DOI: 10.48011/asba.v2i1.1729

Estratégias por Modos Deslizantes para Monitoramento e Compensação de Ciber-Ataques em Sistemas Ciber-Físicos ⋆

Jair Luiz de Azevedo Filho * **Eduardo Vieira Leão Nunes** * **Liu Hsu** *

* *COPPE, Universidade Federal do Rio de Janeiro, RJ, (e-mail: jair.azevedo@ufrj.br; eduardo@coep.ufrj.br;liu@coep.ufrj.br).*

Abstract: In this paper, the problem of monitoring and reconstruction of cyber-attacks in uncertain linear multivariable systems is addressed. The class of cyber-attacks considered in this work can corrupt the states or the outputs of the system. Using higher order sliding mode techniques, a multivariable attack monitor is proposed for attack detection and reconstruction. Furthermore, the use of First Order Approximation Filters ensure global stability and convergence results. The attack rejection problem in square plants is addressed with sliding mode techniques, ensuring finite time output tracking whilst rejecting the effects of the cyber-attack.

Resumo: Neste artigo, o problema de monitoramento e reconstrução de ciber-ataques em sistemas ciber-físicos lineares multivariáveis incertos é estudado. A classe de ataques cibernéticos considerada neste trabalho é capaz de corromper os estados ou as saídas da planta. Usando técnicas de modos deslizantes de ordem superior, um monitor de ataques multivariável é proposto para detecção e reconstrução dos ataques. Além disso, o uso de Filtros de Aproximação de Primeira Ordem garantem resultados globais de estabilidade e convergência. O problema de rejeição de ataques em plantas quadradas é abordado com o uso de técnicas de modo deslizante, garantindo rastreamento da saída em tempo finito enquanto rejeita os efeitos do ciber-ataque.

Keywords: Cyber-Physical Systems; sliding mode control; disturbance rejection; output feedback control; tracking.

Palavras-chaves: Sistemas Ciber-Físicos; controle por modos deslizantes; rejeição de distúrbios; controle por realimentação de saída; rastreamento.

1. INTRODUÇÃO

Encontrados nas principais infraestruturas de uma cidade, sistemas ciber-físicos (*Cyber-Physical Systems* (CPS)) são caracterizados por integrar processos físicos, meios de comunicação e recursos computacionais (Poovendran et al., 2011). A Figura 1 ilustra a estrutura de um sistema ciber-físico e algumas de suas aplicações. Embora CPS apresentem maior eficiência, eles se tornam mais suscetíveis à ataques envolvendo o domínio cibernético, conhecidos como Ciber-Ataques (Sandberg et al., 2015). Conforme apresentado em (Pasqualetti et al., 2013), prejuízos significativos podem ser causados por ciber-ataques sobre os canais de comunicação ou de dados, como por exemplo a falha no sistema de tratamento de esgoto (Slay and Miller, 2007), o acidente na planta nuclear de Taiwan (Lee et al., 2010), os apagões de energia no Brasil (Conti, 2010) e o vírus *Stuxnet* (Langner, 2011). Tais casos elucidaram como os mecanismos de segurança adotados precisavam ser complementados por estratégias de controle capazes de detectar e rejeitar esses ataques.

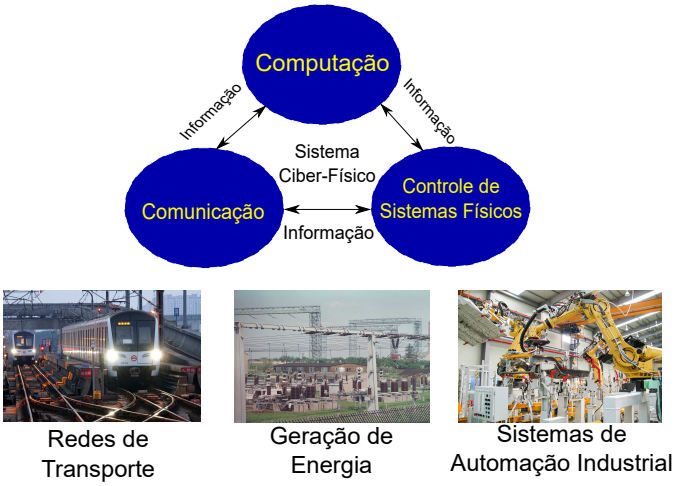

Figura 1. Sistema Ciber-Físico.

