An Abrupt Change Detection Heuristic with Applications to Cyber Data Attacks on Power Systems$^{\pi}$

Borhan M. Sanandaji,* Eilyan Bitar,$^{\diamond}$ Kameshwar Poolla,* and Tyrone L. Vincent$^{\circ}$

Abstract— **We present an analysis of a heuristic for abrupt change detection of systems with bounded state variations. The proposed analysis is based on the Singular Value Decomposition (SVD) of a history matrix built from system observations. We show that monitoring the largest singular value of the history matrix can be used as a heuristic for detecting abrupt changes in the system outputs. We provide sufficient detectability conditions for the proposed heuristic. As an application, we consider detecting malicious cyber data attacks on power systems and test our proposed heuristic on the IEEE 39-bus testbed.**

I. Introduction

Fault detection and supervisory control are essential to ensure that a dynamical system is operating in normal conditions. These monitoring mechanisms are of higher importance for critical systems such as power systems. Any propagation of faults in a power system may have severe consequences in the electricity generation, transmission, or distribution. To this end, Supervisory Control and Data Acquisition (SCADA) systems are designed for controlling and monitoring different parts of a power grid. Traditionally, within SCADA or other conventional supervisory control and monitoring centers, the state of the system under study is estimated at every sample time. The condition of the system is then tested by monitoring a metric based on the estimated state. An abrupt change in that metric is an indicator of the occurrence of some malfunctioning in the system dynamics.

A. Designated Data Attacks

In power systems, as an example, changes of the system dynamics have been traditionally considered as a result of meter aging and malfunctioning, electrical breakdown, or natural causes such as storm, lightening, etc. However, such changes might be the result of a *designated* cyber data attack to the system. In particular, with the emergence of smart grids and its smart hardware and software components such as smart meters, Phasor Measurement Units (PMUs), intelligent control devices, etc., power systems (and other similar large-scale dynamical systems) are more vulnerable to such malicious data attacks. In fact, it has been recently shown that an attacker can design attacks that do not appear in the detection metrics and can pass conventional detection algorithms. Such attacks, namely called *unobservable attacks*, require a careful compromise of meter readings by the attacker. Altogether, these have motivated a great amount of research to address cyber data attack detection within smart grids.

B. Related Work

Recently, Liu et al. [1] considered scenarios in which an attacker designs attacks carefully such that the conventional bad data detection algorithms are not capable of detecting them. Inspired by their work, many other papers targeted this problem [1]–[9]. An adversary attack has an impact on the real-time and day-ahead electricity markets. Such situations have been studied by [10], [11], among others.

Kosut et al. [5] assume a Bayesian model on the state variables and consider a binary detection problem. In particular, they assume that the state variables have a zero-mean Gaussian distribution. Fawzi et al. [2] impose a linear state-space representation on the power system state evolution and propose a decoder that corrects for the compromised meters. In their plant model, they assume they know the state transition and measurement matrices.

C. Main Contributions

In this paper, we assume no a-priori distribution on the attack vector. We assume that the state variations (under normal conditions) are unknown but bounded within an ℓ_2-norm. The time of attack (modeled as an abrupt change added to the unknown systems dynamics under normal conditions) and its magnitude is unknown to us as well. We present a heuristic for detecting such changes. The proposed heuristic is based on the Singular Value Decomposition (SVD) of a history matrix built form system observations. We show that monitoring the largest singular value of the history matrix is a good heuristic for detecting abrupt changes in the system outputs. In particular, we provide sufficient detectability conditions for the proposed heuristic. While the results of this paper can be applied to any system with a similar linear model with bounded state variations and generic faults, of our particular interest are power systems and unobservable attacks where such fault detection schemes play an important role in maintaining the safety and stability of the system.

*Borhan M. Sanandaji and Kameshwar Poolla are with Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA, email: {sanandaji, poolla}@berkeley.edu.

$^{\diamond}$Eilyan Bitar is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850, USA, email: eyb5@cornell.edu.

$^{\circ}$Tyrone L. Vincent is with the Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO 80401, USA, email: tvincent@mines.edu.

$^{\pi}$This work was supported in part by EPRI and CERTS under sub-award 09-206; PSERC S-52; NSF under Grants EECS-1129061, CPS-1239178, and CNS-1239274; the Republic of Singapore National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore for the SinBerBEST Program; Robert Bosch LLC through its Bosch Energy Research Network funding program.

D. Notation

A column vector is shown as $\boldsymbol{x} \in \mathbb{R}^N$ with boldface letters. An element of a vector is shown as $x(i)$. A matrix is shown with a capital letter as $A \in \mathbb{R}^{M\times N}$. The elements of a matrix is shown as $A(i,j)$. The transpose and pseudo-inverse of A are denoted by A^T and $A^\dagger$, respectively. All variables are real-valued unless mentioned otherwise.

II. Setup

While the proposed analysis can be applied to any system with a linear model, we present our problem formulation based on the DC power flow model of a power system and a transmission network.

