Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-22 (2016) 055–060

Resilience and Performance Analysis for State Estimation against Integrity Attacks

Duo Han * **Yilin Mo** * **Lihua Xie** *

* *School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798, (e-mail: dhanaa,ylmo,elhxie@ntu.edu.sg)*

Abstract: We consider the problem of resilient dynamical state estimation in the presence of integrity attacks. We conduct resilience and performance analysis for a convex optimization based estimator. The employed approach for analyzing resilience of an estimator is novel and generic for a wide class of estimators and thus can achieve greater generality, as long as an estimator can be decomposed into a convex optimization based form. We show sufficient and necessary conditions for resilience with a trivial gap. The tradeoff between minimum mean square error (MMSE) optimality and resilience is well investigated. Due to the constructive proof of the main results using the force analogy, we present an upper bound on the damage an attacker can cause when the sufficient condition is satisfied. Simulation results are also given to validate the resilience and performance analysis.

Keywords: Cyber-physical security; resilient estimation; integrity attack; convex optimization

1. INTRODUCTION

Cyber-physical security has received much research attention (Cárdenas et al., 2008) in last decade. Typically, the sensors are vulnerable to integrity attacks since in most cases they are spatially distributed and cannot be fully protected. Compromised data detection via fault detection and isolation based methods has been extensively studied, (Pasqualetti et al., 2010, 2011; Fawzi et al., 2012; Chong et al., 2015). However, in most of these works, the system is assumed to be noiseless, which greatly favors the failure detector. In (Mo et al., 2010), the authors studied the worst bias an attack can cause through reachability analysis and ellipsoid approximation. In (Mishra et al., 2015; Chong et al., 2015; Shoukry and Tabuada, 2015; Mo and Murray, 2015), they designed different robust estimators grounded on the so-called sparsity observability conditions.

If the fundamental sparsity observability conditions is violated, no resilient estimator exists. This, however, is not enough. Instead of designing resilient estimators like (Fawzi et al., 2012; Mishra et al., 2015; Chong et al., 2015; Shoukry and Tabuada, 2015; Mo and Murray, 2015), we conduct the research in a reverse way. We make our efforts to answer the questions: given an estimator, what can we talk about its resilience and performance? And are there any systematic analysis procedures we can use? We illustrate an analytical approach to solve the problem by taking a convex optimization based estimator as an example. Our proposed approach can be applied to analyze the resilience of a class of commonly used estimators and the estimation performance can be quantified, which is always ignored in the existing works.

The significance of this work is threefold. (i) We propose a framework for studying cyber-physical security problems by formally defining three key ingredients, *i.e.*, (p, m)-sparse attacks, resilience and translation invariance. (ii) The employed approach for analyzing resilience of an estimator is novel and generic for a class of estimators. Though we take an estimator with L_1-penalty as an example, the convex optimization based analysis applies to a large number of estimators. We analyze the sufficient and necessary conditions for resilience and the conclusion that benign sensors must be more than malicious sensors aligns with the results in different research scenarios (Fawzi et al., 2012; Mishra et al., 2015; Chong et al., 2015; Shoukry and Tabuada, 2015; Mo and Murray, 2015). (iii) The tradeoff between MMSE optimality and resilience is well studied. The condition that the resilient estimator gives the MMSE estimate without attacks is given. A nontrivial upper bound on the gap between the resilient estimate and the MMSE estimate is also derived.

Notations: The ith entry of the vector u is denoted as $u[i]$. The L_p norm of the vector u is denote as $\|u\|_p$. If unspecified, $\|u\|$ means the L_2 norm of u by default. $\lfloor v \rfloor$ means the largest integer that is less than the scalar v. For a given set $\mathcal{X}$, $|\mathcal{X}|$ denotes its cardinality.

