Resilient Detection in the Presence of Integrity Attacks

Yilin Mo*, João Hespanha†, Bruno Sinopoli*

***Abstract*—We consider the detection of a binary random state based on m measurements that can be manipulated by an attacker. The attacker is assumed to have full information about the true value of the state to be estimated as well as the values of all the measurements. However, the attacker can only manipulate n of the m measurements. The detection problem is formulated as a minimax optimization, where one seeks to construct an optimal detector that minimizes the "worst-case" probability of error against all possible manipulations by the attacker. We show that if the attacker can manipulate at least half the measurements ($n \geq m/2$) then the optimal worst-case detector should ignore *all* m measurements and be based solely on the a-priori information. When the attacker can manipulate less than half of the measurements ($n < m/2$), we show that the optimal detector is a threshold rule based on a Hamming-like distance between the (manipulated) measurement vector and two appropriately defined sets. For the special case where $n = (m-1)/2$, our results provide a constructive procedure to derive the optimal detector. We design a heuristic detector for the case where $n \ll m$, and prove the asymptotic optimality of the detector when $m \to \infty$. Finally we apply the proposed methodology in the case of i.i.d. Gaussian measurements.**

I. Introduction

The increasing use of networked embedded sensors to monitor and control critical infrastructures provides potential malicious agents with the opportunity to disrupt their operations by corrupting sensor measurements. Supervisory Control And Data Acquisition (SCADA) systems, for example, run a wide range of safety critical plants and processes, including manufacturing, water and gas treatment and distribution, facility control and power grids. A successful attack to such kind of systems may significantly hamper the economy, the environment, and may even lead to the loss of human life. The first-ever SCADA system malware (called Stuxnet) was found in July 2010 and rose significant concern about SCADA system security [1], [2]. While most SCADA systems are currently running on dedicated networks, next generation SCADA will make extensive use of widespread sensing and networking, both wired and wireless, making critical infrastructures susceptible to cyber security threats. The research community has acknowledged the importance of addressing the challenge of designing secure detection, estimation and control systems [3].

We consider a robust detection problem inspired by security concerns that arise from the possible manipulation of sensor data. We focus our attention on the detection of a binary random variable θ from independent measurements collected by m sensors, with the caveat that some of these measurements can be manipulated by an attacker. The attacker is assumed to have full information about the true value of θ and all the measurements and uses this information to manipulate the data available to the detector. We assume that the attacker has total control over n corrupted sensors, where $n \leq m$, and he can change their values arbitrarily. To minimize the detector's performance degradation in the face of such attacks, we construct minimax detectors that minimize the "worst-case" probability of detection error, where worst-case refers to all possible manipulations available to the attacker.

We want to analyze the detector design problem for all the cases where $n \leq m$. We start by considering the case $n \geq m/2$, in which the attacker can manipulate at least half ofthe measurements. We show that in this scenario the optimal worst-case detector should ignore *all* m measurements and be based solely on the a-priori distribution of θ. This result is in sharp contrast with non-adversarial detection theory where even very noisy data can provide some information. This also highlights the power of adversarial manipulation of sensor data since an attacker that has the ability to manipulate only half of the sensors, effectively destroys all the information that can be inferred from the full set of sensors.

For the case $n < m/2$, in which the attacker can manipulate strictly less than half of the sensors, the optimal estimator typically depends on the sensor data. Moreover, we show that the optimal estimator consists of a threshold rule that compares a Hamming-like distance between the (manipulated) measurement vector and two appropriately defined sets. In general, these sets may be difficult to compute. In the boundary case $n = (m-1)/2$ we provide a procedure to construct the optimal estimator, which turns out to be a simple voting scheme. If the percentage of compromised sensors are small, i.e., $n \ll m$, we designed a heuristic detector based on truncated sum, which achieves asymptotic optimality when m goes to ∞. The proposed methodology is then applied for the case of Gaussian i.i.d. sensors measurements (prior to the adversarial manipulation).

Related Work

Minimax robust detection problems have been extensively studied in the past decades [4]–[8]. A classical approach assumes that the conditional distribution of sensor measurements under each hypothesis lies in a set of probability distributions, which is called an uncertainty class. One then identifies a pair of "least favorable distributions" (LFDs) from

This research is supported in part by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the Army Research Office Foundation and grant NGIT2009100109 from Northrop Grumman Information Technology Inc Cybersecurity Consortium. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of ARO, CMU, or the U.S. Government or any of its agencies.

*: Yilin Mo and Bruno Sinopoli are with the ECE department of Carnegie Mellon University, Pittsburgh, PA. Email: ymo@andrew.cmu.edu, brunos@ece.cmu.edu

†: João Hespanha is with the ECE department of University of California, Santa Barbara, CA. Email: hespanha@ece.ucsb.edu

the uncertainty class, which conceptually represents the most similar and hardest to distinguish pair of distributions. The robust detector is then designed as a naive-Bayes or Neymann-Pearson detector between the LFDs. The main difficulty in applying the LFD-based method to our scenario is that there is no systematic procedure to construct the LFDs and hence the corresponding detector. As a result, in this paper, we attempt to directly compute the optimal detector instead of seeking LFDs.

Basar et al. [9], [10] consider the problem of transmitting and decoding Gaussian signals over a communication channel with unknown input from a so-called "jammer". The unknown input is assumed to be mean square bounded by a constant, which depends upon the capability of the "jammer". Although this set-up is reasonable for analog communications where the attacker is energy-constrained, it is not practical for cyber attacks on digital communications, where the attacker can change the data arbitrarily when the integrity of the sensor is compromised.

Bayram and Gezici [11] propose a restricted Neyman-Pearson approach for composite hypothesis-testing in the presence of uncertainty in the prior probability distribution. Mutapcic and Kim [12] consider the problem of detecting two Gaussian signals, where the mean and covariance of the signal are uncertain. They prove that the robust linear detector design problem can be formulated as a convex optimization problem. In [13], [14], the authors consider the problem of detecting the presence of a signal with low Signal to Noise Ratio (SNR). The authors prove that there exists an "SNR wall", below which a detector fails to be robust with respect to the uncertainties in the fading and noise model. However, these robustness results cannot be directly applied to the secure detector design problem, as their uncertainty models are in general quite different from a cyber attack model.

The rest of paper is organized as follows. In Section II we formulate the problem of robust detection with n manipulated measurements from m total measurements. In Section III and IV, we consider the optimal detector design for the cases $n \geq m/2$ and $n < m/2$ respectively. In Section V we discuss a special case where $n = (m-1)/2$, formulate the problem of optimal detector design and provide a closed form solution. In Section VI we propose a heuristic detector design and prove its asymptotic optimality. In Section VII we provide a numerical example of i.i.d. Gaussian signals. Section VIII finally concludes the paper.

II. Problem Formulation

The goal is to detect a binary random variable (r.v.) θ with distribution

$$\theta = \begin{cases} -1 & \text{w.p. } p^- \\ +1 & \text{w.p. } p^+ \end{cases}$$

where $p^-, p^+ \geq 0$ and $p^- + p^+ = 1$. Without loss of generality, we assume that $p^+ \geq p^-$. To detect θ we have available a vector $y \triangleq [y_1, \ldots, y_m]' \in \mathbb{R}^m$ of m sensor measurements $y_i \in \mathbb{R}$, $i \in \{1, 2, ..., m\}$, each of which is conditionally independent from the others given θ. Let us assume that the probability measure generated by random variable y_i is ν_i when $\theta = -1$ and μ_i when $\theta = 1$. In other words, for any Borel-measurable set S, the following holds:

$$\nu_i(S) = P(y_i \in S|\theta = -1), \mu_i(S) = P(y_i \in S|\theta = 1).$$

Moreover, let us define the product measure

$$\nu \triangleq \nu_1 \times \ldots \times \nu_m,\ \mu \triangleq \mu_1 \times \ldots \times \mu_m.$$

Let us define the inner measures induced by ν, μ as $\underline{\nu}$, $\underline{\mu}$ respectively. Therefore, for an arbitrary set $W \subseteq \mathbb{R}^m$ (not necessarily Borel-measurable),

$$\underline{\nu}(W) \triangleq \sup\{\nu(S) : S \in \mathcal{B}(\mathbb{R}^m), S \subseteq W\},$$
$$\underline{\mu}(W) \triangleq \sup\{\mu(S) : S \in \mathcal{B}(\mathbb{R}^m), S \subseteq W\},$$

where $\mathcal{B}(\mathbb{R}^m)$ is the Borel-algebra on $\mathbb{R}^m$. We further assume that measures ν_i and μ_i are absolutely continuous with respect to each other for any $1 \leq i \leq m$. Hence the log-likelihood ratio $\Lambda_i : \mathbb{R} \to \mathbb{R}$ of y_i is well defined as

$$\Lambda_i(y_i) \triangleq \log\left(\frac{d\mu_i}{d\nu_i}\right),$$

where $d\mu_i/d\nu_i$ is the Radon-Nikodym derivative. Further we define the log-likelihood ratio of y as

$$\Lambda(y) \triangleq \sum_i \Lambda_i(y_i) = \log\left(\frac{d\mu}{d\nu}\right).$$

We assume that an attacker wants to increase the probability that we make an error in detecting θ. To this end, the attacker has the ability to manipulate n of the m sensor measurements, but we do not know which n of the m measurements have been manipulated. Formally, this means that our estimate of θ has to rely on a vector $y' \in \mathbb{R}^m$ of *manipulated measurements* defined by

$$y' = y + \gamma \circ u, \tag{1}$$

where the attacker chooses the *sensor-selection* vector γ taking values in

$$\Gamma \triangleq \{\gamma \in \mathbb{R}^m : \gamma_i = 0 \text{ } or \text{ } 1, \sum_{i=1}^m \gamma_i \leq n\}$$

and the *bias* vector u taking values in $\mathbb{R}^m$. The $\circ$ in (1) denotes entry-wise multiplication of two vectors. By selection which entries of γ are nonzero, the attacker chooses which of the n sensors will be manipulated. The "magnitude" of manipulation is determined by u.

