# SECURE PARAMETER ESTIMATION: FUNDAMENTAL TRADEOFFS

*Saurabh Sihag and Ali Tajer*

Electrical, Computer, and Systems Engineering Department,
Rensselaer Polytechnic Institute, Troy, NY, USA

**ABSTRACT**

The problem of parameter estimation in an adversarial setting, in which an active adversary might decide to compromise the data for the purpose of subverting the estimation decisions, is considered. Forming secure estimation decisions entails two intertwined inference decisions. Specifically, on one hand, deciding whether the data is compromised, like any detection decision, is never perfect. On the other hand, missing any attack translates to degradation in the estimation quality. Based on these two observations, the paper aims to characterize the interplay between two figures of merit $q$ and $\beta$, where $q$ captures how much estimation quality degrades when the objective is to miss the presence of an attacker with a probability not exceeding $\beta$. The paper characterizes the optimal decision rules and compares the results with the existing literature.

***Index Terms*—** Combined estimation and detection, parameter estimation, secure inference.

## 1. INTRODUCTION

### 1.1. Overview

Consider the canonical parameter estimation problem in a sensor network consisting of $n$ sensors. Each sensor $i \in \{1, \ldots, n\}$ monitors a stochastic scalar $X \in \mathbb{R}$ and reports it to a fusion center (FC). The measurement reported by sensor $i \in \{1, \ldots, n\}$ is denoted by

$$Y_i = h_i X + N_i \ , \tag{1}$$

where $h_i$ is the gain corresponding to the channel between sensor $i$ and the FC, and $N_i$ accounts for the additive channel noise. The FC aims to form an estimate of $X$ based on the measurements.

The potential presence of an active adversary introduces a new decision dimension in the parameter estimation problem, which does not exist in the attack-free settings. Specifically, on the one hand, compromising the measurements alters the stochastic model of the measurements, and on the other hand, the measurements model determines the optimal structure of the estimator. Hence, along with forming an estimate for $X$, the FC should also make a detection decision regarding whether the measurements are compromised.

Based on this observation, in this paper, we introduce and analyze a framework for secure parameter estimation, which is formalized based on the premise that detecting and countering the adversaries is never perfect. Any uncertainty about the presence of the adversaries translates into an uncertainty about the model governing the measurements, which, consequently, leads to degradation in the fidelity of the estimate that can be achieved in an attack-free setting. In order to quantify this interplay, we call an estimator $(q, \beta)$-secure if its estimation quality is larger than that of the attack-free setting by a factor $q \geq 1$, while it misses at most $\beta \in (0, 1]$ of the attacks. Based on this notion, we quantify the fundamental tradeoff between $q$ and $\beta$, and characterize the decision rules that achieve the optimal tradeoff. The analysis demonstrates that the estimation and detection decisions are strongly coupled and should be formed jointly.

This research was supported in part by the U. S. National Science Foundation under Grants ECCS-1455228 and the CAREER Award ECCS-1554482.

### 1.2. Related Studies

Of particular relevance to the scope of our paper is the approach in [1], which considers parameter estimation in a network consisting of a secured sensor and a vulnerable sensor. A two-step detection-driven approach is designed, in which in the first step, a decision is made about whether the vulnerable sensor has been compromised, and based on that, a decision is formed based on the measurements provided by the secured sensor (when the adversary is deemed to be active), or based on the measurements provided by both sensors (when the adversary is deemed to be inactive). We provide a case study that compares the qualities of the decisions made based on the joint designs of the detector and estimator in this paper to those of [1].

The adversarial setting defined in this paper is also related to the Byzantine attack models in sensor networks. In [2], the impact of Byzantine attacks on inferences in sensor networks, as well as relevant mitigation strategies are discussed. Detection-driven estimation strategies in the presence of Byzantine attacks (i.e., the estimation step is preceded by a Byzantine attack detection step) are developed in [3–6], in which multiple sensors send quantized information to the FC, and a subset of the sensors are compromised with pre-specified probabilities, as a result of which the information bits sent by the sensors are randomly flipped. Strategies for isolating the compromised nodes are analyzed in [7–9], which are different from the scope of our work, which places the emphasis on the estimation routine.

