attack power used by the adversary.

1.3 Dissertation Outline

This dissertation is organized as follows. In Chapter 2, a general background for estimation theory is presented. The estimators and the theoretical bounds for the estimator's performance are presented. In Chapter 3, the Maximum Likelihood estimator based on quantized received signal strength for source location estimation in the presence of false information attacks, is proposed. The corresponding CRLB is derived. In Chapter 4, the maximum likelihood estimator based on quantized received signal strength for the case where the attack probability p_a, and the attack power σ_2^2 are random and follow certain distributions, is proposed. The corresponding CRLB is also derived in this case. In Chapter 5, the relationship between the defender and the attacker has been modeled as minimax optimization problem for attack on the Bayesian linear system estimator. Then, a numerical solution for the minimax problem investigated. Conclusion is presented in Chapter 6.

CHAPTER 2

BACKGROUND OF ESTIMATION THEORY

2.1 Introduction

In this chapter, a general background for estimation theory is presented. We start with the problem of estimating a fixed but unknown parameter from observed data. This type of estimation problem is also called a non-Bayesian estimation problem. We present the maximum likelihood estimator, which is a well known non-Bayesian estimator. Next, we present the case of estimating random parameters from observed data, which is also called Bayesian estimation problem. We present the Minimum Mean Squared Error (MMSE) as an estimation criterion for this case which we also will use later in our dissertation. We discuss the estimator performance measure Mean Squared Error (MSE), as well as the Fisher information matrix (FIM) and the Cramer Rao Lower Bound (CRLB). Biased and unbiased estimators are also discussed.

2.2 Non-Bayesian Estimation

Estimating an unknown deterministic parameter is a subject that has a wide range of applications. For example, the localization problem in wireless sensor network (WSNs). In the localization problem, our objective is to estimate the source location $\hat{\boldsymbol{\theta}}(\boldsymbol{Z})$ based on observed sensor measurements $\boldsymbol{Z} = [z_1, ..., z_N]^T$, where N is the number of measurements, which are usually considered to be independent and identically distributed $i.i.d$ and follow a statistical model $f(\boldsymbol{Z}|\boldsymbol{\theta})$. According to the $i.i.d$ assumption the joint likelihood function $L(\boldsymbol{Z}|\boldsymbol{\theta})$ can be found as [24]

$$L(\boldsymbol{Z}|\boldsymbol{\theta}) = f(\boldsymbol{Z}|\boldsymbol{\theta}) = \prod_{i=1}^{N} f(\boldsymbol{\theta}|\boldsymbol{Z}) \tag{2.1}$$

For simplicity of calculation it is also common to write the likelihood function eq.(2.1) in the logarithm form as

$$\log L(\boldsymbol{Z}|\boldsymbol{\theta}) = \sum_{i=1}^{N} \log f(\boldsymbol{\theta}|\boldsymbol{Z}) \tag{2.2}$$

Now the objective is to find the estimate $\hat{\boldsymbol{\theta}}(\boldsymbol{Z})$ based on the observed data $\boldsymbol{Z}$, we denote for $\boldsymbol{\theta}$ here as bold symbol since we are considering the estimation of a vector of unknown deterministic parameters.

For the unknown deterministic parameter, a widely used estimator to estimate the parameter $\boldsymbol{\theta}$ based on the likelihood or log-likelihood function is called the maximum likelihood estimator (MLE) which tries to maximize the log-likelihood function over the prameter $\boldsymbol{\theta}$ as [24], [25]

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log f(\boldsymbol{Z}|\boldsymbol{\theta}) \tag{2.3}$$

Thus, we have defined the process for estimating any unknown deterministic parameters based on received non-linear random data and the likelihood function. The (MLE) can be applied and it can find the estimate for all kinds of measurements with.

Next, we will present the estimator's evaluation methods.

2.3 Bayesian Estimation

Let us suppose that we have a set of measuremements $\boldsymbol{Z} = [z_1, ..., z_N]^T$, with a likelihood function $f(\boldsymbol{Z}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of the parameters to be estimated. Suppose we also have the prior probability density function $f(\boldsymbol{\theta})$, then using the Bayes' theorem the posterior distribution can be written as [25]

$$f(\boldsymbol{\theta}|\boldsymbol{Z}) = \frac{f(\boldsymbol{Z}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{Z})} \tag{2.4}$$

where

$$f(\boldsymbol{Z}) = \int f(\boldsymbol{Z}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{2.5}$$

Note that the denominator in the Bayes' formula does not depend on $\boldsymbol{\theta}$ and can be considered as a constant c, and that the posterior PDF is proportional likelihood function $f(\boldsymbol{Z}|\boldsymbol{\theta})$ and the prior pdf $f(\boldsymbol{\theta})$ as

$$f(\boldsymbol{\theta}|\boldsymbol{Z}) = \frac{1}{c}\boldsymbol{L}(\boldsymbol{Z}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \tag{2.6}$$

where $\boldsymbol{L}(\boldsymbol{\theta}) = f(\boldsymbol{Z}|\boldsymbol{\theta})$ is the likelihood function of $\boldsymbol{\theta}$. Thus, the posterior can be calculated from the Bayes' formula based on the observed data $\boldsymbol{Z}$ and the prior pdf $f(\boldsymbol{\theta})$. A well known Bayesian estimator is the maximum a posteriori estimator (MAP) which maximize the posterior pdf over $\boldsymbol{\theta}$. Another popular Bayesian estimator is the conditional mean $E[\boldsymbol{\theta}|\boldsymbol{Z}]$ mean $E[\boldsymbol{\theta}|\boldsymbol{Z}]$ which is also known as the minimum mean squared error (MMSE) as

$$E[\boldsymbol{\theta}|\boldsymbol{Z}] = \int \boldsymbol{\theta} f(\boldsymbol{\theta}|\boldsymbol{Z}) d\boldsymbol{\theta} = \frac{\int \boldsymbol{\theta} \boldsymbol{L}(\boldsymbol{Z}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int \boldsymbol{L}(\boldsymbol{Z}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{2.7}$$

In this dissertation we will use the MMSE estimator for linear measurements subjected to an injected false information. Next, we will give a background of the MMSE derivation based on linear measurements.

2.3.1 Bayesian Estimation in Linear and Gaussian Systems

In this section, the widely used linear system model is presented and the MMSE estimator is presented along with the covariance matrix under the assumption of Gaussian measurements. Since we will use the MMSE in linear system in Chapter 5 so we find it important to give a brief introduction to the MMSE estimator under linear system in this chapter to make our work more self - contained.

Suppose we have a set of observed data $\boldsymbol{z} = [z_1, ..., z_N]^T$, $z_i^\prime s$ are $i.i.d$ measurements and

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w} \tag{2.8}$$

the above model is widely used model for linear systems. It is common to denote $\boldsymbol{x}$ as the state vector, and $\boldsymbol{w}$ as the noise vector. $\boldsymbol{H}$ is the measurements matrix, and $\boldsymbol{z}$ is the measurements vector. Since we are considering the Bayesian case, the state vector to be estimated is assumed to be random parameter with a Gaussin prior as $\boldsymbol{x} \sim N(\bar{\boldsymbol{x}}, \boldsymbol{P}_{xx})$, where $\bar{\boldsymbol{x}}$ is the mean and the $\boldsymbol{P}_{xx} = E[(\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^T]$ is the covariance matrix of the state vector respectively. $\boldsymbol{w} \sim N(\bar{\boldsymbol{w}}, \boldsymbol{P}_{\boldsymbol{ww}})$ is denoted as the sensor measurements noise and has a Gaussian distribution with $\bar{\boldsymbol{w}}$ mean and covariance matrix $\boldsymbol{P}_{\boldsymbol{ww}} = E[(\boldsymbol{w}-\bar{\boldsymbol{w}})(\boldsymbol{w}-\bar{\boldsymbol{w}})^T]$ respectively. Let $\boldsymbol{x}$ and $\boldsymbol{w}$ be independent thus

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{w} \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{\boldsymbol{x}} \\ \bar{\boldsymbol{w}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{P}_{xx} & 0 \\ 0 & \boldsymbol{P}_{ww} \end{bmatrix}\right) \tag{2.9}$$

By using matrix form, the system equations can be written as

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{z} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & 0 \\ \boldsymbol{H} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{w} \end{bmatrix} \tag{2.10}$$

then the measurements mean $E[\boldsymbol{z}]$ can be written as

$$E[\boldsymbol{z}] = \boldsymbol{H}E[\boldsymbol{x}] + E[\boldsymbol{w}] = \boldsymbol{H}\bar{\boldsymbol{x}} + \bar{\boldsymbol{w}} \tag{2.11}$$

The covariance matrix for the measurements $\boldsymbol{P}_{\boldsymbol{zz}} = E[(\boldsymbol{z}-\bar{\boldsymbol{z}})(\boldsymbol{z}-\bar{\boldsymbol{z}})^T]$ can be calculated as

$$E\begin{bmatrix} \boldsymbol{x}-\bar{\boldsymbol{x}} \\ \boldsymbol{z}-\bar{\boldsymbol{z}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}-\bar{\boldsymbol{x}} \\ \boldsymbol{z}-\bar{\boldsymbol{z}} \end{bmatrix}^T = \begin{bmatrix} \boldsymbol{I} & 0 \\ \boldsymbol{H} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_{\boldsymbol{xx}} & 0 \\ 0 & \boldsymbol{P}_{\boldsymbol{ww}} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & 0 \\ \boldsymbol{H} & \boldsymbol{I} \end{bmatrix}^T = \begin{bmatrix} \boldsymbol{P}_{\boldsymbol{xx}} & \boldsymbol{P}_{\boldsymbol{xx}}\boldsymbol{H}^T \\ \boldsymbol{H}\boldsymbol{P}_{\boldsymbol{xx}} & \boldsymbol{H}\boldsymbol{P}_{\boldsymbol{xx}}\boldsymbol{H}^T + \boldsymbol{P}_{\boldsymbol{ww}} \end{bmatrix} \tag{2.12}$$

Thus, $\boldsymbol{P}_{\boldsymbol{zz}} = E[(\boldsymbol{z}-\bar{\boldsymbol{z}})(\boldsymbol{z}-\bar{\boldsymbol{z}})^T]$ can be written as

$$\boldsymbol{P}_{\boldsymbol{zz}} = E[(\boldsymbol{z}-\bar{\boldsymbol{z}})(\boldsymbol{z}-\bar{\boldsymbol{z}})^T] = \boldsymbol{H}\boldsymbol{P}_{\boldsymbol{xx}}\boldsymbol{H}^T + \boldsymbol{P}_{\boldsymbol{ww}} \tag{2.13}$$

and the mutual covariance which is $\boldsymbol{P_{xz}} = E[(\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{z} - \bar{\boldsymbol{z}})^T] = [E[(\boldsymbol{x} - \bar{\boldsymbol{x}})^T(\boldsymbol{z} - \bar{\boldsymbol{z}})]]^T = \boldsymbol{P_{xx}}\boldsymbol{H}^T = \boldsymbol{P_{zx}}^T$. We can notice that the measurements covariance $\boldsymbol{P_{zz}}$ is a combination of the state covariance $\boldsymbol{H}\boldsymbol{P_{xx}}\boldsymbol{H}^T$ and the noise covariance matrix $\boldsymbol{P_{ww}}$ under the assumption of an $i.i.d$ measurements. For jointly Gaussian vector then the conditional mean $E[\boldsymbol{x}|\boldsymbol{Z}]$ is given as [25]

$$E[\boldsymbol{x}|\boldsymbol{Z}] := \hat{\boldsymbol{x}} = \bar{\boldsymbol{x}} + \boldsymbol{P_{xz}}\boldsymbol{P_{zz}}^{-1}(\boldsymbol{z} - \bar{\boldsymbol{z}}) \tag{2.14}$$

also the conditional covariance matrix is given as

$$Cov(\boldsymbol{x}|\boldsymbol{Z}) := \boldsymbol{P_{xx|z}} = \boldsymbol{P_{xx}} - \boldsymbol{P_{xz}}\boldsymbol{P_{zz}}^{-1}\boldsymbol{P_{zx}} \tag{2.15}$$

Thus, we have presented the fundamental theory and equations for the MMSE Bayesian estimator. In Chapter 5 we will use the presented equations for investigating the problem of false data injection on linear systems.

Next, we will present the methods for evaluating the performance of an estimator.

2.4 Mean Squared Error - MSE

In this Section, the Mean squared error (MSE) method for evaluating the performance of an estimator presented. First, the case of unknown deterministic parameter (non-Bayesian) MSE presented. The MSE for the multiple parameters is also presented, and the biased and unbiased estimators are explained. Also, the relationship between the MSE and the variance is explained. Next, the MSE for unknown random variable (Bayesian) is also presented.

2.4.1 Non Bayesian Mean Squared Error - NBMSE

In this section, the mean squared error for estimator evaluation is presented. The mean squared error for single parameter non-Bayesian estimator is defined as [25]

$$E(\hat{\theta}(\boldsymbol{Z}) - \theta)^2 = E(\hat{\theta}^2(\boldsymbol{Z}) - 2\ \theta\ \hat{\theta}(\boldsymbol{Z}) + \theta^2) \tag{2.16}$$

where $\boldsymbol{Z} = [z_1, ..., z_N]^T$ is the measurement vector. Simplifying the above equation, one can have

$$E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + E(\theta^2) = E(\hat{\theta}^2) - [E(\hat{\theta})]^2 + [E(\hat{\theta})]^2 - 2\ \theta\ E(\hat{\theta}) + \theta^2 = Var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 \tag{2.17}$$

Thus we notice that the MSE is a composed of the variance and the square of the estimator bias

$$E(\hat{\theta}(\boldsymbol{Z}) - \theta)^2 = Var(\hat{\theta}(\boldsymbol{Z})) + [E(\hat{\theta}(\boldsymbol{Z})) - \theta]^2 \tag{2.18}$$

In this dissertation, we use the mean squared error (MSE) criterion to evaluate the MLE estimator performance and compare it with the Cramer Rao lower bound (CRLB) which we will discuss later in this chapter.

Next, the MSE matrix for an estimation problem with a n-dimensional parameter is provided as follows

$$E[\hat{\boldsymbol{\theta}}(\boldsymbol{Z}) - \boldsymbol{\theta}][\hat{\boldsymbol{\theta}}(\boldsymbol{Z}) - \boldsymbol{\theta}]^T = Cov(\hat{\boldsymbol{\theta}}(\boldsymbol{Z}), \boldsymbol{\theta}) + Bias\ Matrix \tag{2.19}$$

Then let's assume $\boldsymbol{\theta} = [\theta_1\ \theta_2\ ...\ \theta_n]^T$ then $\hat{\boldsymbol{\theta}}(\boldsymbol{Z}) = [\hat{\theta_1}(\boldsymbol{Z})\ \hat{\theta_2}(\boldsymbol{Z})\ ...\ \hat{\theta_n}(\boldsymbol{Z})]^T$. Let

$$\Psi_1 = (\hat{\boldsymbol{\theta}}_1(\boldsymbol{Z}) - \boldsymbol{\theta}_1) \tag{2.20}$$

$$\Psi_2 = (\hat{\boldsymbol{\theta}}_2(\boldsymbol{Z}) - \boldsymbol{\theta}_2) \tag{2.21}$$

$$. \tag{2.22}$$

$$. \tag{2.23}$$

$$\Psi_3 = (\hat{\boldsymbol{\theta}}_3(\boldsymbol{Z}) - \boldsymbol{\theta}_n) \tag{2.24}$$

Thus, the MSE matrix can be written as

$$E[\Psi_1\ \Psi_2\ ...\ \Psi_n][\Psi_1\ \Psi_2\ ...\ \Psi_n]^T = E \begin{bmatrix} \Psi_1^2 & \Psi_1\Psi_2 & ... & \Psi_1\Psi_n \\ \Psi_2\Psi_1 & \Psi_2^2 & ... & \Psi_2\Psi_n \\ : & : & ... & : \\ \Psi_n\Psi_1 & \Psi_n\Psi_2 & ... & \Psi_n^2 \end{bmatrix} \tag{2.25}$$

Note that the MSE matrix is a symmetric matrix. So, the expectation for each in element in the above matrix with considering the symmetric parameters can be found as

$$E(\Psi_1^2) = E(\hat{\theta}_1(\boldsymbol{Z}) - \theta_1(\boldsymbol{Z}))^2 = Var(\hat{\theta}_1(\boldsymbol{Z})) + [E(\hat{\theta}_1(\boldsymbol{Z})) - \theta_1]^2 \tag{2.26}$$

Similarly the $E(\Psi_2^2)$ and $E(\Psi_n^2)$ can be calculated as

$$E(\Psi_2^2) = E(\hat{\theta_2} - \theta_2)^2 = Var(\hat{\theta_2}) + [E(\hat{\theta_2}) - \theta_2]^2 \tag{2.27}$$

$$\vdots \tag{2.28}$$

$$\vdots \tag{2.29}$$

$$E(\Psi_n^2) = E(\hat{\theta_3} - \theta_n)^2 = Var(\hat{\theta_n}) + [E(\hat{\theta_n}) - \theta_n]^2 \tag{2.30}$$

Next, the expectation of the off diagonal elements which represents the covariance and the bias components for MSE matrix can be determined as follows

$$\begin{aligned}E(\Psi_1\Psi_2) = E([\hat{\theta_1}(\boldsymbol{Z}) - \theta_1][\hat{\theta_2}(\boldsymbol{Z}) - \theta_2]) &= E(\hat{\theta_1}(\boldsymbol{Z})\hat{\theta_2}(\boldsymbol{Z}) - \hat{\theta_1}(\boldsymbol{Z})\theta_2 - \theta_1\hat{\theta_2}(\boldsymbol{Z}) + \theta_1\theta_2) \\ &= E(\hat{\theta_1}(\boldsymbol{Z})\hat{\theta_2}(\boldsymbol{Z})) - E(\hat{\theta_1}(\boldsymbol{Z})\theta_2) - E(\theta_1\hat{\theta_2}(\boldsymbol{Z})) + E(\theta_1\theta_2) \\ &= Cov(\hat{\theta_1}(\boldsymbol{Z}), \hat{\theta_2}(\boldsymbol{Z})) - E(\hat{\theta_1}(\boldsymbol{Z}))\theta_2 - \theta_1 E(\hat{\theta_2}(\boldsymbol{Z})) + \theta_1\theta_2 \quad (2.31)\end{aligned}$$

So, it is obvious that the mutual MSE elements is a combination of the covarince of the two elements plus a bias component. Here, the bias components denoted as $\boldsymbol{B}$. Following the same procedure in eq.(2.15), it is possible to calculate the elements of the MSE matrix. Here, the elements will be denoted as $Var(\hat{\boldsymbol{\theta}}(\boldsymbol{Z}))$, $Cov(\hat{\boldsymbol{\theta}}(\boldsymbol{Z}))$, and bias as $\boldsymbol{B}$. Following is the elements of the covariance and the bias matrix.

$$c_{11} = Var(\hat{\theta_1}(\boldsymbol{Z}))$$

$$c_{22} = Var(\hat{\theta_2}(\boldsymbol{Z}))$$

$$\vdots$$

$$\vdots$$

$$c_{nn} = Var(\hat{\theta_n}(\boldsymbol{Z}))$$

$$c_{12} = c_{21} = Cov(\hat{\theta_1}(\boldsymbol{Z}), \hat{\theta_2}(\boldsymbol{Z}))$$

$$\vdots$$

$$\vdots$$

$$c_{1n} = c_{n1} = Cov(\hat{\theta_1}(\boldsymbol{Z}), \hat{\theta_n}(\boldsymbol{Z}))$$

$$\vdots$$

$$\vdots$$

$$c_{2n} = c_{n2} = Cov(\hat{\theta_2}(\boldsymbol{Z}), \hat{\theta_n}(\boldsymbol{Z})) \quad (2.32)$$

Thus, the covariance matrix elements of the MSE found. Next, the bias $\boldsymbol{B}$ matrix elements of the MSE matrix can be calculated as

$$B_{11} = [E(\hat{\theta_1}) - \theta_1]^2$$

$$B_{22} = [E(\hat{\theta_2}) - \theta_2]^2$$

$$:$$

$$:$$

$$B_{nn} = [E(\hat{\theta_n}) - \theta_n]^2$$

$$B_{12} = B_{21} = [\theta_1 \ \theta_2 - E(\hat{\theta_1})\theta_2 - \theta_1 E(\hat{\theta_2})]$$

$$:$$

$$:$$

$$B_{1n} = B_{n1} = [\theta_1 \ \theta_n - E(\hat{\theta_1})\theta_n - \theta_1 E(\hat{\theta_n})]$$

$$:$$

$$:$$

$$B_{2n} = B_{n2} = [\theta_2 \ \theta_n - E(\hat{\theta_2})\theta_n - \theta_2 E(\hat{\theta_n})] \quad (2.33)$$

Then, the MSE matrix can be written as follows

$$E[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}][\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]^T = \begin{bmatrix} c_{11} & c_{12} & ... & c_{1n} \\ c_{21} & c_{22} & ... & c_{2n} \\ : & : & ... & : \\ c_{n1} & c_{n2} & ... & c_{nn} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & ... & B_{1n} \\ B_{21} & B_{22} & ... & B_{2n} \\ : & : & ... & : \\ B_{n1} & B_{n2} & ... & B_{nn} \end{bmatrix} \quad (2.34)$$

which represents the covariance matrix and the bias of the estimator. For unbiased estimators then $\boldsymbol{B} = 0$. In other words, for unbiased estimator the MSE is equal to the covariance of the estimator.

In this dissertation, the MSE method will be used for evaluating the performance of the proposed MLE estimators. Next, the MSE for the Bayesian estimator will be presented.

2.4.2 Bayesian Mean Square Error - BMSE

For the unknown random parameter estimator also known as Bayesian estimator the MSE is defined as [25]

$$MSE[\hat{\theta}(\boldsymbol{Z})] := E[(\hat{\theta}(\boldsymbol{Z}) - \theta)^2] \tag{2.35}$$

where $\hat{\theta}(\boldsymbol{Z})$ is the Bayesian estimate based on the observed sensor measurements vector $\boldsymbol{Z} = [z_1, ..., z_N]^T$, and θ is the true value of the random parameter which has a prior pdf $p(\theta)$. The expectation in the above equation will be calculated over the joint pdf of $\boldsymbol{Z}$ measurements and the prior pdf of random parameter $p(\theta)$.

