Secure Estimation in the Presence of Integrity Attacks

Yilin Mo*, Bruno Sinopoli†

Abstract

We consider the estimation of a scalar state based on m measurements that can be potentially manipulated by an adversary. The attacker is assumed to have full knowledge about the true value of the state to be estimated and about the value of all the measurements. However, the attacker has limited resources and can only manipulate up to l of the m measurements. The problem is formulated as a minimax optimization, where one seeks to construct an optimal estimator that minimizes the "worst-case" expected cost against all possible manipulations by the attacker. We show that if the attacker can manipulate at least half the measurements ($l \geq m/2$), then the optimal worst-case estimator should ignore all measurements and be based solely on the a-priori information. We provide the explicit form of the optimal estimator when the attacker can manipulate less than half the measurements ($l < m/2$), which is based on $\binom{m}{2l}$ local estimators. We further prove that such an estimator can be reduced into simpler forms for two special cases, i.e., either the estimator is symmetric and monotone or $m = 2l+1$. Finally we apply the proposed methodology in the case of 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. The first-ever SCADA

This research is supported in part by grants CNS-1135895 and ECCS-0955111 from NSF.

*: Yilin Mo is with the department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA. Email: yilinmo@caltech.edu

†: Bruno Sinopoli are with the ECE department of Carnegie Mellon University, Pittsburgh, PA. Email: brunos@ece.cmu.edu

system malware (called Stuxnet) was found in July 2010 and rose significant concern about SCADA system security [1], [2]. The research community has acknowledged the importance of addressing the challenge of designing secure estimation and control systems [3].

We consider a secure estimation problem inspired by security concerns that arise from the possible manipulation of sensor data. We focus our attention on the estimation of a scalar state x from measurements collected by m sensors, with the caveat that some of these measurements can be manipulated by a malicious third party. The attacker is assumed to have full information about the true value of x and all the measurements and use this information to manipulate the data available to the estimator. Limitations in the resources available to the attacker enable him to only manipulate l of the m sensors. However, the attacker has total control over the corrupted sensors, as it can change the measurements of the compromised sensors arbitrarily. To minimize the estimator's performance degradation in the presence of such attacks, we construct minimax estimator that minimize the "worst-case" expected cost.

We start by considering the case $l \geq m/2$, in which the attacker can manipulate at least half the measurements. We show that in such a scenario the optimal worst-case estimators should ignore *all* m measurements and be based solely on the a-priori distribution of x. This result provides a fundamental limitation on the estimation performance in adversarial environment and is in sharp contrast with non-adversarial estimation theory where even very noisy data can provide some information.

For the case $l < m/2$, in which the attacker can manipulate less than half the sensors, we provide the explicit form of the optimal estimator, which is based on $\binom{m}{2l}$ local estimators. Therefore, the search space of the optimal estimator is reduced from all possible functions to a special class of functions. We also prove that the expected "worst-case" estimation performance is a convex functional with respect to the local estimators provided that cost function is convex. Hence, convex optimization techniques can be used to compute the optimal estimator when the local estimators are finite-degree polynomials. For general cases, such an estimator may be computationally hard to implement, as it needs to compute all $\binom{m}{2l}$ local estimates. However, we prove that under two special cases, i.e., either the estimator is symmetric and monotone or $m = 2l + 1$, the optimal estimator can be reduced to simpler form, the computational complexity of which is $O(m \log m)$.

Related Work

Robust estimators such as M-estimator, L-estimator, R-estimator and etc. have also been extensively studied in the literature [4], [5], [6]. However, such kinds of approaches usually assume that the outliers of the data are generated *independently* by some other probability distribution different from the model assumptions. Furthermore, the robustness are usually measured by breakdown points [7], [8] or influence functions [9]. In this paper, we assume that the attacker generates the optimal "outliers" to destroy the estimation performance. Since the attacker can take control over multiple sensors, the compromised measurements from these sensors are jointly selected by the adversary to maximize the estimation error. Furthermore, the security of an estimator is measured by the worst-case mean squared error. Hence, a robust estimator may not necessarily be secure and thus the techniques developed for robust estimation need to be reexamined before they can be applied in the context of security.

For dynamical systems, robust estimation techniques such as $\mathcal{H}_\infty$ estimators have also been an active research area for the past decades. The $\mathcal{H}_\infty$ estimator can be seen as the worst-case estimator when the disturbance is in the $\mathcal{L}_2$ space or of bounded power spectral density, as the $\mathcal{H}_\infty$ norm can be interpreted as an induced norm [10]. In security settings, we feel that the sparsity of the disturbance is a better way to characterize the capability of the adversary, since it can change the compromised sensor readings arbitrarily.

Furthermore, bad data detection and identification techniques, which is based on truncating the "atypical" data, have been widely used in large scaled systems such as power grid [11]. While such approaches are very successful in detecting and removing random failures, they are not effective against integrity attacks. Liu et al. [12] illustrate how an adversary can inject a stealthy input into the measurements to change the state estimation, without being detected by the bad data detector. Sandberg et al. [13] consider how to find a sparse stealthy input, which enables the adversary to launch an attack with a minimum number of compromised sensors. Xie et al. [14] further illustrate that the stealthy integrity attacks on state estimation can lead to a financial gain in the electricity market for the adversary.

This paper generalizes our previous works on secure estimation [15], which only consider designing the optimal symmetric estimator which minimizes the "worst-case" mean squared error. In this paper, we derive the optimal estimator (not necessarily symmetric), which minimizes an arbitrary quasiconvex cost function. Furthermore, this paper extends our previous work on hypothesis testing in adversarial environment [16], where the system needs to make a binary decision (instead of real valued state estimation) on which hypothesis is true based on the potentially compromised sensory information.

