Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Sliding mode strategies for monitoring and compensation of cyber-attacks to Cyber-Physical Systems ⋆

Azevedo Filho, Jair L. * **Nunes, Eduardo V. L.** * **Hsu, L.** *

* *COPPE-Federal University of Rio de Janeiro, Rio de Janeiro, RJ, 21941-901, Brazil (e-mail: jair.azevedo@ufrj.br; eduardo@coep.ufrj.br; liu@coep.ufrj.br).*

Abstract: In this paper, the problem of detection and reconstruction of cyber-attacks in linear cyber-physical systems is considered. The class of cyber-attacks described in this paper can corrupt the states or the outputs of a cyber-physical system. An attack monitor based on High-Order sliding mode is proposed to reconstruct the cyber-attack. A First Order Approximation Filter is proposed to ensure global stability and convergence results. Using sliding mode techniques, an attack compensation is developed for square plants, guarateeing finite time convergence to the output tracking while rejecting the effects of the cyber-attack.

Keywords: Cyber-Physical Systems, sliding mode control, disturbance rejection, output feedback control, cyber-attack reconstruction.

1. INTRODUCTION

Cyber-physical Systems (CPS) can be found in the main infrastructures of a society, such as power generation, smart grids, and transportation networks. This class of system integrates physical processes, communication capabilities and computational resources (Poovendran et al., 2011). Although CPS improves efficiency, it becomes more susceptible to attacks envolving the cyber-domain, as known as Cyber-attacks (Sandberg et al., 2015). As presented in (Pasqualetti et al., 2013), great damages can be led by cyber attacks on the communication and data channels, e.g. power blackouts in Brazil (Conti, 2010), and Stuxnet storm (Langner, 2011). These attacks proved that the security mechanisms already used must be complemented with control strategies capable of detecting and rejecting this kind of attacks.

Some results have been published on the context of detection and reconstruction of Cyber-Attacks in recent years. In (Nateghi et al., 2018), nonlinear systems are considered. Using HOSM differentiators, cyber-attacks are approximated through an optimization problem. This technique, however, considers that all states of the system are available. In (Huang et al., 2018), uncertain linear systems are studied. Based on Integral Sliding Mode and adaptation laws for the upper bound of the attack, the proposed controller ensures quasi-optimal performance to the system. However, this work considers that the system without external disturbances and attacks is available before its implementation.

Based on (Ao et al., 2016), an interesting sliding mode strategy was proposed in (Corradini and Cristofaro, 2017) to detect, reconstruct and compensate state attacks and sensor attacks in linear Cyber-Physical Systems. To this end, an attack monitor and a state observer are proposed to ensure that the estimation errors and the monitoring errors are bounded by positive constants, while a compensator based on First Order Sliding Mode is designed.

The following constraints and assumptions restrict the application of the strategy proposed in (Corradini and Cristofaro, 2017):

- The attack monitor may have limited performance since it does not ensure the convergence of the monitoring error to zero. Furthermore, the proposed attack estimation is not defined when the output estimation error becomes lower than a predefined positive constant.
- The attack detection/reconstruction in each state or output channel requires one attack monitor and one state observer. Therefore, monitoring many channels becomes computationally expensive.
- The attack compensation scheme does not ensure the tracking error convergence to zero.
- The attack and its time derivative are assumed to have known upper bounds.
- Only local results of stability and convergence are achieved since an upper bound for the non measurable states is assumed.

Inspired by the strategy presented in (Corradini and Cristofaro, 2017), this paper proposes an alternative sliding mode-based strategy in order to overcome the previous limitations discussed. Here, the following improvements are achieved.

- The output estimation error converges to zero in finite time and the attack reconstruction is fully defined. Furthermore, in the absence of external disturbances, the monitoring error tends exponentially to zero.

⋆ This work was supported in part by CAPES, CNPq, and FAPERJ.

- The attack vector is reconstructed using only one attack monitor and one state observer.
- The proposed attack compensation ensures that the tracking error converges to zero in finite time. Furthermore, since a High Order Sliding Mode technique is considered, the chattering effect is attenuated.
- The upper bound of the attack is no longer required.
- Global stability and convergence results are achieved by using a First Order Approximation Filter to estimate a norm bound for the non measurable states.

Note that the proposed strategy ensures better reconstruction results under less restrictive assumptions that in (Corradini and Cristofaro, 2017). Therefore, the approach proposed in this paper can be applied to a broader class of Cyber-Pysical Systems (CPS) with better performance and global stability and convergence properties.

Preliminaries: The euclidean norm of a vector $\mathbf{y}$ and the corresponding induced norm of a matrix $\mathbf{A}$ are denoted by $||\mathbf{y}||$ and $||\mathbf{A}||$, respectively. I_n is the identity matrix of $\mathbb{R}^{n\times n}$. Here, Fillipov's definition for the solutions of discontinuous differential equations is assumed (Filippov, 1964).

2. SLIDING MODE STRATEGIES APPLIED FOR MONITORING AND COMPENSATION OF CYBER-ATTACKS

This section presents the properties of the class of CPS and cyber-attacks considered in this paper. The attack monitor for deception and stealth attacks is presented in sequence.

2.1 System properties

Consider the following CPS:

$$\begin{aligned}\dot{\boldsymbol{\zeta}}(t) &= \bar{A}\boldsymbol{\zeta(t)} + \bar{B}_u\boldsymbol{u}(t) + \bar{B}_f\boldsymbol{f}(\boldsymbol{\zeta},t) + \bar{D}_d\boldsymbol{d}(\boldsymbol{\zeta},t)\\ \boldsymbol{y}(t) &= \bar{C}\boldsymbol{\zeta}(t) + \bar{D}_u\boldsymbol{u}(t) + \bar{D}_f\boldsymbol{f}(\boldsymbol{\zeta},t),\end{aligned} \tag{1}$$