For the MMSE estimator, the relationship between the MSE and the conditional variance can be explained as follows. The MSE for the MMSE estimator conditioned on a certain set $\boldsymbol{Z}$ measurements can be written as [25]

$$E[(\hat{\theta}^{MMSE}(\boldsymbol{Z}) - \theta)^2|\boldsymbol{Z}] = E[(\theta - E(\theta|\boldsymbol{Z}))^2|\boldsymbol{Z}] = Var(\theta|\boldsymbol{Z}) \tag{2.36}$$

which is the conditional variance $Var(\theta|\boldsymbol{Z})$. Since the expectations are calculated with respect to $f(\theta|\boldsymbol{Z})$ for the equations above, then averaging over $\boldsymbol{Z}$ yields

$$E[Var(\theta|\boldsymbol{Z})] = E[(\theta - E[\theta|\boldsymbol{Z}])^2] \tag{2.37}$$

which also equal to the unconditional MSE given in eq.(1st) for the MMSE estimator. Next, the non - Bayesian and the Bayesian Fisher information matrix (FIM) will be presented.

2.5 Fisher Information Matrix

2.5.1 Non Bayesian Fisher Information Matrix - NBFIM

For non - Bayesian unknown deterministic vector parameters $\boldsymbol{\theta}$, with an estimate based of the random sensors measurements $\hat{\boldsymbol{\theta}}(Z)$. The Fisher information is defined as :

$$I(\boldsymbol{\theta}) := -E\left[\frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right] = E\left[\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2\right] \tag{2.38}$$

where $f(\boldsymbol{Z}|\boldsymbol{\theta})$ is the likelihood function, it is clear that the Fisher information is calculated at the true values of $\boldsymbol{\theta}$. Let $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \theta_3]^T$ is the true parameters vector, and $\nabla = [\frac{\partial f}{\partial \theta_1} \ \frac{\partial f}{\partial \theta_2} \ \frac{\partial f}{\partial \theta_3}]$ is the gradient vector . The Hessian matrix for the FIM can be described as

$$\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \log f(\boldsymbol{Z}|\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1^2} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1 \theta_2} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1 \theta_3} \\ \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_2 \theta_1} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_2^2} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_2 \theta_3} \\ \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_3 \theta_1} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_3 \theta_2} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_3^2} \end{bmatrix} \tag{2.39}$$

The equality of the two terms in eq.(2.19) can be proven for single element as follows

$$-E\left[\frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1^2}\right] = -E\left[\frac{1}{f(\boldsymbol{Z}|\boldsymbol{\theta})} \ \frac{\partial^2 f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1^2} \ + \ \frac{-1}{f(\boldsymbol{Z}|\boldsymbol{\theta})^2}\left(\frac{\partial f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\right)^2\right] \tag{2.40}$$

the expectation for the 1st term of eq.(2.21) can be calculated as

$$\begin{aligned} -E\left[\frac{1}{f(\boldsymbol{Z}|\boldsymbol{\theta})}\ \frac{\partial^2 f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1^2}\right] &= -\int_{-\infty}^{\infty}\frac{1}{f(\boldsymbol{Z}|\boldsymbol{\theta})}\ \frac{\partial^2 f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1^2}\ f(\boldsymbol{Z}|\boldsymbol{\theta})\ d_z \\ &= -\int_{-\infty}^{\infty}\frac{\partial^2 f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1^2}\ d_z \end{aligned} \tag{2.41}$$

using the differentiation and integral interchangeable property and since $\int_{-\infty}^{\infty} f(\boldsymbol{Z}|\boldsymbol{\theta})\, d_z = 1$ we get

$$-\frac{\partial^2}{\partial\theta_1^2}\int_{-\infty}^{\infty} f(\boldsymbol{Z}|\boldsymbol{\theta})\ d_z = -\frac{\partial^2}{\partial\theta_1^2}[1] = 0 \tag{2.42}$$

sub. eq.(2.23) in eq.(2.21) and using $\frac{\partial f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1} = f(\boldsymbol{Z}|\boldsymbol{\theta})\ \frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\theta_1}$, yields

$$-E\left[\frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial{\theta_1}^2}\right] = E\left[\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\theta_1}\right)^2\right] \tag{2.43}$$

Thus, the equality of eq.(2.19) proved. Now, eq.(2.19) can be also shown valid for the off - diagonal elements of the FIM matrix as

$$\begin{aligned} -E\left[\frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2}\right] = -E\Bigg[&\frac{1}{f(\boldsymbol{Z}|\boldsymbol{\theta})}\ \frac{\partial^2 f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2} \\ &+ \frac{-1}{f(\boldsymbol{Z}|\boldsymbol{\theta})}\frac{1}{f(\boldsymbol{Z}|\boldsymbol{\theta})}\left(\frac{\partial f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1}\right)\left(\frac{\partial f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2}\right)\Bigg] \end{aligned} \tag{2.44}$$

following the same procedures for eq.(2.23), the first term of eq.(2.25) is shown equal to zero. And using $\frac{\partial f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1} = \frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1} f(\boldsymbol{Z}|\boldsymbol{\theta})$ and $\frac{\partial f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2} = \frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2} f(\boldsymbol{Z}|\boldsymbol{\theta})$, yields

$$-E\left[\frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2}\right] = E\left[\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1}\right)\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2}\right)\right] \tag{2.45}$$

now we can write the FIM matrix as

$$-E[\nabla\nabla^T \log f(\boldsymbol{Z}|\boldsymbol{\theta})] = -E\begin{bmatrix} \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1^2} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1\theta_2} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1\theta_3} \\ \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2\theta_1} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2^2} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2\theta_3} \\ \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_3\theta_1} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_3\theta_2} & \frac{\partial^2 \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_3^2} \end{bmatrix}$$

$$= E\begin{bmatrix} \left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\theta_1}\right)^2 & \left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1}\right)\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2}\right) & \left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1}\right)\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_3}\right) \\ \left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2}\right)\left(\frac{\partial \log f(\boldsymbol{Z},\theta_1)}{\partial\theta_1}\right) & \left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\theta_2}\right)^2 & \left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2}\right)\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_3}\right) \\ \left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_3}\right)\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_1}\right) & \left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_3}\right)\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial\theta_2}\right) & \left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\theta_3}\right)^2 \end{bmatrix} \tag{2.46}$$

Thus the non - Bayesian NB-FIM calculated . In next section, The Bayesian (NB-CRLB) will be presented

2.5.2 Non - Bayesian Cramer Rao Lower Bound (NB - CRLB)

For unbiased deterministic parameters, it is known form section 2 that the mean of the estimation error is equal to 0. According to that assumption the CRLB is defined as the minimum limit that the MSE or the Variance of the estimator can achieve. The estimator considered efficient if it achieves the CRLB bound. The CRLB for unbiased estimator of a vector of unknown deterministic parameters based on N observed sensor measurements $\boldsymbol{Z} = [z_1, ..., z_N]^T$ is defined as

$$E[\ [\hat{\boldsymbol{\theta}}(\boldsymbol{Z}) - \boldsymbol{\theta_0}][\hat{\boldsymbol{\theta}}(\boldsymbol{Z}) - \boldsymbol{\theta_0}]^T\ ] \geq \boldsymbol{J}^{-1} \tag{2.47}$$

where $\hat{\boldsymbol{\theta}}(\boldsymbol{Z})$ is the estimated value for vector parameters, and $\boldsymbol{\theta}$ is the true value of the parameters, and $\boldsymbol{J}$ is the non - Bayesian fisher information matrix (NB-FIM) calculated in previous section.

The proof of the non Bayesian (NB - CRLB) for single element of the estimated vector $\hat{\boldsymbol{\theta}}(\boldsymbol{Z})$ which is the element $\hat{\theta_1}$ can be done by using the assumption that the mean for the estimation error under unbiased estimator is equal to 0 as

$$E[\hat{\theta_1}(z) - \theta_1] = \int_{-\infty}^{\infty} [\hat{\theta_1}(\boldsymbol{Z}) - \theta_1]\ f(\boldsymbol{Z}|\boldsymbol{\theta})\ dz = 0 \tag{2.48}$$

deriving eq.(2.29) with respect to θ_1, yields

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \int_{-\infty}^{\infty} [\hat{\theta_1}(\boldsymbol{Z}) - \theta_1]\ f(\boldsymbol{Z}|\boldsymbol{\theta})\ dz &= \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta_1}\ \{[\hat{\theta_1}(\boldsymbol{Z}) - \theta_1]\ f(\boldsymbol{Z}|\boldsymbol{\theta})\}\ dz \\ &= \int_{-\infty}^{\infty} [\hat{\theta_1}(\boldsymbol{Z}) - \theta_1]\ \frac{\partial f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\ dz + \int_{-\infty}^{\infty} [-1]\ f(\boldsymbol{Z}|\boldsymbol{\theta})\ dz \\ &= \int_{-\infty}^{\infty} [\hat{\theta_1}(\boldsymbol{Z}) - \theta_1]\ \frac{\partial f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\ dz - 1 = 0 \end{aligned} \tag{2.49}$$

using $\frac{\partial f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1} = \frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\ f(\boldsymbol{Z}|\boldsymbol{\theta})$, we get

$$\int_{-\infty}^{\infty} [\hat{\theta_1}(\boldsymbol{Z}) - \theta_1]\ f(\boldsymbol{Z}|\boldsymbol{\theta})\ \left[\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\right]\ dz = 1 \tag{2.50}$$

eq.(2.31) can be rewritten as

$$\int_{-\infty}^{\infty} [\hat{\theta_1}(\boldsymbol{Z}) - \theta_1] \ \sqrt{f(\boldsymbol{Z}|\boldsymbol{\theta})} \ \sqrt{f(\boldsymbol{Z}|\boldsymbol{\theta})} \ \left[\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\right] \ dz = 1 \tag{2.51}$$

square both sides of the eq.(2.32) as

$$\begin{aligned} &\left[\int_{-\infty}^{\infty} [\hat{\theta_1}(\boldsymbol{Z}) - \theta_1] \ \sqrt{f(\boldsymbol{Z}|\boldsymbol{\theta})} \ \sqrt{f(\boldsymbol{Z}|\boldsymbol{\theta})} \ \left[\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\right] \ dz\right]^2 \\ &= \int_{-\infty}^{\infty} [\hat{\theta_1}(\boldsymbol{Z}) - \theta_1]^2 \ f(\boldsymbol{Z}|\boldsymbol{\theta}) \ dz \int_{-\infty}^{\infty} \left[\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\right]^2 \ f(\boldsymbol{Z}|\boldsymbol{\theta}) \ dz = 1 \end{aligned} \tag{2.52}$$

using Cauchy - Schwartz inequality $[\int f. \int g \geq (\int f \ g)^2]$, yields

$$\int_{-\infty}^{\infty} [\hat{\theta_1}(z) - \theta_1]^2 \ f(\boldsymbol{Z}|\boldsymbol{\theta}) \ dz \int_{-\infty}^{\infty} \left[\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\right]^2 \ f(\boldsymbol{Z}|\boldsymbol{\theta}) \ dz \geq 1 \tag{2.53}$$

it is obvious from eq.(2.34) that $\int_{-\infty}^{\infty} [\hat{\theta_1}(\boldsymbol{Z}) - \theta_1]^2 \ f(\boldsymbol{Z}|\boldsymbol{\theta}) \ dz = E[(\hat{\theta_1}(\boldsymbol{Z}) - \theta_1)^2]$ and $\int_{-\infty}^{\infty} \left[\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\right]^2 \ f(\boldsymbol{Z}|\boldsymbol{\theta}) \ dz = E\left[\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\right)^2\right] = FIM$, then it is clea that the CRLB is

$$E[(\hat{\theta_1}(\boldsymbol{Z}) - \theta_1)^2] \geq E\left[\left(\frac{\partial \log f(\boldsymbol{Z}|\boldsymbol{\theta})}{\partial \theta_1}\right)^2\right]^{-1} \tag{2.54}$$

which proves the non Bayesian (NB - CRLB). Next, the Bayesian Fisher information matrix (B - FIM) and the Bayesian CRLB (B - CRLB) also known as Postrior CRLB (PCRLB) will be presented.

2.5.3 Bayesian Fisher Information Matrix (B-FIM)

Let $\boldsymbol{\theta}$ be a random vector with prior pdf $f(\boldsymbol{\theta})$, then the Baysian FIM (B - FIM) is given as

$$\boldsymbol{J_B} = E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log f(\boldsymbol{Z}, \boldsymbol{\theta})] \tag{2.55}$$

where $\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} := \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T$, and $\nabla_{\boldsymbol{\theta}} = [\frac{\partial}{\partial \theta_1} \; \frac{\partial}{\partial \theta_2} \; \frac{\partial}{\partial \theta_3}]^T$, $\boldsymbol{Z} = [z_1, ..., z_N]^T$ is the sensors measurements vector, and

$$f(\boldsymbol{Z}, \boldsymbol{\theta}) = f(\boldsymbol{Z}|\boldsymbol{\theta}) \; f(\boldsymbol{\theta}) \tag{2.56}$$

where $f(\boldsymbol{\theta})$ is the prior information probability of the random variable, and $f(\boldsymbol{Z}|\boldsymbol{\theta})$ is the likelihood function of the system.

Thus, it is obvious that the above FIM is a combination of two additive FIMs one with respect to the likelihood averaged over the prior probability $f(\boldsymbol{\theta})$ and the other is with respect to the prior probability information as [26, 27, 28]

$$\boldsymbol{J_B} = E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log f(\boldsymbol{Z}|\boldsymbol{\theta})] + E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log f(\boldsymbol{\theta})] = \boldsymbol{J_d} + \boldsymbol{J_p} \tag{2.57}$$

where

$$\boldsymbol{J_p} := E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log f(\boldsymbol{\theta})] \tag{2.58}$$

is the apriori information, and

$$\boldsymbol{J_d} := E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log f(\boldsymbol{Z}|\boldsymbol{\theta})] \tag{2.59}$$

where $\boldsymbol{J_d}$ can be also considered as the non - Bayesian (NB - FIM) averaged of prior pdf as

$$\boldsymbol{J_d} := E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log f(\boldsymbol{D}|\boldsymbol{\theta})] = \int_{\boldsymbol{\theta}} \boldsymbol{J} f(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{2.60}$$

where $\boldsymbol{J}$ is the standard FIM derived in non-Bayesian case. Next, the Bayesian CRLB (B - CRLB) also called a posterior CRLB (PCRLB) will be presented.

2.5.4 Bayesian CRLB (Poterior CRLB)

For Bayesian parameter, the Bayesian Fisher information matrix (B - FIM) used to calculate the (BCRLB) also claaed (PCRLB) . The Bayesian CRLB is given as

$$E[(\hat{\boldsymbol{\theta}}(\boldsymbol{Z}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\boldsymbol{Z}) - \boldsymbol{\theta})^T] \geq \boldsymbol{J_B}^{-1} \tag{2.61}$$

where $\boldsymbol{J_B}$ is the Bayesian Fisher information matrix (B - FIM) calculated in the previous section.

2.6 Consistency and Efficiency of an estimator

For an unknown deterministic non - Bayesian case, an estimator is called consistent if it is eventually converges to the true value of the estimate and it might be determined by using some stochastic criteia like the mean squared criteria to determine the consistency as in [25]

$$\lim_{n\to\infty} E[(\hat{\theta}(n, \boldsymbol{Z}^n) - \theta_0)^2] = 0 \tag{2.62}$$

where $\hat{\theta}(n, \boldsymbol{Z}^n)$ is the estimate , θ_0 is the true value of the deterministic parameter, $\boldsymbol{Z} = [z_1, ..., z_N]^T$ is the sensor measurements vector, and the expectation will be calculated with respect to the likelihood function $f(\boldsymbol{Z^n}|\theta_0)$.

For the the random unknown parameters also called Bayesian parameters, the estimator will be called consistent if the estimated value eventually converges to the true value of the random parameter θ, and it might also be measured using the MSE criteria as in [25]

$$\lim_{n\to\infty} E[(\hat{\theta}(n, \boldsymbol{Z}^n) - \theta)^2] = 0 \tag{2.63}$$

where the expectation in the above equation will be calculated with respect to the joint pdf $f(\boldsymbol{Z^n}, \theta) = f(\boldsymbol{Z^n}|\theta)f(\theta)$.

For the efficiency of the estimator, an estimator is called efficient if its variance attains to the Cramer Rao Lower Bound (CRLB).

CHAPTER 3

MAXIMUM LIKELIHOOD SOURCE LOCATION ESTIMATOR UNDER FALSE INFORMATION INJECTION ATTACKS WITH KNOWN ATTACK POWER AND PROBABILITY

In this chapter, we first formulate the localization problem in a sensor network using QRSS data which are corrupted by noise injected by an adversary with known attack parameters. We introduce the GM model and derive the pmf for the quantized sensor data under the GMM false information injection attack. After that, we derive the maximum likelihood estimator for the proposed source localization problem, and finally we derive the CRLB for the proposed estimator for performance evaluation.

3.1 Mathematical Model

The proposed method can handle any sensor deployment. For simplicity, we consider in this dissertation the uniform sensor deployment as shown in Fig. 1. The target signal intensity attenuation is inversely proportional to the distance from the target with some exponent. We assume an isotropic signal attenuation model as in [6]:

$$a_i^2 = \frac{G_i P_0^{'}}{(\frac{d_i}{d_0})^n} \tag{3.1}$$

where a_i is the signal amplitude at the ith sensor, $P_0^{'}$ is the radiated power by the target at a reference distance d_0, G_i is the gain of the ith sensor, and d_i^n is the Euclidean distance between the target and the ith sensor as given below:

$$d_i = \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2} \tag{3.2}$$

The combinations (x_i, y_i) and (x_t, y_t) are the coordinates of the ith sensor and the target location respectively, and n is the attenuation exponent. For simplicity, we assume that $G_i = G, \quad \forall i$, and $P_0 = GP_0^{'}$. We also assume that $d_0 = 1m$. Then (3.1) can be rewritten as:

$$a_i^2 = \frac{P_0}{(d_i)^n} \tag{3.3}$$

We assume that the presence of a target in the sensor grid with N sensors with known locations as shown in Fig.1, has been detected correctly, and the distance between the target and any sensor is at least d_0. The isotropic power attenuation model has been used widely for modeling signal attenuation for acoustic and electromagnetic waves propagation [6, 29, 30, 31, 32].

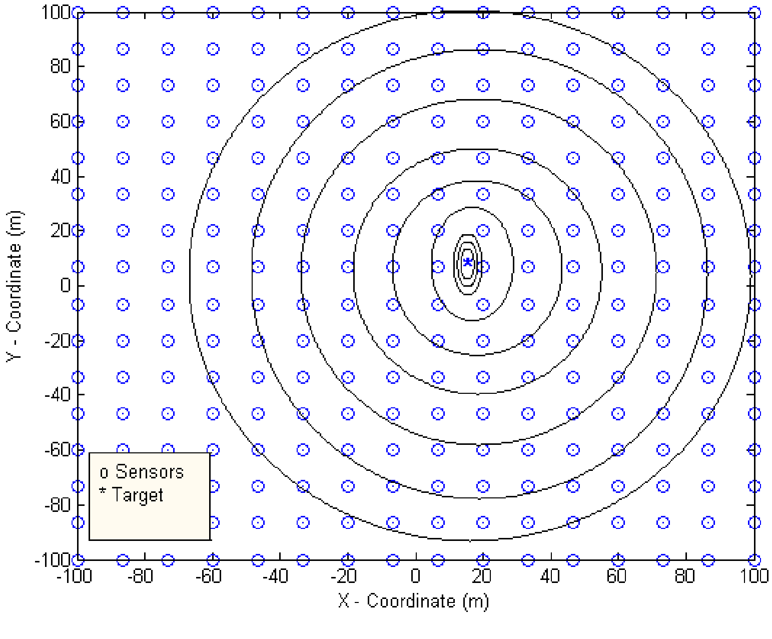


Fig. 1. The sensor deployment in the ROI.

As illustrated in Fig.2, we follow the same strategy for collecting sensors readings which has been developed in [6]. The amplitude at the ith sensor a_i will be attacked and corrupted by noise that follows a GM model, which can be considered as an impulsive noise or a biased noise injected by an attacker. The signal at each sensor is modeled by:

$$r_i = a_i + b_i \tag{3.4}$$

where r_i is the received signal by the ith sensor, a_i is the amplitude of the uncorrupted signal, and b_i is an independent and identically distributed (i.i.d.) noise that follows GM distribution, which affects all the sensors independently.

$$b_i \sim \sum_{k=1}^{K} w_{i,k} \mathcal{N}(0, \sigma_k^2) \tag{3.5}$$

where $w_{i,k}$ is the weight of the kth Gaussian component with zero mean and variance σ_k^2, and K is the number of Gaussian components, which in our proposed work is set as 2. We denote p_a as the attack probability, which is the same for all the sensors, and $p_a = w_{i,2}$ and $w_{i,1} = (1 - w_{i,2})$. We also assume that $\sigma_2^2 >> \sigma_1^2$, meaning that the injected noise is much stronger than the sensor measurement noise.

After having the signal model we are now ready to derive the likelihood for the estimation problem.

3.2 Maximum Likelihood Estimator Based on Quantized Data

In this section, we derive the likelihood function based on the quantized multi-bit data by using the received signal model Q-RSS-GM, presented earlier in this chapter. First, let us denote the desired parameter to be estimated is $\boldsymbol{\theta} = [P_0 \; x_t \; y_t]^T$, which consists of the target signal strength and its location coordinates. We assume

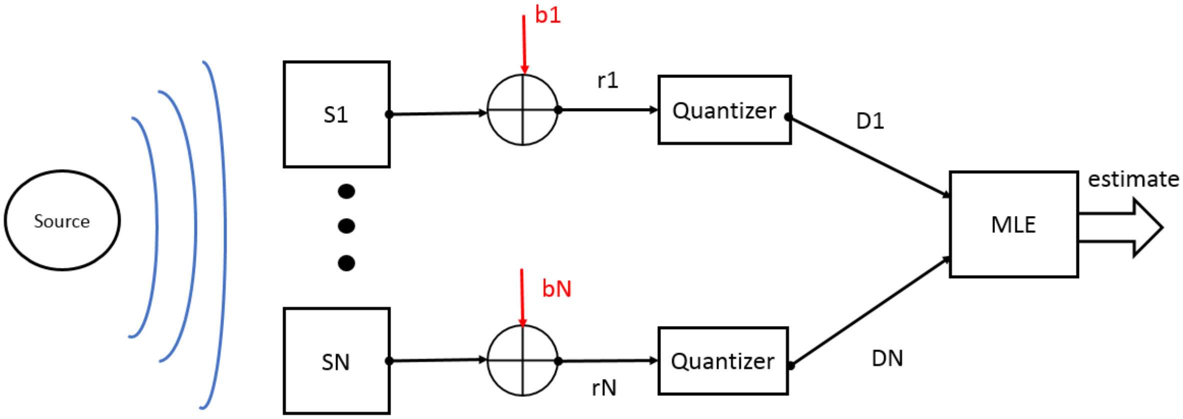


Fig. 2. The system diagram.

that each sensor quantizes its received sensor measurements into M-bit data according to a certain threshold and sends them to the fusion center, which we denote as $\mathbf{D} = [D_1, \cdots, D_N]^T$, where $D_i \in \{0, \cdots, 2^M - 1\}$. For simplicity, we denote $L = 2^M$. The quantization thresholds set is $\boldsymbol{\eta}_i = [\eta_{i0}, \eta_{i1}, ..., \eta_{iL}]^T$, where $\eta_{i0} = -\infty$ and $\eta_{iL} = \infty$, and the quantization process can be performed as follows:

$$D_i = \begin{cases} 0 & -\infty \leq r_i < \eta_{i1} \\ 1 & \eta_{i1} \leq r_i < \eta_{i2} \\ : & : \\ : & : \\ L-1 & \eta_{L-1} \leq r_i < \infty \end{cases} \tag{3.6}$$

According to the Gaussian Mixture model, the probability that D_i takes a specific value l can be derived as:

$$p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta}) = \sum_{k=1}^{K} w_{i,k} \left[Q\left(\frac{\eta_{il} - a_i}{\sigma_k}\right) - Q\left(\frac{\eta_{il+1} - a_i}{\sigma_k}\right) \right] \tag{3.7}$$

where $l \in \{0, \cdots, L-1\}$, and $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt$ is the complementary cumulative distribution function of the standard Gaussian distribution with zero mean and unit variance.

