SCALABLE NETWORK PARAMETER ESTIMATION IN THE PRESENCE OF ANOMALIES

Saurabh Sihag and Ali Tajer

Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

ABSTRACT

This paper considers the problem of parameter estimation in a network in which the stochastic model of its measurements can change due to disruption in an unknown subset of sensors. This uncertainty in the measurements model introduces a new dimension to the estimator design. On one hand, the estimation quality depends on the successful isolation of anomalous sensors, and on the other hand, the detection performance is imperfect because of noisy measurements. Motivated by these two observations, this paper models the problem as a composite hypothesis testing problem and analyzes an optimal estimation framework. In large networks, the dimension of the hypotheses testing problem increases exponentially with the size of the network, and also finding the optimal estimate becomes computationally prohibitive. To counter this, this paper provides a scalable solution that consists of detecting and isolating anomalous sensors followed by a sensor-level estimation routine, and establishes asymptotic optimality of the scalable approach. This paper also formulates the decision rules to establish the reliability of the local estimates formed by each sensor, and the local estimates deemed to be reliable are aggregated to form a global estimate. The optimal and scalable schemes are evaluated and compared in a case study.

***Index Terms*—** Anomaly detection, parameter estimation, detection and isolation, scalable.

1. INTRODUCTION

1.1. Overview

Consider the canonical estimation problem in a sensor network consisting of n sensors and a fusion center (FC). Each sensor observes the stochastic parameter $\boldsymbol{X} \in \mathbb{R}^{p\times 1}$ and reports k measurements to the FC. The measurements of sensor $i \in \{1,\dots n\}$ are denoted by $\boldsymbol{Y}_i \triangleq [Y_i^1,\dots,Y_i^k]$, which are related to $\boldsymbol{X}$ according to

$$Y_i^j = h_i(\boldsymbol{X}) + N_i^j, \quad \text{for} \quad j \in \{1,\dots k\}, \tag{1}$$

where h_i captures the characteristics of the channel linking sensor $i \in \{1,\dots n\}$ to the FC, and N_i^j accounts for the measurement noise distributed according to the probability density function (pdf) g_0. The network might undergo an external disturbance (e.g., measurement unit failures), driven by which the network dynamics deviates from its known nominal model. Specifically, each sensor $i \in \{1,\dots n\}$ might behave anomalously with a certain likelihood, in which case the nominal channel model h_i changes to $\bar{h}_i$. [1]. We denote the joint pdf of the measurement vector $\boldsymbol{Y}_i$ under the nominal and anomalous models by f_i^0 and f_i^1, respectively. We define $K \in \{1,\dots,n\}$ as the maximum number of sensors that are anomalous concurrently. Hence, there exist $T \triangleq \sum_{i=1}^{K}\binom{n}{i}$ possible sets of anomalous sensors. Define $\mathcal{S} \triangleq \{S_1,\dots,S_T\}$ as the set of all such sets. We assume that all the sensors are operating normally with the prior likelihood ϵ_0, in which case the joint pdf of $\boldsymbol{Y} \triangleq [\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n]$ is denoted by f_0. Similarly, we define ϵ_i as the prior likelihood of the sensors in S_i being anomalous, in which case the joint pdf of $\boldsymbol{Y}$ is denoted by f_i, for $i \in \{1,\dots,T\}$.

In this paper, we aim to form an estimate of $\boldsymbol{X}$, for which we have the prior pdf π. Forming such an estimate, is inherently coupled with reliably detecting whether there exists any anomalous sensor in the network, and when such sensors are deemed to exist, isolating them as well.

1.2. Relevant Studies

Inference under uncertainty in sensor networks due to Byzantine attacks is relevant to the scope of the problem discussed in this paper. A literature review of the impact of Byzantine attacks on inferences in sensor networks and mitigation strategies is provided in [1]. Detection in the presence of Byzantine attacks in wireless sensor networks is studied in [2–5]. Detection performance for a binary hypotheses testing problem in a sensor network under Byzantine attacks is analyzed in [2]. An event detection algorithm for a setting with model uncertainty and Byzantine attacks is developed in [3]. Detection-driven estimation strategies are developed in [6], where the detection performance as the number of sensors and their observations increase is analyzed. The problem of robust estimation in linear dynamical systems under Byzantine setting is studied in [7–9]. Specifically, degradation in estimation performance of a single sensor network with the stealthiness of the adversary is analyzed in [9].

Under uncertainty in the true stochastic model of sensor measurements, the detection and isolation rules for identifying the true model and the estimator design are intertwined. Decoupling the two problems does not ensure optimality, as established in [10] and [11]. Since the detection performance cannot be perfect under noisy measurements, the estimator design must incorporate the uncertainty in the detection step for optimality. Optimal joint detection and estimation frameworks are analyzed in [12] and [13], where an optimal estimator is designed under the constraints on error performance. A two-step joint detection and estimation methodology is also proposed in [12] in the context of radar systems, in which the best detector is used to decide upon the true model in the first step, and decision rules are developed to decide on the reliability of the estimate formed in the second step.

