Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-25 (2018) 178–183

Statistical Approach to Detection of Attacks for Stochastic Cyber-Physical Systems

Damián Marelli[1,2], **Tianju Sui**[3], **Minyue Fu**[4,1] ***Fellow IEEE***

[1] *School of Automation, Guandong University of Technology, China.*
[2] *French-Argentinean International Center for Information and Systems Sciences, National Scientific and Technical Research Council, Argentina.*
[3] *Department of Control Science and Engineering, Dalian University of Technology, China.*
[4] *School of Electrical Engineering and Computer Science, University of Newcastle, Australia.*

Abstract: We study the problem of detecting an attack on a stochastic cyber-physical system. We start by proposing a detection criterion based on checking the statistics of the Kalman prediction error. To show the importance of the proposed criterion, we study certain attacking schemes which can be effectively detected by this criterion, but not by checking simpler conditions based on commonly used statistics. We then use the proposed criterion to derive a detection algorithm. This algorithm is based on a statistical confidence test, giving a confidence level for detecting an attack. We present simulation results to illustrate the performance of this algorithm.

© 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: System security, linear systems, cyber-physical systems, integrity attacks, attack detection, statistical hypothesis testing.

1. INTRODUCTION

A cyber-physical systems (CPS) is a physical system which is monitored or controlled via a communication channel. It finds a wide range of applications such as traffic signal systems Puthran et al. [2015], health care Cardenas et al. [2008], energy manufacturing Chen [2010], power system DeMarco et al. [1996], Sridhar et al. [2012], Mohsenian-Rad and Leon-Garcia [2011], Dan and Sandberg [2010], Pasqualetti et al. [2011], the water industry Amin et al. [2010], Eliades and Polycarpou [2010], etc. A CPS is prone to receiving attacks in the form of signals injected in the communication link Cardenas et al. [2008]. These attacks are known to have caused a number of serious accidents around the world Farwell and Rohozinski [2011], Richards [2008], Conti [2010], Slay and Miller [2007], Kuvshinkova [2003]. This has motivated the development of methods for detecting attacks.

In principle, an attack may be regarded as a system fault. This permits using methods for fault tolerant control, such as robust statistics Huber [2011], robust control Zhou et al. [1996] and failure detection and identification Willsky [1976]. However, the essential difference between an attack and a fault is that the design of the former aims at making it difficult for detection. For example, Liu *et al* studied how to inject a stealthy input into the measurement without being detected by the classical failure detector Liu et al. [2011]. Hence, the analysis and development of methods for CPS attack detection need to take special care of this difference.

Early works on CPS attack detection rely on certain prior knowledge of the attacker's model. Among these methods, we find the work in Amin et al. [2009], which deals with a kind of attack called denial of service. The works Liu et al. [2011] and Mo et al. [2010] concentrate on false data injection attacks against state estimation. The authors of Teixeira et al. [2010] introduced stealthy deception attacks, which consist in manipulating the measurements to be processed by a power system state estimator in such a manner that the resulting systematic errors introduced by the adversary are either undetected or only partially detected by a bad data detection method. In Mo and Sinopoli [2009], the effect of replay attacks is studied. This kind of attack consist in recording a history of system measurements, and send them to the estimator when the attacker is controlling the actuator. Smith investigated the behavior of control system under covert attacks Smith [2011], where a malicious agent can access the signals and information within the control loop and use these to disrupt or compromise the controlled plant.

It is often unrealistic to assume in practice that the defender has some knowledge of the attacker's model. To avoid this limitation, recent works study the CPS attack problem without an attacking model assumption. In this line, Pasqualetti *et al* studied the problem of which kind of attacks can be detected Pasqualetti et al. [2013]. They also studied in Pasqualetti et al. [2013, 2012] the design of centralized and distributed attack detection methods. However, a limitation of this approach is that the system model is noiseless, and it fails to work for system models

2405-8963 © 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2018.11.101

containing process noises or measurement noises. The study of systems involving random noises is much more challenging, since these systems present more ambiguities where attacks can be hidden.

Concerning the detection of attacks in systems with noise, Mo and Sinopoli Mo and Sinopoli [2016] analyzed the estimation error introduced by an attack which is not detected by a χ^2 failure detector. They also studied in Mo and Sinopoli [2015] the attacks on scalar systems with multiple sensors. They showed that, if more than half of the sensors are under attack, the optimal worst-case estimator should ignore all measurements and base its estimation only on prior knowledge. Also, they gave the explicit form of the optimal estimator when less than half of the sensors are under attack.