Now the joint probability of the sensor data can be found as follows

$$p\left(\mathbf{D}|\boldsymbol{\theta}\right) = \prod_{i=1}^{N} \prod_{l=0}^{L-1} p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})^{\delta(D_i - l)} \tag{3.8}$$

where $\delta(.)$ is the Kronecker delta function. The log-likelihood function is therefore

$$\log p(\mathbf{D}|\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \delta(D_i - l) \log p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta}) \tag{3.9}$$

We tries to maximize the log-likelihood function over $\boldsymbol{\theta}$. So the ML estimator is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log p(\mathbf{D}|\boldsymbol{\theta}) \tag{3.10}$$

Next, we derive the CRLB for this estimation problem.

3.3 Cramer Rao Lower Bound (CRLB) for the proposed MLE

In this section we will derive the CRLB for the proposed MLE-QRSS-GMM for performance evaluation.

Theorem 1 : For an unbiased estimator $\widehat{\boldsymbol{\theta}}(\mathbf{D})$, the CRLB is given by

$$E\big\{[\widehat{\boldsymbol{\theta}}(\mathbf{D}) - \boldsymbol{\theta}][\widehat{\boldsymbol{\theta}}(\mathbf{D}) - \boldsymbol{\theta}]^T\big\} \geq \mathbf{J}^{-1} \tag{3.11}$$

where $\mathbf{J}$ is the 3×3 Fisher information matrix (FIM).

Proof: The following is the proof of *Theorem* 1, which provides the FIM for the localization problem subjected to the GMM attack assumption.

$$\mathbf{J} = -E[\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \log p(\mathbf{D}|\boldsymbol{\theta})] \tag{3.12}$$

where $\nabla_{\boldsymbol{\theta}}$ is the gradient vector:

$$\nabla_{\boldsymbol{\theta}} = \left[\frac{\partial}{\partial P_0} \; \frac{\partial}{\partial x_t} \; \frac{\partial}{\partial y_t} \right]^T \tag{3.13}$$

the Fisher information matrix $\mathbf{J}$ can be described as

$$\mathbf{J} = -E[\nabla_\theta \nabla_\theta^T \log p(\mathbf{D}|\boldsymbol{\theta})] = -\mathbf{E} \begin{bmatrix} \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial x_t} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial y_t} \\ \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial P_0} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t^2} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial y_t} \\ \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t \partial P_0} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t \partial x_t} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t^2} \end{bmatrix} = \begin{bmatrix} j_{11} & j_{12} & j_{13} \\ j_{21} & j_{22} & j_{23} \\ j_{31} & j_{32} & j_{33} \end{bmatrix} \tag{3.14}$$

from (3.9) we have the log - likelihood function as

$$\log p(\boldsymbol{D}|\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \delta(D_i - l) \log p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta}) \tag{3.15}$$

then, each element of the above matrix will be calculated as follows ,

the 1st derivative for the matrix element (1,1)

$$\frac{\partial \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0} = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial P_0} \right] \tag{3.16}$$

and the 2nd derivative will be

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2} = \sum_{i=1}^{N}\sum_{l=0}^{L-1} \frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})]^2}\left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial P_0}\right]^2 + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2}\right] \tag{3.17}$$

Next, we calculate the (1,1) element of $\mathbf{J}$.

$$\begin{aligned}\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2} = \sum_{i=1}^{N}\sum_{l=0}^{L-1} -\frac{\delta(D_i - l)}{p_{il}^2(\boldsymbol{\eta}_i, \boldsymbol{\theta})}\left[\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial P_0}\right]^2 + \\ \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta}_i, \theta)}\left[\frac{\partial^2 p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial P_0^2}\right]\end{aligned} \tag{3.18}$$

and

$$j_{11} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2}\right] = \sum_{i=1}^{N}\sum_{l=0}^{L-1} \frac{1}{p_{il}}\left[\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial P_0}\right]^2 \tag{3.19}$$

by using the identity $E[\delta(D_i - l)] = p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})$. Next we find $\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial P_0}$ as

$$\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial P_0} = \sum_{k=1}^{K} w_{i,k} \frac{\gamma_{i,l,k}}{2\sqrt{2\pi}\sigma_k a_i d_i^n} \tag{3.20}$$

where

$$\gamma_{i,l,k} = \left[e^{-\frac{(\eta_{il}-a_i)^2}{2\sigma_k^2}} - e^{-\frac{(\eta_{il+1}-a_i)^2}{2\sigma_k^2}}\right] \tag{3.21}$$

Then

$$\left[\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial P_0}\right]^2 = \left[\sum_{k=1}^{K} w_{i,k} \frac{\gamma_{i,l,k}}{2\sqrt{2\pi}\sigma_k a_i d_i^n}\right]^2 \tag{3.22}$$

By substituting (3.22) in (3.19), we get

$$j_{11} = -E\left[\tfrac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2}\right] = \sum_{i=1}^{N} \beta_i a_i^{-2} d_i^{-2n} \tag{3.23}$$

where

$$\beta_i = \frac{1}{8\pi} \sum_{l=0}^{L-1} \frac{1}{p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})} \left[\sum_{k=1}^{K} w_{i,k} \frac{\gamma_{i,l,k}}{\sigma_k}\right]^2 \tag{3.24}$$

Similarly, we can find the j_{12} which is equal to j_{21} elements since the FIM matrix is symmetric

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial x_t} = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})]^2} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial P_0}\right] \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial x_t}\right] + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial x_t}\right] \tag{3.25}$$

then, $j_{12} = j_{21}$:

$$j_{12} = j_{21} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial x_t}\right] = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{1}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial P_0}\right] \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial x_t}\right] \tag{3.26}$$

Since we know the $\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial P_0}$ from (3.20). Next we find $\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial x_t}$ as

$$\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial x_t} = \sum_{k=1}^{K} w_{i,k} \frac{n a_i d_i^{-2} \gamma_{i,l,k}}{2\sqrt{2\pi}\sigma_k} \ (x_i - x_t) \tag{3.27}$$

Then

$$\left[\frac{\partial p_{il}(\boldsymbol{\eta}_i,\boldsymbol{\theta})}{\partial P_0}\right]\left[\frac{\partial p_{il}(\boldsymbol{\eta}_i,\boldsymbol{\theta})}{\partial x_t}\right] = \left[\sum_{k=1}^{K} w_{i,k}\frac{\gamma_{i,l,k}}{2\sqrt{2\pi}\sigma_k a_i d_i^n}\right]\left[\sum_{k=1}^{K} w_{i,k}\frac{n a_i d_i^{-2}\gamma_{i,l,k}}{2\sqrt{2\pi}\sigma_k}\ (x_i - x_t)\right] \tag{3.28}$$

By substituting (3.28) in (3.26), we get

$$j_{12} = j_{21} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2}\right] = n\sum_{i=1}^{N}\beta_i d_i^{-(n+2)}(x_i - x_t) \tag{3.29}$$

Next, the j_{22} element of the FIM matrix can be found as

$$j_{22} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\theta)}{\partial x_t^2}\right] = \sum_{i=1}^{N}\sum_{l=0}^{L-1}\frac{1}{p_{il}(\eta_i,\theta)}\left[\frac{\partial p_{il}(\eta_i,\theta)}{\partial x_t}\right]^2 \tag{3.30}$$

Thus,

$$j_{22} = n^2\sum_{i=1}^{N}\beta_i a_i^2 d_i^{-4}(x_i - x_t)^2 \tag{3.31}$$

Next, Calculating the $j_{13} = j_{31}$ elements

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial y_t} = \sum_{i=1}^{N}\sum_{l=0}^{L-1}\frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})]^2}\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial P_0}\right]\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial y_t}\right] + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial y_t}\right] \tag{3.32}$$

then $j_{13} = j_{31}$ as:

$$j_{13} = j_{31} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial y_t}\right] = \sum_{i=1}^{N}\sum_{l=0}^{L-1}\frac{1}{p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial P_0}\right]\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial y_t}\right] \tag{3.33}$$

where

$$\frac{\partial p_{il}(\boldsymbol{\eta}_i,\boldsymbol{\theta})}{\partial y_t} = \sum_{k=1}^{K} w_{i,k}\frac{n a_i d_i^{-2}\gamma_{i,l,k}}{2\sqrt{2\pi}\sigma_k}\ (y_i - y_t) \tag{3.34}$$

Then, $j_{13} = j_{31}$ elements

$$j_{13} = j_{31} = n\sum_{i=1}^{N}\beta_i d_i^{-(n+2)}(y_i - y_t) \tag{3.35}$$

Now, calculating the elements $j_{23} = j_{32}$:

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial y_t} = \sum_{i=1}^{N}\sum_{l=0}^{L-1}\frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})]^2}\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial x_t}\right]\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial y_t}\right] + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial y_t}\right] \tag{3.36}$$

then $j_{23} = j_{32}$

$$j_{23} = j_{32} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial y_t}\right] = \sum_{i=1}^{N}\sum_{l=0}^{L-1}\frac{1}{p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial x_t}\right]\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial y_t}\right] \tag{3.37}$$

Thus, $j_{13} = j_{31}$ elements

$$j_{23} = j_{32} = n^2 \sum_{i=1}^{N} \beta_i a_i^2 d_i^{-4} (x_i - x_t)(y_i - y_t) \tag{3.38}$$

finally, calculating the j_{33} element

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t^2} = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})]^2} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial y_t} \right]^2 + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t^2} \right] \tag{3.39}$$

then j_{33} :

$$j_{33} = -E \left[\frac{\partial^2 \log p(\mathbf{D}|\theta)}{\partial y_t^2} \right] = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{1}{p_{il}(\eta_i, \theta)} \left[\frac{\partial p_{il}(\eta_i, \theta)}{\partial y_t} \right]^2 \tag{3.40}$$

Thus,

$$j_{33} = n^2 \sum_{i=1}^{N} \beta_i a_i^2 d_i^{-4} (y_i - y_t)^2 \tag{3.41}$$

Thus the elements of the FIM are provided as

$$\begin{aligned}
j_{11} &= \sum_{i=1}^{N} \beta_i d_i^{-2n} a_i^{-2} \\
j_{12} = j_{21} &= n \sum_{i=1}^{N} \beta_i d_i^{-(n+2)} (x_i - x_t) \\
j_{13} = j_{31} &= n \sum_{i=1}^{N} \beta_i d_i^{-(n+2)} (y_i - y_t) \\
j_{22} &= n^2 \sum_{i=1}^{N} \beta_i a_i^2 d_i^{-4} (x_i - x_t)^2 \\
j_{23} = j_{32} &= n^2 \sum_{i=1}^{N} \beta_i a_i^2 d_i^{-4} (x_i - x_t)(y_i - y_t) \\
j_{33} &= n^2 \sum_{i=1}^{N} \beta_i a_i^2 d_i^{-4} (y_i - y_t)^2
\end{aligned} \tag{3.42}$$

where

$$\beta_i = \frac{1}{8\pi} \sum_{l=0}^{L-1} \frac{1}{p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})} \left[\sum_{k=1}^{K} w_{i,k} \frac{\gamma_{i,l,k}}{\sigma_k} \right]^2 \tag{3.43}$$

and

$$\gamma_{i,l,k} = \left[e^{-\frac{(\eta_{il} - a_i)^2}{2\sigma_k^2}} - e^{-\frac{(\eta_{il+1} - a_i)^2}{2\sigma_k^2}} \right] \tag{3.44}$$

Once the FIM is obtained, the CRLB matrix can be readily calculated, by taking the inverse of the FIM.

3.4 Simulation Results

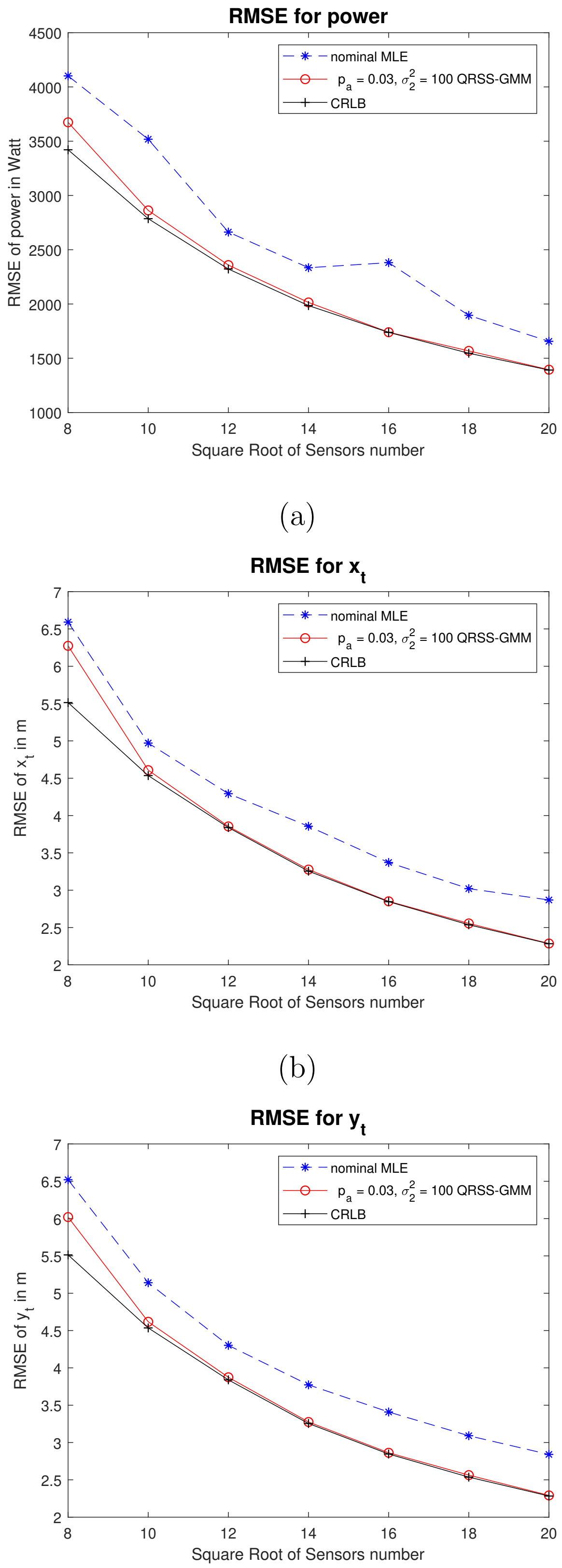
In this section, we present the simulation results of the proposed Q-RSS-GM estimator. We will compare the Q-RSS-GM estimator with the MLE based on the nominal model ignoring any possible false information attacks as in [6], and with the CRLB derived in Section 4.2. A systematic grid search is employed to find an initial state estimate, then we use the nonlinear optimization function in MATLAB to find the MLE $\widehat{\boldsymbol{\theta}} = [\widehat{P}_0 \ \widehat{x}_t \ \widehat{y}_t]^T$. We assume a uniform sensor deployment with known sensor locations in a $200 \times 200\ m^2$ region of interest. The true target state vector is $\boldsymbol{\theta} = [25000\ 15\ 20]$. The GMM parameters used for the attack are assumed to be known with $\sigma_1 = 1, \sigma_2 = 10$, and $\mu_1 = \mu_2 = 0$. The performance of our proposed method is evaluated in terms of the root mean squared error (RMSE) for the estimated parameters. The simulation is based on 1000 Monte-Carlo runs with the sensor square root number ranging from 8 to 20. We use the quaternary quantization with $M = 2$, the attenuation parameter is $n = 2$. The quantization threshold is set as $\boldsymbol{\eta}_i = [0.82, 1.7, 2.72]^T, \ \ \forall i$.

In Fig. 2, the RMSE of the Q-RSS-GM estimator is plotted as a function of the attack probability p_a, which takes one of the values in the following vector

$$[0.01\ \ 0.0246\ \ 0.0605\ \ 0.1488\ \ 0.3659\ \ 0.9]^T.$$

It is assumed that $N = 144$ sensors are deployed in the ROI for this simulation. Note that the increase in attack probability will lead to an increase in the RMSE and the corresponding CRLB in general. It is also clear that the proposed Q-RSS-GM approach provides much better estimation performance especially when the attack probability is large.

In Fig. 3, the RMSE of the Q-RSS-GM estimator is plotted as a function of the number of sensors. The attack probability is assumed to be $p_a = 0.03$. The performance of the proposed Q-RSS-GM estimator is compared with that of the nominal MLE from [6] under the same situation, and the derived CRLB. It is clear that our proposed Q-RSS-GM estimator outperforms the nominal MLE that ignores the attacks. The system performance improves with increasing number of sensors and the RMSE of the proposed Q-RSS-GM estimator approaches the CRLB for all the estimated parameters.

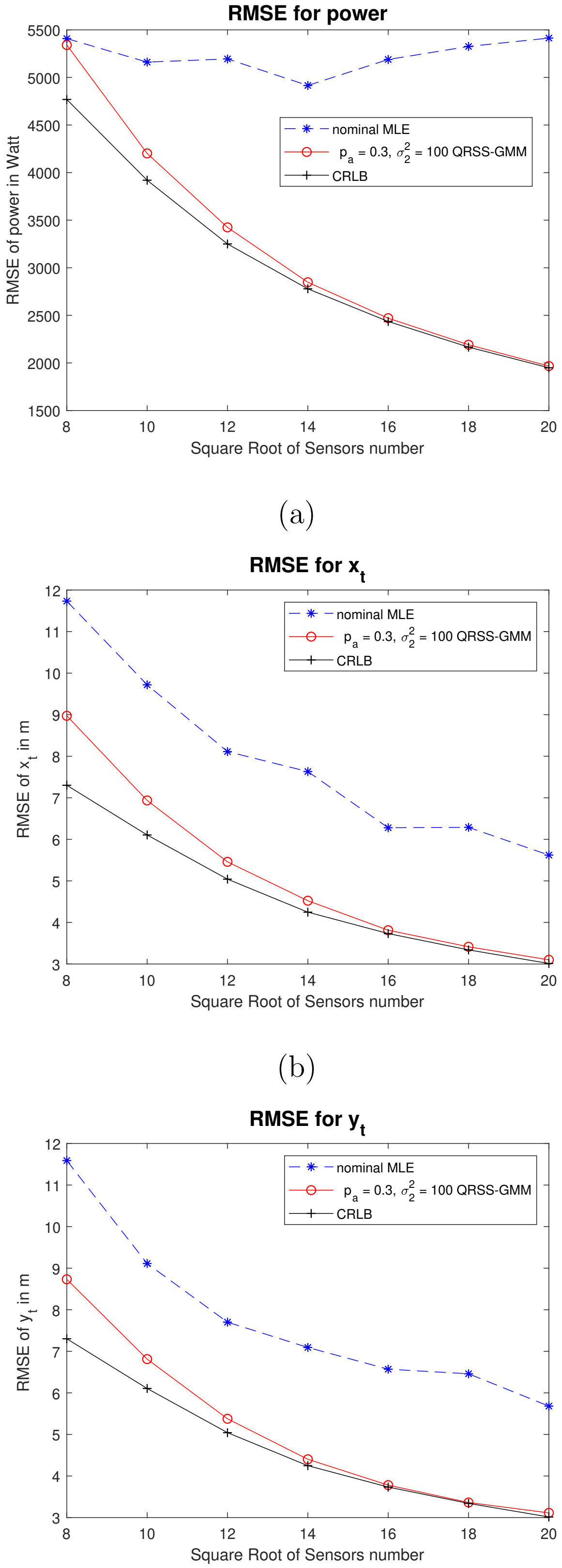


(a)

(b)

(c)

Fig. 3. The RMSE vs. square root of number of sensors ($p_a = 0.03$) .



(a)

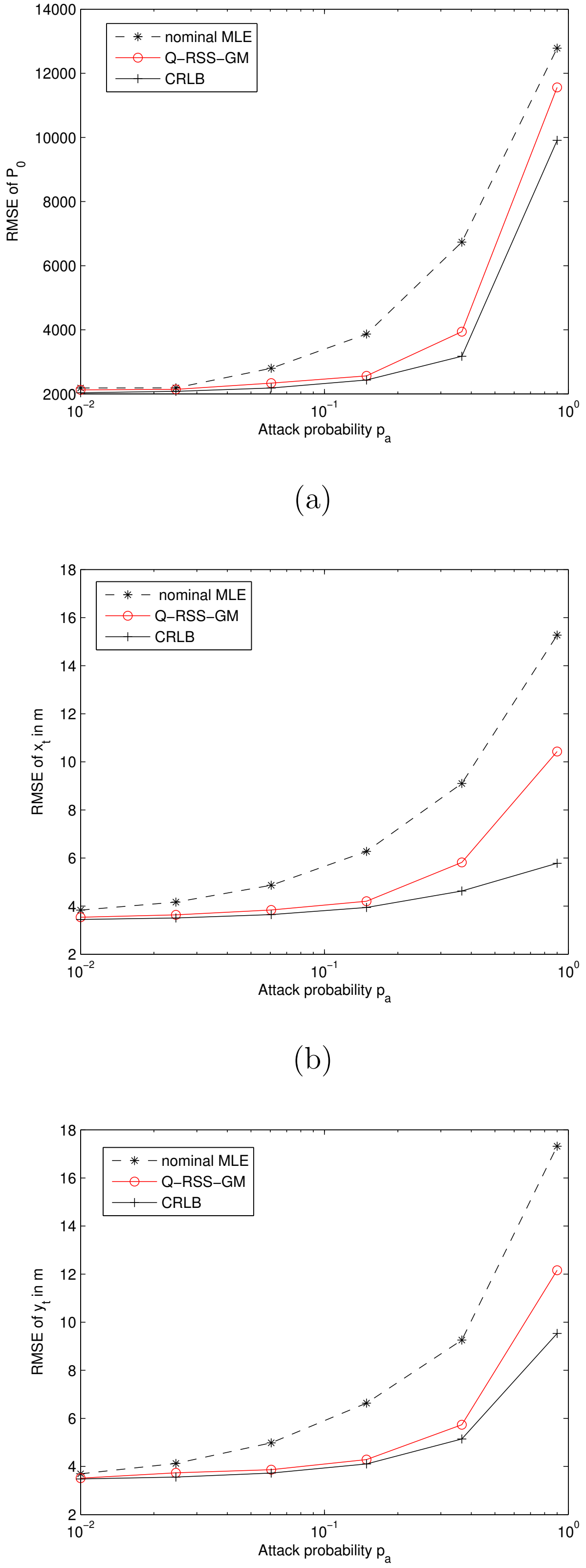
(b)

(c)

Fig. 4. The RMSE vs. square root of number of sensors ($p_a = 0.3$).