The problem of identifying the set of anomalous sensors can be modeled as a $(T+1)-$ary composite hypotheses testing problem. Clearly, the dimension of this problem increases exponentially in large networks. Forming an optimal estimate also can be computationally prohibitive in large networks. Scalable estimation schemes by decentralizing or distributing the estimation routine to sensor level have been investigated in [14–20]. Consensus-based distributed estimation routines are investigated in [14–17], which are applicable in dynamical models. Distributed estimation algorithms under various constraints, such as communication bandwidth between the sensors and the FC, and power constraints, are also studied in [15, 16, 18]. An optimal decentralized linear estimation strategy for sensor network is developed in [19]. Decentralized estimation scheme for a noise affected deterministic parameter in a sensor network with bandwidth constraints is developed in [18]. A minimal energy decentralized estimation scheme with best-linear-unbiased-estimation fusion rule is developed in [20].

[1] For the convenience in notations, in this paper we assume that the anomalous model takes only one form. Extensions to a countable number of models follows the same line of analysis.

This research was supported in part by the U. S. National Science Foundation under the CAREER Award ECCS-1554482 and grant DMS-1737976.

1.3. Contributions

In this paper, we model the problem of anomaly detection and estimation as a composite hypotheses testing problem and provide the optimal decision rules and estimates. However, the number of hypotheses corresponding to all possible set of anomalous sensors increase exponentially with the number of sensors. Therefore, we propose a scalable isolation mechanism that involves forming sensor-level isolation decisions, and subsequently aggregating them. Similarly, forming an optimal estimate in a large network is computationally prohibitive. We also propose a reliability test for the local estimates and utilize the techniques from existing literature to aggregate them. The scalable isolation mechanism and the decentralized estimation scheme jointly form the scalable decision rules proposed in this paper.

2. PARAMETER ESTIMATION MODEL

The structure of the optimal estimator for $\boldsymbol{X}$ varies with the true model of the measurements $\boldsymbol{Y}$. All possible models of the measurements set $\boldsymbol{Y}$, for $i \in \{0, \dots, T\}$, can be listed as

$$\mathsf{H}_i : \ \boldsymbol{Y} \sim f_i(\boldsymbol{Y} \,|\, \boldsymbol{X}), \ \text{with } \boldsymbol{X} \sim \pi(\boldsymbol{X}) \ . \tag{2}$$

To capture the estimation quality, we define the non-negative cost function $\mathsf{C}(\boldsymbol{X}, \boldsymbol{U})$ to measure the fidelity of estimate $\boldsymbol{U}$ for $\boldsymbol{X}$. Under model H_i, the average posterior cost function is defined as

$$\mathsf{C}_{\mathrm{p},i}(\boldsymbol{U} \mid \boldsymbol{Y}) \triangleq \mathbb{E}_i \left[\mathsf{C}(\boldsymbol{X}, \boldsymbol{U}) \mid \boldsymbol{Y} \right] \ , \ \text{for } i \in \{0, \dots, T\} \ . \tag{3}$$

The optimal estimator is defined as

$$\hat{\boldsymbol{X}}_i(\boldsymbol{Y}) \triangleq \arg\inf_{\boldsymbol{U}} \mathsf{C}_{\mathrm{p},i}(\boldsymbol{U} \mid \boldsymbol{Y}) \ , \tag{4}$$

and the optimal estimation cost is given by

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

In Section 3, we provide the optimal rules for estimating $\boldsymbol{X}$, and at the same time isolating the anomalous sensors, when they are deemed to exist. The complexity of such decisions, however, grows exponentially, as the size of the network grows. To circumvent the complexity, in Section 4 we provide a scalable algorithm that has controlled complexity and has asymptotically optimal performance.

3. OPTIMAL DECISION RULES

Forming a reliable estimate for $\boldsymbol{X}$ requires knowing the statistical model of the measurements, since the measurements model faces uncertainty due to potentially having anomalous sensors. Hence, for estimating $\boldsymbol{X}$, we need to also concurrently detect and isolate anomalous sensors. We model this problem as a $(T+1)$-ary composite hypothesis testing problem, where for $i \in \{0, \dots, T\}$ we have

$$\mathsf{H}_i : \ \boldsymbol{Y} \sim f_i(\boldsymbol{Y} \,|\, \boldsymbol{X}), \ \text{with } \boldsymbol{X} \sim \pi(\boldsymbol{X}) \ . \tag{6}$$

The detection, isolation, and estimation routines are intertwined and a decoupled approach cannot guarantee optimal performance. Specifically, performing anomaly detection and isolation followed by optimal estimation does not incorporate the uncertainty of the detection and isolation decisions into the estimation rule. In this section, we provide the optimal combined detection, isolation, and estimation rules.

3.1. Optimal Sensor Isolation and Parameter Estimation

We model the detection rules for the hypothesis testing problem in (20) by a randomized rule $\boldsymbol{\delta} \triangleq [\delta_0(\boldsymbol{Y}), \dots, \delta_T(\boldsymbol{Y})]$, where $\delta_i(\boldsymbol{Y})$ is the probability of deciding in favor of H_i, for $i \in \{0, \dots, T\}$. Define $\mathsf{D}_{\mathrm{c}} \in \{\mathsf{H}_0, \dots, \mathsf{H}_T\}$ as the decision and $\mathsf{T}_{\mathrm{c}} \in \{\mathsf{H}_0, \dots, \mathsf{H}_T\}$ as the true model. Hence, the likelihood of deciding H_i while the true model is H_j is given by

$$\mathbb{P}(\mathsf{D}_{\mathrm{c}} = \mathsf{H}_i \,|\, \mathsf{T}_{\mathrm{c}} = \mathsf{H}_j) = \int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) f_j(\boldsymbol{Y}) \, d\boldsymbol{Y} \ . \tag{7}$$