The detection problem is formalized as a minimax problem where one wants to select an optimal detector

$$\hat{\theta} = f(y') = f(y + \gamma \circ u) \tag{2}$$

so as to minimize the probability of error, for the worst case manipulation by the adversary. Following Kerckhoffs' Principle [15] that security should not rely on the obscurity of the system, our goal is to design the detector $f : \mathbb{R}^m \to \{-1, 1\}$ assuming that f is known to the attacker. We also take the conservative approach that the attacker has full information about the state of the system. Namely, the underlying θ and all the measurements $y_1, \ldots, y_m$ are assumed to be known to the attacker. In addition the attacker can manipulate up to n of the m sensors. We assume that the defender knows how many sensors may be compromised, but cannot identify them. Our goal is to analyze the problem for different values of n, ranging from 1 to m.

Remark 1: The parameter n can be interpreted as a design parameter for the defender. In general, increasing n will increase the resilience of the detector under attack. However, as is shown in the rest of the paper, a large n will result in performance degradation during normal operation when no sensor is compromised. Therefore, there exists a trade-off between resilience and efficiency (under normal operation), which can be tuned by choosing a suitable parameter n.

To compute the worst-case probability of error that we seek to minimize, we consider given values of θ, y and an detector f, for which an optimal policy for the attacker can be written as follows:

$$(u,\gamma)=\begin{cases}\displaystyle\mathop{\arg\min}_{u\in\mathbb{R}^m,\gamma\in\Gamma} f(y+\gamma\circ u) & \theta=1\\ \displaystyle\mathop{\arg\max}_{u\in\mathbb{R}^m,\gamma\in\Gamma} f(y+\gamma\circ u) & \theta=-1,\end{cases}$$

where the selection of the manipulation pair (u,γ) tries to get $\hat{\theta}$ in (2) as low as possible when $\theta=1$ (ideally as low as -1) or as high as possible when $\theta=-1$ (ideally as high as 1). The min and max are attainable since f only takes ± 1.

Under this worst-case attacker policy, a correct decision will be made only when the pair (θ,y) belongs to the set

$$\Big\{(-1,y):y\in Y^-(f)\Big\}\cup\Big\{(+1,y):y\in Y^+(f)\Big\} \tag{3}$$

where $Y^+(f)$ and $Y^-(f)$ denote the set of measurement values $y\in\mathbb{R}^m$ for which the attacker cannot force the estimate to be -1 and $+1$, respectively, i.e.,

$$Y^+(f)\triangleq\big\{y\in\mathbb{R}^m: f(y+\gamma\circ u)=1,\ \forall u\in\mathbb{R}^m,\gamma\in\Gamma\big\},$$
$$Y^-(f)\triangleq\big\{y\in\mathbb{R}^m: f(y+\gamma\circ u)=-1,\ \forall u\in\mathbb{R}^m,\gamma\in\Gamma\big\}.$$

For a given detector f, the worst-case probability of error $P_e(f)$ is then given by the measure of the set defined in (3) and can be expressed as

$$\begin{aligned}P_e(f)&\triangleq(1-\beta(f))P(\theta=1)+\alpha(f)P(\theta=-1)\\&=(1-\beta(f))p^+ +\alpha(f)p^-,\end{aligned} \tag{4}$$

where the false alarm rate $\alpha(f)$ and the probability of detection $\beta(f)$ are defined as

$$\alpha(f)\triangleq 1-\underline{\nu}(Y^-(f)),\ \beta(f)\triangleq\underline{\mu}(Y^+(f)).$$

One should think of $\alpha(f)$ as the measure of the set $\mathbb{R}^m\setminus Y^-(f)$ conditioned to $\theta=-1$ and of $\beta(f)$ as the measure of the set $Y^+(f)$ conditioned to $\theta=+1$. The use of inner measures ensures that P_e is well defined even if these sets are not measurable[1].

Formally, the problem under consideration is to determine the optimal detector f in (2) that minimizes the worst-case probability of error in (4):

$$P_e^*=\inf_f P_e(f).$$

From the discussion above, we can recognize $Y^+(f)$ and $Y^-(f)$ as "good" sets for the detector, in the sense that when measurements fall in these sets the attacker cannot induce errors. From this perspective, good detection policies obviously correspond to these sets being large. This statement is formalized, without proof, in the following lemma:

Lemma 1: Given two functions $f,g:\mathbb{R}^m\to\{-1,1\}$, if $Y^+(g)\supseteq Y^+(f)$ and $Y^-(g)\supseteq Y^-(f)$, then $P_e(g)\le P_e(f)$.

Next section will illustrate the case where half or more of sensors are compromised.

III. Optimal Detector Design for $n\ge m/2$

In this section we consider the case when half or more of the measurements can be manipulated by the attacker. We show that, in this case, the attacker can render the information provided by the manipulated measurement vector y useless, forcing the optimal estimate to be determined exclusively from the a-priori distribution of θ.

Theorem 1: If $n\ge m/2$ then an optimal f_* is given by[2]

$$f_*(y)=1,\quad \forall y\in\mathbb{R}^m,$$

and the corresponding probability of error P_e and sets Y^+ and Y^- are given by

$$P_e(f_*)=p^-,\quad Y^+(f_*)=\mathbb{R}^m,\quad Y^-(f_*)=\emptyset.$$

The following lemma characterizes the relationship between $Y^-(f)$ and $Y^+(f)$ when $n\ge m/2$ and provides a key technical result needed to prove Theorem 1.

Lemma 2: If $n\ge m/2$, then $Y^-(f)\ne\emptyset$ implies that $Y^+(f)=\emptyset$.

Proof of Lemma 2: First note that in this case $m-n\le n$. Assuming by contradiction that neither $Y^+(f)$ nor $Y^-(f)$ is empty. As a result, there exist two measurement vectors

$$y^+=[y_1^+,\dots,y_m^+]'\in Y^+(f),$$
$$y^-=[y_1^-,\dots,y_m^-]'\in Y^-(f).$$

Now let us construct another vector

$$y=[y_1^+,\dots,y_n^+,y_{n+1}^-,\dots,y_m^-]'.$$

Thus,

$$y=y^+ +\gamma_1\circ(y^- -y^+),$$

where

$$\gamma_1=[\underbrace{0,\dots,0}_{n},\underbrace{1,\dots,1}_{m-n}]'.$$

Since γ_1 has $m-n\le n$ nonzero entries (recall that $n\ge m/2$), this vector belongs to Γ. From the facts that $y^+\in Y^+(f)$ and $\gamma_1\in\Gamma$, we conclude from the definition of $Y^+(f)$ that $f(y)=1$. On the other hand,

$$y=y^- +\gamma_2\circ(y^+ -y^-),$$

where

$$\gamma_2=[\underbrace{1,\dots,1}_{n},\underbrace{0,\dots,0}_{m-n}]'.$$

[1] Even if the original function f is measurable, the sets Y^+ and Y^- may not be necessarily measurable.

[2] The optimal detector is not necessarily unique.

Since γ_2 has n nonzero entries, this vector also belongs to Γ. From the facts that $y^- \in Y^-(f)$ and $\gamma_2 \in \Gamma$, we can also conclude from the definition of $Y^-(f)$ that $f(y) = -1$, which contradicts our previous assertion about $f(y)$. ■

Proof of Theorem 1: By Lemma 2, we know that either $Y^+(f)$ or $Y^-(f)$ must be empty. First suppose that $Y^-(f)$ is empty and hence $\alpha(f) = 1$. As a result

$$P_e(f) = p^+(1 - \beta(f)) + p^-.$$

The minimum is achieved when $Y^+(f) = \mathbb{R}^m$, which implies that $f = 1$ and $P_e(f) = p^-$.
On the other hand, if $Y^+(f)$ is empty, then the optimal $Y^-(f) = \mathbb{R}^m$, $f = -1$ and $P_e(f) = p^+$. Since we assume that $p^+ \geq p^-$, the optimal f is $f_* = 1$ and optimal sets are $Y^+(f_*) = \mathbb{R}^m$ and $Y^-(f_*) = \emptyset$. ■

IV. Optimal Detector Design for $n < m/2$

We now consider the case when less than half of the measurements can be manipulated by the attacker, i.e., $n < m/2$. We show that the optimal detector is a threshold rule based on a Hamming-like distance between the (manipulated) measurement vector and two appropriately defined sets.