Similar to Fig. 3, in Fig. 4 the RMSE of the Q-RSS-GM estimator is plotted as a function of the number of sensors. However, the attack probability is increased ten-fold to $p_a = 0.3$. The RMSE of the Q-RSS-GM estimator is compared with that of the nominal method and the CRLB. Note that a larger attack probability will mislead the nominal method more significantly with a much larger RMSE while the proposed method is still able to converge to the CRLB and provide acceptable performance, especially when the number of the deployed sensors increases. The probability of attack ($p_a = 0.3$) is considered to be very high since the attacker in this case is assumed to have enough resources to corrupt 30% of the sensors in the ROI.

In Fig.5, the RMSE of the Q-RSS-GM estimator is plotted as a function of the attack probability pa, which takes one of the values in the following vector $[0.01\ 0.0246\ 0.0605\ 0.1488\ 0.3659\ 0.9]^T$. It is assumed that $N = 144$ sensors are deployed in the ROI for this simulation. Note that the increase in attack probability will lead to an increase in the RMSE and the corresponding CRLB in general. It is also clear that the proposed Q-RSS-GM approach provides much better estimation performance especially when the attack probability is large.



(a)

(b)

(c)

Fig. 5. The RMSE vs. the probability of attack (p_a).

In Tables 1, 2, and 3, we show the performance of a Q-RSS-GM estimator which we have designed assuming a nominal attack probability of $p_{a_n} = 0.05$. Then we assume that the adversary uses a different actual attack probability p_{a_t}. We test the performance of the estimator with a mismatched parameter p_a, and observe its response to various actual p_{a_t}'s. It is clear that the estimator works perfectly when the actual attack probability p_{a_t} is less than or equal to $p_{a_n} = 0.05$, which is the nominal attack probability by design. Note that the estimator is still able to give very good performance even when the actual attack probability is increased to $p_{a_t} = 0.1$. These results show that the proposed approach is robust against parameter mismatch.

Table 1. RMSE of P_0 for the Q-RSS-GM estimator with a mismatched p_a

N	$p_{a_t} = 0$	$p_{a_t} = 0.01$	$p_{a_t} = 0.05$	$p_{a_t} = 0.1$
144	2505.1	2505.6	2594.9	2714.0
256	1827.2	1847.0	1912.2	2026.6
400	1449.6	1460.3	1493.2	1571.1

Table 2. RMSE of x_t for the Q-RSS-GM estimator with a mismatched p_a

N	$p_{a_t} = 0$	$p_{a_t} = 0.01$	$p_{a_t} = 0.05$	$p_{a_t} = 0.1$
144	4.0614	4.1942	4.4723	4.6697
256	3.0083	3.0859	3.3337	3.4781
400	2.3707	2.4047	2.7200	2.8098

Table 3. RMSE of y_t for the Q-RSS-GM estimator with a mismatched p_a

N	$p_{a_t} = 0$	$p_{a_t} = 0.01$	$p_{a_t} = 0.05$	$p_{a_t} = 0.1$
144	4.1767	4.1357	4.4543	4.7472
256	3.0306	3.0286	3.2758	3.5819
400	2.4028	2.4592	2.4864	2.7359

CHAPTER 4

SOURCE LOCATION ESTIMATION UNDER FALSE INFORMATION ATTACK WITH UNKNOWN ATTACK POWER AND PROBABILITY

In this Chapter, we investigate the problem of source location estimation in wireless sensor networks (WSNs) based on quantized data in the presence of false information attacks. We assume that the attack power and probability is unknown for the system, we develop a maximum likelihood estimator (MLE) to locate the source with sensor data corrupted by injected false information. The Cramer-Rao lower bound (CRLB) for this estimation problem is also derived to evaluate the estimator's performance. It is shown that the proposed estimator is robust in various cases with the attack probability and power following a uniform distribution, and it shows excellent performance under the mismatch case.

4.1 Mathematical Modeling

In this section, we use the same mathematical model for the target as in (3.3). We assume here the false information signal is following a randomized Gaussian mixture model RGMM, then the probability density function (pdf) of b_i can be modeled as

$$f(b_i, p_a, \sigma_2^2) = f(b_i|p_a, \sigma_2^2)\ f(p_a, \sigma_2^2) \tag{4.1}$$

where b_i is the injected false information, p_a and σ_2^2 are the attack probability and the attack power respectively, and $f(p_a, \sigma_2^2)$ is the joint pdf of the attack probability and the attack power.

We assume that p_a and σ_2^2 are independent and identically distributed *i.i.d.* We also assume that there are constraints on the p_a and σ_2^2, and both of them follow certain uniform distributions.

$$f(p_a) = \begin{cases} \frac{1}{\alpha} & 0 \leq p_a \leq \alpha \\ \\ 0 & o.w \end{cases} \tag{4.2}$$

where α is the upper limit for the probability of attack p_a which can be assigned according to the possible percentage of the sensors, which the attacker or noise is able to corrupt or control in the targeted WSN.

$$f(\sigma_2^2) = \begin{cases} \frac{1}{\rho - \sigma_1^2} & \sigma_1^2 \leq \sigma_2^2 \leq \rho \\ \\ 0 & o.w \end{cases} \tag{4.3}$$

where ρ is the upper limit for the attack noise power σ_2^2, which can also be assigned according to the power of the attacker or noise in the targeted WSN area.

$f(b_i|p_a, \sigma_2^2)$ is

$$f(b_i|p_a, \sigma_2^2) = \left[(1 - p_a)\mathcal{N}(0, \sigma_1^2) + p_a\mathcal{N}(0, \sigma_2^2)\right] \tag{4.4}$$

where $\mathcal{N}(0, \sigma_1^2)$ is the Gaussian distribution with zero mean, and variance $\sigma_1^2 = 1$, which can be considered the case of no attack. $\mathcal{N}(0, \sigma_2^2)$ is a Gaussian component with zero mean, and σ_2^2 variance which follows a uniform distribution as in (4.3)

Now, we can rewrite the pdf of the RGMM assumption as

$$f(b_i, p_a, \sigma_2^2) = \left[(1 - p_a)\mathcal{N}(0, \sigma_1^2) + p_a\mathcal{N}(0, \sigma_2^2)\right] f(p_a)f(\sigma_2^2) \tag{4.5}$$

We can see that according to (4.5), we can find the b_i pdf by integrating (4.5) over p_a and σ_2^2 respectively. The resulting $f(b_i)$ is therefore

$$f(b_i) = \int_{\sigma_1^2}^{\rho}\int_{0}^{\alpha}\left[(1-p_a)\tfrac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{b_i^2}{2\sigma_1^2}} + p_a\tfrac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{b_i^2}{2\sigma_2^2}}\right]\frac{1}{\alpha}\frac{1}{\rho-\sigma_1^2}dp_a d\sigma_2^2 \tag{4.6}$$

Solving the internal integration over p_a we get

$$(1-\frac{\alpha}{2})\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{b_i^2}{2\sigma_1^2}} + \int_{\sigma_1^2}^{\rho}\left[\tfrac{\alpha}{2}\tfrac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{b_i^2}{2\sigma_2^2}}\tfrac{1}{\rho-\sigma_1^2}d\sigma_2^2\right] \tag{4.7}$$

Solving the integral in (4.7), we get the false information injected distribution $f(b_i)$:

$$\begin{gathered} b_i \sim \bigg[\left(1-\frac{\alpha}{2}\right)\mathcal{N}(0,\sigma_1^2) \\ +\left(\frac{\alpha}{2}\right)\ \left(\frac{1}{\rho-\sigma_1^2}\right)\Bigg[\Bigg[\frac{e^{-\frac{b_i^2}{2\rho}}}{\sqrt{\pi}\sqrt{\frac{1}{2\rho}}} - \frac{e^{-\frac{b_i^2}{2\sigma_1^2}}}{\sqrt{\pi}\sqrt{\frac{1}{2\sigma_1^2}}}\Bigg] \\ +\sqrt{b_i^2}\ \left[erf\left(\frac{\sqrt{b_i^2}}{\sqrt{2\rho}}\right) - erf\left(\frac{\sqrt{b_i^2}}{\sqrt{2\sigma_1^2}}\right)\right]\Bigg] \end{gathered} \tag{4.8}$$

Now we have derived the pdf of the assumed injected noise.

We will use the same signal collecting strategy as in [7], as shown in Fig. (2). The amplitude a_i at the ith sensor will be corrupted by a noise that follows RGM Model, which can be considered as a false information injection or a general spoofing attack by an unknown adversary. The signal received at each sensor can be modeled by :

$$r_i = a_i + b_i \tag{4.9}$$

where r_i is the received corrupted signal at the i_{th} sensor ,a_i is the amplitude of the uncorrupted target signal at i_{th} sensor as given in (3.1), and b_i is the attacker noise that follows the distribution given in eq.(4.8).

Since, we have found the b_i's distribution, we are ready to derive the maximum likelihood estimator (MLE) and the CRLB for the problem where p_a and σ_2^2 are random variables in the next section.

4.2 Maximum Likelihood Estimator (MLE)

In this section, we derive the maximum likelihood estimator for fault-tolerant source location problem where p_a and σ_2^2 are unknown random variables. We assume that the MLE uses QRSS data as in Chapter 3. Where the same quantization process as in (3.6)

$$D_i = \begin{cases} 0 & -\infty \leq r_i < \eta_{i1} \\ 1 & \eta_{i1} \leq r_i < \eta_{i2} \\ : & : \\ : & : \\ L-1 & \eta_{L-1} \leq r_i < \infty \end{cases} \tag{4.10}$$

where $\boldsymbol{D} = [D_1, \cdots, D_N]^T$ is the vector of the quantized data generated by the sensors, and D_i could take any value in the $D_i \in \{0, \cdots, 2^M - 1\}$, which is also denoted by $D_i = l \in [0 \quad L-1]$, where $L = 2^M$.

Now, we need to find the probability that D_i takes a specific value of l according toassumptions made earlier in this Chapter. It is possible to find $p_{il}(\eta_i, \theta)$ by integrating the pdf of b_i over the interval $[\xi_{il}\ \ \xi_{il+1}]$, where $\xi_{il} = (\eta_{il} - a_i)$ and $\xi_{il+1} = (\eta_{il+1} - a_i)$ respectively. For simplicity, we denote $c_1 = (1 - \frac{\alpha}{2})$, and $c_2 = \left(\frac{\alpha}{2}\right)\left(\frac{1}{\rho - \sigma_1^2}\right)$.

$$\int_{\xi_{il}}^{\xi_{il+1}} f(b_i) db_i \tag{4.11}$$

For simplicity, we divide the integration into five terms and we will calculate the integration for each term with respect to b_i.

1st term

$$\int_{\xi_{il}}^{\xi_{il+1}} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{b_i^2}{2\sigma_1^2}} \quad db_i \quad = \quad \left[Q\left(\frac{\xi_{il}}{\sigma_1}\right) - Q\left(\frac{\xi_{il+1}}{\sigma_1}\right)\right] \quad = \quad \lambda_0 \tag{4.12}$$

2nd term

$$\int_{\xi_{il}}^{\xi_{il+1}} \left[\frac{e^{-\frac{b_i^2}{2\rho}}}{\sqrt{\pi}\sqrt{\frac{1}{2\rho}}}\right] db_i = 2\rho \ \left[Q\left(\frac{\xi_{il}}{\sqrt{\rho}}\right) - Q\left(\frac{\xi_{il+1}}{\sqrt{\rho}}\right)\right] = \lambda_1 \tag{4.13}$$

3rd term

$$\int_{\xi_{il}}^{\xi_{il+1}} \left[\frac{e^{-\frac{b_i^2}{2\sigma_1^2}}}{\sqrt{\pi}\sqrt{\frac{1}{2\sigma_1^2}}}\right] db_i = 2\sigma_1^2 \ \left[Q\left(\frac{\xi_{il}}{\sigma_1}\right) - Q\left(\frac{\xi_{il+1}}{\sigma_1}\right)\right] = \lambda_2 \tag{4.14}$$

4th term

$$\int_{\xi_{il}}^{\xi_{il+1}} \sqrt{b_i^2}\ erf\left(\frac{\sqrt{b_i^2}}{\sqrt{2\rho}}\right) db_i$$

$$= \left[\frac{(\xi_{il+1}^2 - \rho)\ \ erf\left(\frac{\xi_{il+1}}{\sqrt{2\rho}}\right)}{2} + \frac{\sqrt{\rho}\ \ \xi_{il+1}\ \ e^{-\frac{\xi_{il+1}^2}{2\rho}}}{\sqrt{2\pi}}\right] - \left[\frac{(\xi_{il}^2 - \rho)\ \ erf\left(\frac{\xi_{il}}{\sqrt{2\rho}}\right)}{2} + \frac{\sqrt{\rho}\ \ \xi_{il}\ \ e^{-\frac{\xi_{il}^2}{2\rho}}}{\sqrt{2\pi}}\right] = \lambda_3$$

Similarly we can find the 5th term result as:

5th term

$$\int_{\xi_{il}}^{\xi_{il+1}} \sqrt{b_i^2}\ erf\left(\frac{\sqrt{b_i^2}}{\sqrt{2\sigma_1^2}}\right) db_i = \left[\frac{(\xi_{il+1}^2 - \sigma_1^2)\ erf\left(\frac{\xi_{il+1}}{\sqrt{2\sigma_1^2}}\right)}{2} + \frac{\sigma_1\ \xi_{il+1}\ e^{-\frac{\xi_{il+1}^2}{2\sigma_1^2}}}{\sqrt{2\pi}}\right]$$

$$- \left[\frac{(\xi_{il}^2 - \sigma_1^2)\ erf\left(\frac{\xi_{il}}{\sqrt{2\sigma_1^2}}\right)}{2} + \frac{\sigma_1\ \xi_{il}\ e^{-\frac{\xi_{il}^2}{2\sigma_1^2}}}{\sqrt{2\pi}}\right] = \lambda_4$$

Then, the probability $p_{il}(\eta_i, \theta)$ is

$$p_{il}(\eta_i, \theta) = c_1\ [\lambda_0] + c_2\ [(\lambda_1 - \lambda_2) + (\lambda_3 - \lambda_4)] \tag{4.15}$$

Now the joint probability of the sensor data can be found as follows

$$p\ (\mathbf{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N} \prod_{l=0}^{L-1} p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})^{\delta(D_i - l)} \tag{4.16}$$

where $\delta(.)$ is the Kronecker delta function. The log-likelihood function for the sensor data is

$$\log p(\mathbf{D}|\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \delta(D_i - l) \log p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta}) \tag{4.17}$$

Our object is to find the value of $\boldsymbol{\theta}$ that maximizes the log-likelihood function.

So the ML estimator is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log p(\mathbf{D}|\boldsymbol{\theta}) \tag{4.18}$$

After we find the MLE for the source location, we can find the CRLB for the proposed estimator.

4.3 Cramer Rao Lower Bound (CRLB)

In this section we derive the CRLB for the proposed MLE for source location estimation for performance evaluation.

Theorem 2 : For an unbiased estimator $\widehat{\boldsymbol{\theta}}(\mathbf{D})$, the CRLB is given by

$$E\big\{[\widehat{\boldsymbol{\theta}}(\mathbf{D}) - \boldsymbol{\theta}][\widehat{\boldsymbol{\theta}}(\mathbf{D}) - \boldsymbol{\theta}]^T\big\} \geq \mathbf{J}^{-1} \tag{4.19}$$

where $\mathbf{J}$ is the 3×3 Fisher information matrix (FIM).

Proof:

The following is the proof of *Theorem* 2, which provides the FIM for the localization problem subjected to the false information injection attacks, where the attack probability and noise variance follow certain uniform distribution.

$$\mathbf{J} = -E[\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \log p(\mathbf{D}|\boldsymbol{\theta})] \tag{4.20}$$

where $\nabla_{\boldsymbol{\theta}}$ is the gradient vector:

$$\nabla_{\boldsymbol{\theta}} = \left[\frac{\partial}{\partial P_0}\ \frac{\partial}{\partial x_t}\ \frac{\partial}{\partial y_t}\right]^T \tag{4.21}$$

The Fisher information matrix $\mathbf{J}$ can be described as

$$\mathbf{J} = -E[\nabla_\theta \nabla_\theta^T \log p(\mathbf{D}|\boldsymbol{\theta})] = -\mathbf{E}\begin{bmatrix} \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial x_t} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial y_t} \\ \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial P_0} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t^2} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial y_t} \\ \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t \partial P_0} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t \partial x_t} & \frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t^2} \end{bmatrix} = \begin{bmatrix} j_{11} & j_{12} & j_{13} \\ j_{21} & j_{22} & j_{23} \\ j_{31} & j_{32} & j_{33} \end{bmatrix} \tag{4.22}$$

From (2.9) we have the log - likelihood function as

$$\log p(\boldsymbol{D}|\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \delta(D_i - l) \log p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta}) \tag{4.23}$$

Then, each element of the above matrix can be calculated as follows ,

the 1st derivative of the log likelihood function with respect to P_0 is

$$\frac{\partial \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0} = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial P_0} \right] \tag{4.24}$$

and the 2nd derivative is

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2} = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})]^2} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial P_0} \right]^2 + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2} \right] \tag{4.25}$$

Next, we calculate the (1,1) element of $\mathbf{J}$.

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2} = \sum_{i=1}^{N} \sum_{l=0}^{L-1} - \frac{\delta(D_i - l)}{p_{il}^2(\boldsymbol{\eta}_i, \boldsymbol{\theta})} \left[\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial P_0} \right]^2 + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta}_i, \theta)} \left[\frac{\partial^2 p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial P_0^2} \right] \tag{4.26}$$

and

$$j_{11} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2}\right] = \sum_{i=1}^{N}\sum_{l=0}^{L-1}\frac{1}{p_{il}}\left[\frac{\partial p_{il}(\boldsymbol{\eta}_i,\boldsymbol{\theta})}{\partial P_0}\right]^2 \tag{4.27}$$

by using the identity $E[\delta(D_i - l)] = p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})$. Since

$$p_{il}(\eta_i, \theta) = c_1[\lambda_0] + c_2[(\lambda_1 - \lambda_2) + (\lambda_3 - \lambda_4)] \tag{4.28}$$

.

we have

$$\frac{\partial p_{il}(\boldsymbol{\eta}_i,\boldsymbol{\theta})}{\partial P_0} = c_1[\frac{\partial \lambda_0}{\partial P_0}] + c_2[(\frac{\partial \lambda_1}{\partial P_0} - \frac{\partial \lambda_2}{\partial P_0}) + (\frac{\partial \lambda_3}{\partial P_0} - \frac{\partial \lambda_4}{\partial P_0})] \tag{4.29}$$

Then

$$\frac{\partial \lambda_0}{\partial P_0} = \left[\frac{\gamma_{il}}{2\sqrt{2\pi}a_i d_i^n \sigma_1}\right] = \frac{k_{il0}}{2a_i d_i^n} \tag{4.30}$$

$$\frac{\partial \lambda_1}{\partial P_0} = \frac{\sqrt{2}\sqrt{\rho}\ \phi_{il}}{2\sqrt{\pi}a_i d_i^n} = \frac{k_{il1}}{2a_i d_i^n} \tag{4.31}$$

$$\frac{\partial \lambda_2}{\partial P_0} = \frac{\sqrt{2}\ \sigma_1\ \gamma_{il}}{2\sqrt{\pi}a_i d_i^n} = \frac{k_{il2}}{2a_i d_i^n} \tag{4.32}$$

$$\frac{\partial \lambda_3}{\partial P_0} = \frac{k_{il3}}{2a_i d_i^n} \tag{4.33}$$

$$\frac{\partial \lambda_4}{\partial P_0} = \frac{k_{il4}}{2a_i d_i^n} \tag{4.34}$$

where

$$k_{il0} = \left[\frac{\gamma_{il}}{\sqrt{2\pi}\sigma_1} \right] \tag{4.35}$$

$$k_{il1} = \frac{\sqrt{2}\sqrt{\rho}}{\sqrt{\pi}} \ \ [\phi_{il}] \tag{4.36}$$

$$k_{il2} = \frac{\sqrt{2}\sigma_1}{\sqrt{\pi}} \ \ [\gamma_{il}] \tag{4.37}$$

$$k_{il3} = \left[\xi_{il+1} \ erf \left(\frac{\xi_{il+1}}{\sqrt{2\rho}} \right) - \xi_{il} \ erf \left(\frac{\xi_{il}}{\sqrt{2\rho}} \right) \right] \tag{4.38}$$

$$k_{il4} = \left[\xi_{il+1} \ erf \left(\frac{\xi_{il+1}}{\sqrt{2\sigma_1^2}} \right) - \xi_{il} \ erf \left(\frac{\xi_{il}}{\sqrt{2\sigma_1^2}} \right) \right] \tag{4.39}$$

and

$$\gamma_{il} = \left[e^{-\left(\frac{\xi_{il}^2}{2\sigma_1^2} \right)} - e^{-\left(\frac{\xi_{il+1}^2}{2\sigma_1^2} \right)} \right] \tag{4.40}$$

$$\phi_{il} = \left[e^{-\left(\frac{\xi_{il}^2}{2\rho} \right)} - e^{-\left(\frac{\xi_{il+1}^2}{2\rho} \right)} \right] \tag{4.41}$$

Then

$$\left[\frac{\partial p_{il}(\boldsymbol{\eta}_i,\boldsymbol{\theta})}{\partial P_0}\right] = \left[\frac{\left[\frac{c_1}{2}[k_{il0}] + \frac{c_2}{2}[(k_{il1}-k_{il2})+(k_{il3}-k_{il4})]\right]}{a_i d_i^n}\right] \tag{4.42}$$

