ROBUST INFERENCE IN WIRELESS SENSOR NETWORKS

by

Santosh Paudel

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Boise State University

August 2022

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Santosh Paudel

Thesis Title: Robust Inference in Wireless Sensor Networks

Date of Final Oral Examination: 5 April 2022

The following individuals read and discussed the dissertation submitted by student Santosh Paudel, and they evaluated the student's presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Hao Chen, Ph.D.	Chair, Supervisory Committee
John Chiasson, Ph.D.	Member, Supervisory Committee
Leming Qu, Ph.D.	External Examiner

The final reading approval of the dissertation was granted by Hao Chen, Ph.D., Chair of the Supervisory Committee. The dissertation was approved by the Graduate College.

DEDICATION

dedicated to all the people in my life who encourage, support, and guide me

ACKNOWLEDGMENT

I am using this opportunity to express my sincere gratitude to my advisor Dr.Hao Chen for his support, guidance, and friendship during the last five years. I would like to express my appreciation to him for belief in me from the very beginning and patient guidance on my research during the course of my Ph.D. study. Dr.Chen is one of the sharpest people I know with great mathematical intuition. I am thankful for him for being a mentor as a caring friend and also giving me great freedom and encouragement to manage an internship as a Data Scientist at a crucial time in my career.

I am grateful to Dr.John Chiasson, and Dr.Leming Qu for serving in my committee, and would like to thank them for encouragement and insightful comments. Thanks for Dr.Chiasson's courses that helped me to understand probability and statistics, and machine learning algorithms from scratch. Furthermore, I would like to thank Becton, Dikinson and Company (BD), and manager Satya Allumallu, Healthsight Inventory Optimization (HSIO) team, for providing an internship opportunity to me for nine months during my graduate study which significantly helped me to gain skills and explore more in the new and upcoming Data Science era.

I would like to thank my colleagues Mojtaba Ahmadi Almasi and Ruthvik Vialla for their support. I am really fortunate to have friends Fnu Tala, Sohil Rana, Md.Kamrul Hassan, and Ziyang Liang during the course of my doctoral study for

their company especially during outdoor activities. Finally, I want to thank the administrative and IT staffs in the ECE department who tirelessly work to keep the department running smoothly.

This dissertation is dedicated to my family for their unconditional love, complete support and belief in me. Without them, this would not have been possible.

ABSTRACT

This dissertation presents a systematic approach to obtain robust statistical inference schemes in unreliable networks. Statistical inference offers mechanisms for deducing the statistical properties of unknown parameters from the data. In Wireless Sensor Networks (WSNs), sensor outputs are transmitted across a wireless communication network to the fusion center (FC) for final decision-making. The sensor data are not always reliable. Some factors may cause anomaly in network operations, such as malfunction, corruption, or compromised due to some unknown source of contamination or adversarial attacks.

Two standard component failure models are adopted in this study to describe the system vulnerability: the probabilistic and static models. In probabilistic models, we consider a widely known $\varepsilon-$contamination model, where each node has ε probability of malfunctioning or being compromised. In contrast, the static model assumes there is up to a certain number of malfunctioning nodes. It is assumed that the decision center/network operator is aware of the presence of anomaly nodes and can adjust the operation rule to counter the impact of the anomaly. The anomaly node is assumed to know that the network operator is taking some defensive actions to improve its performance. Considering both the decision center (network operator) and compromised (anomalous) nodes and their possible actions, the problem is formulated as a two-player zero-sum game. Under this setting, we attempt to discover the worst possible failure models and best possible operating strategies.

First, the effect of sensor unreliability on detection performance is investigated, and robust detection schemes are proposed. The aim is to design robust detectors when some observation nodes malfunction. The detection problem is relatively well known under the probabilistic model in simple binary hypotheses testing with known saddle-point solutions. The detection problem is investigated under the mini-max framework for the static settings as no such saddle point solutions are shown to exist under these settings.

In the robust estimation, results in estimation theory are presented to measure system robustness and performance. The estimation theory covers probabilistic and static component failure models. Besides the standard approaches of robust estimation under the frequentist settings where the interesting parameters are fixed but unknown, the estimation problem under the Bayes settings is considered where the prior probability distribution is known. After first establishing the general framework, comprehensive results on the particular case of a single node network are presented under the probabilistic settings. Based on the insights from the single node network, we investigate the robust estimation problem for the general network for both failure models. A few robust localization methods are presented as an extension of robust estimation theory at the end.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CLLRD Clipped Log-Likelihood Ratio Detector

FC Fusion Center

IRLS Iteratively Reweighted Least Sqaures

LR Likelihood Ratios

LTS Least Trimmed Square

MIMO Multiple Input Multiple Output

ML Maximum Likelihood

MSE Mean Square Error

RCS Radar Cross Section

ROC Receiver Operating Curves

TDOA Time Delay of Arrival

TOA Time of Arrival

WSNs Wireless Sensor Networks

CHAPTER 1:

INTRODUCTION

1.1 Wireless Sensor Networks (WSNs)

Recent advances in wireless communications and electronics have enabled the development of low-cost, low-power, multifunctional sensors that are small in size and can communicate within a short distance. A sensor network consists of sensing, data processing, and communicating components and can accomplish various tasks, including environment monitoring, surveillance, target localization, tracking, and many more [1, 2, 3, 4]. WSNs are widely employed in many applications such as military applications, environmental monitoring, cyber-physical systems, healthcare, diagnostics of complex systems, and so on. It has been the focus of multiple disciplinary research in the past several decades [5, 6, 7, 8, 9, 10, 11].

1.2 Applications of WSNs

Statistical inference in WSNs is a field of study that attempts to determine or estimate a state of nature based on observations regarding that state. It includes detecting phenomena, estimating parameters, or measuring some physical properties of the environment, where sensors are densely deployed to the region of interest [1, 4]. Many WSNs have a dedicated Fusion Center (FC) with more potent computational capability than other sensing nodes because of data processing requirements. The traditional inference problem occurs when all the observations are available at an FC, and will be referred to as centralized inference. A key attribute of centralized inference is that the decision procedure is only applied once for a given set of observations at the FC. Due to the bandwidth and power constraints, sensor data are often needed to be compressed or preporcessed prior to communication, and the FC only have access to received data. This is commonly known as decentralized inference or, in some cases, distributed inference [3].

The problems of inference share much in common with many centralized statistical inference and learning problems such as signal detection and estimation, dimension reduction, and feature extraction [8, 15]. But unlike centralized inference, an essential attribute of decentralized inference is that decision-making typically occurs at multiple locations and layers, where these so-called local decisions are fused at a central FC to make the final decision [12, 13, 14]. The multitude of decision locations and layers introduces coupling among the various decision processes, which dramatically increases complexity regarding the system design and optimization.

We summarize a few applications of WSNs.

1. Environmental monitoring: Temperature monitoring, flood detection, and geophysical research [16, 17, 18]. For example, forest fire detection, where many sensors are densely deployed to a forest to sense weather conditions, including temperature, rain, and relative humidity.

2. Health-related applications: Tracking patients' physiological conditions, movements, and behaviors. For this purpose, patients usually wear different types of wireless sensors to collect data on body conditions [19]. For real-time applications such as monitoring of patients, sensors transmit to the FC securely in

real-time. For offline decision-making, such as future medical diagnostics, and drug administration in hospitals, sensors collect data for a long time and then securely transmit the data to the FC.

3. Autonomous transportation: Radar networks potentially provide highly accurate object detection and localization (range, velocity, and angle) [20, 21]. Thus, they are being increasingly integrated into commercial vehicles as sensors for environmental perception in future (semi) autonomous driver assistance operational modes. In particular, side-looking radars are widely used to support lane change, or keeping assist, blind spot detection, and rear cross-traffic alert. For these applications, a high-resolution radar image by using distributed radar module is key to effective separation of close objects, detection of the spatial extension of traffic participants, and enhanced object recognition [22, 23, 24].

4. Military applications: In the military, the rapid deployment, self-organization, and fault tolerance characteristics of sensor networks make them a very promising sensing technique for military command, control, communications, computing, intelligence, surveillance, and targeting systems, for example, battlefield surveillance and reconnaissance of opposing forces. WSNs also can be deployed to detect, localize and track targets. Moreover, they can be used to assess damage conditions, monitoring equipment, and ammunition [25].

5. A cyber-physical system: WSNs play an important role in sensing and providing information for such systems including smart grids and nuclear power plants [26, 27]. The sensor nodes in such systems are deeply intertwined physical and software components, each operating on different spatial and temporal scales,

exhibiting multiple and distinct behavioral modalities [28]. For smart grids, many sensors are distributed for monitoring long-range power transmission lines to improve transmission efficiency, reliability, and sustainability [29, 30].

1.3 Robustness issues in WSNs

Many problems encountered in engineering practice rely on some assumptions well justified in many situations. This enables a simple derivation of optimal processing. Nominal optimality, however, is useless if it was derived under nominal assumptions that do not hold in practice.

Suppose a signal processing scheme, say a detector for a signal with a known waveform in additive noise, is designed to give optimum performance for noise possessing a specific statistical description, i.e., noise as a Gaussian process [31]. The critical question is, how sensitive is the performance of such an optimum scheme to deviations in the system for which the scheme is designed? Unfortunately, it turns out that in many cases, such nominally optimum designed schemes can suffer a drastic degradation in performance even for minor deviations from nominal assumptions. This basic observation motivates the search for robust inference techniques, that is, techniques with good performance under any nominal conditions and acceptable performance under other than the nominal conditions.

Sensor nodes in WSNs consist of sensing, data processing, and communicating components. However, like other complex networked systems, the functionality and performance of a WSN may be affected when it is not operated under nominal ideal conditions [32]. Various reasons may cause the deviations, such as sensor malfunction, sensor drifts, and compromised sensors. The unreliable nodes may be degraded naturally or exposed to vulnerability such as cybersecurity attacks. If left untreated, such unreliable nodes may cause significant damage to the WSNs inference performance. Unfortunately, not many attempts have been made to study the performance degradation of WSNs and design a robust network to counter such vulnerabilities [31, 33, 34].

1.4 Contribution and Overview of the Dissertation

This dissertation focuses on robust inference approaches to WSNs when the network is unreliable. The overall system performance depends significantly on the knowledge of the malfunctioning nodes, the behavior of nominal systems, and the inference schemes. The more information about the malfunctioning nodes in WSNs, the better system performance it can achieve. Based on the possible abnormal actions of sensor node and action of network operator of WSNs, we model the inference problem in theoretical game settings. We consider the case where the network operator or FC is aware of the presence of anomaly nodes and can adjust the decision rule to counter the impact of unknown malicious nodes. The adversarial nodes are assumed to know that the FC would take some actions to improve its robustness.

The systematic approaches are provided for robust detection and estimation in WSNs. In order to achieve this, first, the measure of the overall performance of a scheme is specified with a wide range of all possible allowable actions. One such measure that has been widely used and leads to valuable results in many situations is the worst-case performance over a wide range of actions of compromised nodes. If its worst-case performance is good, we may say that a given inference in WSNs is robust. The mini-max robust schemes performance is usually not far below that of nominally optimal schemes when the system is under nominal operation.

The rest of this dissertation is organized as follows

- Chapter 2: Background and fundamental concepts- In this chapter, key concepts such as fundamental concepts of robust detection and robust estimation. In addition, two models to describe the component or node reliability of WSNs are presented. Finally, statistical inference in WSNs is formulated under theoretical game settings.

- Chapter 3: Robust detection- In this chapter, the main contributions of robust detection in WSNs are summarized as follows:

  - For detection problems, the main focus is on binary detection, or binary hypotheses testing. That is, among H_0 and H_1 hypotheses, the aim is to design robust detectors when some observation nodes are malfunctioning.
  - The detection problem is investigated under the mini-max framework for the static settings as no such saddle point solutions are shown to exist.
  - Two robust detectors, namely Clipped Log-Likelihood Ratio Detector (CLLRD) and $\alpha-$Trimmed Sum Detector, are proposed. The performance of these detectors using a mini-max game theoretical framework is studied.
  - Applications of these detectors for normally distributed observations and distributed MIMO radars are analyzed. The closed-form solutions of the performance of these two detectors are derived. It is shown that these detectors provide some guaranteed performance despite a fixed number of extreme outliers in the observation, as validated by numerical simulations.

- Chapter 4: Robust estimation- In this chapter, results in robust estimation are presented under both the probabilistic and static models. The main contributions are summarized as follows:

  - Estimation problem is first presented under the frequentist setting where the unknown is a scalar parameter. The max-mini solutions that make the observation partially uninformative and sometimes even completely uninformative are derived. Specifically, the sufficient and necessary condition is derived for the estimator to be completely uninformative about unknown parameters.
  - Next, the estimation problem is considered under the Bayes settings where the prior probability distribution of a parameter is known, we obtain the saddle point solution for a single-node network under the probabilistic unreliability model. Based on the result of a single node, a robust estimator for the multi-node networks is proposed under both probabilistic and static models.

- Chapter 5: Robust regression- In this chapter, robust algorithms are proposed for target localization. The main contributions are:

  - Results in target localization through regression are presented. The robust regression problem is formulated into an equivalent weighted least square regression where weights are based on the robust cost.
  - A method to improve the localization accuracy is proposed by introducing a small set of secured sensors, potentially by spending more resources on those sensors, especially in a potentially hostile environment. By doing so,

we might be able to improve the signal quality and integrity for a small set of sensors and, consequently, reduce the target localization deviations.

- Chapter 6: All presented works are concluded in this chapter and discuss future research directions related to robust inference are discussed.

CHAPTER 2:

BACKGROUND AND BASIC CONCEPTS

Statistical inference in WSNs includes detection, estimation, and tracking. They all assume some knowledge of the states of nature, and the uncertainty of the environment, and typically define these statistically. One of the main differences between detection and estimation is the phenomenon to be inferred by sensors. In detection, the phenomenon observed by sensors is discrete, e.g., binary hypothesis testing, where one aims to decide between two potential hypotheses, $H \varepsilon (H_1, H_0)$. In estimation, the phenomenon is often a parameter in a continuous set [15].

2.1 Robust Detection

Detection is widely used for both military and civilian applications, including distributed array radars intruder detection, anomaly detection, and intelligent transportation systems where the infrastructure sensors detect pedestrians, vehicles, and other anomaly events [35, 36, 37]. As one of the essential aspects of inference, detection is often the initial goal of a pattern recognition system and aims at detecting signals or events as accurately as possible [14]. For example, a WSN of N sensors are densely deployed in forests to observe the temperatures, and through communication channels, these nodes send the compressed data to the FC, where the final decision is made about whether there is forest fire or not [38]. For WSNs, detecting the presence of an event is often the priority of all the other tasks including estimation, tracking, and learning [7].

“A robust detection refers to the determination of events with a guaranteed level of detection performance despite the uncertainties on the underlying statistical models” [32]. Often, a detector is designed to give the best possible performance for a specific statistical model, e.g., to detect the presence of a signal with a known waveform in additive noise with known distribution. However, such idealized settings are seldom met in practical applications. Besides the usual uncertainty regarding signals and noises, there is always a chance of drifting observations from the ideal conditions for various reasons, such as sensor malfunction, sensor drifts, and compromised sensors due to cyberattacks etc. In many cases, nominally optimal detection schemes can suffer a drastic degradation in performance even for minor deviations from nominal assumptions [39]. This motivates the search for robust detectors with good performance under nominal conditions and acceptable performance under conditions other than the nominal one.

2.1.1 Distributed Detection in WSNs

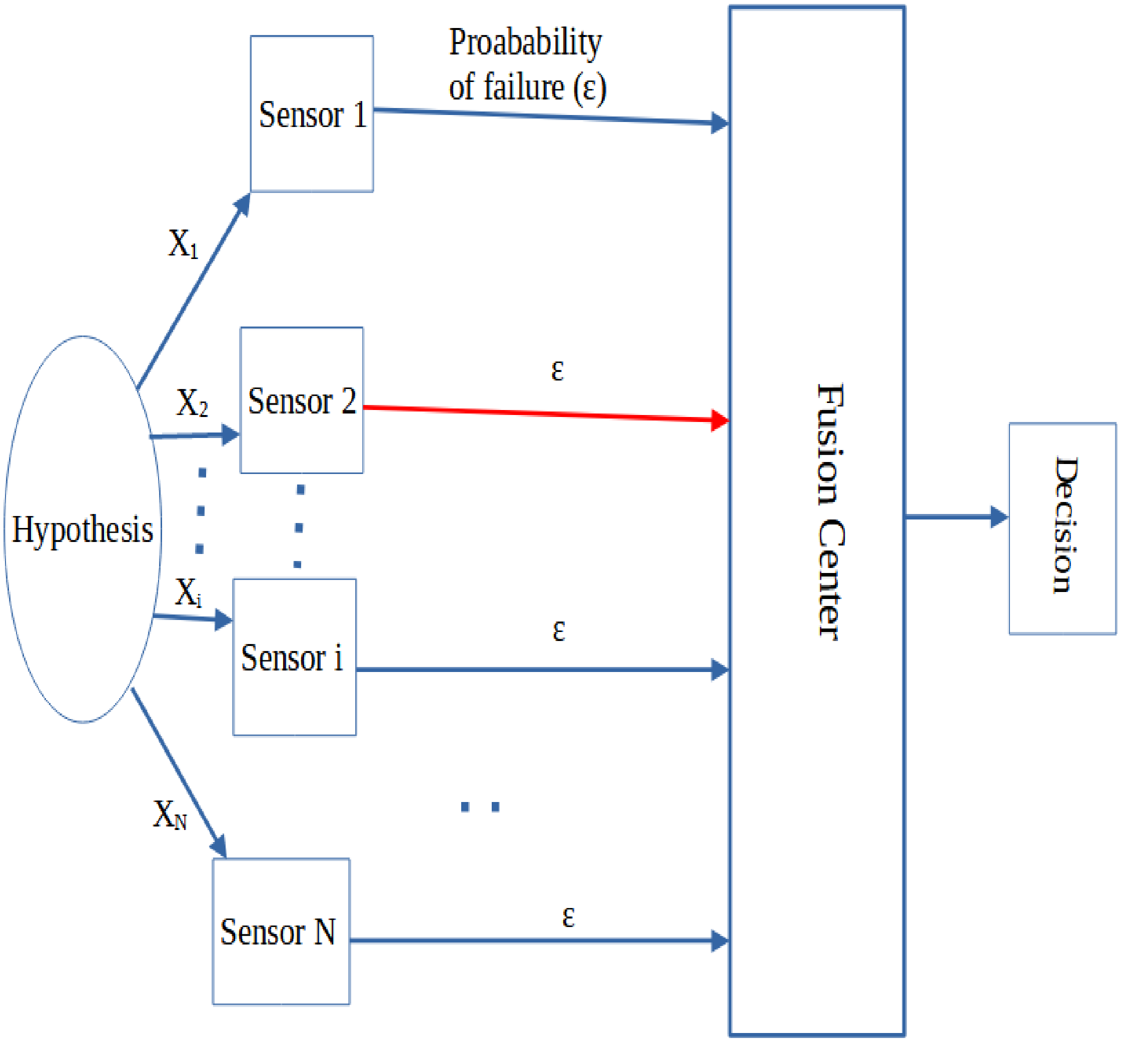

Figure 2.1: Detection in a parallel WSN.

Fig 2.1 shows the structure of a parallel network detection problem where local sensors sense the data X. Sensor i sends compressed or uncompressed data $X_i, (i = 1...N)$ to the outputs transmitted across a channel. Ultimately, the FC makes the decision based on the received data. When communication links are of high capacity and/or the latency requirements of the decision making are low, the local sensors send all the data without compression. The optimum detector is the Likelihood Ratios (LR) [43]. This requires that all the observations of the sensors be independent, and available at the FC. Owing to this assumption, the optimum detector at FC is a likelihood ratio at nominal conditions

$$\Lambda(\boldsymbol{X}) = \sum_{i=1}^{N} \frac{p_{x|H_1}(X_i \mid H_1)}{p_{x|H_0}(X_i \mid H_0)} \mathop{\gtrless}_{H_0}^{H_1} \eta. \tag{2.1}$$

Suppose $\Lambda(\boldsymbol{X})$ is unbounded as a function of X. In that case, a single observation can heavily influence the detection. In WSNs, a single unreliable sensor node can, therefore, completely override the weight of a possibly large number of other reliable sensor nodes in the choice between H_0 and H_1. Various factors may cause such unreliability. For example, the sensor malfunctions when statistical mismatches between the actual distribution and the model assumptions exist. It can also be corrupted or compromised due to the natural degradation of nodes and some unknown source of contamination or adversarial attacks. To counter such undesirable detection sensitivity due to unreliable sensor nodes. It can also be corrupted or compromised due to the natural degradation of nodes and/or some unknown source of contamination or adversarial attacks.

To counter such undesirable sensitivity due to unreliable sensor nodes, we consider a bounded modification (clipping) $\tilde{\Lambda}(\boldsymbol{X})$ of the function $\Lambda(\boldsymbol{X})$ corresponding to the assumed nominal model.

$$\tilde{\Lambda}(\boldsymbol{X}) = \begin{cases} \tau_2 & \Lambda(\boldsymbol{X}) > \tau_2 \\ \Lambda(\boldsymbol{X}) & \tau_1 \leq \Lambda(\boldsymbol{X}) \leq \tau_2 \\ \tau_1 & \Lambda(\boldsymbol{X}) < \tau_1 \end{cases},$$

where τ_1 and τ_2 are constants. One can expect that clipping would degrade the detection performance when WSNs nodes operate ideal nominal conditions. On the other hand, the boundedness builds robustness against the influence of spurious nodes. The range $[\tau_1, \tau_2]$ controls the trade-off between the degree of robustness and performance degradation under nominal operation.

The trimming method is another conservative approach to counter the unreliable nodes. It is based on the rank order $\varLambda_1(X_1) \leq \Lambda_2(X_2) \leq ...\varLambda_N(X_N)$ of likelihood ratios of sensor observations. The influence of the compromised nodes is eliminated after removing some biggest and smallest of log-likelihood ratios. Similar to clipping on log-likelihood ratio, trimming increases robustness against outlier nodes at the cost of performance degradation when the network operates under nominal conditions. In Chapter 3, the detection problem under the mini-max framework is investigated by using clipping and trimming.

2.2 Robust Estimation

Suppose the detection function in WSNs determines the presence of an object, a signal, or an event. In that case, more complicated tasks such as estimation and tracking can be performed. For instance, if an intelligent transportation system detects a vehicle, the following task would be estimating how fast the vehicle is moving and where it is moving. Aiming to estimate the values of a group of parameters based on a network of sensors, centralized and distributed estimation has been an important and active research area over the past several decades [40, 41, 42].

2.2.1 Distributed Estimation in WSNs

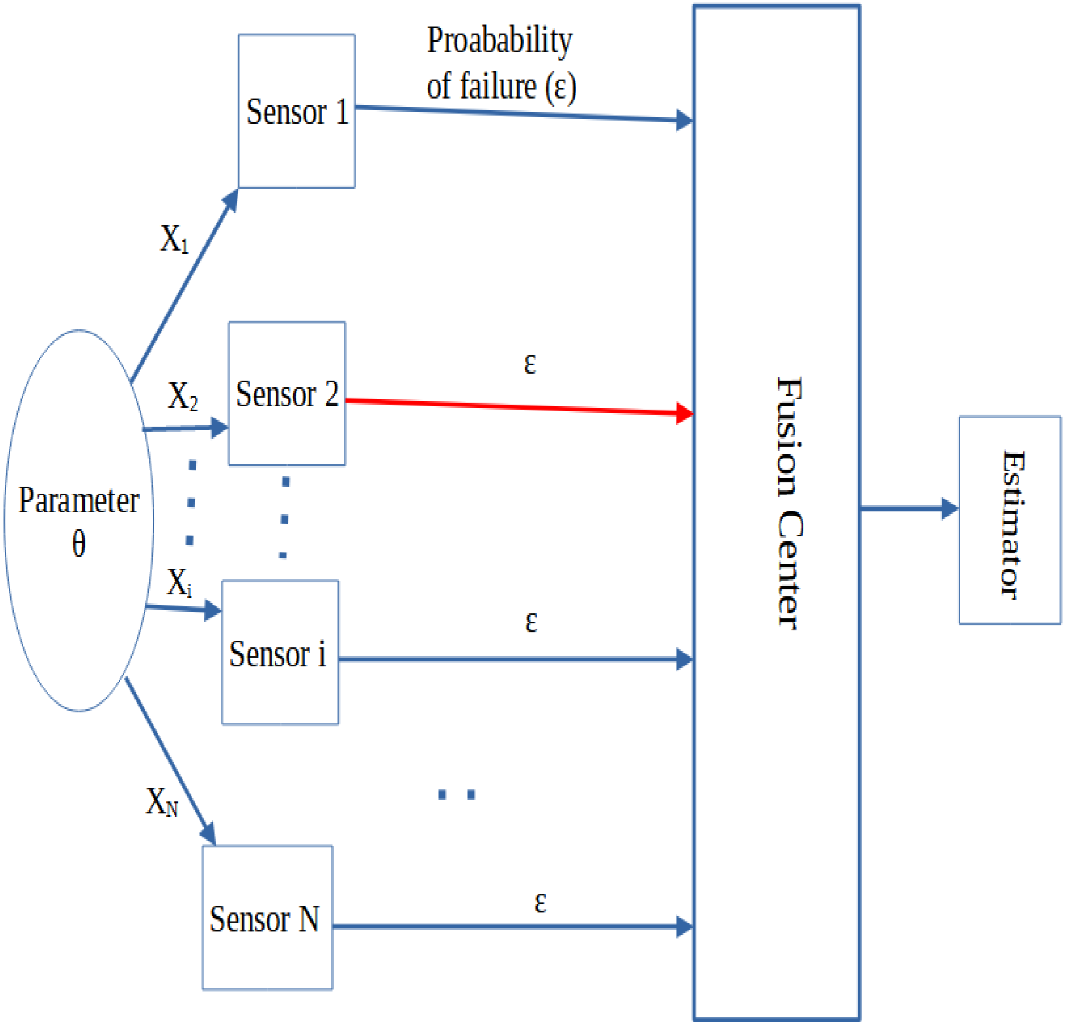

Figure 2.2: Estimation in a parallel WSN.

Similar to the detection setting, sensors observe and send the processed data to the FC through a channel as shown in Fig 2.2.

For example, consider the scalar estimation problem,

$$X_i = \theta + W_i, \; i = 1, ..., N,$$

where θ is the true random unknown variable and W_i is the random noise with the probability distribution f_w. The goal is to obtain $\hat{\theta}$, an estimate of θ from the observations. The estimation accuracy is often measured by a cost function $C(\theta, \hat{\theta})$. For example, for i.i.d. observations with the Gaussian observation noise, the Maximum Likelihood (ML) estimate of θ

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{N} \log f_X(X_i \mid \theta),$$

is the sample mean, $\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} X_i$, which is also optimal least square estimate (LSE). However, it is common in practice that the noise process is non-Gaussian [31] or may contain outliers [33, 34]. The estimation performance would be degraded significantly. To improve the estimation performance under such unreliable conditions, the robust estimator is proposed in Chapter 4. The trade-off is analyzed between the robustness against the node failures and performance under the nominal conditions.

2.3 System Reliability Models

For any WSNs, its performance depends on the reliability of its components. A reliability model describes the vulnerability of one or a group of nodes. Two standard failure models are adopted in this study to describe the WSNs system vulnerability.

2.3.1 Probabilistic Model

When the network nodes operate independently, the failure of one node is unlikely related to the others, and node reliability can be modeled independently. In such cases, we consider a probabilistic model such that the node i has a probability of ε_i being malfunctioning, independent of other nodes,

$$P\left(\mathbf{s}=\mathbf{1}\right)=\prod_{i=1}^{N}\varepsilon_i^{s_i}\left(1-\varepsilon_i\right)^{1-S_i}.$$

In this model, a node (e.g., transmitter, receiver or data link) has an ε_i probability of malfunctioning or being compromised. We further consider the i.i.d. case where ε_i=ε. For instance, $\varepsilon = 0.1$ means each node has a 0.1 probability of being compromised. This model is also known as $\varepsilon-$contamination model proposed by Huber for robust inference theory [43]. For estimation of a parameter θ, we assume a nominal distribution P_θ and an outlier distribution Q where each observation follows a mixture distribution.

$$\tilde{P}_\theta=(1-\varepsilon)P_\theta+\varepsilon Q. \tag{2.2}$$

Under this model, data are drawn from equation (2.2) where each entry has a probability of ε to be contaminated by some arbitrary distribution Q. Given observations from equation (2.2), the objective is to infer θ robustly against Q.

2.3.2 Static Model

While $\varepsilon-$contamination is widely used for modeling, this model may not be a good fit for some scenarios. For example, where either the value of ε is hard to know in a prior or when such probabilities are not independent among observations. Therefore, the binomial distribution of outliers in the $\varepsilon-$contamination model is seldom met. Instead, we often consider redundancy designs in handling up to a few outliers. Therefore, we also consider a static model which assumes there is up to a certain number of malfunction nodes in the system.