2. PROBLEM SETUP

2.1 System Model

Assume that m homogenous sensors are measuring the following LTI system:

$$x(k+1) = Ax(k) + w(k). \tag{1}$$

The measurement equation for the ith sensor is given by

$$y_i(k) = Cx(k) + \varepsilon_i(k),\ i = 1, \dots, m, \tag{2}$$

where $x(k) \in \mathbb{R}^n$ is the state, $y_i(k) \in \mathbb{R}^l$ is the measurement collected by the ith sensor, $w(k) \in \mathbb{R}^n$ and $\varepsilon_i(k) \in \mathbb{R}^l$ are the process noise and measurement noise for the ith

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.10.372

sensor, respectively. The noise $w(k)$ and $\varepsilon_i(k)$'s are Gaussian distributed, *i.e.*, $w(k) \sim \mathcal{N}(0, Q)$, $\varepsilon_i(k) \sim \mathcal{N}(0, R)$. The noises are assumed to be independent from each other across different time instants and sensors. Denote the tall measurement matrix $H \triangleq [C^\top, C^\top, \ldots, C^\top]^\top \in \mathbb{R}^{lm\times n}$ and $y(k) \triangleq [y_1(k)^\top, y_2(k)^\top, \ldots, y_m(k)^\top]^\top$ and $\Sigma \triangleq \mathrm{diag}(R, \ldots, R), R > 0$. The initial state $x(0)$ is Gaussian distributed with mean μ_0 and variance P_0, and is independent from all noises. Assume that (A, C) is observable and $(A, \sqrt{Q})$ is controllable. Denote the index set of the sensors as $\mathcal{S} \triangleq \{1, \ldots, m\}$.

Kalman filter is well known as the recursive minimum mean square error (MMSE) estimator:

$$\hat{x}_{KF}(k) = (A - K(k)HA)\hat{x}_{KF}(k-1) + K(k)y(k),$$
$$P^-(k) = AP(k-1)A^\top + Q, P(k) = (\mathrm{I}_n - K(k)H)P^-(k),$$

where the Kalman gain is given by

$$K(k) = P^-(k)H^\top(HP^-(k)H^\top + \Sigma)^{-1}. \tag{3}$$

The state error covariance $P^-(k)$ converges exponentially fast to $\overline{P}$ which is obtained by solving the following discrete algebraic Riccati equation (DARE):

$$X = AXA^\top - AXH^\top(HXH^\top + \Sigma)^{-1}HXA^\top + Q. \tag{4}$$

Therefore, we assume the Kalman filter to be in the steady state, *i.e.*, $P(k) = (\mathrm{I}_n - KH)\overline{P}$ and $K(k) = K$ from (3).

Due to the homogeneousness of the sensors, we know that K can be written into the form of $[G, \ldots, G]$, $G \in \mathbb{R}^{n\times l}$. The Kalman filter can be equivalently rewritten as:

$$\hat{x}_{KF}(k) = \frac{1}{m}\sum_{i\in\mathcal{S}} \tilde{x}_i(k), \tag{5}$$

where

$$\tilde{x}_i(k) = (A - KHA)\tilde{x}_i(k-1) + mGy_i(k), \tag{6}$$

This means the estimation process at the estimator can be decomposed into m sub-processes each of which only involves measurements from one sensor. This decomposition renders distributed estimation possible. To be specific, the sensor can locally compute $\tilde{x}_i(k)$ based on its own measurements and then the information fusion of all local estimates occurs at the remote estimator. It is worth noting that such distributed estimation is more resilient to attacks than the centralized estimation (all sensors transmit raw measurements to a central estimator). Since each local estimate of one sensor encodes all its historical measurements, corruption of one local estimate at some time instant causes little damage to the estimation.

Even if the sensor lacks computational capability and can only transmit raw measurements, each local estimation process can be computed at the central estimator. Therefore, without loss of generality, we assume each sensor computes a local estimate based on (6) and sends it to the estimator.