Similarly we can find $\frac{\partial p_{il}(\boldsymbol{\eta}_i,\boldsymbol{\theta})}{\partial x_t}$ as

$$\frac{\partial p_{il}(\boldsymbol{\eta}_i,\boldsymbol{\theta})}{\partial x_t} = c_1\left[\frac{\partial \lambda_0}{\partial x_t}\right] + c_2\left[\left(\frac{\partial \lambda_1}{\partial x_t} - \frac{\partial \lambda_2}{\partial x_t}\right) + \left(\frac{\partial \lambda_3}{\partial x_t} - \frac{\partial \lambda_4}{\partial x_t}\right)\right] \tag{4.43}$$

Then

$$\frac{\partial \lambda_0}{\partial x_t} = n a_i d_i^{-2}\left[\frac{\gamma_{il}}{2\sqrt{2\pi}\sigma_1}\right](x_i - x_t) = n a_i d_i^{-2}\frac{k_{il0}}{2}(x_i - x_t) \tag{4.44}$$

$$\frac{\partial \lambda_1}{\partial x_t} = n a_i d_i^{-2}\frac{\sqrt{2}\sqrt{\rho}\ \phi_{il}}{2\sqrt{\pi}}(x_i - x_t) = n a_i d_i^{-2}\frac{k_{il1}}{2}(x_i - x_t) \tag{4.45}$$

$$\frac{\partial \lambda_2}{\partial x_t} = n a_i d_i^{-2}\frac{\sqrt{2}\ \sigma_1\ \gamma_{il}}{2\sqrt{\pi}}(x_i - x_t) = n a_i d_i^{-2}\frac{k_{il2}}{2}(x_i - x_t) \tag{4.46}$$

$$\frac{\partial \lambda_3}{\partial x_t} = n a_i d_i^{-2}\frac{k_{il3}}{2}(x_i - x_t) \tag{4.47}$$

$$\frac{\partial \lambda_4}{\partial x_t} = n a_i d_i^{-2}\frac{k_{il4}}{2}(x_i - x_t) \tag{4.48}$$

Then,

$$\left[\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial x_t}\right] = na_i d_i^{-2} \left(\frac{c_1}{2}\left(k_{il0}\right) + \frac{c_2}{2}\left(\left(k_{il1} - k_{il2}\right) + \left(k_{il3} - k_{il4}\right)\right)\right)(x_i - x_t) \quad (4.49)$$

Now, we can find $\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial y_t}$ as

$$\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial y_t} = c_1 \left[\frac{\partial \lambda_0}{\partial y_t}\right] + c_2 \left[\left(\frac{\partial \lambda_1}{\partial y_t} - \frac{\partial \lambda_2}{\partial y_t}\right) + \left(\frac{\partial \lambda_3}{\partial y_t} - \frac{\partial \lambda_4}{\partial y_t}\right)\right] \quad (4.50)$$

Where

$$\frac{\partial \lambda_0}{\partial x_t} = na_i d_i^{-2} \left[\frac{\gamma_{il}}{2\sqrt{2\pi}\sigma_1}\right](y_i - y_t) = na_i d_i^{-2} \frac{k_{il0}}{2}(y_i - y_t) \quad (4.51)$$

$$\frac{\partial \lambda_1}{\partial y_t} = na_i d_i^{-2} \frac{\sqrt{2}\sqrt{\rho}\ \phi_{il}}{2\sqrt{\pi}}(y_i - y_t) = na_i d_i^{-2} \frac{k_{il1}}{2}(y_i - y_t) \quad (4.52)$$

$$\frac{\partial \lambda_2}{\partial y_t} = na_i d_i^{-2} \frac{\sqrt{2}\ \sigma_1\ \gamma_{il}}{2\sqrt{\pi}}(y_i - y_t) = na_i d_i^{-2} \frac{k_{il2}}{2}(y_i - y_t) \quad (4.53)$$

$$\frac{\partial \lambda_3}{\partial y_t} = na_i d_i^{-2} \frac{k_{il3}}{2}(y_i - y_t) \quad (4.54)$$

$$\frac{\partial \lambda_4}{\partial y_t} = na_i d_i^{-2} \frac{k_{il4}}{2}(y_i - y_t) \quad (4.55)$$

Then,

$$\left[\frac{\partial p_{il}(\boldsymbol{\eta}_i, \boldsymbol{\theta})}{\partial y_t}\right] = n a_i d_i^{-2} \left(\frac{c_1}{2} \left(k_{il0} \right) + \frac{c_2}{2} \left(\left(k_{il1} - k_{il2} \right) + \left(k_{il3} - k_{il4} \right) \right) \right) \left(y_i - y_t \right) \quad (4.56)$$

Now we can return to calculate the **J** matrix elements,

By substituting (4.42) in (4.27), we get

$$j_{11} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2}\right] = \sum_{i=1}^{N} \Omega_i a_i^{-2} d_i^{-2n} \quad (4.57)$$

where

$$\Omega_i = \sum_{l=0}^{L-1} \frac{\left[\frac{c_1}{2}[k_{il0}] + \frac{c_2}{2}[(k_{il1} - k_{il2}) + (k_{il3} - k_{il4})]\right]^2}{p_{il}(\eta_i, \theta)} \quad (4.58)$$

Similarly, we can find the j_{12} which is equal to j_{21} elements since the FIM matrix is symmetric

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial x_t} = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})]^2} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial P_0}\right] \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial x_t}\right] + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial x_t}\right] \quad (4.59)$$

$$j_{12} = j_{21} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial x_t}\right] = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{1}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial P_0}\right] \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial x_t}\right] \quad (4.60)$$

substituting (4.42) and (4.49) in (4.60), we get

$$j_{12} = j_{21} = -E\left[\tfrac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0^2}\right] = n\sum_i \Omega_i d_i^{-(n+2)}(x_i - x_t) \tag{4.61}$$

Next, the j_{22} element of the FIM matrix can be found as

$$j_{22} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\theta)}{\partial x_t^2}\right] = \sum_{i=1}^{N}\sum_{l=0}^{L-1}\frac{1}{p_{il}(\eta_i,\theta)}\left[\frac{\partial p_{il}(\eta_i,\theta)}{\partial x_t}\right]^2 \tag{4.62}$$

substituting (4.49) and (4.62), we get

$$j_{22} = n^2\sum_i \Omega_i a_i^2 d_i^{-4}(x_i - x_t)^2 \tag{4.63}$$

Next, we calculate $j_{13} = j_{31}$ elements

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial y_t} = \sum_{i=1}^{N}\sum_{l=0}^{L-1}\frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})]^2}\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial P_0}\right]\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial y_t}\right] + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial y_t}\right] \tag{4.64}$$

Substituting (4.42) and (4.56) in (4.64), then $j_{13} = j_{31}$ as:

$$j_{13} = j_{31} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial P_0 \partial y_t}\right] = \sum_{i=1}^{N}\sum_{l=0}^{L-1}\frac{1}{p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial P_0}\right]\left[\frac{\partial p_{il}(\boldsymbol{\eta_i},\boldsymbol{\theta})}{\partial y_t}\right] \tag{4.65}$$

Then,

$$j_{13} = j_{31} = n \sum_i \Omega_i d_i^{-(n+2)} (y_i - y_t) \tag{4.66}$$

Now, calculating the elements $j_{23} = j_{32}$:

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial y_t} = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})]^2} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial x_t}\right] \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial y_t}\right] + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial y_t}\right] \tag{4.67}$$

then $j_{23} = j_{32}$

$$j_{23} = j_{32} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial x_t \partial y_t}\right] = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{1}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial x_t}\right] \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial y_t}\right] \tag{4.68}$$

Thus, $j_{13} = j_{31}$ elements

$$j_{23} = j_{32} = n^2 \sum_i \Omega_i a_i^2 d_i^{-4} (x_i - x_t)(y_i - y_t) \tag{4.69}$$

Finally, we calculate the j_{33} element

$$\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t^2} = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{-\delta(D_i - l)}{[p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})]^2} \left[\frac{\partial p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})}{\partial y_t}\right]^2 + \frac{\delta(D_i - l)}{p_{il}(\boldsymbol{\eta_i}, \boldsymbol{\theta})} \left[\frac{\partial^2 \log p(\mathbf{D}|\boldsymbol{\theta})}{\partial y_t^2}\right] \tag{4.70}$$

Then j_{33} :

$$j_{33} = -E\left[\frac{\partial^2 \log p(\mathbf{D}|\theta)}{\partial y_t^2}\right] = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \frac{1}{p_{il}(\eta_i, \theta)} \left[\frac{\partial p_{il}(\eta_i, \theta)}{\partial y_t}\right]^2 \tag{4.71}$$

Thus,

$$j_{33} = n^2 \sum_i \Omega_i a_i^2 d_i^{-4} (y_i - y_t)^2 \tag{4.72}$$

In summary, we have the CRLB for the above MLE as follows :

$$j_{11} = \sum_i \Omega_i d_i^{-2n} a_i^{-2}$$

$$j_{12} = j_{21} = n \sum_i \Omega_i d_i^{-(n+2)} (x_i - x_t)$$

$$j_{13} = j_{31} = n \sum_i \Omega_i d_i^{-(n+2)} (y_i - y_t)$$

$$j_{22} = n^2 \sum_i \Omega_i a_i^2 d_i^{-4} (x_i - x_t)^2$$

$$j_{23} = j_{32} = n^2 \sum_i \Omega_i a_i^2 d_i^{-4} (x_i - x_t)(y_i - y_t)$$

$$j_{33} = n^2 \sum_i \Omega_i a_i^2 d_i^{-4} (y_i - y_t)^2 \tag{4.73}$$

then the FIM matrix can be written as:

$$\boldsymbol{J} = \begin{bmatrix} j_{11} & j_{12} & j_{13} \\ j_{21} & j_{22} & j_{23} \\ j_{31} & j_{32} & j_{33} \end{bmatrix} \tag{4.74}$$

Then, the CRLB can be easily calculated by finding the inverse of **J**.

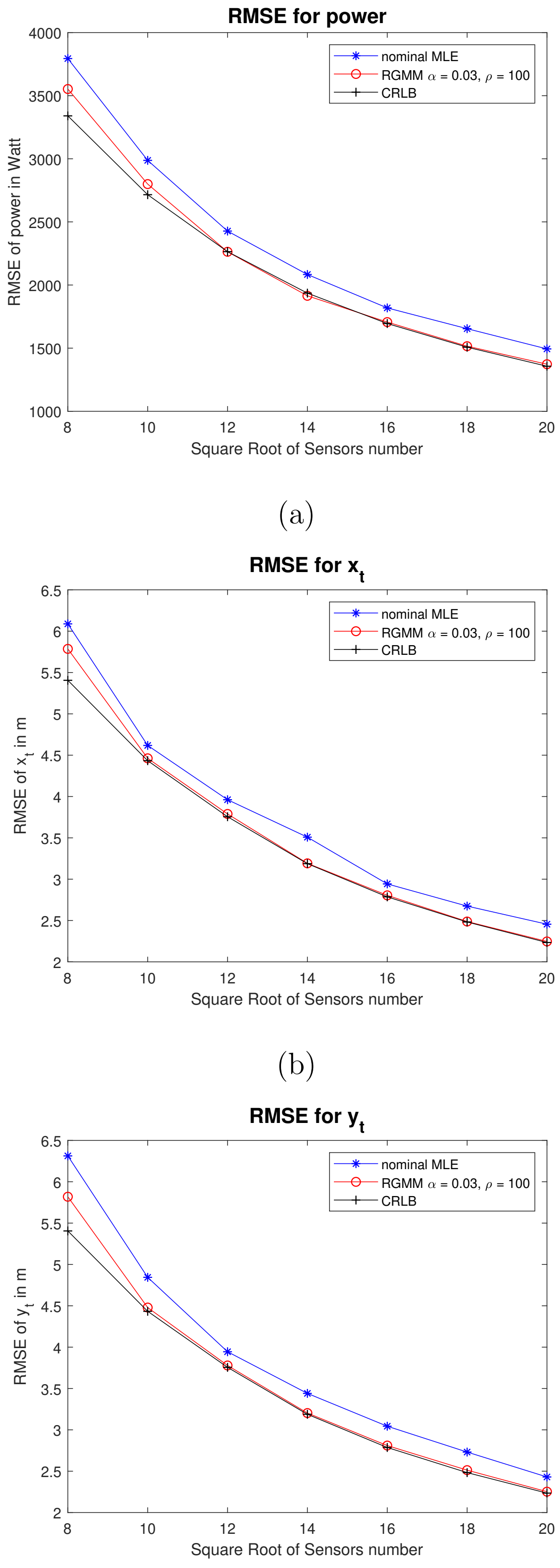
Now, we have derived the closed-from MLE and the CRLB for the false information injection attacks, where the asttack probability and noise variance follow certain uniform distribution.

In the next section, we will present and discuss the simulation results for the proposed estimator.

4.4 Simulation Results

In this section, we present and discuss the simulation results for the proposed source location estimation based on the MLE. Here, we also use a grid search algorithm to find the intial esitmate, then we use the MATLAB nonlinear optimization function $fmincon$ to find the estimate $\widehat{\boldsymbol{\theta}} = [\widehat{P_0}\ \widehat{x_t}\ \widehat{y_t}]$. We also assume a sensor deployed uniformly in the ROI, the threshold set is set as $\boldsymbol{\eta_i} = [1, 2, 3]$, and the state vector of the true parameters is [25000 2 2]. The injected noise attack probability and injected noise variance follow uniform distributions. So $p_a \sim \mathcal{U}(0, \alpha)$ and $\sigma_2^2 \sim \mathcal{U}(\sigma_1^2, \rho)$. Our simulations are based on 1500 Monte-Carlo runs with the sensor square root number ranging from 8 to 20. The quantization used here is also the quaternary quantization with $M = 2$ and the attenuation parameter is set as $n = 2$. We use the root mean square error (RMSE) as a metric for performance evaluation. The ROI area is $200 \times 200\ m^2$.

In Fig.6, the performance of the proposed estimator is compared to the nominal MLE from [6], which was developed for the case of no attack, and the corresponding CRLB derived in the previous section. We can see that the proposed MLE provides a better performance under the assumption of an attack probability $p_a \sim \mathcal{U}(0, \alpha = 0.03)$ and $\sigma_2^2 \sim \mathcal{U}(\sigma_1^2, \rho = 100)$, than the MLE in [6] which is not aware of possible attacks, and the simulation results show that the estimator is able to converge to the CRLB in all the cases even with a relatively small number of sensors.



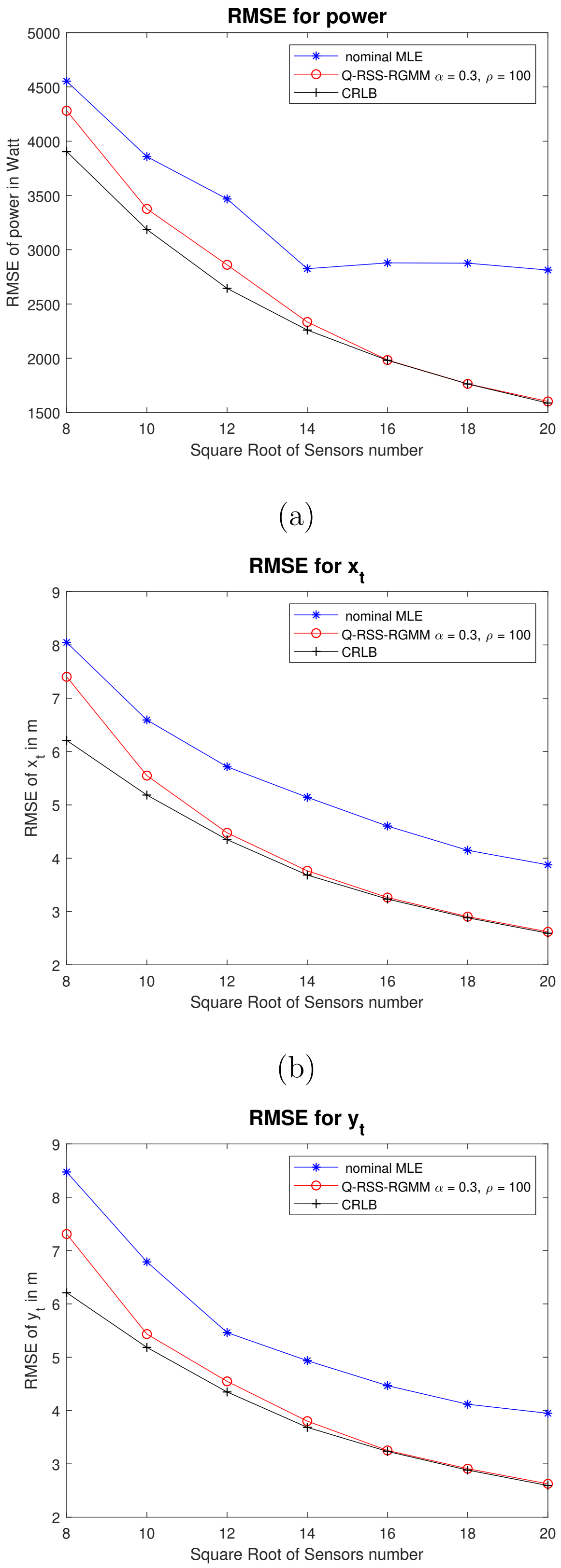
(a)

(b)

(c)

Fig. 6. The RMSE vs. square root of number of sensors $\alpha = 0.03$

In Fig.7, we increase the upper limit for the attack probability to be $p_a \sim \mathcal{U}(0, \alpha = 0.3)$, and we notice from the simulation results that the proposed MLE performance is robust and can approach the CRLB as the number of sensors increases, and it outperforms the nominal MLE which is unaware of the attacks and provides a very poor performance with the increased attack probability p_a.



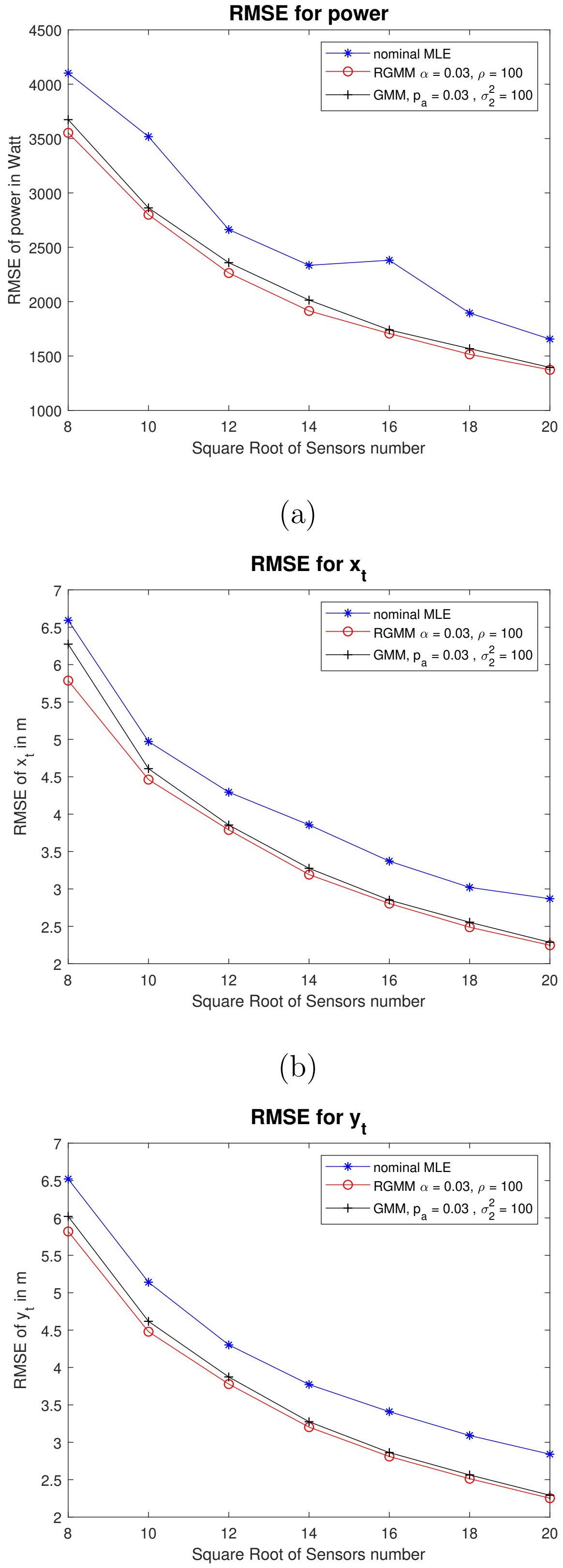
(a)

(b)

(c)

Fig. 7. The RMSE vs. square root of number of sensors $\alpha = 0.3$

In Fig.8 and Fig.9, we illustrate the RMSEs of different MLE algorithms including the nominal MLE, MLE-QRSS-GMM, and MLE-QRSS-RGMM algorithms. The sensor data have been generated according to RGMM model. We can notice from Fig.8 that at small value of p_a the RGMM's performance is slightly better than the GMM's performance, and both algorithms outperform the nominal MLE. In Fig.9, when p_a for the MLE-QRSS-GMM incresed into $p_a = 0.3$ and the α for the MLE-QRSS-RGMM is also increased into $\alpha = 0.3$, we notice that the MLE - QRSS - RGMM has a much better performance, then the MLE - QRSS - GMM which provides an acceptable performance it will still give better results than the nominal case.