where $\boldsymbol{\zeta}(t) \in \mathbb{R}^n$ is the state vector, $\mathbf{u}(t) \in \mathbb{R}^m$ is the input of the system and $\mathbf{y}(t) \in \mathbb{R}^p$ is the output of the system, with $p \geq m$. The matrices $\bar{A}$, $\bar{B}_u$, $\bar{C}$ and $\bar{D}_u$ have compatible dimensions. The vector $\bar{D}_d\mathbf{d}(\boldsymbol{\zeta},t)$ describes any external disturbance or model uncertainty and the terms $\bar{B}_f\mathbf{f}(\boldsymbol{\zeta},t)$ and $\bar{D}_f\mathbf{f}(\boldsymbol{\zeta},t)$ represent the state attack and the sensor attack, respectively. Assuming that $\bar{B}_f$ and $\bar{D}_f$ are known matrices, the attack monitor should detect and reconstruct the attack vector $\mathbf{f}(\boldsymbol{\zeta},t) \in \mathbb{R}^q$. As in (Corradini and Cristofaro, 2017), the state vector $\boldsymbol{\zeta}(t)$ is assumed to include both physical and cyber variables, while $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are known signals.

Since the system is known, except for the uncertainty disturbance vector, state and sensor attacks can be detected and reconstructed by using a Luemberger-like estimator. To this end, the class of CPS considered must respect the following Assumptions, derived from (Ao et al., 2016):

Assumption 1:

1. The pair $(\bar{A},\bar{C})$ is completely observable.
2. $rank(\bar{C}\bar{B}_f) = rank(\bar{B}_f)$, where $\bar{B}_f$ is full column rank.
3. The matrices $\bar{B}_u$ and $\bar{D}_f$ are full column rank.
4. The invariant zeros of $(\bar{A},\bar{B}_u,\bar{C},\bar{D}_u)$ are stable.
5. Attacks are assumed detectable, i.e., the system $(\bar{A},\bar{B}_f,\bar{C},\bar{D}_f)$ has no invariant zeros.

Assumption 1.1 is quite reasonable once an estimator is proposed in this monitoring scheme. Assumption 1.2 implies that $q \leq p$. Therefore, at most p attacks can be detected/reconstructed at the same time. Assumption 1.3 is considered since that there are no reason to use unnecessary inputs in $\bar{B}_u$ and both $\bar{B}_f$ and $\bar{D}_f$ are full rank for observer design. Note that the CPS is a minimum phase system, since Assumption 1.4 holds. Finally, Assumption 1.5 ensures that neither non-zero unstable undetectable attacks nor non-zero stable undetectable attacks can occur.

Note that the choice of $\bar{B}_f$ and $\bar{D}_f$ define in which channels the attack detection occurs. In (Corradini and Cristofaro, 2017), these matrices are defined as column vectors, which means that the attack $f(t)$ is a scalar. Thus, to detect p attacks simultaneously, p observers must be developed, which is expensive from a computational point of view. In this work, it is considered that the matrix $\bar{B}_f \in \mathbb{R}^{n\times p}$ and $\bar{D}_f \in \mathbb{R}^{p\times p}$, i.e., the attack monitor reconstructs an attack vector, with p channels.

Remark 1*:* Note that if $q < p$, the $p-q$ channels of attack monitor in absence of non-zero attacks are zero, since $\bar{B}_f$ and $\bar{D}_f$ are full column rank.

Assumption 1 implies that there exists a linear change of coordinate $\mathbf{x} = T\boldsymbol{\zeta}$, where

$$\begin{aligned}\dot{\mathbf{x}}(t) &= A\mathbf{x}(t) + B_u\mathbf{u}(t) + B_f\mathbf{f}(t) + D_d\mathbf{d}(t)\\ \mathbf{y}(t) &= C\mathbf{x}(t) + D_u\mathbf{u}(t) + D_f\mathbf{f}(t),\end{aligned} \tag{2}$$

with

$$A = \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix} \quad B_u = \begin{bmatrix}B_1\\ B_2\end{bmatrix} \quad B_f = \begin{bmatrix}0\\ B\end{bmatrix} \tag{3}$$

$$C = [0 \ I_p] \quad D_d = \begin{bmatrix}D_1\\ D_2\end{bmatrix} \quad \bar{D}_f = D_f \quad \bar{D}_u = D_u \tag{4}$$

where $B \in \mathbb{R}^{p\times p}$ is a nonsingular matrix and $A_{11} \in \mathbb{R}^{n-p\times n-p}$ is a Hurwitz matrix.

2.2 Class of attacks

The model of attacks considered in this paper are the *Deception Attacks* (Teixeira et al., 2010) and *Stealth Attacks* (Ao et al., 2016).

Deception attacks affects the states of the system, which can change the steady-state system behavior and its stability, as presented in (Teixeira et al., 2010; Ao et al., 2016). *Stealth attacks* corrupts the measurements of the system. This attack can be used to affect output-feedback controllers (affecting indirectly the states of the system) or to hide deception attacks. When these attacks are used together, they are known as *Coordinated Attacks*.

Based on the previously arguments, it is noted that the detection of coordinated attacks could be pratically unfeasible (Pasqualetti et al., 2015). Since by the Assumption 1.5, the attacks are assumed detectable, a suficient condition is given by Assumption 2.

Assumption 2: Coordinated attacks do not occur, i.e., $||\bar{B}_f|| \cdot ||\bar{D}_f|| = 0$.

Note that this assumption implies that B_f or D_f is a zero matrix. Therefore, the attack vector $\mathbf{f}(x,t)$ in (1) does not occur simultaneously as deception and stealth attack. Furthermore, the following assumption holds

Assumption 3: The time derivative of the attack vector has a known upper bound:

$$||\dot{\mathbf{f}}(\mathbf{x},t)|| \leq \rho_2(\mathbf{x},t) \tag{5}$$

where $\rho_2(\mathbf{x},t)$ is a positive function.

