2014 American Control Conference (ACC)
June 4-6, 2014. Portland, Oregon, USA

Attack-Resilient Minimum Mean-Squared Error Estimation

James Weimer, Nicola Bezzo, Miroslav Pajic, Oleg Sokolsky, and Insup Lee

Abstract— **This work addresses the design of resilient estimators for stochastic systems. To this end, we introduce a minimum mean-squared error resilient (MMSE-R) estimator whose conditional mean squared error from the state remains finitely bounded and is independent of additive measurement attacks. An implementation of the MMSE-R estimator is presented and is shown as the solution of a semidefinite programming problem, which can be implemented efficiently using convex optimization techniques. The MMSE-R strategy is evaluated against other competing strategies representing other estimation approaches in the presence of small and large measurement attacks. The results indicate that the MMSE-R estimator significantly outperforms (in terms of mean-squared error) other realizable resilient (and non-resilient) estimators.**

I. Introduction

As cyber physical systems become more integrated into safety critical systems, malicious attacks on sensory information can have catastrophic effects. The fusion of additional sensory information offers a potential for increased security in these systems. For instance, in modern vehicular cruise control systems the fusion of velocity estimates from wheel encoders, inertial measurements units, and GPS information can be incorporated to not only improve the estimate of the vehicle velocity, but may also be used to determine whether any specific sensor measurement has been altered. Thus, establishing techniques for defending against sensor attacks that maliciously alter the measurements can significantly improve both performance and safety in today's critical systems.

While there exists a multitude of methods for securing measurements internally (i.e. encryption), defenses against attacks on the process the sensors are measuring are relatively weak. Such attacks include spoofing of the GPS signal, attaching magnetic wave altering devices to the anti-lock brake sensors of vehicles, and placing a heating coils near temperature sensors. Thus, to ensure the safe operation of systems whose sensory environments can be altered requires attack defenses beyond sensor encryption. To address these issues, we have introduced a design framework for development of high-confidence control systems that can be used in adversarial environments [6]. The framework employs system design techniques that guarantee that the system will maintain a minimum performance, possibly at a reduced efficiency, under several classes of attacks. Specific to this work, we consider the design of estimators which are resilient to maliciously altered sensor measurements.

James Weimer, Nico Bezzo, Miroslav Pajic, Oleg Sokolsky, and Insup Lee are with the PRECISE Center, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104, USA {weimerj, nicbezzo, pajic}@seas.upenn.edu {sokolsky, lee}@cis.upenn.edu

Literature review: The design of estimators which are resilient against faults have been addressed from many points of view, including fault detection [15], robust control [8], adaptive control [1], and more generally from estimation and hypothesis testing [12]. In general, these approaches address the issue of maximizing some performance measure with respect to a known or bounded disturbances. In the context of security against malicious attacks, many of these approaches are not applicable because of their assumption that the attack is either known or bounded, with notable exceptions being approaches which ask for invariance to the unknown parameters (or attacks) [9]. The remainder of this literature review focuses on secure estimation.

Secure estimation and control system design in the presence of disturbances or attacks has received increasing research interest [10], [5], [7], [11], [14], [13]. Most closely related to the work presented herein is [3], which addresses the secure estimation and control of linear deterministic systems under malicious sensor attacks. While the approach in [3] is shown to stabilize the systems under consideration, their approach does not consider any statistical properties of the measurements.

Statement of contributions: Beyond the previous approaches, this work focuses on the design of resilient estimators for stochastic systems. The primary technical contributions of this work are: (a) a mathematical formulation of the design problem for minimum mean-squared error resilient estimation in stochastic systems; (b) a resilient estimator that achieves the minimum variance; (c) an implementation of the minimum mean-squared error resilient estimator using semi-definite programming; (d) an evaluation of the minimum variance resilient estimator against other resilient estimation strategies.

Structure of the paper: Section II identifies notation and preliminary definitions that will be utilized throughout the paper. Section III formulates the minimum variance resilient estimator design problem. A discussion regarding classical fault-tolerant estimation is provided in sectionIV while section V presents a candidate minimum mean-squared error resilient estimator, proves the candidate satisfies the design problem, and provides a semi-definite programming implementation. A simulated comparison against other estimation strategies is included in section VI, and the final section provides discussion and future research directions.

II. Notation and Preliminaries

This section introduces notation and preliminaries that prove useful in this work. We use $\mathbb{E}[y \mid z]$ and $\mathrm{Cov}[y \mid z]$ to denote the expected value and covariance, respectively,

978-1-4799-3274-0/$31.00 ©2014 AACC

of y conditioned on z. For a matrix $\boldsymbol{X}$, we write $\vec{\boldsymbol{X}}$ to denote the vector of the concatenated columns of $\boldsymbol{X}$, and we write $\operatorname{diag}(\boldsymbol{X})$ to denote the vector formed from the diagonal elements of $\boldsymbol{X}$. We employ this notation to recall the definition of a minimum variance unbiased (MVUB) estimator [12]