By Lemma 1, to find the optimal f_*, we should maximize the "volume" of both $Y^-(f)$ and $Y^+(f)$. However, it is easy to see that there is a trade-off between $Y^-(f)$ and $Y^+(f)$. In other words, expanding one set usually results in shrinking the other set. To characterize the exact trade-off between $Y^-(f)$ and $Y^+(f)$, we need to introduce the following notation: We denote by $d : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{N}_0$ the metric induced by the "zero-norm," i.e.,

$$d(x, y) \triangleq \|x - y\|_0,$$

where $\|x\|_0$ is the "zero-norm" of x, which is defined as the number of non-zero entries of the vector x. While the "zero-norm" is not a norm, it is easy to verify that the function d defined above is a metric. In fact, d can be viewed as an extension of the Hamming distance to continuous-valued vectors. The metric d can be generalized to sets in the usual way: given an element x and two subsets X, Y of $\mathbb{R}^m$, we define

$$d(X, Y) \triangleq \min_{x \in X,\, y \in Y} d(x, y) \qquad d(x, Y) \triangleq d(\{x\}, Y). \tag{5}$$

For convenience, we define the distance from any set to the empty set to be infinity: $d(X, \emptyset) = \infty$. The minimum in (5) is always attainable since d takes only integer values.

We also need to introduce a "truncation function": Given an *indexed subset* $\mathcal{I} = \{i_1, i_2, \dots, i_j\}$ of $\{1, 2, ..., m\}$, we define the function $\mathrm{Trunc}_{\mathcal{I}} : \mathbb{R}^m \to \mathbb{R}^{|\mathcal{I}|}$ by

$$\mathrm{Trunc}_{\mathcal{I}}(y) = \begin{bmatrix} y_{i_1} & y_{i_2} & \cdots & y_{i_j} \end{bmatrix}'.$$

For a given set $X \subset \mathbb{R}^m$ and $\mathcal{I} \subset \{1, \dots, m\}$. We can identify the truncated set $X_{\mathcal{I}}$ as

$$X_{\mathcal{I}} = \{y \in \mathbb{R}^{|\mathcal{I}|} : \exists\, x \in X, \text{ such that } y = \mathrm{Trunc}_{\mathcal{I}}(x)\}.$$

On the contrary, suppose that for each indexed subset $\mathcal{I} \subset \{1, \dots, m\}$ of size $m - 2n$ we have available a set $X_{\mathcal{I}} \subseteq \mathbb{R}^{m-2n}$. We want to find the set $X \subseteq \mathbb{R}^m$ such that X is the inverse image of $X_{\mathcal{I}}$ under $\mathrm{Trunc}_{\mathcal{I}}$, for each $\mathcal{I}$. It is easy to see that X can be defined in the following way:

$$X \triangleq \{y \in \mathbb{R}^m : \mathrm{Trunc}_{\mathcal{I}}(y) \in X_{\mathcal{I}}, \forall |\mathcal{I}| = m - 2n\}. \tag{6}$$

We define the class of such a set X parameterized by $X_{\mathcal{I}}$s as $\mathcal{X}_{m,n}$.

Definition 1: Two sets $X^+, X^- \in \mathcal{X}_{m,n}$ are called mutually exclusive if and only if

$$X^+ \triangleq \{y \in \mathbb{R}^m : \mathrm{Trunc}_{\mathcal{I}}(y) \in X_{\mathcal{I}}, \forall |\mathcal{I}| = m - 2n\},$$
$$X^- \triangleq \{y \in \mathbb{R}^m : \mathrm{Trunc}_{\mathcal{I}}(y) \in \mathbb{R}^{m-2n} \backslash X_{\mathcal{I}}, \forall |\mathcal{I}| = m - 2n\},$$

for some $X_{\mathcal{I}}$s.

Example: Let $m = 3$ and $n = 1$, then the following two sets are mutually exclusive:

$$X^+ = \{y \in \mathbb{R}^3 : y_i > 0, \forall i = 1, 2, 3\},$$
$$X^- = \{y \in \mathbb{R}^3 : y_i \leq 0, \forall i = 1, 2, 3\},$$

with $X_{\{i\}} = \mathbb{R}^+$. Furthermore, it is worth noticing that the union of set X^+ and X^- is not the entire space in general.

It turns out to be that the concept of mutually exclusive sets provides the exact characterization of the trade-off between $Y^-(f)$ and $Y^+(f)$, which is illustrated by the following theorems:

Theorem 2: For any detector f, there exists a pair of mutually exclusive sets $\mathcal{Y}^-(f)$, $\mathcal{Y}^+(f) \in \mathcal{X}_{m,n}$, such that

$$Y^-(f) \subseteq \mathcal{Y}^-(f), Y^+(f) \subseteq \mathcal{Y}^+(f). \tag{7}$$

Theorem 3: For any pair of mutually exclusive sets X^-, $X^+ \in \mathcal{X}_{m,n}$, there exists a detector f, defined as

$$f(y) = \begin{cases} 1 & d(y, X^-) \geq d(y, X^+) \\ -1 & d(y, X^-) < d(y, X^+), \end{cases} \tag{8}$$

for which the following inequalities hold:

$$X^+ \subseteq Y^+(f),\; X^- \subseteq Y^-(f). \tag{9}$$

Combining Theorem 2 and Theorem 3, we have the following corollary, which casts the design of the optimal detector as an optimization problem over pairs of mutually exclusive sets:

Corollary 1: An optimal detector f_* is of the following form:

$$f(y) = \begin{cases} 1 & d(y, X_*^-) \geq d(y, X_*^+) \\ -1 & d(y, X_*^-) < d(y, X_*^+), \end{cases} \tag{10}$$

where X_*^+ and X_*^- are the solutions of the following optimization problem:

$$\begin{aligned} \underset{X^+, X^-}{\text{minimize}} \quad & (1 - \underline{\mu}(X^+))p^+ + (1 - \underline{\nu}(X^-))p^- \\ \text{subject to} \quad & X^+,\, X^- \in \mathcal{X}_{m,n}, \\ & X^+,\, X^- \text{ are mutually exclusive.} \end{aligned} \tag{11}$$

Proof: Suppose that X_*^+ and X_*^- are the optimal solutions. By Theorem 3, we know that

$$\begin{aligned} P_e(f_*) &= (1 - \underline{\mu}(Y^+(f_*)))p^+ + (1 - \underline{\nu}(Y^-(f_*)))p^- \\ &\leq (1 - \underline{\mu}(X_*^+))p^+ + (1 - \underline{\nu}(X_*^-))p^-. \end{aligned}$$

Now pick an arbitrary detector f. By Theorem 2, there exists a pair of mutually exclusive sets $\mathcal{Y}^+(f)$ and $\mathcal{Y}^-(f)$ such that

$$\begin{aligned}P_e(f) &= (1-\underline{\mu}(Y^+(f)))p^+ + (1-\underline{\nu}(Y^-(f)))p^- \\ &\geq (1-\underline{\mu}(\mathcal{Y}^+(f)))p^+ + (1-\underline{\nu}(\mathcal{Y}^-(f)))p^-.\end{aligned}$$

Since X_*^+ and X_*^- are the optimal solutions for (11), we have

$$\begin{aligned}&(1-\underline{\mu}(\mathcal{Y}^+(f)))p^+ + (1-\underline{\nu}(\mathcal{Y}^-(f)))p^- \\ &\qquad \geq (1-\underline{\mu}(X_*^+))p^+ + (1-\underline{\nu}(X_*^-))p^-,\end{aligned}$$

which concludes the proof. ■

Remark 2: The key challenge in directly applying Corollary 1 is that it does not provide a method to construct the sets X^+, X^- that lead to the optimal f_*, potentially requiring one to search for the optimal detector by ranging over all possible pairs of mutually exclusive sets in $\mathcal{X}_{m,n}$. In general this result does not yield a computationally viable way to compute the optimal detector, excluding a special case, which is discussed in Section V. Furthermore, even if we could find the optimal X_*^+, X_*^-, it is possible that $d(y, X_*^+)$ and $d(y, X_*^-)$ could be numerically unstable and expensive to compute.