The rest of paper is organized as follows: In Section II we formulate the problem of secure estimation with l manipulated measurements from m total measurements. In Section III, we consider the optimal estimator design for the cases $l \geq m/2$. In Section IV, we provide an explicit form of the optimal estimator when $l < m/2$. Furthermore, in Section V we discuss two special cases, i.e., either the estimator is symmetric and monotone or $m = 2l + 1$, and prove that the optimal estimator can be further reduced to simpler forms. In Section VI we provide a numerical example of Gaussian measurements. Finally Section VII concludes the paper.

II. Problem Formulation

The goal is to estimate a random variable $x \in \mathbb{R}$ from a vector $y \triangleq [y_1, \ldots, y_m]^T \in \mathbb{R}^m$ consisting of m sensor measurements $y_i \in \mathbb{R}$, where the index $i \in \mathcal{S} \triangleq \{1, 2, ..., m\}$. We assume that x and y follow the following joint distribution:

$$P((x, y) \in S) = \mu(S), \tag{1}$$

where $S \subseteq \mathbb{R}^{m+1}$ is a Borel-measurable set and μ is a probability measure on $\mathbb{R}^{m+1}$.

We assume that an attacker wants to disturb the state estimation. To this end, the attacker has the ability to manipulate up to l of the m sensor measurements. Formally, this means that our estimate has to rely on a vector $y^c \in \mathbb{R}^m$ of *manipulated measurements* defined by

$$y^c = y + \gamma \circ y^a, \tag{2}$$

where $\circ$ is element-wise multiplication and the *sensor-selection* vector γ taking values in

$$S_\gamma \triangleq \{\gamma \in \mathbb{R}^m : \gamma_i = 0 \; or \; 1, \sum_{i=1}^{m} \gamma_i = l\} \tag{3}$$

and the *bias* vector y^a taking values in $\mathbb{R}^m$. By selecting which values of γ are nonzero, the attacker chooses which of the n sensors will be manipulated. The "level" of manipulation is determined by y^a.

Define the estimated state based on y^c to be $\hat{x} \in \mathbb{R}$. Now we give a formal definition of the estimator:

Definition 1: An estimator $f : \mathbb{R}^m \to \mathbb{R}$ is a mapping from the compromised sensor measurements y^c to the state estimate $\hat{x}$, i.e.,

$$\hat{x} = f(y^c) = f(y + \gamma \circ y^a) \tag{4}$$

Given the state and corresponding state estimation, the error e is defined as

$$e \triangleq x - \hat{x}. \tag{5}$$

We further define the cost associated with error e to be

$$cost = c(e), \tag{6}$$

where $c : \mathbb{R} \to \mathbb{R}$ is the cost function, which is assumed to be a *quasiconvex* function with respect to e. In other words, the following inequality holds for all $e_1 \leq e_2 \leq e_3$,

$$c(e_2) \leq \max(c(e_1), c(e_3)). \tag{7}$$

Some typical quasiconvex cost functions include the squared error ($c(e) = e^2$) and the absolute error ($c(e) = |e|$).

The estimation problem is formalized as a minimax problem where the system operator wants to select an optimal estimator so as to minimize the expected cost, for the worst case manipulation by the adversary. Following Kerckhoffs' Principle [17] that security should not rely on the obscurity of the system, our goal is to design the estimator f 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 x and all the measurements $y_1, \ldots, y_m$ are assumed to be known to the attacker. However, due to limited resources, the attacker can only manipulate l of the m sensors. We assume that the system operator knowns how many sensors l can be attacked, but cannot identify them.

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

To compute the worst-case expected cost that we seek to minimize, let us first consider that the adversary compromised a subset $\mathcal{I} \subseteq \mathcal{S}$ of sensors. We have the following definitions:

Definition 2: Define the cardinality $|\mathcal{I}|$ of set $\mathcal{I}$ as the number of elements in $\mathcal{I}$.

Definition 3: Define the complement of an index set $\mathcal{I} \subseteq \mathcal{S}$ as $\mathcal{I}^c \triangleq \{x \in \mathcal{S} : x \notin \mathcal{I}\}$. The difference between two index sets $\mathcal{K}$ and $\mathcal{I}$ is defined as

$$\mathcal{K} \backslash \mathcal{I} \triangleq \mathcal{K} \cap \mathcal{I}^c. \tag{8}$$

Definition 4: Define vector $\gamma_{\mathcal{I}} \triangleq [\gamma_0, \dots, \gamma_m]^T \in \mathbb{R}^m$, where $\gamma_i = 1$ if $i \in \mathcal{I}$ and $\gamma_i = 0$ otherwise.

Definition 5: Given the estimator f, define function $f_{\mathcal{I}}^{+} : \mathbb{R}^m \to \mathbb{R}$ and $f_{\mathcal{I}}^{-} : \mathbb{R}^m \to \mathbb{R}$ as

$$f_{\mathcal{I}}^{+}(y) \triangleq \sup_{y^a \in \mathbb{R}^m} f(y + \gamma_{\mathcal{I}} \circ y^a), \tag{9}$$

$$f_{\mathcal{I}}^{-}(y) \triangleq \inf_{y^a \in \mathbb{R}^m} f(y + \gamma_{\mathcal{I}} \circ y^a). \tag{10}$$

Let us further define the following functions:

$$f^{+}(y) \triangleq \max_{|\mathcal{I}|=l} f_{\mathcal{I}}^{+}(y), \; f^{-}(y) \triangleq \min_{|\mathcal{I}|=l} f_{\mathcal{I}}^{-}(y). \tag{11}$$

Remark 2: $f_{\mathcal{I}}^{+}$ ($f_{\mathcal{I}}^{-}$) can be seen as the maximum (minimum) state estimate $\hat{x}$ that the attacker can enforce when the attacker compromised sensors in a fixed set $\mathcal{I}$. Hence, f^{+} (f^{-}) indicates the maximal(minimum) state estimation $\hat{x}$ that an attacker can enforce with the worst possible set of compromised sensors.