In this work we move a step forward in the research line described above. As in Mo and Sinopoli [2016], we also study systems with noise. We start by introducing a detection criterion, which we show to be equivalent to testing that the statistics of the output signal equals those corresponding to the healthy (i.e., non-attacked) system. To appreciate the importance of this criterion, we give two examples of attacks, which cannot be detected by checking commonly used statistics, but are instead detected by our criterion. We also use this criterion to derive a practically feasible detection algorithm, and provide simulation results to illustrate the superiority of our algorithm for detecting attacks that cannot be detected by other simpler methods.

The rest of this paper is organized as follows. In Section 2 we describe the attack detection problem. In Section 3 we state the proposed attack detection criterion. In Section 4, we discuss attack examples which cannot be detected by checking other simpler conditions. In Section 5 we use our condition to derive the detection algorithm. In Section 6 we use simulations to illustrate the superiority of our algorithm for detecting attacks that cannot be detected by other simpler methods. Finally, concluding remarks are stated in Section 7.

2. PROBLEM DESCRIPTION

Notation 1. For a vector x we use $[x]_i$ to denote its i-th entry and for a matrix X we use $[X]_{i,j}$ to denote its (i,j)-th entry. For a vector or matrix X, we use $X^\top$ to denote its transpose. For vectors x and y, $x \prec y$ ($x \preceq y$) means that $[x]_i < [y]_i$ ($[x]_i \leq [y]_i$), for all i, and $z = x \wedge y$ denotes the vector with entries $[z]_i = \min(x_i, y_i)$. We use I to denote the identity matrix. We also use $\phi_{\mu,\Sigma}$ and $\Phi_{\mu,\Sigma}$ to denote the probability density function (PDF) and cumulative distribution function (CDF), respectively, of the normal distribution with mean μ and covariance matrix Σ.

We have the following system in state-space form

$$x_{t+1} = Ax_t + w_t, \tag{1}$$

$$y_t = Cx_t + v_t. \tag{2}$$

The measurement y_t is a D-dimensional random vector, $x_1 \sim \mathcal{N}(0,P)$, $w_t \sim \mathcal{N}(0,Q)$, $v_t \sim \mathcal{N}(0,R)$, and $\{x_1, w_t, v_t : t \in \mathbb{N}\}$ are jointly independent. We assume that

$$P = APA^\top + Q,$$

so that the system is in steady state.

We assume that there is an attacker, which interferes the measurement signal y_t, replaces it by an attacking signal z_t, and sends z_t instead of y_t to the receiver. In order to treat the problem in its full generality, we assume that z_t is generated by an arbitrary (possibly non-linear and non-stationary) measurable function of the whole history of y_s up to time t, i.e.,

$$z_t = h_t\left(y_s : t \geq s \in \mathbb{N}\right).$$

The attack detection problem consists in assessing whether or not $z = y$.

Definition 2. We say that the statistics are *nominal* if they equal those which occur when $z = y$. The probability law and expected value taken with respect to these statistics are denoted by $p_\star(\cdot)$ and $\mathcal{E}_\star\{\cdot\}$, respectively.

3. PROPOSED ATTACK DETECTION CRITERION

Suppose that there is no attack. We then have

$$z_k = C\sum_{n=0}^{\infty} A^n w_{k-1-n} + v_k.$$

Hence, z_k is a normal (vector) random process with $\mathcal{E}_\star\{z_k\} = 0$ and, for $k \geq l$,

$$\begin{aligned}\mathcal{E}_\star\left\{z_k z_l^\top\right\} &= C\sum_{n,m=0}^{\infty} A^n \mathcal{E}_\star\left\{w_{k-1-n}w_{l-1-m}^\top\right\}\left(A^m\right)^\top C^\top \\ &\quad +\mathcal{E}\left\{v_k v_l^\top\right\} \\ &= CA^{k-l}\left[\sum_{m=0}^{\infty} A^m \Sigma_w \left(A^m\right)^\top\right] C^\top + \delta_{k-l}\Sigma_v.\end{aligned}$$

A possible criterion for detecting the presence of an attack consists in verifying that the statistics of z_k differ form the nominal ones. If we run a Kalman filter, once it reaches steady state, we obtain

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + K\left(z_t - C\hat{x}_{t|t-1}\right), \tag{3}$$

$$K = A\Psi C^\top\left(C\Psi C^\top + R\right)^{-1},$$

where Ψ is the solution of

$$\Psi = A\Psi A^\top - A\Psi C^\top\left(C\Psi C^\top + R\right)^{-1} C\Psi A^\top + Q.$$

Let $\hat{z}_{t|t-1} = C\hat{x}_{t|t-1}$, $\tilde{z}_t = z_t - \hat{z}_{t|t-1}$ and $\Gamma = C\Psi C^\top + R$. Let also

$$\check{z}_t = \Gamma^{-1/2}\tilde{z}_t. \tag{4}$$

We then have

$\check{z}_k$ is a normal vector random process with

$$\mathcal{E}_\star\left\{\check{z}_k\check{z}_l^\top\right\} = \begin{cases} I, & k = l, \\ 0, & k \neq l, \end{cases} \tag{5}$$

The next lemma asserts that the aforementioned detection criterion is equivalent to (5). In view of this, our detection criterion consists in verifying that (5) holds.