A. Measurement Model

Let $\boldsymbol{y}^t \in \mathbb{R}^M$ contain the injected power measurements of $n+1$ buses and line power measurements of m branches of a transmission network at time t, where $M := m+n+1$. Under a DC power flow model assumption over a finite time interval $t = t_i, \dots, t_f$, one can consider a linear relation between the measurements $\boldsymbol{y}^t$ and the power systems state vector $\boldsymbol{x}^t \in \mathbb{R}^N$ as:

$$\boldsymbol{y}^t = H\boldsymbol{x}^t + \boldsymbol{e}^t, \quad t = t_i, \dots, t_f, \tag{1}$$

where $\boldsymbol{x}^t \in \mathbb{R}^N$ is the state of the system at time t containing the relative bus phase angles[1] and $\boldsymbol{e}^t$ contains the measurement noise. The matrix H relates the state variables and the meter measurements and in general, is affected by the grid topology and link impedances. It is the job of the control center to construct H. In this paper, we assume H is given and fixed over time and both the attacker and the control center have access to it. We assume the measurement noise is Gaussian with $\boldsymbol{e}^t \sim \mathcal{N}(0, \Lambda)$. In order to incorporate the attack in the model, one can extend (1) as:

$$\boldsymbol{y}^t = H\boldsymbol{x}^t - \theta^t\boldsymbol{a} + \boldsymbol{e}^t, \quad t = t_i, \dots, t_f, \tag{2}$$

where $\boldsymbol{a}$ is the attack vector and θ^t is an indicator variable.

B. Attack Model

The attacker *abruptly* changes the meter readings at time $t = t_a$ where t_a is the time of attack. The indicator variable θ^t is defined as:

$$\theta^t = \begin{cases} 0, & t < t_a, \\ 1, & t \geq t_a. \end{cases} \tag{3}$$

Remark 1: In a more general setting, one can consider an arbitrary function $s(t, t_{a_i}, t_{a_f})$ as the *signature* of the attack where t_{a_i} is when an attack starts and t_{a_f} is when an attack reaches its final value. Apparently, there exists a trade-off for the attacker between the detectability of the attack (in this case, function s should be smooth and gradually increasing rather than a step function) and the harmfulness of the attack (a step function has a larger harmful impact). □

C. Systems with Bounded State Variations

In this paper, we are interested in linear systems whose state variations are bounded within an ℓ_2-norm ball. Formally, we consider systems with

$$\|\boldsymbol{x}^t - \boldsymbol{x}^{t_0}\|_2 \leq \gamma, \tag{4}$$

for any $t, t_0 \in \{t_i, \dots, t_f\}$ and for some $\gamma > 0$.

III. State Estimation

A. Before Attack

Let's consider the attack model (2)-(3). Note that at any given time t before the attack (i.e., $t_i \leq t < t_a$), we have $\boldsymbol{y}^t = H\boldsymbol{x}^t + \boldsymbol{e}^t$. A state estimate, $\widehat{\boldsymbol{x}}^t$, can be found by minimizing the cost function associated with a Weighted Least Squares (WLS) estimator as:

$$\widehat{\boldsymbol{x}}^t = \underset{\boldsymbol{x}^t}{\operatorname{argmin}}\, J(\boldsymbol{x}^t),$$

where

$$J(\boldsymbol{x}^t) := (\boldsymbol{y}^t - H\boldsymbol{x}^t)^T\Lambda^{-1}(\boldsymbol{y}^t - H\boldsymbol{x}^t).$$

It is trivial to find the minimizer of $J(\boldsymbol{x}^t)$. We have

$$\widehat{\boldsymbol{x}}^t = K\boldsymbol{y}^t, \tag{5}$$

where

$$K := (H^T\Lambda^{-1}H)^{-1}H^T\Lambda^{-1}. \tag{6}$$

Substituting the measurement model $\boldsymbol{y}^t = H\boldsymbol{x}^t + \boldsymbol{e}^t$ in (5),

$$\widehat{\boldsymbol{x}}^t = K\boldsymbol{y}^t = KH\boldsymbol{x}^t + K\boldsymbol{e}^t = \boldsymbol{x}^t + K\boldsymbol{e}^t,$$

where we used the fact that $KH = I_N$.

B. After the Attack

After the attack, we have $\boldsymbol{y}_a^t := \boldsymbol{y}^t + \boldsymbol{a} = H\boldsymbol{x}^t + \boldsymbol{e}^t$. A WLS estimate of the state under the attack can be found as:

$$\begin{aligned} \widehat{\boldsymbol{x}}_a^t &= K\boldsymbol{y}_a^t = K\boldsymbol{y}^t + K\boldsymbol{a} \\ &= \widehat{\boldsymbol{x}}^t + (H^T\Lambda^{-1}H)^{-1}H^T\Lambda^{-1}\boldsymbol{a}. \end{aligned} \tag{7}$$

There has been a recent interest in the so-called *unobservable* malicious data attacks on the power system [1]. From (7), one can see that if there exists a vector $\boldsymbol{c} \in \mathbb{R}^N$ such that

$$\boldsymbol{a} = H\boldsymbol{c},$$

then we have

$$\widehat{\boldsymbol{x}}_a^t = \widehat{\boldsymbol{x}}^t + \boldsymbol{c} \tag{8}$$

and consequently,

$$\boldsymbol{y}_a^t - H\widehat{\boldsymbol{x}}_{\boldsymbol{a}}^t = \boldsymbol{y}^t - H\widehat{\boldsymbol{x}}^t.$$

Thus, at each sample time t, if the residual $\boldsymbol{r}^t := \boldsymbol{y}^t - H\widehat{\boldsymbol{x}}^t$ passes any detection metric (e.g., $\|\boldsymbol{y}^t - H\widehat{\boldsymbol{x}}^t\|_2$), the residual under attack $\boldsymbol{r}_a^t := \boldsymbol{y}_a^t - H\widehat{\boldsymbol{x}}_a^t$ would pass that criteria, and consequently such an attack is unobservable from the view point of the control center. To this end, such detection algorithms are namely referred to as "bad" detection algorithms.