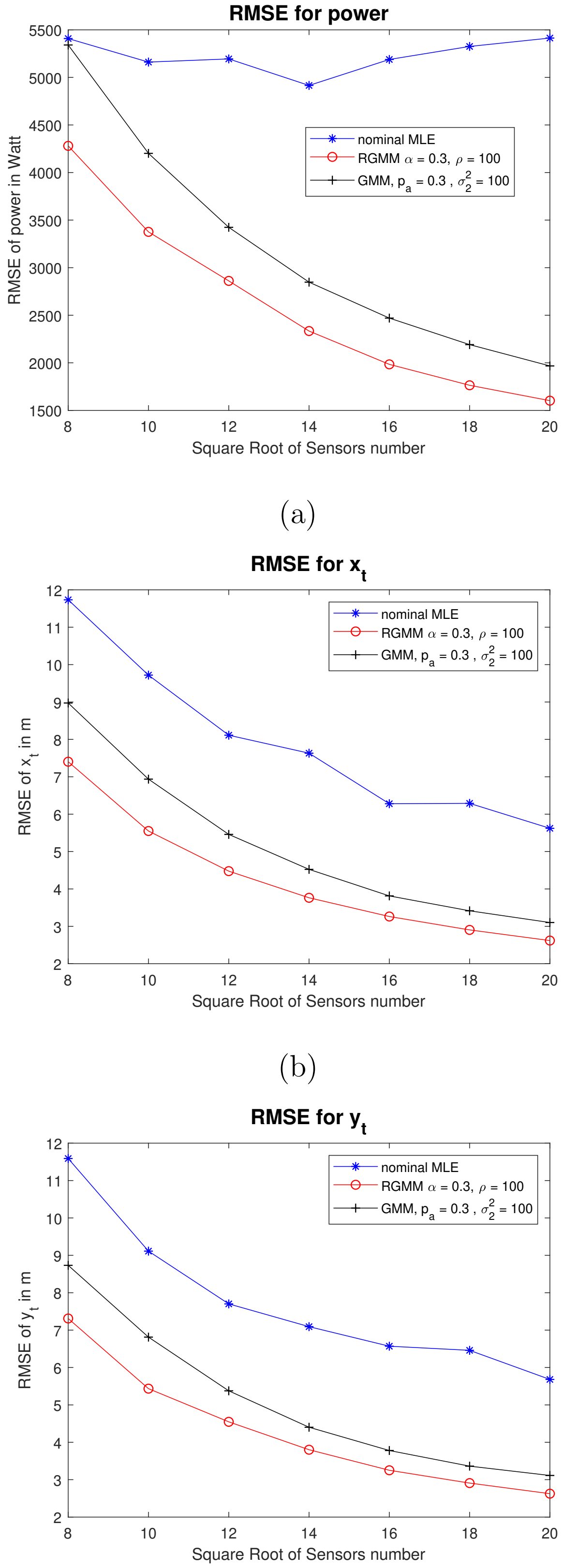


(a)

(b)

(c)

Fig. 8. The RMSE vs. square root of number of sensors $p_a = 0.03, \alpha = 0.03$.



(a)

(b)

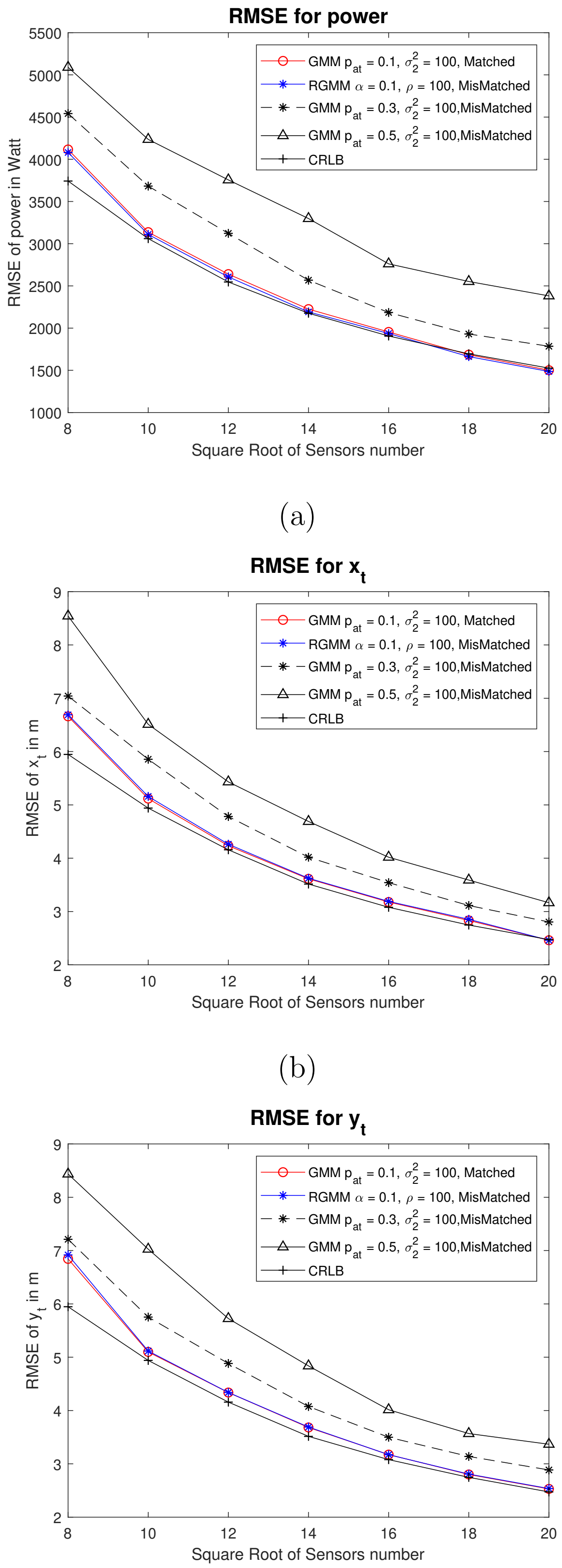
(c)

Fig. 9. The RMSE vs. square root of number of sensors $p_a = 0.3, \alpha = 0.3$.

We also make an assumption of a mismatch between the measurements and the algorithm design parameters to evaluate the algorithm's robustness. The simulation parameters are set as $p_a = 0.1$, and $\sigma_2^2 = 100$ for the original data generation. We have designed algoritms with the follwing parameters, for the 1st algorithm, an MLE-QRSS-GMM, $p_a = 0.1$ and $\sigma_2^2 = 100$. This algorithm is considered as the matched algorithm. The other two algorithms are MLE-QRSS-GMM algorithms with $p_a = 0.3$ and $p_a = 0.5$ respectively, and $\sigma_2^2 = 100$ for both of them. The 4th algorithm is designed according to the MLE-QRSS-RGMM assumption with $p_a \sim U(0, 0.1)$, and $\sigma_2^2 \sim U(\sigma_1^2, 100)$. The CRLB here is calculated for the MLE-QRSS-GMM with matched parameters. In Fig.10, the simulation results are shown for the mismatched situation. We can notice that the MLE-QRSS-RGMM performance is almost identical to the MLE-QRSS-GMM with the parameters $p_a = 0.1$ and $\sigma_2^2 = 100$, while the MLE-QRSS-GMM algorithms with $p_a = 0.3$ and $p_a = 0.5$ give poor performances. It is obvious that the MLE - QRSS -RGMM algorithm is giving a very robust performance under the mismatch situation.

In Fig.11, the simulation parameters are set as $p_a = 0.2$, and $\sigma_2^2 = 100$ for the original data generation. Here, we have also designed four algoritms with the follwing parameters, for the 1st algorithm, an MLE-QRSS-GMM, $p_a = 0.2$ and $\sigma_2^2 = 100$, and this algorithm is considered as the matched algorithm. The other two algorithms are MLE-QRSS-GMM algorithms with $p_a = 0.3$ and $p_a = 0.5$ respectively, and $\sigma_2^2 = 200$ for both of them. The 4th algorithm is designed according to the MLE-QRSS-RGMM assumption with $p_a \sim U(0, 0.4)$, and $\sigma_2^2 \sim U(\sigma_1^2, 200)$. We can see that even without the knowledge of p_a and σ_2^2, the MLE-QRSS-RGMM algorithm will still give an excellent performance and it coincides with the matched algorithm performance. On the other hand, the other mismatched algorithms with $p_a = 0.3$ and $p_a = 0.5$, provide poor performances in this case.

In summary, the proposed MLE-QRSS-RGMM algorithm provides a very robust performance for the cases where p_a and σ_2^2 are unknown, and only their ranges are known.

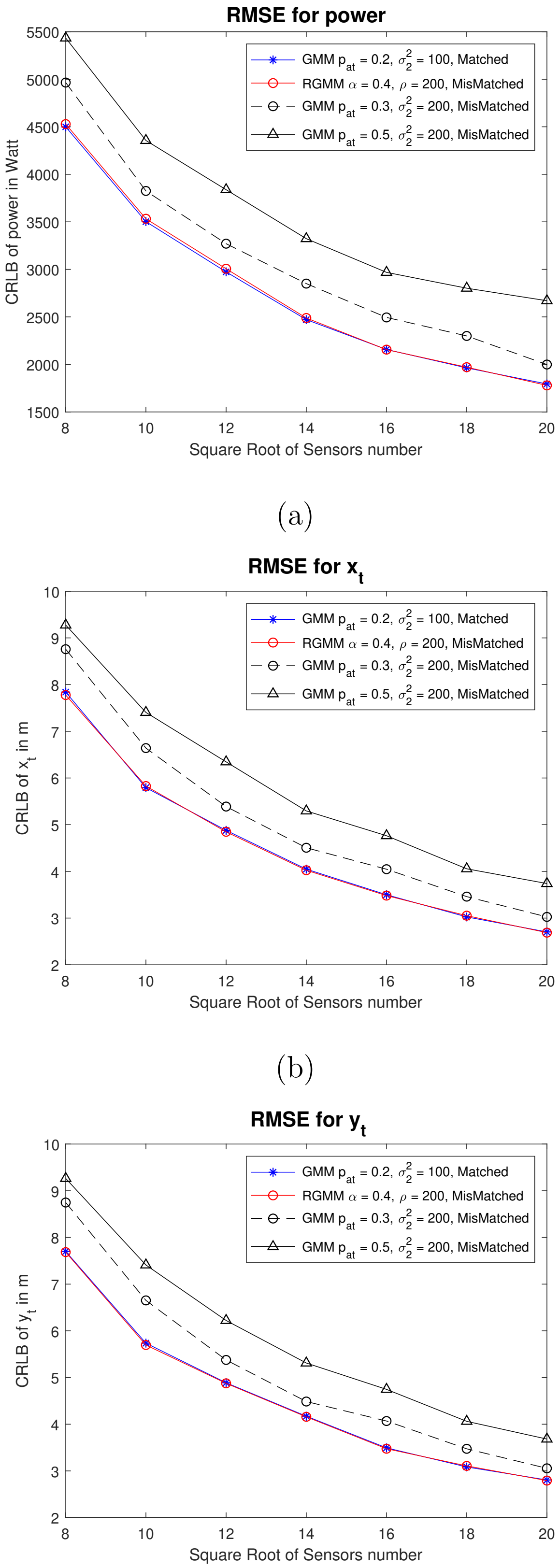


(a)

(b)

(c)

Fig. 10. The RMSE vs. square root of number of sensors for MisMatched assumption.



(a)

(b)

(c)

Fig. 11. The RMSE vs. square root of number of sensors for MisMatched assumption.

CHAPTER 5

MINIMAX BAYESIAN ESTIMATOR UNDER FALSE INFORMATION INJECTION ATTACKS

In this Chapter, we investigate the problem of false information injection attack on the linear Bayesian estimator. We formulated the relationship between the attacker and the defender as a minimax optimization problem. The attacker tries to maximize the cost function by controlling the attack power. On the other hand, the defender tries to minimize the attack effect by optimizing the detection threshold selection. we develop numerical solution for the minimax problem.

5.1 Mathematical Model

Assume that we have N sensors . The general linear and Gaussian system measurements can be modeled as

$$z_i = \boldsymbol{H}_i x + w_i,\ i = 1, ..., N \tag{5.1}$$

where $\boldsymbol{H}_i$ is the measurement matrix, $w_i \sim \mathcal{N}(0, \sigma_w^2)$, and $\mathbf{x}$ is the state vector. Assume that $w_i's$ are $i.i.d$ across the sensors, and b_i is an injected sensor bias by an adversary, then

$$z_i' = \mathbf{H}_i x + w_i + b_i = z_i + b_i \tag{5.2}$$

where z_i' is the corrupted sensor measurement, b_i is the injected false information which can be a deterministic or random bias.

Let $\boldsymbol{z} = [z_1, ..., z_N]^T$ is the total sensor measurement vector, $\boldsymbol{b} = [b_1, ..., b_N]^T$ is the vector includes the injected noise at all the sensors. Similarly let $\boldsymbol{H} = [\boldsymbol{H}_1^T, ..., \boldsymbol{H}_N^T]^T$ be the matrix measurement, and $\boldsymbol{w} = [w_1, ..., w_N]^T$ be the measurement noise vector. Then (5.1) and (5.2) can be written as

$$\begin{aligned} \boldsymbol{z} &= \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w} \\ \boldsymbol{z'} &= \boldsymbol{z} + \boldsymbol{b} \end{aligned} \tag{5.3}$$

Let $\boldsymbol{P}_{ww} = E[\boldsymbol{w}\boldsymbol{w}^T]$ be the covariance matrix of the measurements error. Next, we will formulate the relationship between the attacker an the defender as a *minimax* optimization problem for Bayesian estimator under false information attacks.

5.2 False Information Injection with Deterministic Bias

5.2.1 Mathematical Model

In this section we assume the attacker attacks with a constant deterministic bias $\boldsymbol{b}$, the two hypotheses can be modeled as

$\mathrm{H}_0 : \boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}$

$\mathrm{H}_1 : \boldsymbol{z'} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w} + \boldsymbol{b}$

Under H_0 hypothesis where no attack exists we have :

$$\boldsymbol{z} = \boldsymbol{H}\ \boldsymbol{x} + \boldsymbol{w} \tag{5.4}$$

where $\boldsymbol{H}$ is the known measurement matrix, $\boldsymbol{x} \sim \mathcal{N}(\overline{\boldsymbol{x}}, \boldsymbol{P}_{xx})$ is the state vector, $\boldsymbol{P}_{xx}$ and $\boldsymbol{P}_{ww}$ are covariance matrices for $\boldsymbol{x}$ and $\boldsymbol{w}$.

Then, we assume that the system is using a Chi-square detector to detect the corrupted measurement. The Chi square detector uses the following test statistic

$$t = (\boldsymbol{z} - \overline{\boldsymbol{z}})^T \boldsymbol{S}^{-1} (\boldsymbol{z} - \overline{\boldsymbol{z}}) \tag{5.5}$$

where $\overline{\boldsymbol{z}} = \boldsymbol{H}\overline{\boldsymbol{x}}$ is the measurement mean, and $\boldsymbol{S} = \boldsymbol{P}_{zz,H_0}$, is the covariance matrix for measurement $\boldsymbol{z}$ under H_0 :

$$\boldsymbol{P}_{zz,H_0} = \boldsymbol{S} = E[(\boldsymbol{z} - \overline{\boldsymbol{z}})(\boldsymbol{z} - \overline{\boldsymbol{z}})^T] = \boldsymbol{H}\boldsymbol{P}_{xx}\boldsymbol{H}^T + \boldsymbol{P}_{ww} \tag{5.6}$$

Now, the probability of false alarm can be derived as

$$P_r(D_1|H_0) = P_r(t \geq \psi|H_0) = 1 - \chi^2_{n_z}(\psi) \tag{5.7}$$

where $\chi^2_{n_z}$ is the CDF of a Chi square distributed random variable with n_z degrees of freedom, and ψ is the threshold for the detector.

$$P_r(D_0|H_0) = P_r(t < \psi|H_0) = 1 - P_r(D_1|H_0) = \chi^2_{n_z}(\psi) \tag{5.8}$$

Now, we have derived the probability of false alarm under H_0 hypothesis, where D_1 is the detector decision that there is an attack, and D_0 is the decision that there is no attack.

Under H_1 we have :

$$\boldsymbol{z}' = \boldsymbol{H}\ \boldsymbol{x} + \boldsymbol{w} + \boldsymbol{b} \tag{5.9}$$

where $\boldsymbol{b}$, is a deterministic injected bias.

We need to find the probabilities of detection and miss under H_1 hypothesis. Assume $\boldsymbol{\gamma} \sim \mathcal{N}(0, \boldsymbol{P}_{zz})$, where $\boldsymbol{\gamma} = (\boldsymbol{z} - \overline{\boldsymbol{z}})$, and $\boldsymbol{P}_{zz} = E[(\boldsymbol{z} - \bar{\boldsymbol{z}})(\boldsymbol{z} - \bar{\boldsymbol{z}})^T] = \boldsymbol{H}\boldsymbol{P}_{xx}\boldsymbol{H}^T + \boldsymbol{P}_{ww}$. Then a new parameter can be defined as $\boldsymbol{\gamma}' = (\boldsymbol{\gamma} + \boldsymbol{b})$. Let

$$t = \boldsymbol{\gamma}'^T \boldsymbol{S}^{-1} \boldsymbol{\gamma}' = \boldsymbol{\gamma}'^T \boldsymbol{P}_{zz}^{-1} \boldsymbol{\gamma}' \tag{5.10}$$

where $\boldsymbol{\gamma}' \sim \mathcal{N}(\boldsymbol{b}, \boldsymbol{P}_{zz})$, (5.10) can also be written as

$$t = (\boldsymbol{\gamma} + \boldsymbol{b})^T \boldsymbol{P}_{zz}^{-1} (\boldsymbol{\gamma} + \boldsymbol{b}) \tag{5.11}$$

Since $\boldsymbol{P}_{zz}^{-1}$ can be decomposed as $\boldsymbol{P}_{zz}^{-1} = \boldsymbol{P}_{zz}^{-\frac{1}{2}} \boldsymbol{P}_{zz}^{-\frac{1}{2}}$, (5.11) can be rewritten as

$$t = (\boldsymbol{\gamma} + \boldsymbol{b})^T \boldsymbol{P}_{zz}^{-\frac{1}{2}} \boldsymbol{P}_{zz}^{-\frac{1}{2}} (\boldsymbol{\gamma} + \boldsymbol{b}) \tag{5.12}$$

Therefore, we have

$$t = (\boldsymbol{P}_{zz}^{-\frac{1}{2}} (\boldsymbol{\gamma} + \boldsymbol{b}))^T (\boldsymbol{P}_{zz}^{-\frac{1}{2}} (\boldsymbol{\gamma} + \boldsymbol{b})) \tag{5.13}$$

which can be also written as

$$t = (\boldsymbol{P}_{zz}^{-\frac{1}{2}} \boldsymbol{\gamma}')^T (\boldsymbol{P}_{zz}^{-\frac{1}{2}} \boldsymbol{\gamma}') \tag{5.14}$$

Now, a new parameter can be defined as $\boldsymbol{\omega} = \boldsymbol{P}_{zz}^{-\frac{1}{2}}\boldsymbol{\gamma}' = \boldsymbol{P}_{zz}^{-\frac{1}{2}}(\boldsymbol{\gamma}+\boldsymbol{b})$. Then (5.14) can also be described as

$$t = \boldsymbol{\omega}^T\boldsymbol{\omega} \tag{5.15}$$

Therefore, the mean and the covariance of the parameter $\boldsymbol{\omega}$ can be calculated as

$$E[\boldsymbol{\omega}] = \boldsymbol{P}_{zz}^{-\frac{1}{2}}\boldsymbol{b} = \overline{\boldsymbol{\omega}} \tag{5.16}$$

and

$$E[(\boldsymbol{\omega} - \boldsymbol{P}_{zz}^{-\frac{1}{2}}\boldsymbol{b})(\boldsymbol{\omega} - \boldsymbol{P}_{zz}^{-\frac{1}{2}}\boldsymbol{b})^T] = E[\boldsymbol{P}_{zz}^{-\frac{1}{2}}\boldsymbol{\gamma}\boldsymbol{\gamma}^T\boldsymbol{P}_{zz}^{-\frac{1}{2}}] = \boldsymbol{I} \tag{5.17}$$

where the fact that $E[\boldsymbol{\gamma}\boldsymbol{\gamma}^T] = \boldsymbol{P}_{zz}$ has been used. Let $\boldsymbol{r} = \boldsymbol{P}_{zz}^{-\frac{1}{2}}\boldsymbol{b}$, where $\boldsymbol{P}_{zz}^{-\frac{1}{2}}$ is a $n_z \times n_z$ and $\boldsymbol{b}$ is a $n_z \times 1$ vector.

Now we have

$$\boldsymbol{\omega} \sim \mathcal{N}(\boldsymbol{r}, \boldsymbol{I}) \tag{5.18}$$

where $\boldsymbol{I}$ is the $n_z \times n_z$ identity matrix. So we have

$$\begin{bmatrix} \omega_1 \sim \mathcal{N}(r_1, 1) \\ \omega_2 \sim \mathcal{N}(r_2, 1) \\ . \\ . \\ . \\ \omega_N \sim \mathcal{N}(r_N, 1) \end{bmatrix} \tag{5.19}$$

Since

$$t = (\boldsymbol{\omega})^T(\boldsymbol{\omega}) = \sum_{i=1}^{N} \omega_i^2 \tag{5.20}$$

it can be shown that, t follows a Non-Central Chi squared distribution with n_z degrees of freedom, with non-centrality parameter [33, 34]

$$\lambda = \sum_{i=1}^{N} r_i^2 = ||\boldsymbol{r}||^2 \tag{5.21}$$

where λ is the non centrality parameter. Hence, the probability of detection can be written as

$$P(D_1|H_1) = Pr(t \geq \psi) = Q_{\frac{k}{2}}(\sqrt{\lambda}, \sqrt{\psi}) \tag{5.22}$$

where k is the degrees of freedom which is in our assumption n_z, and $Q_M(a, b)$ is the Marcum Q-function [34]

$$Q_M(a,b) = \int_b^\infty x \left(\frac{x}{a}\right)^{M-1} e^{-\frac{x^2+a^2}{2}} I_{M-1}(ax)\, dx \tag{5.23}$$

where $I_{M-1}(.)$ is the modified Bessel function of order $M-1$. Note that the probability of miss is simply

$$P(D_0|H_1) = 1 - P(D_1|H_1) \tag{5.24}$$

Now, we have found the probability of detection $P(D_1|H_1)$, and miss detection probability $P(D_0|H_1)$, where D_1 is the detector decision that there is an attack, and D_0 is the decision that there is no attack.

We adopt the detection and discard strategy of [13], which was shown as a robust strategy in [13]. This can be explained in the following table

Scenarios	The response
1- if $D_1|H_1$ Correct detection	Discard the sensors' data
2- if $D_0|H_1$ Miss detection - Mismatch	Use the sensors' data
3- if $D_1|H_0$ – False alarm	Discard the sensors' data
4- if $D_0|H_0$ No false alarm	Use the sensors' data

where in $D_1|H_1$ is the case when the detector indicates an attack. In this case the response for the detector will be to discard all the received corrupted measurements and the cost will be determined by the prior information $\boldsymbol{P}_{xx}$. In the $D_0|H_1$ scenario, the detector will indicate that there is no attack and it will incorporate the corrupted data in the system, and the cost function in this scenario will be a combination of the prior information and the corrupted data.