Lemma 3. Condition (5) holds if and only if the statistics of Z_k are nominal.

Proof. The 'if' part is obvious. For the 'only if' part, suppose that (5) is satisfied. Now,

$$\begin{aligned}\hat{x}_{k+1} &= A\hat{x}_k + AK\left(z_k - C\hat{x}_{k|k-1}\right) \\ &= A\hat{x}_k + AK\Gamma^{1/2}\check{z}_k,\end{aligned}$$

with $\hat{x}_1 = 0$. This means that the statistics of $\left\{\hat{X}_k, \check{Z}_k\right\}$ are nominal. The result then follows since

$$z_k = \Gamma^{1/2}\check{z}_k + C\hat{x}_k.$$

Notice that assessing (5) is not equivalent to assessing

$$\check{z}_k \sim \mathcal{N}(0,I) \text{ and } \mathcal{E}_\star\left\{\check{z}_k\check{z}_l^\top\right\} = 0, \forall k \neq l. \tag{6}$$

The condition (5) is stronger in the sense that, for any set of indexes $k_1, \cdots, k_M$, the vector $[\check{z}_{k_1}, \cdots, \check{z}_{k_M}]$ is required to be (jointly) normal. In particular, (5) implies that $\check{z}_k$ and $\check{z}_l$ are statistically independent, when $k \neq l$, while (6) does not. In the next section we provide two examples of situations in which an attacker is detected by checking (5), but not by checking (6).

4. ATTACK EXAMPLES WHOSE DETECTION REQUIRES (5)

Checking condition (5) essentially means checking that a block of samples of $\check{z}_t$ has joint standard normal distribution. As it is known, from a theoretical point of view, checking this is a stronger requirement than doing some other more practical checks, e.g., for pairwise independence or uncorrelation. However, the question arise as to whether, for the purposes of detecting an attack, it is really necessary to carry out a full distribution check, or if instead, a simpler test would be enough. In this section we provide two examples showing how a stochastic detectable attack can pass undetected if a test checking only for uncorrelation or pairwise independence is used. This supports our claim that checking condition (5) is indeed needed.

4.1 Checking for uncorrelation

By combining a normality test for i.i.d. samples Thode [2002], together with a test for uncorrelation [Wasserman, 2004, S 14.2], we can verify if condition (6) holds. Condition (5) is stronger that (6), in the sense that the former requires that $\check{z}_t$ and $\check{z}_s$ are statistically independent, rather than uncorrelated, when $t \neq s$. In this section we describe an attack example which can be detected by a method verifying (5), but not by one verifying (6).

Suppose that we feed the output y_t to the Kalman filter (3). Let $\check{y}_t \sim \mathcal{N}(0,I)$ denote the resulting normalized prediction error, obtained as in (4). Let $\tau \in \mathbb{N}$, $0 < \alpha < 1$ and γ_t be an i.i.d. sequence of random variables with Bernoulli distribution $\mathcal{B}(0.5)$. Let $r_0 \sim \mathcal{N}(0,I)$ and

$$\begin{aligned} r_t &= \alpha r_{t-\tau} + \sqrt{1-\alpha^2}\check{y}_t, \\ \check{z}_t &= \gamma_t r_t. \end{aligned} \tag{7}$$

Since $\check{y}_t \sim \mathcal{N}(0,I)$ is i.i.d., it is straightforward to see that $r_t \sim \mathcal{N}(0,I)$. Hence

$$\check{z}_t \sim \mathcal{N}(0,I).$$

Also, if $t \neq s$,

$$\mathcal{E}\left\{\check{z}_t\check{z}_s^\top\right\} = \left(1-\alpha^2\right)\mathcal{E}\left\{\gamma_t\right\}\mathcal{E}\left\{\gamma_s\right\}\mathcal{E}\left\{r_t r_s\right\} = 0.$$

Hence, the process $\check{z}$ satisfies (6). However, since

$$\begin{aligned} \check{z}_t &= \gamma_t\left(\alpha r_{t-\tau} + \sqrt{1-\alpha^2}\check{y}_t\right), \\ \check{z}_{t-\tau} &= \gamma_{t-\tau} r_{t-\tau}, \end{aligned}$$

the vector $\left[\check{z}_t^\top, \check{z}_{t-\tau}^\top\right]^\top$ is clearly not Gaussian. Hence, $\check{z}$ does not satisfy (5). Since the attacker knows y_s, for all $s \leq t$, it can always build the attacking signal so that the normalized prediction error $\check{z}$ at the receiver equals the one described above. Such an attack can be detected by (5) but not by (6).