Definition 1 ***Minimum Variance Unbiased (MVUB) Estimator***:
Given stochastic measurements $\boldsymbol{y}$, a stochastic parameter (or state) $\boldsymbol{x}$, and an estimator, $\hat{\boldsymbol{x}}(\boldsymbol{y})$, the estimator $\hat{\boldsymbol{x}}$ is said to be *unbiased* if

$$\mathbb{E}\left[\hat{\boldsymbol{x}}(\boldsymbol{y})\right] = \mathbb{E}\left[\boldsymbol{x}\right]$$

and, assuming $\mathcal{T} := \{\boldsymbol{t} \mid \mathbb{E}\left[\boldsymbol{t}(\boldsymbol{y})\right] = \mathbb{E}\left[\boldsymbol{x}\right]\}$, have *minimum variance* if

$$\forall \boldsymbol{t} \in \mathcal{T}, \;\; \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{y})\|^2\right] \leq \mathbb{E}\left[\|\boldsymbol{x} - \boldsymbol{t}(\boldsymbol{y})\|^2\right].$$

Assuming the first two central moments of $\boldsymbol{x}$ and $\boldsymbol{y}$ are

$$\mathbb{E}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} = \begin{bmatrix}\tilde{\boldsymbol{m}}_x\\ \tilde{\boldsymbol{m}}_y\end{bmatrix}, \;\; \operatorname{Cov}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xy}\\ \boldsymbol{\Sigma}_{xy}^\top & \boldsymbol{\Sigma}_y\end{bmatrix}$$

then the minimum variance unbiased estimator, $\tilde{\boldsymbol{x}}(\boldsymbol{y})$, and its corresponding covariance, $\tilde{\boldsymbol{\Sigma}}$, of $\boldsymbol{x}$, as a function of $\boldsymbol{y}$, is

$$\begin{aligned}\tilde{\boldsymbol{x}}(\boldsymbol{y}) &= \tilde{\boldsymbol{m}}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}\left(\boldsymbol{y} - \tilde{\boldsymbol{m}}_y\right)\\ \tilde{\boldsymbol{\Sigma}}_x &= \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{xy}^\top.\end{aligned} \tag{1}$$

The notation and definitions introduced in this section are used in the following to motivate and formulate the minimum mean-squared error (MMSE) estimator.

III. Problem Formulation

In this work, we consider the problem of designing resilient estimators when an unknown subset of the measurements are altered by unmodeled additive attacks (possibly malicious in nature). Abstractly, we wish to solve the problem of finding an estimator, restricted to the class of potential attack-resilient estimators, that generates a *minimum mean-squared error* estimate of the parameter (or state).

We assume there exists a signal $\boldsymbol{s} \in \mathcal{S} \subseteq \mathbb{R}^N$, a zero-mean noise $\boldsymbol{n} \in \mathcal{N} \subseteq \mathbb{R}^M$ having covariance $\boldsymbol{\Sigma}_n$, an unmodeled sparse attack vector $\boldsymbol{d} \in \mathcal{D} \subset \mathbb{R}^N$, an observable state, $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^N$, and measurements, $\boldsymbol{y} \in \mathcal{Y} \subseteq \mathbb{R}^M$, linearly related as

$$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} = \begin{bmatrix}\boldsymbol{F}_{xs} & \boldsymbol{F}_{xn} & \boldsymbol{0}\\ \boldsymbol{F}_{ys} & (\boldsymbol{I} - \boldsymbol{\Gamma})\boldsymbol{F}_{yn} & \boldsymbol{\Gamma}\end{bmatrix}\begin{bmatrix}\boldsymbol{s}\\ \boldsymbol{n}\\ \boldsymbol{d}\end{bmatrix} \tag{2}$$

where $\boldsymbol{F}_{xs} \in \mathbb{R}^{N\times N}$, $\boldsymbol{F}_{xn} \in \mathbb{R}^{N\times M}$, $\boldsymbol{F}_{ys} \in \mathbb{R}^{M\times N}$ and $\boldsymbol{F}_{yn} \in \mathbb{R}^{M\times M}$ denote the linear mapping of the signal and noise, $(\boldsymbol{s}, \boldsymbol{n})$, into the observable state and measurements, $(\boldsymbol{x}, \boldsymbol{y})$ and $\boldsymbol{\Gamma} \in \operatorname{diag}\left(\{0,1\}^M\right)$ is the *attack matrix* denoting the elements of $\boldsymbol{d}$ which are non-zero, such that

$$\boldsymbol{d} = \boldsymbol{\Gamma}\boldsymbol{d} \;\; \text{and} \;\; \boldsymbol{0} = (\boldsymbol{I} - \boldsymbol{\Gamma})\boldsymbol{d}.$$