By (7), for all $\gamma \in S_\gamma, y^a \in \mathbb{R}^m$, the following inequality holds:

$$c(x - f(y + \gamma \circ y^a)) \leq \max\left[c(x - f^{+}(y)), c(x - f^{-}(y))\right], \tag{12}$$

which indicates that the maximum cost is achieved either when $\hat{x} = f^{+}(y)$ or $\hat{x} = f^{-}(y)$. Thus, given values of x, y and an estimator f, an optimal policy for the attacker can be described as follows:

The attacker computes $f^{+}(y)$ and $f^{-}(y)$ and compare them with x. If $c(x - f^{+}(y)) \geq c(x - f^{-}(y))$, then the attacker chooses γ and y^a such that $f(y + \gamma \circ y^a)$ equals $f^{+}(y)$ (or as close as possible). Otherwise, the attacker chooses γ and y^a such that $f(y + \gamma \circ y^a)$ equals $f^{-}(y)$ (or as close as possible).

Under this worst-case attacker policy, we can define the worst-case expected cost $\mathcal{C}$ for an estimator f[1]:

$$\mathcal{C}(f) \triangleq \mathbb{E}\left\{\max\left[c(x - f^+(y)), c(x - f^-(y))\right]\right\}. \tag{13}$$

Before continuing on, we would like to state the following theorem regarding the optimality of an estimator f, which will be used in future analysis.

Theorem 1: For any estimator f, if there exist functions h^+, h^- and g^+, g^-, such that

$$g^+ \geq f^+ \geq h^+ \geq h^- \geq f^- \geq g^-, \tag{14}$$

then

$$\mathcal{C}(f) \geq \mathbb{E}\left\{\max\left[c(x - h^+(y)), c(x - h^-(y))\right]\right\}, \tag{15}$$

$$\mathcal{C}(f) \leq \mathbb{E}\left\{\max\left[c(x - g^+(y)), c(x - g^-(y))\right]\right\}. \tag{16}$$

Proof: We only prove (15), as (16) follows the same argument. Since c is quasiconvex, it is easy to see that

$$c(x - h^+(y)) \leq \max\left[c(x - f^+(y)), c(x - f^-(y))\right],$$

$$c(x - h^-(y)) \leq \max\left[c(x - f^+(y)), c(x - f^-(y))\right],$$

which implies that

$$\begin{aligned} &\max\left[c(x - f^+(y)), c(x - f^-(y))\right] \\ &\qquad \geq \max\left[c(x - h^+(y)), c(x - h^-(y))\right]. \end{aligned} \tag{17}$$

By taking expectation on both sides, we can conclude the proof. ■

The following corollary can be immediately proved by Theorem 1:

Corollary 1: For two estimators f_1 and f_2, if the following inequalities hold

$$f_1^+ \geq f_2^+ \geq f_2^- \geq f_1^-, \tag{18}$$

then

$$\mathcal{C}(f_1) \geq \mathcal{C}(f_2). \tag{19}$$

[1] Even if the function f is measurable, f^+ and f^- may not be necessarily measurable. In that case, we can use upper Darboux integral instead of Lebesgue integral to define the expected cost. However, all the discussion in this paper will hold regardless.

Theorem 1 implies that in order to find the optimal f, we should make $f^{+}(y)$ and $f^{-}(y)$ as "close" as possible to each other.

III. Optimal Estimator Design for $l \geq m/2$

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 x, which is formalized as the following theorem:

Theorem 2: If $l \geq m/2$, then the optimal estimator[2] f^* is given as $f^* = \delta^*$, where δ^* is the solution of the following optimization problem:

$$\underset{\delta \in \mathbb{R}}{\text{minimize}} \quad \mathbb{E}\left[c(x-\delta)\right]. \tag{20}$$

The rest of the section is devoted to the proof of Theorem 2, which requires an intermediate result:

Lemma 1: If $\mathcal{I} \cup \mathcal{J} = \mathcal{S}$, then there exists a constant δ independent of y, such that

$$f_{\mathcal{I}}^{+}(y) \geq \delta \geq f_{\mathcal{J}}^{-}(y), \forall y. \tag{21}$$

Proof: We will prove Lemma 1 by contradiction. It is easy to see that (21) is equivalent to

$$f_{\mathcal{I}}^{+}(y) \geq f_{\mathcal{J}}^{-}(y'), \forall y, y' \in \mathbb{R}^m.$$

Suppose that on the contrary, there exist $y = [y_1, \dots, y_m]^T$ and $y' = [y'_1, \dots, y'_m]^T$, such that $f_{\mathcal{I}}^{+}(y) < f_{\mathcal{J}}^{-}(y')$. Now consider another $y^o \in \mathbb{R}^m$, such that

$$y_i^o = \begin{cases} y'_i & \text{if } i \in \mathcal{I} \\ y_i & \text{if } i \in \mathcal{I}^c \end{cases}$$

Since $\mathcal{I} \cup \mathcal{J} = \mathcal{S}$, $\mathcal{I}^c \subseteq \mathcal{J}$. It is easy to verify that

$$y^o = y + \gamma_{\mathcal{I}} \circ (y^o - y),\; y^o = y' + \gamma_{\mathcal{J}} \circ (y^o - y').$$

[2] The optimal estimators discussed in this section and later sections may not necessarily be unique.

Therefore, from the definition of $f_{\mathcal{I}}^{+}$ and $f_{\mathcal{J}}^{-}$, we have

$$f(y^o) \leq f_{\mathcal{I}}^{+}(y) < f_{\mathcal{J}}^{-}(y') \leq f(y^o),$$

which is impossible. ■

Now we are ready to prove Theorem 2.