However, Corollary 1 provides a general guideline for designing detector in adversarial environments, as it effectively reduces the search space of optimal f from all possible functions to the functions of the special form (10). In fact we shall see in Section V that we can use this general result to find the optimal detector for the case $m = 2n + 1$, as the computation of the sets X^+, X^- becomes trivial. In Section VI, we propose the design of a heuristic detector of form (10) for the general case and prove that our design is asymptotically optimal when n is fixed and m goes to infinity. Both of these detectors can be efficiently computed.

The remainder of this section is mostly devoted to the proof of Theorem 2 and 3.

A. Proof of Theorem 2

For a given set $\mathcal{I}$ and detector f, in the sequel we denote by $Y_\mathcal{I}^-(f)$ and $Y_\mathcal{I}^+(f)$ the image of $Y^-(f)$ and $Y^+(f)$, respectively, under the function $\mathrm{Trunc}_\mathcal{I}$. As stated in the following result, it turns out that these sets are always disjoint:

Lemma 3: if $n < m/2$, for every detector f and index subset $\mathcal{I}$ of cardinality $|\mathcal{I}| = m - 2n$, we have

$$Y_\mathcal{I}^-(f) \bigcap Y_\mathcal{I}^+(f) = \emptyset.$$

Proof of Lemma 3: We prove the statement by contradiction. Without loss of generality, we assume that $\mathcal{I} = \{1, \dots, m-2n\}$, and

$$Y_\mathcal{I}^- \bigcap Y_\mathcal{I}^+ \neq \emptyset.$$

As a result, there exist two vectors

$$\begin{aligned}y^+ &= [y_1, \dots, y_{m-2n}, y_{m-2n+1}^+, \dots, y_m^+]' \in Y^+, \\ y^- &= [y_1, \dots, y_{m-2n}, y_{m-2n+1}^-, \dots, y_m^-]' \in Y^-.\end{aligned}$$

Now let us construct another vector

$$y = [y_1, \dots, y_{m-2n}, y_{m-2n+1}^+ \dots, y_{m-n}^+, y_{m-n+1}^-, \dots, y_m^-]'.$$

It can be easily seen that there at most are n elements in y that differ from $y^+ \in Y^+$. As a result, $f(y) = 1$ from the definition of Y^+. However, there are also at most n elements in y that differ from $y^- \in Y^-$ which implies that $f(y) = -1$ from the definition of Y^-, leading to a contradiction. ■

Now we are ready to prove Theorem 2:

Proof of Theorem 2: For every detector f, it is easy to see that

$$\begin{aligned}&Y^-(f) \subseteq \mathcal{Y}^-(f) \\ &\qquad \triangleq \{y \in \mathbb{R}^m : \mathrm{Trunc}_\mathcal{I}(y) \in Y_\mathcal{I}^-(f), \forall |\mathcal{I}| = m - 2n\},\end{aligned}$$

By Lemma 3,

$$Y_\mathcal{I}^+(f) \bigcap Y_\mathcal{I}^-(f) = \emptyset,$$

which implies that $Y_\mathcal{I}^+(f) \subseteq R^{m-2n} \backslash Y_\mathcal{I}^-(f)$ and therefore, $Y^+(f)$ is upper bounded by

$$\begin{aligned}&Y^+(f) \subseteq \mathcal{Y}^+(f) \\ &\triangleq \{y \in \mathbb{R}^m : \mathrm{Trunc}_\mathcal{I}(y) \in \mathbb{R}^{m-2n} \backslash Y_\mathcal{I}^-(f), \forall |\mathcal{I}| = m - 2n\}.\end{aligned}$$

Therefore, $\mathcal{Y}^-(f), \mathcal{Y}^+(f) \in \mathcal{X}_{m,n}$ are mutually exclusive. ■

B. Proof of Theorem 3

First let us prove an important inequality on the distance between any pair of mutually exclusive sets:

Lemma 4: For any pair of mutually exclusive sets X^-, $X^+ \in \mathcal{X}_{m,n}$,

$$d(X^-, X^+) \geq 2n + 1.$$

Proof of Lemma 4: Without loss of generality, we assume that X^- and X^+ are not empty. By contradiction, assume that there exist $y^- \in X^-$, $y^+ \in X^+$ for which $d(y^-, y^+) \leq 2n$, which implies that y^- and y^+ share at least $m - 2n$ equal elements. As a result, there exists $\mathcal{I} = \{i_1, \dots, i_{m-2n}\}$, such that $y_i^- = y_i^+$, $\forall i \in \mathcal{I}$. Therefore, $\mathrm{Trunc}_\mathcal{I}(y^-) = \mathrm{Trunc}_\mathcal{I}(y^+)$. Thus,

$$\mathrm{Trunc}_\mathcal{I}(X^-) \bigcap \mathrm{Trunc}_\mathcal{I}(X^+) \neq \emptyset,$$

which contradicts with the fact that

$$\mathrm{Trunc}_\mathcal{I}(X^-) \subseteq X_\mathcal{I},\ \mathrm{Trunc}_\mathcal{I}(X^+) \subseteq \mathbb{R}^{m-2n} \backslash X_\mathcal{I}.$$

■

The following lemma characterizes the distance between an arbitrary vector y and the sets X^- and X^+.

Lemma 5: For any pair of mutually exclusive sets X^-, $X^+ \in \mathcal{X}_{m,n}$ and any vector $y \in \mathbb{R}^m$,

$$d(y, X^-) + d(y, X^+) \geq 2n + 1.$$

Proof of Lemma 5: Without loss of generality, we assume that X^- and X^+ are not empty. Since the minimum in (5) is attainable, there exist $y^- \in X^-$ and $y^+ \in X^+$, such that

$$d(y, X^-) = d(y, y^-),\ d(y, X^+) = d(y, y^+).$$

By Lemma 4,

$$d(y^-, y^+) \geq d(X^-, X^+) \geq 2n + 1,$$

and therefore, using the triangle inequality, we conclude that

$$\begin{aligned} d(y,X^-)+d(y,X^+) &= d(y,y^-)+d(y,y^+) \\ &\ge d(y^-,y^+) \ge 2n+1. \end{aligned}$$

■

We are now ready to prove Theorem 3:

Proof of Theorem 3: We first prove that

$$X^- \subseteq Y^-(f).$$

Consider an arbitrary vector $y \in X^-$. To show that $y \in Y^-(f)$, we need to prove that $f(y+\gamma\circ u) = -1$, $\forall u \in \mathbb{R}^m, \gamma \in \Gamma$. Since $y \in X^-$,

$$d(y+\gamma\circ u, X^-) \le d(y+\gamma\circ u, y) = \|\gamma\circ u\|_0 \le n,$$

and therefore, because of Lemma 5, we must have that

$$d(y+\gamma\circ u, X^+) \ge 2n+1-d(y+\gamma\circ u, X^-) \ge n+1.$$

Consequently $d(y+\gamma\circ u, X^-) < d(y+\gamma\circ u, X^+)$ and we indeed have that $f(y+\gamma\circ u) = -1$. The proof that $X^+ \subseteq Y^+(f)$ follows similar steps. ■

V. Optimal Detector for $n = (m-1)/2$

In this section, we construct the optimal detector for the case where $n = (m-1)/2$. From Corollary 1, we know that the optimal detector can be constructed by choosing an ‘appropriate" family of sets $X_{\mathcal{I}}$. It turns out that when $n = (m-1)/2$ this family of sets has a particularly simple structure.

Theorem 4: If $m - 2n = 1$, the family of sets $X_{\{i\}}$ that gives the optimal detector f_* in (10) is of the form

$$X_{\{i\}} = T_i(\eta_i), \quad \forall i \in \{1,2,...,m\} \tag{12}$$

where $\eta_i \in \mathbb{R}$,

$$T_i(\eta_i) \triangleq \Big\{ y_i \in \mathbb{R} : \Lambda_i(y_i) < \eta_i \Big\},$$

where $\Lambda_i : \mathbb{R} \to \mathbb{R}$ is the log-likelihood ratio of the distribution of the ith measurement y_i. By convention, $T_i(\infty) = \mathbb{R}$ and $T_i(-\infty) = \emptyset$[3].