Essa crescente necessidade motivou o desenvolvimento de trabalhos recentes no contexto de detecção e reconstrução de ciber-ataques. Em (Nateghi et al., 2018), sistemas não-lineares são considerados. Usando diferenciadores baseados em modos deslizantes de ordem superior, os ciber-ataques são aproximados através de um problema de otimização.

⋆ Este trabalho foi financiado pelas seguintes agências de fomento à pesquisa: CAPES, CNPq e FAPERJ.

No entanto, assume-se que todos os estados do sistema estão disponíveis. Em (Huang et al., 2018), sistemas lineares incertos são considerados. Através de técnicas de modo deslizante integral e leis de adaptação, a estimação de um limitante superior para o ataque é obtida, garantindo desempenho quase-ótimo para o sistema com o controlador proposto. Entretanto, a estratégia citada considera que o sistema livre de perturbações externas e ataques estará disponível antes de sua implementação.

Baseado em (Ao et al., 2016), uma estratégia por modos deslizantes interessante é proposta em (Corradini and Cristofaro, 2017) para detecção, reconstrução e compensação de ataques aos estados e às saídas de sistemas ciber-físicos lineares incertos. Para tal, um monitor de ataques e um observador de estados são propostos, visando a garantir que os erros de estimação e monitoramento sejam limitados por constantes positivas, enquanto um compensador baseado em modos deslizantes de primeira ordem é projetado.

As seguintes restrições e hipóteses limitam a aplicação da estratégia proposta em (Corradini and Cristofaro, 2017):

- O monitor de ataques pode apresentar desempenho limitado, uma vez que não garante a convergência do erro de monitoramento para zero. Além disso, a estimação de ataque proposta não é definida quando o erro de estimação de saída se torna menor que uma constante positiva pré-definida.
- A detecção/reconstrução de ataques em cada canal de estado ou saída requer um monitor de ataques e um observador de estados. Portanto, o monitoramento de diversos canais se torna computacionalmente custoso.
- O esquema para compensação de ataques não garante a convergência do erro de rastreamento para zero.
- Assume-se que o ataque e sua derivada temporal possuem limitantes superiores conhecidos a priori.
- Garante-se somente resultados locais de estabilidade e convergência, uma vez que assume-se o conhecimento de um limitante superior para os estados não-medidos.

Inspirado pela estratégia apresentada em (Corradini and Cristofaro, 2017), esse trabalho propõe uma abordagem alternativa baseada em modos deslizantes a fim de contornar as limitações previamente discutidas. As seguintes características são asseguradas:

- O erro de estimação de saída converge para zero em tempo finito e a reconstrução de ataques é completamente definida. Além disso, na ausência de perturbações externas, o erro de monitoramento tende exponencialmente para zero.
- O vetor de ataque é reconstruído usando um único monitor de ataques e um estimador de estados.
- A compensação de ataques proposta garante que o erro de rastreamento converge para zero em tempo finito. Além disso, como se utiliza uma técnica baseada em modo deslizante de ordem superior, o efeito do *chattering* é atenuado.
- Limitantes superiores para o ataque não são mais necessários.
- Resultados globais de estabilidade e convergência são garantidos com o uso de um Filtro de Aproximação de Primeira Ordem para a estimação de um limitante superior para os estados não medidos.

Note que a estratégia aqui proposta garante resultados de reconstrução superiores com o uso de menos hipóteses quando comparado aos resultados obtidos em (Corradini and Cristofaro, 2017). Portanto, além da abordagem proposta nesse trabalho poder ser aplicada à uma classe mais ampla de sistemas ciber-físicos, são asseguradas performance superior para monitoramento e compensação e propriedades globais de convergência e estabilidade.

Preliminares: A norma euclideana de um vetor $\mathbf{y}$ e a norma induzida correspondente de uma matriz $\mathbf{A}$ são denotadas por $||\mathbf{y}||$ e $||\mathbf{A}||$, respectivamente. I_n é a matriz identidade do $\mathbb{R}^{n\times n}$. Aqui, a definição de Fillipov para a solução de equações diferenciais descontínuas é assumida (Filippov, 1964).

2. ESTRATÉGIAS POR MODOS DESLIZANTES APLICADAS AO MONITORAMENTO E COMPENSAÇÃO DE CIBER-ATAQUES

Esta seção apresenta as propriedades da classe de CPS, os ciber-ataques considerados nesse trabalho e o projeto dos monitores e do compensador propostos aqui.