We define P_{md} as the aggregate likelihood of erroneously deciding about the model under the presence of anomalous sensors, i.e.,

$$\mathsf{P}_{\mathrm{md}} \triangleq \mathbb{P}(\mathsf{D}_{\mathrm{c}} \neq \mathsf{T}_{\mathrm{c}} \,|\, \mathsf{T}_{\mathrm{c}} \neq \mathsf{H}_0) \tag{8}$$

$$= \frac{1}{1 - \mathbb{P}(\mathsf{T}_{\mathrm{c}} = \mathsf{H}_0)} \sum_{i=1}^{T} \mathbb{P}(\mathsf{D}_{\mathrm{c}} \neq \mathsf{H}_i, \mathsf{T}_{\mathrm{c}} = \mathsf{H}_i) \ . \tag{9}$$

By leveraging (7) and the definition of ϵ_i, we have

$$\mathsf{P}_{\mathrm{md}} = \frac{1}{1 - \epsilon_0} \sum_{i=1}^{T} \sum_{\substack{j=0 \\ j \neq i}}^{T} \epsilon_i \int_{\boldsymbol{Y}} \delta_j(\boldsymbol{Y}) f_i(\boldsymbol{Y}) \, d\boldsymbol{Y} \ . \tag{10}$$

Similarly, we define P_{fa} as the likelihood of erroneously detecting anomaly when no sensor is operating anomalously, i.e.,

$$\mathsf{P}_{\mathrm{fa}} \triangleq \mathbb{P}(\mathsf{D}_{\mathrm{c}} \neq \mathsf{T}_{\mathrm{c}} \,|\, \mathsf{T}_{\mathrm{c}} = \mathsf{H}_0) = \sum_{i=1}^{T} \mathbb{P}(\mathsf{D}_{\mathrm{c}} = \mathsf{H}_i \,|\, \mathsf{T}_{\mathrm{c}} = \mathsf{H}_0) \tag{11}$$

$$= \sum_{i=1}^{T} \int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) f_0(\boldsymbol{Y}) \, d\boldsymbol{Y} \ . \tag{12}$$

Note that the estimation cost $\mathsf{C}(\boldsymbol{X}, \boldsymbol{U}_i)$ for any generic estimator $\boldsymbol{U}_i$ of $\boldsymbol{X}$ under model H_i is relevant only when the decision is H_i. Therefore, we define $J_i(\delta_i, \boldsymbol{U}_i)$ as the average estimation cost when the decision is H_i, i.e., $J_i(\delta_i, \boldsymbol{U}_i) \triangleq \mathbb{E}_i[\mathsf{C}(\boldsymbol{X}, \boldsymbol{U}_i) \mid \mathsf{D}_{\mathrm{c}} = \mathsf{H}_i]$. We also define an overall estimation cost as

$$J(\boldsymbol{\delta}, \boldsymbol{U}) \triangleq \max_{i \in \{0, \dots, T\}} J_i(\delta_i, \boldsymbol{U}_i) \ , \tag{13}$$

where $\boldsymbol{U} \triangleq [\boldsymbol{U}_0, \dots, \boldsymbol{U}_T]$. For making combined decisions, we aim to minimize the estimation cost $J(\boldsymbol{\delta}, \boldsymbol{U})$ under constraints on the error rates P_{md} and P_{fa}, i.e.,

$$\mathcal{P}(\alpha, \beta) \triangleq \begin{cases} \min_{(\boldsymbol{\delta}, \boldsymbol{U})} & J(\boldsymbol{\delta}, \boldsymbol{U}) \\ \text{s.t.} & \sum\limits_{i=1}^{T} \sum\limits_{\substack{j=0 \\ j \neq i}}^{T} \frac{\epsilon_i}{1 - \epsilon_0} \int_{\boldsymbol{Y}} \delta_j(\boldsymbol{Y}) f_i(\boldsymbol{Y}) d\boldsymbol{Y} \leq \beta \\ & \sum\limits_{i=1}^{T} \int_{\boldsymbol{Y}} \delta_i(\boldsymbol{Y}) f_0(\boldsymbol{Y}) \, d\boldsymbol{Y} \leq \alpha \end{cases} . \tag{14}$$

Remark 1 (Feasibility)**.** *The constraints α and β on P_{fa} and P_{md}, respectively, cannot be made arbitrarily small simultaneously. It can be readily verified that under the constraint α on P_{fa}, there exists a minimum feasible value for P_{md} denoted by $\beta^*(\alpha) \in (0, 1)$. Therefore, we must have $\beta \geq \beta^*(\alpha)$ for the problem in* (14) *to be feasible.*

Since the estimation cost appears only in $J(\boldsymbol{\delta}, \boldsymbol{U})$ in (14), solving $\mathcal{P}(\alpha, \beta)$ can be broken down into two sub-problems. The estimators $\boldsymbol{U}_i$ and decision rules $\delta_i(\boldsymbol{Y})$, for $i \in \{0, \dots, T\}$, that solve (14) are given in Theorem 1.