Before proving Theorem 4, we note that one can implement the optimal detector in (10) without actually computing $d(y,X^-)$ and $d(y,X^+)$. When either X^+ or X^- is empty, then one of the distances in (10) is $+\infty$ and f_* is simply a constant. When none of these sets is empty, it is straightforward to show that

$$d(y,X^-) = \big|\{i : y_i \in X_{\{i\}}\}\big|, \quad d(y,X^+) = \big|\{i : y_i \notin X_{\{i\}}\}\big|,$$

where $|\cdot|$ is the number of elements in a set. The detection algorithm can be implemented as the following voting process:

- The detector computes m individual estimates $\hat{\theta}_i$ by a Neymann-Pearson detector based on individual (possibly manipulated) measurements y'_i:
$$\hat{\theta}_i \triangleq \begin{cases} -1 & \Lambda_i(y'_i) < \eta_i \\ 1 & \Lambda_i(y'_i) \ge \eta_i \end{cases}.$$
- The optimal estimate $\hat{\theta}$ is obtained by voting:
$$\hat{\theta} = \begin{cases} -1 & \text{at least } n+1 \text{ sensors estimate } \hat{\theta}_i = -1 \\ +1 & \text{less than } n+1 \text{ sensors estimate } \hat{\theta}_i = -1 \end{cases}$$

A. Proof of Theorem 4

We start by noting that when $m - 2n = 1$ the sets X^- and X^+ are especially simple to compute:

$$\begin{aligned} X^- &= \big\{ y \in \mathbb{R}^m : \mathrm{Trunc}_i(y) \in X_{\{i\}}, \forall i \big\} = \prod_{i=1}^m X_{\{i\}} \\ X^+ &= \big\{ y \in \mathbb{R}^m : \mathrm{Trunc}_i(y) \in \mathbb{R}\backslash X_{\{i\}}, \forall i \big\} \\ &= \prod_{i=1}^m \mathbb{R}\backslash X_{\{i\}}, \end{aligned}$$

where $\prod_i X_{\{i\}}$ is the Cartesian product. The following result is a straightforward consequence of the fact that X^- and X^+ can be written as Cartesian products:

Lemma 6: If $X^+ \neq \emptyset$ and $X^- \neq \emptyset$, then $\mathrm{Trunc}_i(X^-) = X_{\{i\}}$ and $\mathrm{Trunc}_i(X^+) = \mathbb{R}\backslash X_{\{i\}}$.

The following result essentially states that the inner measure of Cartesian products is the product of inner measure of each set, which is trivial for measurable sets. The detail of the proof is reported in the appendix for the sake of legibility.

Lemma 7: Let

$$\begin{aligned} \alpha_i &= 1 - \sup\{\nu_i(S) : S \in \mathcal{B}(\mathbb{R}), S \subseteq X_{\{i\}}\}, \\ \beta_i &= \sup\{\mu_i(S) : S \in \mathcal{B}(\mathbb{R}), S \subseteq \mathbb{R}\backslash X_{\{i\}}\}. \end{aligned}$$

If X^- and X^+ are both not empty, then the following holds:

$$\alpha(f) = 1 - \prod_{i=1}^m (1-\alpha_i),\ \beta(f) = \prod_{i=1}^m \beta_i.$$

We are now ready to prove Theorem 4 by leveraging the independence of y_i.

Proof of Theorem 4: It is easy to see that if X^+ (X^-) is empty, then $f = -1$ ($f = 1$), which implies that $X_{\{i\}} = T_i(\infty)$ ($X_{\{i\}} = T_i(-\infty)$). Now assume that X^+ and X^- are not empty, by Lemma 7,

$$P_e(f) = 1 - p^+ \prod_{i=1}^m \beta_i - p^- \prod_{i=1}^m (1-\alpha_i).$$

Suppose the optimal α_i, β_i are α_i^*, β_i^*. As a result, we know that

$$\begin{aligned} P_e^* &= 1 - \left[p^+ \prod_{j\neq 1} \beta_j^* \right] \beta_1^* - \left[p^- \prod_{j \neq 1} (1-\alpha_j^*) \right] (1-\alpha_1^*) \\ &= a_1^* \alpha_1^* - b_1^* \beta_1^* + c_1^*, \end{aligned}$$

where

$$a_1^* = \left[p^- \prod_{j\neq 1}(1-\alpha_j^*) \right], b_1^* = \left[p^+ \prod_{j \neq 1} \beta_j^* \right],\ c_1^* = 1 - a_1^*.$$

By the Bayes Risk Criterion [16], we could change $X_{\{1\}}$ to the form (12), with $\eta_1 = log(a_1^*/b_1^*)$, without increasing the probability of error. Similarly, we can sequentially change

[3] The threshold η_i is not the threshold for Bayesian detector based on y_i in general, as illustrated in Section VII.

$X_{\{2\}}, \ldots, X_{\{m\}}$ to the form (12) without increasing the probability of error, which concludes the proof. ■

Remark 3: It is clear that β_i and α_i are functions of η_i, when $X_{\{i\}}$ is of the form (12). Due to Corollary 1, we know that

$$P^* = \min_{\eta_i} 1 - p^+ \prod_{i=1}^{m} \beta_i - p^- \prod_{i=1}^{m} (1 - \alpha_i), \tag{13}$$

which can be solved numerically. Therefore, we effectively reduce the search space from all pairs of mutually exclusive sets in $\mathcal{X}_{m,n}$ to a m-dimensional vector of thresholds.

VI. A Heuristic Detector for General n

In this section we propose a heuristic detector, where we design the set $X_\mathcal{I}$ as

$$X_\mathcal{I} = \{y \in \mathbb{R}^{m-2n} : \sum_{i \in \mathcal{I}} \Lambda_i(y_i) < \eta\},$$

where $\eta = \log(p^-/p^+)$. It is easy to see that in that case, X^- and X^+ are given by:

$$X^- = \{y \in \mathbb{R}^m : \sum_{i \in \mathcal{I}} \Lambda_i(y_i) < \eta, \forall |\mathcal{I}| = m - 2n\}, \tag{14}$$

and

$$X^+ = \{y \in \mathbb{R}^m : \sum_{i \in \mathcal{I}} \Lambda_i(y_i) \geq \eta, \forall |\mathcal{I}| = m - 2n\}. \tag{15}$$

We show that the corresponding f has a simple structure, if the following assumption is satisfied:

(H1) $\inf_{y_i} \Lambda_i(y_i) = -\infty$, $\sup_{y_i} \Lambda_i(y_i) = \infty$.

Moreover, we prove that such detector is asymptotically optimal for i.i.d. measurements when n is fixed and $m \to \infty$.

Remark 4: (H1) is satisfied when y_is are Gaussian random variables with different means under each hypothesis but with the same variance.

Let us first arrange y_is as $y_{i_1}, \ldots, y_{i_m}$, such that

$$\Lambda_{i_1}(y_{i_1}) \geq \Lambda_{i_2}(y_{i_2}) \geq \ldots \geq \Lambda_{i_m}(y_{i_m}).$$

Also define the index set

$$\mathcal{I}_k(y) = \{i_k, \ldots, i_{k+m-2n-1}\},\ k = 1, \ldots, 2n+1.$$

It is easy to check that $|\mathcal{I}_k(y)| = m - 2n$. Now let us define function $h_k(y)$ as

$$h_k(y) \triangleq \begin{cases} -1 & \sum_{i \in \mathcal{I}_k(y)} \Lambda_i(y_i) < \eta \\ 1 & \sum_{i \in \mathcal{I}_k(y)} \Lambda_i(y_i) \geq \eta. \end{cases} \tag{16}$$

The following theorem claims that the heuristic detector has the following form:

Theorem 5: Consider a heuristic detector f_0, such that

$$f_0(y) = \begin{cases} 1 & d(y, X^-) \geq d(y, X^+) \\ -1 & d(y, X^-) < d(y, X^+), \end{cases} \tag{17}$$

with

$$X^- = \{y \in \mathbb{R}^m : \sum_{i \in \mathcal{I}} \Lambda(y_i) < \eta, \forall |\mathcal{I}| = m - 2n\},$$

and

$$X^+ = \{y \in \mathbb{R}^m : \sum_{i \in \mathcal{I}} \Lambda(y_i) \geq \eta, \forall |\mathcal{I}| = m - 2n\}.$$

If assumption (H1) is satisfied, then the heuristic detector can be computed as

$$f_0(y) = h_{n+1}(y).$$

The proof of Theorem 5 is quite technical and hence is reported in the appendix for the sake of legibility. We would like to remark that the function $\sum_{i \in \mathcal{I}_{n+1}(y)} \Lambda_i(y_i)$ is the truncated sum of log-likelihood ratio since it is the sum of $m - 2n$ log-likelihood ratios whose values are in the middle. As a result, $h_{n+1}(y)$ can be computed very efficiently by the following procedures:

- The detector sorts all the log-likelihood ratio $\Lambda_i(y_i)$ of individual measurements in descending order.
- The detector throws away n measurements with the largest $\Lambda_i(y_i)$s and n measurements with the smallest $\Lambda_i(y_i)$s.
- The detector sums the remaining $m - 2n$ $\Lambda_i(y_i)$s and compares it to η. The detector chooses $\hat{\theta} = -1$ if the truncated sum is less than η, otherwise the detector chooses $\hat{\theta} = 1$.

The complexity of such detector is $O(mlog(m))$, which is the complexity for sorting the likelihood ratio.

A. Asymptotic Optimality

In this subsection, we prove that the heuristic detector f_0 is asymptotically optimal. Throughout this subsection, we assume that the measurements y_i are identically distributed. However, *we do not require that assumption (H1) holds.* Let us define $\alpha_{m,n}$ and $\beta_{m,n}$ as the probability of false alarm and detection of detector f_0 defined in (17) with m sensors and $n < m/2$ corrupted measurements.