[1] The key state variables in a power grid contain bus voltage magnitudes and angles. However, in a DC power flow model the state variables are usually the bus voltage angles only.

IV. Analysis of an SVD-based Heuristic for Abrupt Change Detection

In this section, we analyze an SVD-based heuristic which can be used for abrupt change detection of systems with bounded state variations. In our proposed approach, we collect a trace of the measurements over a finite time window.

A. History Matrix Δ^t

Given the measurements $\boldsymbol{y}^t$ over a finite horizon of time, at any given time t one can build a *history matrix* that contains the changes of the measurements as:

$$\Delta^t = \begin{bmatrix} (\boldsymbol{y}^t - \boldsymbol{y}^{t-1})^T \\ (\boldsymbol{y}^t - \boldsymbol{y}^{t-2})^T \\ \vdots \\ (\boldsymbol{y}^t - \boldsymbol{y}^{t-w})^T \end{bmatrix}^T \in \mathbb{R}^{M\times w}, \tag{9}$$

where w is the size of the considered time window. Define

$$E^t := \begin{bmatrix} \boldsymbol{e}^{tT} \\ \boldsymbol{e}^{tT} \\ \vdots \\ \boldsymbol{e}^{tT} \end{bmatrix}^T \in \mathbb{R}^{M\times w}, \quad G^t := \begin{bmatrix} -\boldsymbol{e}^{t-1T} \\ -\boldsymbol{e}^{t-2T} \\ \vdots \\ -\boldsymbol{e}^{t-wT} \end{bmatrix}^T \in \mathbb{R}^{M\times w},$$

$$X^t := \begin{bmatrix} (\boldsymbol{x}^t - \boldsymbol{x}^{t-1})^T \\ (\boldsymbol{x}^t - \boldsymbol{x}^{t-2})^T \\ \vdots \\ (\boldsymbol{x}^t - \boldsymbol{x}^{t-w})^T \end{bmatrix}^T \in \mathbb{R}^{N\times w}, \text{ and } \ A := \begin{bmatrix} -\boldsymbol{a}^T \\ -\boldsymbol{a}^T \\ \vdots \\ -\boldsymbol{a}^T \end{bmatrix}^T,$$

for any t. It is trivial to see that Δ^t can be decomposed as

$$\Delta^t = E^t + G^t + HX^t, \quad (\forall t < t_a). \tag{10}$$

At $t = t_a$ (i.e., when the attack happens),

$$\Delta^{t_a} = E^{t_a} + G^{t_a} + HX^{t_a} + A. \tag{11}$$

Similarly, at $t = t_a + 1$, we have

$$\Delta^{t_a+1} = E^{t_a+1} + G^{t_a+1} + HX^{t_a+1} + \begin{bmatrix} \mathbf{0}^T \\ -\boldsymbol{a}^T \\ \vdots \\ -\boldsymbol{a}^T \end{bmatrix}^T \tag{12}$$

and for $t \geq t_a + w$,

$$\Delta^t = E^t + G^t + HX^t, \qquad (t \geq t_a + w). \tag{13}$$

Note that the structure of the history matrix for $t \geq t_a + w$ in (13) is similar to the one for $t < t_a$ given in (10).

B. Singular Value Analysis on Δ^t

Based on the structure of Δ^t and how it changes over time (before and after the attack), one can consider a heuristic for detecting abrupt changes in $\boldsymbol{y}^t$. While the rank of Δ^t (the number of non-zero singular values) does not change before and after the attack, the distribution of the singular values of Δ^t changes (due to the addition of a rank-1 matrix) after the attack. In particular, there exists a large jump in the largest singular value of the history matrix at the time of attack and afterwards.[2] In what follows, we monitor this jump and provide bounds on its magnitude before and after the attack. In particular, note that E^{t_a} and A are rank-1 matrices. Our goal is to exploit such a structure (rank-1 structure of A and E^{t_a}) in evaluating the changes in the first singular value of Δ^t. In order to keep the paper self-contained, we provide all required lemmas and theorems in proving the main theorems.

Let $\sigma_i(\Delta^t)$ denote the ith singular value of Δ^t. The following theorems present probability tail bounds on $\sigma_1(\Delta^t)$ (the largest singular value of Δ^t). The first theorem shows that the largest singular value of Δ^t is bounded from above with exponentially high probability when $t < t_a$.

Theorem 1: Let $\tau > 0$ and $\epsilon > 0$. Consider a linear system described by (2) and with bounded state variations as described by (4). Let M be the number of measurements and w be the window size. Assume $\boldsymbol{a}$ be an unknown attack vector and $\boldsymbol{e^t} \sim \mathcal{N}(0, \nu^2)$. Let Δ^t and G^t be defined as in section IV-A. Then, for $t < t_a$

$$\mathbf{P}\left\{\sigma_1(\Delta^t) \geq \ell\right\} \leq 2\exp\left(-\frac{\tau^2}{2}\right) + \left((1+\epsilon)e^{-\epsilon}\right)^{M/2},$$

where

$$\ell := \nu\sqrt{w}\sqrt{M}(1+\epsilon) + \nu(\sqrt{M} + \sqrt{w} + \tau) + \gamma\sqrt{w}\|H\|.$$

Proof: See Appendix. ■

The second theorem shows that $\sigma_1(\Delta^{t_a})$ is bounded from below with exponentially high probability.