2.3 Detection and reconstruction of deception attacks

Note that, according to Assumption 2, $D_f = 0$. Rewriting (2) it follows that:

$$\begin{aligned}\dot{\mathbf{x}}_1(t) &= A_{11}\mathbf{x}_1(t) + A_{12}\mathbf{x}_2(t) + B_1\mathbf{u}(t) + D_1\mathbf{d}(\mathbf{x},t)\\ \dot{\mathbf{x}}_2(t) &= A_{21}\mathbf{x}_1(t)+A_{22}\mathbf{x}_2(t)+B_2\mathbf{u}(t)+B\mathbf{f}(\mathbf{x},t)+D_2\mathbf{d}(\mathbf{x},t)\\ \mathbf{y}(t) &= \mathbf{x}_2(t) + D_u\mathbf{u}(t),\end{aligned} \tag{6}$$

Since $\mathbf{y}(t)$ and $D_u\mathbf{u}(t)$ are known signals, consider the following estimator

$$\begin{aligned}\dot{\hat{\mathbf{x}}}_1(t) &= A_{11}\hat{\mathbf{x}}_1(t) + B_1\mathbf{u}(t) + A_{12}\left(\mathbf{y}(t) - D_u\mathbf{u}(t)\right)\\ \dot{\hat{\mathbf{x}}}_2(t) &= A_{21}\hat{\mathbf{x}}_1(t)+B_2\mathbf{u}(t)+B\hat{\mathbf{f}}(x,t)+A_{22}(\mathbf{y}(t)-D_u\mathbf{u}(t))\\ \hat{\mathbf{y}}(t) &= \hat{\mathbf{x}}_2(t) + D_u\mathbf{u}(t),\end{aligned} \tag{7}$$

Defining the estimation errors as $\mathbf{e}_1(t) = \mathbf{x}_1(t) - \hat{\mathbf{x}}_1(t)$ and $\mathbf{e}_2(t) = \mathbf{x}_2(t) - \hat{\mathbf{x}}_2(t)$, it follows that:

$$\begin{aligned}\dot{\mathbf{e}}_1(t) &= A_{11}\mathbf{e}_1(t) + D_1\mathbf{d}(\mathbf{x},t)\\ \dot{\mathbf{e}}_2(t) &= A_{21}\mathbf{e}_1(t) + B\left(\mathbf{f}(\mathbf{x},t) - \hat{\mathbf{f}}(\mathbf{x},t)\right) + D_2\mathbf{d}(\mathbf{x},t),\end{aligned} \tag{8}$$

where $e_2(t)$ is a measurable signal, since $y(t) - \hat{y}(t) = e_2(t)$. Thus, the estimation error e_2 can be applied in the monitoring scheme. Since that A_{11} is a Hurwitz matrix, $\mathbf{e}_1(t)$ can be interpreted as a stable filter for the signal $\mathbf{d}(\mathbf{x},t)$. In addition to this, the following assumption holds.

Assumption 4: The disturbance vector and its derivative have known upper bounds:

$$||\mathbf{d}(\mathbf{x},t)|| \leq \rho_d(\mathbf{x},t) \qquad ||\dot{\mathbf{d}}(\mathbf{x},t)|| \leq \bar{\rho}_d(\mathbf{x},t) \tag{9}$$

where $\rho_d(\mathbf{x},t)$ and $\bar{\rho}_d(\mathbf{x},t)$ are known bounded functions.

Based on the above results, it is known that for any bounded initial condition of $\mathbf{e}_1(0)$, $\mathbf{e}_1(t)$ remains bounded for all $t \geq 0$, since that:

$$\begin{aligned}||\mathbf{e}_1(t)|| &\leq \alpha_1 e^{-\gamma t}||\mathbf{e}_1(0)|| + \alpha_2\int_0^t e^{-\gamma(t-\tau)}||D_d||\cdot||\mathbf{d}(\mathbf{x},t)||d\tau\\ &\leq \alpha_1 e^{-\gamma t}||\mathbf{e}_1(0)|| + \alpha_2||D_d||\int_0^t e^{-\gamma(t-\tau)}\rho_d(\mathbf{x},t)d\tau \leq \rho\end{aligned} \tag{10}$$

where α_1, α_2, $||\mathbf{e}_1(0)||$, and ρ are bounded positive constants. Note that the constant upper bound ρ is considered known in (Corradini and Cristofaro, 2017). Since $||\mathbf{e}_1(0)||$ is unknown, in this work, we consider that ρ is *unknown.*

Remark 2*:* Note from (8) that, although the exponentially convergence of $\mathbf{e}_1(t)$ is hindered by $\mathbf{d}(\mathbf{x},t)$, its boundedness is ensured by Assumption 4.

Chosing $\hat{\mathbf{f}}(\mathbf{x},t) = -B^{-1}\bar{\mathbf{f}}(\mathbf{x},t)$, the estimation error e_2 in (8) can be written as

$$\dot{\mathbf{e}}_2(t) = \bar{\mathbf{f}}(\mathbf{x},t) + \boldsymbol{\Delta}(\mathbf{e}_1,\mathbf{x},t) \tag{11}$$

with

$$\boldsymbol{\Delta}(\mathbf{e}_1,\mathbf{x},t) = A_{21}\mathbf{e}_1(t) + B\mathbf{f}(\mathbf{x},t) + D_2\mathbf{d}(\mathbf{x},t). \tag{12}$$

From (12) and (8), it follows that:

$$\begin{aligned}\dot{\boldsymbol{\Delta}}(\mathbf{e}_1,\mathbf{x},t) &= A_{21}\dot{\mathbf{e}}_1(t) + B\dot{\mathbf{f}}(\mathbf{x},t) + D_2\dot{\mathbf{d}}(\mathbf{x},t)\\ \dot{\boldsymbol{\Delta}} &= A_{21}(A_{11}\mathbf{e}_1(t)+D_1\mathbf{d}(\mathbf{x},t))+B\dot{\mathbf{f}}(\mathbf{x},t)+D_2\dot{\mathbf{d}}(\mathbf{x},t)\end{aligned} \tag{13}$$

Note that (13) can be upper bounded by

$$||\dot{\boldsymbol{\Delta}}|| \leq a_1||\mathbf{e}_1(t)|| + a_2||\mathbf{d}(\mathbf{x},t)|| + a_3||\dot{\mathbf{f}}(\mathbf{x},t)|| + a_4||\dot{\mathbf{d}}(\mathbf{x},t)|| \tag{14}$$

where $a_1 = ||A_{21}A_{11}||$, $a_2 = ||A_{21}D_1||$, $a_3 = ||B||$, and $a_4 = ||D_2||$.