Moreover we define

$$P(m,n) \triangleq p^+(1 - \beta_{m,n}) + p^- \alpha_{m,n},$$

and $P^*(m,n)$ as the probability error of the optimal detector with m sensors and n corrupted measurements. Since when $n = 0$, i.e. no measurement is corrupted, f_0 is the optimal detector by the Bayes rule, we have the following lemma:

Lemma 8: $P(m,n) \geq P^*(m,n) \geq P^*(m,0) = P(m,0)$.

Let us define the rate function as

$$\overline{I}_n \triangleq \limsup_{m \to \infty} \frac{-\log(P(m,n))}{m},\ \underline{I}_n \triangleq \liminf_{m \to \infty} \frac{-\log(P(m,n))}{m}.$$

Moreover let us define $I_n \triangleq \overline{I}_n$ when $\overline{I}_n = \underline{I}_n$. Similarly one can define the rate function of the optimal detector.

Remark 5: If I_n exists, then from definition

$$e^{(-I_n - \delta)m} \leq P(m,n) \leq e^{(-I_n + \delta)m},$$

for arbitrary small δ and sufficiently large m. As a result, $P(m,n)$ converges to 0 as "fast" as the exponential function $e^{-I_n m}$. Therefore, larger I_n indicates better asymptotic performance. Moreover, if $I_n = I_n^*$, then the probability of error for the heuristic detector converges to 0 as "fast" as that of the optimal detector.

It is well known that I_0 exists, which is formalized by the following lemma [17]:

Lemma 9 (Chernoff Lemma): The optimal decay rate for $n = 0$ is given by

$$I_0 = \inf_{0<t<1} \log \left[E\left(e^{t\Lambda(y)} | \theta = -1 \right) \right].$$

The following theorem proves that the heuristic detector decays as "fast" as the optimal detector:

Theorem 6: I_n and I_n^* exist. Moreover, the following equality holds:

$$I_n = I_n^* = I_0. \tag{18}$$

Proof: Due to Lemma 8, we know that $\overline{I}_n \le \overline{I}_n^* \le I_0$. As a result, we only need to prove that $\underline{I}_n \ge I_0$. From the definition of $\alpha_{m,n}$ and the fact that $X^- \subseteq Y^-(f_0)$[4],

$$\alpha_{m,n} \le \nu(\mathbb{R}^m \backslash X^-).$$

Now by the definition of X^-, we know that

$$\begin{aligned}\nu(\mathbb{R}^m \backslash X^-) &= \nu(\bigcup_{|\mathcal{I}|=m-2n} \{y \in \mathbb{R}^m : \sum_{i\in\mathcal{I}} \Lambda_i(y_i) \ge \eta\}) \\ &\le \sum_{|\mathcal{I}|=m-2n} \nu(\{y \in \mathbb{R}^m : \sum_{i\in\mathcal{I}} \Lambda_i(y_i) \ge \eta\}).\end{aligned}$$

Since y_is are i.i.d. distributed, we know that

$$\begin{aligned}\alpha_{m,n} &\le \left(\begin{array}{c} m \\ m-2n \end{array} \right) \nu(\{y \in \mathbb{R}^m : \sum_{i=1}^{m-2n} \Lambda_i(y_i) \ge \eta\}) \\ &\le m^{2n} \alpha_{m-2n,0}\end{aligned}$$

Similarly, one can prove that

$$1 - \beta_{m,n} \le m^{2n}(1 - \beta_{m-2n,0}).$$

Therefore,

$$P(m, n) = p^-\alpha_{m,n} + p^+(1 - \beta_{m,n}) \le m^{2n} P(m - 2n, 0),$$

which implies that

$$\begin{aligned}\underline{I}_n &= \liminf_{m\to\infty} \frac{-\log(P(m,n))}{m} \\ &\ge -\limsup_{m\to\infty} \frac{2n\log(m)}{m} + \liminf_{m\to\infty} \frac{-\log(P(m-2n,0))}{m} \\ &= 0 + \liminf_{m\to\infty} \frac{-\log(P(m-2n,0))}{m-2n} \times \frac{m-2n}{m} = I_0.\end{aligned}$$

As a result, $I_n = I_n^* = I_0$. ■

Remark 6: Theorem 6 claims that the heuristic detector f_0 achieves the same convergence rate as the optimal detector f_*. As a result, if $m \gg n$, which means a small percentage of the measurements are corrupted, then the heuristic detector is a good approximation of the optimal detector and should be used due to its low computational complexity.

Table I summarizes the discussion on the optimal detector until this point .

[4] We use the fact that X^- and X^+ defined in (14),(15) are measurable.

VII. I.I.D. Gaussian Case

We now specialize our results for *i.i.d. Gaussian measurement* y_i. In particular, we assume that

$$y_i = a\theta + v_i,$$

where $a > 0$ is constant and v_is denote i.i.d. Gaussian variables. Without loss of generality, we assume that v_is have zero mean and unit variance.

A. Optimal Detector for $n = (m - 1)/2$

We first consider the case where $m - 2n = 1$. It is easy to prove that $X_{\{i\}}$s in Theorem 4 are of the form

$$X_{\{i\}} = T(\eta_i) = \{y_i \in \mathbb{R} : y_i < \zeta_i\}, \tag{19}$$

with $\zeta_i = \eta_i/2a$. Moreover, the following results use symmetry to provide an even tighter characterization of the sets corresponding to the optimal detector.

Theorem 7: In the case of i.i.d. Gaussian measurements and $m - 2n = 1$, the optimal worst-case probability of error is given by

$$P_e^* = 1 - \sup_{\zeta} \left(p^+ \left[Q(\zeta - a)\right]^m + p^- \left[Q(-\zeta - a)\right]^m \right), \tag{20}$$

where

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du.$$

Moreover, the $X_{\{i\}}$s of the optimal detector f_* in (10) are symmetric and of the form

$$X_{\{i\}} = \{y_i \in \mathbb{R} : y_i < \zeta\}, \quad \forall i \in \{1, 2, ..., m\} \tag{21}$$

for any $\zeta \in \mathbb{R} \cup \{-\infty, +\infty\}$ that achieves the supremum in (20).

Remark 7: The main difference between Theorem 7 and 4 is that all the individual thresholds in Theorem 7 are essentially the same, which reduces the search space further from $\mathbb{R}^m$ to $\mathbb{R}$.

Before proving Theorem 7, we need the following lemma, which characterizes one important property of Q:

Lemma 10: Q satisfies the following inequality:

$$Q(x)Q(y) \le \left[Q\left(\frac{x+y}{2} \right) \right]^2. \tag{22}$$

Moreover, the equality holds only when $x = y$.

Proof: It is easy to see that (22) holds if $\log(Q(x))$ is strictly concave. Consider the second derivative of $\log(Q(x))$, we have

$$\frac{d^2}{dx^2} \log(Q(x)) = \frac{Q(x)Q''(x) - Q'(x)Q'(x)}{Q^2(x)}.$$

Therefore, we only need to prove that

$$Q(x)Q''(x) - Q'(x)Q'(x) < 0, \forall x$$

It is easy to derive that

$$Q'(x) = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \; Q''(x) = \frac{1}{\sqrt{2\pi}} x e^{-x^2/2}.$$

m, n	Detection algorithm	Optimality
$n \geq m/2$	Choose the hypothesis with larger prior probability	Optimal
$n = (m-1)/2$	Compute m estimates θ_i based on y_i by N-P detectors, then vote	Optimal
$n \ll m$	Compare the truncated sum of log-likelihood ratio to threshold η	Asymptotically optimal
$n = 0$	Perform Naive Bayes detection	Optimal

TABLE I

THE OPTIMAL AND HEURISTIC DETECTORS FOR DIFFERENT m, n

Thus,

$$\begin{aligned} & Q(x)Q''(x) - Q'(x)Q'(x) \\ &\quad = \frac{1}{2\pi} e^{-x^2/2} \left(\int_x^\infty x e^{-u^2/2} du - e^{-x^2/2} \right) \\ &\quad < \frac{1}{2\pi} e^{-x^2/2} \left(\int_x^\infty u e^{-u^2/2} du - e^{-x^2/2} \right) \\ &\quad = \frac{1}{2\pi} e^{-x^2/2} \left(e^{-x^2/2} - e^{-x^2/2} \right) = 0. \end{aligned}$$

As a result, $\log(Q(x))$ is strictly concave, which concludes the proof. ■

Proof of Theorem 7: From the definition of Q function, it is trivial to see that

$$\alpha_i = Q(\zeta_i + a), \qquad \beta_i = Q(\zeta_i - a).$$

Therefore,

$$\begin{aligned} P_e(f) &= 1 - p^+ \prod_{i=1}^m Q(\zeta_i - a) - p^- \prod_{i=1}^m (1 - Q(\zeta_i + a)) \\ &= 1 - p^+ \prod_{i=1}^m Q(\zeta_i - a) - p^- \prod_{i=1}^m Q(-\zeta_i - a). \quad (23) \end{aligned}$$

Let us denote by ζ_i^* the constant ζ_i in (19) that corresponds to the optimal detector f_*.