Theorem 2: Let $\tau > 0$ and $\epsilon > 0$. Consider a linear system described by (2) and with bounded state variations as described by (4). Let M be the number of measurements and w be the window length. Assume $\boldsymbol{a}$ be an unknown attack vector and $\boldsymbol{e^t} \sim \mathcal{N}(0, \nu^2)$. Assume $\|\boldsymbol{a}\|_2 \geq \|\boldsymbol{e}^{t_a}\|_2$. Let Δ^{t_a} and G^{t_a} be defined as in section IV-A. Then,

$$\mathbf{P}\left\{\sigma_1(\Delta^{t_a}) \leq u\right\} \leq 2\exp\left(-\frac{\tau^2}{2}\right) + \left((1+\epsilon)e^{-\epsilon}\right)^{M/2},$$

where

$$u =: \sqrt{w}\|\boldsymbol{a}\|_2 - \ell,$$

and ℓ is as defined in Theorem 1.

Proof: See Appendix. ■

Remark 2: Theorems 1 and 2 provide probability tail bounds on $\sigma_1(\Delta^t)$ before and at the time of attack, respectively. The results have a probabilistic notion with exponential bounds. When we say "with high probability" it refers to such exponential behavior. With reasonable choices of τ and ϵ for a given M, the probability term $2\exp\left(-\frac{\tau^2}{2}\right) + \left((1+\epsilon)e^{-\epsilon}\right)^{M/2}$ can be pushed to be very close to 0. □

Based on these results, a detection rule can be considered as follows. For any given ℓ and u such that $\ell < u$, if $\sigma_1(\Delta^t) < \ell$ for all $t_i \leq t < t_a$ and $\sigma_1(\Delta^{t_a}) > u$, then an attack has happened at time t_a. Based on this detection rule, we can derive the detection probability as follows.

[2] While it is still noticeable, this jump starts to decrease at later times after the attack and vanishes at $t = t_a + w$.

Theorem 3: Let $\tau > 0$ and $\epsilon > 0$. Consider a linear system described by (2) and with bounded state variations as described by (4). Let M be the number of measurements and w be the window length. Assume $\boldsymbol{a}$ be an unknown attack vector and $\boldsymbol{e^t} \sim \mathcal{N}(0, \nu^2)$. Let ℓ and u be as defined in Theorems 1 and 2, respectively. Then, an attack $\boldsymbol{a}$ can be detected at t_a with detection probability

$$\mathbf{P}\left\{\text{detection}\right\} \geq 1 - 2\big[2\exp\big(-\frac{\tau^2}{2}\big) + \big((1+\epsilon)e^{-\epsilon}\big)^{M/2}\big].$$

if

$$\|\boldsymbol{a}\|_2 > 2\big[\nu\sqrt{M}\big(1+\epsilon+\frac{1}{\sqrt{w}}+\frac{1}{\sqrt{M}}+\frac{\tau}{\sqrt{M}\sqrt{w}}\big)+\gamma\|H\|\big].$$

Proof: See Appendix. ■

Remark 3: Theorem 3 provides a sufficient condition on $\|\boldsymbol{a}\|_2$ for detectability. The derived bound illustrates how different factors affect the detectability. For example, the larger the noise level (i.e., large ν value), the harder the detection. The number of measurements M and the window size w are also affecting the detectability. Detection of abrupt changes in systems with smaller state variations (i.e., smaller γ) is also easier as can be interpreted from this result. □

V. Case Study - IEEE 39-Bus Testbed

In this section, we examine our proposed heuristic and detection condition on the IEEE 39-bus testbed [12]. Consider a 4-sparse unobservable attack happening at $t_a = 129$. Fig. 1 illustrates how $\sigma_1(\Delta^t)$ evolves over time, where $M = 85$ and $w = 16$. As an alternative to measurements, one could build the history matrix based on the state estimates as given by (5). To this end, we consider two cases. Once we construct the history matrix based on the measurmenets (Fig. 1(a)) and once based on the state estimates (Fig.1(b)). As can be seen, the jump in $\sigma_1(\Delta^{t_a})$ is more distinguishable when the history matrix is built based on the measurements. In all of the simulations of this section, we assume $\gamma = 0$. However, similar results can be achieved for the case with $\gamma \neq 0$.

Also, note that how $\sigma_1(\Delta^t)$ starts to decrease at times after the attack (i.e., $t \geq 129$). In fact, this can be understood by looking at the structure of Δ^t at times after the attack. As shown in (12), the first column of the change matrix is zero at $t = t_a + 1$. This makes the Frobenius norm (and consequently ℓ_2-norm) of the history matrix smaller. This decrease in $\sigma_1(\Delta^t)$ continues on until $t = t_a + w$. The effect of the attack disappears for $t \geq t_a + w$ when the characteristics of Δ^t are similar to the ones at $t < t_a$.