Proof of Theorem 2: Since $l \geq m/2$, for any set $\mathcal{I}$ with cardinality l, we could always find another set $\mathcal{J}$, such that $|\mathcal{J}| = l$ and $\mathcal{I} \cup \mathcal{J} = \mathcal{S}$. By Lemma 1, there exist two constants $\delta_{\mathcal{I}}^{+}$ and $\delta_{\mathcal{J}}^{-}$, such that

$$f_{\mathcal{I}}^{+}(y) \geq \delta_{\mathcal{I}}^{+} \geq \delta_{\mathcal{J}}^{-} \geq f_{\mathcal{J}}^{-}(y).$$

Therefore,

$$f^{+}(y) = \max_{|\mathcal{I}|=l} f_{\mathcal{I}}^{+}(y) \geq \max_{|\mathcal{I}|=l} \delta_{\mathcal{I}}^{+},$$

$$f^{-}(y) = \min_{|\mathcal{I}|=l} f_{\mathcal{I}}^{-}(y) \leq \min_{|\mathcal{I}|=l} \delta_{\mathcal{I}}^{-}.$$

Thus, we could always find a constant δ, such that

$$f^{+}(y) \geq \delta \geq f^{-}(y).$$

By Corollary 1, the expected cost of the estimator $f_1(y) = \delta$ is less than or equals to $f(y)$. Hence, the optimal estimator is a constant estimator and it is straight forward to see, from the definition of the expected cost, that the optimal δ is the solution of (20). ■

IV. Optimal Estimator Design for $l < m/2$

We now consider the case when less than half the measurements can be manipulated by the attacker, i.e., $l < m/2$.

Definition 6: Let $\mathcal{I} = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$ be a subset of $\mathcal{S}$. Define the projection function $\mathrm{Proj}_{\mathcal{I}} : \mathbb{R}^m \to \mathbb{R}^k$ as

$$\mathrm{Proj}_{\mathcal{I}}(y) \triangleq [y_{i_1}, \ldots, y_{i_k}]^T. \tag{22}$$

Definition 7: Let $\mathcal{K}$ be a subset of $\mathcal{S}$ with cardinality $2l$. An estimator $\varphi_{\mathcal{K}} : \mathbb{R}^m \to \mathbb{R}$ is called a local estimator if the following equality holds

$$\varphi_{\mathcal{K}}(y) = \varphi_{\mathcal{K}}(y + \gamma_{\mathcal{K}} \circ y^a), \forall y^a \in \mathbb{R}^m. \tag{23}$$

Remark 3: The locality of $\varphi_{\mathcal{K}}$ is based on the fact that $\varphi_{\mathcal{K}}$ only depends on those measurements whose indices are not in $\mathcal{K}$. It is trivial to prove that a local estimator $\varphi_{\mathcal{K}}$ can be written as a function of projected measurements:

$$\varphi_{\mathcal{K}}(y) \triangleq \Phi_{\mathcal{K}}(\mathrm{Proj}_{\mathcal{K}^c}(y)), \tag{24}$$

where $\Phi_{\mathcal{K}} : \mathbb{R}^{m-2l} \to \mathbb{R}$.

We are now ready to state the main theorem, which relates the optimal estimator f^* with a set of local estimators.

Theorem 3: If $l < m/2$, then the optimal estimator f^* is of the following form

$$f^*(y) = \min_{|\mathcal{I}|=l} \left[\max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}=\emptyset} \varphi^*_{\mathcal{I}\cup\mathcal{J}}(y) \right], \tag{25}$$

where $\{\varphi^*_{\mathcal{K}}\}$ is a set of $\binom{m}{2l}$ local estimators, given by the solution of the following optimization problem:

$$\underset{\{\varphi_{\mathcal{K}}\}\text{ set of local estimators}}{\text{minimize}} \quad \mathbb{E}\left[\max_{|\mathcal{K}|=2l} c(x - \varphi_{\mathcal{K}}(y)) \right]. \tag{26}$$

The rest of the section is devoted to proving Theorem 3, which requires several intermediate results:

Lemma 2: If $l < m/2$, then for any estimator f, there exists a set of local estimators $\{\varphi_{\mathcal{K}}\}$, such that

$$\mathcal{C}(f) \geq \mathbb{E}\left[\max_{|\mathcal{K}|=2l} c(x - \varphi_{\mathcal{K}}(y)) \right]. \tag{27}$$

Proof: Let f be an arbitrary estimator and $\mathcal{K}$ be a subset of $\mathcal{S}$ with cardinality $2l$. Let us find a subset $\mathcal{I} \subset \mathcal{K}$ with cardinality l and define subset

$$\mathcal{J} \triangleq \mathcal{K} \backslash \mathcal{I}.$$

Now define functions $\varphi_{\mathcal{K}}, \varphi^-_{\mathcal{K}}$ to be

$$\varphi_{\mathcal{K}}(y) \triangleq \inf_{y^b} \sup_{y^a} f(y + \gamma_{\mathcal{I}} \circ y^a + \gamma_{\mathcal{J}} \circ y^b)$$

$$\varphi^-_{\mathcal{K}}(y) \triangleq \sup_{y^a} \inf_{y^b} f(y + \gamma_{\mathcal{I}} \circ y^a + \gamma_{\mathcal{J}} \circ y^b)$$

It is easy to see that $\varphi_{\mathcal{K}}$ is a local estimator, since it does not depend on measurement y_i, where $i \in \mathcal{I} \cup \mathcal{J}$. Moreover, we have the following inequalities:

$$f^+ \geq f^+_{\mathcal{I}} \geq \varphi_{\mathcal{K}} \geq \varphi^-_{\mathcal{K}} \geq f^-_{\mathcal{J}} \geq f^-. \tag{28}$$