We assume that J sensors are used to collect the M measurements such that the non-zero entries of $\boldsymbol{d}$ (or equivalently the unit entries of $\boldsymbol{\Gamma}$) equates to whether the corresponding sensor used to collect the measurement was attacked. Mathematically, we represent this by assuming an *attack vector* , $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_J]^\top \in \Theta \subseteq \{0,1\}^J$, exists where $\theta_j = 1$ ($\theta_j = 0$) if sensor j is attacked (not attacked), such that by claiming[1] $\boldsymbol{y} = \left[\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_J^\top\right]^\top$, where $\boldsymbol{y}_j$ denotes the measurements corresponding to sensor j, and $\boldsymbol{\Gamma}$ can be defined by $\boldsymbol{\theta}$ as

$$\boldsymbol{\Gamma} = \begin{bmatrix}\theta_1\boldsymbol{I} & & \\ & \ddots & \\ & & \theta_J\boldsymbol{I}\end{bmatrix}. \tag{3}$$

In the work, we assume that the attack, $\boldsymbol{d}$, is stealthy:

Definition 2 ***Stealthy attack*:** An attack is considered stealthy if:

$$\forall \boldsymbol{\theta} \in \Theta, \; \Pr\left[\boldsymbol{\theta}|\boldsymbol{y}\right] < 1 \tag{4}$$

where, in words, an attack is stealthy if, after sampling, there is more than one attack vector, $\boldsymbol{\theta} \in \Theta$, with a non-zero probability.

We aim to design an estimator for $\boldsymbol{x}$ of the form,

$$\hat{\boldsymbol{x}}(\boldsymbol{L}) = \left(\boldsymbol{F}_{xs} - \boldsymbol{L}\boldsymbol{F}_{ys}\right)\boldsymbol{s} + \boldsymbol{L}\boldsymbol{y} \tag{5}$$

where $\boldsymbol{L} \in \mathcal{L} \subseteq \mathbb{R}^{N\times M}$ denotes the estimator gain, selected such that the estimate $\hat{\boldsymbol{x}}(\boldsymbol{L})$ is:

- *Resilient*: $\forall \boldsymbol{\theta} \in \Theta$,

$$\mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\hat{\boldsymbol{L}})\|^2|\boldsymbol{\theta}\right] \leq \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\hat{\boldsymbol{L}})\|^2|\hat{\boldsymbol{\theta}}\right]$$

- *Minimum Variance*: $\forall \boldsymbol{L} \in \mathcal{L}$,

$$\mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\hat{\boldsymbol{L}})\|^2|\hat{\boldsymbol{\theta}}\right] \leq \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{L})\|^2|\hat{\boldsymbol{\theta}}\right].$$

The resilient requirement specifies that the minimum mean-squared error (MMSE) of the estimate is maximized at $\hat{\boldsymbol{\theta}}$, while the minimum variance requirement enforces that the MMSE of the estimate is minimized at $\hat{\boldsymbol{L}}$. Thus, the problem of finding the minimum mean-squared error resilient (MMSE-R) estimator can be stated as

Problem 1 ***Design a Minimum Mean-Squared Error Resilient (MMSE-R) Estimator*:** Minimize $\mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\hat{\boldsymbol{L}})\|^2|\hat{\boldsymbol{\theta}}\right]$ subject to the constraint:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{L}} = \arg\max_{\boldsymbol{\theta}\in\Theta}\min_{\boldsymbol{L}\in\mathcal{L}} \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{L})\|^2|\boldsymbol{\theta}\right].$$

[1]We note that the structured claim on $\boldsymbol{y}$ does not limit the generality of this formulation since it can always be satisfied through a re-ordering of the elements of $\boldsymbol{y}$, which is equivalent to a rotation in $\mathcal{Y}$ (known to be an information preserving one-to-one mapping).

The problem in 1 is a maximin (or minimax) problem, where the solution for the resilient estimator gain, $\boldsymbol{L}$, occurs at the saddle point of the mean-squared error function. We note that problem 1 represents a statistical game, where the defender first chooses the estimator gain, $\boldsymbol{L}$, conservatively such that the attackers best option is to select the attack vector, θ, equal to $\hat{\boldsymbol{\theta}}$. Before deriving the MMSE-R estimator in section V, the following section discusses the classical approach to fault tolerant estimation typically used to determine whether sensors are faulty.

IV. Classical Fault-Tolerant Estimation

In this section we review the classical fault-tolerant estimation approach employed for detecting, identifying, and removing faulty sensor measurements. The classical fault tolerant estimator consists of two sequential steps: (a) fault-detection and isolation and (b) MMSE estimator design. The fault-detection and isolation scheme is employed to identify the subset of sensors functioning properly (i.e. not attacked). Once a subset of sensors are identified as functioning, then a MMSE estimator is designed. There are a multitude of approaches to fault-tolerant estimation [15], [10], [5], [7], [11], where some strategies employ temporal reasoning to correct incorrect decisions. In the most general of senses, a fault-tolerant estimator can be expressed mathematically as:

- ***Fault-Detection and Identification*:** Given $\hat{\boldsymbol{y}} \in \mathcal{Y}$,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \Pr\left[\boldsymbol{\theta}|\hat{\boldsymbol{y}}\right] \tag{6}$$

- ***MMSE Estimator Design*:** Given $\hat{\boldsymbol{\theta}} \in \Theta$,

$$\hat{\boldsymbol{L}} = \arg\min_{\hat{\boldsymbol{L}}\in\mathcal{L}} \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{L})\|^2|\hat{\boldsymbol{\theta}}\right] \tag{7}$$