4.2 Checking for pairwise independence

A combination of a normality test Thode [2002] with a test for pairwise independence [Wasserman, 2004, S 15], Bakirov et al. [2006] permits checking the following condition

$$\left[\check{z}_t^\top \check{z}_s^\top\right]^\top \sim \mathcal{N}(0,I), \forall t \neq s. \tag{8}$$

As it is known, assessing that (5) holds is not equivalent to assessing (8). This is because pairwise independence does not imply joint independence in general. We describe below an attacking scheme which would be detected by (5), but not by (8).

Let the measurement dimension $D = 1$. As before, we feed the output y to the Kalman filter (3), and let $\check{y}_t \sim \mathcal{N}(0,1)$ denote the normalized prediction error. Draw $[\check{z}_0, \check{z}_{-1}]$ from the distribution $\mathcal{N}(0,I)$. Then, for $t \in \mathbb{N}$, we compute

$$\check{z}_t = \begin{cases} \check{y}_t, & t \text{ even}, \\ \operatorname{sign}\left(\check{z}_{t-1}\check{z}_{t-2}\right)\left|\check{y}_t\right|, & t \text{ odd}, \end{cases} \tag{9}$$

We first analyze pairwise independence. If t is even $\check{z}_t$ is obviously independent of $\check{z}_s$, for all $s \neq t$. So we assume that t is odd. Suppose that at time t, the vector $[\check{z}_{t-1}, \check{z}_{t-2}]$ has distribution $\mathcal{N}(0,I)$. We have

$$p\left(\check{z}_t, \check{z}_{t-1}\right) = p\left(\check{z}_t|\check{z}_{t-1}\right)p\left(\check{z}_{t-1}\right).$$

Now

$$\begin{aligned} p\left(\check{z}_t = \beta|\check{z}_{t-1}\right) &= \begin{cases} \dfrac{1}{2}p\left(\left|\check{y}_t\right| = \beta\right), & \beta > 0 \\ \dfrac{1}{2}p\left(-\left|\check{y}_t\right| = \beta\right), & \beta \leq 0 \end{cases} \\ &= \phi_{0,1}(\beta). \end{aligned}$$

Also, $\check{z}_{t-1} \sim \mathcal{N}(0,1)$. Hence $[\check{z}_t, \check{z}_{t-1}] \sim \mathcal{N}(0,I)$. By symmetry, we also have that $[\check{z}_t, \check{z}_{t-2}] \sim \mathcal{N}(0,I)$. Since clearly $[\check{z}_t, \check{z}_s] \sim \mathcal{N}(0,I)$ for any even s, it remains to be shown that $[\check{z}_t, \check{z}_{2s+1}] \sim \mathcal{N}(0,I)$, for all s. This follows immediately from (9), since $\check{z}_{t-2}$ is independent of $\check{z}_{2s+1}$. Then, by induction on t, we have that (9) holds for all t and s.

Now, clearly, if t is even, $[\check{z}_t, \check{z}_{t-1}, \check{z}_{t-2}] \sim \mathcal{N}(0,I)$, However, for any odd t,

$$\check{z}_t\check{z}_{t-1}\check{z}_{t-2} \geq 0.$$

Then,

$$[\check{z}_t, \check{z}_{t-1}, \check{z}_{t-2}] \nsim \mathcal{N}(0,I). \tag{10}$$

Hence, in view of (10), $\check{z}_t$ does not satisfy (5). We can then draw the same conclusions as those in Section 4.1.

5. ATTACK DETECTION ALGORITHM

In this section we derive a numerically tractable algorithm for checking condition (5). In doing so, we make use of certain statistical definitions which we introduce below.