As described in (10), the knowledge of an upper bound for $||\mathbf{e}_1(t)||$ is not considered, because this assumption restricts the convergence and stability results to a local result, since $||\mathbf{e}_1(0)|| \leq \mu$, where $\mu \in \mathbb{R}$. To circumvent this constraint, in this work we propose the application of a First Order Approximation Filter (FOAF) (Cunha et al., 2003; Hsu et al., 1997):

$$\dot{\hat{\eta}}_e(t) = -\lambda_f\hat{\eta}_e(t) + c_f||\mathbf{d}(\mathbf{x},t)|| \tag{15}$$

where, $\hat{\eta}_e(t) + |\pi(t)| > ||\mathbf{e}_1(t)||$ and

$$|\pi(t)| = c_\eta||\mathbf{e}_\eta(t_0)||e^{-\lambda_\eta(t-t_0)} \tag{16}$$

for some c_η,$\lambda_\eta > 0$ and $\mathbf{e}_\eta(t) = [\mathbf{e}_1^T\ \hat{\eta}_e]^T$. Thus, it can be noted that after some finite time t_1, $\hat{\eta}_e(t) > ||\mathbf{e}_1(t)||$,$\forall t > t_1$

Based on (8) the parameters of (15) can be defined as:

$$\lambda_f = \min_j\{-Re(\lambda_j)\} \qquad c_f = ||D_1||$$

where $\{\lambda_j\}$ are the eigenvalues of A_{11}.

Therefore, from (15) and Assumptions 3 and 4, after some finite time, (14) can be further upper bounded by

$$\boldsymbol{\Delta}^\star(t) := a_1\hat{\eta}_e(t)+a_2\rho_d(\mathbf{x},t)+a_3\rho_2(\mathbf{x},t)+a_4\bar{\rho}_d(\mathbf{x},t) \tag{17}$$

where $\boldsymbol{\Delta}^\star(t)$ is a known upper bound for $||\dot{\boldsymbol{\Delta}}(t)||$, with a_1, a_2, a_3, and a_4 given by (14).

Finally, note that (11) is a first order system with a perturbation term whose time-derivative has a known upper bound. For this class of system there is a wide range of controllers that can reject the disturbance term and drive the sliding variable to the origin. In this work, a well-known sliding-mode strategy is considered, the Variable-Gain Super Twisting Algorithm.

2.4 Deception attack reconstruction using Multivariable Global Variable Gains Super-Twisting Algorithm

Super-Twisting algorithm is a second-order sliding mode technique that ensures finite time convergence for the sliding variable and its derivative. This algorithm became popular since the time derivative of the sliding variable is not required. Furthermore, as a High Order Sliding Mode technique, the Super Twisting can attenuate the chattering effect present in First Order Sliding Mode techniques.

In this work, the Variable Gains Super-Twisting Algorithm (VGSTA) for MIMO systems proposed in Vidal et al. (2017), is considered. The algorithm is given by:

$$\bar{\mathbf{f}}(\mathbf{x},t) = -k_1(\mathbf{x},\hat{\eta}_e,t)\boldsymbol{\phi}_1(\mathbf{e}_2) - \int_{t_0}^t k_2(\mathbf{x},\hat{\eta}_e,t)\boldsymbol{\phi}_2(\mathbf{e}_2)dt \tag{18}$$

where

$$\begin{aligned}\boldsymbol{\phi}_1(\mathbf{e}_2) &= \frac{\mathbf{e}_2}{||\mathbf{e}_2||^{\frac{1}{2}}} + k_3\mathbf{e}_2\\ \boldsymbol{\phi}_2(\mathbf{e}_2) &= \frac{1}{2}\frac{\mathbf{e}_2}{||\mathbf{e}_2||} + \frac{3k_3}{2}\frac{\mathbf{e}_2}{||\mathbf{e}_2||^{\frac{1}{2}}} + k_3^2\mathbf{e}_2\end{aligned}, \ k_3 > 0 \tag{19}$$

From the control law given by (18) and the disturbance term (12), the closed-loop system dynamics is given by:

$$\begin{aligned}\dot{\hat{\eta}}_e(t) &= -\lambda_f \hat{\eta}_e(t) + c_f||\mathbf{d}(\mathbf{x},t)||\\ \dot{\mathbf{e}}_1(t) &= A_{11}\mathbf{e}_1(t) + D_1\mathbf{d}(\mathbf{x},t)\\ \dot{\mathbf{e}}_2(t) &= -k_1(\mathbf{x},\hat{\eta}_e,t)\boldsymbol{\phi}_1(\mathbf{e}_2) + \mathbf{z}(t) + |\pi_1(t)|\boldsymbol{\phi}_1(\mathbf{e}_2)\\ \dot{\mathbf{z}}(t) &= -k_2(\mathbf{x},\hat{\eta}_e,t)\boldsymbol{\phi}_2(\mathbf{e}_2) + \dot{\boldsymbol{\Delta}}(\mathbf{e}_1,\mathbf{x},t) + |\pi_2(t)|\boldsymbol{\phi}_2(\mathbf{e}_2)\end{aligned} \tag{20}$$

where the terms $|\pi_1(t)|\boldsymbol{\phi}_1(\mathbf{e}_2)$ and $|\pi_2(t)|\boldsymbol{\phi}_2(\mathbf{e}_2)$ are used to account for the presence of the FOAF in the closed-loop system, with $|\pi_1(t)|$ and $|\pi_2(t)|$ defined similarly as in (16).