Hence, we could find $\binom{m}{2l}$ local estimator $\varphi_{\mathcal{K}}$s, such that the following inequalities hold:

$$f^{+} \geq \varphi_{\mathcal{K}} \geq f^{-}, \forall |\mathcal{K}| = 2l.$$

Since the cost function c is quasiconvex,

$$\max\left[c(x - f^{+}(y)), c(x - f^{-}(y))\right] \geq c(x - \varphi_{\mathcal{K}}(y)), \forall |\mathcal{K}| = 2l,$$

which implies that

$$\max\left[c(x - f^{+}(y)), c(x - f^{-}(y))\right] \geq \max_{|\mathcal{K}|=2l}\left[c(x - \varphi_{\mathcal{K}}(y))\right].$$

By taking the expectation on both sides, we can conclude the proof. ■

Remark 4: Lemma 2 provides a lower bound for the expected cost of any estimator, while the following lemma, which can be seen as the converse of Lemma 2, indicates that such a lower bound is achievable.

Lemma 3: For an arbitrary set of local estimators $\{\varphi_{\mathcal{K}}\}$, define the following estimator f to be:

$$f(y) = \min_{|\mathcal{I}|=l}\left[\max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}}(y)\right]. \tag{29}$$

The expected cost of f satisfies the following inequality:

$$\mathcal{C}(f) \leq \mathbb{E}\left[\max_{|\mathcal{K}|=2l}\, c(x - \varphi_{\mathcal{K}}(y))\right]. \tag{30}$$

Proof: The proof is divided into four steps:

1) We first prove the following inequality:

$$f(y) \geq \max_{|\mathcal{J}|=l}\left[\min_{|\mathcal{I}|=l,\, \mathcal{J}\cap\mathcal{I}=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}}(y)\right]. \tag{31}$$

which is equivalent to the following inequality:

$$\max_{|\mathcal{J}|=l, \mathcal{J}\cap\mathcal{I}_0} \varphi_{\mathcal{I}_0\cup\mathcal{J}}(y) \geq \min_{|\mathcal{I}|=l, \mathcal{I}\cap\mathcal{J}_0} \varphi_{\mathcal{I}\cup\mathcal{I}_0}(y), \tag{32}$$

for all index sets $|\mathcal{I}_0| = |\mathcal{J}_0| = l$. Let us find an index set $\mathcal{K}$, such that $\mathcal{I}_0 \subset \mathcal{K}$, $\mathcal{J}_0 \subset \mathcal{K}$ and $|\mathcal{K}| = 2l$. Therefore, we have

$$\varphi_{\mathcal{K}}(y) = \varphi_{\mathcal{I}_0\cup(\mathcal{K}\backslash\mathcal{I}_0)}(y) = \varphi_{(\mathcal{K}\backslash\mathcal{J}_0)\cup J_0}(y).$$

Hence, for all $|\mathcal{I}_0| = |\mathcal{J}_0| = l$,

$$\max_{|\mathcal{J}|=l, \mathcal{J}\cap\mathcal{I}_0} \varphi_{\mathcal{I}_0\cup\mathcal{J}}(y) \geq \varphi_{\mathcal{K}}(y) \geq \min_{|\mathcal{I}|=l, \mathcal{I}\cap\mathcal{J}_0} \varphi_{\mathcal{I}\cup\mathcal{J}_0}(y),$$

which implies (32) and hence (31).

2) By (29) and (31), it is easy to verify that for arbitrary sets $\mathcal{I}_0, \mathcal{J}_0 \subseteq \mathcal{S}$ with cardinality l, the following inequalities hold:

$$f(y) \leq \max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}_0=\emptyset} \varphi_{\mathcal{I}_0\cup\mathcal{J}}(y), \tag{33}$$

$$f(y) \geq \min_{|\mathcal{I}|=l,\, \mathcal{J}_0\cap\mathcal{I}=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}_0}(y). \tag{34}$$

3) As a result of (33),

$$\begin{aligned} f^+_{\mathcal{I}_0}(y) &= \sup_{y^a} f(y+\gamma_{\mathcal{I}_0}\circ y^a) \\ &\leq \sup_{y^a} \max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}_0=\emptyset} \varphi_{\mathcal{I}_0\cup\mathcal{J}}(y+\gamma_{\mathcal{I}_0}\circ y^a) \\ &= \max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}_0=\emptyset} \varphi_{\mathcal{I}_0\cup\mathcal{J}}(y). \end{aligned} \tag{35}$$

Similarly, one can prove that

$$f^-_{\mathcal{I}_0}(y) \geq \min_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}_0=\emptyset} \varphi_{\mathcal{I}_0\cup\mathcal{J}}(y). \tag{36}$$

4) By (35),

$$\begin{aligned} f^+(y) &= \max_{|\mathcal{I}|=l} f^+_{\mathcal{I}}(y) \leq \max_{|\mathcal{I}|=l} \max_{|\mathcal{J}|=l, \mathcal{J}\cap\mathcal{I}=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}}(y) \\ &= \max_{|\mathcal{K}|=2l} \varphi_{\mathcal{K}}(y). \end{aligned}$$

Similarly,

$$f^-(y) \geq \min_{|\mathcal{K}|=2l} \varphi_{\mathcal{K}}(y).$$

Thus, (30) holds by Theorem 1,

■

Now we are ready to prove Theorem 3.