Definition 4. Let z_t be a random process whose probability distribution $p_\star$ is determined by certain hypothesis (called null hypothesis) made on its statistical model. For a given $T \in \mathbb{N}$, a test statistic is a positive random variable

$$v_T = g_T \left(z_t : T \geq t \in \mathbb{N}\right),$$

which is built using all the available samples of z_t up to time T. Let $H_\star$ denote the CDF of v_T under the null hypothesis. We define the confidence ψ_T of rejecting the null hypothesis at time T by

$$\psi_T = H_\star \left(v_T\right).$$

Suppose that a null hypothesis rejection alarm is triggered whenever $\psi_T \geq \alpha$, for some threshold $\alpha > 0$. The false alarm rate is defined by

$$p_\star \left(\psi_T \geq \alpha\right) = 1 - \alpha.$$

We now derive the proposed algorithm. For a given fixed time horizon $L \in \mathbb{N}$, we define a set of sample points on $\mathbb{R}^{LD}$ at which we will test the joint distribution of $\zeta_t^{(L)} = \left[\check{z}_t^\top, \cdots, \check{z}_{t+L-1}^\top\right]^\top$. Let $\rho_i = \left[\rho_{i,0}^\top, \cdots, \rho_{i,L-1}^\top\right]^\top$, $i = 1, \cdots, I$, be these points, with $\rho_{i,l} \in \mathbb{R}^D$, $l = 0, \cdots, L-1$. Let

$$F_\star^{(L)}(\rho) = p_\star \left(\zeta_1^{(L)} \preceq \rho\right) = \Phi_{0,\mathbf{I}}\left(\rho\right), \tag{11}$$

be the nominal CDF of $\zeta_1^{(L)}$. For each $T \in \mathbb{N}$, let

$$\hat{F}_T^{(L)}\left(\rho\right) = \frac{1}{T} \sum_{t=1}^{T} \chi_{\left\{\zeta_t^{(L)} \preceq \rho\right\}}(\rho), \tag{12}$$

be a sample approximation of the nominal CDF. We then sample $\hat{F}_T^{(L)}$ and $F^{(L)}$ at the points ρ_i, $i = 1, \cdots, I$, forming the vectors $\hat{u}_T \in \mathbb{R}^I$ and $\bar{u} \in \mathbb{R}^I$, respectively, defined by

$$\left[\hat{u}_T\right]_i = \hat{F}_T^{(L)}\left(\rho_i\right), \tag{13}$$

$$\left[\bar{u}\right]_i = F^{(L)}\left(\rho_i\right). \tag{14}$$

We next define the following weighted difference between the above vectors

$$v_T = T\left(\hat{u}_T - \bar{u}\right)^\top \Sigma^{-1}\left(\hat{u}_T - \bar{u}\right), \tag{15}$$

where $\Sigma = \sum_{\tau \in \mathbb{Z}} \Sigma(\tau)$, with

$$\Sigma(\tau) = \mathcal{E}\left\{\left(u_\tau - \bar{u}\right)\left(u_0 - \bar{u}\right)^\top\right\}.$$

We then have the following result, whose proof is omitted in this conference version.

Theorem 5. Using the test statistic v_T, the confidence of asserting that there is an attack, is

$$\psi_T = p_\star\left(\xi \leq v_T\right), \tag{16}$$

where $\xi \sim \chi^2\left(I\right)$ is a random variable with chi-squared distribution with I degrees of freedom.

In view of Theorem 5, for a given alarm triggering threshold $\alpha > 0$, the false alarm rate, i.e., the probability of triggering an alarm when there is no attack, is given by $1 - \alpha$.

In order to compute ψ_T we need expressions for u_t, $\bar{u}$ and Σ. These are given in the next proposition, whose proof is also omitted in this conference version.

Proposition 6. For each $i, j = 1, \cdots, I$, we have

$$\left[u_t\right]_i = \prod_{l=0}^{L-1} \chi_{\left\{\check{z}_{t+l} \preceq \rho_{i,l}\right\}}, \tag{17}$$

$$\left[\bar{u}\right]_i = \prod_{l=0}^{L-1} \Phi_{0,\mathbf{I}}\left(\rho_{i,l}\right), \tag{18}$$

and

$$\begin{aligned}
\left[\Sigma\right]_{i,j} = & \sum_{t=0}^{-L+1} \prod_{\tau=0}^{-t-1} \Phi_{0,I}\left(\rho_{i,t+L+\tau}\right) \Phi_{0,I}\left(\rho_{j,\tau}\right) \times \\
& \times \prod_{\tau=0}^{L+t-1} \Phi_{0,I}\left(\rho_{i,\tau} \wedge \rho_{j,-t+\tau}\right) \\
& + \sum_{t=1}^{L-1} \prod_{\tau=0}^{t-1} \Phi_{0,I}\left(\rho_{i,\tau}\right) \Phi_{0,I}\left(\rho_{j,L-t+\tau}\right) \times \\
& \times \prod_{\tau=0}^{L-t-1} \Phi_{0,I}\left(\rho_{i,t+\tau} \wedge \rho_{j,\tau}\right) \\
& - \left(2L-1\right) \prod_{\tau=0}^{L-1} \Phi_{0,I}\left(\rho_{i,\tau}\right) \Phi_{0,I}\left(\rho_{j,\tau}\right).
\end{aligned} \tag{19}$$