The advantage of the fault-tolerant estimator design is that once $\hat{\boldsymbol{\theta}}$ is identified, the MMSE estimator can be designed similar to the MVUB estimator in section II by assuming $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ (i.e. $\hat{\boldsymbol{\theta}}_j = 0 \leftrightarrow \boldsymbol{d}_j = \boldsymbol{0}$) and using only the safe sensors to estimate $\boldsymbol{x}$ [15]. Moreover, as $\Pr\left[\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{y}}\right]$ approaches one, the fault-tolerant estimator becomes MVUB

A disadvantage of the fault-tolerant estimator, and one of the primary contributions of this work, is given in the following proposition:

Proposition 1 The fault-tolerant estimator in (6) and (7) is not a minimum mean-squared error resilient (MMSE-R) estimator.

Proof: To be an MMSE-R estimator requires $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{L}}$ to be chosen according to the constraint in (1). Thus. the fault-tolerant estimator can only be MMSE-R if there exists a $\hat{\boldsymbol{\theta}} \in \Theta$ such that

$$\min_{\hat{\boldsymbol{L}}\in\mathcal{L}} \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{L})\|^2|\hat{\boldsymbol{\theta}}\right] = \max_{\boldsymbol{\theta}\in\Theta} \min_{\boldsymbol{L}\in\mathcal{L}} \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{L})\|^2|\boldsymbol{\theta}\right] \tag{8}$$

implying that $\Theta = \{\hat{\boldsymbol{\theta}}\}$, or equivalently,

$$\Pr\left[\hat{\boldsymbol{\theta}}|\boldsymbol{y}\right] = 1. \tag{9}$$

This is contradicted by the assumption that the attack, $\boldsymbol{d}$, is stealthy in definition 2 ■

The above proposition illustrates that as the probability of error for the fault-detection and identification portion of the fault-tolerant estimator increases, it is less likely to be a resilient estimate. When $\boldsymbol{d}$ represents a non-malicious disturbance, it may be likely that the probability of identifying the faulty sensors is high and the fault-tolerant estimate will yield satisfactory results (this is consistent with its overwhelming use in real-world applications). However, when $\boldsymbol{d}$ is malicious in nature, one goal of the attacker is to remain undetected (i.e. maximize the probability that the fault-detector will yield an incorrect result). In the presence of potential malicious attacks, an estimator that minimizes the worst case scenario (such as the MMSE-R estimator), is likely to yield more accurate estimates. This point will be emphasized through a case study in section VI.

V. Minimum Mean-Squared Error Resilient Estimation

This section formulates the minimum mean-squared error resilient (MMSE-R) estimator and presents a semidefinite-programming implementation. We begin by presenting the MMSE-R estimator, a primary contribution of this work, in the following proposition.

Proposition 2 ***MMSE-R Estimator*** **:** The MMSE-R estimator is $\hat{\boldsymbol{x}} = \boldsymbol{m}_x + \hat{\boldsymbol{L}}(\boldsymbol{y} - \boldsymbol{m}_y)$, where

$$\hat{\boldsymbol{L}}, \hat{\sigma} = \arg\min_{\boldsymbol{L},\sigma} \quad \sigma$$
$$s.t. \quad \sigma \geq \vec{\boldsymbol{L}}^\top \boldsymbol{A}_{\boldsymbol{\theta}} \vec{\boldsymbol{L}} - 2\vec{\boldsymbol{B}}_{\boldsymbol{\theta}}^\top \vec{\boldsymbol{L}} + \mathrm{Tr}\left[\boldsymbol{\Sigma}_x\right], \quad \forall \boldsymbol{\theta} \in \Theta$$

assuming

$$\begin{aligned}
\boldsymbol{A}_{\boldsymbol{\theta}} &= \left(\left(\boldsymbol{I} - \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\right)\boldsymbol{\Sigma}_y\left(\boldsymbol{I} - \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\right) + \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{r}\boldsymbol{r}^\top\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\right) \otimes \boldsymbol{I}_N \\
\boldsymbol{B}_{\boldsymbol{\theta}} &= \boldsymbol{\Sigma}_{xy}\left(\boldsymbol{I} - \boldsymbol{\Gamma}\right), \\
\boldsymbol{\Sigma}_x &= \boldsymbol{F}_{xn}\boldsymbol{\Sigma}_n\boldsymbol{F}_{xn}^\top, \qquad \boldsymbol{r} = \hat{\boldsymbol{y}} - \boldsymbol{F}_{ys}\boldsymbol{s} \\
\boldsymbol{\Sigma}_y &= \boldsymbol{F}_{yn}\boldsymbol{\Sigma}_n\boldsymbol{F}_{yn}^\top \qquad \boldsymbol{\Sigma}_{xy} = \boldsymbol{F}_{xn}\boldsymbol{\Sigma}_n\boldsymbol{F}_{yn}^\top.
\end{aligned}$$

Proof: The proof is provided in the appendix. ■

The minimization problem used to determine the MMSE-R estimator is a convex quadratically constrained optimization problem which can be evaluated efficiently using convex optimization techniques.