Theorem 1. *The estimator under the model H_i, which minimize $J(\boldsymbol{\delta}, \boldsymbol{U})$, is $\hat{\boldsymbol{X}}_i(\boldsymbol{Y})$ defined in* (4)*. Furthermore, for the optimal decision rule $\boldsymbol{\delta}$, we have $\delta_{i^*}(\boldsymbol{Y}) = 1$, where $i^* \triangleq \operatorname{argmin}_{i \in \{0, \dots, T\}} A_i$, A_0 is defined as*

$$A_0 \triangleq \ell_0 f_0(\boldsymbol{Y}) (\hat{\mathsf{C}}_{\mathrm{p},0}(\boldsymbol{Y}) - u) + \ell_{T+1} \sum_{i=1}^{T} \frac{\epsilon_i}{1 - \epsilon_0} f_i(\boldsymbol{Y}) \ , \tag{15}$$

and $\{A_i : i \in \{1, \dots, T\}\}$ are given by

$$A_i \triangleq \ell_i f_i(\boldsymbol{Y}) (\hat{\mathsf{C}}_{\mathrm{p},i}(\boldsymbol{Y}) - u) + \ell_{T+1} \sum_{\substack{j=1, \\ j \neq i}}^{T} \frac{\epsilon_j}{1 - \epsilon_0} f_j(\boldsymbol{Y}) + \ell_{T+2} f_0(\boldsymbol{Y}) \ , \tag{16}$$

where the non-negative constants $\{\ell_i : i \in \{0, \dots T+2\}\}$ are the Lagrangian multipliers selected such that $\sum_{i=0}^{T+2} \ell_i = 1$, and the

constraints in a convex optimization problem equivalent to the problem in (14) *are satisfied.*

4. SCALABLE DECISION RULES

The number of possible models in set $\mathcal{S}$ increases exponentially with the number of sensors. Also, forming an optimal estimate for $\boldsymbol{X}$ by the FC has high computational complexity. Motivated by controlling these complexities, we provide a scalable approach, in which detecting the presence of anomalies is followed by isolating the anomalous sensors and estimating $\boldsymbol{X}$ at the sensor level.

To formalize the approach, let H_0^i and H_1^i represent the normal and anomalous models, respectively, for the measurements of sensor i. We denote an estimate for $\boldsymbol{X}$ under the model H_j^i, for $j \in \{0,1\}$, by $\boldsymbol{U}_j^i$. Under model H_j^i, the average *local* posterior cost function is defined as

$$\mathsf{C}_{\mathrm{p,i}}^{\mathrm{d}}(\boldsymbol{U}_i^j \mid \boldsymbol{Y}_j) \triangleq \mathbb{E}_i[\mathsf{C}(\boldsymbol{X}, \boldsymbol{U}_i^j) \mid \boldsymbol{Y}_j] \ , \tag{17}$$

and the optimal local estimator and estimation cost are given by

$$\hat{\boldsymbol{X}}_i^j(\boldsymbol{Y}_j) \triangleq \arg\inf_{\boldsymbol{U}_i^j} \mathsf{C}_{\mathrm{p,i}}^{\mathrm{d}}(\boldsymbol{U}_i^j \mid \boldsymbol{Y}_j) \ , \tag{18}$$

$$\text{and} \quad \hat{\mathsf{C}}_{\mathrm{p,i}}^j(\boldsymbol{Y}_j) \triangleq \inf_{\boldsymbol{U}_i^j} \mathsf{C}_{\mathrm{p,i}}^{\mathrm{d}}(\boldsymbol{U}_i^j \mid \boldsymbol{Y}_j) \ . \tag{19}$$

4.1. Detecting Anomalies

We model detecting the presence of anomalous sensors in the network as the binary composite hypotheses testing problem

$$\begin{array}{lll} \hat{\mathsf{H}}_0 & : & \boldsymbol{Y} \sim f_0(\boldsymbol{Y} \mid \boldsymbol{X}), \ \text{ with } \boldsymbol{X} \sim \pi(\boldsymbol{X}) \\ \hat{\mathsf{H}}_1 & : & \boldsymbol{Y} \not\sim f_0(\boldsymbol{Y} \mid \boldsymbol{X}), \ \text{ with } \boldsymbol{X} \sim \pi(\boldsymbol{X}) \end{array} \ . \tag{20}$$

By defining $\mathsf{D}_{\mathrm{m}} \in \{\hat{\mathsf{H}}_0, \hat{\mathsf{H}}_1\}$ as the decision and $\mathsf{T}_{\mathrm{m}} \in \{\hat{\mathsf{H}}_0, \hat{\mathsf{H}}_1\}$ as the true hypothesis, we are interested in designing a decision rule for each sensor that solves

$$\mathcal{P}(\hat{\alpha}) \triangleq \begin{cases} \min & \mathbb{P}(\mathsf{D}_{\mathrm{m}} = \hat{\mathsf{H}}_0 \,|\, \mathsf{T}_{\mathrm{m}} = \hat{\mathsf{H}}_1) \\ \text{s.t.} & \mathbb{P}(\mathsf{D}_{\mathrm{m}} = \hat{\mathsf{H}}_1 \,|\, \mathsf{T}_{\mathrm{m}} = \hat{\mathsf{H}}_0) \leq \hat{\alpha} \end{cases} , \tag{21}$$

the solution to which is the Neyman-Pearson test.