Unlike parameter estimation in stationary systems, which is the focus of this paper, the problem of estimating the state of linear *dynamical* systems in an adversarial setting in the context of cyber-physical systems has been studied more extensively (c.f. [10–16]). The studies more relevant to the scope of this paper include robust estimation of the states in dynamic systems studied in [11], [15], and [16]. Also, inference in dynamical systems from the perspective of the adversaries is studied in [17], where the bounds on degradation in estimation performance in a single-sensor network with degrees of stealthiness of the attacker is characterized.

Characterizing the optimal secure estimator necessitates forming an entangled detection decision about whether the adversary is active, which cannot be reduced to independent estimation and detection routines as such decoupling approaches lead to sub-optimal performance ([18] and [19]). In the detection-driven approach, initially a detection decision is made, which reduces the problem to a purely estimation one, at which point an estimator is optimized (e.g., Neyman-Pearson detection followed by Bayesian estimation). In the estimation-driven approach, first an estimate for the unknown parameter is made, which is used to circumvent the uncertainties associated with the unknown parameter, and then a detection decision is made (e.g., the generalized likelihood ratio test). Both these approaches admit optimality properties only at the asymptote of having an infinite number of measurements.

Recent development on the non-asymptotically optimal com-

bined detection and estimation rules are investigated in [20] and [21], where [20] considers a binary hypothesis testing problem in which under one of the hypotheses the measurements follow a composite model and involve an unknown parameter that is of interest to be estimated. Extensions to having composite models under both hypotheses, where the parameters to be estimated under the different hypotheses are distinct and independent in nature, are studied in [21].

## 2. SYSTEM MODEL

We first provide a canonical model for parameter estimation in an attack-free setting. This furnishes a baseline for modeling and assessing the fundamental performance limits in the adversarial setting.

### 2.1. Attack-free Setting

Consider the problem of estimating a scalar random parameter $X$ with known probability density function (pdf) $\pi$ from $n \in \mathbb{N}$ noisy observations collected by a fusion center (FC) from $n$ independent sensors. The noisy observation collected from sensor $i \in \{1, \ldots, n\}$ is denoted by

$$Y_i = h_i X + N_i \ , \tag{2}$$

where $h_i \in \mathbb{R}$ captures the fixed gain of the channel linking sensor $i \in \{1, \ldots, n\}$ to the FC and is known to the FC[1]. The independent and identically distributed (i.i.d) random variables $\{N_i \colon i \in \{1, \ldots, n\}\}$ account for the measurement noise. We denote the joint pdf of the collected measurements $\boldsymbol{Y} \triangleq [Y_1, \ldots, Y_n]$ by $f_0$. Based on measurements $\boldsymbol{Y}$, the FC forms an estimate $\hat{X}(\boldsymbol{Y})$ for $X$. We define a non-negative cost function $\mathsf{C}(X, U)$ to measure the fidelity of any generic estimator $U$. A popular cost function based on the mean-squared error criterion is $\mathsf{C}(X, U) = \|X - U\|^2$. We also define the average posterior cost function when $\boldsymbol{Y}$ is distributed according to $f_0$ as

$$\mathsf{C}_{\mathrm{p},0}(U \mid \boldsymbol{Y}) \triangleq \mathbb{E}_0 \left[ \mathsf{C}(X, U) \mid \boldsymbol{Y} \right] \ , \tag{3}$$

where the expectation is with respect to $X$ for given $\boldsymbol{Y}$. Therefore, the optimal estimate, which minimizes $\mathsf{C}_{\mathrm{p},0}(U \mid \boldsymbol{Y})$, is

$$\hat{X}_0(\boldsymbol{Y}) \triangleq \arg\inf_U \mathsf{C}_{\mathrm{p},0}(U \mid \boldsymbol{Y}) \ . \tag{4}$$

We define $\hat{\mathsf{C}}_{\mathrm{p},0}$ as the minimum cost in an attack-free setting, i.e.,

$$\hat{\mathsf{C}}_{\mathrm{p},0}(\boldsymbol{Y}) \triangleq \inf_U \mathsf{C}_{\mathrm{p},0}(U \mid \boldsymbol{Y}) \ . \tag{5}$$

Also, for any estimator $U$, we define the average cost function

$$J_0(U) \triangleq \mathbb{E}_0[\mathsf{C}(X, U)] \ , \tag{6}$$

where the expectation is taken over X and Y.