2.2 Attack Model

The attacker launches an integrity attack to the sensory data in different fashions. For example, it can change the physical environment to mislead the sensors or it hacks the onboard sensor chip or it can manipulate the data packet during the sensor-to-estimator transmission. No matter in which way the attack is launched, we can rewrite the measurement equation into

$$z_i(k) = \tilde{x}_i(k) + a_i(k), \tag{7}$$

where $z_i(k) \in \mathbb{R}^n$ is the "manipulated" local estimate and $a_i(k) \in \mathbb{R}^n$ is the attack vector. In other words, the attacker can change the local estimate of the ith sensor by $a_i(k)$. If the sensor is safe, then $a_i(k) = 0$. Define the local estimation error as $e_i(k) \triangleq \tilde{x}_i(k) - x_i(k)$. Then we have

$$z_i(k) = x(k) + e_i(k) + a_i(k). \tag{8}$$

For concise notations, denote

$$\tilde{x}(k) \triangleq [\tilde{x}_1(k)^\top, \tilde{x}_2(k)^\top, \ldots, \tilde{x}_m(k)^\top]^\top. \tag{9}$$

Similarly we can define $z(k), e(k), a(k)$. For any index set $\mathcal{I} \subseteq \mathcal{S}$, define the complement set to be $\mathcal{I}^c \triangleq \mathcal{S}\backslash\mathcal{I}$. In our attack model, we assume that the attacker can only compromise at most p sensors but can arbitrarily choose $a_i(k)$. The index set of malicious sensors is assumed to be time invariant. Formally, a (p, m)-sparse attack can be defined as

Definition 1. ((p, m)-sparse attack). A vector a is called a (p, m)-sparse attack if there exists an index set $\mathcal{I} \subset \mathcal{S}$, such that (i) $\|a_i(k)\| = 0,\ \forall i \in \mathcal{I}^c$; (ii)$|\mathcal{I}| \leq p$, both hold.

Define the collection of a possible index set of malicious sensors as $\mathbb{C} \triangleq \{\mathcal{I} : \mathcal{I} \subset \mathcal{S}, |\mathcal{I}| = p\}$. The set of all possible (p, m)-sparse attacks is denoted as $\mathcal{A} = \mathcal{A}(k) \triangleq \bigcup_{\mathcal{I}\in\mathbb{C}}\{a(k) : \|a_i(k)\| = 0, i \in \mathcal{I}^c\}, \forall k$.

After introducing the (p, m)-sparse attack, we need to formally define what we mean by resilience.

Definition 2. (Resilience). An estimator $g : \mathbb{R}^{mn} \mapsto \mathbb{R}^n$ which maps the measurements $z(k)$ to a state estimate $\hat{x}(k)$ is said to be resilient to the (p, m)-sparse attack if it satisfies the following condition:

$$\|g(\tilde{x}(k)) - g(\tilde{x}(k) + a(k))\| \leq \mu(\tilde{x}(k)),\ \forall a \in \mathcal{A}, \tag{10}$$

where $\mu : \mathbb{R}^{mn} \mapsto \mathbb{R}$ is a real-valued mapping on $\tilde{x}(k)$.

The resilience implies that the disturbance on the state estimate caused by an arbitrary attack is bounded. A trivial resilient estimator is $g(y) = 0$ which provides a very poor estimate. Therefore, another desirable property for an estimator is translation invariance defined as follows, where $E \triangleq [\mathrm{I}_n, \ldots, \mathrm{I}_n]^\top$:

Definition 3. (Translation invariance). An estimator g is translation invariant if $g(z + Eu) = u + g(z),\ \forall u \in \mathbb{R}^n$.

2.3 A Generic Resilient Estimator

Apparently, the linear estimator (5) cannot give an estimate with bounded error even when only one estimate is arbitrarily manipulated. In other words, there is a conflict between the MMSE optimality and the resilience against attacks. One of the main tasks is to analyze an estimator which may achieve a desirable tradeoff between MMSE optimality and resilience, and investigate the sufficient and necessary conditions to be resilient to (p, m)-sparse attacks. To this end, a general estimator is proposed as follows:

$$\hat{x}(k) \triangleq g(z(k)) = \arg\min_{\hat{x}(k)} \sum_{i\in\mathcal{S}} \varphi_i(z_i(k) - \hat{x}(k)), \tag{11}$$

where $\varphi_i : \mathbb{R}^n \mapsto \mathbb{R}$. We notice that to recover Kalman filter we can choose φ_i to be L_2 norm. The candidate