Proof of Theorem 3: Let $\{\varphi^*_{\mathcal{K}}\}$ be the optimal solution of (26). By Lemma 3, the expected cost of the estimator f^* satisfies:

$$\mathcal{C}(f^*) \leq \mathbb{E}\left[\max_{|\mathcal{K}|=2l} c(x-\varphi^*_{\mathcal{K}}(y))\right].$$

On the other hand, by Lemma 2, for any estimator f, there exists a set of local estimators $\{\varphi_{\mathcal{K}}\}$, such that

$$\mathcal{C}(f) \geq \mathbb{E}\left[\max_{|\mathcal{K}|=2l} c(x-\varphi_{\mathcal{K}}(y))\right].$$

Since $\{\varphi^*_{\mathcal{K}}\}$ is the optimal solution of (26),

$$\mathbb{E}\left[\max_{|\mathcal{K}|=2l} c(x - \varphi^*_{\mathcal{K}}(y))\right] \leq \mathbb{E}\left[\max_{|\mathcal{K}|=2l} c(x - \varphi_{\mathcal{K}}(y))\right],$$

which implies that

$$\mathcal{C}(f^*) \leq \mathcal{C}(f).$$

Therefore, we can conclude the proof. ■

By Theorem 3, in order to derive the optimal estimator f^*, one needs to solve the optimization problem (26), the convexity of which is proved by the following theorem:

Theorem 4: Let $\{\varphi_{\mathcal{K}}\}$, $\{\psi_{\mathcal{K}}\}$ be two sets of local estimators. Define a new set of local estimators $\{\rho_{\mathcal{K}}\}$, such that for every $\mathcal{K}$,

$$\rho_{\mathcal{K}} = \alpha\varphi_{\mathcal{K}} + \beta\psi_{\mathcal{K}},$$

where $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$. If the cost function c is convex in e, then the following inequality holds:

$$\begin{aligned}&\mathbb{E}\left[\max_{|\mathcal{K}|=2l} c(x - \rho_{\mathcal{K}}(y))\right]\\ &\leq \alpha\mathbb{E}\left[\max_{|\mathcal{K}|=2l} c(x - \varphi_{\mathcal{K}}(y))\right] + \beta\mathbb{E}\left[\max_{|\mathcal{K}|=2l} c(x - \psi_{\mathcal{K}}(y))\right].\end{aligned} \tag{37}$$

Proof: Since we restrict the cost function c to be convex,

$$c(x - \rho_{\mathcal{K}}(y)) \leq \alpha c(x - \varphi_{\mathcal{K}}(y)) + \beta c(x - \psi_{\mathcal{K}}(y)).$$

Thus,

$$\begin{aligned}&\max_{|\mathcal{K}|=2l} c(x - \rho_{\mathcal{K}}(y)) \leq \max_{|\mathcal{K}|=2l}\left[\alpha c(x - \varphi_{\mathcal{K}}(y)) + \beta c(x - \psi_{\mathcal{K}}(y))\right]\\ &\leq \alpha \max_{|\mathcal{K}|=2l} c(x - \varphi_{\mathcal{K}}(y)) + \beta \max_{|\mathcal{K}|=2l} c(x - \psi_{\mathcal{K}}(y)).\end{aligned}$$

By taking the expectation on both sides, we can finish the proof. ■

By Theorem 4, we know that the objective function of (26) is a convex functional with respect to $\varphi_{\mathcal{K}}$ provided that c is convex. If we further assume that $\varphi_{\mathcal{K}}$ belongs to some finite dimensional linear space, e.g., all polynomials of degree less than k, then $\varphi_{\mathcal{K}}$ can be written as

$$\varphi_{\mathcal{K}} = a_1\phi_1 + a_2\phi_2 + \cdots + a_k\phi_k,$$

where a_is are scalars and $\{\phi_1, \dots, \phi_k\}$ is the basis of the space, then (26) becomes a convex optimization problem. As a result, algorithms such as interior point method [18] can be used to find the optimal a_is and thus the optimal $\varphi_{\mathcal{K}}$.

It is also worth noticing that even if we could derive the optimal $\varphi_{\mathcal{K}}$, (25) is still computationally hard. In particular, to determine $f(y)$, we need to compute all $\binom{m}{2l}$ different $\varphi_{\mathcal{K}}(y)$s, which could be a huge burden if m is large. In the next section, we consider two special cases, under which (25) can be simplified and thus computed efficiently.

V. Special Cases

In this section, we prove that under certain conditions, (25) can be simplified. We first consider the case where the estimator f is symmetric and monotone, which is given by the following definition:

Definition 8: $f(y)$ is symmetric if

$$f(y_1, \dots, y_m) = f(y_{i_1}, \dots, y_{i_m}),$$

for any permutation $(i_1, \dots, i_m)$ of $\mathcal{S}$.

Definition 9: $f(y)$ is monotonically increasing if the following inequality hold:

$$f(y_1, \dots, y_m) \geq f(y_1', \dots, y_m'), \text{ if } y_i \geq y_i', \forall i.$$

f is monotonically decreasing if $-f$ is monotonically increasing. f is monotone if it is either monotonically increasing or monotonically decreasing.

Remark 5: The symmetry assumption is reasonable if the joint distribution of x and y is also symmetric on y, which implies that the sensors are statistically identical.

We further define the following function:

Definition 10: Define the function $\text{Med}_l^m : \mathbb{R}^m \to \mathbb{R}^{m-2l}$ as a symmetric function, which satisfies[3]

$$\text{Med}_l^m([y_1, \dots, y_m]^T) \triangleq [y_{l+1}, \dots, y_{m-l}]^T, \tag{38}$$

when $y_1 \leq \cdots \leq y_m$.

[3] Due to symmetry, we only need to define the function when $y_1 \leq \cdots \leq y_m$.