Theorem 2. *The solution of $\mathcal{P}(\hat{\alpha})$ is given by*

$$\frac{f_m(\boldsymbol{Y})}{f_0(\boldsymbol{Y})} \underset{\hat{\mathsf{H}}_0}{\overset{\hat{\mathsf{H}}_1}{\gtrless}} \gamma \ , \tag{22}$$

where the pdf f_m is given by

$$f_m(\boldsymbol{Y}) = \frac{1}{1-\epsilon_0} \sum_{i=1}^{T} \epsilon_i f_i(\boldsymbol{Y}) \ , \tag{23}$$

and the threshold γ is chosen such that the constraint of $\mathcal{P}(\hat{\alpha})$ is satisfied with equality.

4.2. Isolating Anomalous Sensors

If the network is deemed to contain anomalous sensors, we isolate them in the second step. Isolating anomalous sensors is equivalent to selecting the correct model in a $T-$ hypotheses testing problem, modeled by H_i, for $i \in \{1, \ldots, T\}$, defined in (2). Define $\mathsf{D}_{\mathrm{is}} \in \{\mathsf{H}_1, \ldots \mathsf{H}_T\}$ as the decision formed and $\mathsf{T}_{\mathrm{is}} \in \{\mathsf{H}_1, \ldots \mathsf{H}_T\}$ as the true model. Let P_{is} denote the probability of isolating the wrong set of sensors given that anomaly exists, given by

$$\mathsf{P}_{\mathrm{is}} \triangleq \mathbb{P}(\mathsf{D}_{\mathrm{is}} \neq \mathsf{T}_{\mathrm{is}} \,|\, \mathsf{T}_{\mathrm{is}} \neq \mathsf{H}_0) \tag{24}$$

$$= \sum_{i=1}^{T} \sum_{\substack{j=1, \\ j \neq i}}^{T} \mathbb{P}(\mathsf{D}_{\mathrm{is}} = \mathsf{H}_j \,|\, \mathsf{T}_{\mathrm{is}} = \mathsf{H}_i) \mathbb{P}(\mathsf{T}_{\mathrm{is}} = \mathsf{H}_i) \ . \tag{25}$$

Furthermore, define $\mathsf{P}_{\mathrm{is}}^{\mathrm{l}}$ as the value of P_{is} when $\boldsymbol{X}$ is known perfectly. Also, define $\mathsf{P}_{\mathrm{is}}^{\mathrm{u}}(\boldsymbol{X}_c)$ as the value of P_{is} when $\boldsymbol{X}$ is set to $\boldsymbol{X}_c$. These two terms can be leveraged to establish trackable bounds on P_{is}.

Lemma 1. *There exists $\boldsymbol{X}_c \in \mathbb{R}^{p \times 1}$ such that $\mathsf{P}_{\mathrm{is}}^{\mathrm{l}} \leq \mathsf{P}_{\mathrm{is}} \leq \mathsf{P}_{\mathrm{is}}^{\mathrm{u}}(\boldsymbol{X}_c)$.*

Lemma 1 is instrumental to analyzing the error exponent of P_{is} established in Theorem 3 and Theorem 4. For $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots k\}$, we assume $(Y_i^j - h_i(\boldsymbol{X}))$ to be independent of each other and distributed according to pdf $\bar{g}_i$ and g_0 when the sensor i has anomalous behavior and normal behavior, respectively.

Theorem 3. *The decision rule that minimizes P_{is} is given by*

$$\mathsf{D}_{\mathrm{I}} = \mathsf{H}_{i^*}, \ \textit{where } i^* = \arg\max_{i \in \{1,\ldots,T\}} f_i(\boldsymbol{Y}) \cdot \mathbb{P}(\mathsf{T}_{\mathrm{is}} = \mathsf{H}_i) \ . \tag{26}$$

Also, under the assumption that $\mathrm{supp}(g_0) = \mathbb{R}$ *and* $\mathrm{supp}(\bar{g}_i) = \mathbb{R}$,

$$\lim_{k \to \infty} -\frac{\log(\mathsf{P}_{\mathrm{is}})}{k} = \begin{cases} \min\limits_{i \neq j} C(\bar{g}_i, g_0) + C(g_0, \bar{g}_j), \ \textit{if } K = 1 \ , \\ \min\limits_{i} \{\min\{C(\bar{g}_i, g_0), C(g_0, \bar{g}_i)\}\} \ , \\ \qquad\qquad \textit{if } K > 1 \end{cases} , \tag{27}$$

for $i, j \in \{1, \ldots, n\}$, where $C(\bar{g}_i, g_0)$ denotes the Chernoff information between $\bar{g}_i$ and g_0.

The optimal decision rule defined in Theorem 3 can become computationally prohibitive in large networks. Therefore, we provide an alternative decision rule at each sensor to isolate the anomalous sensors. For this purpose, define η_i as the probability of sensor i being anomalous, given that anomaly exists. Clearly, $\eta_i = \sum_{j \in V_i} \frac{\epsilon_j}{1-\epsilon_0}$, where V_i is the subset of the set of hypotheses $\{\mathsf{H}_1 \ldots \mathsf{H}_T\}$ in which sensor i is anomalous. Define $\mathsf{LR}_i(\boldsymbol{Y}_i)$ as a marginal likelihood ratio defined for sensor i, i.e.,

$$\mathsf{LR}_i(\boldsymbol{Y}_i) \triangleq \frac{\eta_i f_i^1(\boldsymbol{Y}_i)}{(1-\eta_i) f_i^0(\boldsymbol{Y}_i)} \ . \tag{28}$$