If either X^- or X^+ is empty, then $\zeta_i^* = -\infty$ or $\zeta_i^* = \infty$ and the proof is trivial . As a result, we assume that X^- and X^+ are non-empty. By contradiction assume that $\zeta_1^* \neq \zeta_2^*$. By Lemma 10, we know that

$$\prod_{i=1}^m Q(\zeta_i^* - a) < Q^2\left(\frac{\zeta_1^* + \zeta_2^*}{2} - a\right) \prod_{i=3}^m Q(\zeta_i^* - a),$$

$$\prod_{i=1}^m Q(-\zeta_i^* - a) < Q^2\left(-\frac{\zeta_1^* + \zeta_2^*}{2} - a\right) \prod_{i=3}^m Q(-\zeta_i^* - a).$$

Therefore, the following thresholds $(\zeta_1^* + \zeta_2^*)/2$, $(\zeta_1^* + \zeta_2^*)/2$, $\zeta_3, \dots, \zeta_m$ are strictly better than $\zeta_1^*, \zeta_2^*, \dots, \zeta_m^*$, which contradicts the optimality of f_*. We thus conclude that all the ζ_i^* must be equal since the same argument could have been made for any pair of ζ_i^*'s, which proves (21). The result then follows from this and (23) . ■

In Figure VII-A we plot the probability of error versus the threshold ζ for different pairs of m, n. The parameters are chosen as follows:

$$p^+ = 0.6,\ p^- = 0.4,\ a = 1.$$

The optimum for $m = 1$, $n = 0$ is the pair $\zeta = -0.202$, $P_e = 0.154$. The optimum for $m = 3$, $n = 1$ is $\zeta = -0.508$, $P_e = 0.380$. For the case $m = 5$, $n = 2$, the optimal ζ is actually $-\infty$. Therefore, the optimal detector is simply $f_* = 1$.

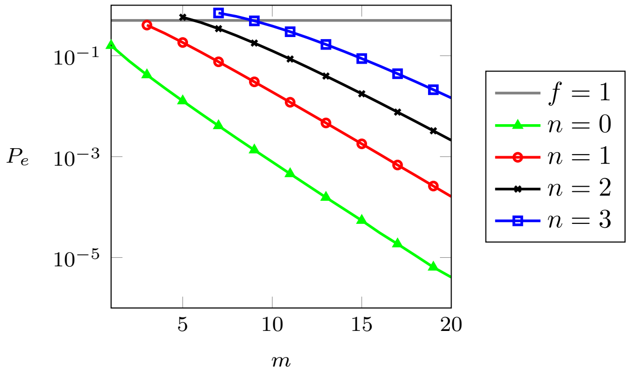


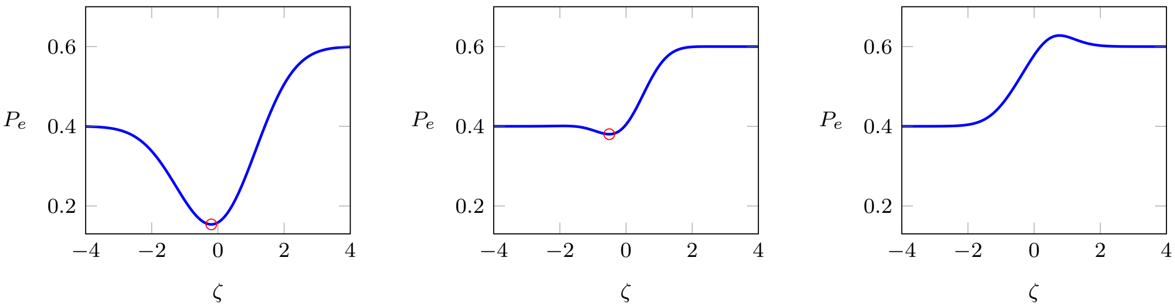
Fig. 2. Probability of Error v.s. Number of Sensors (m). The horizontal line corresponds to the probability of error of the constant detector $f = 1$. The green line corresponds to the probability of error of the heuristic detector f_0 when $n = 0$. The red line corresponds to $P_e(f_0)$ when $n = 1$. The black line corresponds to $P_e(f_0)$ when $n = 2$. The blue line corresponds to $P_e(f_0)$ when $n = 3$.

B. Heuristic Detector for General $n < m/2$

We assume that $p^+ = p^- = 0.5$ and $a = 1$. Figure 2 shows the probability of error $P(m, n)$ versus m for the heuristic detector proposed in Section VI and a constant detector $f = 1$. The $P(m, n)$ is computed as the empirical probability by averaging 10^8 random experiments. It can be seen when n is close to $m/2$, the heuristic detector is worse than constant detector, which shows that the heuristic detector is not necessarily optimal. However, as m goes to infinity. $P(m, n)$ decays as "fast" as $P(m, 0)$, which illustrates that the heuristic detector is asymptotically optimal.

VIII. Conclusion

In this paper we consider the problem of designing detectors able to minimize the probability of error with n corrupted measurements due to integrity attacks on a subset of the sensor pool. The problem is posed as a minimax optimization where the goal is to design the optimal detector against all possible attacker's strategies. We show that if the attacker can manipulate at least half of the m measurements ($n \geq m/2$) then the optimal worst-case detector should ignore *all* m measurements and be based solely on the a-priori information. When the attacker can manipulate less than of half of the measurements ($n < m/2$), we show that the optimal detector is a threshold rule based on a Hamming-like distance between the manipulated measurement vector and two appropriately defined sets. For a particular case ($m = 2n + 1$) we were able to compute the optimal detector, showing that it consists of a simple voting scheme. A heuristic detector, which is



(a) $m = 1$, $n = 0$

(b) $m = 3$, $n = 1$

(c) $m = 5$, $n = 2$

Fig. 1. Probability of Error v.s. threshold ζ

asymptotically optimal when $m \to \infty$, is proposed for general $n < m/2$. We further apply the results to i.i.d. Gaussian case.

IX. Appendix

A. Proof of Lemma 7

Before proving Lemma 7, we first prove a preliminary lemma on the relationship between X^+, X^- and the corresponding $Y^+(f), Y^-(f)$:

Lemma 11: Let X^-, $X^+ \in \mathcal{X}_{m,n}$ be mutually exclusive sets generated as follows

$$X^- = \{y \in \mathbb{R}^m : \text{Trunc}_{\mathcal{I}}(y) \in X_{\mathcal{I}}, \forall |\mathcal{I}| = m - 2n\},$$
$$X^+ = \{y \in \mathbb{R}^m : \text{Trunc}_{\mathcal{I}}(y) \in \mathbb{R}^{m-2n} \backslash X_{\mathcal{I}}, \forall |\mathcal{I}| = m - 2n\}.$$

If the following holds for all $|\mathcal{I}| = m - 2n$:

$$\text{Trunc}_{\mathcal{I}}(X^-) = X_{\mathcal{I}},\ \text{Trunc}_{\mathcal{I}}(X^+) = \mathbb{R}^{m-2n} \backslash X_{\mathcal{I}},$$

then $X^- = Y^-(f)$, $X^+ = Y^+(f)$, where f is defined in (10).

Proof: By Theorem 3, we know that

$$X^+ \subseteq Y^+(f),\ X^- \subseteq Y^-(f).$$

Consider $\text{Trunc}_{\mathcal{I}}(Y^-(f))$ with $|\mathcal{I}| = m - 2n$, we know that

$$X_{\mathcal{I}} = \text{Trunc}_{\mathcal{I}}(X^-) \subseteq \text{Trunc}_{\mathcal{I}}(Y^-(f)).$$

By Theorem 3, we have

$$\text{Trunc}_{\mathcal{I}}(Y^+(f)) \subseteq \mathbb{R}^{m-2n} \backslash X_{\mathcal{I}}. \tag{24}$$

Since equation (24) is true for every $|\mathcal{I}| = m - 2n$, we know that $Y^+(f) \subseteq X^+$, which implies that $Y^+(f) = X^+$. Similarly we can prove that $Y^-(f) = X^-$. ■