### 2.2. Adversarial Setting

In an adversarial setting, an adversary might attempt to degrade the estimation quality of $X$ by corrupting the observations received by the FC. We assume that the adversary can attack at most one sensor at any instant [2]. Denote the probability that sensor $i$ is compromised by $\epsilon_i$, and the probability that no sensor is compromised by $\epsilon_0$, where we have $\sum_{i=0}^{n} \epsilon_i = 1$ .
By defining $Z_i$ as the disturbance imposed by the adversary on the measurement from sensor $i \in \{1, \ldots, n\}$, the observation model in (2) changes to:

$$Y_i = h_i X + N_i + Z_i \ , \tag{7}$$

and the joint pdf of $\boldsymbol{Y}$ changes from $f_0$ to $f_i$. Similar to the attack-free setting, we define the average posterior cost function when $\boldsymbol{Y}$ is distributed according to $f_i$ as

$$\mathsf{C}_{\mathrm{p},i}(U \mid \boldsymbol{Y}) \triangleq \mathbb{E}_i \left[ \mathsf{C}(X, U) \mid \boldsymbol{Y} \right] \ . \tag{8}$$

Similarly, the optimal estimate of $X$ when an attack is deemed to exist on sensor $i$ is given by

$$\hat{X}_i(\boldsymbol{Y}) \triangleq \arg\inf_U \mathsf{C}_{\mathrm{p},i}(U \mid \boldsymbol{Y}) \ , \tag{9}$$

and the optimal estimation cost is

$$\hat{\mathsf{C}}_{\mathrm{p},i}(\boldsymbol{Y}) \triangleq \inf_U \mathsf{C}_{\mathrm{p},i}(U \mid \boldsymbol{Y}) \ . \tag{10}$$

## 3. PROBLEM FORMULATION

The adversary may or may not choose to inject an attack, and therefore, it is imperative for the FC to also form a decision about the presence of an attack. The problem of detecting an attack and providing an estimate for $X$ can be modeled as an $(n+1)$-ary composite hypothesis testing problem given by:

$$\begin{aligned} \mathsf{H}_0 \ : \quad & \boldsymbol{Y} \sim f_0(\boldsymbol{Y} \mid X), \ \text{ with } X \sim \pi(X) \\ \mathsf{H}_i \ : \quad & \boldsymbol{Y} \sim f_i(\boldsymbol{Y} \mid X), \ \text{ with } X \sim \pi(X) \end{aligned} \ , \tag{11}$$

where $\mathsf{H}_0$ is the hypothesis corresponding to the attack-free setting, and $\mathsf{H}_i$ is the hypothesis corresponding to experiencing an attack on sensor $i \in \{1, \ldots, n\}$. We solve the estimation and detection problems jointly to design an optimum estimate for $X$ in the adversarial setting.

### 3.1. Definitions

Define $\mathsf{T} \in \{\mathsf{H}_0, \ldots, \mathsf{H}_n\}$ as the true hypothesis, and $\mathsf{D} \in \{\mathsf{H}_0, \ldots, \mathsf{H}_n\}$ as the decision about the hypothesis. Therefore, $\mathbb{P}(\mathsf{D} = \mathsf{H}_i \mid \mathsf{T} = \mathsf{H}_j)$ captures the likelihood of deciding in favour of $\mathsf{H}_i$ while the true hypothesis is $\mathsf{H}_j$, for $i, j \in \{0, \ldots, n\}$. Define $\mathsf{P}_{\mathrm{md}}$ as the probability of missing the attack when it exists, i.e.,

$$\mathsf{P}_{\mathrm{md}} \triangleq \mathbb{P}(\mathsf{D} = \mathsf{H}_0 \,|\, \mathsf{T} \neq \mathsf{H}_0) \ . \tag{12}$$

By invoking $\mathbb{P}(\mathsf{T} = \mathsf{H}_i) = \epsilon_i$, we readily find

$$\mathsf{P}_{\mathrm{md}} \triangleq \frac{1}{1 - \epsilon_0} \sum_{i=1}^{n} \epsilon_i \cdot \mathbb{P}(\mathsf{D} = \mathsf{H}_0 \,|\, \mathsf{T} = \mathsf{H}_i) \ . \tag{13}$$

Also, let $\mathsf{P}_{\mathrm{fa}}$ denote the probability of false alarms, which is given by

$$\mathsf{P}_{\mathrm{fa}} \triangleq \sum_{i=1}^{n} \mathbb{P}(\mathsf{D} = \mathsf{H}_i \,|\, \mathsf{T} = \mathsf{H}_0) \ . \tag{14}$$

[1] For the convenience in notations we are assuming that each sensor generates one measurement, and the results can be readily generalized to having multiple measurements per sensor.