In $D_1|H_0$, the detector will indicate an attack under H_0 hypothesis, which is the no attack hypothesis. The response for this scenario will be discarding the measurement and keep only the prior information, and the cost function will be determined by the prior information $\boldsymbol{P}_{xx}$. In $D_0|H_0$, the detector will indicate that there is no attack under H_0 hypothesis. In this case, the cost will be determined by the traditional MMSE. the average cost function has been provided by [13] as

$$\begin{aligned} c = [&P(H_1)P(D_1|H_1)\,\mathrm{c}_1 \\ &+P(H_1)P(D_0|H_1)\,\mathrm{c}_2 \\ &+P(H_0)P(D_1|H_0)\,\mathrm{c}_3 \\ &+P(H_0)P(D_0|H_0)\,\mathrm{c}_4] \end{aligned} \tag{5.25}$$

where c_i as $i \in [1, 2, 3, 4]$ are the traces of the mean squared error matrices, $P(D_i|H_j)$ where $i, j \in \{0, 1\}$ are the detection, miss detection, false alarm, and no false alarm probabilities. $P(H_0)$ and $P(H_1)$ are the prior probabilities of attack and no attack, respectively.

The cost elements have been derived in [13] as :

$$c_1 = c_3 = Tr[\boldsymbol{P}_{xx}] \tag{5.26}$$

and

$$c_4 = Tr[\boldsymbol{P}_{xx} - \boldsymbol{P}_{xz,H_0}\boldsymbol{P}^{-1}_{zz,H_0}\boldsymbol{P}_{zx,H_0}] \tag{5.27}$$

where c_4 is the trace of the conditional covariance of the MMSE estimator given by (2.15) in Chapter 2 .

In [13], the injected bias was assumed to be a random Gaussian vector. In contrast to [13], we assume the injected noise $\boldsymbol{b}$ to be a deterministic vector. So we derive c_2 in the cost function as follows. First we have

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}} + \boldsymbol{P}_{xz}\boldsymbol{P}_{zz,H_0}^{-1}(\boldsymbol{z} - \bar{\boldsymbol{z}} + \boldsymbol{b}) \tag{5.28}$$

where $\hat{\boldsymbol{x}}$ is the Bayesian MMSE estimator, and $\boldsymbol{P}_{xz,H_0} = \boldsymbol{P}_{zx,H_0}^T = \boldsymbol{P}_{xx}H^T$. Let $\tilde{\boldsymbol{x}} = (\boldsymbol{x} - \bar{\boldsymbol{x}})$ and $\tilde{\boldsymbol{z}} = (\boldsymbol{z} - \bar{\boldsymbol{z}})$, then

$$(\boldsymbol{x} - \hat{\boldsymbol{x}}) = (\tilde{\boldsymbol{x}} - \boldsymbol{P}_{xz}\boldsymbol{P}_{zz,H_0}^{-1}\tilde{\boldsymbol{z}} - \boldsymbol{P}_{xz}\boldsymbol{P}_{zz,H_0}^{-1}\boldsymbol{b}) \tag{5.29}$$

Now $E[(\boldsymbol{x} - \hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}})^T]$

$$E[(\boldsymbol{x}-\hat{\boldsymbol{x}})(\boldsymbol{x}-\hat{\boldsymbol{x}})^T] = E[(\tilde{\boldsymbol{x}}-\boldsymbol{P}_{xz}\boldsymbol{P}_{zz,H_0}^{-1}\tilde{\boldsymbol{z}}-\boldsymbol{P}_{xz}\boldsymbol{P}_{zz,H_0}^{-1}\boldsymbol{b})(\tilde{\boldsymbol{x}}^T-\tilde{\boldsymbol{z}}^T\boldsymbol{P}_{zz,H_0}^{-1}\boldsymbol{P}_{zx}-\boldsymbol{b}^T\boldsymbol{P}_{zz,H_0}^{-1}\boldsymbol{P}_{zx})] \tag{5.30}$$

Thus

$$E[(\boldsymbol{x} - \hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}})^T] = \boldsymbol{P}_{xx} - \boldsymbol{P}_{xz,H_0}\boldsymbol{P}_{zz,H_0}^{-1}(I - [\boldsymbol{b}\boldsymbol{b}^T]\boldsymbol{P}_{zz,H_0}^{-1})\boldsymbol{P}_{zx} \tag{5.31}$$

where the fact that $E[\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^T] = \boldsymbol{P}_{xx}$, $E[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}^T] = \boldsymbol{P}_{zz,H_0}$, and $E[\boldsymbol{b}\boldsymbol{b}^T] = \boldsymbol{b}\boldsymbol{b}^T$ has been used,thus under the mismatch case the cost can be calculated as the trace of (5.31) as :

$$c_2 = Tr[\boldsymbol{P}_{xx} - \boldsymbol{P}_{xz,H_0}\boldsymbol{P}_{zz,H_0}^{-1}(I - [\boldsymbol{b}\boldsymbol{b}^T]\boldsymbol{P}_{zz,H_0}^{-1})\boldsymbol{P}_{zx}] \tag{5.32}$$

Since $P(D_1|H_1)$,$P(D_0|H_1)$, $P(D_1|H_1)$, and $P(D_0|H_0)$ are functions of ψ, the cost function defined in (5.25) is a function of both ψ and the attacker injected bias $\boldsymbol{b}$. So we have formulated the relationship between the detector threshold ψ and the attacker injected bias $\boldsymbol{b}$ as a minimax optimization problem as follows

$$\begin{aligned} \arg\min_{\boldsymbol{\psi}}\max_{\boldsymbol{b}\boldsymbol{b}^T}[P(H_1)P(D_1|H_1)\,\mathrm{c}_1 \\ +P(H_1)P(D_0|H_1)\,\mathrm{c}_2 \\ +P(H_0)P(D_1|H_0)\,\mathrm{c}_3 \\ +P(H_0)P(D_0|H_0)\,\mathrm{c}_4] \\ s.t.\;\; \boldsymbol{b}^T\boldsymbol{b} = a^2 \end{aligned} \tag{5.33}$$

where the attacker is trying to maximize the attack effect by changing $\boldsymbol{b}\boldsymbol{b}^T$, while the defender is trying to minimize the attack effect by changing ψ, where $\boldsymbol{b}^T\boldsymbol{b}$ is the attack power, which is under constraint a^2. Next, some numerical results for this minimax problem are shown.

5.2.2 Numerical Results for Deterministic False Information Injection

In this section, we present the numerical results for the attack with a deterministic bias on the Bayesian estimator. The attacker bias will be a deterministic vector $\boldsymbol{b} = [b_1\ b_2]^T$, and the false information power is $b_1^2 + b_2^2 = a^2$, where $a^2 \in [0, 500]$. Let us assume that $b_1^2 = \kappa a^2$, and κ is a power allocation parameter. The attack matrix $\boldsymbol{b}\boldsymbol{b}^T$ can be written as

$$\boldsymbol{b}\boldsymbol{b}^T = \begin{bmatrix} \kappa a^2 & \upsilon\sqrt{\kappa(1-\kappa)}a^2 \\ \upsilon\sqrt{\kappa(1-\kappa)}a^2 & (1-\kappa)a^2 \end{bmatrix} \tag{5.34}$$

where $\upsilon \in \{-1, 1\}$, the prior information will be $\boldsymbol{x} \sim \mathcal{N}(\overline{\boldsymbol{x}}, \boldsymbol{P}_{xx})$, with $\overline{\boldsymbol{x}} = [5\ 10]^T$, and $\boldsymbol{P}_{xx} = \begin{bmatrix} 50 & 0 \\ 0 & 50 \end{bmatrix}$. $\boldsymbol{w} \sim \mathcal{N}(0, \boldsymbol{P}_{ww})$ is the Gaussian noise with $\boldsymbol{P}_{ww} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ and $\overline{\boldsymbol{w}} = [0\ 0]^T$ (the mean of the measurements noise is zero). $\boldsymbol{H}$ is the measurements matrix which is a 2×2 identity matrix. The prior probability of attack p_1 and no attack p_0 are set as $p_1 = p_0 = 0.5$.

The cost function of (5.25) and the probabilities of detection and miss detection given in (5.22) and (5.24) are calculated using non-Central Chi squared assumption. The probabilities of false alarm is calculated by (5.7). The cost function elements $[c_1, c_2, c_3, c_4]$ are calculated according to the equations (5.26), (5.27), and (5.32). The detection threshold ψ will be in the range of $\psi \in [0, 30]$.

We have used grid search algorithm to find the combination of (κ, υ) that maximizes the cost function for multiple combinations of $(\psi, (\boldsymbol{b}^T\boldsymbol{b}) = a^2)$ values. It always holds that choosing $\kappa = 1$, or the total power being allocated to the first state element, will achieve maximum cost for the multiple combinations of thresholds ψ and $(\boldsymbol{b}^T\boldsymbol{b} = a^2)$. So, in our numerical results we will always set the value of $\kappa = 1$. When $\kappa = 1$, the value of υ will not affect the cost function since the total power will be assigned to the first element of matrix $\boldsymbol{b}\boldsymbol{b}^T$. Therefore, we will always set the value of $\upsilon = 1$ in our numerical results.

In Fig.12, the cost function versus the attacker power a^2 and the detection threshold ψ is shown. We can notice that, when the threshold value is $\psi = 0$, the cost function will be equal to $c = 100$, which is a reasonable outcome since when $\psi = 0$, both probabilities of detection and false alarm will be one, and the cost function will be combination of c_1, and c_3 which is $Tr[\boldsymbol{P}_{xx}]$. When ψ value increases, we notice three behaviours from the figure with respect to the attacker power based on Fig.12. First, when the attacker power is low, i.e lower than or equal to 100, then the cost function will decrease with the increase in the detector threshold. This means that when the injected bias is weak, then incorporating the observed sensors' measurements will reduce the estimation error and improve the over all system performance under low power injected bias. When the attacker power is in between 100 to 300 then the cost function will have an interesting result which is a decrease in the range from 0 to some value of threshold lower than 5, and then it starts increasing, that means incorporating the observed sensors' measurements to a certain limit will improve the system performance and after that threshold the performance will be degraded by incorporating the sensors' data.

The third region is when the attacker power is greater than 300 then the cost function will start at 100 and it will be increasing with the increase of the threshold, that means that when the injected bias power is greater than 300, then it is better for the detector to discard the sensors' data. It could be also noted that for a high value of threshold then the cost function will be increasing linearly with the increase in the attacker power. This is because when the injected bias is strong, it is better to discard the corrupted sensor data.

In Fig.13, the three regions of the cost function are illustrated, where we notice that at first sub-figure, the cost function will keep decreasing with increasing the threshold when a^2 is small. In the second sub-figure, we notice the decrements in the cost function for the range of threshold between 0 and some value lower than 5, and it will start increasing afte that. The third sub-figure shows the case when the attacker power is large enough then the cost function will keep increasing with the increase of the detection threshold.

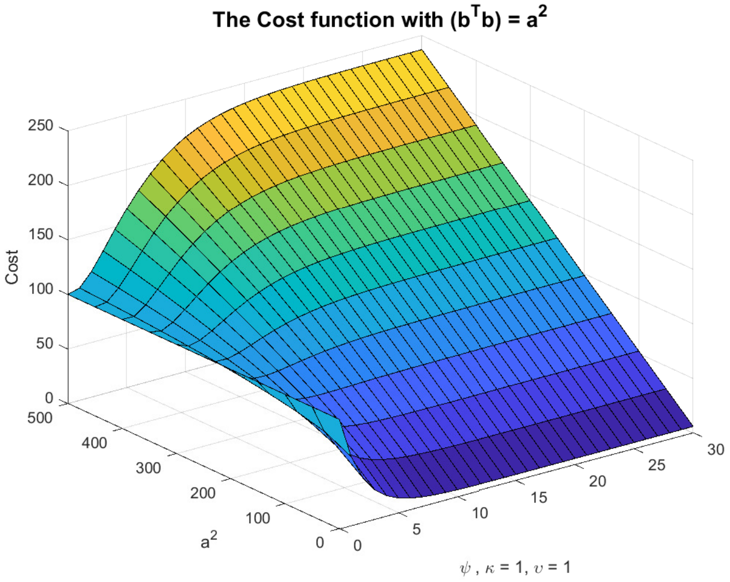
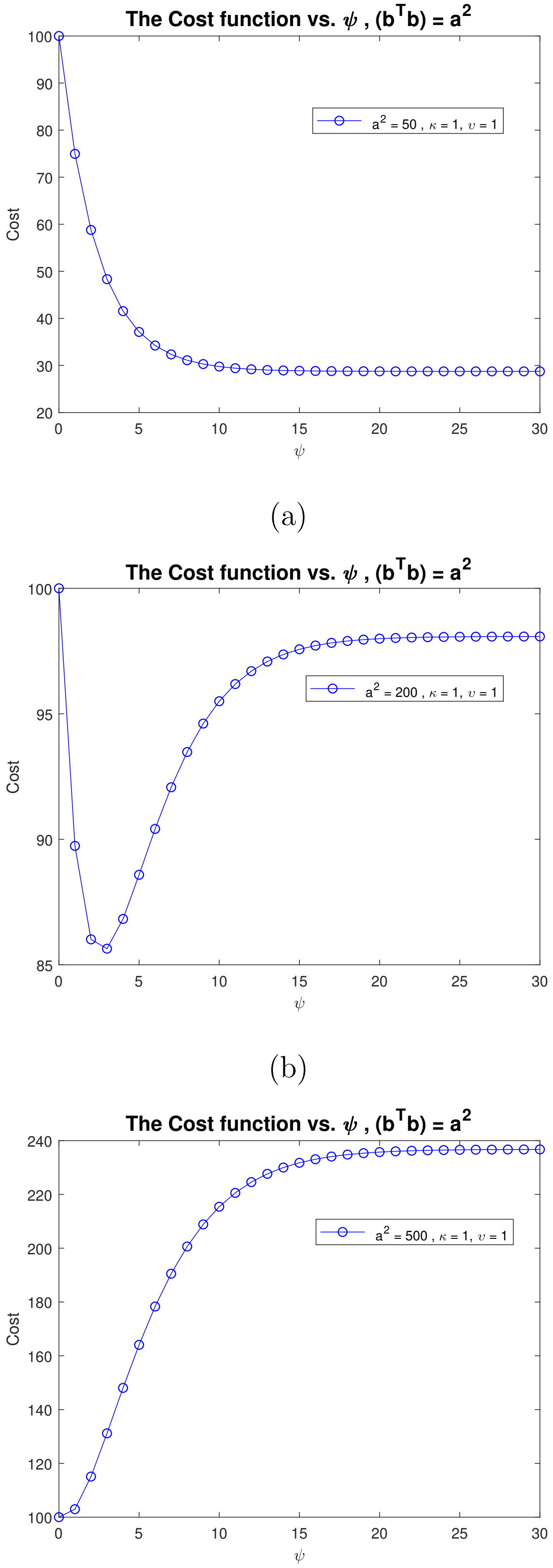


Fig. 12. The Cost function for attack with deterministic bias.



(a)

(b)

(c)

Fig. 13. The cost function for deterministic bias vs. detection threshold ψ, attacker power $a^2 = 50, 200, 500$.

5.3 False Information Injection with Random Bias

5.3.1 Mathematical Model

Assume the attacker is attacking with a random bias $\boldsymbol{b}$ that has a Gaussian distribution $\boldsymbol{b} \sim \mathcal{N}(\bar{\boldsymbol{b}}, \boldsymbol{P}_{bb})$, where $\bar{\boldsymbol{b}}$ and $\boldsymbol{P}_{bb}$ are the mean and the covariance matrix for the injected bias respectively.

Under H_0, which is the no attack hypothesis, the probabilities of false alarm and its complement can be calculated as in (5.7) and (5.8), based on a Chi square distribution with n_z degrees of freedom.

Under H_1 hypothesis, where the sensor measurements will be corrupted by injected bias $\boldsymbol{b}$ as

$$\boldsymbol{z} = \boldsymbol{H}\ \boldsymbol{x} + \boldsymbol{w} + \boldsymbol{b} \tag{5.35}$$

Hence, the probability of detection can be calculated as

$$P(D_1|H_1) = \int_{b_2}\int_{b_1} P_d(\boldsymbol{b})P_b(\boldsymbol{b})db_1db_2 \tag{5.36}$$

according to the non-central Chi squared distribution then $P_d(\boldsymbol{b})$ can be described as

$$P_d(\boldsymbol{b}) = Q_{\frac{k}{2}}(\sqrt{\lambda}, \sqrt{\psi}) \tag{5.37}$$

where $Q_M(a,b)$ is the Marcum function with M degrees of freedom [34], $\lambda = ||\boldsymbol{P}_{zz}^{-\frac{1}{2}}\boldsymbol{b}||^2$, and $\boldsymbol{b} \sim \mathcal{N}(0, \mathbf{P_{bb}})$ with

$$P_{\boldsymbol{b}}(\boldsymbol{b}) = |2\pi \boldsymbol{P}_{bb}|^{-\frac{1}{2}} e^{-\frac{1}{2}\boldsymbol{b}^T \boldsymbol{P}_{bb}^{-1} \boldsymbol{b}} \tag{5.38}$$

Therefore, the probability of detection can be calculated as

$$P(D_1|H_1) = \int_{b_2} \int_{b_1} Q_{\frac{k}{2}}(\sqrt{\lambda}, \sqrt{\psi}) \; |2\pi \boldsymbol{P}_{bb}|^{-\frac{1}{2}} e^{-\frac{1}{2}\boldsymbol{b}^T \boldsymbol{P}_{bb}^{-1} \boldsymbol{b}} db_1 db_2 \tag{5.39}$$

and the probability of miss is

$$P(D_0|H_1) = 1 - P(D_1|H_1) \tag{5.40}$$

Solving the above integrals numerically we get the $P(D_1|H_1)$ and $P(D_0|H_1)$.

Now, cost function parameters of $[c_1, c_2, c_3]$ can be calculated as given in (5.26) and (5.27), and under the mismatch case the cost has been derived in [4] as the $c_2 = Tr(E[(\hat{\boldsymbol{x}} - \boldsymbol{x})(\hat{\boldsymbol{x}} - \boldsymbol{x})^T])$ which is

$$\begin{aligned} c_2 &= Tr[\boldsymbol{P}_{xx} - \boldsymbol{P}_{xz,H_0} \boldsymbol{P}_{zz,H_0}^{-1} \boldsymbol{P}_{zx,H_0} + \boldsymbol{P}_{xz,H_0} \boldsymbol{P}_{zz,H_0}^{-1} \boldsymbol{P}_{bb} \boldsymbol{P}_{zz,H_0}^{-1} \boldsymbol{P}_{zx,H_0}] \\ &= Tr[\boldsymbol{P}_{xx} - \boldsymbol{P}_{xz,H_0} \boldsymbol{P}_{zz,H_0}^{-1} (\boldsymbol{I} - \boldsymbol{P}_{bb} \boldsymbol{P}_{zz,H_0}^{-1}) \boldsymbol{P}_{zx,H_0}] \end{aligned} \tag{5.41}$$

where $\boldsymbol{P}_{xz,H_0} = \boldsymbol{P}_{zx,H_0}^T = \boldsymbol{P}_{xx} \boldsymbol{H}^T$.

Now, we have found all the terms of the cost function. The attack probability is assumed to be known and it is set as $P(H_0) = p_0 = 0.5$ and $P(H_1) = p_1 = 0.5$.

Next, the relationship between the attacker and the defender can be formulated as a minimax optimization problem over $C(\psi, \boldsymbol{P}_{bb})$, as

$$
\begin{aligned}
\arg\min_{\boldsymbol{\psi}}\max_{\boldsymbol{P_{bb}}}[&P(H_1)P(D_1|H_1)\,\mathrm{c}_1\\
&+P(H_1)P(D_0|H_1)\,\mathrm{c}_2\\
&+P(H_0)P(D_1|H_0)\,\mathrm{c}_3\\
&+P(H_0)P(D_0|H_0)\,\mathrm{c}_4]\\
&s.t.\ Tr[\boldsymbol{P}_{bb}]=a^2
\end{aligned}
\tag{5.42}
$$

where the attacker is trying to maximize the attack effect by changing $\boldsymbol{P}_{bb}$, while the defender is trying to minimize the attack effect by changing ψ. Here, the detection and discard strategy will be used as in the deterministic case, where the defender will discard the sensors' data under $D_1|H_1$ and $D_1|H_0$ as explained earlier. Next, some numerical results for this minimax problem will be presented.

5.3.2 Numerical Results

In this section, we present the numerical results for the attack with a random bias on the Bayesian estimator. The attacker bias will be a random Gaussian vector $\boldsymbol{b}\sim\mathcal{N}(\overline{\boldsymbol{b}},\boldsymbol{P}_{bb})$, where $\overline{\boldsymbol{b}}=[0\ 0]^T$ and $\boldsymbol{P}_{bb}=\begin{bmatrix}\sigma_{1b}^2 & \rho\sigma_{1b}\sigma_{2b}\\ \rho\sigma_{1b}\sigma_{2b} & \sigma_{2b}^2\end{bmatrix}$, $\rho\in[-1,1]$ and the false information power is $\sigma_{1b}^2+\sigma_{2b}^2=a^2$, where $a^2\in[0,500]$, $\sigma_{1b}^2=\kappa a^2$, and κ is the power allocation parameter.

The prior information will be $\boldsymbol{x}\sim\mathcal{N}(\overline{\boldsymbol{x}},\boldsymbol{P}_{xx})$, with $\overline{\boldsymbol{x}}=[5\ 10]^T$, and $\boldsymbol{P}_{xx}=\begin{bmatrix}50 & 0\\ 0 & 50\end{bmatrix}$. $\boldsymbol{w}\sim\mathcal{N}(0,\boldsymbol{P}_{ww})$ is the Gaussian noise with $\boldsymbol{P}_{ww}=\begin{bmatrix}2 & 0\\ 0 & 4\end{bmatrix}$ and $\overline{\boldsymbol{w}}=[0\ 0]^T$, and $\boldsymbol{H}$ is the measurements matrix which is a 2×2 identity matrix. The prior probabilities of attack p_1 and no attack p_0 are set as $p_1=p_0=0.5$.

The cost function of (5.42) and the probabilities of detection and miss detection given in (5.39) and (5.40) are calculated using numerical integration function in Matlab. The probabilities of false alarm is calculated by (5.7). The cost function elements $[c_1, c_2, c_3, c_4]$ are calculated according to (5.26), (5.27), and (5.41).

We have used grid search algorithm to investigate the combination of (κ, ρ) that maximizes the cost function for multiple combinations of $(\psi, Tr(\boldsymbol{P}_{bb}) = a^2)$ values. It always holds that choosing $\kappa = 1$, or equivalently the total power being allocated to the first state element, will achieve maximum cost for the multiple combinations of thresholds ψ and $(Tr(\boldsymbol{P}_{bb}) = a^2)$. So, in our numerical results we will always set the value of $\kappa = 1$. $\kappa = 1$ means that the value of ρ will not affect the cost function, since the total power will be allocated to the first element of $\boldsymbol{P}_{bb}$ matrix. Therefore, we will always set the $\rho = 1$ in our numerical results.