Note that, for $\mathbf{e}_2(t) \neq \mathbf{0}$, one has that:

$$||\dot{\boldsymbol{\Delta}}|| = 2||\dot{\boldsymbol{\Delta}}|| \cdot \frac{1}{2}\left|\left|\frac{\mathbf{e}_2}{||\mathbf{e}_2||}\right|\right| \leq \varrho_2(\mathbf{e}_1,\mathbf{x},t)||\boldsymbol{\phi}_2(\mathbf{e}_2)|| \tag{21}$$

where, from (17), $\varrho_2(\mathbf{e}_1,\mathbf{x},t)$ can be defined as $2\boldsymbol{\Delta}^\star(t)$. From (20), it follows that:

$$\left|\left|\dot{\boldsymbol{\Delta}} + |\pi_2(t)|\boldsymbol{\phi}_2(\mathbf{e}_2)\right|\right| \leq (\varrho_2(\mathbf{e}_1,\mathbf{x},t) + |\pi_2(t)|)||\boldsymbol{\phi}_2(\mathbf{e}_2)|| \tag{22}$$

Therefore, if the variable gains in (18) are defined as

$$\begin{aligned}k_1(\mathbf{x},\hat{\eta}_e,t) &= \delta + \frac{1}{\beta}\left(\frac{\varrho_2^2}{4\varepsilon} + 2\varepsilon\varrho_2 + \varepsilon + 8\varepsilon^3 + 2\varepsilon\beta\right)\\ k_2(\mathbf{x},\hat{\eta}_e,t) &= \beta + 4\varepsilon^2 + 2\varepsilon k_1(\mathbf{x},\hat{\eta}_e,t)\end{aligned} \tag{23}$$

where δ, β, and ε are arbritrary positive constants, then a second order sliding mode $\mathbf{e}_2 = \dot{\mathbf{e}}_2 = 0$ is reached in finite time (Vidal et al., 2017, Theorem 1).

Note that when the sliding mode takes place, it follows from (20) that $\mathbf{z}(t) = 0$. Then, after a finite time, it follows from (11) and (12) that $\hat{\mathbf{f}}(\mathbf{x},t) = B^{-1}\boldsymbol{\Delta}(\mathbf{e}_1,\mathbf{x},t)$. Furthermore, from (12), one has that:

$$\begin{aligned}\hat{\mathbf{f}}(\mathbf{x},t) - \mathbf{f}(\mathbf{x},t) &= B^{-1}\left(A_{21}\mathbf{e}_1(t) + D_2\mathbf{d}(\mathbf{x},t)\right)\\ \left|\left|\hat{\mathbf{f}}(\mathbf{x},t) - \mathbf{f}(\mathbf{x},t)\right|\right| &\leq b_1||e_1(t)|| + b_2||d(x,t)||\end{aligned} \tag{24}$$

where $b_1 = ||B^{-1}|| \cdot ||A_{21}||$ and $b_2 = ||B^{-1}|| \cdot ||D_2||$.

From (10) and Assumption 4, $\mathbf{d}(\mathbf{x},t)$ and $\mathbf{e}_1(t)$ are bounded signals. Thus, it follows that the monitoring error $\left|\left|\hat{\mathbf{f}}(\mathbf{x},t) - \mathbf{f}(\mathbf{x},t)\right|\right|$ is bounded. Furthermore, if $\mathbf{d}(\mathbf{x},t) \equiv 0$, it follows from (8) that $\hat{\mathbf{f}}(\mathbf{x},t)$ converges to $\mathbf{f}(\mathbf{x},t)$ exponentially. Hence, the attack function $\mathbf{f}(\mathbf{x},t)$ is reconstructed exponentially.

Most of the existing cyber-attacks and fault scenarios can be modeled by unknown additive inputs affecting the state and the measurements. Besides reflecting the genuine failure of system components, these disturbances model the effect of attacks against the cyberphysical system (Pasqualetti et al., 2015). Thus, our approach can be applied to both fault-tolerant and cyber-physical systems.

2.5 Deception attack reconstruction using Predefined Layer Algorithm

The attack monitor proposed in (Corradini and Cristofaro, 2017) is considered in this paper for comparison purposes. This method imposes that

$$\frac{d^2\left(||\mathbf{e}_2(t)||^2\right)}{dt^2} < 0 \quad \text{if} \quad ||\mathbf{e}_2(t)|| > \epsilon \tag{25}$$

with the following attack monitor:

$$\begin{aligned}\dot{\hat{f}}(t) &= \frac{\gamma k\left(\alpha(\mathbf{x},t)||\mathbf{e}_2|| + \beta(\mathbf{x},t)^2 + \kappa(\mathbf{x},t) + \eta\right)}{k\mathbf{e}_2^T B - \epsilon sign(\mathbf{e}_2^T B)}, \text{ if } |\mathbf{e}_2^T B| > \epsilon\\ \alpha(\mathbf{x},t) &:= a_1\rho + a_2\rho_d(t) + a_3\rho_2(\mathbf{x},t) + a_4\bar{\rho}_d(t)\\ \beta(\mathbf{x},t) &:= ||A_{21}||\rho + ||B||\rho_1(\mathbf{x},t) + ||D_2||\rho_d(t)\\ \kappa(\mathbf{x},t) &:= ||B||\left(||B||\hat{f}(t)^2 + 2\beta(\mathbf{x},t)|\hat{f}(t)|\right)\end{aligned}$$

where γ, k, ϵ and η are positive constants, $\rho_1(\mathbf{x},t)$ is an upper-bound for $||\mathbf{f}(\mathbf{x},t)||$ and ρ, $\rho_2(\mathbf{x},t)$, $\rho_d(t)$, and $\bar{\rho}_d(t)$ are given by (10), (5), and (9), respectively.

The imposed condition (25) ensures that, in finite time, the norm of $\mathbf{e}_2(t)$ decreases until that the layer $||\mathbf{e}_2(t)|| < \epsilon$ is reached. From now on in this paper, this method is referred by *Method 2*. Note that the authors did not specify $\hat{f}(\mathbf{x},t)$ inside the predefined layer. In the present work, we have assumed that $\hat{f}(\mathbf{x},t)$ was kept constant inside such layer. The simulation results of (Corradini and Cristofaro, 2017) and the ones presented here are similar.