Fig. 2 illustrates how tight are the probability tail bounds of Theorems 1 and 2. As mentioned earlier, a 4-sparse unobservable attack with $\|\boldsymbol{a}\|_2 = 2$ has occurred at $t_a = 129$. With a given number of measurements $M = 85$, we choose $\tau = 4$ and $\epsilon = 0.75$ to make the probability term $2\exp\big(-\frac{\tau^2}{2}\big) + \big((1+\epsilon)e^{-\epsilon}\big)^{M/2} = 6.7 \times 10^{-4}$. We stick to these values in all simulations provided in this section. However, different values of τ and ϵ can be chosen to achieve a desired tail probability. A window size of $w = 8$ and a noise level of $\nu = 0.01$ are considered in Fig. 2(a). We repeat the simulations for 300 realizations at each sample time. A window size of $w = 64$ and a noise level of $\nu = 0.04$ are considered in Fig. 2(b). As can be seen, the gap between the provided bounds and the actual $\sigma_1(\Delta^t)$ magnitude is larger for cases with larger noise levels.

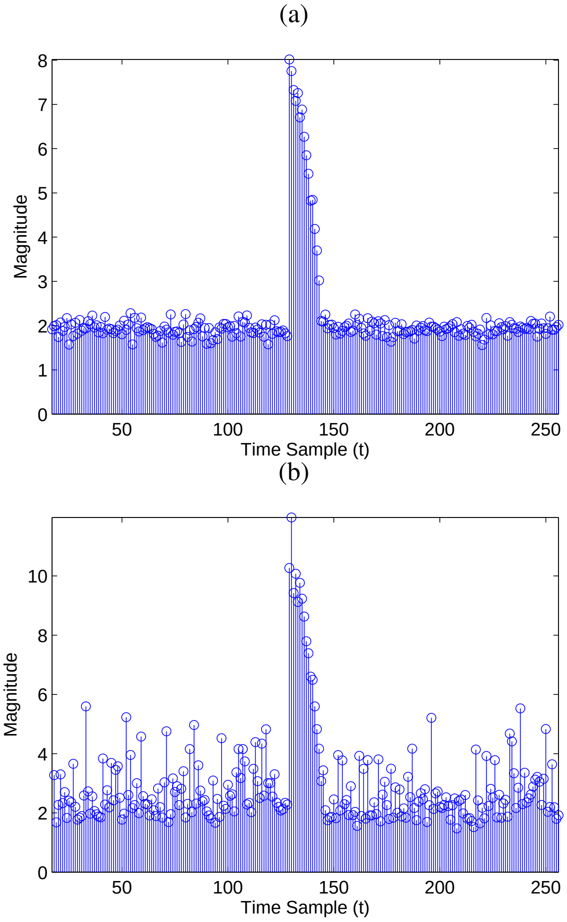

Fig. 1. Illustration of how $\sigma_1(\Delta^t)$ changes over time when an attack happens. In this example, an unobservable attack happens at $t = 129$. A window size of $w = 16$ and a noise level with $\nu = 0.05$ are considered. The history matrix is built based on (a) measurements and (b) state estimates.

Fig. 3 shows how different parameters affect the detectability condition proposed in Theorem 3. We first assume $\|\boldsymbol{a}\| = 2$, $M = 85$, $\tau = 4$, and $\epsilon = 0.75$, and we are interested in the relation between window size w and noise level ν such that the sufficient condition of Theorem 3 is satisfied. Fig. 3(a) shows the result. As can be seen, for larger values of ν, larger window sizes w should be considered such that the attack can be detected with exponentially high probability. In another scenario, we consider the case where $\nu = 0.05$, $M = 85$, $\tau = 4$, and $\epsilon = 0.75$ are fixed and we are interested in finding the relation between window size w and $\|\boldsymbol{a}\|_2$. Fig. 3(b) shows the result. For any $\|\boldsymbol{a}\|_2$, the curve determines the minimum required window size for having detectability. For example, when $\|\boldsymbol{a}\|_2 = 2$ one need to construct the history matrix Δ^t with $w \geq 22$ such that the sufficient detectability condition of Theorem 3 is satisfied. In other words, attacks with larger magnitude may be detected