[2] Generalization to attacks on multiple sensors follows the same line of analysis, albeit with an increase in the dimension of the parameters to be introduced.

We use a randomized test $\boldsymbol{\delta}= [\delta_0(\boldsymbol{Y}), \ldots, \delta_n(\boldsymbol{Y})]$ to design the decision rule for discerning the correct hypothesis, where $\delta_i(\boldsymbol{Y})$ is defined as the probability of deciding in favour of $\mathsf{H}_i$. Clearly $\delta_i(\boldsymbol{Y}) \in [0,1]$ and $\sum_{i=0}^{n} \delta_i(\boldsymbol{Y}) = 1$ . For any generic estimator of $X$ under the hypothesis $\mathsf{H}_i$, which we denote by $U_i$, the estimation cost is $\mathsf{C}(X, U_i)$, and accordingly define $\boldsymbol{U} \triangleq [U_0, \ldots, U_n]$. The estimate of $X$ is different under different hypotheses and the estimation cost $\mathsf{C}(X, U_i)$ is relevant only when deciding in favour of $\mathsf{H}_i$. To characterize the overall estimation cost in the adversarial setting, let $J_i(\delta_i, U_i)$ denote the average estimation cost given that the decision is in favor of $\mathsf{H}_i$. For given $\delta_i(\boldsymbol{Y})$ and $U_i$, define

$$J_i(\delta_i, U_i) \triangleq \mathbb{E}\left[\mathsf{C}(X, U_i) \,|\, \mathsf{D}=\mathsf{H}_i\right] \ , \tag{15}$$

where the expectation is with respect to $X$ and $\boldsymbol{Y}$. The overall estimation cost, $J(\boldsymbol{\delta}, \boldsymbol{U})$ is defined as the maximum of the average costs $J_i(\delta_i, U_i)$, and is given by

$$J(\boldsymbol{\delta}, \boldsymbol{U}) \triangleq \max_{i \in \{0,\ldots,n\}} \{J_i(\delta_i, U_i)\} \ . \tag{16}$$

### 3.2. Secure Parameter Estimation

The quality of estimation routine hinges on successfully detecting the underly data models $\{\mathsf{H}_i : i = 0, \ldots, n\}$, but detecting the presence of an attack is never perfect because of noisy measurements. Consequently, any potential presence of an attack degrades the estimation quality. In this paper, we aim to characterise the fundamental interplay between the qualities of of estimating $X$ and detecting the presence of an attacker. For this purpose, we provide the following definition.

**Definition 1.** *We call an estimation procedure $(q, \beta)$-secure, if the estimation cost is larger than that of the attack-free setting by factor $q \geq 1$, i.e.,*

$$q \triangleq \frac{\min_{(\boldsymbol{\delta}, \boldsymbol{U})} J(\boldsymbol{\delta}, \boldsymbol{U})}{\min_U J_0(U)} \ , \tag{17}$$

*while we aim to miss at most $\beta \in (0, 1]$ fraction of the attacks, i.e., $\mathsf{P}_{\mathsf{md}} \leq \beta$.*

Based on this definition, we specifically aim to characterize the region of all simultaneously achievable values of $q$ and $\beta$. Such a region can be found as the solution to

$$\mathcal{P}_1(\beta) \triangleq \begin{cases} \min_{(\boldsymbol{\delta}, U)} & J(\boldsymbol{\delta}, \boldsymbol{U}) \\ \text{s.t.} & \mathsf{P}_{\mathsf{md}} \leq \beta \end{cases} \ . \tag{18}$$

**Remark 1.** *Note that minimizing q defined in (17) is equivalent to minimizing $J(\boldsymbol{\delta}, \boldsymbol{U})$, since the average cost function in the attack-free setting, i.e., $\min_U J_0(U)$ becomes a constant for the given distributions $f_0$ and $\pi$.*

Problem $\mathcal{P}_1(\beta)$ is concerned with only the likelihood of missing the attacks. Sometimes it can be of interest to also control the likelihood of false alarm, which happens when the detector decides that an attack exists in an attack free model. To further accommodate such settings, we provide the following definition.