Under the static setting, it is assumed that there are K outliers each from unknown pdf $q(X_i) \in Q$ and nominal $N-K$ observations, each with known probability density function $p(X_i) \in P_\theta$, respectively. It gives intuitive results with good interoperability to describe and model the severeness of anomaly. The system design goal is to provide robustness against such failure scenarios, and also provide good performance when the system is normal. Let Ω_K be the set of all subsets of K nodes, the probability of the observation $\mathbf{s} = [s_1, s_2, ..., s_N]$ is given by

$$P\left(X_K|\theta\right) = \sum_{A_K \in \Omega_K} P\left(A_K\right) \prod_{i=1}^{N} q(X_i)_{I_{i \in A_K}(S_i=1)} p(X_i)_{I_{i \notin A_K}(S_i=0)}, \qquad (2.3)$$

where $P\left(A_K\right)$ is probability of the subset nodes A_K being compromised, usually depends on the system architecture. When $P\left(A_K\right)$ is not specified, one might either consider a mini-max approach where the design goal is to minimize the impact of the worst possible K subset or assigning a uniform distribution on $P\left(A_K\right) = 1/\left(\begin{array}{c} N \\ K \end{array}\right)$, where each node is equally likely to be malfunctioning.

Both can capture some system failure scenarios, and they are equivalent in one extreme case. $\varepsilon = 0$ means none of nodes fail $\varepsilon = 1$ means all nodes fail. Designs for one model may be used for the other. We would like to point out that the probabilistic setting is different from static settings in system robustness, where one assumes up to a fixed number of the node being compromised. For instance, for a MIMO radar with number of the transmitter M=10, with $\varepsilon = 0.1$ is not the same as 1 out of 10 nodes is controlled by the attacker, as in the former case, the attacker might end up controlling 0 transmitter or all 10 transmitters, with specific probabilities. However, the probabilistic and static settings are almost equivalent in an asymptotic sense when ε is fixed and $M \to \infty$ for fixed number of receiver.

2.4 Problem Formulation for Statistical Inference

The inference problems in a WSN with unreliable nodes is investigated under the game-theoretical settings. Specifically, a network of nodes in a zero-sum game is considered with 2 players with single stage. The goal is to infer the underlying status of $\theta = [\theta_1, \theta_2, \cdots, \theta_L]$ drawn from $p(\theta)$ distribution. The node status $\mathbf{s} = [S_1, S_2, \cdots, S_N], S_n \in \{0, 1\}$ is a binary random vector and fixed throughout the duration of the game. $S_n = 0$ denotes a nominal and $S_n = 1$ denotes an abnormal node state. Node observations $\mathbf{x} = [X_1, X_2, \cdots, X_N]^T$ are conditionally independent given by the node status s and the underlying status θ such that

$$p_{\mathbf{x}}(\mathbf{x}|\mathbf{s}, \theta) = \prod_{n=1}^{N} p_{Xn}(X_n|S_n, \theta).$$

Framed as a “personalized” attacker, Player 1 denotes the anomaly states the system experiences. Under the zero-sum game settings, Player 1 “attempts” to make the inference performance as bad as possible. As Player $1'$s counterpart, Player 2 is the “personalized” system operator/designer who attempts to optimize the system performance under the potential anomaly caused by Player 1.

When $S_n = 0$, the nth node functions normally with a known pdf $p_{Xn}\left(x_n|s_n = 0, \theta\right)$. When $S_n = 1$, the node n malfunctions, and its observation distribution follows $p_{Xn}\left(x_n|S_n = 1, \theta\right)$, which is "picked"/"designed" by Player 1. Player 1 also has some additional private information about θ through its observation X_{p1}, obtained either through own observations or in some privacy/cybersecurity settings where Player 1 is the target itself and knows θ perfectly. Similarly, it is assumed that Player 2 has some additional private information about θ via its observation X_{p2} which is made available through his own secured observation nodes.

The system vulnerability to malfunctioning is described by the distribution of $p_{\mathbf{s}}\left(s\right)$. Such information is assumed to be known to both players. The system operator Player 2 aims to design the best strategy to infer $\hat{\theta} = \hat{\theta}\left(\mathbf{x}\right)$ from $\mathbf{x}$. As a zero-sum game, the payoff to Player 1 is the inference performance $d\left(\theta, \hat{\theta}\right)$ and the payoff to Player 2 is $-d\left(\theta, \hat{\theta}\right)$.

Table 2.1: A two player, zero-sum game.

Player	Knowledge	Strategy	Payoff
1	$X_{p1}, \left\{p_{Xn}\left(x_n\|S_n = 0, \theta\right)\right\}, p_{\mathbf{s}}\left(s\right)$	$\left\{p_{Xn}\left(x_n\|S_n = 1, \theta\right)\right\}$	$E\left(d\left(\theta, \hat{\theta}\right)\right)$
2	$X_{p2}, \left\{p_{Xn}\left(x_n\|S_n = 0, \theta\right)\right\}, p_{\mathbf{s}}\left(s\right)$	$\hat{\theta}\left(\mathbf{x}\right)$	$-E\left(d\left(\theta, \hat{\theta}\right)\right)$

2.4.1 Three Game Solutions

Under the settings in table 2.1, depending on the availability of any additional knowledge, the game can be briefly categorized into three types and seek three solutions.

1. Max-mini: In the game between two players, Player 2 is assumed to know Player $1'$s strategy $\left\{p_{Xn}\left(x_{nl}|S_n = 1, \theta\right)\right\}$, and can use it to design $\hat{\theta}$. Therefore, the best Player 1 can do is to maximize the minimal $E\left(d\left(\theta, \hat{\theta}\right)\right)$ through his choice of

$\{p_{Xn}(x_n|S_n = 1, \theta)\}$.

2. Mini-max: In this case, Player 1 is assumed to know Player 2's strategy $\hat{\theta}(\mathbf{x})$, and use such knowledge to design $\{p_{Xn}(x_n|S_n = 1, \theta)\}$ to maximize the $E\left(d\left(\theta, \hat{\theta}\right)\right)$. From Player $2'$s perspective, the designing of $\hat{\theta}(\mathbf{x})$ is a Mini-max approach.

3. Equilibrium (if existing): In this case, neither player knows about the other's strategy and the goal is to determine an equilibrium solution between two players $\left(\{p_{Xn}(x_n|S_n = 1, \theta)\}, \hat{\theta}(\mathbf{x})\right)$, if existing, such that neither player can gain by deviating from his strategy. We also refer to such equilibrium solutions as saddle-point solutions as no player can unilaterally increase his payoff by choosing a different strategy.

2.4.2 Performance Metrics

One critical part of the problem formulation is the choice of payoff function $d\left(\theta, \hat{\theta}\right)$. Depending on the inference settings, there are three most widely used performance measures among the popular choices of payoff $d\left(\theta, \hat{\theta}\right)$:

1. Probability of error: Probability of error P_e is a popular choice for detection problems under Bayes settings when θ is the set of hypotheses to be tested by:

   $$d\left(\theta, \hat{\theta}\right) = Pr\left(\theta \neq \hat{\theta}\right)$$

   In most cases, more general linear cost functions can be reformulated and normalized to the P_e via a new set of prior probabilities. For binary hypotheses testing, the Neyman-Pearson criterion, which seeks to maximize the probability of detection under a probability of false constraint, is also of critical importance.

This can be considered in the current formulation via the additional constraints.

2. Mean Square Error (MSE): MSE is widely used for parameter estimation. Specifically, when θ is the set of continuous parameters to be estimated, the performance of the estimation is given by

$$d\left(\theta, \hat{\theta}\right) = E\left(\left\|\theta - \hat{\theta}\right\|_2^2\right),$$

where $\|\cdot\|_2$ is the second order vector norm. A potential extension Frobenius Norm can be used when θ is a matrix.

3. Mutual information: Although it is not monotonically correlated to the inference performance, due to its generality and versatility, mutual information $I\left(\theta; \mathbf{x}\right)$ is often used to measure the information contained in $\mathbf{x}$ about θ. To be consistent with the MSE and probability of error, we can use the negative mutual information as the distance measure

$$d\left(\theta, \hat{\theta}\right) = -I\left(\theta; \mathbf{x}\right).$$

As the calculations of $d\left(\theta, \hat{\theta}\right)$ requires the complete distribution information, this metric is most suitable applicable to the Maxi-mini settings where Player 2 knows $\{p_{Xn}\left(x_n|S_n = 1, \theta\right)\}$.

CHAPTER 3:

ROBUST DETECTION

For detection problems, we consider binary detection or hypotheses testing problems. That is, among H_0 and H_1 hypotheses, we aim to design robust detectors when some of the observation nodes are malfunctioning, as in Fig 3.1. As the detection problem is relatively well known under the probabilistic model in simple binary hypotheses testing with a known saddle-point solution [39, 44], the focus is given on the static setting, where there is a fixed number of malfunctioning nodes. The detection problem under the mini-max framework is considered for the static settings as no such saddle-point solutions are shown to exist. Most of the results are under the single-stage settings where the detection game is only played once, and the adversary is not concerned by revealing the malfunctioning node identity.

This chapter is organized in the following manner. First, the target detection model and problem statement are outlined. In section-3.2, two robust detectors are introduced: Huber's Clipped Log-likelihood Ratio Detector and $\alpha-$Trimmed Sum Detector. These detectors are analyzed and evaluated in two different cases: one is target detection with normally distributed observations and the other is target detection in a distributed MIMO radar in section-3.3, and section-3.4, respectively. The Effectiveness of these robust detectors is evaluated and compared through Monte Carlo simulations in section-3.5.

3.1 Problem Statement

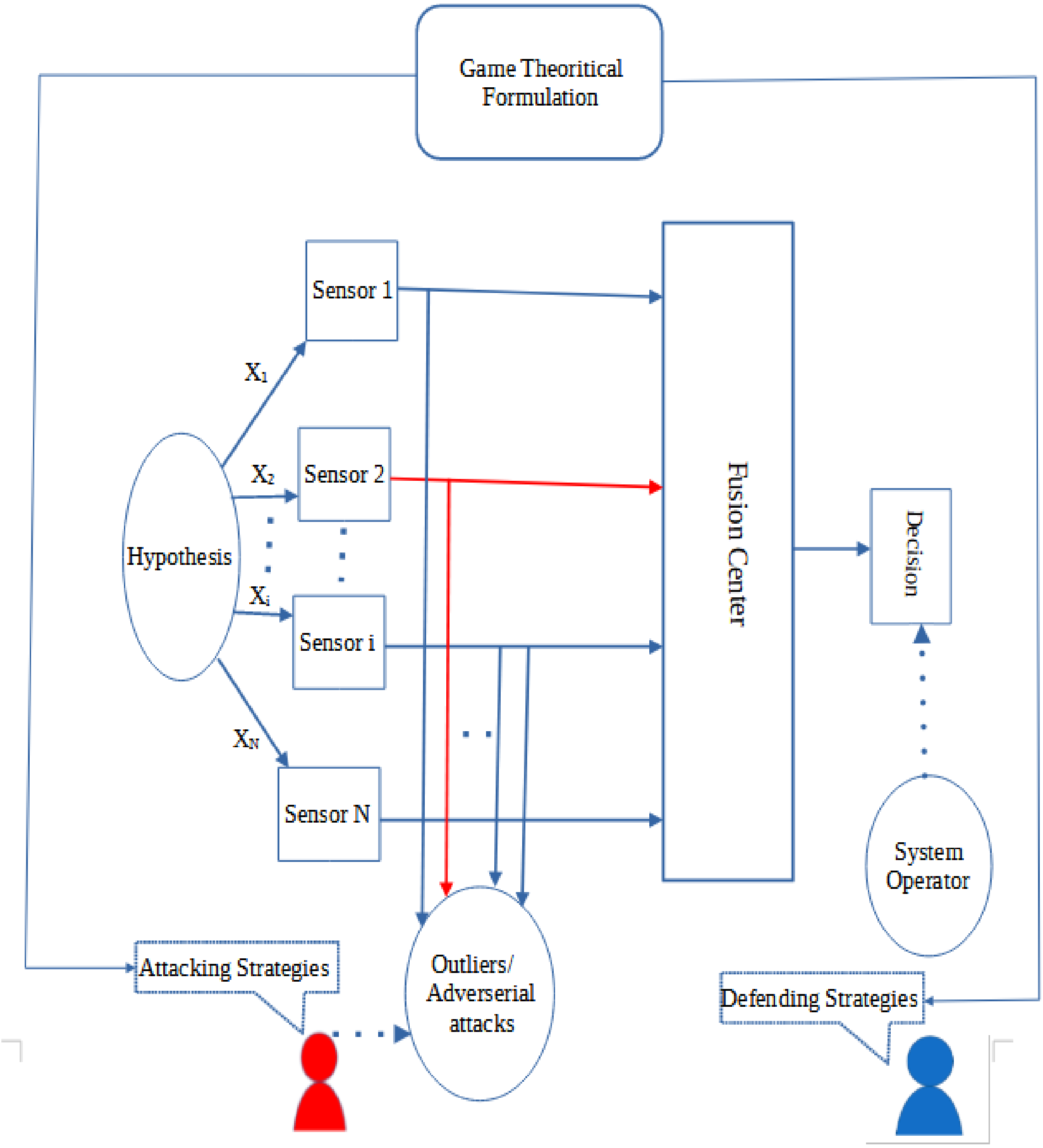

Figure 3.1: Robust Detection Framework.

Let $(\Omega, \mathcal{A})$ be a measurable space. The observation $X \sim [X_1, X_2, ..., X_N]$ are assumed to be independent and identically distributed (i.i.d.), and each observation is assumed to be equally likely to be compromised. Consider the binary hypotheses testing problem H_0 vs H_1, where there are K outliers with unknown pdf $q(X_i/H_1) \in Q(H_1)$ and $q(X_i/H_0) \in Q(H_0)$, and the remainder $N - K$ observations follow nominal distributions, each with known pdf $p(X_i/H_1) \in P(X_i/H_1)$ and $p(X_i/H_0) \in P(X_i/H_0)$, respectively. The overall detection problem is

$$\begin{aligned} H_1 : p(X/H_1) &= \sum_{S \epsilon \Omega_K} Pr(S) \prod_{j=j_1}^{j_K} q(X_j/H_1) \prod_{j=j_1}^{j_{N-K}} p(X_j/H_1) \\ H_0 : p(X/H_0) &= \sum_{S \epsilon \Omega_K} Pr(S) \prod_{j=j_1}^{j_K} q(X_j/H_0) \prod_{j=j_1}^{j_{N-K}} p(X_j/H_0), \end{aligned} \tag{3.1}$$

where Pr(S) is the probability of the set S of observations with cardinality K to be the outliers and $\Omega_K = \{j_1, ..., j_K \mid 1 \leq j_1 < j_2..., < j_K \leq N\}$ is the set of all possible subset with size K.

3.2 Robust Detectors

3.2.1 Clipped Log-Likelihood Ratio Detector CLLRD

The Huber's detector is considered first, which is basically a truncated log-likelihood ratio, $l(X_i) = ln\left(\frac{p_1(X_i)}{p_0(X_i)}\right)$ with two threshold $\tau_1 \geq \tau_0$ such that

$$\lambda(X_i) = \begin{cases} \tau_1 & l_i(X_i) \geq \tau_1 \\ l_i(X_i) & \tau_0 < l(X_i) < \tau_1 \\ \tau_0 & l_i(X_i) \leq \tau_0 \end{cases} . \tag{3.2}$$

The test statistic is the sum of clipped log likelihood ratios

$$T_T(X) = \sum_{i=1}^{N} \left(\lambda_i(X_i)\right),$$

and the threshold is set to η, then the detection rule can be expressed as

$$T_T(X) \gtrless^{H_1}_{H_0} \eta, \tag{3.3}$$

where η is the threshold. The resulting probability of detection P_d and probability of false alarm P_f are given by

$$\begin{aligned} P_f &= Pr\left\{T_T(X) > \eta \mid H_0\right\} \\ P_d &= Pr\left\{T_T(X) > \eta \mid H_1\right\}. \end{aligned} \tag{3.4}$$

Since the distribution of $\lambda(X_i)$ contains a point mass, randomization might be needed for a better detection performance.

Obviously, the detection performance is the worst when $T_T(X)$ is statistically as small as possible under H_1 and is as big as possible under H_0. As such, when $Q(H_1) = Q(H_0) = \Omega$, that is, the outlier distribution can be arbitrary, the optimal outlier $q^0(X_i/H_1)$ is to put all the probability mass at the set $\left\{X_i^I \mid l_i(X_i) \leq \tau_0\right\}$ and $q^0(X_i/H_0)$ is to put all the probability mass at the set $\left\{X_i^U \mid l_i(X_i) \geq \tau_1\right\}$. Under this worst possible scenario, the test statistic becomes

$$T_T(X) = \begin{cases} \sum_{i=1}^{N-K} \lambda(X_i) + K\tau_0 : & H_1 \\ \sum_{i=1}^{N-K} \lambda(X_i) + K\tau_1 : & H_0. \end{cases} \tag{3.5}$$

The corresponding minimum guaranteed detection performance can be analyzed by evaluating equation (3.5) with the distribution of the sum of clipped log-likelihood ratios. When N and K are sufficiently large, some approximated asymptotic results can be obtained by employing large deviation analysis such as the central limit theorem (CLT). For example, let u_j and v_j be the mean and variance of $\lambda(X_i)$ under H_j with perspective nominal distributions, $j = 0, 1$, respectively. By CLT, it can be shown that

$$\begin{aligned} P_f &\simeq Q\left\{\frac{\eta - K\tau_1 - (N-K)u_0}{\sqrt{(N-K)v_0}}\right\} \\ P_d &\simeq Q\left\{\frac{\eta - K\tau_0 - (N-K)u_1}{\sqrt{(N-K)v_1}}\right\}, \end{aligned} \tag{3.6}$$

where $Q\,(.)$ is the complementary distribution function of a standard normal random variable. The thresholds τ_1 and τ_0 can be tuned based on equation (3.6). Alternatively, one may adopt the thresholds by Huber's test with the corresponding $\epsilon = K/N$. Although, this problem is easier to solve, this set of thresholds are not guaranteed to be optimal.

3.2.2 Trimmed Sum Detector

Trimmed means, which are less affected by outliers, often provide a better estimation of the location of the bulk of the observations than the mean. Inspired by such robust estimation, a detector as an α-trimmed sum on log-likelihood ratios is considered.

Without loss of generality, assume $l_1(X_1) \leq l_2(X_2) \leq ...l_N(X_N)$. The α-trimmed sum on log likelihood ratios is the sum of log likelihood ratios after removing the largest $K_1 \geq K$ and the smallest K_1 values, and used as a test statistic.

$$T_\alpha(X) = \sum_{i=K_1+1}^{N-K_1} l_i(X_i). \tag{3.7}$$

When $Q(H_1) = Q(H_0) = \Omega$, the worst possible outlier $q^0(X_i/H_1)$ is to put all the probability mass at the lower end $X_i^I = \arg\inf(l_i(X_i))$ and $q^0(X_i/H_0)$ is to put all the probability mass at the upper end $X_i^S = \arg\sup(l_i(X_i))$. As a result, $T_\alpha(X)$ becomes the sum of the lowest ranked $N - 2K_1$ log likelihood ratio values out of $N - K_1$ nominal observations under H_1, and the sum of the highest ranked $N - 2K$ log likelihood ratio values out of $N - K_1$ nominal observations under H_0. Next, these two robust detectors are evaluated and analyzed.

3.3 Detection in Normally Distributed Observations

Let us first consider a detection problem where the observation $X \sim [X_1, X_2, ..., X_N]$ are normally distributed such that

$$X_i \sim \begin{cases} N(\mu, \sigma^2): & H_1 \\ N(-\mu, \sigma^2): & H_0 \end{cases}, \ \mu > 0.$$

We consider the case where K out of N observations are compromised. Due to the symmetry of this problem, it can be shown that the threshold is symmetric such that

$\tau_1 = -\tau_0 = \tau$ as in [43], and

$$T_c(X) = \begin{cases} \sum_{i=1}^{N-K} \lambda(X_i) - K\tau : & H_1 \\ \sum_{i=1}^{N-K} \lambda(X_i) + K\tau : & H_0 \end{cases}$$

The respective pdfs under either hypothesis are given by

$$\begin{aligned} f\left(\lambda(X_i) \mid H_1\right) &= Q\left(\frac{\tau-\mu}{\sigma}\right)\delta(X_i-\tau) + \phi(X_i-\mu)I_{X_i\epsilon[-\tau,\tau]} + Q\left(\frac{\tau+\mu}{\sigma}\right)\delta(X_i+\tau) \\ f\left(\lambda(X_i) \mid H_0\right) &= Q\left(\frac{\tau+\mu}{\sigma}\right)\delta(X_i-\tau) + \phi(X_i-\mu)I_{X_i\epsilon[-\tau,\tau]} + Q\left(\frac{\tau-\mu}{\sigma}\right)\delta(X+\tau), \end{aligned} \tag{3.8}$$

where $\delta(.)$ is used to denote the probability mass function. For $N = 3$, $K = 1$, the probability of detection P_d and false alarm P_f given by

$$\begin{aligned} P_d &= Pr\left\{\lambda\left(X_1\right) + \lambda\left(X_2\right) > \eta + \tau \mid H_1\right\} \\ P_f &= Pr\left\{\lambda\left(X_1\right) + \lambda\left(X_2\right) > \eta - \tau \mid H_0\right\}. \end{aligned}$$

For 3 observations, the closed-form solution for P_d and P_f can be calculated using convolution sum as in appendix B. However, for larger N, it is difficult to get distribution of $T_c(X)$, so asymptotic result of P_d and P_f is approximated by equation (3.8), and can be expressed as

$$P_d \simeq Q\left\{Q^{-1}(P_f) - 2\frac{d(\tau)}{\sqrt{(N-K)v_1(\tau)}}\right\}, \tag{3.9}$$

where $d(\tau) = (N-K)u_1(\tau) - K\tau$, $u_1(\tau))$ and $v_1(\tau)$ are the mean and variance of $\lambda\left(X_i/H_1\right)$. For any given P_f, the threshold $\tau^* = \arg\max_\tau \left(d(\tau)\right)$ that maximizes P_d

is the one that maximizes $\frac{d(\tau)}{\sqrt{(N-K)v_1(\tau)}}$ respectively.

Similarly, in order to analyze the performance of $\alpha-$Trimmed Sum Detector, we need the distribution of sum of ranked log likelihood ratios. A closed form solution is possible for small N and K. For example, consider 1 out of 3 observations is contaminated, the $\alpha-$trimmed sum on log likelihood ratios is the median of log likelihood ratio after removing the biggest and smallest and used as test statistics

$$T_\alpha(X) = median\,\{Y_{1,}Y_{2,}Y_3\}^{>^{H_1}}_{<_{H_0}}\eta,$$

$T_\alpha(X)$ is statistically as small as possible under H_1 and as large as possible under H_0. Extreme outliers would be to send $-\infty$ under H_1 and ∞ under H_0,

$$T_\alpha(X) = \begin{cases} min\,\{X_1, X_2\}: & H_1 \\ max\,\{X_1, X_2\}: & H_0. \end{cases}$$

The detection performance is

$$\begin{aligned} P_f &= 1 - \left[1 - Q\left(\frac{\eta+\mu}{\sigma}\right)\right]^2, \\ P_d &= \left[Q\left(\frac{\eta-\mu}{\sigma}\right)\right]^2 \end{aligned} \tag{3.10}$$

3.3.1 Performance analysis

The detection performance depends on N, K, and signal quality under nominal conditions. From equation (3.6), for any given P_f, it can find thresholds τ_0, τ_1 that maximize P_d. For a fixed N, a greater K results in worse detection performance. There exists a K_0 such that when $K \geq K_0$, the detector can not provide any meaningful results, i.e. $P_d \leq P_f$. In such case, we denote $\beta_0 = f(K_0, N)$ to be the break down point. It indicates the maximal fraction of outliers in the observations that a detector can handle without breaking down and is used to characterize the quantitative robustness of a detector.

From equation (3.9), the detector is meaningful if $d(\tau) > 0$. For the particular N and K, the break down point is reached if $d(\tau) = 0$, i.e. $P_d \leq P_f$. Therefore, if $u_1(\tau) \leq \frac{K}{N-K}\tau$, then the detector can not provide any meaningful result. As a result, the breakdown point K_0 can be maximized by maximizing $\frac{u_1(\tau)}{\tau}$. For a particular N and K, we want to find minimum break down μ such that $d(\tau) > 0$. At $\tau = 0$, $d(0) = 0$. If there exist τ such that $d(\tau) > 0$. Hence,

$$\frac{\partial d(\tau)}{\partial \tau} \mid_{\tau=0} > 0 = \frac{\partial u_1(\tau)}{\partial \tau} \mid_{\tau=0} - \frac{K}{N-K} > 0,$$

where, $\frac{\partial u_1(\tau)}{\partial \tau} \mid_{\tau=0} = Q\left(-\frac{\mu}{\sigma}\right) - Q\left(\frac{\mu}{\sigma}\right)$. The minimum breaking down point μ, above which the detector works properly for given N and K can be obtained by solving $Q\left(-\frac{\mu}{\sigma}\right) - Q\left(\frac{\mu}{\sigma}\right) = \frac{\beta}{1-\beta}$, where $\beta = \frac{K}{N}$.

For $\alpha-$Trimmed Sum Detector, the breakdown happens when the sum of ranked log likelihood ratio under H_0 is statistically greater than that under H_1. A closed-form solution is possible for small N and K. For example, consider 1 out of 3 observation is contaminated, detection performance is derived in equation(3.10). If $P_d \leq P_f$, then $\alpha-$Trimmed Sum Detector is useless. The minimum breaking down signal-to-noise ration (SNR), u_0 for the above detector to work properly can be obtained by solving $P_d = P_f$ at $\eta = 0$ and yielding.

$$u_0 = Q^{-1}\left(1 - \frac{1}{\sqrt{2}}\right).$$

For larger N, the distribution of order sum does not appear to have a closed form expression for normal distribution, and the analyses of breakdown points are currently unavailable.

3.4 Target Detection in Distributed MIMO radar

Since its introduction in the early 2000s, MIMO radar has quickly become an important and active radar research arena. MIMO radar is based on the idea of employing multiple antennas to transmit multiple waveforms and multiple antennas to receive echoes reflected by targets [45, 46, 47, 48, 49, 50]. With greater design flexibility in the choice of transmit waveforms, the placement of transmitting and receiving antennas, as well as the design of receiver processing algorithms, MIMO radars can potentially exhibit significantly improved performance characteristic relative to conventional radars [51, 52].

While enjoying a great potential in system performance gain, MIMO radar posts some challenges in system design and implementation. Compared with traditional radar systems, MIMO radars are often more complicated and require a high degree of synchronization among all their components [45]. Furthermore, when the receivers are distributed, the data collected by different receivers have to be transmitted to FC, where all the receiver data are jointly processed to make a final decision. These synchronization and data transmission tasks must be carried out via a communication network connecting the transmitters, receivers, and the FC.

Like many complex network systems, the functionality and performance of a MIMO radar are likely to be affected when the radar is not operated under the ideal nominal conditions. Unlike monostatic radar systems, MIMO radars are more vulnerable to malicious cybersecurity attacks due to the distributed setting and stringent system synchronization and data communications requirements. One failed component may affect many other components. For example, if one transmitter is unreliable, the radar signal at all receivers will be affected. If left untreated, this uncertainty may cause great damage to MIMO radar system performance. There are some results available to incorporate the statistical model mismatches by robust transmitter and receivers designs [53, 54], detect and monitor the sensor drifting [55, 56] and sensor fault [57, 58]. Unfortunately, only a few attempts have been made to study the performance degradation and robust system in the presence of severe degradations, such as cybersecurity attacks in MIMO radars. This chapter also addresses a challenge in distributed MIMO radars target detection under unreliable networks.

Let us now consider a target detection problem in a spatially distributed MIMO radar, similar to the one proposed and studied in [51]. The radar consists of well-spaced M transmitters, N receivers, and one fusion center which receives the signals from the receivers and makes the final decision. Similar to the signal model in [39, 59], the scope of the problem is restricted under the following conditions.

1. The transmitter waveform $S = [S_1, S_2, ..., S_M]$ is a set of narrow band signals, where S_m of size $L \times 1$ is the waveform of m^{th} transmitter and $L > M$ is the number of time samples. The waveform are normalized and orthogonal.

2. The background is clutter-free and target is stationary. At the n^{th} receiver, the sampled echo is given by

$$X_n = SG_n + W_n,$$

where X_n is an $L \times 1$ sampled signal, $G_n = [G_{1n}, ..., G_{Mn}]$ is the gain vector where $G_{mn} = r_{mn}\sigma_{mn}$ is the product of the path gain from m^{th} transmitter reflected at the target to n^{th} with corresponding Radar Cross Section (RCS)) $\sigma_{mn}, and W_n = [W_{1n}, W_{2n}, ..., W_{Ln}]$ represents the observation noise.