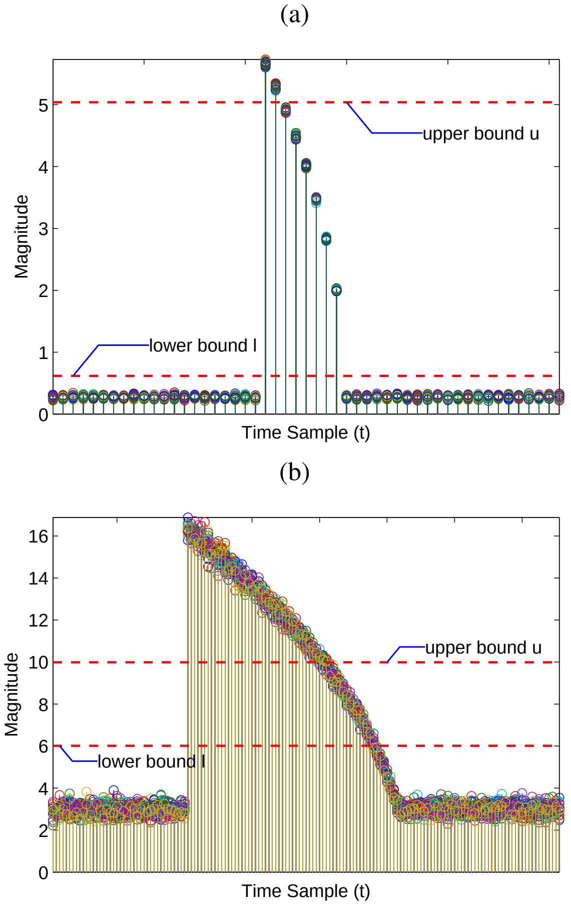

Fig. 2. Illustration of the performance of the provided bounds of Theorems 1 and 2. A 4-sparse unobservable attack with $\|\boldsymbol{a}\|_2 = 2$ has occurred at $t_a = 129$. Plots depict 300 iterations at each sample time. (a) A window size of $w = 8$ and a noise level of $\nu = 0.01$ are considered. (b) A window size of $w = 64$ and a noise level of $\nu = 0.04$ are considered.

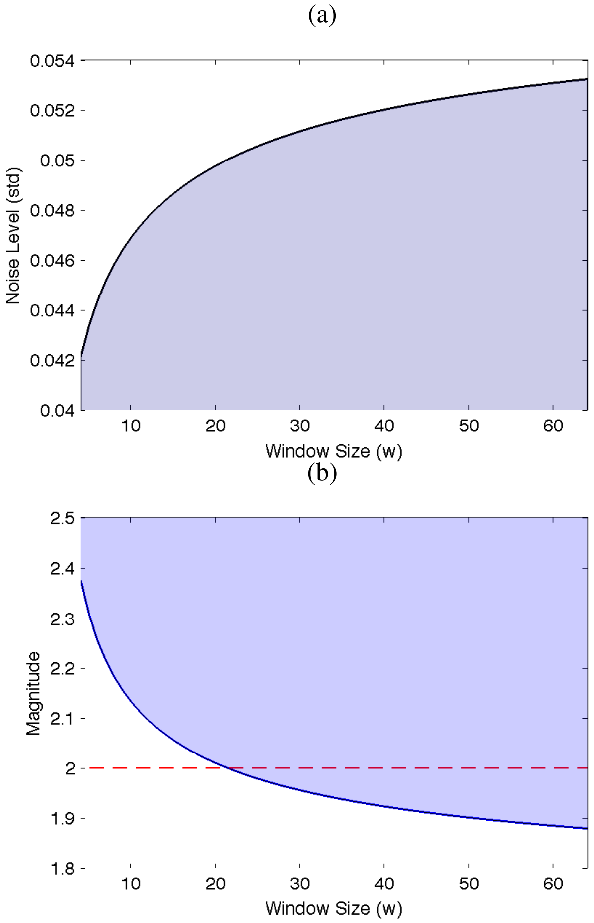

Fig. 3. Illustration of how different parameters affect the detectability condition of Theorem 3. (a) $\|\boldsymbol{a}\|_2 = 2$ is fixed. One needs to increase w as noise level ν increases. (b) $\nu = 0.05$ is fixed. For any $\|\boldsymbol{a}\|_2$, the curve determines the minimum required window size for having detectability. Attacks with larger magnitudes may be detected with smaller window sizes.

with smaller window sizes and similarly, one one needs to increase the window size as $\|\boldsymbol{a}\|_2$ gets smaller.

Appendix

Proof of Theorem 1 Considering the measurement model given in (2), for $t < t_a$ (i.e., before the attack), we have

$$\Delta^t = G^t + E^t + HX^t, \qquad (t < t_a).$$

We are interested in showing that there exists an ℓ such that $\mathbf{P}\left\{\sigma_1(\Delta^t) \geq \ell\right\}$ is very small for $\forall t < t_a$. We have

$$\begin{aligned}\sigma_1(\Delta^t) &= \sigma_1(E^t + G^t + HX^t)\\ &\leq \sigma_1(E^t) + \sigma_1(G^t) + \sigma_1(HX^t)\\ &\leq \sqrt{w}\|\boldsymbol{e}^t\|_2 + \sigma_1(G^t) + \gamma\sqrt{w}\|H\|, \qquad (14)\end{aligned}$$

where we used the assumption that $\|\boldsymbol{x}^t - \boldsymbol{x}^{t-t_0}\| \leq \gamma$, for all t and $t_0 \in \{t_i, \ldots, t_f\}$. Given $\nu, \tau, \epsilon, M, w$, and any $t < t_a$, let's define event A as

$$\mathcal{E}(A) := \left\{\sigma_1(G^t) < \nu(\sqrt{M} + \sqrt{w} + \tau)\right\}$$

and event B as

$$\mathcal{E}(B) := \left\{\|\boldsymbol{e}^t\|_2 < \nu\sqrt{M}(1+\epsilon)\right\}.$$

It is trivial to see that if events $\mathcal{E}(A)$ and $\mathcal{E}(B)$ happen, then event C defined as

$$\mathcal{E}(C) := \left\{\sqrt{w}\|\boldsymbol{e}^t\|_2 + \sigma_1(G^t) + \gamma\sqrt{w}\|H\| < \ell\right\}$$

happens where

$$\ell := \nu\sqrt{w}\sqrt{M}(1+\epsilon) + \nu(\sqrt{M} + \sqrt{w} + \tau) + \gamma\sqrt{w}\|H\|.$$

Using results from Concentration of Measure (CoM) phenomenon of random processes [13]–[15], we first show that for any t, random variables $\sigma_1(G^t)$ and $\|\boldsymbol{e}^t\|_2$ are highly concentrated around their expected value. The following lemmas provide such CoM bounds.