**Definition 2.** *We call an estimation procedure $(q, \alpha, \beta)$-secure if it is $(q, \beta)$-secure and the rate of false alarms is below $\alpha \in (0, 1]$, i.e., $\mathsf{P}_{\mathsf{fa}} \leq \alpha$.*

All achievable secure strategies belong to a region characterized by solving

$$\mathcal{P}_2(\alpha, \beta) = \begin{cases} \min_{(\boldsymbol{\delta}, \boldsymbol{U})} & J(\boldsymbol{\delta}, \boldsymbol{U}) \\ \text{s.t.} & \mathsf{P}_{\mathsf{md}} \leq \beta \\ & \mathsf{P}_{\mathsf{fa}} \leq \alpha \end{cases} \ . \tag{19}$$

**Remark 2.** *Clearly, we have the connection $\mathcal{P}_1(\beta) = \mathcal{P}_2(1, \beta)$.*

**Remark 3** (Feasibility)**.** *In principle, probabilities $\mathsf{P}_{\mathsf{md}}$ and $\mathsf{P}_{\mathsf{fa}}$ cannot be made arbitrarily small simultaneously. By the Neyman-Pearson theory [22], it can be readily verified that for any given $\alpha$, there exists a value $\beta^*(\alpha)$, which specifies the smallest feasible value for $\beta$.*

## 4. DECISION RULES

In this section, we characterize the optimal decision rules, i.e., the estimators $\{U_i : i \in \{0, \ldots, n\}\}$ and the detectors $\{\delta_i : i \in \{0, \ldots, n\}\}$. We focus on the more general problem $\mathcal{P}_2(\alpha, \beta)$, solving which also provides the solution to $\mathcal{P}_1(\beta) = \mathcal{P}_2(1, \beta)$. Throughout the analysis, we assume that the combination of $\alpha$ and $\beta$ is a feasible combination. We start by explicitly specifying the dependence of $\mathsf{P}_{\mathsf{md}}$ and $\mathsf{P}_{\mathsf{fa}}$ on $\{\delta_i : i \in \{0, \ldots, n\}\}$. By noting that

$$\mathbb{P}(\mathsf{D}=\mathsf{H}_j \,|\, \mathsf{T}=\mathsf{H}_i) = \int_{\boldsymbol{Y}} \delta_j(\boldsymbol{Y}) f_i(\boldsymbol{Y}) d\boldsymbol{Y} \ , \tag{20}$$

and leveraging (13), we have

$$\mathsf{P}_{\mathsf{md}} = \frac{1}{1-\epsilon_0} \sum_{i=1}^{n} \epsilon_i \int_{\boldsymbol{Y}} \delta_0(\boldsymbol{Y}) f_i(\boldsymbol{Y}) d\boldsymbol{Y} \ . \tag{21}$$

Similarly, by noting (20) and based on (14), we have

$$\mathsf{P}_{\mathsf{fa}} = \sum_{i=1}^{n} \int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) f_0(\boldsymbol{Y}) d\boldsymbol{Y} \ . \tag{22}$$

By using expansions in (21) and (22), the secure parameter estimation of interest becomes

$$\mathcal{P}_2(\alpha, \beta) = \begin{cases} \min_{(\boldsymbol{\delta}, \boldsymbol{U})} & J(\boldsymbol{\delta}, \boldsymbol{U}) \\ \text{s.t.} & \frac{1}{1-\epsilon_0} \sum\limits_{i=1}^{n} \epsilon_i \int_{\boldsymbol{Y}} \delta_0(\boldsymbol{Y}) f_i(\boldsymbol{Y}) d\boldsymbol{Y} \leq \beta \\ & \sum\limits_{i=1}^{n} \int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) f_0(\boldsymbol{Y}) d\boldsymbol{Y} \leq \alpha \end{cases} \ . \tag{23}$$

Close scrutiny of (23) indicates that the effect of the estimators $\{U_i : i \in \{0, \ldots, n\}\}$ appear only in the utility function. This allows for breaking the optimization problem in (23) into two subproblems, as formalized in the next theorem.