3. The receiver noise W_{ln} are normalized independently and identically distributed (i.i.d.) complex Gaussian noise such that $W_{ln} \sim C\mathcal{N}(\mathit{0}, \mathit{1})$

4. The path gains are assumed to be identical for all paths and the target is assumed to consists of a large number of scatters, as a result $\sigma_{mn} \sim C\mathcal{N}(\mathit{0}, \gamma)$, where γ is known as signal-to-noise ratio.

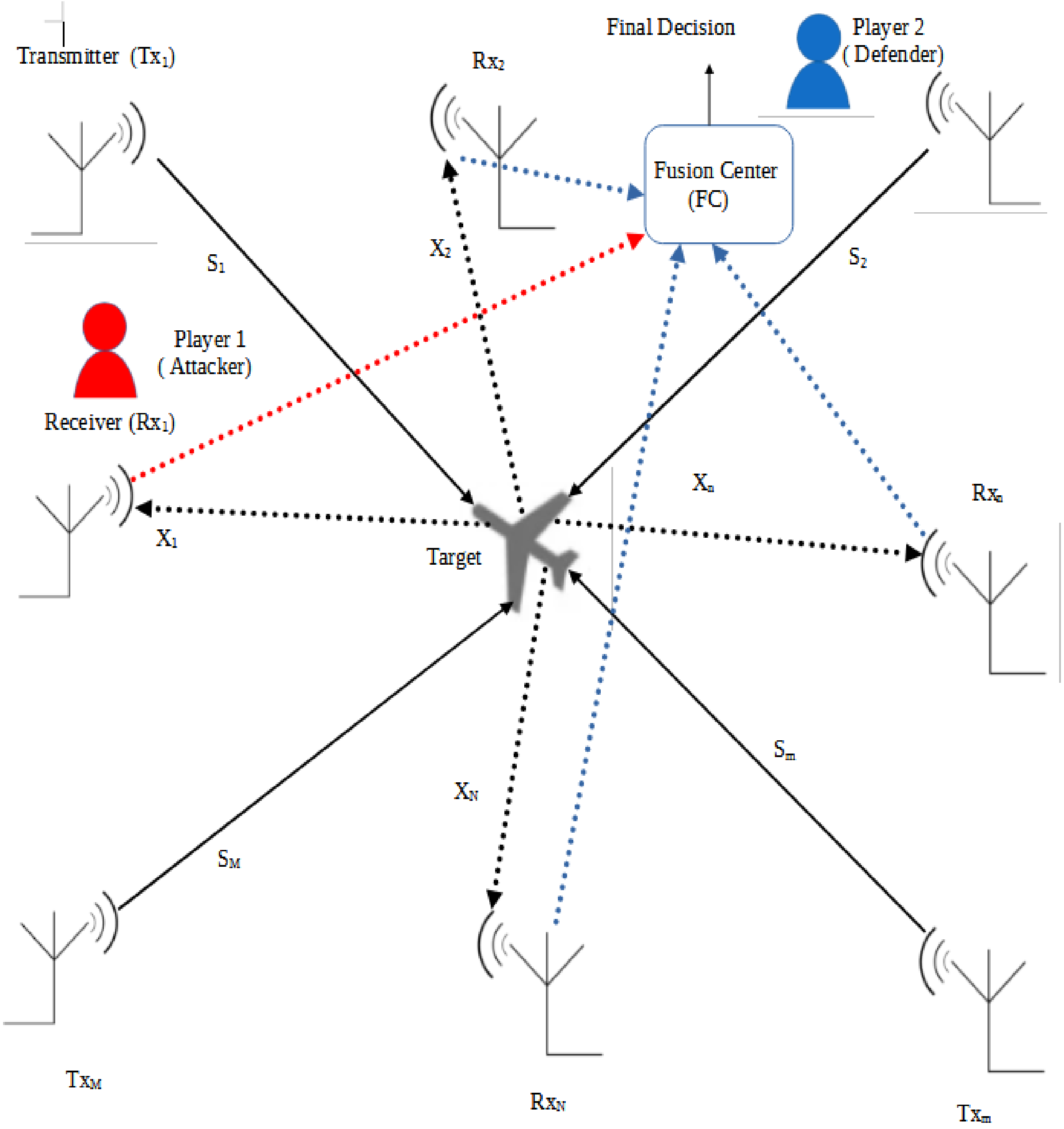

Figure 3.2: Detection in Distributed MIMO Radar.

For the detection problem under the aforementioned assumptions, the sufficient statistics is $Y = [Y_1^T, Y_2^T, ..., Y_N^T]$ such that

$$Y_{mn} = S_m^* X_n \sim \begin{cases} \mathcal{CN}(\mathit{0}, \gamma + \mathit{1}) : & H_1 \\ \mathcal{CN}(\mathit{0}, \mathit{1}) : & H_0 \end{cases}. \tag{3.11}$$

The optimal detector at fusion center under nominal condition is a non coherent detector [51] such that

$$T(X) = \sum_{m=1}^{M} \sum_{n=1}^{N} \|Y_{mn}\|^2$$

For centralized detection, since Y_{mn} follows a complex Gaussian distribution, $\frac{\|Y_{mn}\|^2}{1+\gamma} \sim \chi_{2MN}$ under H_1, and $\|Y_{mn}\|^2 \sim \chi_{2MN}$ under H_0, where χ_{2MN} is chi-square distribution with $2MN$ degrees of freedom.

The functionality and performance of distributed MIMO radar my be affected when radar is not operated under ideal nominal conditions. The detector performance in probabilistic setting is studied in [39]. As one receiver receives signals emanating from all M transmitters, one needs to jointly consider all data from all receive-transmit pairs, i.e.

$$\sum_{m=1}^{M} \sum_{n=1}^{N} \|Y_{mn}\|^2 = \left(\sum_{m=1}^{M} \|Y_{m2}\|^2, .., \sum_{m=1}^{M} \|Y_{mN}\|^2 \right). \tag{3.12}$$

Here, the case where K out of N receivers are unreliable is considered.

$$T_c(X) = \begin{cases} \sum_{i=1}^{N-K} \lambda\left(Y_{Mi}\right) + K\tau_0 & H_1 \\ \sum_{i=1}^{N-K} \lambda\left(Y_{Mi}\right) + K\tau_1 & H_0 \end{cases}, \tag{3.13}$$

where $Y_{Mi} = \sum_{m=1}^{M} \left\|Y_{mi}\right\|^2$ and $\lambda\left(Y_{Mi}\right)$ is a clipped log likelihood ratio

$$\lambda\left(Y_{Mi}\right) = \begin{cases} \tau_1 & Y_{Mi} \geq \tau_1 \\ Y_{Mi} & \tau_0 < Y_{Mi} < \tau_1 \\ \tau_0 & Y_{Mi} \leq \tau_0. \end{cases} \tag{3.14}$$

As a result, the pdf under H_0 and pdf under H_1 are given by

$$\begin{aligned} p\left(\lambda(Y_{Mi}) \mid H_0\right) = 1 - F_{\chi^2}(\tau_1, 2M)\delta\left(Y_{Mi} - \tau_1\right) + f_{\chi^2}(Y_{Mi}, 2M)I_{Y_{Mi}\varepsilon[\tau_0,\tau_1]} \\ + f_{\chi^2}(Y_{Mi}, 2M)\delta\left(Y_{Mi} - \tau_0\right), \end{aligned} \tag{3.15}$$

and

$$\begin{aligned} p\left(\lambda(Y_{Mi}) \mid H_1\right) = 1 - F_{\chi^2}(\frac{\tau_1}{1+\gamma}, 2M)\delta\left(Y_{Mi} - \tau_1\right) + \\ f_{\chi^2}(Y_{Mi}, 2M)[1+\gamma]I_{Y_{Mi}\varepsilon[\tau_0,\tau_1]} + F_{\chi^2}(\frac{\tau_0}{1+\gamma}, 2M) \\ \delta\left(Y_{Mi} - \tau_0\right). \end{aligned} \tag{3.16}$$

For smaller values of N, consider the case with $N = 3$, $K = 1$, the P_d and P_f are

$$\begin{aligned} P_d = Pr\left\{\lambda\left(Y_{M1}\right) + \lambda\left(Y_{M2}\right) > \eta - \tau_0 \mid H_1\right\}, \\ P_f = Pr\left\{\lambda\left(Y_{M1}\right) + \lambda\left(Y_{M2}\right) > \eta - \tau_1 \mid H_0\right\}. \end{aligned} \tag{3.17}$$

For a sufficiently large MN, the approximated asymptotic result of P_d and P_f is possible by equation (3.6). For any given P_f, the thresholds τ_0, τ_1 that maximize P_d is

$$\tau_0^*, \tau_1^* = \arg\min_{\tau_0,\tau_1} Q\left(Q^{-1}(P_f)\sqrt{\frac{v_0}{v_1}} + \Phi\right), \tag{3.18}$$

where $\Phi = \frac{K(\tau_1-\tau_0)+(N-K)(u_0-u_1)}{\sqrt{(N-K)v_1}})$.

Similarly, for the $\alpha-$Trimmed Sum Detector, the distribution of the sum of ranked log-likelihood ratios is needed. A closed-form solution is possible for small N and K. For example, consider 1 out of 3 observations is contaminated. The $\alpha-$trimmed sum on the log-likelihood ratios is the median of the log-likelihood ratio after removing the biggest and smallest values. If Y_3 is compromised, the test statistic is

$$T_\alpha(X) = median\left\{Y_{1,}Y_{2,}Y_3\right\} \gtrless^{H_1}_{H_0} \eta. \tag{3.19}$$

$T_\alpha(X)$ is as small as possible under H_1 and as large as possible under H_0. When the compromised observation is $-\infty$ under H_1 and ∞ under H_0,

$$T_\alpha(X) = \begin{cases} min\left\{Y_1, Y_2\right\}: & H_1 \\ max\left\{Y_1, Y_2\right\}: & H_0, \end{cases} \tag{3.20}$$

then,

$$\begin{aligned} P_f &= 1 - \left[F_{\chi^2}(\eta, 2M)\right]^2, \\ P_d &= \left[1 - F_{\chi^2_a}(\frac{\eta}{1+\gamma}, 2M)\right]^2. \end{aligned} \tag{3.21}$$

For larger N, from equation (3.12), all MN entries of the test statistic Y_{mn} are needed to make the final decision. Assuming all MN data are carried in MN independent parallel (virtual or real) channels and the attacker can modify the data content Y_{mn} arbitrarily according to some designed distributions (strategies). The attacker may degrade the MIMO radar detection performance by falsifying the data in the compromised communication link to send false information to the FC. The analysis of communication link attack can be treated as a receiver attack in a virtual single-input multiple-output (SIMO) radar with 1 transmitter and MN receivers. In this case, each distribution $\sum_{m=1}^{M}\sum_{n=1}^{N}\|Y_{mn}\|^2 = (Y_1, Y_2, ..., Y_N)$ follows the exponential distribution. In such cases, we can obtain the exact closed-form solution of the order statistic and is derived in appendix A.

From equation (3.18), the detection performance depends on M, N, K, and thresholds τ_0, and τ_1. The detector is meaningful if $P_d > P_f$. For the particular M, N, K, the breakdown point is reached if $P_d \leq P_f$. Due to the asymmetry of this problem, it is very difficult to derive the breakdown point exactly therefore, it can used equation (3.4) to analyze the breakdown point.

For the $\alpha-$Trimmed Sum Detector, the breakdown point happens when the sum of ranked log likelihood ratios under H_0 is statistically greater than that under H_1. For example, consider 1 out of 3 observations is contaminated, the detection performance is derived in equation (3.21). If $P_d \leq P_f$, then the $\alpha-$Trimmed Sum Detector is useless. The minimum breakdown signal-to-noise ratio (SNR), γ_0 for the above detector to work properly can be obtained by solving $P_d = P_f$, which yields

$$\gamma_0 = \frac{F_{\chi^2}^{-1}(\sqrt{1-P_d}, 2M)}{F_{\chi_a^2}^{-1}(1-\sqrt{P_d}, 2M)} - 1.$$

For large N, it does admit a closed-form solution for target detection in distributed MIMO radar case and is evaluated using appendix (B.1).

3.5 Simulation and Validation

In this section, some numerical simulation results of CLLRD and $\alpha-$Trimmed Sum Detector, obtained by Monte-Carlo simulation are presented in the presence of worst possible adversary in term of Receiver Operating Characteristic curves (ROC). In order to prepare for worst possible scenario, it is assumed that first K compromised observations are very small numbers under H_1, and large numbers under H_0. The attacker employs an independent and identical attack strategy at each compromised component.

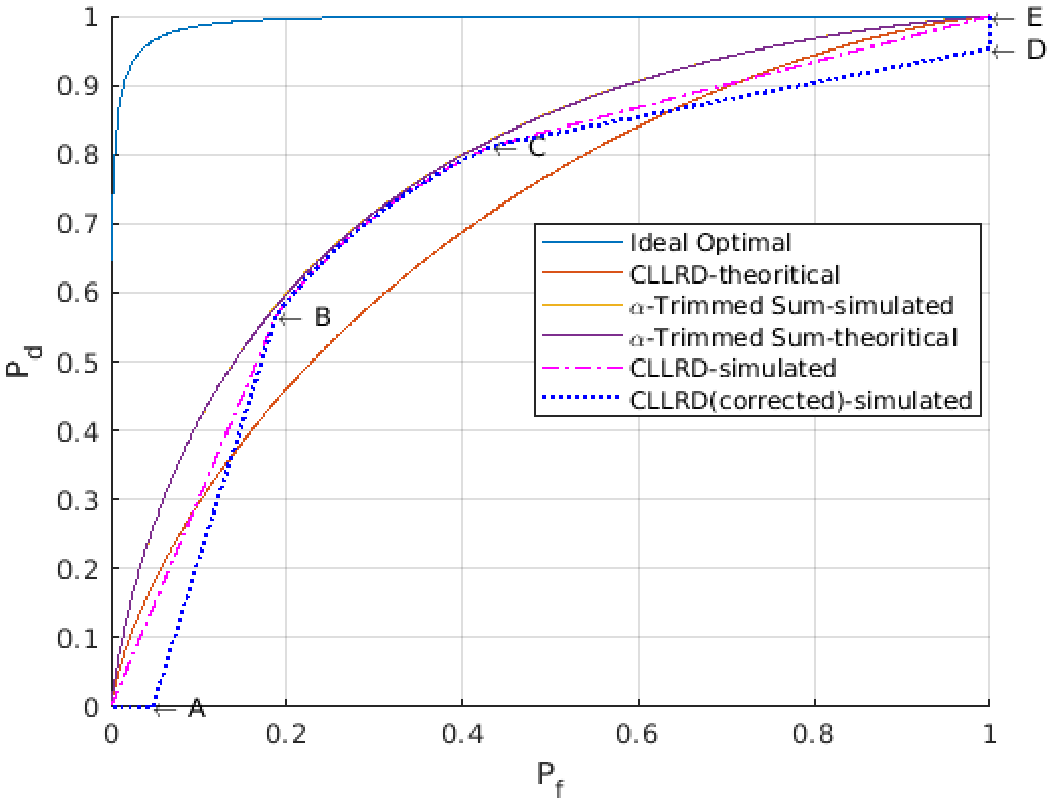

Figure 3.3: ROC curves at $N = 3$, $K = 1$, $\gamma = 0$ dB for normally distributed observation.

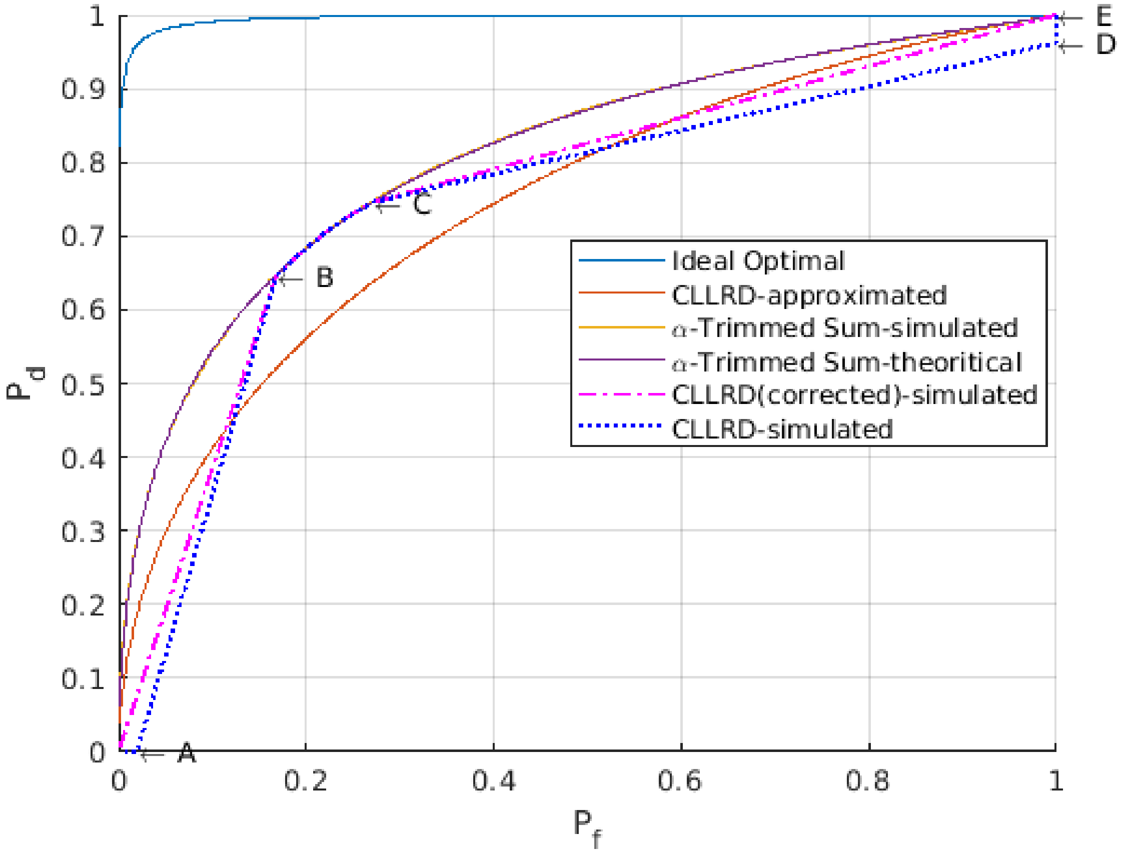

Figure 3.4: ROC curves at $N = 3$, $M = 10$, $K = 1$, $\gamma = 0$ dB for MIMO radar.

Fig 3.3 and Fig 3.4 show the performance of CLLRD and $\alpha-$Trimmed Sum Detector for $N = 3$, $M = 10$, $K = 1$, and $SNR(\gamma) = 0$ dB. The test statistics $T_c(X) \in [-3\tau, 3\tau]$ for the worst case outlier are considered and evaluated. If the fusion center is aware of the attacking strategy but the detector designer does not want to change the CLLRD, then the detection is based on equation(3.5). Under H_0, the test statistics $T_c(X) \in [-\tau, 3\tau]$ and under H_1, the test statistics $T_c(X) \in [-3\tau, \tau]$. When $\eta = \tau + \varepsilon$ for any $\varepsilon > 0, P_d = 0$, but $P_f > 0$ as shown in the region from

$(0,0)$ to A, and when $\eta = \tau - \varepsilon$, $P_d < 1$, but $P_f = 1$ as shown in region D to E in Fig 3.3. The resulting offsets in P_f and P_d due to the small N can be corrected by randomization in between $(0,0)$ to B and C to E as shown in Fig 3.3. It is observed that the $\alpha-$Trimmed Sum Detector performs better for a few number of observations while the CLLRD performs better with the larger number of observations. These results show that *neither detector* is mini-max optimally.

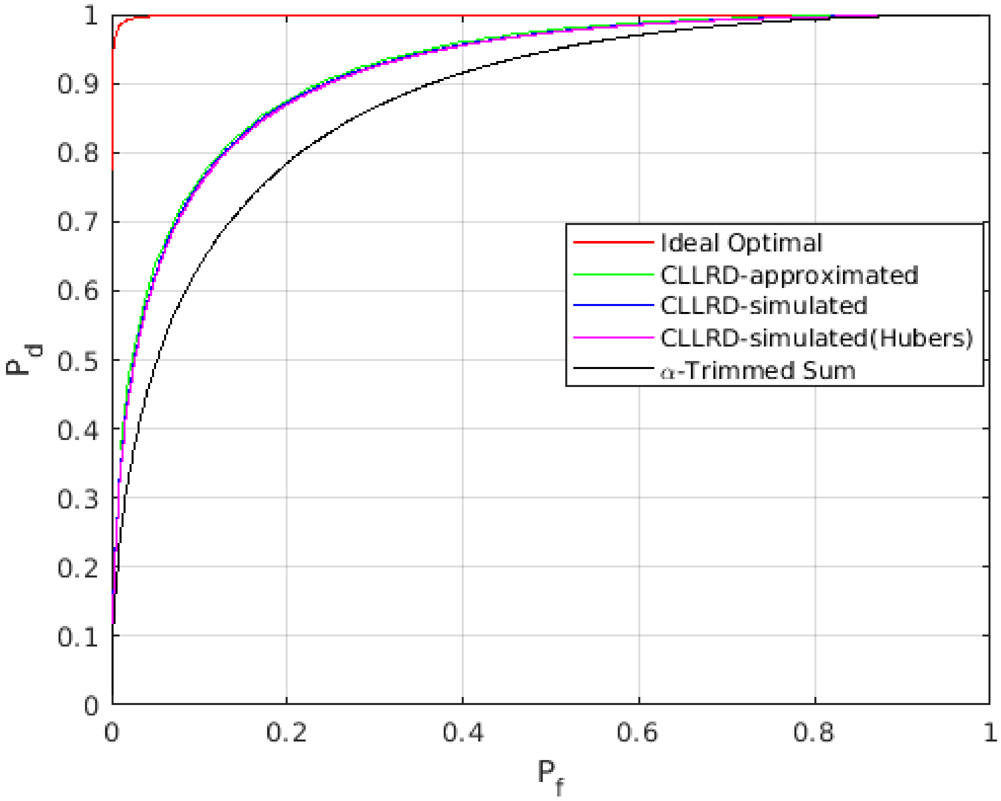

Figure 3.5: ROC curves at $N = 20$, $K = 3$, $\gamma = -3$ dB for normally distributed observation.

Fig 3.5 shows the performance CLLRD and $\alpha-$trimmed sum detector at $N = 20$,

$K = 3$, $\gamma = -3$ dB for normally distributed observation. Here we also consider the CLLRD in static case by using threshold proposed by Huber [44] for least favorable distribution in probabilistic setting. The CLLRD with threshold obtained by equation (3.9) and that proposed by Huber's performs almost same.

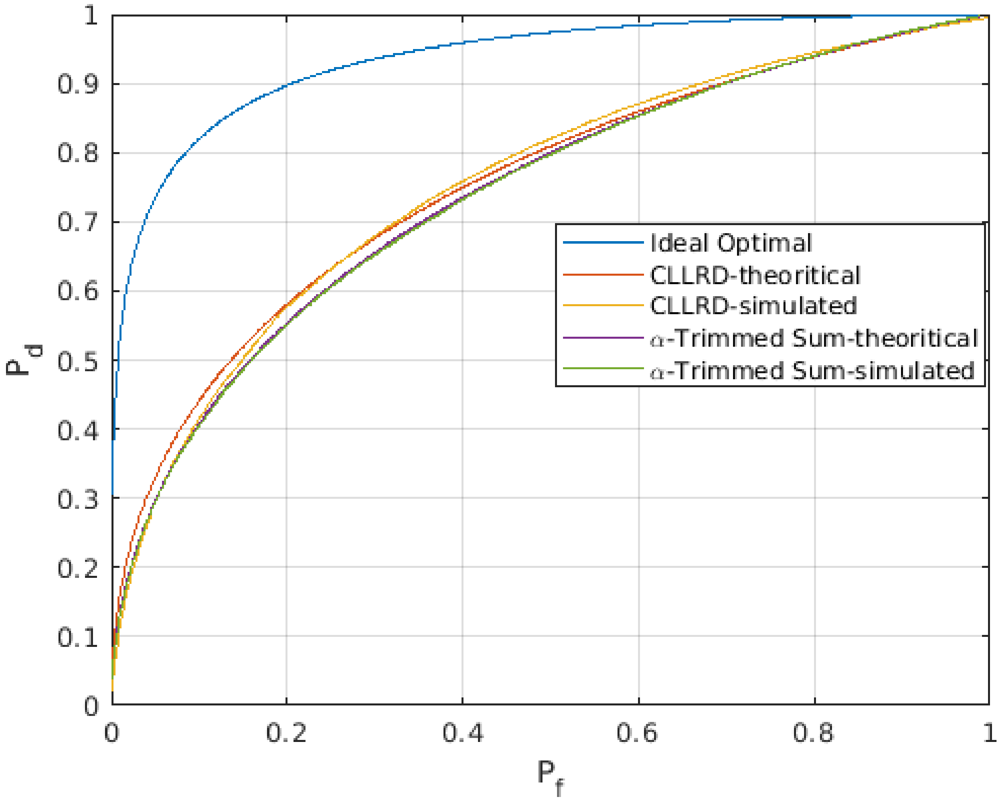

Figure 3.6: ROC curves at $N = 10$,$M = 1$, $K = 1$, $\gamma = 0$ dB for MIMO radar.

In the case of MIMO radar, computation of threshold for truncated likelihood ratio is obtained by solving equation (3.18). As their performance is almost similar with Huber's thresholds, we use it to simulate CLLRD for simplicity and comparative performance with $\alpha-$trimmed sum detector is as shown in Fig 3.6 for $N = 10$, $M = 1$,

$K = 1$, $\gamma = 0$ dB.

The $\alpha-$Trimmed Sum Detector is simpler to design and performs better at smaller value of N and K. The CLLRD performs better for larger N.

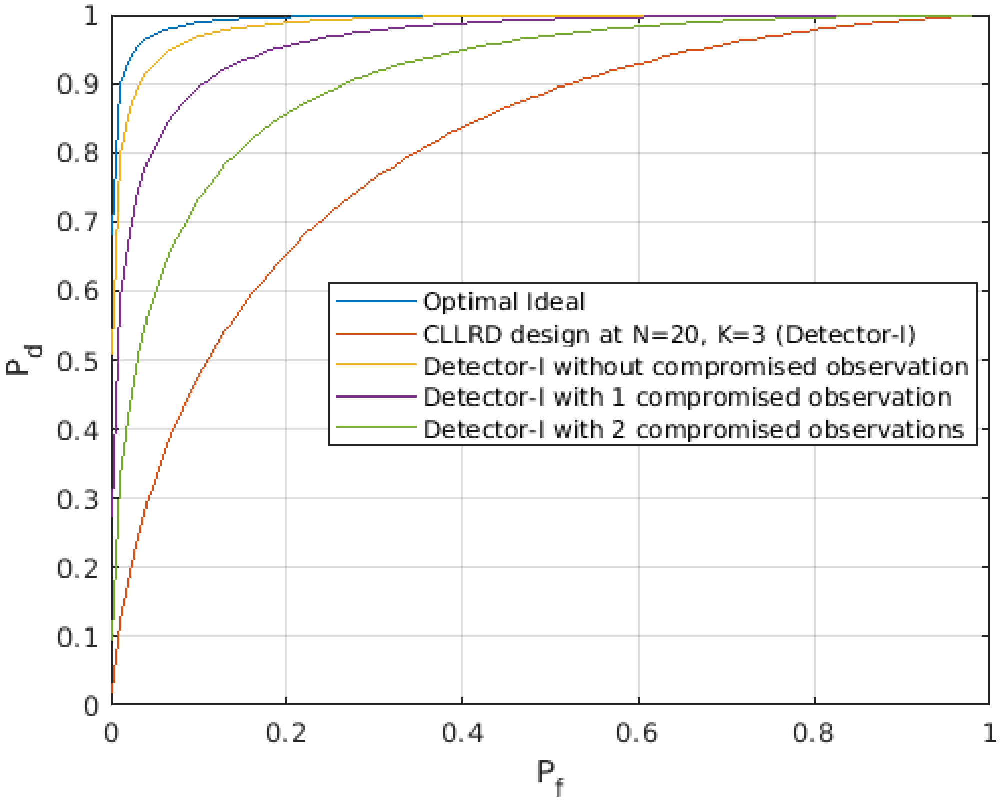

Figure 3.7: ROC curves at $N = 20$, $K = 3$, $\gamma = -4$ dB for normally distributed observation with different outliers.

Fig 3.7 shows the performance at various levels of compromised observations. We design our detector with the case where there are $K = 3$ outliers in observations. In case 1 observation and 2 observations are compromised, the performance are as shown in ROC curves. The gap in performance curves for CLLRD with no observation compromised, 1 observation compromised and 2 observations compromised in a static setting with $K = 3$ show the prices that have to pay for detection performance due to robustness.