The proposed method assesses the presence of an attack by measuring the squared distance between the nominal and empirical CDFs of $\zeta_1^{(L)}$. Since the domain of these functions is $\mathbb{R}^{LD}$, their distance is measured over the grid sample points $\rho_i \in \mathbb{R}^{LD}$, $i = 1, \cdots, I$. In order to complete the description of the method, we need a criterion for choosing these points. To this end, we apply the generalized Lloyd's algorithm [Gersho and Gray, 1991, S 11.3] to the nominal probability distribution of $\zeta_1^{(L)}$, i.e., $\mathcal{N}\left(0, \mathbf{I}\right)$. We then obtain the algorithm summarized in Algorithm 1.

```
Algorithm 1 Attack detection test
Initialization: choose $L, T, I \in \mathbb{N}$ and a threshold $\alpha > 0$.
  (1) Run Lloyd's algorithm on the $LD$-dimensional distribution $\mathcal{N}(0, \mathbf{I})$, to obtain $\rho_i \in \mathbb{R}^{LD}$, $i = 1, \cdots, I$.
  (2) Using the points $\rho_i$, $i = 1, \cdots, I$, compute $\bar{u}$ and $\Sigma$.
Main loop: at time $t$, let $\tau = t - L + 1$ and run the following steps.
  (1) Run the Kalman filter (3) to obtain $\hat{z}_{t|t-1}$.
  (2) Compute $\check{z}_t$ using (4).
  (3) Compute $\zeta_\tau^{(L)}$ using $\check{z}_\tau, \cdots, \check{z}_t$.
  (4) Compute $\hat{u}_T$ using $\zeta_{\tau-T+1}^{(L)}, \cdots, \zeta_\tau^{(L)}$.
  (5) Using $\hat{u}_T$ compute $w_T$ and then $v_T$.
  (6) Compute $\psi_T$ using $v_T$ and (16).
  (7) Trigger an alarm if $\psi_T \geq \alpha$.
```

6. SIMULATION

In this section we illustrate our proposed method. Since this method checks that the *joint* statistics (JS) of a block of contiguous samples equal their nominal values, we refer to it as JS. We compare the JS method with other two methods. The first method is the one described in Section 4.2, which checks for the normality as well as pairwise independence (NPI) of samples of $\check{z}$. To do so, the method compares the joint CDFs of the vector $\left[\check{z}_t^\top, \check{z}_{t-l}^\top\right]^\top$, for all values $l = 1, \cdots, L-1$, using a procedure similar to the one described in Section 5. This yields the $L-1$ statistics $v_T^{(l)}$, $l = 1, \cdots, L-1$, which are computed as in (15). We refer to this method as NPI. The second method is the one described in [Mo and Sinopoli, 2016, eqs. (6)-(7)]. This method checks that the second order

(SO) statistics of samples of the prediction error $\tilde{z}$ equal their nominal values. In our notation, it defines

$$v_T = \tilde{z}_T^\top \Gamma \tilde{z}_T \sim \chi(D).$$

We refer to this method as SO.

To do the comparison, we use a system with $A = 0.98$, $C = 1$, $R = 0.1$ and $Q = 0.1$. Also, for the JS and NPI methods we use $I = 100$, $L = 3$ and $T = 100$.

In the first experiment we consider the attack described in (7), with $\tau = 1$ and $\alpha = 1/\sqrt{2}$. As described in Section 4.1, this attack introduces statistical dependence between samples of $\tilde{z}$ which are τ samples away from each other. However, these samples remain uncorrelated. Figures 1, 2 and 3 show the values of the statistic v_T for the methods JS, NPI and SO, respectively. We see how the SO method is unable to detect the appearance of the attack at time $t = 2.5 \times 10^4$.

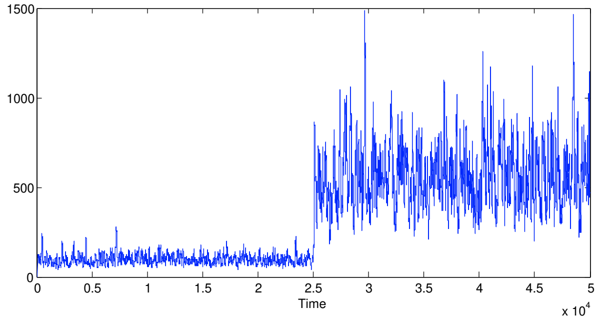

Fig. 1. Evolution of v_T yield by the JS detection method under the attack (7).