**Theorem 1.** *The solution of $\mathcal{P}_2(\alpha, \beta)$ can be found by equivalently solving*

$$\mathcal{P}_2(\alpha, \beta) = \begin{cases} \min_{\boldsymbol{\delta}} & \tilde{J}(\boldsymbol{\delta}, \hat{\boldsymbol{X}}) \\ \text{s.t.} & \frac{1}{1-\epsilon_0} \sum\limits_{i=1}^{n} \epsilon_i \int_{\boldsymbol{Y}} \delta_0(\boldsymbol{Y}) f_i(\boldsymbol{Y}) d\boldsymbol{Y} \leq \beta \\ & \sum\limits_{i=1}^{n} \int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) f_0(\boldsymbol{Y}) d\boldsymbol{Y} \leq \alpha \end{cases} \ , \tag{24}$$

$$\textit{where} \qquad \tilde{J}(\boldsymbol{\delta}, \hat{\boldsymbol{X}}) = \min_{\boldsymbol{U}} J(\boldsymbol{\delta}, \boldsymbol{U}) \ , \tag{25}$$

$$\textit{and} \qquad \hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{U}} J(\boldsymbol{\delta}, \boldsymbol{U}) \ . \tag{26}$$

By using the result of Theorem 1, next we provide the design of the optimal estimators.

**Theorem 2** (Secure Estimator)**.** *The optimal secure estimator under model* $\mathsf{H}_i$ *is given by*

$$\hat{X}_i(\boldsymbol{Y}) = \arg\inf_{U_i} \mathsf{C}_{\mathrm{p},i}(U_i \mid \boldsymbol{Y}) \ , \tag{27}$$

*and correspondingly, the cost function* $\tilde{J}(\boldsymbol{\delta}, \hat{\boldsymbol{X}})$ *is given by*

$$\tilde{J}(\boldsymbol{\delta}, \hat{\boldsymbol{X}}) = \max_i \left\{ \frac{\int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) \hat{\mathsf{C}}_{\mathrm{p},i}(\boldsymbol{Y}) f_i(\boldsymbol{Y}) d\boldsymbol{Y}}{\int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) f_i(\boldsymbol{Y}) d\boldsymbol{Y}} \right\} . \tag{28}$$

Given the optimal estimators $\hat{\boldsymbol{X}}$, the corresponding optimal decision rules are obtained by solving the minimization problem in (24).

**Theorem 3.** *The optimal decision rules are given by:* $\delta_{i^*}(\boldsymbol{Y}) = 1$ *and* $\delta_i(\boldsymbol{Y}) = 0$ *for* $i \neq i^*$*, where*

$$i^* = \underset{i \in \{0,\ldots,n\}}{\operatorname{argmin}} \ A_i \ ,$$

*and*

$$A_0 \triangleq \ell_0 f_0(\boldsymbol{Y})(\hat{\mathsf{C}}_{\mathrm{p},0}(\boldsymbol{Y}) - \mathcal{P}_2(\alpha,\beta)) + \ell_{n+1} \sum_{i=1}^{n} \frac{\epsilon_i f_i(\boldsymbol{Y})}{1-\epsilon_0} \ ,$$

*and for* $i \in \{1,\ldots,n\}$*,*

$$A_i \triangleq \ell_i f_i(\boldsymbol{Y})(\hat{\mathsf{C}}_{\mathrm{p},i}(\boldsymbol{Y}) - \mathcal{P}_2(\alpha,\beta)) + \ell_{n+2} f_0(\boldsymbol{Y}) \ .$$

*Non-negative constants* $\ell_i$*, for* $i \in \{0,\ldots,n+2\}$ *are the Lagrange multipliers selected such that* $\sum_{i=0}^{n+2} \ell_i = 1$ *and the constraints in the following convex optimization problem (that is equivalent to* (24)*) are satisfied.*

$$\mathcal{P}_2(\alpha,\beta) \triangleq \begin{cases} \min_{\boldsymbol{\delta}} & u \\ \text{s.t.} & \int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) f_i(\boldsymbol{Y})(\hat{\mathsf{C}}_{\mathrm{p},i}(\boldsymbol{Y}) - u) d\boldsymbol{Y} \leq 0 \ , \\ & \forall\, i \in \{0,\ldots,n\} \\ & \frac{1}{1-\epsilon_0} \sum_{i=1}^{n} \epsilon_i \int_{\boldsymbol{Y}} \delta_0(\boldsymbol{Y}) f_i(\boldsymbol{Y}) d\boldsymbol{Y} \leq \beta \ , \\ & \sum_{i=1}^{n} \int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) f_0(\boldsymbol{Y}) d\boldsymbol{Y} \leq \alpha \ . \end{cases} \tag{29}$$

*Also, the optimum estimation cost is given by* $\mathcal{P}_2(\alpha,\beta)$*.*

The secure estimation approach developed in this section leads to the optimal estimation performance under the given constraints on $\mathsf{P}_{\mathsf{md}}$ and $\mathsf{P}_{\mathsf{fa}}$. The methodology developed so far is illustrated using a case study presented in the next section.