Theorem 4. *The anomalous sensors can be isolated using the following decision rule*

$$\mathsf{D}_{\mathrm{is}} = \mathsf{H}_{i^*}, \ \textit{where } i^* = \arg\max_{i \in \{1,\ldots,T\}} \prod_{v \in S_i} \mathsf{LR}_v(\boldsymbol{Y}_v) \ . \tag{29}$$

Under the assumption that $\mathrm{supp}(g_0) = \mathbb{R}$ *and* $\mathrm{supp}(\bar{g}_i) = \mathbb{R}$, *the error exponent of this decision rule is given by*

$$\lim_{k \to \infty} -\frac{\log(\mathsf{P}_{\mathrm{is}})}{k} = \begin{cases} \min\limits_{i \neq j} C(\bar{g}_i, g_0) + C(g_0, \bar{g}_j), \ \textit{if } K = 1 \ , \\ \min\limits_{i} \{\min\{C(\bar{g}_i, g_0), C(g_0, \bar{g}_i)\}\} \ , \\ \qquad\qquad \textit{if } K > 1 \end{cases} , \tag{30}$$

for $i, j \in \{1, \ldots, n\}$.

Therefore, P_{is} decays at the same rate that the optimal decision rule given in Theorem 3, does.

4.3. Estimation

In the final step, we design a test to decide on the reliability of the estimates of $\boldsymbol{X}$ formed by different sensors. Based on the outcome of this test, the estimates deemed to be reliable are aggregated to form a global estimate of $\boldsymbol{X}$. For this purpose, we provide some definitions relevant to the analysis in this step. The problem of deciding whether a sensor $i \in \{1, \ldots n\}$ is anomalous is a binary hypothesis testing problem

$$\begin{array}{lll} \mathsf{H}_0^i & : & \boldsymbol{Y}_i \sim f_i^0(\boldsymbol{Y}_i \mid \boldsymbol{X}), \ \text{ with } \boldsymbol{X} \sim \pi(\boldsymbol{X}) \\ \mathsf{H}_1^i & : & \boldsymbol{Y}_i \sim f_i^1(\boldsymbol{Y}_i \mid \boldsymbol{X}), \ \text{ with } \boldsymbol{X} \sim \pi(\boldsymbol{X}) \end{array} \ . \tag{31}$$

Define $\boldsymbol{\delta}^i \triangleq [\delta_0^i(\boldsymbol{Y}_i), \delta_1^i(\boldsymbol{Y}_i)]$ as the corresponding randomized decision rule for sensor i. Also, define $\rho_0^i \in (0,1)$ and $\rho_1^i \in (0,1)$

as the false alarm rate and miss-detection rate for sensor i based on the decision rules in detection and isolation steps. Given a decision $\mathsf{D}_{\mathrm{s}}^i \in \{\mathsf{H}_0^i, \mathsf{H}_1^i\}$ for sensor i, we accept the local estimate from sensor i only if the estimate is deemed to be reliable. Let $\mathsf{H}_{\mathrm{r}}^i$ denote the hypothesis that the estimate is reliable, and $\mathsf{H}_{\mathrm{u}}^i$ denote the hypothesis that the estimate is not reliable. We define $\mathsf{D}_{\mathrm{r}}^i \in \{\mathsf{H}_{\mathrm{r}}^i, \mathsf{H}_{\mathrm{u}}^i\}$ as the decision formed on the reliability of the estimate, and define $\boldsymbol{\Delta}^i \triangleq [\Delta_{\mathrm{r}}^i(\boldsymbol{Y}_i), \Delta_{\mathrm{u}}^i(\boldsymbol{Y}_i)]$ as the randomized decision rule for deciding about the reliability of the estimate from sensor i, where $\Delta_{\mathrm{r}}^i(\boldsymbol{Y}_i)$ is the probability that the estimate is reliable and $\Delta_{\mathrm{u}}^i(\boldsymbol{Y}_i)$ is the probability that the estimate is not reliable. Define $\mathbb{P}_j(\mathsf{D}_{\mathrm{r}}^i = \mathsf{H}_{\mathrm{r}}^i)$ as the likelihood of forming a reliable estimate, given that the decision on the sensor measurements is H_j^i, for $j \in \{0,1\}$. Since the estimate is accepted only under $\mathsf{H}_{\mathrm{r}}^i$, we define the estimation cost under the decision H_j^i for $j \in \{0,1\}$ and $\mathsf{H}_{\mathrm{r}}^i$ as

$$J_j^i(\delta_j^i(\boldsymbol{Y}_i), \Delta_{\mathrm{r}}^i(\boldsymbol{Y}_i), \boldsymbol{U}_j^i) \triangleq \mathbb{E}_j[\mathsf{C}(\boldsymbol{X}, \boldsymbol{U}_j^i) \mid \mathsf{D}_{\mathrm{r}}^i = \mathsf{H}_{\mathrm{r}}^i] \,. \quad (32)$$

For a given $\boldsymbol{Y}$, let the detection and isolation steps decide on the true model of $\boldsymbol{Y}$ as H_j, for $j \in \{0, \ldots, T\}$. Define $\hat{\boldsymbol{X}}^d(\boldsymbol{Y})$ as the estimate formed by fusing the reliable local estimates at the FC. The average posterior estimation cost for the estimator $\hat{\boldsymbol{X}}^d(\boldsymbol{Y})$ is

$$J_j(\tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\Delta}}) \triangleq \mathbb{E}_j[\mathsf{C}(\boldsymbol{X}, \hat{\boldsymbol{X}}^d) \mid \mathsf{H}_j], \text{ for } j \in \{0, \ldots, T\} \,, \quad (33)$$

where $\tilde{\boldsymbol{\delta}} \triangleq \{\boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^n\}$ and $\tilde{\boldsymbol{\Delta}} \triangleq \{\boldsymbol{\Delta}^1, \ldots, \boldsymbol{\Delta}^n\}$.