In Fig.14, the cost function is plotted as a function of the detection threshold $\psi = [0 : 1 : 30]$ and the attacker power $a^2 = [0 : 50 : 500]$, with the power allocation parameter $\kappa = 1$ that means $\sigma_{b1}^2 = a^2$, and the correlation coefficient is set as $\rho = 1$. It is clear from Fig.14 that there are three regions in which the cost function changes differently. The first region is when the power $a^2 \in [0, 100]$ in this region the cost will decrease with increasing ψ until it reaches to its minimum value when $\psi = 30$. This means when the power of the injected bias is low, it is better for the defender to use the sensors' measurements.

The second region lies in the range of $a^2 = [150, 300]$, in this region the cost function will decrease at the beginning in the range of ψ from 0 to a value lower than 5. Then when ψ goes beyond 5 the cost function will increase until it reaches its maximum when $\psi = 30$. Here, incorporating the sensors' measurements below certain threshold will improve the estimation performance.

The third region lies in the range of $a^2 = [350, 500]$, in this region the cost function will increase with the increase of the value of the threshold until it reaches its maximum when $\psi = 30$. In this case the cost function will reach its maximum value. In this case, the best strategy for the defender is to reject the sensors' data and only use the prior information available for the state vector.

Also, we can note from Fig.14, that the cost function will increase with the increase of attacker power when the value of threshold greater than 10.

In Fig.15 $a,\ b,\ and\ c$, we plot the three regions of the cost function with respect to detection threshold ψ and the values of $a^2 = 50,\ 200,\ and\ 500$ respectively. In Fig.15, it is clear that the cost function will be increasing in general with respect to the increase in the attacker power, and there are some regions in which the increase in cost function will be controlled by the detection threshold when $a^2 \in [0, 300]$, and when a^2 crosses the value of 300 then the cost function will keep increasing with the increase of attacker power.

We can notice from the numerical results that the best strategy for the attacker is to attack with a highest power available a^2 since increasing the attack power will generally increase the cost function. From the defender's perspective, if there is information about the attacker power, then the best strategy for the defender is to select the optimal threshold to minimze the attacker effect. But if the defender has no information about the attacker power then the best solution for the detector is to discard the corrupted data, and use only the prior information.

5.4 The Solution for the Minimax Problem

In this section, we solve the minimax optimization problems in (5.33) and (5.42). By solving these problems, we will know what the optimal strategy is for the defender when it is under the worst false information injection attack, if it has a knowledge about the attacker power a^2. On the other hand, we will know which the best strategy is for the attacker to use if it has knowledge about the defender detection threshold.

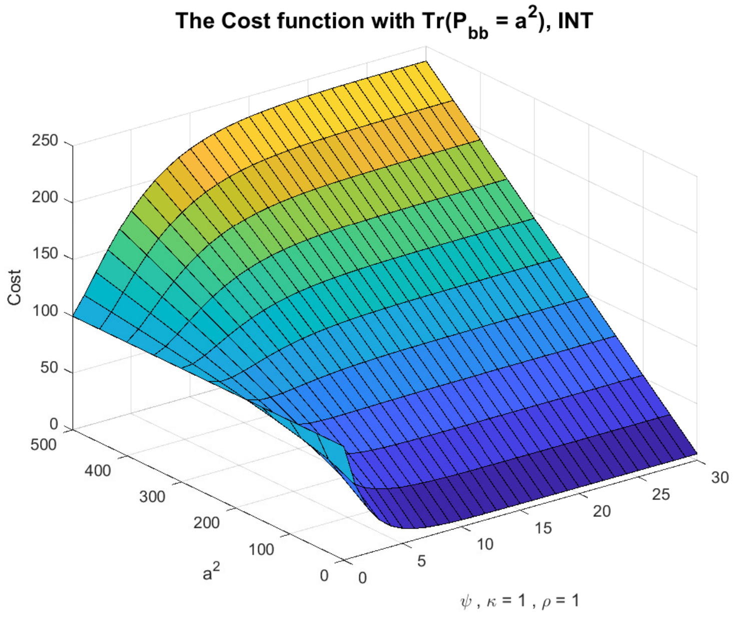
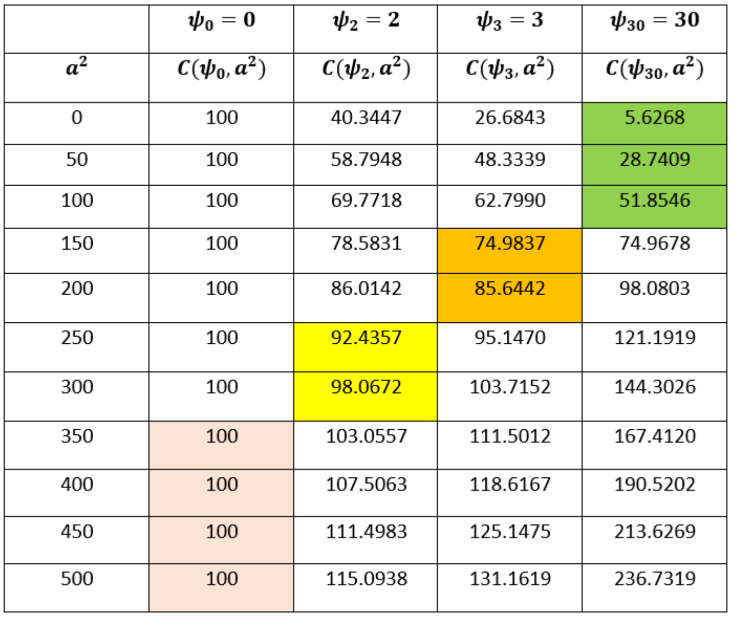


Fig. 14. The Cost function for attack with random bias INT.

Here, we denote to the numerical integration approach as $\boldsymbol{INT}$. Tables 5.1 and 5.2 show the numerical results for the solutions of minimax problems in (5.33) and (5.42) for different values of attacker power a^2 versus the defender's detection threshold ψ. In these tables, we have assumed that the attacker knows the threshold ψ, and it attacks with the worst possible strategy by setting $\kappa = 1$, meaning that it allocates all the power to the first state element.

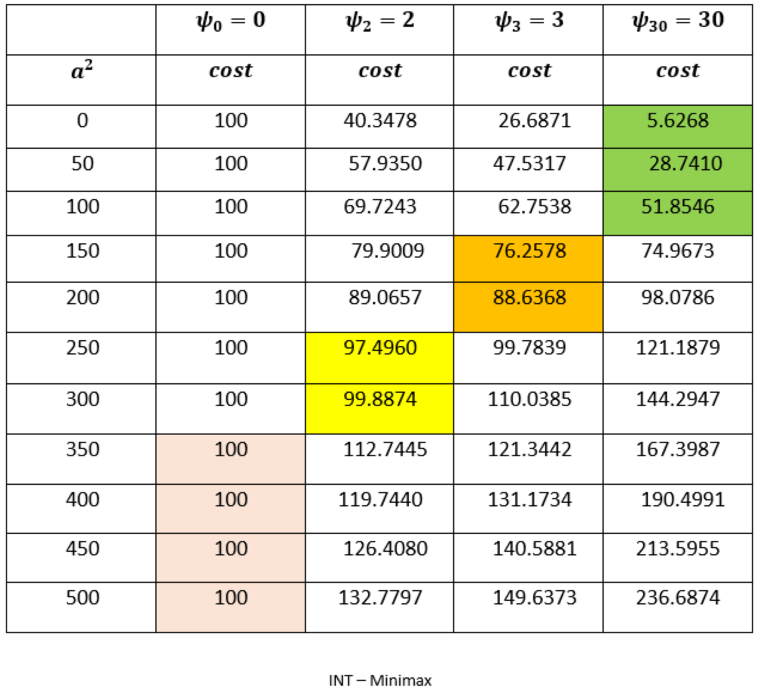
In Tables 5.1 and 5.2, the results of the attack strategy with the deterministic and random bias $\boldsymbol{b}$ are shown by solving (5.33) and (5.42), respectively. The attacker power is determined by a^2, which is $\boldsymbol{b}^T\boldsymbol{b} = a^2$ for the deterministic case and $Tr[\boldsymbol{P}_{bb}] = a^2$ for the random case. Then we maximize the cost function in (5.33) and (5.42) with respect to a^2 instead of $\boldsymbol{b}\boldsymbol{b}^T$ and $\boldsymbol{P}_{bb}$ matrices, where $a^2 \in [0 : 50 : 500]$. On the other hand, for the defender and based on the discussion done for Fig.12 and Fig.14, we have selected four values of the thresholds $\psi = \{0, 2, 3, 30\}$, which have been shown to achieve the minimum cost function results with respect to a^2.



	$\psi_0 = 0$	$\psi_2 = 2$	$\psi_3 = 3$	$\psi_{30} = 30$
a^2	$C(\psi_0, a^2)$	$C(\psi_2, a^2)$	$C(\psi_3, a^2)$	$C(\psi_{30}, a^2)$
0	100	40.3447	26.6843	5.6268
50	100	58.7948	48.3339	28.7409
100	100	69.7718	62.7990	51.8546
150	100	78.5831	74.9837	74.9678
200	100	86.0142	85.6442	98.0803
250	100	92.4357	95.1470	121.1919
300	100	98.0672	103.7152	144.3026
350	100	103.0557	111.5012	167.4120
400	100	107.5063	118.6167	190.5202
450	100	111.4983	125.1475	213.6269
500	100	115.0938	131.1619	236.7319

Det – Minimax

Table 5.1. The solution to the minimax problem with deterministic bias, $(\boldsymbol{b}^T\boldsymbol{b}) = a^2$

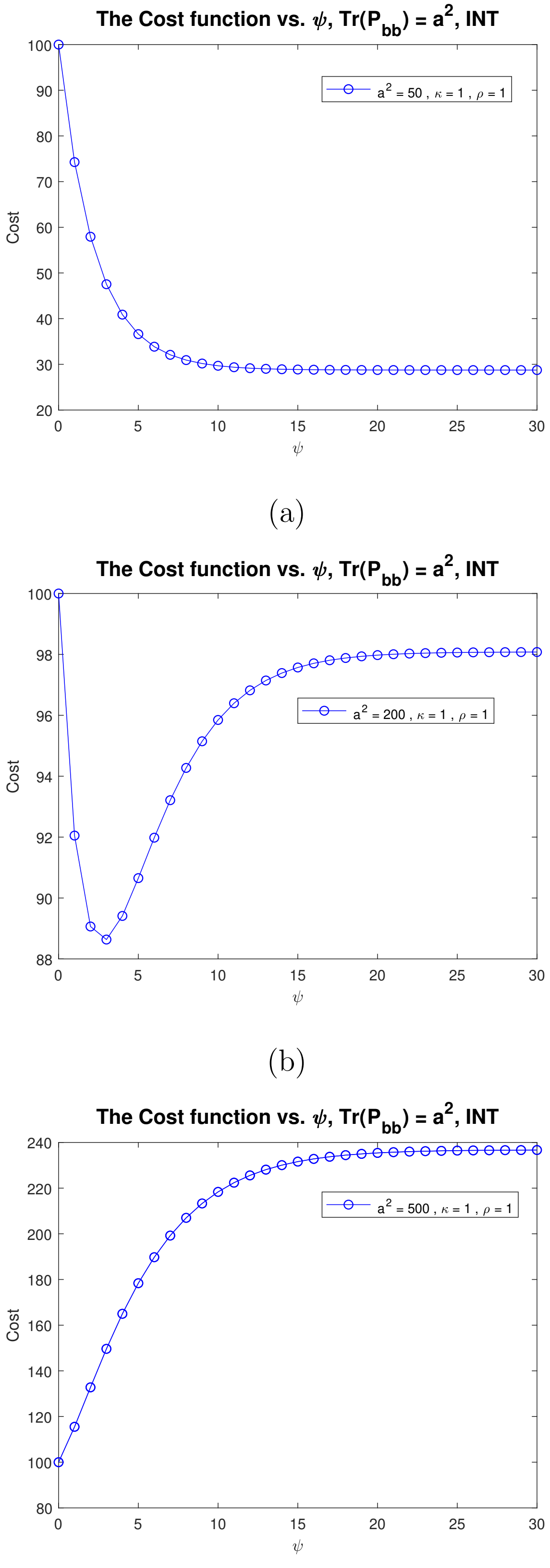


	$\psi_0 = 0$	$\psi_2 = 2$	$\psi_3 = 3$	$\psi_{30} = 30$
a^2	$cost$	$cost$	$cost$	$cost$
0	100	40.3478	26.6871	5.6268
50	100	57.9350	47.5317	28.7410
100	100	69.7243	62.7538	51.8546
150	100	79.9009	76.2578	74.9673
200	100	89.0657	88.6368	98.0786
250	100	97.4960	99.7839	121.1879
300	100	99.8874	110.0385	144.2947
350	100	112.7445	121.3442	167.3987
400	100	119.7440	131.1734	190.4991
450	100	126.4080	140.5881	213.5955
500	100	132.7797	149.6373	236.6874

INT – Minimax

Table 5.2. The solution to the minimax problem with for random bias, $Tr(\boldsymbol{P}_{bb}) = a^2$

If the defender has knowledge of the attacker power, and the attacker power is $0 \leq a^2 \leq 100$, the optimum threshold for the detector will be $\psi = 30$, which will achieve the minimum cost value. Next, if the attacker power is $150 \leq a^2 \leq 200$, the optimal threshold for the defender will be $\psi = 3$, that will minimize the worst cost function. We can also notice that when the attacker power is $250 \leq a^2 \leq 300$, the optimum threshold will be $\psi = 2$. Finally, if the attacker power is $350 \leq a^2 \leq 500$, the optimal threshold for the defender will be $\psi = 0$, which means that the to minimize the cost function the defender needs to reject the corrupted sensor measurements. On the other hand, if the defender has no knowledge of the attacker power, then the optimum threshold for the defender will be $\psi = 0$, regardless of the attack power a^2 that the attacker uses.



(a)

(b)

(c)

Fig. 15. The cost function for random bias vs. detection threshold ψ, attacker power $a^2 = 50, 200, 500$.

CHAPTER 6

CONCLUSION

In this dissertation, we have investigated the problem of secure localization under a false information injection attack. We have considered the localization system based on quantized received signal strength. We have proposed two estimators based on the QRSS data in WSNs. The first estimator is a maximum likelihood estimator based on QRSS in which we have addressed the problem of false information injection using a Gaussian mixture model. We have derived the maximum likelihood estimator and its corresponding CRLB. Simulation results show that the prposed estimator is able to outperform the nominal MLE, which is unaware of the attack, and it gives a very robust performance even with a large probability of attack. We have discussed the problem of the mismatched measurements, and we have shown that the MLE-QRSS-GMM is robust in the mismatch situation but it requires the knowledge of attack probability and attack power.

Thus, another estimator was proposed and it is also based on the Gaussian mixture assumption. This time we assumed that the probability of the attack and the power of the attack are both random variables following certain uniform distributions. Starting from such assumptions we derived the mathematical model and the relevant distribution for the proposed problem. We derived the MLE-QRSS-RGMM estimator and its corresponding CRLB. Simulation results showed that the MLE-QRSS-RGMM estimator provides an excellent and robust performance, without the knowledge of the attack power and probability.

Next, the problem of Bayesian estimation subjected to false information injection attacks is considered. We have formulated the relationship between the attacker and the defender from a minimax perspective, in which the attacker tries to maximize the system estimation error by controlling the bias vector, whereas the defender tries to minimize the system estimation error by optimally selecting the detection threshold.

We assumed two different scenarios. First, we assumed the attacker will attack with a deterministic bias. For this assumption we have derived the probabilities of detection and miss detection, based on non-central Chi squared distribution. Secondly, it is assumed that the attacker will attack with a random bias which follows a Gaussian distribution with some known prior. For this assumption we have derived the probabilities of detection and miss detection, by using numerical integration.

The minimax problem was numerically solved for both the cases with the deterministic and random biases. Numerical results showed that if the defender has prior knowledge of the attacker power, it will be possible to select the optimum detection threshold that minimizes the worst possible cost function accordingly. On the other hand, if the defender has no prior knowledge of the attacker's power, then the best strategy is to always reject the corrupted sensors' measurements.

REFERENCES

[1] B. Alnajjab and R. S Blum. “After-attack performance of parameter estimation systems”. In: *Proc. 2014 48th Annual Conference on Information Sciences and Systems (CISS).* Princeton, NJ, USA, Mar. 2014, pp. 1–6.

[2] J. Zhang et al. “Asymptotically Optimum Distributed Estimation in the Presence of Attacks.” In: *IEEE Trans. Signal Processing* 63.5 (2015), pp. 1086–1101.

[3] J. Zhang et al. “Functional forms of optimum spoofing attacks for vector parameter estimation in quantized sensor networks”. In: *IEEE Transactions on Signal Processing* 65.3 (2017), pp. 705–720.

[4] J. Lu and R. Niu. “False information injection attack on dynamic state estimation in multi-sensor systems”. In: *17th International Conference on Information Fusion (FUSION).* IEEE. 2014, pp. 1–8.

[5] J. Zhang, R. S Blum, and H V. Poor. “Approaches to Secure Inference in the Internet of Things: Performance Bounds, Algorithms, and Effective Attacks on IoT Sensor Networks”. In: *IEEE Signal Processing Magazine* 35.5 (2018), pp. 50–63.

[6] R. Niu and P. K Varshney. “Target location estimation in sensor networks with quantized data”. In: *IEEE Transactions on Signal Processing* 54.12 (2006), pp. 4519–4528.

[7] M. Al-Salman and R. Niu. “Source Location with Quantized Sensor Data Corrupted by False Information”. In: *2018 21st International Conference on Information Fusion (FUSION).* IEEE. 2018, pp. 391–397.

[8] R. Niu, A. Vempaty, and P. K Varshney. “Received-signal-strength-based localization in wireless sensor networks”. In: *Proceedings of the IEEE* 106.7 (2018), pp. 1166–1182.

[9] J. Zhang et al. “Attack detection in sensor network target localization systems with quantized data”. In: *IEEE Transactions on Signal Processing* (2018).

[10] J. H. Lee and R M. Buehrer. “Characterization and detection of location spoofing attacks”. In: *Journal of Communications and Networks* 14.4 (2012), pp. 396–409.

[11] K. Agrawal et al. “Target localization in wireless sensor networks with quantized data in the presence of byzantine attacks”. In: *Proc. 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR).* Pacific Grove, CA, USA, Nov. 2011, pp. 1669–1673.

[12] B. Alnajjab, J. Zhang, and R. S Blum. “Attacks on sensor network parameter estimation with quantization: Performance and asymptotically optimum processing”. In: *IEEE Transactions on Signal Processing* 63.24 (2015), pp. 6659–6672.

[13] R. Niu and J. Lu. “False information detection with minimum mean squared errors for Bayesian estimation”. In: *2015 49th Annual Conference on Information Sciences and Systems (CISS).* IEEE. 2015, pp. 1–6.

[14] J. Zhang and R. S Blum. “Distributed joint spoofing attack identification and estimation in sensor networks”. In: *Proc. 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP).* Chengdu, China, July 2015, pp. 701–705.

[15] A. Vempaty, Y. S Han, and P. K Varshney. “Target localization in wireless sensor networks using error correcting codes in the presence of Byzantines”. In: *Proc. 2013 IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Vancouver, CA, Oct. 2013, pp. 5195–5199.

[16] Z. Li et al. “Robust statistical methods for securing wireless localization in sensor networks”. In: *Proceedings of the 4th international symposium on Information processing in sensor networks*. IEEE Press. 2005, p. 12.

[17] F. Yin et al. “Cooperative localization in WSNs using Gaussian mixture modeling: Distributed ECM algorithms”. In: *IEEE Transactions on Signal Processing* 63.6 (2015), pp. 1448–1463.

[18] F. Yin et al. “Robust cooperative sensor network localization via the EM criterion in LOS/NLOS environments”. In: *2013 IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE. Darmstadt, Germany, June 2013, pp. 505–509.

[19] P. Pfaff, C. Plagemann, and W. Burgard. “Gaussian mixture models for probabilistic localization”. In: *Proc. 2008. ICRA IEEE International Conference on Robotics and Automation*. IEEE. Pasadena, CA, USA, May 2008, pp. 467–472.

[20] Y. Zhang et al. “Base station localization in search of empty spectrum spaces in cognitive radio networks”. In: *Mobile Ad-hoc and Sensor Networks, 2009. MSN’09. 5th International Conference on*. IEEE. Fujian, China, Dec. 2009, pp. 94–101.

[21] J. N Ash and R. L Moses. “Outlier compensation in sensor network self-localization via the EM algorithm”. In: *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP’05). IEEE International Conference on.* Vol. 4. IEEE. Philadelphia, PA, USA, Mar. 2005, pp. 749–752.

[22] Y. Zhang et al. "RSS-based localization in WSNs using Gaussian mixture model via semidefinite relaxation". In: *IEEE Communications Letters* 21.6 (2017), pp. 1329–1332.

[23] R. Niu and L. Huie. "System state estimation in the presence of false information injection". In: *Statistical Signal Processing Workshop (SSP), 2012 IEEE.* IEEE. 2012, pp. 385–388.

[24] G. Casella and R. L. Berger. *Statistical inference.* Vol. 2. Duxbury Pacific Grove, CA, 2002.

[25] Y. Bar-Shalom, X R. Li, and T. Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software.* John Wiley & Sons, 2004.

[26] E. Masazade et al. "Energy aware iterative source localization for wireless sensor networks". In: *IEEE Transactions on Signal Processing* 58.9 (2010), pp. 4824–4835.

[27] H.L. Trees Van, K Bell, Z Tiany, et al. "Detection Estimation and Modulation Theory". In: *Detection, Estimation, and Filtering Theory.* Wiley & Sons, Inc., 2013.

[28] P. Tichavsky, C. H. Muravchik, and A. Nehorai. "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering". In: *IEEE Transactions on signal processing* 46.5 (1998), pp. 1386–1396.

[29] Y. Zheng, R. Niu, and P. K Varshney. "Closed-form performance for location estimation based on quantized data in sensor networks". In: *Proc. 2010 13th IEEE Conference on Information Fusion (FUSION).* Edinburgh, UK, July 2010, pp. 1–7.

[30] X. Sheng and Yu-Hen Hu. “Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks”. In: *IEEE Transactions on Signal Processing* 53.1 (2005), pp. 44–53.

[31] L.E. Kinsler and A.R. Frey. *Fundamentals of Acoustics.* New York, NY: John Wiley & Sons, Inc., 1962.

[32] N. Levanon. *Radar Principles.* New York, NY: John Wiley & Sons, Inc., 1988.

[33] D. A. Shnidman. “The calculation of the probability of detection and the generalized Marcum Q-function”. In: *IEEE Transactions on Information Theory* 35.2 (1989), pp. 389–400.

[34] J. G. Proakis and M. Salehi. *Digital communications.* Vol. 4. McGraw-hill New York, 2001.