Different from the fault-tolerant estimator discussed in the previous section, the MMSE-R estimator minimizes a worst case bounds on the mean-squared error. This results in an estimate that simultaneously minimizes the expected deviation for all attack vectors, $\boldsymbol{\theta} \in \Theta$. The MMSE-R estimator only accepts measurements when they are likely to improve the mean-squared error of the estimate. For this reason, and unlike other approaches to estimation, the

MMSE-R is likely to reject all the sensors if the prior signal, $\boldsymbol{s}$, is biased. In this work, we focus only on the condition when the prior signal is known, and leave the design of resilient statistical estimators for unknown priors as a subject of future work. The remainder of this section discusses the robustness of the MVR estimator.

While the MMSE-R estimator ensures the estimate is resilient to measurement attacks, a quantitative measure of its resilience is preferable. Since the MMSE-R estimator was designed to minimize the worst case mean-squared error, the following lemma provides an upper bounds for the probability of the state estimate diverging:

Lemma 1 ***Robustness of the MMSE Estimator*** **:** For any positive threshold η, the probability that the MVR estimator error exceeds η is:

$$\Pr\left[\|\boldsymbol{x}-\hat{\boldsymbol{x}}(\hat{\boldsymbol{L}})\|^2 \geq \eta|\boldsymbol{\theta}\right] \leq \frac{\hat{\sigma}}{\eta}$$

Proof: A direct consequence of the Markov inequality [2]. ∎

The robustness of the MMSE-R estimator identifies an upper bound on the probability that the estimated state diverges from the true state for a given measurement, $\hat{\boldsymbol{y}}$. While a runtime evaluation of the robustness is useful to identify when the MMSE-R estimator is inaccurate, a worst-case evaluation of the estimator robustness is preferred to ensure the state estimate remains accurate for all attacks. For the MMSE-R estimator, the worst case robustness is provided by solving the following minimization problem

$$\max_{\boldsymbol{y}\in\mathcal{Y}} \min_{\boldsymbol{L},\sigma} \quad \sigma$$
$$s.t. \quad \sigma \geq \vec{\boldsymbol{L}}^{\top}\boldsymbol{A}_{\boldsymbol{\theta}}\vec{\boldsymbol{L}} - 2\vec{\boldsymbol{B}}_{\boldsymbol{\theta}}^{\top}\vec{\boldsymbol{L}} + \mathrm{Tr}\left[\boldsymbol{\Sigma}_x\right], \quad \forall\boldsymbol{\theta}\in\Theta$$

Similar to the MMSE-R design problem in 1, solving the worst-case robustness problem requires solving a maximin optimization problem. Since the set of potential measurements $\mathcal{Y}$ is not countable, the maximin problem can not be solved in the same manner as the MMSE-R design problem (i.e. by generating a constraint for each element of the set). Thus, we leave the problem of finding the worst-case mean-squared error as a subject of future work.

VI. Simulation Results

In this section, we present a case study to illustrate the performance of the MMSE-R estimator. For comparison, we compare the MMSE-R estimator to two other approaches, namely:

- **Non-Resilient Estimator:** The minimum variance unbiased estimator from section II,
- **Optimal Fault Tolerant Estimator:** The fault tolerate detector described in section IV where the detector exactly removes the attack (i.e. no errors in the fault detection and identification step).

Each table below represents different attacks scenarios. Specifically, each entry of the table contains two values: the upper number represents the Mean Square Error (MSE) while the lower number is the ratio between the MSE of the specific state estimator under analysis and the MSE of an optimal estimator which has full knowledge of the attack. In the tables, we denote the optimal detector-estimator approach as *ODE*, the minimum mean-squared error resilient estimator as *MMSE-R*, and the non-resilient estimator as *N-R*. In this evaluation, we consider the following linear system

$$\begin{aligned} x(k+1) &= .8x(k) + 1 + w(k) \\ \boldsymbol{y}(k) &= \boldsymbol{1}x(k) + \boldsymbol{v}(k) \end{aligned} \tag{10}$$

where $\boldsymbol{y} \in \mathbb{R}^5$, $x \in \mathbb{R}$, and w and $\boldsymbol{v}$ are the zero-mean i.i.d. process and measurement noises, respectively. We consider different combinations of the process covariance, $\sigma_w \in \{\sigma_w(1), \sigma_w(2)\}$, and measurement covariance, $\Sigma_v \in \{\Sigma_v(1), \Sigma_v(2)\}$, namely

$$\begin{aligned} \sigma_w(1) &= 0.1 \\ \sigma_w(2) &= 1.0 \\ \Sigma_v(1) &= \mathrm{diag}\left([1,1,1,1,1,]\right) \\ \Sigma_v(2) &= \mathrm{diag}\left([10,1,10,1,10]\right) \end{aligned} \tag{11}$$

This system is evaluated assuming a window of 6 samples for the MVR and DR strategies under 5 sensor attack scenarios, namely : no attacks, 1 small attack, 2 small attacks, 1 large attack, and 2 large attacks, where a small attack is assumed to have a magnitude of less than 1 while a large attack has a magnitude of 10. The results of those evaluations are shown in the following tables.