CHAPTER 4:
ROBUST ESTIMATION

This chapter presents the robust estimation theory under the probabilistic and static models. Besides the standard approaches of robust estimation under the frequentist settings where the parameters of interest are fixed but unknown, the estimation problem is mostly considered under the Bayes settings, where the prior probability distribution of the parameter is known. First, we establish the framework, and present the comprehensive result in the case of a single node network under the probabilistic setting. The estimation problem is then investigated for the general multi-node framework based on the insights shed by the single node network.

4.1 Problem formulation

Consider an estimation problem with N sensors where the ith sensor observations is $X_i,\ i = 1, 2, \cdots, N$. Under the nominal operating conditions, X_i is conditionally independent from other sensors and follows the model

$$p_i\left(X_i|\theta, \theta_a\right), \tag{4.1}$$

where $\theta \in \Theta$ is the parameter of interest, $\theta_a \in \Theta_a$ is the auxiliary parameter, and $p_i\left(X_i|\theta, \theta_a\right)$ is the conditional probability density function. When the sensors are unreliable, we assume that with $\varepsilon_i > 0$ probability X_i may follow a different model $q_i\left(X_i|\theta, \theta_a\right)$, the overall sensor observation model becomes

$$\tilde{p}\left(X_i = x_i|\theta, \theta_a\right) = \left(1 - \varepsilon_i\right) p_i\left(X_i = x_i|\theta, \theta_a\right) + \varepsilon_i q_i\left(X_i = x_i|\theta, \theta_a\right). \tag{4.2}$$

The performance departure from the nominal settings is due to the uncertainty ε_i and the observation model q_i. We first analyze the impact of ε_i and assume q_i can be completely known. This assumption is roughly valid where the sensors operate under binary models, and the chance of switching to the abnormally model is ε_i.

To estimate $g\left(\theta\right)$, a known function of θ, the goal of the estimation problem is to form $\hat{g}\left(\theta\right)$ based on the data $\mathbf{x} = \left[X_1, \cdots X_N\right]^T$. Let Q_i be the set of all possible q_i and $F = \left\{f|\hat{\theta} = f\left(X_{1,}X_2, \cdots, X_N\right)\right\}$ be the set of all possible estimates, and $d\left(\left(g\left(\theta\right), \hat{g}\left(\theta\right)\right)\right) \geq 0$ or $C\left(\hat{g}\right) = E\left(d\left(g\left(\theta\right), \hat{g}\left(\theta\right)\right)\right)$ when the prior of θ known, is the estimation cost function. In many applications, $g\left(\theta\right) = \theta$ is the parameter itself. The performance change is due to the uncertainty ε_i and the new observation model q_i.

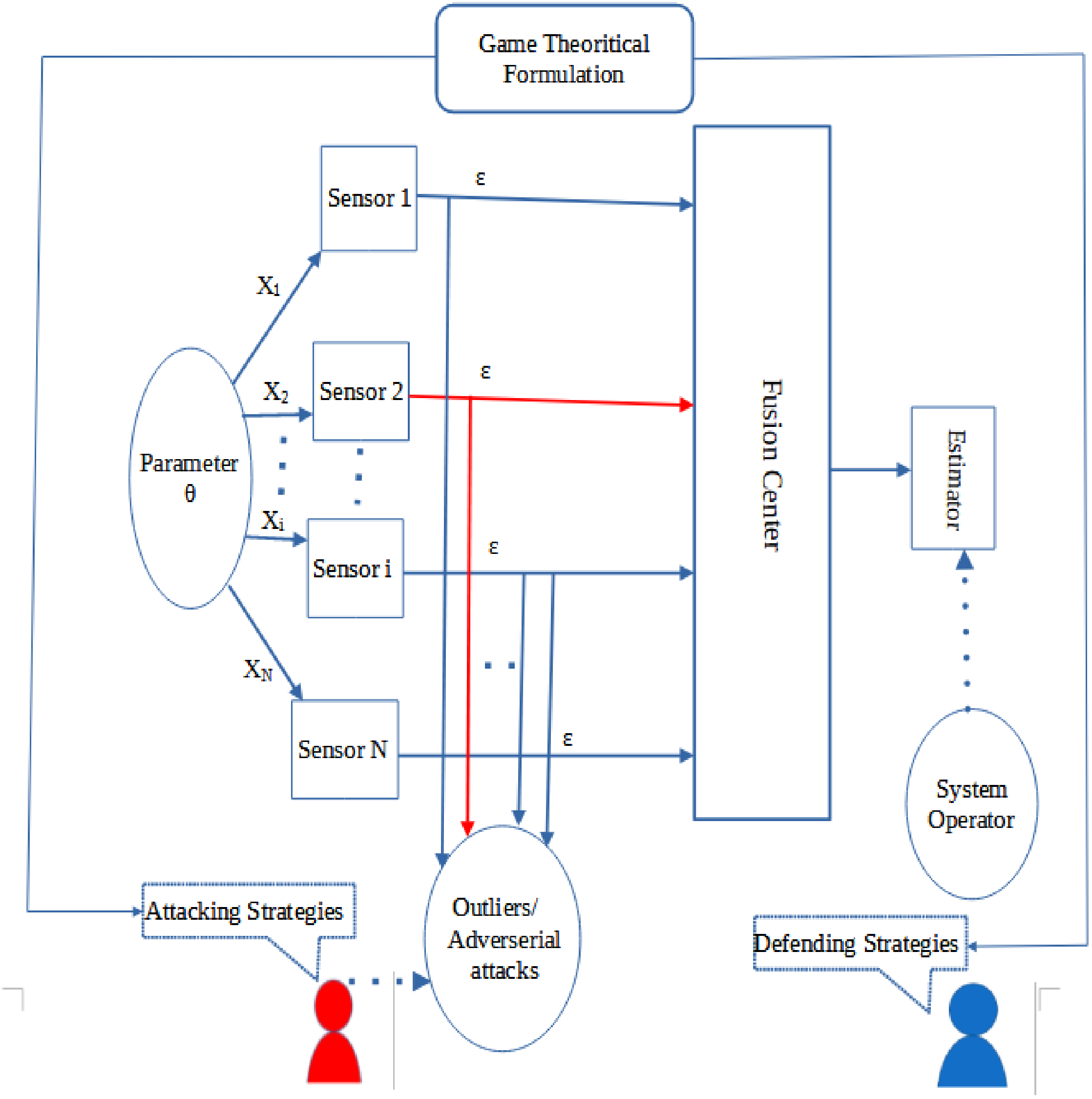

Figure 4.1: Robust Estimation Framework

This problem can be formulated as a two-player zero-sum game: the $q_i \in Q$ choice maker as Player 1, the estimator $\hat{\theta} \in$ F as Player 2, and the estimation error C as the reward to the Player 1 and the negative of error as the reward to Player 2. We assume that p_i, the pdf under the nominal condition, and N, the sensor network size, are known to both players. Usually, the estimator can arbitrarily choose its estimate. However, the choices of q_i depend on the knowledge of Player 1 about θ, θ_a.

In practice, there are situations where either Player 1 or Player 2 has private information about the parameter θ. Player 1 may have a prior knowledge in terms of a prior distribution $\rho(\theta, \theta_a)$, e.g., Player 1 is the target itself in radar detection. Such information helps Player 1 design a proper strategy to optimize performance. Player 2 may have its observations, which are inaccessible to Player 2. For example, some secured sensor observations are typically unknown to the malfunctioning nodes in a parallel network setting.

While it is safe to say that the consistent estimation of $g(\theta)$ at Player 2 is almost always impossible due to the uncertainty in the observation model, a necessary analysis is warranted to quantify the exact impact due to the node unreliability. Depending on Player $1'$s information, we can categorize such knowledge into three categories: complete, partial, and none.

1. Complete: In this case, Player 1 has a complete knowledge of θ, θ_a, or equivalently, $\rho(\theta, \theta_a)$ degrades into a point mass function at the true θ and θ_a. The choice of $Q_i = \Omega$ which is the set of all possible distributions.

2. Partial: In this case, Player 1 has some knowledge of θ, θ_a and $Q_i = \{q | q = q(X_i, \rho)\}$ is the set of X_i distribution.

3. None: In this case, Player 1 doesn't have any knowledge of θ, θ_a, and $Q_i = \{q | q = q(X_i)\}$ is the set of any X_i distribution that is independent of the parameters

4.2 Estimation in Frequencist Setting

In this section, we investigate robust estimation under the frequentist settings where the parameters of interest are fixed but unknown.

4.2.1 Complete Informative

For the complete information case where θ and θ_a are known to Player 1, he can choose a q_i which minimizes the information carried by X_i. This is essentially a mini-max approach. In particular, if there exists a q_i such that for a pair of $\theta_1 \neq \theta_2$, $\tilde{p}\left(X_i|\theta_1,\theta_a\right) = \tilde{p}\left(X_i|\theta_2,\theta_a\right)$, i.e., the conditional observation distributions are the same. Then, a consistent estimation will not be possible since one cannot gain any information from the observation to distinguish between θ_1 and θ_2.

Theorem 4.2.1 *Under the complete informative case where the unreliable sensor can choose any arbitrary distribution, the sufficient and necessary condition for observation X_i being completely uninformative about θ is*

$$\varepsilon \geq \varepsilon_0\left(\Theta\right) = 1 - \frac{1}{\int \sup_{\theta\in\Theta} p_i\left(X_i|\theta,\theta_a\right) dX_i} \tag{4.3}$$

Proof: Under the $\varepsilon-$contamination model, this condition is equivalent to

$$\begin{aligned}\left(1-\varepsilon_i\right) p_i\left(X_i|\theta_1,\theta_a\right) + \varepsilon_i q_i\left(X_i|\theta_1,\theta_a\right) &= \left(1-\varepsilon_i\right) p_i\left(X_i|\theta_2,\theta_a\right) + \varepsilon_i q_i\left(X_i|\theta_2,\theta_a\right) \\ &= \tilde{p}\left(X_i|\theta_a\right)\end{aligned} \tag{4.4}$$

When q_i can be picked arbitrarily, notice that

$$\tilde{p}\left(X_i|\theta_a\right) \geq \max\left\{\left(1-\varepsilon_i\right)p_i\left(X_i|\theta_2,\theta_a\right),\left(1-\varepsilon_i\right)p_i\left(X_i|\theta_1,\theta_a\right)\right\},$$

a necessary and sufficient conditional for condition (4.4) observation model to be satisfied is

$$\int \max\left\{\left(1-\varepsilon_i\right)p_i\left(X_i|\theta_2,\theta_a\right),\left(1-\varepsilon_i\right)p_i\left(X_i|\theta_1,\theta_a\right)\right\}dX_i \leq 1$$

or equivalently,

$$\begin{aligned}
\varepsilon \geq \varepsilon_0\left(\theta_1,\theta_2\right) &= 1-\frac{1}{\int \max\left\{p_i\left(X_i|\theta_2,\theta_a\right),p_i\left(X_i|\theta_1,\theta_a\right)\right\}dX_i}\\
&= 1-\frac{1}{\int \frac{1}{2}\left\{p_i\left(X_i|\theta_2,\theta_a\right)+p_i\left(X_i|\theta_1,\theta_a\right)+\left|p\left(X_i|\theta_2,\theta_a\right)-p_i\left(X_i|\theta_1,\theta_a\right)\right|\right\}dX_i}\\
&= 1-\frac{1}{1+\frac{1}{2}d\left(p_i\left(X_i|\theta_2,\theta_a\right),p_i\left(X_i|\theta_1,\theta_a\right)\right)}\\
&= \frac{d\left(p_i\left(X_i|\theta_2,\theta_a\right),p_i\left(X_i|\theta_1,\theta_a\right)\right)}{2+d\left(p_i\left(X_i|\theta_2,\theta_a\right),p_i\left(X_i|\theta_1,\theta_a\right)\right)},
\end{aligned} \tag{4.5}$$

where $d\left(p\left(x\right),q\left(x\right)\right)=\int\left|p\left(x\right)-q\left(x\right)\right|dx$ is the first order distance between to two pdfs. More over, if ε is big enough such that for the mixed distribution function can be made the same for all θ, i.e.,

$$\tilde{p}\left(X_i|\theta,\theta_a\right)=\left(1-\varepsilon_i\right)p_i\left(X_i|\theta,\theta_a\right)+\varepsilon_i q_i\left(X_i|\theta,\theta_a\right)=\tilde{p}\left(X_i|\theta_a\right), \tag{4.6}$$

and the resulting observation distribution is independent of θ, then the observation X_i is completely uninformative about θ. The corresponding distribution q_i is given by

$$q_i\left(X_i|\theta,\theta_a\right)=\frac{1}{\varepsilon_i}\left(\tilde{p}\left(X_i|\theta_a\right)-\left(1-\varepsilon_i\right)p_i\left(X_i|\theta,\theta_a\right)\right). \tag{4.7}$$

Example 1 : Binomial Random Variables

Under the nominal case, let X_i be an independent and identically distributed (i.i.d.) Bernoulli random variables with a success probability θ, i.e., $p_i\left(X_i=1|\theta\right)=\theta$, and $\varepsilon_i=\varepsilon$. $\Theta=\left(\theta_l,\theta_u\right)$, $0\leq\theta_l<\theta_u\leq 1$, there is no auxiliary random variable. In order to make X_i completely uninformative, one needs $\tilde{p}\left(X_i|\theta\right)=\left(1-\varepsilon\right)$
$p\left(X_i|\theta\right)+\varepsilon q\left(X_i|\theta\right)$ be a constant, which can be met when

$$\begin{aligned}\varepsilon\geq\varepsilon_0&=\frac{p\left(X_i=1|\theta_u\right)-p\left(X_i=1|\theta_l\right)}{1+p\left(X_i|\theta_u\right)-p\left(X_i|\theta_l\right)}=\frac{\theta_u-\theta_l}{1+\theta_u-\theta_l}\\ q\left(X_i|\theta\right)&=\frac{1-\varepsilon_0}{\varepsilon}\left(p\left(X_i=1|\theta_u\right)-p\left(X_i=1|\theta\right)\right)+\frac{\varepsilon-\varepsilon_0}{\varepsilon}p\left(X_i=1|\theta\right)\\ &=\frac{1-\varepsilon_0}{\varepsilon}\left(p\left(X_i=1|\theta_u\right)-p\left(X_i=1|\theta\right)\right)+\frac{\varepsilon-\varepsilon_0}{\varepsilon}p\left(X_i=1|\theta\right).\end{aligned}$$

This results in

$$\tilde{p}\left(X_i=1|\theta\right)=\left(1-\varepsilon\right)p\left(X_i=1|\theta\right)+\varepsilon q\left(X_i=1|\theta\right)=\left(1-\varepsilon_0\right)\theta_u,$$

and X_i becomes completely uninformative. Fig 4.2 shows the minimal required ε_0 for varies θ_u when the lower limit $\theta_l=0.1$. The narrower θ range is, the easier the observation becomes completely uninformative. When $\varepsilon<\varepsilon_0$, complete uninformative becomes unachievable since $\tilde{p}\left(X_i=1|\theta_l\right)<\tilde{p}\left(X_i=1|\theta_u\right)$ regardless the choice of q. However, partial uninformative is still well within reach. This can be achieved by a "stair-case" approach with the division of the range of θ into K intervals such that $\theta_l=\theta_1<\theta_1<\theta_2\cdots<\theta_{K+1}=\theta_u$, which results in a constant $\tilde{p}\left(X_i=1|\theta\right)=p_k$ for

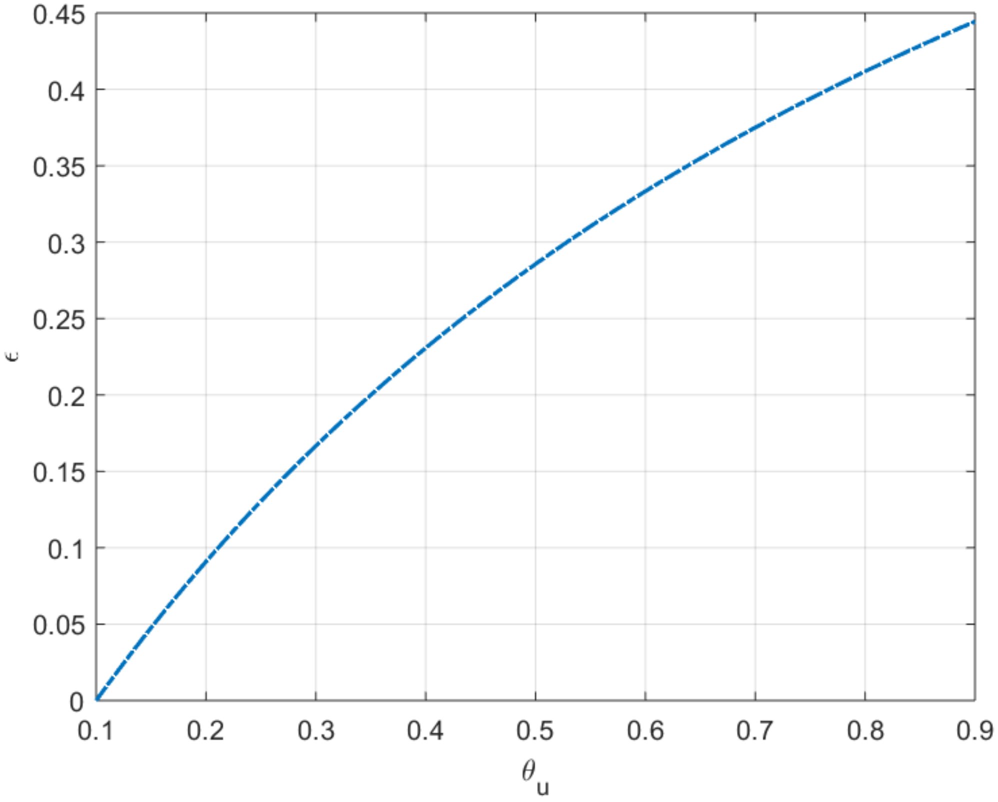

Figure 4.2: Minimal ε_0 to achieve complete uninformative when $\theta_l = 0.1$.

any $\theta \in (\theta_k, \theta_{k+1}), k = 1, 2, \cdots, K$. This requires that $\varepsilon \geq \max(\varepsilon_0(1), \cdots, \varepsilon_0(K))$, where $\varepsilon(k) = \frac{\theta_{k+1}-\theta_k}{1+\theta_{k+1}-\theta_k}$. When the intervals are evenly spaced, $\theta_{k+1}-\theta_k = \frac{1}{K}(\theta_u - \theta_l)$ and the minimal required $\varepsilon_0 = \varepsilon_0(k) = \frac{\theta_u-\theta_l}{K+\theta_u-\theta_l}$ which can be made arbitrary small by increasing K. In this case, the best one can learn is which interval θ lies, increasing number of sensors will not be able to further reduce any uncertainty and provide any further information.

For single sensor scenarios, the optimal MSE and optimal estimate can be derived when the observation is unreliable. Under the nominal settings, $\epsilon = 0$, the estimate $\hat{\theta} = E[\theta|U] = \frac{2}{3}$.MSE=$\int_0^1 \left\{\theta(\theta - \frac{2}{3})^2 + (1-\theta)(\theta - \frac{1}{3})^2\right\} d\theta = \frac{1}{18}$. Under complete unreliable,ε is 1, MSE=$\int_0^1 \left\{(\theta - \frac{1}{2})^2\right\} d\theta = \frac{1}{12}$. Hence the error is in between $\frac{1}{12}$ to $\frac{1}{18}$ when the observation is unreliable. The Player 1 strategy q to maximize the error would be to send

$$q = \begin{cases} 1 & \theta < \frac{1}{2} \\ 0 & \theta \geq \frac{1}{2} \end{cases}$$

Example 2: Location Estimation in Gaussian Observation Noise

Under the nominal case, let X_i be i.i.d. Gaussian random variables with mean θ and variance σ^2, i.e., $p_i\left(X_i = x|\theta\right) = \frac{1}{\sqrt{2\pi\sigma^2}}$ $\exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)$, and $\varepsilon_i = \varepsilon$. $\Theta = \left(\theta_l, \theta_u\right), -\infty < \theta_l < \theta_u < \infty$, there is no auxiliary random value.

In this case, for any $-\theta_u < \theta_1 < \theta_2 < \theta_u$,

$$\int \max\left\{p_i\left(X_i|\theta_2, \theta_a\right), p_i\left(X_i|\theta_1, \theta_a\right)\right\} dX_i = 2\Phi\left(\frac{\theta_2 - \theta_1}{2\sigma}\right), \tag{4.8}$$

where $\Phi\left(x\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$ is the standard Gaussian commutative distribution function. Therefore, as long as

$$\varepsilon \geq \varepsilon_0\left(\theta_1, \theta_2\right) = 1 - \frac{1}{2\Phi\left(\frac{\theta_2 - \theta_1}{2\sigma}\right)},$$

then one can make θ_1 and θ_2 indistinguishable based on X_i.

Notice that $\sup_{\theta(\theta_l,\theta_u)} \in p_i\left(X_i|\theta\right)$ is flat at $\frac{1}{\sigma\sqrt{2\pi}}$ for $X_i \in \left(\theta_l, \theta_u\right)$ and a standard Gaussian centered at θ_l or θ_u at either side, we have $\int \sup_{\theta\in\Theta} p_i\left(X_i|\theta\right) dX_i = \frac{\theta_u - \theta_l}{\sigma\sqrt{2\pi}} + 1$

and therefore to achieve complete uninformative, the minimum sufficient ε_0 is

$$\varepsilon_0 = 1 - \frac{1}{\frac{\theta_u - \theta_l}{\sigma\sqrt{2\pi}} + 1} = \frac{\theta_u - \theta_l}{\theta_u - \theta_l + \sigma\sqrt{2\pi}}. \tag{4.9}$$

Notice that $\varepsilon_0 = 1$ when $\theta_l = -\infty$ and $\theta_u = \infty$, i.e., the complete uninformative is not achievable, if the nodes are at least somewhat reliable.

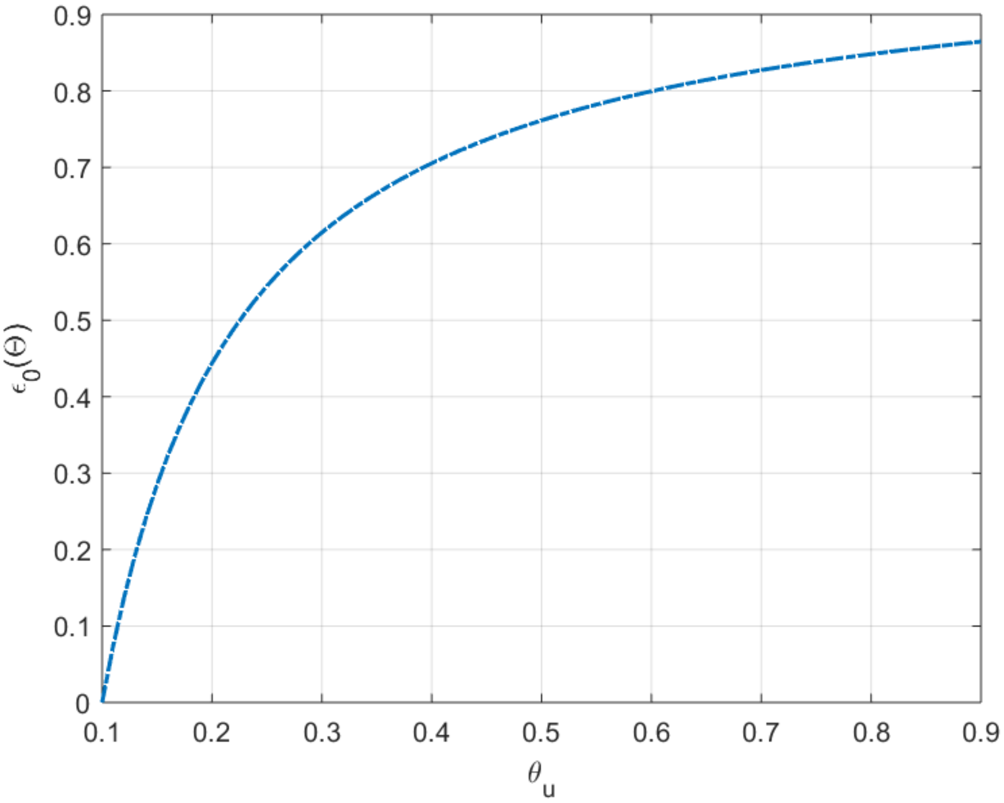

Figure 4.3: Minimal ε_0 to achieve complete uninformative.

Fig 4.3 depicts the minimal required ε_0 for complete uninformative as a function of range $\theta_u = -\theta_l$ where $\sigma^2 = 1$. As also seen in equation (4.9) Gaussian Complete uninformative, ε_0 increases and approaches to 1, when θ_u increases and approaches

∞.

4.2.2 No Informative Case

In this case, player 1 doesn't have any knowledge of θ, θ_a, $Q_i = \{q | q = q(X_i)\}$ is the set of any X_i distribution that is independent of the parameters X_i is pure noise.

Example: Binomial Random Variables

Consider the binary information under the attack with transition probability ρ_0 and ρ_1. Received data will be completely useless when

$$\begin{aligned}\gamma(\theta) = \tilde{p}(X_i = 1|\theta) = (1-\varepsilon)\, p(X_i = 1|\theta) + \\ \varepsilon q(X_i = 1|\theta)(1-\varepsilon)\,\theta \\ = (1-\varepsilon)\,\theta + \varepsilon(1-\rho_1) + \varepsilon\theta(1-\rho_0),\end{aligned}$$

is least informative. The information contained in the received bit $I(\theta, \gamma(\theta)) = \left[\frac{(\delta(\gamma(\theta))^2}{\gamma(\theta)(1-\gamma(\theta)}\right] = \frac{(1-\varepsilon(\rho_1+\rho_0))^2}{\gamma(\theta)(1-\gamma(\theta))}$ is monotonic decreasing function so the $I(\theta, \gamma(\theta))$ will be the least informative at $\rho_0 = \rho_1 = 1$. Hence flipping of the binary information is the best attacking strategy when Player 1 does not have any knowledge of θ.

For single sensor observation, when Player 1 does not have any information about the parameter θ, the best way to maximize the MSE is to flip the information with some probability $\rho = \rho_1 = \rho_0$. Under the non-informative case,

$$MSE = \int_0^1 \{\theta(1-\rho\varepsilon) + (1-\theta)(\varepsilon\rho)\}\,(\theta - \hat{\theta})^2 d\theta,$$

and it is maximum when $\rho = 1$. The optimal estimate $\hat{\theta}$ can be obtained by minimizing MSE $\left(\frac{\partial MSE}{\partial\hat{\theta}} = 0\right)$, i.e, $\hat{\theta} = \frac{2-\varepsilon}{3}$.

4.2.3 Simulation Results for The Frequentist Setting

Here, some results based on our derivations are presented for a complete informative and non-informative scenario. Fig 4.4 shows the plot between optimal θ and sensor unreliability parameter ε. It shows that the estimation result becomes useless for lower unreliable parameter ε when attacker knows the information about θ compared to the case when the attacker does not have any information about it.

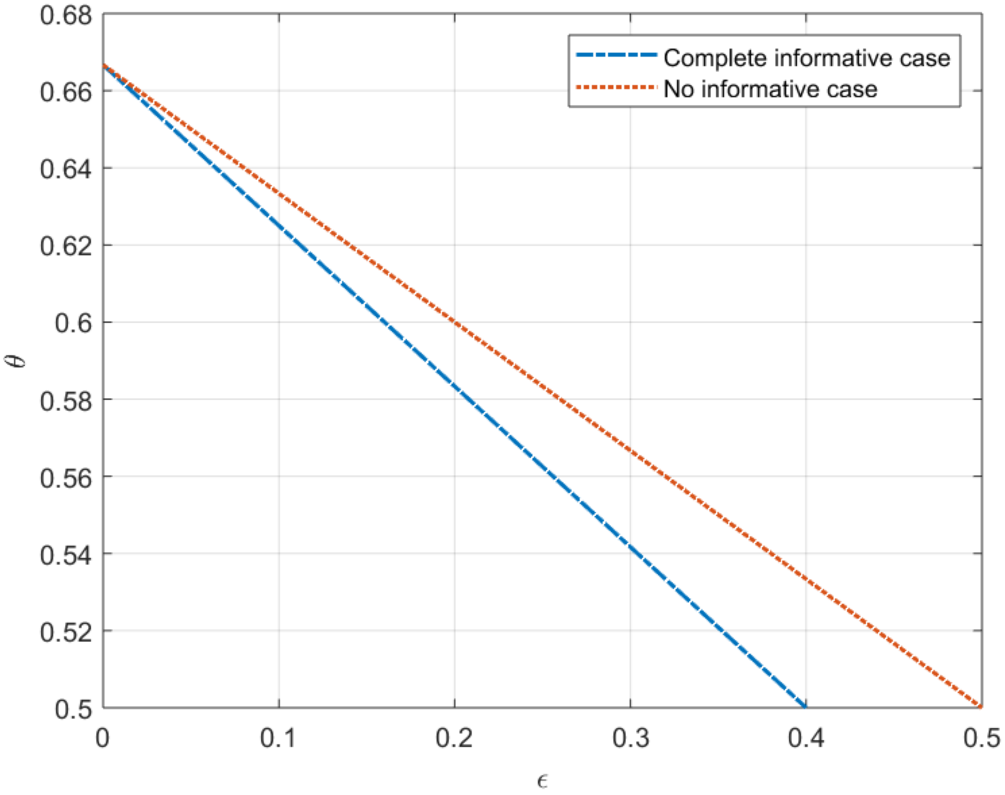

Figure 4.4: Estimation under sensor unreliability for one sensor.