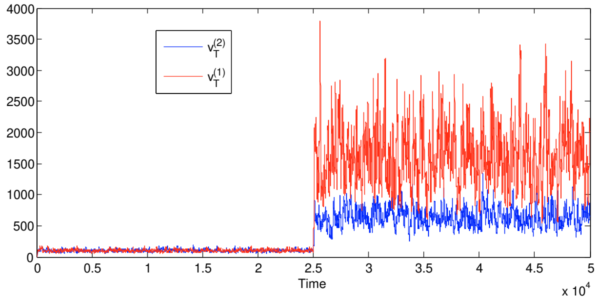

Fig. 2. Evolution of $v_T^{(1)}$ and $v_T^{(2)}$ yield by the NPI detection method under the attack (7).

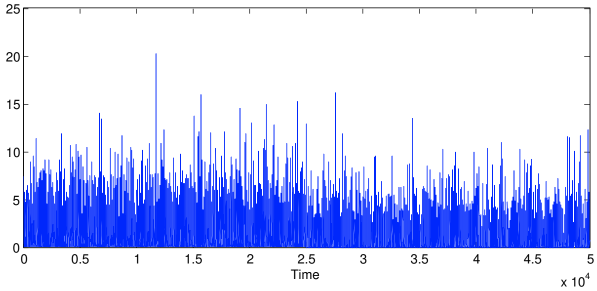

Fig. 3. Evolution of v_T yield by the SO detection method Mo and Sinopoli [2016] under the attack (7).

In the second experiment, we consider the attack described in (9). As explained in that section, this attack introduces statistical dependence between three consecutive samples of $\tilde{z}$, while leaving all samples from $\tilde{z}$ being pairwise independent. Again, the values of v_T for the methods JS, NPI and SO, are shown in Figures 4, 5 and 6, respectively. We see that in this case, only the JS method is able to detect the appearance of the attack.

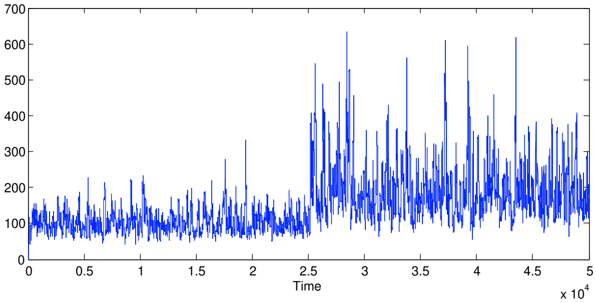

Fig. 4. Evolution of v_T yield by the JS detection method under the attack (9).

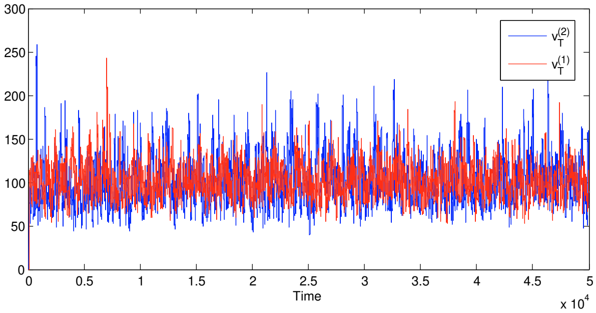

Fig. 5. Evolution of $v_T^{(1)}$ and $v_T^{(2)}$ yield by the NPI detection method under the attack (9).

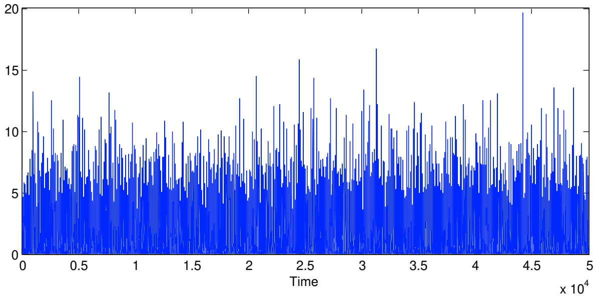

Fig. 6. Evolution of v_T yield by the SO detection method Mo and Sinopoli [2016] under the attack (9).

7. CONCLUSION

We studied the attack detection problem on stochastic cyber-physical systems. We proposed a detection criterion, and showed that it is equivalent to verifying that the output statistics correspond to a system without attack. Using this criterion, we derived a practically realizable attack detection algorithm. We present simulation results showing how our algorithm can detect attacks that cannot be detected by some simpler methods.