TABLE I
MEAN SQUARED ERROR OF STATE ESTIMATE ASSUMING NO ATTACKS

Approach	$\sigma_w(1)$ $\Sigma_v(1)$	$\sigma_w(2)$ $\Sigma_v(1)$	$\sigma_w(1)$ $\Sigma_v(2)$	$\sigma_w(2)$ $\Sigma_v(2)$
ODE	0.0540 1.00	0.1044 1.00	0.0947 1.00	0.2443 1.00
MMSE-R	0.0995 1.84	0.2895 2.77	0.1518 1.60	0.4424 1.81
N-R E	0.0540 1.00	0.1044 1.00	0.0947 1.00	0.2443 1.00

In Table I, we consider the case of no attacks and observe that the ODE and N-R estimators performs better than the MMSE-R estimator obtaining the same MSE for all different noise profile combinations. This is a direct result of the fact that the ODE and N-R estimators being optimal when no attack is present, while the MMSE-R estimator occasionally rejects information that was actually non-attacked. We observe that the MMSE-R has a worst case performance that is within a factor of 2.77 of the optimal strategies.

When one small attack on one sensor is considered, as in Table II, we notice that both ODE and MMSE-R estimators have a small difference. Recalling the ODE approach is the optimal detector-estimator, the proximity of the MMSE-R to the ODE approach suggests that it may be better than

TABLE II

MEAN SQUARED ERROR OF STATE ESTIMATE ASSUMING ONE SMALL ATTACK

Approach	$\sigma_w(1)$ $\Sigma_v(1)$	$\sigma_w(2)$ $\Sigma_v(1)$	$\sigma_w(1)$ $\Sigma_v(2)$	$\sigma_w(2)$ $\Sigma_v(2)$
ODE	0.0632 1.17	0.1587 1.52	0.0952 1.01	0.2884 1.18
MMSE-R	0.0651 1.21	0.1612 1.54	0.0953 1.01	0.2932 1.20
N-R E	0.1154 2.14	0.2124 2.03	0.1214 1.28	0.3172 1.29

a realizable ODE. Studying when the MMSE-R provides a better MSE than a detector-estimator approach is a subject of future work and will be evaluated on an application-by-application basis. The non-resilient estimator has about a factor of 2 worse performance than the MMSE-R approach.

TABLE III

MEAN SQUARED ERROR OF STATE ESTIMATE ASSUMING TWO SMALL ATTACKS

Approach	$\sigma_w(1)$ $\Sigma_v(1)$	$\sigma_w(2)$ $\Sigma_v(1)$	$\sigma_w(1)$ $\Sigma_v(2)$	$\sigma_w(2)$ $\Sigma_v(2)$
ODE	0.0729 1.35	0.1077 1.03	0.2520 2.66	0.4369 1.79
MMSE-R	0.0998 1.85	0.2899 2.78	0.3366 3.55	0.6271 2.57
N-R E	0.3768 6.98	0.6186 5.93	0.3692 3.90	0.6688 2.74

When we increase the number of small attacks, as in Table III, we notice that the relative performance of the MMSE-R strategy to the ODE is about the same as when considering only a single attack; however, the non-resilient and deterministic performance degrades significantly. We notice that the relative performance decrease of the non-resilient estimator is significantly greater when $\Sigma_v = \Sigma_v(1)$ as opposed to when $\Sigma_v = \Sigma_v(2)$. This is a direct result of the fact that the non-resilient estimator places greater faith in sensor 2 (assumed under attack in this scenario) when assuming $\Sigma_v(1)$ as opposed to $\Sigma_v(2)$. This illustrates the importance of ensuring that the most reliable sensors are secure prior to including them in the state estimate.

Table IV shows the performance of the three estimators when one large attack is injected in one of the sensors. Consistent with the previous results, we observe that the MMSE-R and ODE approaches have much better relative performance than the non-resilient approaches. In comparison to the results when one small attack is assumed (i.e. Table II), we notice that the MMSE-R and oracle have identical performances under both scenarios.

TABLE IV

MEAN SQUARED ERROR OF STATE ESTIMATE ASSUMING ONE LARGE ATTACK

Approach	$\sigma_w(1)$ $\Sigma_v(1)$	$\sigma_w(2)$ $\Sigma_v(1)$	$\sigma_w(1)$ $\Sigma_v(2)$	$\sigma_w(2)$ $\Sigma_v(2)$
ODE	0.0632 1.17	0.1587 1.52	0.0952 1.01	0.2844 1.16
MMSE-R	0.0745 1.38	0.2798 2.68	0.1353 1.43	0.3081 1.26
N-R E	2.362 43.74	3.6254 34.72	0.2071 2.19	0.9257 3.80