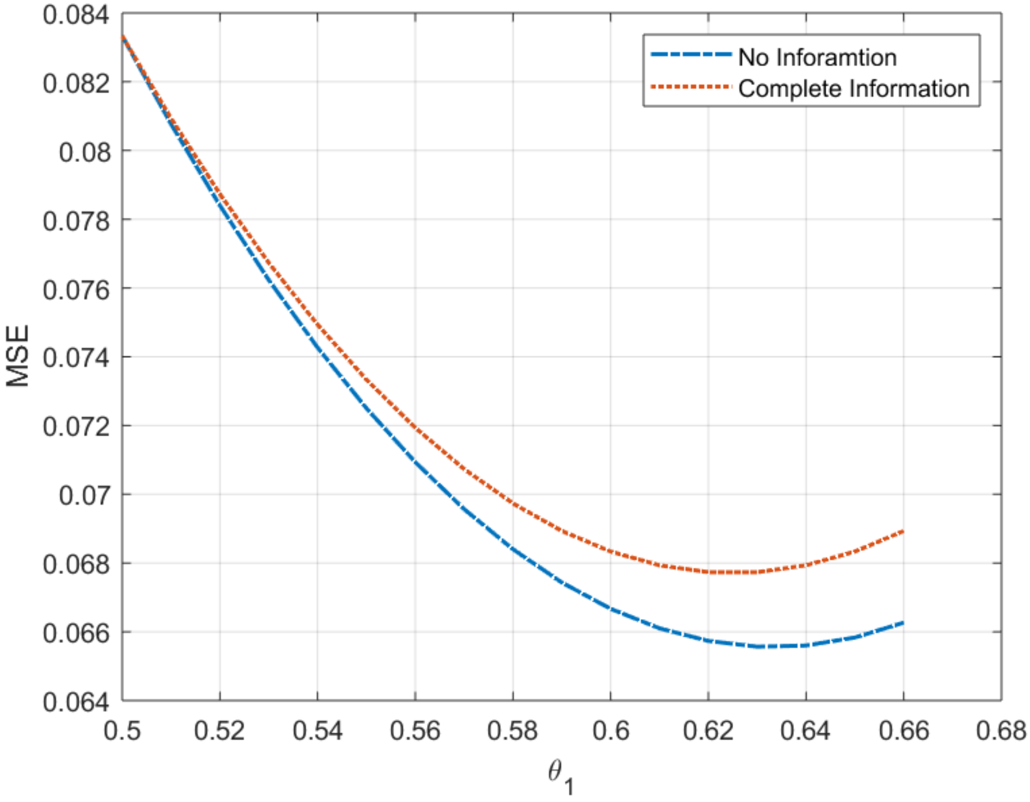

Figure 4.5: MSE comparison: complete and no informative case for single sensor observation.

The MSE with respect to optimal estimate θ is as shown in Fig 4.5 at particular sensor unreliability ($\varepsilon = 0.1$). At the worst case, i.e. $\theta_1 = 0.5$, MSE=$\frac{1}{12}$ is same for both complete and non informative case. The complete informative situation is always better than non informative for all optimal estimate other than the case when $\theta_1 = 0.5$. The difference between MSE for this two cases is increases as estimate θ increases from $\frac{1}{2}$ to $\frac{2}{3}$.

4.3 Bayes Estimation

In this section, we investigate the estimation problem in a system with unreliable nodes under a game theoretical setting. Specifically, a network of nodes is considered in a zero-sum game framework for Bayesian estimation. The saddle-point solution is obtained for a single-node network. It is extended for the multi-node network and obtains strong robustness with a minor performance degradation under the nominal conditions.

Let us first consider the single node case under the Bayesian settings to minimize the MSE $E\left|g\left(\theta\right)-\hat{g}\left(\theta\right)\right|_2^2$ where $N=1$, the parameter of interest $g\left(\theta\right)\in\left(\theta_l,\theta_u\right)$,$-\infty\leq\theta_l<\theta_u\leq\infty$ is real, and $p\left(\theta\right)$ is assumed to be known with no point mass, $q\left(x|\theta\right)\in Q_\theta$ is an unknown pdf belongs to a set of possible distribution Q_θ. Obviously $N=1$ is too simple network, it can provide some insights to the research problem and serve as a baseline performance benchmark for the multi-node case by treating the entire set of nodes as one "super" node.

This problem can be formulated as a two-player zero-sum game: the $q\in Q$ choice maker as Player 1, the estimator $\hat{\theta}\in F$ as Player 2, and the estimation error C as the reward to Player 1 and the negative of the error as the reward to Player 2. It is assumed that p, the pdf under the nominal condition, is known. Usually, the estimator can arbitrarily choose its estimate f. However, the choices of Q_i depend on the knowledge of player 1 about θ. We seek a mini-max estimator that minimizes the cost under the worst case. Our main result shows that the clipped Bayesian estimator can obtain a unique mini-max estimator for any strategy. Furthermore, the least favorable Player 1 strategy is unique.

4.3.1 Min-Max Solution

Our goal is to design an estimator $\hat{g}_m(\theta)$ under the MSE criterion such that

$$\begin{aligned}\hat{g}_m(\theta) &= \inf_{\hat{g}} \sup_{q \in Q_\theta} E\left(d\left(g(\theta), \hat{g}(\theta)\right)\right) \\ &= \inf_{\hat{g}} \sup_{q \in Q_\theta} E\left(g(\theta) - \hat{g}(\theta)\right)_2^2 .\end{aligned} \tag{4.10}$$

For ease of presentation, in the following, we drop (θ) and use g and $\hat{g}$ respectively. Notice that

$$\begin{aligned}E\left(d\left(g, \hat{g}\right)\right) &= E_\theta E_X \left(g - \hat{g}(x)\right)^2 \\ &= (1-\epsilon) E_\theta E_p \left(g - \hat{g}(x)\right)^2 + \epsilon E_\theta E_q \left(g - \hat{g}(x)\right)^2 .\end{aligned}$$

We first notice that the estimate $\hat{g}(x)$ must be bounded such that $-\infty < \eta_l \leq \hat{g}(x) \leq \eta_u < \infty$. Otherwise, one can make the MSE arbitrarily large by letting $q(x|\theta)$ be a degraded point mass at the point x where $\hat{g}(x) = \infty$.

Due to the convexity of the cost function $\left(g - \hat{g}(x)\right)^2$, the second part of the cost $E_\theta E_q \left(g - \hat{g}(x)\right)^2$ is maximized by choosing q that results in extreme values of $\hat{g}(x)$. In this particular MSE case, the worst possible q is such that $\hat{\theta}(x)$ is either at extreme values such that $\hat{g}(x) = \eta_l$ with probability 1 when $g(\theta) \geq \frac{\eta_l + \eta_u}{2}$, or $\hat{g}(x) = \eta_u$ when $g(\theta) < \frac{\eta_l + \eta_u}{2}$. The resulting error due to the node unreliability is

$$\begin{aligned}d(\eta_l, \eta_u) = E_\theta &\left[I\left(g(\theta) < \frac{\eta_l + \eta_u}{2}\right) \left(g(\theta) - \eta_u\right)^2 \right] \\ +E_\theta &\left[I\left(g(\theta) \geq \frac{\eta_l + \eta_u}{2}\right) \left(g(\theta) - \eta_l\right)^2 \right],\end{aligned} \tag{4.11}$$

where $I\left(x\right)$ is the identity function. Notice that

$$E_{\theta}E_{p}\left(g\left(\theta\right)-\hat{g}\left(x\right)\right)^{2}=E_{\theta}\left(g^{2}\right)-E_{p_{0}}\left(E_{\theta|p_{0}}\left(g|x\right)\right)^{2}+E_{p_{0}}\left(\hat{g}\left(x\right)-E_{\theta|p_{0}}\left(g|x\right)\right)^{2}, \tag{4.12}$$

where $E_{\theta}\left(g\left(\theta\right)\right)-E_{p_{0}}\left(E_{\theta|p_{0}}\left(g|x\right)\right)^{2}$, independent of $\hat{g}$, is the minimum MSE (MMSE) under the nominal setting, where $\varepsilon=0$.

To minimize equation (4.12), the optimal choice of $\hat{g}_{m}\left(x_{1};\eta_{l},\eta_{u}\right)$ given the range $(\eta_{l}$,$\eta_{u})$ is a simple "clipped" estimator which caps the well-known MMSE from both above and below such that

$$\hat{g}_{m}\left(x;\eta_{l},\eta_{u}\right)=\begin{cases}\eta_{l} & E_{\theta|p}\left(g\left(\theta\right)|x\right)<\eta_{l}\\ E_{\theta|p}\left(g\left(\theta\right)|x\right) & \eta_{l}\leq E_{\theta|p}\left(g\left(\theta\right)|x\right)\leq\eta_{u}\\ \eta_{u} & E_{\theta|p}\left(g\left(\theta\right)|x\right)>\eta_{u}\end{cases}, \tag{4.13}$$

with the resulting "extra" MSE $E_{p}\left(\hat{g}\left(x\right)-E_{\theta|p}\left(g|x\right)\right)^{2}$ given by

$$E_{p}\left(\hat{g}_{m}\left(x;\eta_{l},\eta_{u}\right)-E_{\theta|p}\left(g|x\right)\right)^{2}=E_{p}\left\{\left(\eta_{u}-E_{\theta|p}\left(g|x\right)\right)^{2}I\left(E_{\theta|p}\left(g|x\right)\geq\eta_{u}\right)\right\}+\\ E_{p}\left\{\left(\eta_{l}-E_{\theta|p}\left(g|x\right)\right)^{2}I\left(E_{\theta|p}\left(g|x\right)\leq\eta_{l}\right)\right\}.$$

When the estimator $\hat{g}\left(\theta\right)$ is bounded in $\left(\eta_l, \eta_u\right)$, the mini-max MSE is

$$\begin{aligned}&E\left(g-\hat{g}_m\left(x;\eta_l,\eta_u\right)\right)^2 = \\ &\left(1-\varepsilon\right)E_p\left\{\left(\eta_u - E_{\theta|p}\left(g|x\right)\right)^2 I\left(E_{\theta|p}\left(g|x\right)\geq\eta_u\right)\right\} \\ &+\left(1-\varepsilon\right)E_p\left\{\left(\eta_l - E_{\theta|p}\left(g|x\right)\right)^2 I\left(E_{\theta|p}\left(g|x\right)\geq\eta_l\right)\right\} \\ &+\varepsilon d\left(\eta_l,\eta_u\right),\end{aligned} \tag{4.14}$$

achieved by $\hat{g}_m$ in equation (4.13). One can further optimize parameters η_u, η_l, and the optimal $\hat{\theta}_m$ can be obtained by minimizing equation (4.14).

When the system is completely normal, such clipping of the estimation by η_u and η_l certainly degrades the estimation performance by an extra

$$\begin{aligned}MSE = &E_p\left\{\left(\eta_u - E_{\theta|p}\left(g|x\right)\right)^2 I\left(E_{\theta|p}\left(g|x\right)\geq\eta_u\right)\right\} \\ &+E_p\left\{\left(\eta_l - E_{\theta|p}\left(g|x\right)\right)^2 I\left(E_{\theta|p}\left(g|x\right)\leq\eta_l\right)\right\}.\end{aligned}$$

However, it is often a relatively small price to pay to obtain strong robustness when the system is indeed malfunctioning.

4.3.2 Saddle-Point Solution

Next, we show that for any ε, there exists a suitable $\left(\eta_u, \eta_l\right)$ such that the estimator equation (4.13) is actually a saddle-point solution or Nash Equilibrium. It can be proven by finding a distribution $q^*\left(x|\theta\right)$ such that the resulting MMSE estimator $\hat{g}\left(\theta\right)_{MMSE} = E\left(g\left(\theta\right)|x\right)$ admits the form in equation (4.13). We show that this max-min solution is the same as the min-max solution. Hence, both Player's strategies form a saddle-point solution or Nash Equilibrium.

Let us define

$$\begin{aligned}\mathcal{X}_l &= \left\{x|E_{\theta|p}\left(g|x\right) < \eta_l\right\}\\ \mathcal{X}_u &= \left\{x|E_{\theta|p}\left(g|x\right) > \eta_u\right\}\\ \Theta_u &= \left\{\theta|g \geq \frac{\eta_l+\eta_u}{2}\right\}\\ \Theta_l &= \left\{\theta|g < \frac{\eta_l+\eta_u}{2}\right\}.\end{aligned} \tag{4.15}$$

From the optimal attacking strategy, we know that the optimal q puts all the probability mass in $\mathcal{X}_l$ (respectively $\mathcal{X}_u$) when $\theta \in \Theta_u$ (respectively $\theta \in \Theta_l$), or equivalently,

$$\begin{aligned}&\int_{\mathcal{X}_l} q\left(x|\theta\right)dx = 1, \theta \in \Theta_u\\ &\int_{\mathcal{X}_u} q\left(x|\theta\right)dx = 1, \theta \in \Theta_l.\end{aligned} \tag{4.16}$$

When $x_1 \in \mathcal{X}_l$, under the mixture model (4.2)

$$\begin{aligned}\hat{g}_{MMSE}\left(x\right) &= E\left(g\left(\theta\right)|x\right)\\ &= \int_{\theta} g\left(\theta\right)p\left(\theta|x\right)d\theta\\ &= \int_{\Theta} g\left(\theta\right)\frac{p\left(x|\theta\right)}{p\left(x\right)}p\left(\theta\right)d\theta\\ &= \eta_l.\end{aligned}$$

Hence,

$$\begin{aligned}\int_{\Theta} g\left(\theta\right)p\left(x|\theta\right)p\left(\theta\right)d\theta &= \left(1-\varepsilon\right)\eta_l\int_{\Theta} p_0\left(x|\theta\right)p\left(\theta\right)d\theta + \varepsilon\eta_l\int_{\Theta} q\left(x|\theta\right)p\left(\theta\right)d\theta\\ &= \left(1-\varepsilon\right)\eta_l p_0\left(x\right) + \varepsilon\eta_l q\left(x\right).\end{aligned}$$

Therefore we have

$$q\left(x\right)\left[\frac{\int_{\Theta} gq\left(x|\theta\right)d\theta}{q\left(x\right)}-\eta_l\right]=\frac{1-\varepsilon}{\varepsilon}p\left(x\right)\left[\eta_l-\frac{\int_{\Theta}\theta p\left(x|\theta\right)d\theta}{p\left(x\right)}\right].$$

That is

$$q\left(x\right)=\frac{1-\varepsilon}{\varepsilon}p\left(x\right)\frac{\eta_l-E_p\left(g|x\right)}{E_q\left(g|x\right)-\eta_l}, x\in\mathcal{X}_l, \tag{4.17}$$

where $E_p\left(g|x\right)=\frac{\int_{\Theta} gp(x|\theta)d\theta}{p(x)}$ is the conditional mean under the nominal distribution $E_q\left(g|x\right)=\frac{\int_{\Theta} gq^*(x|\theta)d\theta}{q^*(x)}$ is the conditional mean under the contaminating distribution q. Recall the condition equation (4.16), for $q\left(x|\theta\right)$ to be a valid pdf, we have

$$\begin{aligned}\int_{\mathcal{X}_l} q\left(x\right)dx&=\int_{\mathcal{X}_l}\int_{\Theta_u} q\left(x|\theta\right)p\left(\theta\right)d\theta dx\\&=\int_{\Theta_u} p\left(\theta\right)d\theta=Pr\left(g\left(\theta\right)>\frac{\eta_l+\eta_u}{2}\right).\end{aligned}\tag{4.18}$$

A similar condition can be obtained for $x\in\mathcal{X}_u$.

We can now summarize the overall conditions for $q\left(x|\theta\right)$ or equivalently $q\left(x\right)$ for the saddle point solution as follows:

$$\begin{cases} q\left(x\right)=\begin{cases}\frac{1-\varepsilon}{\varepsilon}p\left(x\right)\frac{\eta_l-E_p(g|x)}{E_q(g|x)-\eta_l} & x\in\mathcal{X}_l\\ 0 & x\in\overline{\mathcal{X}_l\cup\mathcal{X}_u}\\ \frac{1-\varepsilon}{\varepsilon}p\left(x\right)\frac{\eta_u-E_p(g|x)}{E_q(g|x)-\eta_u} & x\in\mathcal{X}_u\end{cases}\\ \frac{1-\varepsilon}{\varepsilon}\int_{\mathcal{X}_l} p\left(x\right)\frac{\eta_l-E_p(g|x)}{E_q(g|x)-\eta_l}dx=Pr\left(g\left(\theta\right)>\frac{\eta_l+\eta_u}{2}\right)\\ \frac{1-\varepsilon}{\varepsilon}\int_{\mathcal{X}_u} p\left(x\right)\frac{\eta_u-E_p(g|x)}{E_{q_1}(g|x)-\eta_u}dx=Pr\left(g\left(\theta\right)<\frac{\eta_l+\eta_u}{2}\right).\end{cases}\tag{4.19}$$

This set of equations also defines the conditions on the triplet $(\varepsilon, \eta_l, \eta_u)$ for the saddle-point solutions. The values of η_l and η_u cannot be chosen arbitrarily, and often they are unique. In terms of $q\left(x|\theta\right)$, there are many possible solutions of so long as equation (4.19) is satisfied. Since this is the MMSE of the $q^*\left(x|\theta\right)$ which also admits the minimax condition (4.13), this solution is a saddle point solution, or Nash Equilibrium. Next, we present a simple solution of such $q_s^*\left(\cdot\right)$ by dropping the condition of $q_s^*\left(x|\theta\right)$ on θ such that $q_s^*\left(x|\theta\right) = q_s^*\left(x\right)$ is semi-independent of θ. With this simplification, we have

$$E_{q_s^*}\left(g|x\right) = E\left(g|\theta \in \Theta_u\right) = \eta_{su}, x \in \mathcal{X}_l, \tag{4.20}$$

and

$$E_{q_s^*}\left(g|x\right) = E\left(g|\theta \in \Theta_l\right) = \eta_{sl}, x \in \mathcal{X}_u. \tag{4.21}$$

Recall the condition 4.19, and taking the integration of x over $\mathcal{X}_l$ and $\mathcal{X}_u$, we have

$$\begin{aligned}
\varepsilon &= \frac{\int_{\mathcal{X}_u} p\left(x\right) \frac{\eta_u - E_p(g|x)}{\eta_{sl}-\eta_u} dx}{Pr\left(g\left(\theta\right) < \frac{\eta_l+\eta_u}{2}\right) + \int_{\mathcal{X}_u} p\left(x\right) \frac{\eta_u - E_p(g|x)}{\eta_{sl}-\eta_u} dx} \\
&= \frac{\int_{\mathcal{X}_u} p\left(x\right) \frac{\eta_u - E_p(g|x)}{\eta_{sl}-\eta_u} dx}{Pr\left(g\left(\theta\right) > \frac{\eta_l+\eta_u}{2}\right) + \int_{\mathcal{X}_l} p\left(x\right) \frac{\eta_l - E_p(g|x)}{\eta_{su}-\eta_l} dx} \\
&= \frac{\int_{\mathcal{X}_u} p\left(x\right) \frac{\eta_u - E_p(g|x)}{\eta_{sl}-\eta_u} dx + \int_{\mathcal{X}_l} p\left(x\right) \frac{\eta_l - E_p(g|x)}{\eta_{su}-\eta_l} dx}{1 + \int_{\mathcal{X}_u} p\left(x\right) \frac{\eta_u - E_p(g|x)}{\eta_{sl}-\eta_u} dx + \int_{\mathcal{X}_l} p\left(x\right) \frac{\eta_l - E_p(g|x)}{\eta_{su}-\eta_l} dx}.
\end{aligned} \tag{4.22}$$

4.3.3 Saddle Point Solution Under the No Information Scenario

The previous derivation is based on the assumption that the malfunctioning node has the complete knowledge of θ, and thus can choose any arbitrary distribution $q_1\left(x_1|\theta\right)$.

Next, we consider the case where the node only knows about its own observation X_1, and all he can do is to modify it to another value Z_1 based on the conditional distribution $p\left(Z_1|x_1\right)$. This puts a considerable constraints on the possible $q_1\left(x_1|\theta\right)$. Nevertheless, if we allow $p\left(z_1|x_1\right)$ to be arbitrarily, we have can have a similar result. To see that, we rewrite the MSE in terms of X_1, Z_1 and utilizing the fact that $\theta \rightarrow X_1 \rightarrow Z_1 \rightarrow \hat{g}\left(\theta\right) = \hat{g}\left(Z_1\right)$ forms a Markov chain (under the nominal condition, $Z_1 = X_1$) such that

$$\begin{aligned}
E\left|g\left(\theta\right)-\hat{g}\left(\theta\right)\right|_2^2 &= E\left(g^2\left(\theta\right)-2g\left(\theta\right)\hat{g}\left(Z_1\right)+\hat{g}^2\left(\theta\right)\right)\\
&= E\left(g^2\left(\theta\right)\right)+E\left(-2g\left(\theta\right)\hat{g}\left(\theta\right)+\hat{g}^2\left(\theta\right)\right)\\
&= E\left(g^2\left(\theta\right)\right)+E_{\theta X_1 Z_1}\left(-2g\left(\theta\right)\hat{g}\left(\theta\right)+\hat{g}^2\left(\theta\right)\right)\\
&= E\left(g^2\left(\theta\right)\right)+E_{X_1 Z_1}E_{\theta|X_1}\left(-2g\left(\theta\right)\hat{g}\left(\theta\right)+\hat{g}^2\left(\theta\right)\right)\\
&= E\left(g^2\left(\theta\right)\right)+E_{X_1 Z_1}\left(-2E_{\theta|X_1}g\left(\theta\right)\hat{g}\left(\theta\right)+\hat{g}^2\left(\theta\right)\right)\\
&= E\left(g^2\left(\theta\right)\right)+E_{X_1}E_{Z_1|X_1}\left(-2\tilde{g}\left(X_1\right)\hat{g}\left(Z_1\right)+\hat{g}^2\left(Z_1\right)\right)\\
&= E\left(g^2\left(\theta\right)\right)-E_{X_1}\left(\tilde{g}^2\left(X_1\right)\right)+E_{X_1}\left(\tilde{g}\left(X_1\right)-\hat{g}\left(Z_1\right)\right)^2,
\end{aligned}$$

where $\tilde{g}\left(X_1\right) = E_{\theta|X_1}g\left(\theta\right)$ is the conditional mean of $g\left(\theta\right)$ given X_1. Notice that $E\left(g^2\left(\theta\right)\right)-E_{X_1}\left(\tilde{g}^2\left(X_1\right)\right)$ is a constant independent of the choice of estimate $\hat{g}\left(Z_1\right)$, the optimization problem is equivalent to estimate $\tilde{g}\left(X_1\right)$ based on the observation Z_1. Therefore, the problem of partial information can be treated in a similar fashion by incorporating the partial information into $\tilde{g}\left(X_1\right)$.

To sum up, the estimation problem for a single node network with “no information” becomes a particular case of the “complete information,” where the parameter X_1 is observed directly to estimate $\tilde{g}\left(X_1\right)$, a function of X_1, subject to a potential node malfunction that changes X_1 to a different value Z_1.

4.3.4 Probabilistic Setting: Location Estimation in Gaussian Observation Noise

This section presents the scalar parameter estimation problem in Gaussian noise when an adversary may control the observation. The prior distribution $p\left(\theta\right)$ is also assumed to be Gaussian distributed. For this specific observation model, we derive the breakdown point and efficiency of our proposed robust estimator. The efficiency measures the performance loss due to the clipping compared to the nominal case.

Location estimation in a single node

Suppose $\theta \in \mathcal{N}\left(0,1\right)$ is a normalized standard Gaussian random variable. Under the nominal condition with $1-\varepsilon$ probability, node 1 observes a noisy data with variance σ^2, i.e. $X_1 = \theta + W_1$ where $W_1 \sim \mathcal{N}\left(0,\sigma^2\right)$; and with ε probability, X_1 is replaced by another value generated from an unknown distribution $q\left(x|\theta\right)$. The goal is to estimate θ that achieves the minimal possible MSE.

Under the nominal condition, $X_1 \sim \mathcal{N}\left(0,1+\sigma^2\right)$, the a posterior distribution $\theta|X_1 \sim \mathcal{N}\left(\frac{1}{1+\sigma^2}X_1, \frac{\sigma^2}{1+\sigma^2}\right)$. The MMSE estimator $\hat{\theta}_{MMSE} = E_{p_1}\left(\theta|X_1\right) = \frac{1}{1+\sigma^2}X_1$ under the nominal condition is well known with the resulting MSE $\frac{\sigma^2}{1+\sigma^2}$. Under the node uncertainty, a robust estimator clips $\hat{\theta}_{MMSE}$ from both above and below at $\left(\eta_l, \eta_u\right)$. Due to the symmetry of this problem, we let $-\eta_l = \eta_u = \eta \geq 0$ and try to determine the optimal threshold η. That is,

$$\hat{\theta}(\eta) = \begin{cases} -\eta & \frac{1}{1+\sigma^2}X_1 < -\eta \\ \frac{1}{1+\sigma^2}X_1 & -\eta \le \frac{1}{1+\sigma^2}X_1 \le \eta \\ \eta & \frac{1}{1+\sigma^2}X_1 > \eta. \end{cases} \tag{4.23}$$

In this case, we have

$$\begin{aligned} \mathcal{X}_l &= \left(-\infty, -\left(1+\sigma^2\right)\eta\right) \\ \mathcal{X}_u &= \left(-\infty, \left(1+\sigma^2\right)\eta\right) \\ \times_l &= (-\infty, 0)\,\Theta_u = (0, \infty)\,, \end{aligned}$$

and equation (4.21) become

$$\begin{aligned} \eta_{sl} &= E\left(\theta|\theta \in \Theta_l\right) = -\sqrt{\frac{2}{\pi}} \\ &= -\eta_{su}\int_{(1+\sigma^2)\eta}^{\infty} p_1\left(x_1\right)\frac{\eta - \frac{1}{1+\sigma^2}x_1}{-\sqrt{\frac{2}{\pi}} - \eta}dx_1 \\ &= \frac{1}{\sqrt{1+\sigma^2}\left(\sqrt{\frac{2}{\pi}}+\eta\right)}\phi\left(\eta\sqrt{1+\sigma^2}\right) - \eta\sqrt{1+\sigma^2}Q\left(\eta\sqrt{1+\sigma^2}\right) \end{aligned}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2\sigma^2}$ is the pdf and $Q(x)$ are the complementary distribution function of a standard Gaussian random variable, respectively. With $Pr\left(g(\theta) > \frac{\eta_l+\eta_u}{2}\right) = Pr(\theta > 0) = \frac{1}{2} = Pr\left(g(\theta) < \frac{\eta_l+\eta_u}{2}\right)$, the relationship between ε and the optimal η is given by equation (4.22)

$$\begin{aligned}\varepsilon &= \frac{\int_{(1+\sigma^2)\eta}^{\infty} p_1\left(x_1\right) \frac{\eta-\frac{1}{1+\sigma^2}x_1}{-\sqrt{\frac{2}{\pi}}-\eta} dx_1}{\frac{1}{2}+\int_{(1+\sigma^2)\eta}^{\infty} p_1\left(x_1\right) \frac{\eta-\frac{1}{1+\sigma^2}x_1}{-\sqrt{\frac{2}{\pi}}-\eta} dx_1} \\ &= \frac{2\phi\left(\eta\sqrt{1+\sigma^2}\right)-\eta\sqrt{1+\sigma^2}Q\left(\eta\sqrt{1+\sigma^2}\right)}{\sqrt{1+\sigma^2}\left(\sqrt{\frac{2}{\pi}}+\eta\right)+2\phi\left(\eta\sqrt{1+\sigma^2}\right)-\eta\sqrt{1+\sigma^2}Q\left(\eta\sqrt{1+\sigma^2}\right)},\end{aligned} \tag{4.24}$$

which can be used to determine the optimal η for a given ε.