Proof of Lemma 7: We only prove that $\beta = \prod_{i=1}^m \beta_i$. The other equality follows the same argument. Since X^- and X^+ are not empty, by Lemma 11 and Lemma 6, we know that $X^- = Y^-(f)$ and $X^+ = Y^+(f)$. Therefore

$$\beta = \sup\{\mu(Q) : Q \subseteq X^+,\ Q \in \mathcal{B}(\mathbb{R}^m)\}.$$

Consider $Q_i \subseteq \mathbb{R} \backslash X_i$ and $Q_i \in \mathcal{B}(\mathbb{R})$. It is trivial to see that

$$\prod_{i=1}^m Q_i \subseteq \prod_{i=1}^m \mathbb{R} \backslash X_i = X^+.$$

Hence,

$$\beta \ge \mu(\prod_{i=1}^m Q_i) = \prod_{i=1}^m \mu_i(Q_i).$$

Taking the supremum on the right hand side over all measurable $Q_i \subseteq R \backslash X_i$, we have

$$\beta \ge \prod_{i=1}^m \beta_i.$$

On the other hand, suppose that $Q \subseteq X^+$ and $Q \in \mathcal{B}(\mathbb{R}^m)$. Therefore

$$\text{Trunc}_{\{i\}}(Q) \subseteq \text{Trunc}_{\{i\}}(X^+) = X_i.$$

Let us write $\text{Trunc}_{\{i\}}(Q)$ as Q_i. It can be proved that Q_i is universally measurable. In other words, there exist measurable sets $\overline{Q}_i$ and $\underline{Q}_i$, such that

$$\underline{Q}_i \subseteq Q_i \subseteq \overline{Q}_i,\ \mu_i(\underline{Q}_i) = \mu_i(\overline{Q}_i).$$

As a result,

$$\mu(Q) \le \mu(\prod_{i=1}^m \overline{Q}_i) = \prod_{i=1}^m \mu_i(\overline{Q}_i) = \prod_{i=1}^m \mu_i(\underline{Q}_i) \le \prod_{i=1}^m \beta_i.$$

The last inequality holds since $\underline{Q}_i \subseteq Q_i \subseteq X_i$. Take the supremum on the left side, we have

$$\beta \le \prod_{i=1}^m \beta_i,$$

which concludes the proof. ■

B. Proof of Theorem 5

Before proving Theorem 5, let us define

$$n^-(y) \triangleq \sum_{k=1}^{2n+1} \mathbb{I}_{h_k(y)=1}, n^+(y) \triangleq \sum_{k=1}^{2n+1} \mathbb{I}_{h_k(y)=-1},$$

where $\mathbb{I}$ is the indicator function. We have the following lemma:

Lemma 12: $n^-(y) + n^+(y) = 2n + 1$, $n^-(y) = d(y, X^-)$, $n^+(y) = d(y, X^+)$.

Proof: Without loss of generality, let us assume that

$$\Lambda_1(y_1) \ge \Lambda_2(y_2) \ge \ldots \ge \Lambda_m(y_m).$$

It is trivial to prove that $n^-(y) + n^+(y) = 2n + 1$. To prove $n^-(y) = d(y, X^-)$, we first show that $n^-(y) \ge d^-(y, X^-)$. Let

$$M = (m - 2n - 1) \max_i |\Lambda_i(y_i)| + \eta.$$

Consider another vector $y^- \in \mathbb{R}^m$. If $i \leq n^-(y)$, then we choose y_i^- such that

$$\Lambda_i(y_i^-) < -M,\, i = 1, \ldots, n^-(y),$$

and

$$\Lambda_i(y_i^-) \leq \Lambda_{i-1}(y_{i-1}^-),\, i = 2, \ldots, n^-(y),$$

which is always possible due to assumption *(H1)*. If $i > n^-(y)$, we choose $y_i^- = y_i$. It is easy to see that $d(y^-, y) = n^-(y)$. As a result, we only need to prove that $y^- \in X^-$. From the construction of y^-,

$$\begin{aligned}\Lambda_{n^-(y)+1}(y^-_{n^-(y)+1}) &\geq \ldots \geq \Lambda_m(y^-_m) \\ &\geq \Lambda_1(y^-_1) \geq \ldots \Lambda_{n^-(y)}(y^-_{n^-(y)}).\end{aligned} \tag{25}$$

Since

$$X^- = \{y \in \mathbb{R}^m : \sum_{i \in \mathcal{I}} \Lambda(y_i) < \eta, \forall |\mathcal{I}| = m - 2n\}.$$

We know that $y^- \in X^-$ if the sum of the largest $m-2n$ terms in (25) is less than η. First suppose that $n^-(y) < 2n+1$, which implies that $n^-(y) + m - 2n \leq m$. Therefore

$$\sum_{i=1}^{m-2n} \Lambda(y^-_{n^-(y)+i}) = \sum_{i=1}^{m-2n} \Lambda(y_{n^-(y)+i}).$$

From the definition of $n^-(y)$ and monotonicity of $h_k(y)$[5], we know that $h_1(y), \ldots, h_{n^-(y)}(y) = 1$ and $h_{n^-(y)+1}, \ldots, h_{2n+1}(y) = -1$. Since,

$$h_{n^-(y)+1}(y) = \begin{cases} -1 & \sum_{i=1}^{m-2n} \Lambda(y^-_{n^-(y)+i}) < \eta \\ 1 & \sum_{i=1}^{m-2n} \Lambda(y^-_{n^-(y)+i}) \geq \eta. \end{cases}$$

we know that

$$\sum_{i=1}^{m-2n} \Lambda(y^-_{n^-(y)+i}) - \eta < 0,$$

which implies that $y^- \in X^-$. Now suppose that $n^-(y) = 2n+1$. It can be proved that

$$\begin{aligned}\sum_{i=m-2n+2}^{m} \Lambda(y_i^-) + \Lambda(y_1^-) &= \sum_{i=m-2n+2}^{m} \Lambda(y_i) + \Lambda(y_1^-) \\ &< (m-2n-1)\max_i |\Lambda(y_i)| - M \leq \eta,\end{aligned}$$

which implies that $y^- \in X^-$. Hence, $n^-(y) \geq d(y, X^-)$. Similarly one can prove that $n^+(y) \geq d(y, X^+)$. By Lemma 5, we have

$$\begin{aligned}&n^-(y) + n^+(y) = 2n + 1 \leq d(y, X^-) + d(y, X^+), \\ &n^-(y) \geq d(y, X^-),\, n^+(y) \geq d(y, X^-).\end{aligned}$$

Therefore, $n^-(y) = d(y, X^-)$, $n^+(y) = d(y, X^+)$. ■

Now we are ready to prove Theorem 5:

Proof of Theorem 5: By Lemma 12,

$$f_0(y) = \begin{cases} 1 & n^-(y) \geq n^+(y) \\ -1 & n^-(y) < n^+(y). \end{cases}$$

Since $n^-(y) + n^+(y) = 2n + 1$, we know that

$$f_0(y) = \begin{cases} 1 & n^-(y) \geq n + 1 \\ -1 & n^-(y) < n. \end{cases}$$

Due to monotonicity of $h_k(y)$, $n^-(y) \geq n + 1$ if and only if $h_{n+1}(y) = 1$. Therefore

$$f_0(y) = h_{n+1}(y),$$

which concludes the proof. ■

References

[1] T. M. Chen, "Stuxnet, the real start of cyber warfare? [editor's note]," *IEEE Network*, vol. 24, no. 6, pp. 2–3, 2010.

[2] D. P. Fidler, "Was stuxnet an act of war? decoding a cyberattack," *IEEE Security & Privacy*, vol. 9, no. 4, pp. 56–59, 2011.

[3] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *HOTSEC'08: Proceedings of the 3rd conference on Hot topics in security*. Berkeley, CA, USA: USENIX Association, 2008, pp. 1–6.

[4] P. J. Huber, "A robust version of the probability ratio test," *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. pp. 1753–1758, 1965.

[5] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *The Annals of Statistics*, vol. 1, no. 2, pp. 251–263, 1973.

[6] S. Verdú and H. V. Poor, "On minimax robustness: A general approach and applications," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 328–340, 1984.

[7] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, 1985.

[8] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Wiely, 2009.

[9] T. Basar and Y. W. Wu, "Solutions to a class of minimax decision problems arising in communication systems," *Journal of Optimization Theory and Applications*, vol. 51, pp. 375–404, 1986.

[10] R. Bansal and T. Basar, "Communication games with partially soft power constraints," *Journal of Optimization Theory and Applications*, vol. 61, pp. 329–346, 1989.

[11] S. Bayram and S. Gezici, "On the restricted Neyman–Pearson approach for composite hypothesis-testing in presence of prior distribution uncertainty," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 5056–5065, 2011.

[12] A. Mutapcic and S.-J. Kim, "Robust signal detection under model uncertainty," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 287–290, 2009.

[13] R. Tandra, "Fundamental limits on detection in low SNR," Master's thesis, University of California, Berkeley, 2005.

[14] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 4–17, 2008.

[15] A. Kerckhoffs, "La cryptographie militairie," *Journal des Sciences Militaires*, vol. IX, pp. 5–38, 1883.

[16] L. Scharf, *Statistical Signal Processing*. Prentice Hall, 1990.

[17] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *The Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 493–507, Dec 1952.

[5] $h_k(y)$ is monotonically decreasing.