Remark 6: The Med_l^m function can be computed by removing the largest l measurements and smallest l measurements. In particular, if $m = 2l + 1$, then $\text{Med}_l^m(y)$ is simply the median of y. The following theorem characterizes the optimal symmetric and monotone estimator f:

Theorem 5: If $l < m/2$, then the optimal symmetric and monotone estimator is of the following form:

$$f^*(y) = \Phi^*(\text{Med}_l^m(y)), \tag{39}$$

where $\Phi^* : \mathbb{R}^{m-2l} \to \mathbb{R}$ is the solution of the following optimization problem:

$$\underset{\Phi \text{ symmetric and monotone}}{\text{minimize}} \quad \mathbb{E}\left\{ \max_{|\mathcal{K}|=2l} c\left[x - \Phi(\text{Proj}_{\mathcal{K}^c}(y))\right] \right\}. \tag{40}$$

Before proving Theorem 5, we need the following lemma:

Lemma 4: Let $\Phi : \mathbb{R}^{m-2l} \to \mathbb{R}$ be a symmetric and monotone function. If the local estimator $\varphi_{\mathcal{K}}$ satisfies

$$\varphi_{\mathcal{K}}(y) = \Phi(\text{Proj}_{\mathcal{K}^c}(y)),$$

then the corresponding estimator $f(y)$ given by:

$$f(y) = \min_{|\mathcal{I}|=l} \left[\max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}}(y) \right]$$

satisfies the following equality:

$$f(y) = \Phi(\text{Med}_l^m(y)). \tag{41}$$

Proof: We will assume that Φ is monotonically increasing, since the case where Φ is monotonically decreasing can be proved by similar arguments. We can further assume that $y_1 \leq \cdots \leq y_m$ due to symmetry.

Let us define $\mathcal{I}_0 = \{m - l + 1, \ldots, m\}$ and $\mathcal{J}_0 = \{1, \ldots, l\}$. It is clear that

$$\begin{aligned} f(y) &= \min_{|\mathcal{I}|=l} \left[\max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}}(y) \right] \\ &\leq \max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}_0=\emptyset} \varphi_{\mathcal{I}_0\cup\mathcal{J}}(y) \end{aligned}$$

Now by monotonicity of Φ, we know that

$$\max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}_0=\emptyset} \varphi_{\mathcal{I}_0\cup\mathcal{J}}(y) = \varphi_{\mathcal{I}_0\cup\mathcal{J}_0}(y) = \Phi(\text{Med}_l^m(y)),$$

which implies that $f(y) \leq \Phi(\text{Med}_l^m(y))$. On the other hand, we know that

$$\begin{aligned} f(y) &= \min_{|\mathcal{I}|=l} \left[\max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}}(y) \right] \\ &\geq \max_{|\mathcal{J}|=l} \left[\min_{|\mathcal{I}|=l,\, \mathcal{I}\cap\mathcal{J}=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}}(y) \right] \\ &\geq \min_{|\mathcal{I}|=l,\, \mathcal{I}\cap\mathcal{J}_0=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}_0}(y). \end{aligned}$$

By monotonicity of φ, we have that

$$\min_{|\mathcal{I}|=l,\, \mathcal{I}\cap\mathcal{J}_0=\emptyset} \varphi_{\mathcal{I}\cup\mathcal{J}_0}(y) = \varphi_{\mathcal{I}_0\cup\mathcal{J}_0}(y) = \Phi(\text{Med}_l^m(y)),$$

which implies that $f(y) \geq \Phi(\text{Med}_l^m(y))$. Thus $f(y) = \Phi(\text{Med}_l^m(y))$, which concludes the proof.

■

We are now ready to prove Theorem 5:

Proof: Denote the local estimators for the optimal symmetric and monotone estimator f^* as $\varphi^*_{\mathcal{K}}$. Recall that from the definition of local estimators, there exists $\Phi^*_{\mathcal{K}}$, such that the following equality holds:

$$\varphi^*_{\mathcal{K}}(y) = \Phi^*_{\mathcal{K}}(\text{Proj}_{\mathcal{K}^c}(y)).$$

By similar argument as in the proof of Lemma 2, we know that $\Phi^*_{\mathcal{K}}$ is identical, symmetric and monotone. Therefore, by Lemma 4, we can conclude the proof. ■

Now we consider the optimal estimator for the boundary case, where $m = 2l + 1$.

Theorem 6: If $m = 2l + 1$, then the optimal estimator $f^*(y)$ can be expressed as

$$f^*(y) = \text{Median}(\tau_1^*(y_1), \ldots, \tau_m^*(y_m)), \tag{42}$$

where $\tau_i : \mathbb{R} \to \mathbb{R}$ and solves the following optimization problem:

$$\underset{\tau_1,\ldots,\tau_m}{\text{minimize}} \;\; \mathbb{E}\left[\max_i \; c(x - \tau_i(y_i)) \right]. \tag{43}$$

Proof: Let f^* be the optimal estimator with local estimator $\varphi^*_{\mathcal{K}}$s. From the definition of local estimators, there exists $\Phi^*_{\mathcal{K}}$, such that $\varphi^*_{\mathcal{K}}(y) = \Phi^*_{\mathcal{K}}(\text{Proj}_{\mathcal{K}^c}(y))$. Since $m = 2l+1$, we could define

$$\tau_i^*(y_i) = \Phi^*_{\mathcal{K}}(y_i), \tag{44}$$

where $\{i\} = \mathcal{K}^c$. Now consider a function $\Phi : \mathbb{R} \to \mathbb{R}$, such that $\Phi(t) = t$ and the corresponding local estimators

$$\varphi_{\mathcal{K}}(y) = \Phi(\mathrm{Proj}_{\mathcal{K}^c}(y)) = \mathrm{Proj}_{\mathcal{K}^c}(y).$$

Define an estimator g as

$$g(y) \triangleq \min_{|\mathcal{I}|=l} \left[\max_{|\mathcal{J}|=l,\, \mathcal{J}\cap\mathcal{I}=\emptyset} \mathrm{Proj}_{(\mathcal{I}\cup\mathcal{J})^c}(y) \right].$$

By Lemma 4 and the fact that $m = 2l + 1$, $g(y) = \mathrm{Med}_l^m(y) = \mathrm{Median}(y)$. Furthermore, it is easy to verify that

$$f^*(y) = g(\tau_1^*(y_1), \ldots, \tau_m^*(y_m)).$$

Hence, $f^*(y) = \mathrm{Median}(\tau_1^*(y_1), \ldots, \tau_m^*(y_m))$. ■

Remark 7: It is worth noticing that given the optimal Φ^* or $\{\tau_i^*\}$, the computational complexity of (39) and (42) is $O(m \log m)$, which is the complexity of sorting m numbers.