When the optimal threshold is $\eta = 0$, the robust estimate is always 0, the mean based on the prior distribution $p\left(\theta\right)$, and the observation is completely disregarded due to the node uncertainty. The corresponding breaking down point ε_0^* becomes

$$\varepsilon_0^* = \frac{1}{1+\sqrt{1+\sigma^2}}, \tag{4.25}$$

that is, when the node uncertainty $\varepsilon \geq \varepsilon_0^* = \frac{1}{1+\sqrt{1+\sigma^2}}$, the node observation is too ambiguous to be used to form a meaningful estimation and becomes useless. On the other sides, $\eta = \infty$ when $\varepsilon = 0$, the clipped estimator reduces to the nominal setting. The MSE under the nominal condition for $\hat{\theta}\left(\eta\right)$ in equation (4.24) is

$$MSE_N\left(\eta\right) = \frac{\sigma^2}{1+\sigma^2}+\frac{2}{1+\sigma^2}\left[\left(\eta^2\left(1+\sigma^2\right)+1\right)\left(\eta\sqrt{1+\sigma^2}\right)-\eta\sqrt{1+\sigma^2}\phi\left(\eta\sqrt{1+\sigma^2}\right)\right].$$

Clearly, $\frac{2}{1+\sigma^2}\left[\left(\eta^2\left(1+\sigma^2\right)+1\right)\left(\eta\sqrt{1+\sigma^2}\right)-\eta\sqrt{1+\sigma^2}\phi\left(\eta\sqrt{1+\sigma^2}\right)\right]$, monotonic decreasing with η, is the performance degradation due to the clipping. The efficiency of the clipped estimator can be measured by the relative performance between the clipped estimator and the MMSE estimator such that

$$\begin{aligned} e_N(\eta) =& \frac{1 - MSE_N(\eta)}{1 - MSE_N(\infty)} \\ =& 1 - 2\left[\left(\eta^2\left(1+\sigma^2\right)+1\right)\left(\eta\sqrt{1+\sigma^2}\right) - \eta\sqrt{1+\sigma^2}\phi\left(\eta\sqrt{1+\sigma^2}\right)\right]. \end{aligned}$$

The worst possible MSE when the node is malfunctioning is

$$\begin{aligned} MSE_M(\eta) &= p(\theta < 0) E\left[(\eta-\theta)^2 \mid \theta < 0\right] + p(\theta \leq 0) E\left[(\eta+\theta)^2 \mid \theta > 0\right] \\ &= \eta^2 + 1 + \frac{4\eta}{\sqrt{2\pi}}, \end{aligned}$$

is a monotonic increasing function of η, and the overall estimation performance

$$MSE_P(\eta) = (1-\varepsilon) MSE_N(\eta) + \varepsilon MSE_M(\eta),$$

achieves the minimum value at the optimal η from equation (4.24).

Location estimation in multi nodes

Extending the results of the single-node to multiple-node networks is not straightforward. To make some progress on this front, the performance of the clipped estimator is investigated for the location estimation problem in multiple-node networks.

The noisy observations $X \sim [X_1, X_2, ..., X_N]$ collected by N nodes. The MMSE estimator under nominal condition when noise in observations is i.i.d. Gaussian noise $W \sim \mathcal{N}(0, \sigma^2)$ is

$$\hat{\theta} = E(\theta/\mathbf{x}) = \frac{1}{1+\sigma^2/N} \frac{\sum_{i=1}^{N} X_i}{N}. \tag{4.26}$$

MMSE estimmator under nominal condition gives all observations the same weight. However, our intuition suggests that we weigh the observation X_i, $i = 1, ..., N$, in equation (4.26) such that we give more weight to data that is close to the measurement model as compared to the one that is unlikely to occur for unreliable network. For the case of $N > 1$, an intuitive estimator is

$$\hat{\theta} = \frac{1}{1 + \sigma^2/N} \frac{\sum_{i=1}^{N} Y_i}{N}, \tag{4.27}$$

where $Y_i = f\left(X_i, \eta\right)$, is the clipped observation of X_i.

4.3.5 Static Setting: Location Estimation in Gaussian Observation Noise

There are K outliers each from unknown pdf $q(X_i/\theta) \in Q(\theta)$ and nominal $N - K$ observations with each known pdf $p(X_i/\theta) \in P(X_i/\theta)$ respectively. The overall estimation problem is

$$g(X/\theta) = \prod_{j=j_1}^{j_K} q(X_j/\theta) \prod_{j=j_1}^{j_{N-K}} p(X_j/\theta), \tag{4.28}$$

where $\Omega_K = \{j_1, ..., j_K \mid 1 \leq j_1 < j_2 ..., < j_K \leq N\}$ is the set of all possible subset with size K.

For the quadratic loss, $d(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$ estimation error is

$$C = \int_{-\infty}^{\infty} d\theta \int_{-\infty}^{\infty} d(\theta, \hat{\theta}) p_{\theta,X}(\theta, X) dX. \tag{4.29}$$

Consider the clipped estimator defined in section-4.3.4 as $\hat{\theta} = E\left(\theta/\mathbf{x}\right) = \frac{1}{1+\sigma^2/N}\frac{\sum_{i=1}^{N} Y_i}{N}$. Due to the symmetry

$$Y_i = \begin{cases} X_i & -\tau < X_i < \tau \\ \tau & X_i > \tau \\ -\tau & -\tau < X_i \end{cases}.$$

The optimal adversary strategy $q(X_i/\theta)$ is the one that makes $f(X)$ minimum as much as possible for $\theta > 0$ and as large as possible for $\theta < 0$ to make C statistically large. When $Q(\theta) = \Omega$, the optimal outlier $q^0(X_i/\theta)$ is to put all the probability mass at the set $\left\{X_i^I \mid X_i \leq -\tau\right\}$ for $\theta > 0$ and to put all the probability mass at the set $\left\{X_i^U \mid l_i(X_i) \geq \tau\right\}$ for $\theta < 0$. As a result, MSE is

$$MSE(\theta, \hat{\theta}) = E\left(\left(E(\hat{\theta} \mid \theta) - \theta\right)^2 + var(\hat{\theta} \mid \theta)\right),$$

where

$$E(\hat{\theta} \mid \theta) = \begin{cases} \left(\frac{(N-K)}{N}\right) E(Y_i \mid \theta) - \frac{K\tau}{N} & \theta > 0 \\ \left(\frac{(N-K)}{N}\right) E(Y_i \mid \theta) + \frac{K\tau}{N} & \theta < 0, \end{cases}$$

and

$$E(Y_i \mid \theta) = \tau Q\left(\frac{\tau - \theta}{\sigma_w}\right) - \tau Q\left(\frac{\tau + \theta}{\sigma_w}\right) + \sigma_w\left[\phi\left(\frac{-\tau - \theta}{\sigma_w}\right) - \phi\left(\frac{\tau - \theta}{\sigma_w}\right)\right] + \theta Z.$$

As a result,

$$\begin{aligned} var(\hat{\theta} \mid \theta) &= \tau^2 Q\left(\frac{\tau-\theta}{\sigma_w}\right) + \tau^2 Q\left(\frac{\tau+\theta}{\sigma_w}\right) \\ &+ \sigma_w^2 \left[\left(\frac{-\tau-\theta}{\sigma_w}\right)\phi\left(\frac{-\tau-\theta}{\sigma_w}\right) - \left(\frac{\tau-\theta}{\sigma_w}\right)\phi\left(\frac{\tau-\theta}{\sigma_w}\right) + Z\right] \\ &= 2\theta\sigma_w\left[\phi\left(\frac{-\tau-\theta}{\sigma_w}\right) - \phi\left(\frac{\tau-\theta}{\sigma_w}\right)\right] + \theta^2 Z \end{aligned}$$

where $Z = \left[Q\left(\frac{-\tau-\theta}{\sigma_w}\right) - Q\left(\frac{\tau-\theta}{\sigma_w}\right)\right]$. With the prior $\phi(\theta)$, the MSE can be expressed as

$$MSE(\theta, \hat{\theta}) = \int\limits_{-\infty}^{\infty} \left(\left(E(\hat{\theta} \mid \theta) - \theta\right)^2 + var(\hat{\theta} \mid \theta)\right)\phi(\theta)d\theta.$$

For a given K and N, the value of threshold, τ that minimizes MSE can be obtained solving $\frac{\partial MSE(\theta,\hat{\theta})(\tau,K,N)}{\partial \tau} = 0$.

4.3.6 Simulation and Validation

In this section we present the simulation results for location estimation form single and multiple Gaussian observations. Given the estimator equation (4.23), the most damage for the malfunctioning node to do is to send an observation which results in $Y_i = -\eta$ when $\theta > 0$ and η when $\theta \leq 0$.

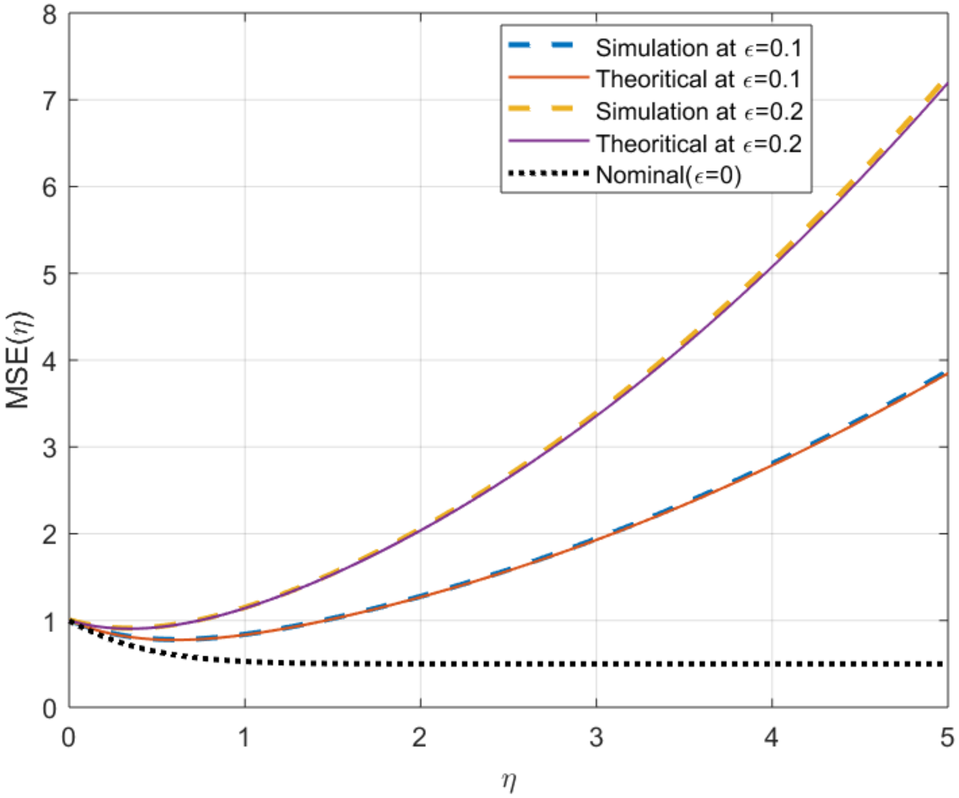

Figure 4.6: Performance of Robust estimator at $\varepsilon = 0.1$ and $\varepsilon = 0.2$ for single node.

Fig 4.6 shows MSE vs. η, for the with different unreliability ε for the single node case. As expected, a higher ε results in a worse estimation performance for the same η. However, the optimal η are pretty close to each other under these two ε's. That is, a suitable choice of η is able to provide a reasonable good performance for a wide range of vulnerable systems. Fig 4.7 shows the validation result based on the derivation in section (4.3.5) for static model at $N = 20$ and $K = 3$ for prior as standard normal distribution.

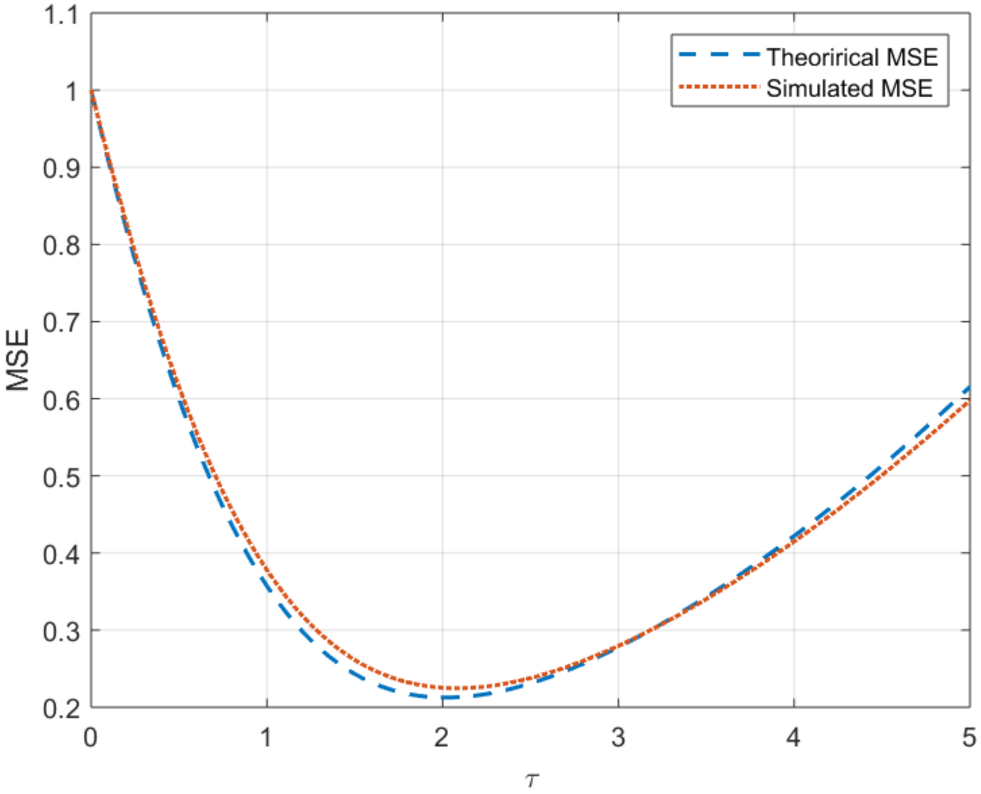

Figure 4.7: MSE of clipped estimator at $N = 20$, $K = 3$ for prior $p_\theta(\theta) \sim N(0, 1)$ and $W \sim N(0, 1)$.

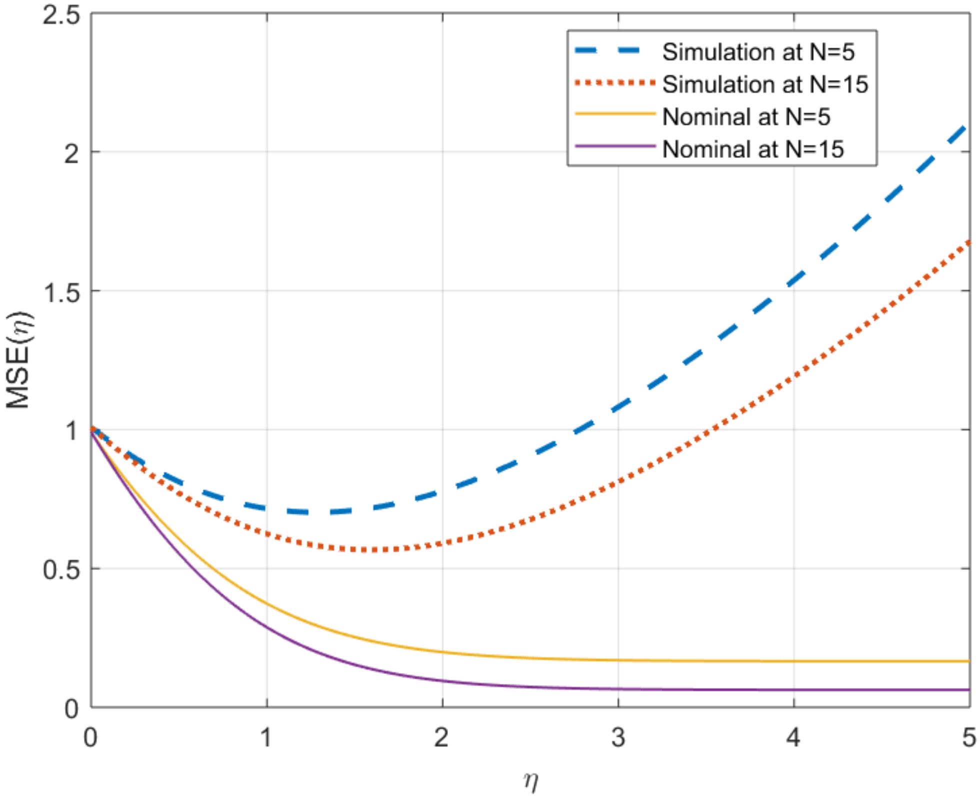

Figure 4.8: Estimation performance as a function of η at $\varepsilon = 0.2$ for multi nodes.

As the number of observations N increases for a fixed ε, the estimation performance improves as shown in Fig 4.8. Further, the optimal robust threshold η increases as N increases. This means that even as the percentage of malfunctioning nodes remains the same, more nodes enable the designer to relax the constraints on each sensor. As a result, the efficiency of the estimator at the optimal η also improves as N increases. The clipped estimator is about 81.08% and 91.35% under the nominal settings, for $N = 5$ and 15, respectively. This means that as the estimation performance is also lower-bounded by the node uncertainty ϵ, and there are still merits in having a larger N; one can obtain a robust design that performs not only closer to the performance bound but also performs better under the nominal conditions when all nodes are normal.

We consider the probabilistic case where each node has a ε chance of malfunctioning and the static case where K out of N nodes are malfunctioning. These two cases are relatively comparable for $\varepsilon = \frac{K}{N}$. The simulation result plotted in Fig 4.9 shows that the performance of static setting $\frac{K}{N}$, is better than its corresponding probabilistic counterpart ε. In a static setting, there is a fixed number of compromised nodes whereas, in a probabilistic setting a node is compromised with probability ε, hence total compromised node is a random number, which may end with all nodes compromised or none of the nodes in under compromised.

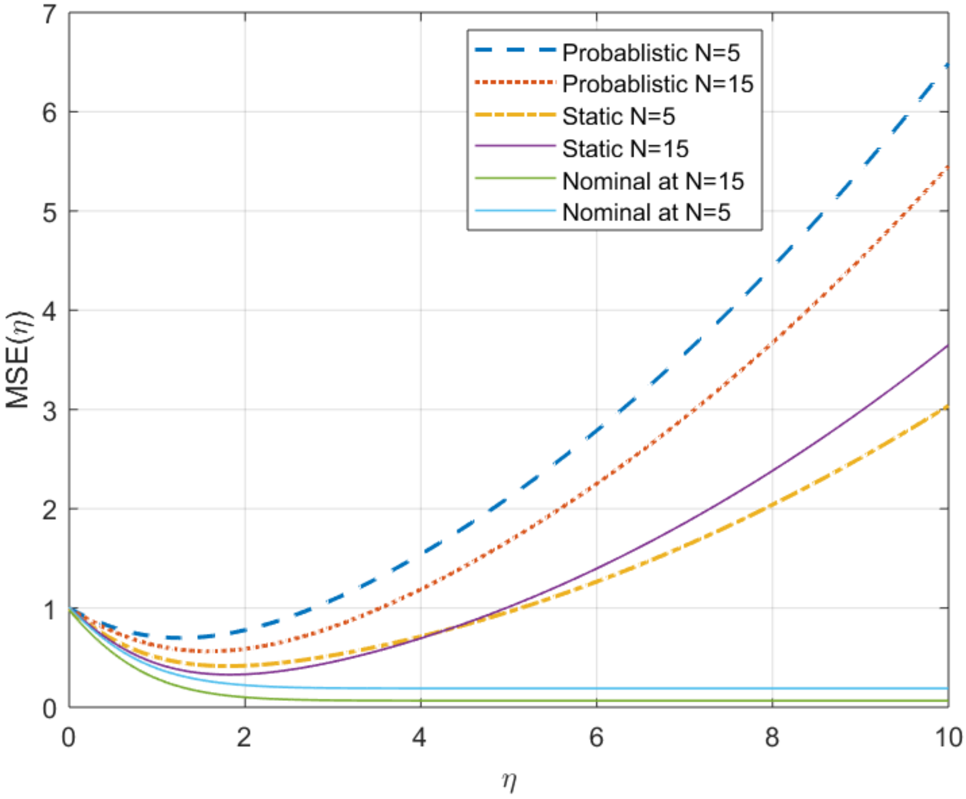

Figure 4.9: Trade off between η, and N for fixed $\varepsilon/(K)$.

Trade off between η, $\epsilon/(K)$ and N:

Fig 4.9 shows MSE vs. η, as N increases for the same ε. As the number of observation increases, for the fixed $\varepsilon/(K)$ the performance increases. The optimal robust thresholds increases, as N increases, which includes more number of observations.

Table 4.1: Efficiency of robust estimator.

N	K	ε	Efficiency(Static)	Efficiency(Probabilistic)
5	1	0.2	88.04%	81.08%
15	3	0.2	94.38%	91.35%

To investigate the robustness of the choice of threshold η when the system vulnerability ε or $\frac{K}{N}$ is unknown, we calculate the estimation performance with respect to η for $N = 10$, $\varepsilon = \frac{K}{N} = 0.1$ and 0.2 respectively, as shown in Fig 4.10. A higher ε or K results in a worse estimation performance for the same η. However, the optimal η are close to each other under these two ε. That is, a suitable choice of η can provide reasonably good performance for a wide range of vulnerable systems.

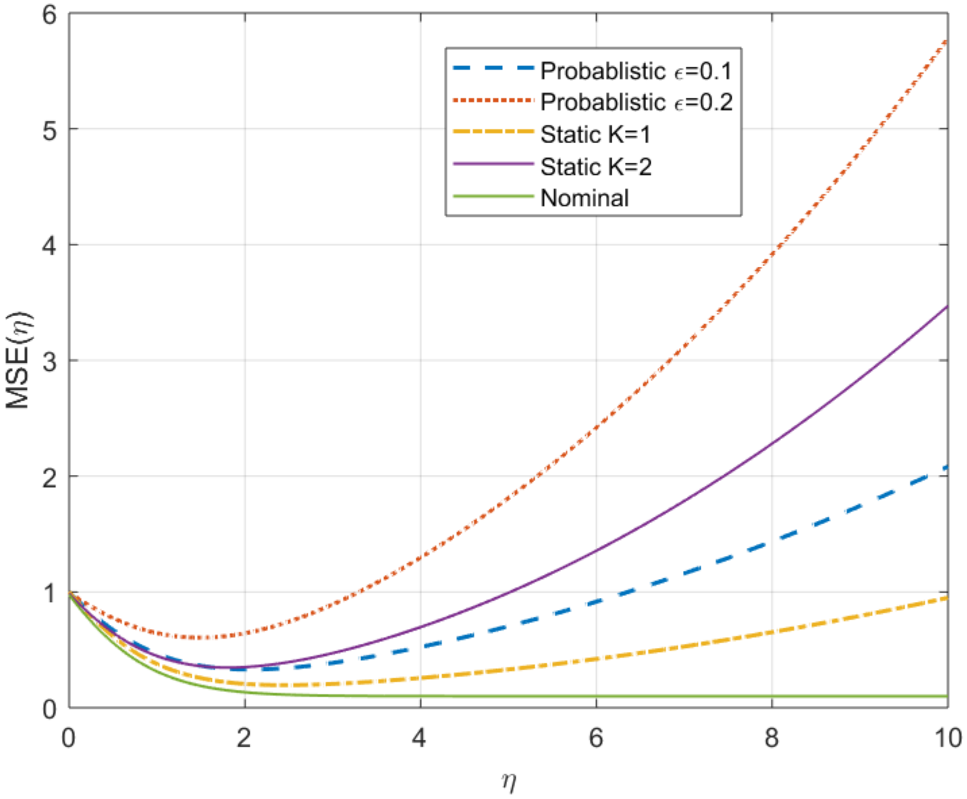

Figure 4.10: Estimation performance as a function of η, for $N = 10$, $\varepsilon = \frac{K}{N} = 0.1, 0.2$ respectively.

CHAPTER 5:
ROBUST LOCALIZATION

In this chapter, results in robust localization through linear regression are presented. The target localization in a distributed co-located radar system is considered by optimizing a robust cost function. In particular, a statistical tool is developed that may be stable to adversarial measurement and provide a more accurate target location. The algorithm does not compromise much accuracy, even without outliers. In particular, we consider the following settings.

1. The robust regression problem is formulated into its equivalent weighted least square regression where the weights are based on the considered robust cost.

2. To improve the localization accuracy of robust localization, a small set of secured sensors is presented, potentially by spending more resources on those sensors, especially in a potentially hostile environment. By doing so, we might be able to improve the signal quality and integrity for a small set of sensors and, consequently, reduce the target localization errors.

5.1 Introduction

Accurate target localization is an important task in various applications such as wireless communication, and surveillance. Broadly speaking, there are two main categories of localization techniques, those that involve range estimation, and those that do not [60]. Target localization in radar is often range based schemes [61, 62, 63] which involves estimation of target location based on distance between the targets and the transmitters, and receivers. Depending on the applications, range information can be obtained using Time of Arrival (TOA) [64], Time Delay of Arrival (TDOA) [65], or a combination of the two [66].

Like other networked systems, the functionality and performance of a radar system may be affected when the radar is not operated under ideal nominal conditions. For example, it has been observed [67] that a slight error in receiver locations can lead to a significant error in target location estimate. In many situations, mainly when a moving platform is involved, there may be random errors in transmitters/receivers positions that would introduce bias in target locations. Another potential deviation may happen if some of the transmitters and receivers are compromised by an adversary. In hostile environments, an adversary may wish to prevent accurate localization of the target and thus prevent the entire radar system from functioning properly. The adversary may compromise some nodes and thereby gain access to inject false TOA/TDOA to provide misleading information to the target that prevents accurate target localization.

As a way to get a robust estimation of location in the presence of outliers, Least Median Squares (LMS) has been proposed [68]. It uses several subsets of nodes to identify candidate locations and then chooses the solution that minimizes the median of the residues [69, 70]. It is robust for scenarios where less than 50% of the nodes are malicious. This method shares similarities with the random sampling consensus (RANSAC) algorithm [71].

5.1.1 Overview of this work

Robust localization problems for radar systems is considered where direct measurements of the distance between transmitters and target are available trough TOA measurements. The main idea behind this work is to minimize a robust cost function using iterative approaches. The cost function is dynamically updated to remove the outliers or reduce the effect of such inconsistent measurements arising from outliers. In this study, multiple distributed collocated radar system is consider with single stationary target P_0 at position β. Let N be the number of transmitters whose locations are known. The estimate of distance between the transmitters P_i at $\mathbf{x_i}$, and target P_0 is denoted as $\{D_i\}_{1\leq i\leq N}$. In radar system, set of data points $\{d_i\}_{1\leq i\leq N}$, may be obtained through a nonlinear sensing system,

$$d_i \approx D_i(\boldsymbol{\beta}), 1 \leq i \leq N, \tag{5.1}$$

where β is the target location and $D_i's$ are nonlinear maps the distances between target and transmitters. Given set of noisy measurements, d_i

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} h\left(d_i - D_i(\boldsymbol{\beta})\right), \tag{5.2}$$

where, $h(.)$ is cost function. Solving this nonlinear, non-convex function usually involves some iterative searching techniques, such as gradient decent or Newton's method. Therefore, to avoid local minimum as much as possible, it is necessary to rerun the algorithm using several initial starting points, and as a result the computation is relatively expensive. Another way to find target location is to convert the system of non linear equations [72] into linear forms and solving for β as a simple linear least square problem. However, both non-linear and linear approaches are not well suited in the presence of outliers to d_i values.

We consider the optimizing robust cost function by converting equation (5.2) into a linear form. We are interested in the statistical linear regression problem $b = A\beta+\epsilon$, where $b \in \mathcal{R}^{\mathrm{N}}$ is the observation vector, $A \in R^{N\times p}$ is the given design matrix, $\beta\in \mathcal{R}^{\mathrm{p}}$ is the regression coefficient vector and $\epsilon \in \mathcal{R}^{\mathrm{N}}$ is the noise. In the robust framework, we assume that a proportion of the observations may be atypical data corresponding to the outliers. These outliers might be due to excess noise, transmit sensors malfunction, and compromised nodes, controlled by Player 1.

There may be two types of outliers with different objectives. The first kind of adversarial outliers can compromise multiple nodes but have limited communication and computational resources to coordinate among the different radar nodes. It is referred to as a non-coordinated outlier. Player 1 is assumed to act independently at each transmitter node and prevent accurate localization by perturbing the estimated distance by TOA. Without loss of generality, it is assumed that each malfunctioning node modifies by adding a value c_i. Thus, the observation model is defined as

$$b_i = \begin{cases} a_i^T\beta + \epsilon_i + c_i & i \in outlier \\ a_i^T\beta + \epsilon_i & otherwise, \end{cases} \tag{5.3}$$

where ϵ_i is independent measurement noise and c_i perturbation is introduced due to Player 1. By nature, outliers are assumed to be distributed far away from the predicted model. We model this perturbation as a high value of constant bias.