TABLE V

MEAN SQUARED ERROR OF STATE ESTIMATE ASSUMING TWO LARGE ATTACKS

Approach	$\sigma_w(1)$ $\Sigma_v(1)$	$\sigma_w(2)$ $\Sigma_v(1)$	$\sigma_w(1)$ $\Sigma_v(2)$	$\sigma_w(2)$ $\Sigma_v(2)$
ODE	0.0629 1.16	0.1077 1.03	0.1243 1.31	0.3215 1.32
MMSE-R	0.1440 2.67	0.2952 2.83	0.1838 1.94	0.4527 1.85
N-R E	9.6730 179.12	14.5986 139.83	9.0457 95.52	18.2776 74.82

The final table, Table V, shows the case of large attacks on two sensors measurements. Overall we notice that the ODE performs better than the other approaches followed in order by the MMSE-R estimator and the N-R estimators. The MMSE-R performs well when the variance of the noise is low and the attacks are small. The N-R estimator performs well if there are no attacks otherwise its error diverges as we increase the number of sensor under attacks and the magnitude of the attacks.

VII. DISCUSSION AND FUTURE WORK

In this work, we introduced a minimum mean-squared error resilient (MMSE-R) estimator for stochastic systems. The results indicate that the MMSE-R estimator performs well as compared to an optimal detector-estimator which assumes full knowledge of the attack space (i.e. which sensors are attacked). Observing that the optimal detector-estimator is unrealizable it is a subject of future work to evaluate the performance of the MMSE-R estimator on an application-by-application basis to determine scenarios when the MMSE-R outperforms a detector-estimator approach and vice versa. Under the assumption that the MMSE-R estimator has accurate knowledge of the statistical profile of the sensors, it is shown to significantly outperform a deterministic resilient estimator (i.e. a resilient estimator that does not consider the statistical profile of the measurements). As another future research direction, we intend to analytically evaluate how errors in the noise profile assumptions affect

the resilience of the MMSE-R estimator.

ACKNOWLEDGMENTS

This work is based on research sponsored by DARPA under agreement number FA8750-12-2-0247. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

REFERENCES

[1] Karl Johan Astrom and Bjorn Wittenmark. *Adaptive Control*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 1994.

[2] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 1991.

[3] Hamza Fawzi, Paulo Tabuada, and Suhas N. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *CoRR*, abs/1205.5073, 2012.

[4] M. Grant and S. Boyd. Cvx: Matlab software for disciplined convex programming [web page and software], October 2007.

[5] A. Gupta, C. Langbort, and T. Basar. Optimal control in the presence of an intelligent jammer with limited actions. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 1096 –1101, dec. 2010.

[6] Miroslav Pajic, Nicola Bezzo, James Weimer, Rajeev Alur, Rahul Mangharam, Nathan Michael, George J Pappas, Oleg Sokolsky, Paulo Tabuada, Stephanie Weirich, et al. Towards synthesis of platform-aware attack-resilient control systems. In *Proceedings of the 2nd ACM international conference on High confidence networked systems*, pages 75–76. ACM, 2013.

[7] Fabio Pasqualetti, Florian Drfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems – part ii: Centralized and distributed monitor design. Technical Report arXiv:1202.6049, Feb 2012.

[8] L. Qiu. Essentials of robust control Kemin Zhou, John C. Doyle Prentice-Hall, Englewood Cliffs, NJ, 1998, ISBN: 0-13-790874-1. *Automatica*, 38(5):910–912, May 2002.

[9] Louis L. Scharf. *Statistical Signal Processing, Detection, Estimation, and Time Series Analysis*. Addison-Welsley Publishing Company Inc., Reading, Massachusetts, 1991.

[10] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S.S. Sastry. Foundations of control and estimation over lossy networks. *Proceedings of the IEEE*, 95(1):163 –187, jan. 2007.

[11] Shreyas Sundaram, Miroslav Pajic, Christoforos N. Hadjicostis, Rahul Mangharam, and George J. Pappas. The wireless control network: Monitoring for malicious behavior. In *CDC*, pages 5979–5984, 2010.

[12] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory*. John Wiley & Sons, Inc., New York, 1968.

[13] James Weimer, Seyed A. Ahmadi, Jose Araujo, Francesca M. Mele, Dario Papale, Iman Shames, Henrik Sandberg, and Karl H. Johansson. Active actuator fault detection and diagnostics in hvac systems. In *4th ACM Workshop On Embedded Systems For Energy-Efficiency In Buildings (BuildSys)*, Toronto, Canada, 2012.

[14] James Weimer, Soummya Kar, and Karl Henrik Johansson. Distributed detection and isolation of topology attacks in power networks. In *Proceedings of the 1st international conference on High Confidence Networked Systems*, HiCoNS '12, pages 65–72, New York, NY, USA, 2012. ACM.

[15] A. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12:601–611, 1976.