## 5. CASE STUDY

Consider a sensor network consisting of two sensors monitoring a parameter $X$ distributed according to $\mathcal{N}(0,\sigma_x^2)$, and reporting their observations $Y_1$ and $Y_2$ to the FC. Therefore, under an attack-free setting, for $i \in \{1,2\}$, we have

$$Y_i = h_i X + N_i \ ,$$

where $h_i$ is a fixed scalar, and $N_i$ is an i.i.d random variable corresponding to the measurement noise with distribution $\mathcal{N}(0,\sigma_n^2)$. The probability distributions of $X$ and the noise are known to the FC. We consider a setting similar to the one studied in [1], and assume that only sensor 2 is vulnerable to an attack by an adversary, based on which for $\epsilon_i$ defined in Section 2.2, we have $\epsilon_1 = 0$ and $\epsilon_0 + \epsilon_2 = 1$. Therefore, in the adversarial setting, the measurement of sensor 2 is given by

$$Y_2 = h_2 X + N_2 + Z_2 \ .$$

We assume $Z_2$ to be uniformly distributed according to $\mathrm{Unif}\,[a,b]$. Therefore, the composite hypothesis testing problem in (11) for this case is given by

$$\begin{array}{lll} \mathsf{H}_0 : & \boldsymbol{Y} \sim f_0(\boldsymbol{Y} \mid X), & \text{with } X \sim \mathcal{N}(0,\sigma_x^2) \\ \mathsf{H}_1 : & \boldsymbol{Y} \sim f_1(\boldsymbol{Y} \mid X), & \text{with } X \sim \mathcal{N}(0,\sigma_x^2) \end{array} \ , \tag{30}$$

where $\mathsf{H}_0$ corresponds to an attack-free setting, and $\mathsf{H}_1$ corresponds to the attack experienced on sensor 2. This setting is fundamentally similar to the one explored in [1], when $|Z_2| \in [0,\infty)$.

Figure 1 depicts the variations of the estimation quality, captured by $q$ versus the tolerable miss-detection rate $\beta$, where it is observed that the estimation quality improves monotonically as $\beta$ increases. This observation is in line with what is expected analytically from the formulation of the secure parameter estimation problems in (18) and (19).

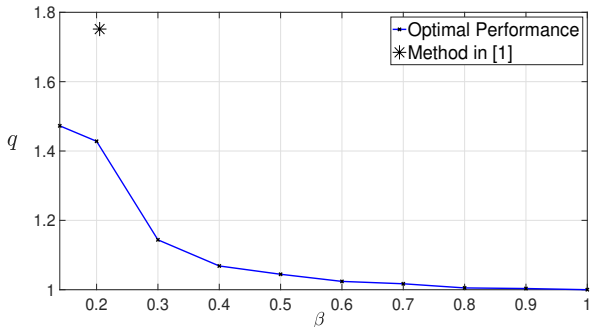


**Fig. 1**. $q$ versus $\beta$ for fixed $\alpha^* = 0.1$.

Figure 1 also compares the estimation quality of the methodology developed in this paper, with that obtained by applying the methodology of [1], which characterizes a single point in the $(q,\beta)$ plane. Specifically, in [1], an estimator is designed to obtain the most robust estimate by exploring the dependence of the estimation quality on the false alarm probability, using which an optimal false alarm probability $\alpha^*$ is obtained, which in turn fixes the miss-detection error probability, and does not provide the flexibility to change the miss-detection rate $\beta$. For the results presented in Fig. 1, we have set $\sigma_x = 3$, $\sigma_n = 1$, $h_1 = 1$, $h_2 = 4$, and $Z_2$ is uniformly distributed $\mathrm{Unif}[-40,40]$. The upper bound on $\mathsf{P}_{\mathsf{fa}}$ is set to $\alpha^*$, which is the optimal false alarm probability obtained for this setting using the methodology in [1].