VI. Numerical Examples: Gaussian Case

We assume the state $x \sim \mathcal{N}(0, 1)$ is normal distributed with zero mean and unit variance. The measurements equation for sensor i is given as

$$y_i = x + v_i,$$

where $v_i \sim \mathcal{N}(0, 1)$ is also normal distributed with zero mean and unit variance. We further assume that x and v_is are independent. The cost function $c(e) = e^2$. We first consider the case where $m = 3\, l = 1$. By Theorem 6, the optimal $f(y)$ is of the following form

$$f(y) = \mathrm{Median}(\tau_1(y_1), \tau_2(y_2), \tau_3(y_3))$$

We seek to find the optimal τ_is over all polynomials of degree less than 5. We will approximate $\mathcal{C}(f)$ by Monte-Carlo method. To be specific, we first randomly generate a set of $(x^{(k)}, y_1^{(k)}, y_2^{(k)}, y_3^{(k)})$, where $k = 1, \ldots, N$, based on the Gaussian assumption. The expected cost $\mathcal{C}(f)$ is then approximated by

$$\mathcal{C}(f) \approx \frac{1}{N} \sum_{k=1}^{N} \left[\max_{i=1,2,3} (x^{(k)} - \tau_i(y_i^{(k)}))^2 \right]$$

We use gradient descent to find the optimal a_ks, such that $\tau_i(y_i) = \sum_{k=0}^{4} a_k y_i^k$. In this simulation, we choose the size of the training set to be $N = 100000$. The optimal τ_i found by gradient descent is given by $\tau_i(y_i) = 0.29y_i$, and the worst case cost is $\mathcal{C}(f) = 0.91$. On the other hand, if no sensor is compromised, then the optimal estimator is $\hat{x} = (y_1 + y_2 + y_3)/4$, which has a mean squared error of 0.25. Thus, there is a more than threefold increase in the expected cost caused by the attacker. Further, the expected cost of a constant estimator $\hat{x} = \mathbb{E}x = 0$ is 1. Hence, it can be seen that one compromised sensor can potentially cause a large degradation in the estimation performance.

It is also worth noticing that the optimal τ_is are linear (within the numerical error). However, as we only search over the space of polynomials of degree less than 5, it is still an open problem to find the true optimal τ_is.

Next we consider $\mathcal{C}(f)$ when m increases and $l = 1$ is fixed. By Theorem 5, the optimal estimator is of the following form $f^*(y) = \Phi^*(\text{Med}_l^m(y))$. In this paper, we only consider a symmetric and linear Φ. In other words, we only consider $f^*(y)$ of the following form: $f(y) = a\mathbf{1}^T\text{Med}_l^m(y)$, where $\mathbf{1}$ is an all one vector of proper dimension and a is a constant. The optimal $\mathcal{C}(f)$ for estimators with a symmetric and linear Φ function is listed in Table I.

m	3	4	5	6	7	8	9	10
$\mathcal{C}(f)$	0.91	0.82	0.66	0.57	0.49	0.43	0.38	0.33

TABLE I

OPTIMAL $\mathcal{C}(f)$ FOR ESTIMATORS WITH A SYMMETRIC AND LINEAR Φ WHEN $l = 1$.

VII. CONCLUSION AND FUTURE WORK

In this paper we consider the problem of designing estimator able to minimize the mean squared error in the presence of l 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 estimator against all possible attacker's strategies. We show that if the attacker can manipulate at least half of the m measurements ($l \geq m/2$) then the optimal worst-case estimator 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 ($l < m/2$), we show that the optimal estimator is based on $\binom{m}{2l}$ local estimators. We further prove that such an estimator can be reduced into simpler forms for two special cases. We are planning to expand our research to the case of multidimensional state estimation in the future.

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] 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.

[5] R. A. Maronna, D. R. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. Wiley, 2006.

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

[7] F. R. Hampel, “A general qualitative definition of robustness,” *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1887–1896, Dec 1971.

[8] D. L. Donoho and P. J. Huber, “The notion of breakdown point,” *A Festschrift for Erich L. Lehmann*, pp. 157–184, 1983.

[9] F. R. Hampel, “The influence curve and its role in robust estimation,” *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974.

[10] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control*. Prentice Hall New Jersey, 1996, vol. 272.

[11] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[12] Y. Liu, M. Reiter, and P. Ning, “False data injection attacks against state estimation in electric power grids,” in *Proceedings of the 16th ACM conference on Computer and communications security*, 2009.

[13] H. Sandberg, A. Teixeira, and K. H. Johansson, “On security indices for state estimators in power networks,” in *First Workshop on Secure Control Systems*, 2010.

[14] L. Xie, Y. Mo, and B. Sinopoli, “Integrity data attacks in power market operations,” *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, 2011.

[15] Y. Mo and B. Sinopoli, “Robust estimation in the presence of integrity attacks,” in *52nd IEEE Conference on Decision and Control*, 2013, p. submitted.

[16] Y. Mo, J. Hespanha, and B. Sinopoli, “Resilient detection in the presence of integrity attacks,” *Signal Processing, IEEE Transactions on*, vol. 62, no. 1, pp. 31–43, Jan 2014.

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

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.