A second type of outliers can not only prevent the network from precisely locate the target, but also try to shift the location estimation to some desired position. We refer these as coordinated outliers. This stronger outlier against the radar network may be due to multiple compromised nodes to make a target location estimate its position at β_f, which is determined by adversary. In this outlier model, $c_i = a_i^T(\beta_f - \beta^*)$, where β^* is the optimal target location. As a result, estimate of target location is forced towards β_f. Thus, the outlier model is define as

$$b_i = \begin{cases} a_i^T\beta + \epsilon_i + a_i^T(\beta_f - \beta^*) & i\epsilon outlier \\ a_i^T\beta + \epsilon_i & otherwise \end{cases} \tag{5.4}$$

5.1.2 Problem Formulation

A multiple distributed collocated radar system is considered with single stationary target P_0 at position $\beta = [x, y]$. Given set of noisy measurements, d_i in equation (5.2) can be expressed as

$$d_i = r_i + n_i, i = 1, 2, ..., N,$$

where $r_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$ is actual distance between i^{th} transmitter and target, n_i is the measurment noise. Let the noise associated term $2r_in_i = \epsilon_i$, the above equation can be expressed in matrix form as

$$b = A\beta + \epsilon,$$

where, $A = \begin{bmatrix} x_1 & y_1 & -0.5 \\ : & : & : \\ x_N & y_N & -0.5 \end{bmatrix}$, $\beta = \begin{bmatrix} x \\ y \\ R^2 \end{bmatrix}$ and $b = \frac{1}{2}\begin{bmatrix} x_1^2 + y_1^2 - d_1^2 \\ : \\ x_N^2 + y_N^2 - d_N^2 \end{bmatrix}$, with $R^2 = x^2 + y^2$.

Given the set of noisy observations with outliers, it is possible to estimate β from linear form (5.1.2) by minimizing empirical loss

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} h(b_i - a_i^T \beta), \tag{5.5}$$

where $h(.)$ is the cost function. Generally, a least square criterion is minimized where $h(.)$ is a square function. A large number of techniques for the minimization of equation (5.5) where robust cost functions such as L_1 cost function, are employed and use complex optimization strategy to minimize these cost function. In this work, we propose simple optimization strategies that iteratively solve the weighted least square cost function to find the robust solution. Thus, in addition of being robust, the proposed techniques inherit the advantage in term of accuracy as least square approach. This is achieved by iteratively minimizing the weighted least square function

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} w_i (b_i - a_i^T \beta)^2, \tag{5.6}$$

where w_i is the scalar weight associated with cost function $h(.)$ in equation (5.5).

5.2 Proposed Method for Robust Localization

The algorithm is developed to make make localization techniques robust to adversarial corruption of measurement data. The $M-$estimator for $\boldsymbol{\beta}$ is that which maximizes the likelihood function is

$$\prod_{i=1}^{n} f(\epsilon_i) = \prod_{i=1}^{n} h(b_i - \mathbf{a_i^T}\beta),$$

or equivalently, to maximize the log-likelihood function

$$\sum_{i=1}^{n} ln\left(f(\epsilon_i)\right) = \sum_{i=1}^{n} ln\left(h(b_i - \mathbf{a_i^T}\beta)\right), \tag{5.7}$$

where $h(u)$ is a suitable cost function, which can be Huber's cost, truncated least squares cost or any other robust cost function. Minimizing equation equation (5.7) to find an estimate requires partial differentiation with respect to each of the parameters in turn, resulting in a system of p equations;

$$\sum_{i=1}^{n} \psi(b_i - \mathbf{a_i^T}\beta)a_{ij} = 0, j = 1, 2, ..., p,$$

where $\frac{\partial h(u)}{\partial u} = \psi(u)$. Let's define the weight $w_i = \frac{1}{u_i}\psi(u_i) * sign(u_i)$. Then

$$\sum_{i=1}^{n} \psi(b_i - \mathbf{a_i^T}\beta)a_{ij} = \sum_{i=1}^{n} w_i(b_i - \mathbf{a_i^T}\beta)a_{ij} = 0,$$

and

$$\sum_{i=1}^{n} w_i b_i a_{ij} = \sum_{i=1}^{n} w_i \mathbf{a_i^T}\beta a_{ij}, j = 1, 2, ..., p.$$

Define weight matrix $W = diag(w_i) = \begin{pmatrix} w_1 & 0 & ... & 0 \\ 0 & w_2 & ... & 0 \\ ... & 0 & ... & 0 \\ 0 & 0 & ... & w_n \end{pmatrix}$. The system of equations can be expressed as the penalization by weight at which the the greatest residuals are penalized by relatively smaller weighting functions. Solution of the robust regression is defined in equation (5.6)as

$$A^T W A \beta = A^T W b,$$

hence the location estimate is

$$\hat{\beta} = (A^T W A)^{-1} A^T W b. \tag{5.8}$$

Therefore, this is very similar to the solution for the least squares estimator, but introduces a weight matrix to reduce the influence of outliers. Hence optimization of equation (5.5) is equivalent to optimization of equation (5.6).

Generally, two scalar weights exist for the robust cost family $h(.)$, non-truncated and truncated. The main advantage of using non-truncated weight such as the Huber cost is that global minimization is assured. However, the influence of outliers is always present as they are never entirely discarded.

In the following, we propose two different truncated weight approaches. The first one is regularized least square regression, which aims to assign the weight dynamically based on the inverse of the absolute value of residue. The second is Least Trimmed Square (LTS) with truncated weighting that keeps smaller residuals discarding significant residuals.

5.2.1 Regularized Robust Least-Squares Regression

Given a set of data points in design matrix A and the corresponding observation vector b, the goal is to estimate a parameter vector β based on the robust regression equation (5.7). Notice that the corrupted points are likely to suffer from the larger residues, we define the weights to every residue as $s_i = \frac{1}{|b_i - \mathbf{a_i^T}\beta|}$ and regularized the weights as $w_i = min(s_i, \delta)$, where $\delta > 0$ is a lower bound for the weight. Now, the robust regression can be solved using following two steps

1. Weighting: For the given model, assign weight to every residues as $w_i = min\left(\frac{1}{|b_i - \mathbf{a_i^T}\beta|}, \delta\right)$
2. Solve the weighted least square problems, i.e. $argmin_\beta \sum_{i=1}^{n} w_i(b_i - \mathbf{a_i^T}\beta)^2$ with above weights to obtain next estimate as $\beta = \left[A^T W A\right]^{-1} A^T W b$, where $W = diag(w_i)$

The intuition behind this procedure is that outliers in the measurements, such as the points with more significant residuals should get down-weighted. If the regularized parameter δ is too small, no data points get considerable weight due to aggressive truncation. However, to converge to the actual β, the non-corrupted residues should get a considerable weight; hence setting a small value of δ could not guarantee the estimation converge towards the optimum β. If we always use the considerable value of δ, and are unlucky enough to initialize the iterative method close to β_f, the wrong target location is introduced by Player 1 through coordinated outliers, then a set of outliers get larger weights. In contrast, the nominal points initially get comparatively smaller weights which will cause the algorithm to converge towards β_f, instead of the actual local β.

The above limitations of Iteratively Reweighted Least Sqaures (IRLS) were well explained in [70]. To remedy this, we can execute the algorithm in stages, with initial stages employing aggressive truncation with a small value of δ and later stages successively relaxing the truncation. At the convergence stage, the larger residuals are weighted as close to the L_1 norm and nominal nodes are weighted close to the L_2 norm.

Algorithm 1 Regularized Weighted IRLS algorithm

```
1. Select initial estimate $\beta^{(0)}$, $A$, $b$, and regularized parameter $\delta$, $\gamma > 1$.
2. Do:
    (a) compute $w^t$ using
            $w_i^t = min\left(\frac{1}{\mid b_i - \mathbf{a_i^T}\beta^{\mathbf{t}} \mid}, \delta\right), W^t = diag(w_i^t)$
    (b) predict: $\hat{\beta}^{t+1} = (A^T W^t A)^{-1} A^T W^t b$
    (c) update: $\delta = \delta * \gamma$
    (d) While $\left\|\hat{\beta}^{t+1} - \hat{\beta}^t\right\| >$ error tolerance
3. Report the final estimate $\hat{\beta}^{t+1}$.
```

Successively relaxing the δ is that if we initialize unfortunately at β_f, the nominal points receive relatively smaller weights. However, when the δ is relaxed later, will allow these nominal points to assign large weights. The algorithm hopefully converges towards β^*.

5.2.2 Least Trimmed Regression

LTS is another robust regression that can be seen as an $M-$estimators but with a truncating weight function. It keeps only the smaller residual values and discards the others. Although this can not assure a global minimization, LTS approach is more robust to outliers as it does not take them into account, provided they are correctly discarded.

The classical LS estimate of β aim at solving

$$\hat{\beta}_{LS} = \arg\min_{\beta} \sum_{i=1}^{n} (r_i(\beta))^2 = \arg\min_{\beta} \sum_{i=1}^{n} (b_i - a_i^T \beta)^2. \tag{5.9}$$

In the robust framework, we assume that a portion of the observations may be corrupted, i.e., outliers. This results large residuals that can spoil the least square estimator. LTS is a classical robust estimator that discards the largest residuals. For the given β, let $r_i(\beta)$ for $i = 1, ..., n$ be the ordered absolute residuals such that $\mid r_{(1)} \mid \leq \mid r_{(2)} \mid, ... \leq \mid r_{(n)} \mid$. Let w_i be the indicator for whether observation i is a good observation or not. Then, it is easy to see that the least trimmed squares estimation problem can be reformulated as the following problem

$$\begin{aligned} &\arg\min_{\beta, w_i} \sum_{i=1}^{n} w_i (b_i - a_i^T \beta)^2, \\ &s.t, \mathbf{w}e^T = k \\ &w_i = \{0, 1\}^n \end{aligned} \tag{5.10}$$

This optimization is now performed jointly on β and the binary weight vector $w\epsilon\{0,1\}^n$. Therefore

$$F(\beta, w_i) = \arg\min_{\beta, w_i} \left\{(b - A\beta)^T W (b - A\beta)\right\}. \tag{5.11}$$

Consequently, the LTS regression problem is a minimization problem in $n + p$ real variables. The objective function in equation (5.11) is continuously differentiable. We can use the Kuhn-Tucker (KKT) conditions to characterize the local minimum. The resulting system of equations and inequalities are nonlinear in its variable, which means that it has to be solved iterativetly. For the fixed w, the global minimizer over β is given by

$$\hat{\beta} = (A^T W A)^{-1} A^T W b,$$

where $W = diag(w_i)$ is the diagonal weighting matrix. For a fixed β, the global minimizer over w is the binary vector such that, for $i = 1, 2, ..., n$

$$w_i^t = \begin{cases} 1 & \mid r_{(i)} \mid \leq \mid r_{(h)} \mid \\ 0 & \text{otherwise.} \end{cases}$$

Ideally, when the squared residual $(b_i - a_i^T \beta)$ is larger, the corresponding weight $w_i = 0$. The integer constraints can be relaxed to linear constraints as follow

$$\begin{aligned} \arg\min_{\beta, w_i} \sum_{i=1}^{n} w_i (b_i - a_i^T \beta)^2, \\ s.t, \mathbf{w}e^T = k \\ 0 \leq w_i \leq 1. \end{aligned} \tag{5.12}$$

The equation (5.10) and (5.12) are equivalent. When we fix $\boldsymbol{\beta}$, the optimization problem in equation (5.12) is linear problem in $\mathbf{w}$. Therefore, the optimal solution of $\mathbf{w}$ must be achieved at an extreme point of the feasible set, for which w_i can only be either 0 or 1.

Algorithm 2 Least Trimmed Square algorithm

```
1. $A$ ,$b$, number of non- outliers $h$, $\hat{\beta}^0$ (initialization)
2. Do:
    (a) (a) compute $w^t$ using
        $$w_i^t = \begin{cases} 1 & \mid r_{(i)} \mid \leq \mid r_{(h)} \mid \\ 0 & \text{otherwise} \end{cases}$$
    (b) $W^t = diag(w_i^t)$
    (c) predict: $\hat{\beta}^{t+1} = (A^T W^t A)^{-1} A^T W^t b$
3. $\left\| \hat{\beta}^{t+1} - \hat{\beta}^t \right\| >$ error tolerance
4. Report the final estimate $\hat{\beta}^{t+1}$
5. output: $\hat{\beta}^t$
```

5.3 Localization Accuracy Improvement using Anchor Nodes

Many existing robust methods assume the signal quality/uncertainty is the same among different measurements. In practice, especially in a potentially hostile environments, we might be able to improve the signal quality and integrity for a small set of sensors by spending more resources on these devices. In such cases, the existence of such secured sensors can be employed to increase localization accuracy. That is, let S_T be secured out of N sensors where no chance to be controlled by Player 1. We define such nodes as “anchor nodes”. For the sake of simplicity, the rest of node $S \in S_T^C$ are assumed to be equally unreliable. As a result, we can trust the anchor nodes observations more by assigning some larger weights on their observations. In this preliminary study, we consider the estimate of $\boldsymbol{\beta}$ in this case by assigning trusted weight, w_i^T for $S \in S_T$ and rest, $S \in S_T^C$ by robust weights, w_i^R. The robust location estimate in the presence of anchor node is weighted-least square estimates

$$\beta^{(t)} = \left[AW^{(t-1)}A\right]^{-1} AW^{(t-1)}b, \tag{5.13}$$

where A is the model matrix, with $a_i^{'}$ as i^{th} row, and $W^{(t-1)} = diag\left\{w_i^{(t-1)}\right\}$ such that if $i \in S_T, w_i = w_i^T$, otherwise $w_i = w_i^R$ for any of least trimmed method or regularized robust least square method.

Algorithm 3 Anchored sensor based robust localization

```
1. $A$, $b$, number of non- outliers $h$ (for least trimmed), $\delta$, $\gamma > 0$ (for regularized robust), $\hat{\beta}^0$ (initialization)
2. Do:
    (a) Compute $w^t$
        i. For least trimmed
            $w_i^t = \begin{cases} 1 & \mid r_{(i)} \mid \leq \mid r_{(h)} \mid \\ 0 & \text{otherwise} \end{cases}$
        ii. For regularized robust
            $w_i^t = min\left(\frac{1}{\mid b_i - \mathbf{a_i^T}\beta^{\mathbf{t}} \mid}, \delta\right)$
    (b) Predict $\hat{\beta}^{t+1} = (A^T W^t A)^{-1} A^T W^t b$, where
            $W^t = diag(w_i^t), w_i^t = \begin{cases} w_i^{tR} & i \in S_R \\ w_i^{tT} & i \in S_T \end{cases}$
3. While $\left\| \hat{\beta}^{t+1} - \hat{\beta}^t \right\| >$ error tolerance
4. Output: $\hat{\beta}^t$
```

5.4 Simulation Results

To test the performance of the above proposed localization algorithms, we assume that Player 1 successfully gained the ability to modify the distance measurements for a fraction of total transmitters arbitrarily. The adversary aims to drive the location estimate as far away from the true location as possible. We consider a 10 transmitters geometry with the coordinates $[0,0]$, $[3\sqrt{3},3]$, $[0,6]$, $[-3\sqrt{3},3]$, $[-3\sqrt{3},-3]$, $[4,4]$,

$[5,5]$, $[7,-5]$, $[9,6]$, $[10,-60]$. All results are averages of 10000 independent runs. The following figures show the mean square errors of two robust methods discussed in the above section with actual target location $[x,y]=[1,2]$ with all distance units in km.

For regularized robust algorithms, the starting truncation parameter $\delta=0.00001$ in such way that the algorithm assigns comparatively smaller weights for outliers and larger weights for nominal measurements.

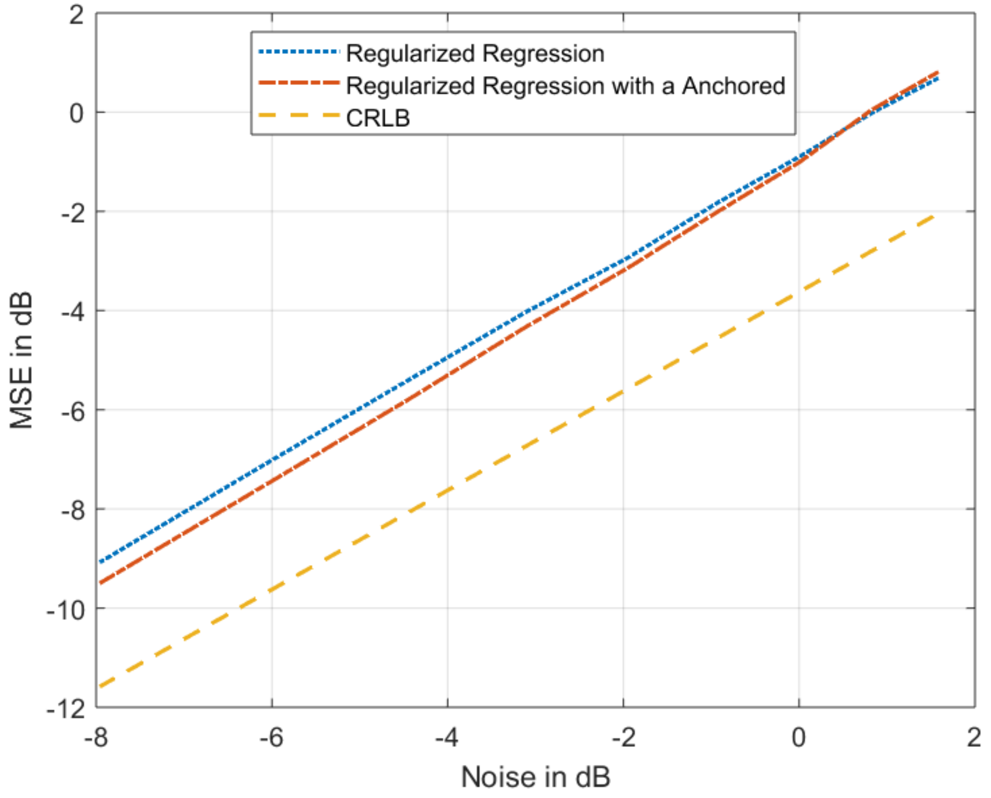

Figure 5.1: Regularized regression under nominal condition.

Fig 5.1 shows the performance of regularized regression under the nominal condition. Here we set the location estimation by non-robust least square at the nominal condition as the baseline for comparison for regularized regression both under nominal and attack conditions. Under nominal conditions, there is at least -2.5dB performance loss for the regularized regression compared with least square regression. It is seen that the performance loss can be decreased by introducing the anchor node. Here, we set transmitter 1 as the anchor node for this particular simulations and assigned larger weights for that node, e.g., $w_1 = 1$ for the regularized robust regression.

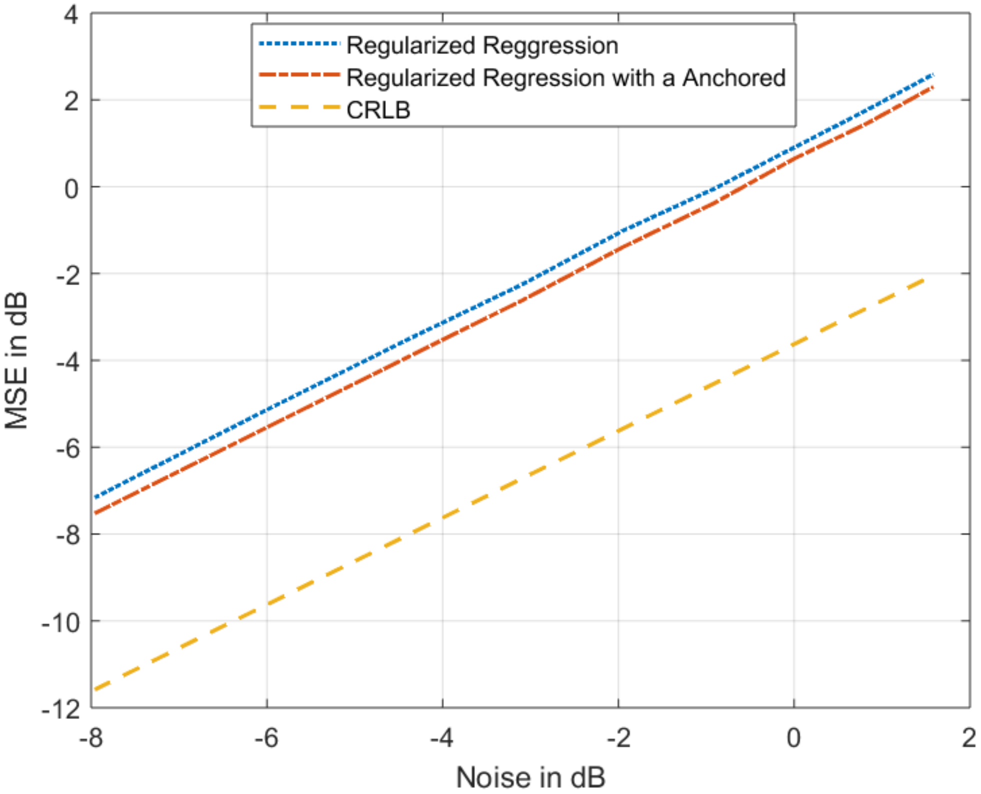

Figure 5.2: Regularized regression under 1 out of 10 transmitter is compromised.

For the least trimmed localization, the simulation is preformed at trimmed rate 10% (1 out of 10 observations are compromised) for least trimmed regression. We set transmitter 1 as the anchor node and assigned arbitrary larger weight for that node. In particular we assigned weight, $w_1 = 2$.

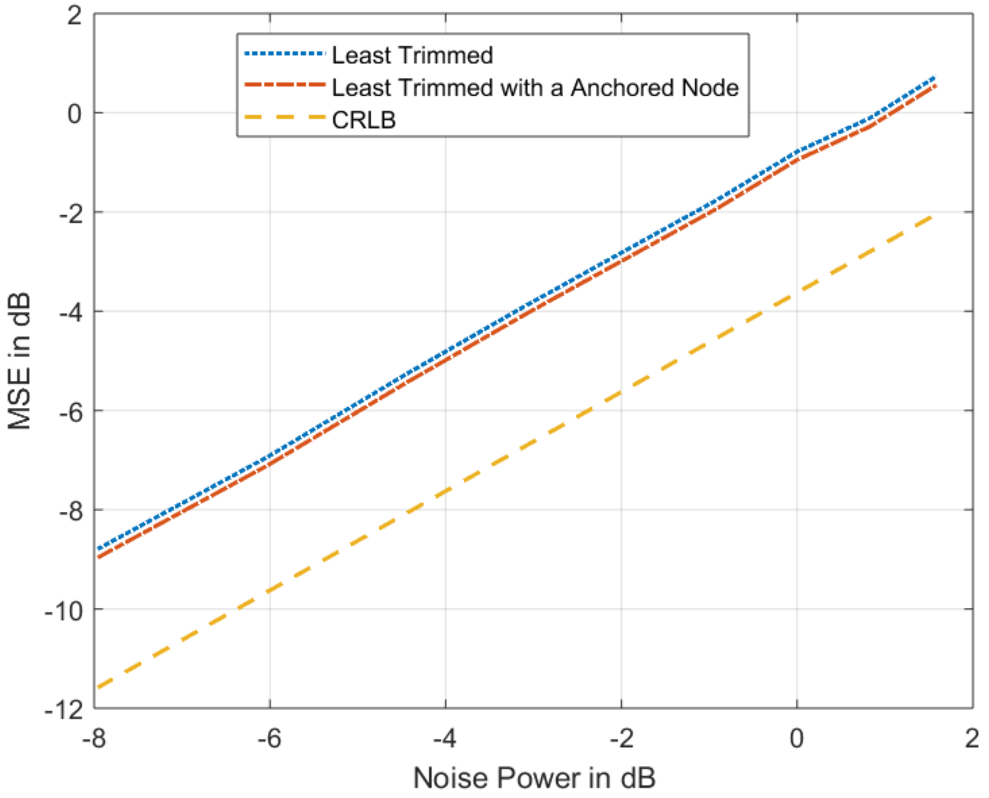

Figure 5.3: Regularized regression under 1 out of 10 transmitter is compromised.

Fig 5.3 shows the performance of least trimmed regression under 1 out of 10 transmitters case. Even in such adverse conditions, it is possible to estimate the location with a certain performance accuracy. It is seen that the presence of anchor nodes can reduce performance loss.

CHAPTER 6:
CONCLUSION

6.1 Summary

The main focus of this dissertation is robust inference, mainly detection and estimation methods. To make some progress on the robust system design in WSNs, the node venerability is quantified using both probabilistic and static models. The inference problem is formulated as a two-player static zero-sum game under the game theoretical framework. The zero-sum game is solved through robust inference theory, and the mini-max solutions and their related inference performance are determined. The proposed robust inference methods performed significantly better than nominal ones under the compromised nodes.

In Chapter 3, two detectors were proposed under the static setting. A preliminary analysis of these two detectors was conducted where some nodes in WSNs could be compromised and provide false information. Through both numerical and theoretical analysis for the case of target detection in a distributed MIMO radar, it was shown that these detectors provide some level of guaranteed performance despite nodes venerability.

In Chapter 4, both the Frequentist's and Bayesian frameworks were investigated. Main results were obtained under the complete information case where Player 1 knows the parameter of interest with some limited results under the no informative setting. Under the frequentist framework, the max-mini solutions were derived that make the observation partially uninformative and sometimes even complete uninformative. This revealed the sensor network performance limits due to the nodes' reliability. The saddle point solution was obtained for a single-node network for the Bayesian estimation where a prior distribution is known. The breaking down point that results in complete uninformative was lower than the Frequentist's one. In addition, the no information case was shown to be a particular case of complete information and solved. Inspired by the single-node closed-form solution, a robust estimator for the multi-node networks was purposed and strong robustness was obtained with a minor performance degradation under the nominal scenario.

In Chapter 5, robust algorithms were proposed for target localization. The hostile environment was considered, where some observations in TOA measurements could be compromised and provide false information. It was validated through numerical simulations for target localization for distributed radars that these proposed algorithms provide some guaranteed localization performance despite extreme outliers in the observation.

6.2 Some Open Future Research Topics

We investigated robust detection and estimation in WSNs and extended our result for a few practical applications for multi-static radars. Still, many other research topics within this framework can also be investigated.

- An immediate research problem to be solved is to complete the estimation problem for multi-node networks. One needs first to determine whether a saddle point solution indeed exists or not. While we conjured that it does not work for general cases, it is critical to seek a definitive answer for some important practical problems, e.g., localization in MIMO radars.

- As in the other inference tasks, more difficulty arises when one goes beyond the single-stage game to consider the problem for multiple stages where Player 1's overly aggressive may yield a more considerable gain at the beginning but diminishing return at the latter stages. The strategy evolution during the different stages is challenging. It is relatively easy to analyze the performance for specific strategies but somewhat complicated to solve for either the Max-mini or Mini-max strategies. The existence of a saddle point solution is still unknown. The inference problem becomes much more complicated and exciting when the inference game is played for many stages. Often, Player 1 must seek a balance between inserting false information and keeping the identity of the malfunctioning node secret. Such trade-off depends mainly on the relative values of current and future payoffs. While Player 1 can still cause significant performance degradation in earlier stages by behaving aggressive in the beginning, such abnormal behavior may be observed and explored by Player 2 to design more targeted algorithms in the later stages to identify and limit the impact of the malfunctioning nodes, and if possible, to use the information of those nodes.

- We proposed two robust target localization methods in unreliable nodes where some of the measurements could get compromised and provide false information. It was shown that the presence of anchor nodes could further reduce the performance loss, and we verified it through numerical simulations. The optimal trust in terms of weights for the anchor nodes for a given number of compromised nodes needs to be explored.

- Robust inference is yet to be explored. In statistical learning, one often relies on a set of data samples to train a model to fit the data and consequently to use the model for future applications. The application performance relies heavily on the model's accuracy. One common primitive in statistical learning is that the collected data are i.i.d. according to the underlying model encountered in practical applications. Therefore, the more data one collects, the better the learned model's accuracy, and consequently, the better the application performs. However, the fundamental assumption that the collected data sample represents the accurate application model is questionable in some cases. For example, inconsistency may arise when the training samples are mislabeled, the dynamic model has changed between the time of learning and application, the bias in the sampling process, or the lack of sufficient training samples. The important direction for further exploration would be to investigate the problems of the mismatched training model in learning and decision making.

REFERENCES

[1] Ian F. Akyildiz, Tommaso Melodia, and Kaushik R. Chowdury. Wireless multimedia sensor networks: A survey. *IEEE Wireless Communications*, 14(6):32–39, 2007. doi: 10.1109/MWC.2007.4407225.