Lemma 1: ([15], [16, Lemma 2]) Let $\boldsymbol{e}$ be a vector in $\mathbb{R}^M$ whose entries are independent Gaussian random variables with zero mean and ν^2 variance. Then for every $\epsilon \geq 0$,

$$\mathbf{P}\left\{\|\boldsymbol{e}\|_2 \geq \nu\sqrt{M}(1+\epsilon)\right\} \leq \left((1+\epsilon)e^{-\epsilon}\right)^{M/2}.$$

Lemma 2: ([17]) Let G be an $M \times w$ matrix whose entries are independent Gaussian random variables with zero mean and ν^2 variance. Then for every $\tau \geq 0$,

$$\mathbf{P}\left\{\sigma_1(G) \geq \nu(\sqrt{M}+\sqrt{w}+\tau)\right\} \leq 2\exp\left(-\frac{\tau^2}{2}\right).$$

Using Lemma 1 and Lemma 2, and noting that $\mathbf{P}\left\{\mathcal{E}(C)^c\right\} \leq \mathbf{P}\left\{\mathcal{E}(A)^c\right\} + \mathbf{P}\left\{\mathcal{E}(B)^c\right\}$, we have

$$\begin{aligned}\mathbf{P}\left\{\sigma_1(\Delta^t) \geq \ell\right\} &\leq \mathbf{P}\left\{\sqrt{w}\|\boldsymbol{e}^t\|_2 + \sigma_1(G^t) \geq \ell\right\} \\ &\leq 2\exp\left(-\frac{\tau^2}{2}\right) + \left((1+\epsilon)e^{-\epsilon}\right)^{M/2},\end{aligned}$$

where we used (14) in showing the first inequality. ■

Proof of Theorem 2 First we need a lower bound on $\sigma_1(\Delta^{t_a})$. Modifying [18, Theorem 6], we can derive a lower bound on the first singular value of Δ^{t_a} as

$$\begin{aligned}\sigma_1(\Delta^{t_a}) &= \sigma_1(E^{t_a} + G^{t_a} + HX^{t_a} + A) \\ &\geq \left|\sigma_i(E^{t_a} + A) - \sigma_i(G^{t_a} + HX^{t_a})\right| \qquad (15)\end{aligned}$$

for all $1 \leq i \leq \min\{M, w\}$. In particular, for $i = 1$

$$\sigma_1(\Delta^{t_a}) \geq \left|\sigma_1(E^{t_a} + A) - \sigma_1(G^{t_a} + HX^{t_a})\right|. \qquad (16)$$

Assuming $\sigma_1(E^{t_a} + A) > \sigma_1(G^{t_a} + HX^{t_a})$, we have

$$\begin{aligned}\sigma_1(\Delta^{t_a}) &\geq \sqrt{w}\|\boldsymbol{e}^{t_a} + \boldsymbol{a}\|_2 - \sigma_1(G^{t_a} + HX^{t_a}) \\ &\geq \sqrt{w}\|\boldsymbol{e}^{t_a} + \boldsymbol{a}\|_2 - \sigma_1(G^{t_a}) - \gamma\sqrt{w}\|H\|,\end{aligned}$$

where we used the fact that $\sigma_1(E^{t_a} + A) = \sqrt{w}\|\boldsymbol{e}^{t_a} + \boldsymbol{a}\|_2$ and $\sigma_1(G^{t_a} + HX^{t_a}) \leq \sigma_1(G^{t_a}) + \sigma_1(HX^{t_a})$. Assuming $\|\boldsymbol{a}\|_2 \geq \|\boldsymbol{e}^{t_a}\|_2$ and using the reverse triangle inequality,

$$\|\boldsymbol{e}^{t_a} + \boldsymbol{a}\|_2 \geq \left|\|\boldsymbol{a}\|_2 - \|\boldsymbol{e}^{t_a}\|_2\right| = \|\boldsymbol{a}\|_2 - \|\boldsymbol{e}^{t_a}\|_2. \qquad (17)$$

Therefore, from (16)

$$\begin{aligned}\sigma_1(\Delta^{t_a}) &\geq \sqrt{w}\|\boldsymbol{e}^{t_a} + \boldsymbol{a}\|_2 - \sigma_1(G^{t_a}) - \gamma\sqrt{w}\|H\| \qquad (18) \\ &\geq \sqrt{w}\|\boldsymbol{a}\|_2 - \sqrt{w}\|\boldsymbol{e}^{t_a}\|_2 - \sigma_1(G^{t_a}) - \gamma\sqrt{w}\|H\|.\end{aligned}$$