## 6. CONCLUSIONS

We have introduced and analyzed a secure parameter estimation framework in order to form estimation decisions in the environments where the collected measurements used for inference are vulnerable to be compromised by active adversaries. The analysis has revealed that designing the optimal estimators is fundamentally intertwined with designing detection rules for deciding whether an adversary has compromised the measurements. This necessitates forming compound estimation and detection decisions, and creates a fundamental tradeoff between the qualities of the estimation and detection rules. We have characterized the optimal estimators and detectors that can achieve these fundamental tradeoffs. We have also provided a case study for two-sensor networks in order to assess the performance of the proposed framework, and compared it with those of the existing approaches.

## 7. REFERENCES


[1] C. Wilson and V. V. Veeravalli, "MMSE estimation in a sensor network in the presence of an adversary," in *Proc. IEEE International Symposium on Information Theory*, Barcelona, Spain, Jul. 2016, pp. 2479–2483.

[2] A. Vempaty, L. Tong, and P. K. Varshney, "Distributed inference with Byzantine data: State-of-the-art review on data falsification attacks," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 65–75, Sep. 2013.

[3] A. Vempaty, O. Ozdemir, K. Agrawal, H. Chen, and P. K. Varshney, "Localization in wireless sensor networks: Byzantines and mitigation techniques," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1495–1508, Mar. 2013.

[4] P. Ebinger and S. D. Wolthusen, "Efficient state estimation and Byzantine behavior identification in tactical MANETs," in *Proc. IEEE Military Communications Conference*, Boston, MA, Oct. 2009.

[5] J. Zhang, R. S. Blum, X. Lu, and D. Conus, "Asymptotically optimum distributed estimation in the presence of attacks," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1086–1101, Mar. 2015.

[6] J. Zhang and R. S. Blum, "Distributed estimation in the presence of attacks for large scale sensor networks," in *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2014.

[7] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Countering Byzantine attacks in cognitive radio networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, Mar. 2010, pp. 3098–3101.

[8] E. Soltanmohammadi, M. Orooji, and M. Naraghi-Pour, "Decentralized hypothesis testing in wireless sensor networks in the presence of misbehaving nodes," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 205–215, Jan. 2013.

[9] A. Vempaty, K. Agrawal, P. Varshney, and H. Chen, "Adaptive learning of Byzantines' behavior in cooperative spectrum sensing," in *Proc. IEEE Wireless Communications and Networking Conference*, Cancun, Mexico, Mar. 2011, pp. 1310–1315.

[10] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure state-estimation for dynamical systems under active adversaries," in *Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2011, pp. 337–344.

[11] ——, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[12] S. Z. Yong, M. Zhu, and E. Frazzoli, "Resilient state estimation against switching attacks on stochastic cyber-physical systems," in *Proc. IEEE Conference on Decision and Control*, Osaka, Japan, Dec. 2015, pp. 5162–5169.

[13] M. Pajic, P. Tabuada, I. Lee, and G. J. Pappas, "Attack-resilient state estimation in the presence of noise," in *Proc. IEEE Conference on Decision and Control*, Osaka, Japan, Dec. 2015, pp. 5827–5832.

[14] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber physical systems under sensor attacks: a satisfiability modulo theory approach," *IEEE Transactions on Automatic Control*, vol. PP, no. 99, 2017.

[15] S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, and P. Tabuada, "Secure state estimation: Optimal guarantees against sensor attacks in the presence of noise," in *Proc. IEEE International Symposium on Information Theory*, Hong Kong, China, Jun. 2015, pp. 2929–2933.

[16] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *Proc. IEEE International Conference on Cyber-Physical Systems*, Berlin, Germany, Apr. 2014, pp. 163–174.

[17] C. Z. Bai and V. Gupta, "On Kalman filtering in the presence of a compromised sensor: Fundamental performance bounds," in *Proc. American Control Conference*, Portland, OR, Jun. 2014, pp. 3029–3034.

[18] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 434–444, May 1968.

[19] O. Zeitouni, J. Ziv, and N. Merhav, "When is the generalized likelihood ratio test optimal?" *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1597–1602, Sep. 1992.

[20] G. V. Moustakides, G. H. Jajamovich, A. Tajer, and X. Wang, "Joint detection and estimation: Optimum tests and applications," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4215–4229, Jul. 2012.

[21] G. H. Jajamovich, A. Tajer, and X. Wang, "Minimax-optimal hypothesis testing with estimation-dependent costs," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6151–6165, Dec. 2012.

[22] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1998.