[2] M. Joharan Beevi. A fair survey on internet of things (iot). In *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, pages 1–6, 2016. doi: 10.1109/ICETETS.2016.7603005.

[3] R. Viswanathan and P.K. Varshney. Distributed detection with multiple sensors part i. fundamentals. *Proceedings of the IEEE*, 85(1):54–63, 1997. doi: 10.1109/5.554208.

[4] Hoon Kim and Sang-wook Han. An efficient sensor deployment scheme for large-scale wireless sensor networks. *IEEE Communications Letters*, 19(1):98–101, 2015. doi: 10.1109/LCOMM.2014.2372015.

[5] R.S. Blum, S.A. Kassam, and H.V. Poor. Distributed detection with multiple sensors ii. advanced topics. pages 64–79. doi: 10.1109/5.554209.

[6] Jean-francois Chamberland and Venugopal V. Veeravalli. Wireless sensors in distributed detection applications. 24(3):16–25, 2007. doi: 10.1109/MSP.2007.361598.

[7] J.-F. Chamberland and V.V. Veeravalli. Decentralized detection in sensor networks. 51:407–416, 2003. doi: 10.1109/TSP.2002.806982.

[8] Z. Chair and P.K. Varshney. Optimal data fusion in multiple sensor detection systems. *IEEE Transactions on Aerospace and Electronic Systems*, AES-22:98–101, 1986. doi: 10.1109/TAES.1986.310699.

[9] Muhammad Ayaz, Mohammad Ammad-uddin, Imran Baig, and el-Hadi M. Aggoune. Wireless sensor's civil applications, prototypes, and future integration possibilities: A review. *IEEE Sensors Journal*, 18(1):4–30, 2018. doi: 10.1109/JSEN.2017.2766364.

[10] Ismail Butun, Salvatore D. Morgera, and Ravi Sankar. A survey of intrusion detection systems in wireless sensor networks. *IEEE Communications Surveys Tutorials*, 16(1):266–282, 2014. doi: 10.1109/SURV.2013.050113.00191.

[11] R.S. Blum, S.A. Kassam, and H.V. Poor. Distributed detection with multiple sensors ii. advanced topics. *Proceedings of the IEEE*, 85(1):64–79, 1997. doi: 10.1109/5.554209.

[12] J. Tsitsiklis and M. Athans. On the complexity of decentralized decision making and detection problems. 30(5):440–446, 1985. doi: 10.1109/TAC.1985.1103988.

[13] J.N. Tsitsiklis. Extremal properties of likelihood-ratio quantizers. 41:550–558. doi: 10.1109/26.223779.

[14] Venugopal V. Veeravalli and Pramod K. Varshney. Distributed inference in wireless sensor networks. *Philosophical Transactions of the Royal Society A: Math-*

ematical, Physical and Engineering Sciences, 370(1958):100–117, January 2012. ISSN 0962-8428. doi: 10.1098/rsta.2011.0194.

[15] X. Nguyen, M.J. Wainwright, and M.I. Jordan. Nonparametric decentralized detection using kernel methods. *IEEE Transactions on Signal Processing*, (11): 4053–4066, 2005. doi: 10.1109/TSP.2005.857020.

[16] L.A. Rossi, B. Krishnamachari, and C.-C.J. Kuo. Hybrid data and decision fusion techniques for model-based data gathering in wireless sensor networks [environmental monitoring applications]. In *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, volume 7, pages 4616–4620 Vol. 7, 2004. doi: 10.1109/VETECF.2004.1404965.

[17] G. Sudha, R. Prakash, A. Balaji Ganesh, and Siva V Girish. Network coding based real time wireless sensor network for environmental monitoring. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pages 1269–1272, 2016. doi: 10.1109/WiSPNET.2016.7566340.

[18] Jing Wang, In Soo Ahn, Yufeng Lu, and Gennady Staskevich. A new distributed algorithm for environmental monitoring by wireless sensor networks with limited communication. In *2016 IEEE SENSORS*, pages 1–3, 2016. doi: 10.1109/ICSENS.2016.7808976.

[19] Asyraf Hakimi, Najmuddin Hassan, Khairul Anwar, Ammar Zakaria, and Ahmad Ashraf. Development of real-time patient health (jaundice) monitoring using wireless sensor network. In *2016 3rd International Conference on Electronic Design (ICED)*, pages 404–409, 2016. doi: 10.1109/ICED.2016.7804678.

[20] Wael Abdullah Ahmad, Maciej Kucharski, Arzu Ergintav, Salah Abouzaid, Jan Wessel, Herman Jalli Ng, and Dietmar Kissinger. Multimode w-band and d-band mimo scalable radar platform. *IEEE Transactions on Microwave Theory and Techniques*, 69(1):1036–1047, 2021. doi: 10.1109/TMTT.2020.3038532.

[21] Igal Bilik, Oded Bialer, Shahar Villeval, Hasan Sharifi, Keerti Kona, Marcus Pan, Dave Persechini, Marcel Musni, and Kevin Geary. Automotive mimo radar for urban environments. In *2016 IEEE Radar Conference (RadarConf)*, pages 1–6, 2016. doi: 10.1109/RADAR.2016.7485215.

[22] Brian B. Tierney and Christopher T. Rodenbeck. 3d-sensing mimo radar for uav formation flight and obstacle avoidance. In *2019 IEEE Radio and Wireless Symposium (RWS)*, pages 1–3, 2019. doi: 10.1109/RWS.2019.8714287.

[23] Jens Klare, Oliver Biallawons, and Delphine Cerutti-Maori. Uav detection with mimo radar. In *2017 18th International Radar Symposium (IRS)*, pages 1–8, 2017. doi: 10.23919/IRS.2017.8008140.

[24] Jiho Seo, SeongJun Hwang, Yong-gi Hong, Jaehyun Park, Sunghyun Hwang, and Woo-Jin Byun. Bayesian matching pursuit-based distributed fmcw mimo radar imaging. *IEEE Systems Journal*, 15(3):4623–4634, 2021. doi: 10.1109/JSYST.2020.3031912.

[25] Wichai Pawgasame. A survey in adaptive hybrid wireless sensor network for military operations. In *2016 Second Asian Conference on Defence Technology (ACDT)*, pages 78–83, 2016. doi: 10.1109/ACDT.2016.7437647.

[26] Ramasamy Mariappan, P. V. Narayana Reddy, and Chang Wu. Cyber physical

system using intelligent wireless sensor actuator networks for disaster recovery. In *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, pages 95–99, 2015. doi: 10.1109/CICN.2015.28.

[27] Myounggyu Won, HoKyeong Ra, Taejoon Park, and Sang H. Son. Modeling random deployment in wireless sensor networks for infrastructure-less cyber physical systems. In *2014 IEEE International Conference on Cyber-Physical Systems, Networks, and Applications*, pages 81–86, 2014. doi: 10.1109/CPSNA.2014.26.

[28] Siddhartha Kumar Khaitan and James D. McCalley. Design techniques and applications of cyberphysical systems: A survey. *IEEE Systems Journal*, 9(2): 350–365, 2015. doi: 10.1109/JSYST.2014.2322503.

[29] Yufei Wang, Weimin Lin, and Tao Zhang. Study on security of wireless sensor networks in smart grid. In *2010 International Conference on Power System Technology*, pages 1–7, 2010. doi: 10.1109/POWERCON.2010.5666729.

[30] Benazir Fateh, Manimaran Govindarasu, and Venkataramana Ajjarapu. Wireless network design for transmission line monitoring in smart grid. *IEEE Transactions on Smart Grid*, 4(2):1076–1086, 2013. doi: 10.1109/TSG.2013.2241796.

[31] K. Kim and G. Shevlyakov. Why gaussianity? pages 102–113. doi: 10.1109/MSP.2007.913700.

[32] S. A. Kassam and H. V. Poor. Robust techniques for signal processing: A survey. *Proceedings of the IEEE*, 73:433–481, 1985. doi: 10.1109/PROC.1985.13167.

[33] T.K. Blankenship, D.M. Kriztman, and T.S. Rappaport. Measurements and simulation of radio frequency impulsive noise in hospitals and clinics. In *1997*

IEEE 47th Vehicular Technology Conference. Technology in Motion, volume 3, pages 1942–1946 vol.3, 1997. doi: 10.1109/VETEC.1997.605897.

[34] D. Middleton. Non-gaussian noise models in signal processing for telecommunications: new methods an results for class a and class b noise models. *IEEE Transactions on Information Theory*, 45(4):1129–1149, 1999. doi: 10.1109/18.761256.

[35] George Dimitrakopoulos and Panagiotis Demestichas. Intelligent transportation systems. *IEEE Vehicular Technology Magazine*, 5(1):77–84, 2010. doi: 10.1109/MVT.2009.935537.

[36] Mounib Khanafer, Mouhcine Guennoun, and Hussein T. Mouftah. Wsn architectures for intelligent transportation systems. In *2009 3rd International Conference on New Technologies, Mobility and Security*, pages 1–8, 2009. doi: 10.1109/NTMS.2009.5384685.

[37] Athanasios Maimaris and George Papageorgiou. A review of intelligent transportation systems from a communications technology perspective. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 54–59, 2016. doi: 10.1109/ITSC.2016.7795531.

[38] L.A. Rossi, B. Krishnamachari, and C.-C.J. Kuo. Hybrid data and decision fusion techniques for model-based data gathering in wireless sensor networks [environmental monitoring applications]. In *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, volume 7, pages 4616–4620 Vol. 7, 2004. doi: 10.1109/VETECF.2004.1404965.

[39] H. Chen and B. Himed. Analyzing and improving mimo radar detection perfor-

mance in the presence of cybersecurity attacks. In *2016 IEEE Radar Conference (RadarConf)*, pages 1–4, 2016. doi: 10.1109/RADAR.2016.7485177.

[40] Y.A. Chau and E. Geraniotis. Distributed multisensor parameter estimation in dependent noise. *IEEE Transactions on Communications*, 40:373–384, 1992. doi: 10.1109/26.129199.

[41] J.A. Gubner. Distributed estimation and quantization. 39:1456–1459, 1993. doi: 10.1109/18.243470.

[42] A. Willsky, M. Bello, D. Castanon, B. Levy, and G. Verghese. Combining and updating of local estimates and regional maps along sets of one-dimensional tracks. *IEEE Transactions on Automatic Control*, 27:799–813, 1982. doi: 10.1109/TAC.1982.1103019.

[43] Peter J. Huber. pages 73–101. Number 1. Institute of Mathematical Statistics, March . doi: 10.1214/aoms/1177703732.

[44] Peter J. Huber. A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, 36(6):1753–1758, 1965. ISSN 00034851. URL `http://www.jstor.org/stable/2239116`.

[45] E. Fishler, A. Haimovich, R. S. Blum, L. J. Cimini, D. Chizhik, and R. A. Valenzuela. Spatial diversity in radars—models and detection performance. *IEEE Transactions on Signal Processing*, 54(3):823–838, March 2006. doi: 10.1109/TSP.2005.862813.

[46] B. J. Donnet and I. D. Longstaff. Mimo radar, techniques and opportunities.

In *2006 European Radar Conference*, pages 112–115, Sep. 2006. doi: 10.1109/EURAD.2006.280286.

[47] L. Pescosolido, S. Barbarossa, and G. Scutari. In *2008 IEEE Radar Conference*.

[48] Qian He, Rick S. Blum, and Alexander M. Haimovich. Noncoherent mimo radar for location and velocity estimation: More antennas means better performance. *Trans. Sig. Proc.*, 58(7), 2010. ISSN 1053-587X.

[49] I. Bekkerman and J. Tabrikian. Target detection and localization using mimo radars and sonars. *IEEE Transactions on Signal Processing*, 54(10):3873–3883, 2006.

[50] D. W. Bliss and K. W. Forsythe. Multiple-input multiple-output (mimo) radar and imaging: degrees of freedom and resolution. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 1, pages 54–59 Vol.1, 2003.

[51] A. M. Haimovich, R. S. Blum, and L. J. Cimini. Mimo radar with widely separated antennas. 25(1):116–129, 2008. ISSN 1053-5888. doi: 10.1109/MSP.2008.4408448.

[52] J. Li and P. Stoica. Mimo radar with colocated antennas. *IEEE Signal Processing Magazine*, 24(5):106–114, 2007.

[53] Cong Xiang, Da-Zheng Feng, Hui Lv, Jie He, and Yang Cao. Robust adaptive beamforming for mimo radar. 90(12):3185 – 3196, 2010. doi: https://doi.org/10.1016/j.sigpro.2010.05.022. URL `http://www.sciencedirect.com/science/article/pii/S0165168410002240`.

[54] Sergiy A. Vorobyov. Principles of minimum variance robust adaptive beamforming design. 93:3264 – 3277, 2013. doi: https://doi.org/10.1016/j.sigpro.2012.10.021. URL `http://www.sciencedirect.com/science/article/pii/S0165168412003830`. Special Issue on Advances in Sensor Array Processing in Memory of Alex B. Gershman.

[55] N. Kitbutrawat, P. Lopattanakij, P. Tiwatthanont, S. Thirachai, and J. Suwatthikul. Sensor drift detection by utilizing multi-sensor signals. In *SICE Annual Conference 2011*, pages 1523–1527, Sep. .

[56] Sungwhan Cho and Jin Jiang. Detection and estimation of sensor drifts using kalman filters with a demonstration on a pressurizer. *Nuclear Engineering and Design*, 242:389 – 398. ISSN 0029-5493. doi: https://doi.org/10.1016/j.nucengdes.2011.10.018. URL `http://www.sciencedirect.com/science/article/pii/S0029549311008910`.

[57] Jinran Chen, Shubha Kher, and Arun Somani. Distributed fault detection of wireless sensor networks. 2006. doi: 10.1145/1160972.1160985.

[58] Ihab Samy, Ian Postlethwaite, and Da-Wei Gu. Survey and application of sensor fault detection and isolation schemes. *Control Engineering Practice*, 19(7): 658 – 674, 2011. ISSN 0967-0661. doi: https://doi.org/10.1016/j.conengprac.2011.03.002. URL `http://www.sciencedirect.com/science/article/pii/S0967066111000414`.

[59] Y. Yang and R. S. Blum. Minimax robust mimo radar waveform design. *IEEE Journal of Selected Topics in Signal Processing*, pages 147–155. ISSN 1941-0484.

[60] Koen Langendoen and Niels Reijers. Distributed localization in wireless sensor networks: a quantitative comparison. *Computer Networks*, 43 (4):499 – 518, 2003. ISSN 1389-1286. doi: https://doi.org/10.1016/S1389-1286(03)00356-6. URL `http://www.sciencedirect.com/science/article/pii/S1389128603003566`. Wireless Sensor Networks.

[61] G. Naddafzadeh-Shirazi, M. B. Shenouda, and L. Lampe. Second order cone programming for sensor network localization with anchor position uncertainty. *IEEE Transactions on Wireless Communications*, 13(2):749–763, 2014.

[62] Pratik Biswas, Tzu-Chen Lian, Ta-Chung Wang, and Yinyu Ye. Semidefinite programming based algorithms for sensor network localization. *ACM Trans. Sen. Netw.*, 2(2):188–220, May 2006. ISSN 1550-4859. doi: 10.1145/1149283.1149286. URL `https://doi.org/10.1145/1149283.1149286`.

[63] K. W. Cheung, W. K. Ma, and H. C. So. Accurate approximation algorithm for toa-based maximum likelihood mobile location using semidefinite programming. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages ii–145, 2004.

[64] M. A. Spirito. On the accuracy of cellular mobile station location estimation. *IEEE Transactions on Vehicular Technology*, 50(3):674–685, May 2001. ISSN 1939-9359. doi: 10.1109/25.933304.

[65] J. Caffery and G. L. Stuber. Subscriber location in cdma cellular networks. *IEEE Transactions on Vehicular Technology*, 47(2):406–416, May 1998. ISSN 1939-9359. doi: 10.1109/25.669079.

[66] K. Yang, G. Wang, and Z. Luo. Efficient convex relaxation methods for robust target localization by a sensor network using time differences of arrivals. *IEEE Transactions on Signal Processing*, 57(7), 2009.

[67] K. C. Ho, X. Lu, and L. Kovavisaruch. Source localization using tdoa and fdoa measurements in the presence of receiver location errors: Analysis and solution. *IEEE Transactions on Signal Processing*, 55(2):684–696, 2007.

[68] Z. Li, W. Trappe, Y. Zhang, and Badri Nath. Robust statistical methods for securing wireless localization in sensor networks. In *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.*, pages 91–98, 2005.

[69] S. A. Vorobyov, A. B. Gershman, and Zhi-Quan Luo. Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem. *IEEE Transactions on Signal Processing*, 51(2):313–324, 2003.

[70] Khurrum Aftab and Richard Hartley. Convergence of iteratively re-weighted least squares to robust m-estimators. In *2015 IEEE Winter Conference on Applications ofComputer Vision*, pages 480–487, 2015. doi: 10.1109/WACV.2015.70.

[71] P. C. Niedfeldt and R. W. Beard. Convergence and complexity analysis of recursive-ransac: A new multiple target tracking algorithm. *IEEE Transactions on Automatic Control*, 61(2):456–461, 2016.

[72] K.W. Cheung, H.C. So, W.-K Ma, and Y.T. Chan. Least squares algorithms for time-of-arrival-based mobile location. *Signal Processing, IEEE Transactions on*, 52:1121 – 1130, 05 2004. doi: 10.1109/TSP.2004.823465.

APPENDIX A:
PDF OF SUM OF TWO CLIPPED NORMAL DISTRIBUTION

In this section, we derive the PDF of sum of two clipped normal distribution. As the PDF if clipped normal distribution is continuous distribution with two probability mass at ends of clipped points, the clipped normal distribution can be expressed in term of piece-wise expression as in. Therefore the sum of two clipped normal distribution can be expressed as

$$\begin{aligned}
f\left(\lambda(X)+\lambda(X)=t \mid H_0\right) &= f\left(\lambda(X)\right) * f\left(\lambda(X)\right) \\
&= Q^2\left(\frac{\tau+\mu}{\sigma}\right)\delta\left(X-2\tau\right) \\
&+ 2Q\left(\frac{\tau+\mu}{\sigma}\right)Q\left(\frac{\tau-\mu}{\sigma}\right)\delta\left(X\right) \\
&+ \phi(X+\mu)Q\left(\frac{\tau+\mu}{\sigma}\right)\delta\left(X-\tau\right) \\
&+ Q^2\left(\frac{\tau-\mu}{\sigma}\right)\delta\left(X+2\tau\right) \\
&+ \phi(X+\mu) * \phi(X+\mu),
\end{aligned}$$

where

$$\begin{aligned}\phi(X+\mu)*\phi(X+\mu) &= \int_{-\infty}^{\infty} f_X(u)f_X(v-u)du \\ &= \frac{1}{2\pi\sigma^4}\int\limits_{-\infty}^{\infty} \exp\left(-\frac{(u+\mu)^2}{2\sigma^2}\right) I_{u\epsilon[-\tau,\tau]} \\ &\exp\left(-\frac{(v-u+\mu)^2}{2\sigma^2}\right) I_{v-u\epsilon[-\tau,\tau]}du,\end{aligned}$$

with $I_{u\epsilon[-\tau,\tau]} = \begin{cases} 1 & -\tau \le u \le \tau, \\ 0 & otherwise \end{cases}$ is the indicator function.

Similarly, $f\left(\lambda(X)+\lambda(X)=t \mid H_1\right)$ can be evaluated.

APPENDIX B: SUM OF I.I.D. EXPONENTIAL RANDOM VARIABLES

In this section, we derive the general form of a weighted sum of order statistics of i.i.d. exponentially distributed random variables, and apply our results to obtain the PDF of the testing statistics under either hypothesis H_1 or H_0, respectively. For the exponential distribution, $Y_{(1)} \sim Exp\left(\frac{1}{N-K}\right), Y_{(i)} - Y_{(i-1)} \sim Exp\left(\frac{1}{N-K+1-i}\right)$ and $Y_{(1)}, Y_{(i)} - Y_{(i-1)}$ for $i = 2, 3, ..., N-K$ are independent [43, 61]. Define $Z_1 = Y_{(1)}$ and $Z_i = Y_{(i)} - Y_{(i-1)}$, therefore $Z = A\mathbf{Y}, \mathbf{Y} = A^{-1}Z$, where $A = \begin{pmatrix} 1 & 0 & 0 & . & . & 0 & 0 \\ -1 & 1 & 0 & . & . & 0 & 0 \\ 0 & -1 & 1 & . & . & .0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & -1 & 1 \end{pmatrix}$. Let $W = [W_1, W_2, ..., W_{N-K}]$, then $W\mathbf{Y} =$

$WA^{-1}Z$

$$X = \sum_{j=1}^{N-K} W_j Y_{(j)} = [V_1, V_2, ..., V_{N-K}] \begin{bmatrix} Z_1 \\ Z_2 \\ . \\ Z_{N-K} \end{bmatrix}$$
$$= \sum_{j=1}^{N-K} V_j Z_j,$$

which is a sum of independent exponentially distributed random variables $\left(\frac{1}{V_j(N-K+1-i)}\right)$, where $V = WA^{-1}$. Using Laplace operator,

$$\mathcal{L}\left(f_X(X)\right) = \prod_{j=1}^{N-K} \mathcal{L}(V_j Z_j),$$

the exact distribution $f_X\left(X\right)$ can be obtained by using partial fraction and applying the inverse Laplace transform. When $V_j(N-K+1-i)$ are different, the resulting pdf is a weighted sum of exponential distributions.

$$f_X(X) = \mathcal{L}^{-1}\left(\sum_{j=1}^{N-K}\left(\frac{\alpha_j}{s + \frac{V}{N-K-i+1}}\right)\right). \tag{B.1}$$

Under H_1, the smallest ordered sum is obtained by $W_1, W_2, ...W_K = 1$, and the factors $W_{K+1}, W_{K+2}, ..., W_{N-K} = 0$. Similarly, under H_0, the PDF of the largest ordered sum can be obtained by $W_1, W_2, ...W_K = 0$, and $W_{K+1}, W_{K+2}, ..., W_{N-K} = 1$.

APPENDIX C:

AN ALTERNATIVE DERIVATION OF SADDLE POINT SOLUTION

Notice that given X_1, p_1, and q, the posterior probability of the status of node 1 being malfunctioning is:

$$\begin{aligned} P\left(s=1|X_1=x_1\right) &= \frac{p\left(X_1|s=1\right)P\left(s=1\right)}{p\left(X_1|s=0\right)P\left(s=0\right)+p\left(X_1|s=0\right)P\left(s=0\right)} \\ &= \frac{\epsilon q_1\left(x_1\right)}{\left(1-\epsilon\right)p_1\left(x_1\right)+\epsilon q_1\left(x_1\right)} \end{aligned}$$

From the defination,

$$\begin{aligned} E\left(g|X_1=x_1\right) &= E_s E_{\theta|s}\left(g|X_1=x_1,s\right) \\ &= P\left(s=0|X_1=x_1\right)E_{\theta|s}\left(g|X_1=x_1,0\right) \\ &+ P\left(s=1|X_1=x_1\right)E_{\theta|s}\left(g|X_1=x_1,1\right) \\ &= E_{\theta|s}\left(g|x_1,0\right)+P\left(s=1|x_1\right)\left[E_{\theta|s}\left(g|x_1,1\right)-E_{\theta|s}\left(\theta|x_1,0\right)\right] \\ &= E_{p_1}\left(g|x_1\right)+\frac{\epsilon q_1\left(x_1\right)}{\left(1-\epsilon\right)p_1\left(x_1\right)+\epsilon q_1\left(x_1\right)}\left[E_{q_1}\left(g|x_1\right)-E_{p_1}\left(g|x_1\right)\right]. \end{aligned}$$

Hence, $q_1\left(x_1\right)=\frac{1-\epsilon}{\epsilon}p_1\left(x\right)\frac{\hat{g}_m-E_{p_1}(g|x_1)}{E_{q_1}(g|x_1)-\hat{g}_m}$.

Replacing $\hat{\theta}_m$ as in (4.13) and enforcing the domain condition on q_1 (4.16), one gets the identical final expression (4.19).

APPENDIX D: ESTIMATION IN BINOMIAL RANDOM VARIABLES (COMPLETE INFORMATIVE): TWO SENSOR

For the 2 sensor binary observation, lets denote $p(U = 11|\theta) = \theta_{11}$, $p(U = 01|\theta) = \theta_{01}$, $p(U = 10|\theta) = \theta_{10}$, $p(U = 00|\theta) = \theta_{00}$ with $\theta_{11} \geq \theta_{10} \geq \theta_{01} \geq \theta_{00}$. For this particular condition, $\theta_{11} = 1 - \theta_{00}$ and $\theta_{01} = \theta_{10}$ and the sensor unreliability is same for both sensor node, i.e. $\epsilon_1 = \epsilon_2 = \epsilon$. In the worst case scenario, $\theta_{11} = \theta_{10} = \theta_{01} = \theta_{00} = \frac{1}{2}$, decision would be discard all the observations. The attacker choice q to maximize the error would be to send

$$\begin{cases} 1 & \theta < \frac{1}{2} \\ 0 & \theta \geq \frac{1}{2}. \end{cases}$$

Therefore the MSE can be expressed as

$$\begin{aligned} E(d(\hat{\theta}, \theta)) &= \int_0^{\frac{1}{2}} E_{u_1,u_2|\theta}(\theta - \hat{\theta}(u_1, u_2))^2 d\theta + \int_{\frac{1}{2}}^1 E_{u_1,u_2|\theta}(\theta - \hat{\theta}(u_1, u_2))^2 d\theta \\ &= \int_0^{\frac{1}{2}} a^2(\theta_{11} - \theta)^2 d\theta + \int_0^{\frac{1}{2}} (1-a)^2(1 - \theta_{11} - \theta)^2 d\theta \\ &+ \int_{\frac{1}{2}}^1 a_1^2(\theta_{11} - \theta)^2 d\theta + \int_{\frac{1}{2}}^1 2a_1(1-a_1)(\frac{1}{2} - \theta)^2 d\theta + \int_{\frac{1}{2}}^1 (1-a_1)^2(1 - \theta_{11} - \theta)^2 d\theta, \end{aligned}$$

where $a = (1-\epsilon)\theta + \epsilon$ and $a_1 = (1-\epsilon)\theta$. The optimal estimate $\hat{\theta}_{11}$ can be obtained by minimizing MSE $\left(\frac{\partial E(d(\hat{\theta},\theta))}{\partial \theta_{11}} = 0\right)$,i.e,

$$\hat{\theta}_{11} = \frac{1}{2} + \frac{(2-5\epsilon)}{14\epsilon^2 - 10\epsilon + 8}$$

.

APPENDIX E: PERFORMANCE COMPARISON BENCHMARK FOR NOMINAL AND UNDER COMPROMISED CASE FOR TARGET LOCALIZATION

Under nominal and compromised cases, we will derive the performance comparison benchmark for above discussed robust localization algorithm. From eq:4.4

$$\hat{\beta} = (A^TWA)^{-1}A^TWb = (A^TWA)^{-1}A^TW(Ax+\epsilon)$$

Therefore error $\mathbf{e}$ $=\hat{x}-x = (A^TWA)^{-1}A^TW\epsilon$. The MSE is

$$E(\mathbf{e^T e}) = E(tr(\mathbf{e^T e}) = E\left[tr\left(A^TWA)^{-1}A^TW\epsilon\epsilon^TW^TA(A^TW^TA)^{-1}\right)\right]$$

Under the nominal condition $E(\epsilon\epsilon^T) = \sigma^2 I, W = diag(w_1, w_2, ...w_N)$. Therefore MSE$= \sigma^2 tr\left(A^TWA)^{-1}A^TWW^TA(A^TW^TA)^{-1}\right)$. One can optimize W to get minimum MSE. When W=I, then MSE$= tr(A^TA)^{-1}$.

Under the model departure, the $E(\epsilon\epsilon^T)$ will change. For example if k out of N transmitter are compromised by new distribution of error. Let us consider now $\epsilon_1, ..., \epsilon_N$

are still independent and node 1 is anchored but 1 out of $N-1$ is compromised by $N(0,\sigma_1^2)$, instead of σ^2, then $E(\epsilon_1\epsilon_j)=\sigma^2\delta_{1j}$, $E(\epsilon_j^2)=\frac{N-2}{N-1}\sigma^2+\frac{1}{N-1}\sigma_1^2$, $E(\epsilon_i\epsilon_j)=0$, hence

$$\begin{aligned}E(\epsilon\epsilon^T) &= diag\left\{\sigma^2,\sigma^2+\frac{1}{N-1}(\sigma_1^2-\sigma^2),....,\frac{1}{N-1}(\sigma_1^2-\sigma^2)\right\}\\ &= \sigma^2 I+\frac{(\sigma_1^2-\sigma^2)}{N-1}diag\left\{0,1,...,1\right\}.\end{aligned}$$

Inserting $E(\epsilon\epsilon^T)$ in eq:3-1 the MSE is the function of W which can be optimized for minimum MSE.