4.3.1. Local Estimators

The likelihood of forming a reliable estimate at sensor i under the decision H_j^i, for $j \in \{0,1\}$, is upper bounded by

$$\mathbb{P}_j(\mathsf{D}_{\mathrm{r}}^i = \mathsf{H}_{\mathrm{r}}^i) = \int_{\boldsymbol{Y}_i} \delta_j^i(\boldsymbol{Y}_i) \Delta_r^i(\boldsymbol{Y}_i) f_j^i(\boldsymbol{Y}_i) d\boldsymbol{Y}_i \leq 1 - \rho_j^i \,. \quad (34)$$

This implies that only a fraction of the decisions on the true model of $\boldsymbol{Y}_i$ will provide reliable estimates. We control this fraction of decisions by choosing $\hat{\rho}_0^i \geq \rho_0^i$ and $\hat{\rho}_1^i \geq \rho_1^i$, such that

$$\mathbb{P}_0(\mathsf{D}_{\mathrm{r}}^i = \mathsf{H}_{\mathrm{r}}^i) \geq 1 - \hat{\rho}_0^i \text{ and } \mathbb{P}_1(\mathsf{D}_{\mathrm{r}}^i = \mathsf{H}_{\mathrm{r}}^i) \geq 1 - \hat{\rho}_1^i \,. \quad (35)$$

The decision rules $\boldsymbol{\Delta}^i$ and estimators $\boldsymbol{U}_j^i$, for $j \in \{0,1\}$, are determined by solving

$$\mathcal{P}_j(\hat{\rho}_j^i) = \begin{cases} \min_{\boldsymbol{\Delta}^i, \boldsymbol{U}_j^i} & J_j^i(\delta_j^i(\boldsymbol{Y}_i), \Delta_{\mathrm{r}}^i(\boldsymbol{Y}_i), \boldsymbol{U}_j^i) \\ \text{s.t.} & \mathbb{P}_j(\mathsf{D}_{\mathrm{r}}^i = \mathsf{H}_{\mathrm{r}}^i) \geq 1 - \hat{\rho}_j^i \end{cases} , \quad (36)$$

for $j \in \{0,1\}$. Note that the estimators $\boldsymbol{U}_j^i$ appear only in the cost $J_j^i(\delta_j^i(\boldsymbol{Y}_i), \Delta_{\mathrm{r}}^i(\boldsymbol{Y}_i), \boldsymbol{U}_j^i)$, which allows for solving the problem in (36) by decoupling it into two sub-problems.

Theorem 5. *The decision rules that minimize $\mathcal{P}_0(\hat{\rho}_0^i)$ and $\mathcal{P}_1(\hat{\rho}_1^i)$ are*

$$\hat{\mathsf{C}}_{\mathrm{p},0}^i(\boldsymbol{Y}_i) \underset{\mathsf{H}_{\mathrm{r}}^i}{\overset{\mathsf{H}_{\mathrm{u}}^i}{\gtrless}} \gamma_0^i \quad \textit{under } \mathsf{H}_0^i \,, \quad (37)$$

$$\hat{\mathsf{C}}_{\mathrm{p},1}^i(\boldsymbol{Y}_i) \underset{\mathsf{H}_{\mathrm{r}}^i}{\overset{\mathsf{H}_{\mathrm{u}}^i}{\gtrless}} \gamma_1^i \quad \textit{under } \mathsf{H}_1^i \,, \quad (38)$$

where γ_0^i and γ_1^i are selected such that the constraints in (35) *are satisfied with equality, and the estimation costs $\hat{\mathsf{C}}_{\mathrm{p},0}^i(\boldsymbol{Y}_i)$ and $\hat{\mathsf{C}}_{\mathrm{p},1}^i(\boldsymbol{Y}_i)$ are defined in* (19)*. The optimal estimators that minimize $J_0^i(\delta_0^i(\boldsymbol{Y}_i), \Delta_{\mathrm{r}}^i(\boldsymbol{Y}_i), \boldsymbol{U}_0^i)$ and $J_1^i(\delta_1^i(\boldsymbol{Y}_i), \Delta_{\mathrm{r}}^i(\boldsymbol{Y}_i), \boldsymbol{U}_1^i)$ are $\hat{\boldsymbol{X}}_i^0(\boldsymbol{Y}_i)$ and $\hat{\boldsymbol{X}}_i^1(\boldsymbol{Y}_i)$, respectively, defined in* (18)*.*

5. CASE STUDY

We illustrate the asymptotic optimality of the decision rule in Theorem 4 by comparing variations of P_{is} with respect to the number of measurements for the decision rules in Theorem 3 and Theorem 4. For the plots in Fig. 1, we consider a two-sensor network, where one sensor is anomalous at all instants. We set the function h_1 and h_2 to be linear, such that, $h_i(\boldsymbol{X}) = H_i \boldsymbol{X}$, for $i \in \{1,2\}$. We fix $\boldsymbol{X}$ to be a scalar with pdf $\mathcal{N}(0,4)$ and $H_1 = 1$ and $H_2 = 4$. Under the normal setting, the sensor measurements are affected by noise components with pdf $\mathcal{N}(0,2)$. Under the anomalous setting, we set the distributions $\bar{g}_1$ and $\bar{g}_2$ to $\mathcal{N}(0,4)$. As observed in Fig. 1, the error probabilities for both decision rules decay exponentially at a similar rate with the increase in the number of sensor measurements.