2.6 Detection and reconstruction of stealth attacks

Invoking Assumption 2, it is known that during a stealth attack, there is no deception attack ($B_f = 0$). Rewriting (1), it follows that:

$$\begin{aligned}\dot{\boldsymbol{\zeta}}(t) &= \bar{A}\boldsymbol{\zeta}(t) + \bar{B}_u\mathbf{u}(t) + \bar{D}_d\mathbf{d}(\boldsymbol{\zeta},t)\\ \mathbf{y}(t) &= \bar{C}\boldsymbol{\zeta}(t) + \bar{D}_u\mathbf{u}(t) + \bar{D}_f\mathbf{f}(\boldsymbol{\zeta},t),\end{aligned} \tag{26}$$

Consider the following low-pass filter to the output of (26) proposed in (Ao et al., 2016).

$$\begin{aligned}\dot{\mathbf{x}}_f(t) &= A_f\mathbf{x}_f(t) + \mathbf{y}(t)\\ \mathbf{y}_f(t) &= \mathbf{x}_f(t)\end{aligned} \tag{27}$$

where $A_f \in \mathbb{R}^{p\times p}$ is a designed Hurwitz matrix. Note that with this "new state" $\mathbf{x}_f(t)$ the following augmented system is obtained:

$$\begin{aligned}\dot{\boldsymbol{\omega}}_1(t) &= \bar{A}\boldsymbol{\omega}_1(t) + \bar{B}_u\mathbf{u}(t) + \bar{D}_d\mathbf{d}(\mathbf{x},t)\\ \dot{\boldsymbol{\omega}}_2(t) &= \bar{C}\boldsymbol{\omega}_1(t) + A_f\boldsymbol{\omega}_2(t) + \bar{D}_u\mathbf{u}(t) + \bar{D}_f\mathbf{f}(\mathbf{x},t)\\ \mathbf{y}_f(t) &= \boldsymbol{\omega}_2(t)\end{aligned} \tag{28}$$

where $\boldsymbol{\omega}_1(t) = \boldsymbol{\zeta}(t)$ and $\boldsymbol{\omega}_2(t) = \mathbf{x}_f(t)$. For this augmented system the following observer is proposed:

$$\begin{aligned}\dot{\hat{\boldsymbol{\omega}}}_1(t) &= \bar{A}\hat{\boldsymbol{\omega}}_1(t) + \bar{B}_u\mathbf{u}(t)\\ \dot{\hat{\boldsymbol{\omega}}}_2(t) &= \bar{C}\hat{\boldsymbol{\omega}}_1(t) + A_f\mathbf{y}_f(t) + \bar{D}_u\mathbf{u}(t) + \bar{D}_f\hat{\mathbf{f}}(\mathbf{x},t)\\ \hat{\mathbf{y}}_f(t) &= \hat{\boldsymbol{\omega}}_2(t),\end{aligned} \tag{29}$$

Thus, defining the estimation errors as $\mathbf{e}_{\omega_1}(t) = \boldsymbol{\omega}_1(t) - \hat{\boldsymbol{\omega}}_1(t)$ and $\mathbf{e}_{\omega_2}(t) = \boldsymbol{\omega}_2(t) - \hat{\boldsymbol{\omega}}_2(t)$, it follows that:

$$\begin{aligned}\dot{\mathbf{e}}_{\omega_1}(t) &= \bar{A}\mathbf{e}_{\omega_1}(t) + \bar{D}_d\mathbf{d}(\mathbf{x},t)\\ \dot{\mathbf{e}}_{\omega_2}(t) &= \bar{C}\mathbf{e}_{\omega_1}(t) + \bar{D}_f\left(\mathbf{f}(\mathbf{x},t) - \hat{\mathbf{f}}(\mathbf{x},t)\right)\end{aligned} \tag{30}$$

where $\mathbf{e}_{\omega_2}$ is a measurable error, since $\mathbf{y}_f(t) - \hat{\mathbf{y}}_f(t) = \mathbf{e}_{\omega_2}(t)$. Comparing (30) to (8), since, from Assumption 4, $\mathbf{d}(\mathbf{x},t)$ is a bounded signal, the boundedness of $\mathbf{e}_{\omega_1}(t)$ depends on $\bar{A}$. Therefore, for Stealth attacks the following assumption is considered

Assumption 5: The matrix $\bar{A}$ presented in (1) is Hurwitz.

Remark 3: Note that with this assumption, a deduction similar to that made in (10) ensures that $\mathbf{e}_{\omega_1}(t)$ is a bounded signal. Furthermore, if $\mathbf{d}(\mathbf{x},t) \equiv \mathbf{0}$, $\mathbf{e}_{\omega_1}(t)$ converges exponentially to zero.

Comparing the dynamics of $\mathbf{e}_{\omega_2}(t)$ in (30) and (11), it is noted that the application of the VGSTA algorithm presented in this paper to stealth attacks is straightforward.

2.7 Compensation of deception attacks

Once the cyber-attack is detected and reconstructed, the estimated attack can be used to ensure that the effect of the cyber-attack is rejected and the system output converges to some desired trajectory $\mathbf{y}_d(t)$. Thus, the objective is to ensure that the tracking error $e_t(t) = y(t) - y_d(t)$ becomes zero after some finite time.

Note that the sliding-mode strategies presented here can be applied to systems with, at least, relative degree one. Therefore, here we consider that $D_u = 0$. It is noteworthing that the same condition is required in Method 2.

Remark 4*:* It is worth noting that this condition is only required to the attack compensation. Therefore, the proposed attack reconstruction does not require this condition.

Defining the control law as in (Corradini and Cristofaro, 2017), with $\Psi(t) = A_{21}\hat{x}_1(t) + B\hat{f}(x,t) + A_{22}y(t) - \dot{y}_d(t)$ and

$$u(t) = (GB_2)^{-1}(\bar{u} - G\Psi(t)) \tag{31}$$

when the sliding mode ($\mathbf{e}_2 = \dot{\mathbf{e}}_2 = 0$) takes place, the variable $\sigma_c(t) = G\boldsymbol{\epsilon}_t(t)$ presents the following dynamics:

$$\dot{\boldsymbol{\sigma}}_{\mathbf{c}}(t) = \bar{\mathbf{u}}(t). \tag{32}$$

where $\boldsymbol{\epsilon}_t(t) = \hat{\mathbf{y}}(t) - \mathbf{y}_{\mathbf{d}}(t)$ and $G \in \mathbb{R}^{m\times p}$ is defined such that GB_2 is a square full rank matrix (for more details, see Corradini and Cristofaro (2017)).