REFERENCES

S Amin, X Litrico, SS Sastry, and AM Bayen. Stealthy deception attacks on water scada systems. In *Proceedings of the 13th ACM international conference on Hybrid systems: computation and control*, pages 161–170, 2010.

Saurabh Amin, Alvaro Cardenas, and Shankar Sastry. Safe and secure networked control systems under denial-of-service attacks. *HSCC*, 5469:31–45, 2009.

Nail K Bakirov, Maria L Rizzo, and Gábor J Székely. A multivariate nonparametric test of independence. *Journal of multivariate analysis*, 97(8):1742–1756, 2006.

Alvaro Cardenas, Saurabh Amin, and Shankar Sastry. Research challenges for the security of control systems. In *HotSec*, 2008.

Thomas Chen. Stuxnet, the real start of cyber warfare? *IEEE Network*, 24(6):2–3, 2010.

Pablo Conti. The day the samba stopped [power blackouts]. *Engineering and Technology*, 5(4):46–47, 2010.

Gyorgy Dan and Henrik Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *First IEEE International Conference on Smart Grid Communications*, pages 214–219, 2010.

Christopher DeMarco, J. V. Sariashkar, and Fernando Alvarado. The potential for malicious control in a competitive power systems environment. In *Proceedings of the IEEE International Conference on Control Applications*, pages 462–467, 1996.

Demetrios Eliades and Marios Polycarpou. A fault diagnosis and security framework for water systems. *IEEE Transactions on Control Systems Technology*, 18(6): 1254–1265, 2010.

James Farwell and Rafal Rohozinski. Stuxnet and the future of cyber war. *Survival*, 53(1):23–40, 2011.

Allen Gersho and Robert M. Gray. *Vector Quantization and Signal Compression*. Springer, 1991.

Peter Huber. *Robust statistics*. Springer Berlin Heidelberg, 2011.

Svetlana Kuvshinkova. Sql slammer worm lessons learned for consideration by the electricity sector. *North American Electric Reliability Council*, 1(2):4–5, 2003.

Yao Liu, Peng Ning, and Michael Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security*, 14(1):13–14, 2011.

Y. Mo, E. Garone, A. Casavola, and B Sinopoli. False data injection attacks against state estimation in wireless sensor networks. In *49th IEEE Conference on Decision and Control*, pages 5967–5972, 2010.

Yilin Mo and Bruno Sinopoli. Secure control against replay attacks. In *47th Annual Allerton Conference on Communication, Control, and Computing*, pages 911–918, 2009.

Yilin Mo and Bruno Sinopoli. Secure estimation in the presence of integrity attacks. *IEEE Transactions on Automatic Control*, 60(4):1145–1151, 2015.

Yilin Mo and Bruno Sinopoli. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, 61 (9):2618–2624, 2016.

Amir-Hamed Mohsenian-Rad and Alberto Leon-Garcia. Distributed internet-based load altering attacks against smart power grids. *IEEE Transactions on Smart Grid*, 2(4):667–674, 2011.

Fabio Pasqualetti, Florian Dorfler, and Francesco Bullo. Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design. In *50th IEEE Conference on Decision and Control*, pages 2195–2201, 2011.

Fabio Pasqualetti, Florian Dorfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems–part ii: Centralized and distributed monitor design. *arXiv preprint arXiv:1202.6049*, 2012.

Fabio Pasqualetti, Florian Dorfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58 (11):2715–2729, 2013.

Monish Puthran, Sangeet Puthur, and Radhika Dharulkar. Smart traffic signal. *Int. Journal of Comp Sc and Info Tech*, 6(2):1360–1363, 2015.

Guy Richards. Hackers vs slackers-[control security]. *Engineering and Technology*, 3(19):40–43, 2008.

Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. *Critical infrastructure protection*, pages 73–82, 2007.

Roy Smith. A decoupled feedback structure for covertly appropriating networked control systems. *IFAC Proceedings*, 44(1):90–95, 2011.

Siddharth Sridhar, Adam Hahn, and Manimaran Govindarasu. Cyberšcphysical system security for the electric power grid. *Proceedings of the IEEE*, 100(1):210–224, 2012.

A. Teixeira, S. Amin, H. Sandberg, K. Johansson, and S. Sastry. Cyber security analysis of state estimators in electric power systems. In *49th IEEE Conference on Decision and Control*, pages 5991–5998, 2010.

Henry C Thode. *Testing for normality*, volume 164. CRC press, 2002.

Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Science & Business Media, 2004.

Alan Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12(6):601–611, 1976.

Kemin Zhou, John Doyle, and Keith Glover. *Robust and optimal control*. New Jersey: Prentice hall, 1996.