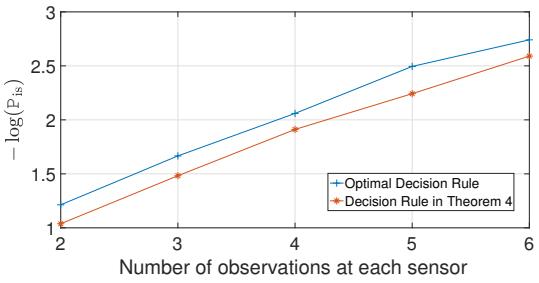

Fig. 1. Error exponent versus number of measurements.

Next, we compare the estimation quality obtained using the scalable approach. For a fair comparison between the optimal scheme and the scalable scheme, we set α and β in the optimal scheme corresponding to $\hat{\rho}_0^1, \hat{\rho}_1^1, \hat{\rho}_0^2$, and $\hat{\rho}_1^2$,and evaluate q, which is the ratio of estimation cost in the scalable approach to that obtained using the optimal approach.

Table 1. Estimation quality in the optimal and scalable approaches

$\hat{\rho}_0^1$	$\hat{\rho}_1^1$	$\hat{\rho}_0^2$	$\hat{\rho}_1^2$	q	α	β
0.3	0.45	0.35	0.46	1.124	0.18	0.38
0.25	0.42	0.41	0.53	1.158	0.192	0.46
0.21	0.58	0.26	0.64	1.174	0.154	0.51
0.35	0.45	0.34	0.52	1.104	0.214	0.41

For the results presented in Table 1, we consider a 2-sensor network with both sensors vulnerable. We set $\epsilon_1 = 0.25$, $\epsilon_2 = 0.15$ and $\epsilon_3 = 0.2$. We also consider $\mathsf{H}_1 = 1$ and $\mathsf{H}_2 = 1$, assume $\boldsymbol{X}$ to be distributed according to $\mathrm{Unif}[0,4]$, consider the distribution g_0 to be $\mathcal{N}(0, 0.75)$, and set the distributions $\bar{g}_1$ and $\bar{g}_2$ to $\mathcal{N}(0,2)$. In the results summarized in Table 1, the optimal scheme consistently performs better than the scalable scheme in terms of estimation quality.

6. CONCLUSION

In this paper, we have analyzed a state estimation problem in a sensor network in which the measurements of an unknown subset of sensors may undergo change due to any external disturbance or disruption. Through the analysis, we have illustrated that the detection and estimation routines are intertwined and provided the optimal framework for estimators. Considering the infeasibility of optimal design in large networks, we have also provided a scalable scheme consisting of three steps: detection of anomalous behavior in the network, isolation of anomalous sensors and a decentralized estimation routine with reliability test on each local estimate. We have also provided a scalable decision rule for isolation of affected sensors and established its asymptotic optimality by error exponent analysis. The theory developed in this paper is evaluated in a case study.

7. REFERENCES

[1] 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.

[2] M. Stefano, M. Vincenzo, and L. Tong, "Distributed detection in the presence of byzantine attacks," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 16–29, Jan. 2009.

[3] P. Zhang, J. Y. Koh, S. Lin, and I. Nevat, "Distributed event detection under Byzantine attack in wireless sensor networks." in *Proc. IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Singapore, Apr. 2014, pp. 1–6.

[4] 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.

[5] B. Kailkhura, S. Brahma, Y. S. Han, and P. K. Varshney, "Optimal distributed detection in the presence of Byzantines," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013, pp. 2925–2929.

[6] 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.

[7] H. Fawzi, P. Tabuada, and S. Diggavi, "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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] C. Chen, S. Zhu, X. Guan, and X. S. Shen, *Wireless sensor networks: Distributed consensus estimation.* Springer, 2014.

[15] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion." *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1631–1639, 2005.

[16] F. Jun and H. Li, "Power constrained distributed estimation with correlated sensor data." *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3292–3297, 2009.

[17] A. S. Behbahani, A. M. Eltawil, and H. Jafarkhani, "Decentralized estimation under correlated noise." *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5603–5614, Sep. 2014.

[18] X. Jin-Jun and L. Zhi-Quan, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3564–3575, 2005.

[19] X. Yuzhe, V. Gupta, and C. Fischione, "Distributed estimation," available at https://people.kth.se/~carlofi/Publications/e-reference-distributed-estimation.pdf, Tech. Rep., 2012.

[20] W. Jwo-Yuh, H. Qian-Zhi, and L. Ta-Sung, "Minimal Energy Decentralized Estimation via exploiting the statistical knowledge of sensor noise variance," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 2171–2176, 2008.