Comparing (32) and (11) it is known that finite time convergence for the sliding variable $\boldsymbol{\sigma}_{\mathbf{c}}(t)$ can be achieved using the VGSTA technique presented in this paper. Note that if G is full rank, in the square case, $\boldsymbol{\epsilon}_t$ also converges to zero in finite time. Thus, the finite time convergence for the tracking error is achieved.

2.8 Example of state attack monitoring and compensation: the IEEE 39 bus power system

Consider the IEEE 39 bus power system presented in (Mei et al., 2011; Zimmerman et al., 2010) with ten generators. A linear state-space representation can be achieved using the strategy described in (Dorfler and Bullo, 2012), known as Kron-reduction. In this simulation example, the same additional simplifications and numerical values of (Corradini and Cristofaro, 2017, Section 5.1) have been adopted. Furthermore, the matrix $\bar{B}_f$ is defined as $[\mathbf{0}_{10\times 10}\ I_{10}]^T$

A deception attack of the form $f(t) = \frac{1}{2}sin(0.2\pi t) \times (|x_{11}(t)|+1)$ has been considered to corrupt the first output of (6), starting from $t = 2$ s. The control objective is the tracking of the reference output $\mathbf{y}_d = [1\ 1\ 2\ \mathbf{0}_{1\times 7}]^T$. Note that, for this case, B_2 is square and full rank. Therefore, from (32), the matrix G was defined as the identity matrix.

Two techniques have been used as attack monitor: VGSTA and Method 2. In this simulation example, the upper bounds required for the Method 2 are defined as $\rho = 2.5$, $\rho_1(\mathbf{x},t) = 1.75$, and $\rho_2(\mathbf{x},t) = 2$. Note that the implementation of the VGSTA algorithm only requires the knowledge of $\rho_2(\mathbf{x},t)$. The initial condition of the plant (2) and the estimator (7) are selected as $x(0) = [-0.5\cdot\mathbf{1}_{1\times 5}\ 0.25\cdot\mathbf{1}_{1\times 5}]^T$ and $\hat{x}(0) = \mathbf{0}_{10\times 1}$.

The parameters used in the VGSTA are tuned as $\delta = 1$, $\varepsilon = 0.3$ and $\beta = 0.1$, while the parameters for Method 2 are $k = 1.1$, $\gamma = 1.1$, $\eta = 0.2$ and $\epsilon = 0.1$, as in (Corradini and Cristofaro, 2017). The results are reported in the following Figures.

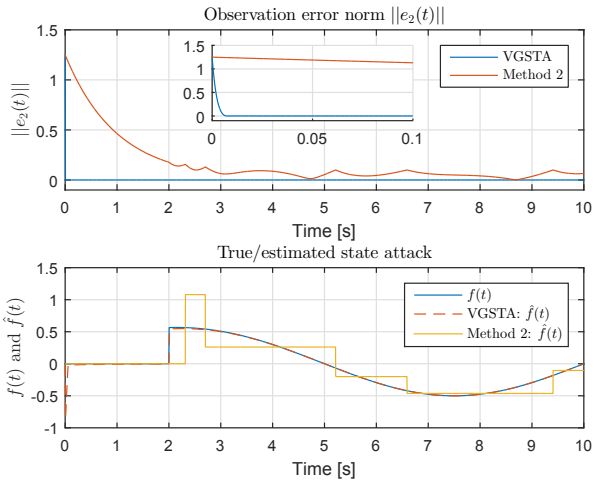

Fig. 1. (a) Observation error norm $||\mathbf{e}_2(t)||$ and (b) True $f(t)$ versus estimated $\hat{f}(t)$ state attack.

It is noted in Fig. 1 that the sliding mode is achieved in finite time using the VGSTA strategy adopted in this work. Furthermore, when the sliding mode is reached, the state attack reconstruction converges exponentially to the sinusoidal attack.

In the context of compensation of attacks, no signal must be reconstructed, since the attack was already estimated by the attack monitor $\hat{\mathbf{f}}(\mathbf{x},t)$. Therefore, other techniques based on sliding mode can be used in (32).

In this simulation example, the attack reconstructed by the VGSTA is compensated by a control law also based on VGSTA. The controller is given by (33).

$$\bar{\mathbf{u}}(\boldsymbol{\sigma}_{\mathbf{c}},t) = -k_1(\mathbf{x},\hat{\eta}_e,t)\boldsymbol{\phi}_1(\boldsymbol{\sigma}_{\mathbf{c}}) - \int_{t_0}^{t} k_2(\mathbf{x},\hat{\eta}_e,t)\boldsymbol{\phi}_2(\boldsymbol{\sigma}_{\mathbf{c}})dt, \tag{33}$$

where $\boldsymbol{\phi}_1(\boldsymbol{\sigma}_{\mathbf{c}})$ and $\boldsymbol{\phi}_2(\boldsymbol{\sigma}_{\mathbf{c}})$ are defined as in (19), with $k_3 = 1$ and the variable gains $k_1(\mathbf{x},\hat{\eta}_e,t)$ and $k_2(\mathbf{x},\hat{\eta}_e,t)$ as in (23), with the arbitrary constants tuned as $\delta = \beta = 0.5$ and $\varepsilon = 0.25$. For comparison purposes, the controller for Method 2, $\bar{\mathbf{u}}(\boldsymbol{\sigma}_{\mathbf{c}},t) = -\eta_c\frac{\boldsymbol{\sigma}_{\mathbf{c}}}{||\boldsymbol{\sigma}_{\mathbf{c}}||}$ was designed with $\eta_c = 1$. The results are reported in the following Figures.

From Fig. 2, it is noted that the FOSM-based controller lead to a discontinuous control and is more prone to present the chattering phenomena, restricting their application in physical systems. In turn, note that the control signal is continuous using a High Order Sliding Mode technique, keeping important features of FOSM, as disturbance rejection, finite time convergence and avoiding the appearance of chattering.