Given $\nu, \tau, \epsilon, M, w$, and t_a, let's define event A as

$$\mathcal{E}(A) := \left\{\sigma_1(G^{t_a}) < \nu(\sqrt{M} + \sqrt{w} + \tau)\right\}$$

and event B as $\mathcal{E}(B) := \left\{\|\boldsymbol{e}^{t_a}\|_2 < \nu\sqrt{M}(1+\epsilon)\right\}$. It is trivial to see that if events $\mathcal{E}(A)$ and $\mathcal{E}(B)$ happen, then event C defined as

$$\mathcal{E}(C) := \left\{\sqrt{w}\|\boldsymbol{a}\|_2 - \sqrt{w}\|\boldsymbol{e}^{t_a}\|_2 - \sigma_1(G^{t_a}) - \gamma\sqrt{w}\|H\| > u\right\}$$

happens where $u := \sqrt{w}\|\boldsymbol{a}\|_2 - \ell$ and

$$\ell = \nu\sqrt{w}\sqrt{M}(1+\epsilon) + \nu(\sqrt{M} + \sqrt{w} + \tau) + \gamma\sqrt{w}\|H\|.$$

Using Lemma 1 and Lemma 2, and noting that $\mathbf{P}\left\{\mathcal{E}(C)^c\right\} \leq \mathbf{P}\left\{\mathcal{E}(A)^c\right\} + \mathbf{P}\left\{\mathcal{E}(B)^c\right\}$, we have

$$\begin{aligned}&\mathbf{P}\left\{\sigma_1(\Delta^{t_a}) \leq u\right\} \leq \\ &\qquad \mathbf{P}\left\{\sqrt{w}\|\boldsymbol{a}\|_2 - \sqrt{w}\|\boldsymbol{e}^{t_a}\|_2 - \sigma_1(G^{t_a}) \leq u\right\} \\ &\qquad\qquad \leq 2\exp\left(-\frac{\tau^2}{2}\right) + \left((1+\epsilon)e^{-\epsilon}\right)^{M/2},\end{aligned}$$

where we used (18) in showing the first inequality. ■

Proof of Theorem 3 For any ℓ and u such that $\ell < u$,

$$\mathbf{P}\left\{\text{detection}\right\} = \mathbf{P}\left\{\sigma_1(\Delta^t) < \ell \text{ and } \sigma_1(\Delta^{t_a}) > u\right\}.$$

Using similar techniques as used in the proof of Theorems 1 and 2, we have

$$\begin{aligned}\mathbf{P}\left\{\text{not detection}\right\} &\leq \mathbf{P}\left\{\sigma_1(\Delta^t) \geq \ell\right\} + \mathbf{P}\left\{\sigma_1(\Delta^{t_a}) \leq u\right\} \\ &\leq 2\left[2\exp\left(-\frac{\tau^2}{2}\right) + \left((1+\epsilon)e^{-\epsilon}\right)^{M/2}\right],\end{aligned}$$

where ℓ and u are chosen as given in Theorems 1 and 2, respectively. If

$$\|\boldsymbol{a}\|_2 > 2\left[\nu\sqrt{M}\left(1+\epsilon+\frac{1}{\sqrt{w}}+\frac{1}{\sqrt{M}}+\frac{\tau}{\sqrt{M}\sqrt{w}}\right)+\gamma\|H\|_2\right],$$

then $\ell < u$ and this completes the proof. ■

References

[1] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[2] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure state-estimation for dynamical systems under active adversaries," *Proceedings of the* 49*-th Annual Allerton Conference on Communication, Control, and Computing*, pp. 337–344, 2011.

[3] F. Pasqualetti, F. Dorfler, and F. Bullo, "Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design," *Proceedings of the* 50*-th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pp. 2195–2201, 2011.

[4] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: characterizations and countermeasures π," *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, pp. 232–237, 2011.

[5] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "On malicious data attacks on power system state estimation," *Proc. of the* 45*-th International Universities Power Engineering Conference*, pp. 1–6, 2010.

[6] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 326–333, 2011.

[7] D. Gorinevsky, S. Boyd, and S. Poll, "Estimation of faults in dc electrical power system," *Proceedings of American Control Conference*, pp. 4334–4339, 2009.

[8] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on dc state estimation," *Preprints of the First Workshop on Secure Control Systems*, 2010.

[9] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," *Preprints of the First Workshop on Secure Control Systems*, 2010.

[10] L. Jia, R. J. Thomas, and L. Tong, "Malicious data attack on real-time electricity market," *Proc. of the* 2011 *IEEE International Conf. on Acoustics, Speech and Signal Processing*, pp. 5952–5955, 2011.

[11] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," *Proceedings of the* 2010 *IEEE International Conference on Smart Grid Communications*, pp. 226–231, 2010.

[12] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[13] M. Ledoux, *The concentration of measure phenomenon.* Amer Mathematical Society, 2001.

[14] G. Lugosi, "Concentration-of-measure inequalities," *Lecture Notes*, 2004.

[15] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, "Concentration of measure inequalities for Toeplitz matrices with applications," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 109–117, 2013.

[16] B. M. Sanandaji, T. L. Vincent, K. Poolla, and M. B. Wakin, "A tutorial on recovery conditions for compressive system identification of sparse channels," *Proceedings of the* 51*-th IEEE Conference on Decision and Control (CDC)*, pp. 6277–6283, 2012.

[17] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *Arxiv preprint arxiv:1011.3027*, 2011. [Online]. Available: http://arxiv.org/abs/1011.3027

[18] J. K. Merikoski and R. Kumar, "Inequalities for spreads of matrix sums and products," *Applied Mathematics E-Notes*, vol. 4, pp. 150–159, 2004.