APPENDIX

To prove the estimator in proposition 2 is the MMSE-R that solves problem 1, we begin by writing the first and second central moments of $\boldsymbol{x}$ and $\boldsymbol{y}$ as

$$\begin{aligned}\mathbb{E}\left[\begin{array}{c}\boldsymbol{x}\\ \boldsymbol{y}\end{array}\right] &= \left[\begin{array}{c}\boldsymbol{F}_{xs}\boldsymbol{s}\\ \boldsymbol{F}_{ys}\boldsymbol{s}+\boldsymbol{d}\end{array}\right]\\ \mathrm{Cov}\left[\begin{array}{c}\boldsymbol{x}\\ \boldsymbol{y}\end{array}\right] &= \left[\begin{array}{cc}\boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xy}(\boldsymbol{I}-\boldsymbol{\Gamma})\\ (\boldsymbol{I}-\boldsymbol{\Gamma})\boldsymbol{\Sigma}_{xy}^\top & (\boldsymbol{I}-\boldsymbol{\Gamma})\,\boldsymbol{\Sigma}_y\,(\boldsymbol{I}-\boldsymbol{\Gamma})\end{array}\right]\\ &\quad+\left[\begin{array}{cc}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \mathbb{E}\left[\boldsymbol{d}\boldsymbol{d}^\top|\boldsymbol{\theta}\right]\end{array}\right]\end{aligned} \tag{12}$$

where, given a realization (sampling) of $\boldsymbol{y}$, $\hat{\boldsymbol{y}}$, the maximum likelihood estimate of the covariance of $\boldsymbol{d}$ is given by the sample covariance,

$$\mathbb{E}\left[\boldsymbol{d}\boldsymbol{d}^\top|\boldsymbol{\theta}\right] = \boldsymbol{\Gamma}\left(\hat{\boldsymbol{y}}-\boldsymbol{m}_y\right)\left(\hat{\boldsymbol{y}}-\boldsymbol{m}_y\right)^\top\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\boldsymbol{r}\boldsymbol{r}^\top\boldsymbol{\Gamma}.$$

It then follows that:

$$\begin{aligned}&\mathbb{E}\left[\|\boldsymbol{x}-\hat{\boldsymbol{x}}(\boldsymbol{L})\|^2|\boldsymbol{\theta}\right]\\ &=\mathbb{E}\left[\|\boldsymbol{x}-\left(\left(\boldsymbol{F}_{xs}-\boldsymbol{L}\boldsymbol{F}_{ys}\right)\boldsymbol{s}+\boldsymbol{L}\boldsymbol{y}\right)\|^2|\boldsymbol{\theta}\right]\\ &=\mathbb{E}\left[\|\boldsymbol{x}-\boldsymbol{F}_{xs}\boldsymbol{s}-\boldsymbol{L}\left(\boldsymbol{y}-\left(\boldsymbol{F}_{ys}\boldsymbol{s}+\boldsymbol{d}\right)+\boldsymbol{d}\right)\|^2|\boldsymbol{\theta}\right]\\ &=\mathbb{E}\left[\|\boldsymbol{x}-\mathbb{E}\left[\boldsymbol{x}\right]-\boldsymbol{L}\left(\boldsymbol{y}-\mathbb{E}\left[\boldsymbol{y}\right]\right)-\boldsymbol{L}\boldsymbol{d}\|^2|\boldsymbol{\theta}\right]\\ &=\mathrm{Tr}\left[\boldsymbol{\Sigma}_x-2\boldsymbol{L}(\boldsymbol{I}-\boldsymbol{\Gamma})\boldsymbol{\Sigma}_{xy}^\top\right]\\ &\quad+\mathrm{Tr}\left[\boldsymbol{L}(\boldsymbol{I}-\boldsymbol{\Gamma})\boldsymbol{\Sigma}_y(\boldsymbol{I}-\boldsymbol{\Gamma})\boldsymbol{L}^\top\right]\\ &\quad+\mathrm{Tr}\left[\boldsymbol{L}\boldsymbol{\Gamma}\left(\mathbb{E}\left[\boldsymbol{d}\boldsymbol{d}^\top|\boldsymbol{\theta}\right]\right)\boldsymbol{\Gamma}\boldsymbol{L}^\top\right]\\ &=\mathrm{Tr}\left[\boldsymbol{\Sigma}_x\right]-2\mathrm{Tr}\left[(\boldsymbol{I}-\boldsymbol{\Gamma})\boldsymbol{\Sigma}_{xy}^\top\boldsymbol{L}\right]\\ &\quad+\mathrm{Tr}\left[\boldsymbol{L}\left((\boldsymbol{I}-\boldsymbol{\Gamma})\boldsymbol{\Sigma}_y(\boldsymbol{I}-\boldsymbol{\Gamma})+\boldsymbol{\Gamma}\boldsymbol{r}\boldsymbol{r}^\top\boldsymbol{\Gamma}\right)\boldsymbol{L}^\top\right]\\ &=\vec{\boldsymbol{L}}^\top\boldsymbol{A}_{\boldsymbol{\theta}}\vec{\boldsymbol{L}}-2\vec{\boldsymbol{B}}_{\boldsymbol{\theta}}^\top\vec{\boldsymbol{L}}+\mathrm{Tr}\left[\boldsymbol{\Sigma}_x\right]\end{aligned}$$