Finally, it can be observed from Fig. 3 that although the controller of Method 2 ensures that $\epsilon_t(t)$ converges to zero

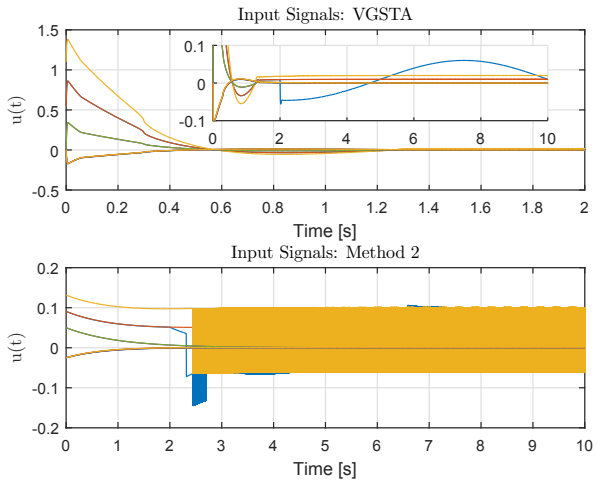

Fig. 2. Control signal components of $\mathbf{u}(t)$.

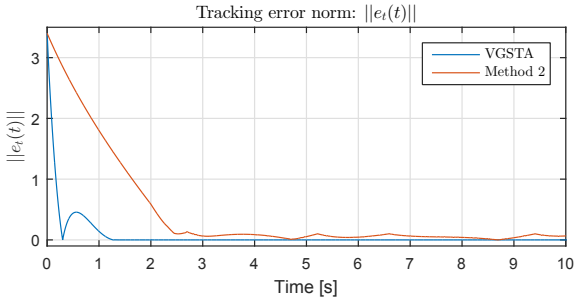

Fig. 3. Tracking error norm.

in finite time, it follows that $e_t(t)$ may not converge to the origin, since Method 2 only ensures the boundedness of $\mathbf{e}_2(t)$. In turn, the technique proposed here ensures finite time convergence for both errors. Also note that proposed monitor can be combined with a FOSM controller to achieve a smoother transient at cost of a discontinuous control law.

3. CONCLUSION

In this paper, a strategy for the state/sensor attack monitor design is proposed, which allows that the attacks to be reconstructed using a single multivariable estimator. Moreover, a sliding mode technique is used to guarantee finite time convergence of the sliding variables and exponential reconstruction of the attacks, in the absence of external disturbances. It should be emphasized that, by means of a First Order Approximation Filter, the knowledge of an upper bound for the non-measurable states is circumvented making the stability and convergence properties global.

REFERENCES

Ao, W., Song, Y., and Wen, C. (2016). Adaptive cyber-physical system attack detection and reconstruction with application to power systems. *IET Control Theory & Applications*, 10(12), 1458–1468.

Conti, J.P. (2010). The day the samba stopped [power blackouts]. *Engineering & Technology*, 5(4), 46–47.

Corradini, M.L. and Cristofaro, A. (2017). Robust detection and reconstruction of state and sensor attacks for cyber-physical systems using sliding modes. *IET Control Theory & Applications*, 11(11), 1756–1766.

Cunha, J.P.V.S., Costa, R.R., and Hsu, L. (2003). Design of first order approximation filters applied to sliding mode control. In *Proceedings of the IEEE Conference on Decision and Control*, volume 4, 3531–3536.

Dorfler, F. and Bullo, F. (2012). Kron reduction of graphs with applications to electrical networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 60(1), 150–163.

Filippov, A.F. (1964). Differential equations with discontinuous right-hand side. *Amer. Math. Soc. Trans.*, 42, 199–231.

Hsu, L., Lizarralde, F., and De Araujo, A.D. (1997). New results on output-feedback variable structure model-reference adaptive control: design and stability analysis. *IEEE Transactions on Automatic Control*, 42(3), 386–393.

Huang, X., Zhai, D., and Dong, J. (2018). Adaptive integral sliding-mode control strategy of data-driven cyber-physical systems against a class of actuator attacks. *IET Control Theory & Applications*, 12(10), 1440–1447.

Langner, R. (2011). Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, 9(3), 49–51.

Mei, S., Zhang, X., and Cao, M. (2011). *Power grid complexity*. Springer Science & Business Media.

Nateghi, S., Shtessel, Y., Barbot, J.P., Zheng, G., and Yu, L. (2018). Cyber-attack reconstruction via sliding mode differentiation and sparse recovery algorithm: Electrical power networks application. In *2018 15th International Workshop on Variable Structure Systems (VSS)*, 285–290. IEEE.

Pasqualetti, F., Dörfler, F., and Bullo, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE transactions on automatic control*, 58(11), 2715–2729.

Pasqualetti, F., Dorfler, F., and Bullo, F. (2015). Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems. *IEEE Control Systems Magazine*, 35(1), 110–127.

Poovendran, R., Sampigethaya, K., Gupta, S.K.S., Lee, I., Prasad, K.V., Corman, D., and Paunicka, J.L. (2011). Special issue on cyber-physical systems [scanning the issue]. *Proceedings of the IEEE*, 100(1), 6–12.

Sandberg, H., Amin, S., and Johansson, K.H. (2015). Cyberphysical security in networked control systems: An introduction to the issue. *IEEE Control Systems Magazine*, 35(1), 20–23.

Teixeira, A., Sandberg, H., and Johansson, K.H. (2010). Networked control systems under cyber attacks with applications to power networks. In *Proceedings of the 2010 American Control Conference*, 3690–3696. IEEE.

Vidal, P.V., Nunes, E.V., and Hsu, L. (2017). Output-feedback multivariable global variable gain super-twisting algorithm. *IEEE Transactions on Automatic Control*, 62(6), 2999–3005.

Zimmerman, R.D., Murillo-Sánchez, C.E., and Thomas, R.J. (2010). Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems*, 26(1), 12–19.