2.1 Propriedades do Sistema

Considere o seguinte sistema ciber-físico:

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

onde $\boldsymbol{\zeta}(t) \in \mathbb{R}^n$ é o vetor de estados, $\mathbf{u}(t) \in \mathbb{R}^m$ é a entrada do sistema, e $\mathbf{y}(t) \in \mathbb{R}^p$ é a saída do sistema, com $p \geq m$. As matrizes $\bar{A}$, $\bar{B}_u$, $\bar{C}$ e $\bar{D}_u$ possuem dimensões compatíveis. O vetor $\bar{D}_d\mathbf{d}(\boldsymbol{\zeta},t)$ representa qualquer perturbação externa ou incerteza no modelo e os termos $\bar{B}_f\mathbf{f}(\boldsymbol{\zeta},t)$ e $\bar{D}_f\mathbf{f}(\boldsymbol{\zeta},t)$ descrevem os ataques aos estados e à saída, respectivamente. Assumindo que $\bar{B}_f$ e $\bar{D}_f$ são matrizes conhecidas, o monitor de ataques deve detectar e reconstruir o vetor de ataques $\mathbf{f}(\boldsymbol{\zeta},t) \in \mathbb{R}^q$. Como em (Corradini and Cristofaro, 2017), assume-se que o vetor de estados $\boldsymbol{\zeta}(t)$ inclui as variáveis físicas e cibernéticas, enquanto $\mathbf{u}(t)$ e $\mathbf{y}(t)$ são sinais conhecidos.

Uma vez que o sistema é conhecido, exceto pelo vetor de perturbações externas/incertezas no modelo, os ataques aos estados e aos sensores podem ser detectados e reconstruídos com o auxílio de um estimador de estados de Luemberger. Para tal, a classe de CPS considerada deve satisfazer as seguintes hipóteses, baseadas em (Ao et al., 2016):

Hipótese 1:

1. O par $(\bar{A},\bar{C})$ é observável.
2. $rank(\bar{C}\bar{B}_f) = rank(\bar{B}_f)$, onde $\bar{B}_f$ possui posto completo por colunas.
3. As matrizes $\bar{B}_u$ e $\bar{D}_f$ possuem posto completo por colunas.
4. Os zeros invariantes de $(\bar{A},\bar{B}_u,\bar{C},\bar{D}_u)$ são estáveis
5. Assume-se que todos os ataques são detectáveis, isto é, o sistema $(\bar{A},\bar{B}_f,\bar{C},\bar{D}_f)$ não possui zeros invariantes.

A Hipótese 1.1 é razoável visto que é proposto o uso de um estimador de estados no esquema de monitoramento.

A Hipótese 1.2 implica que $q \leq p$. Logo, até p ataques podem ser detectados/reconstruídos simultaneamente. A Hipótese 1.3 é considerada pois não há sentido em usar entradas desnecessárias em $\bar{B}_u$, assim como $\bar{B}_f$ e $\bar{D}_f$ que possuem posto completo, visto que serão usadas no projeto do estimador. Note ainda que a classe de sistemas estudada é de fase mínima, visto que a Hipótese 1.4 é satisfeita. Por fim, a Hipótese 1.5 garante que não ocorre ataque não-detectável, quer seja estável ou instável.

Observe que a escolha de $\bar{B}_f$ e $\bar{D}_f$ define em quais canais a detecção aos ataques irá ocorrer. Em (Corradini and Cristofaro, 2017), essas matrizes são vetores coluna, implicando que o ataque monitorado $f(t)$ é um escalar. Logo, a detecção de p ataques simultâneos requer que p estimadores sejam desenvolvidos, tornando-se custoso do ponto de vista computacional. Aqui, considera-se que as matrizes $\bar{B}_f \in \mathbb{R}^{n\times p}$ e $\bar{D}_f \in \mathbb{R}^{p\times p}$, isto é, o monitor de ataques reconstrói um vetor de ataques, com p canais.

Observação *1.* Note que se $q < p$, os $p - q$ canais do vetor $\boldsymbol{f}(\boldsymbol{\zeta}, t)$ livres de ataques serão reconstruídos como zero, visto que $\bar{B}_f$ e $\bar{D}_f$ possuem posto completo por colunas.

A Hipótese 1 implica que há uma mudança linear de coordenadas $\mathbf{x} = T\boldsymbol{\zeta}$, tal que

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

com

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

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

onde $B \in \mathbb{R}^{p\times p}$ é uma matriz não-singular e $A_{11} \in \mathbb{R}^{n-p\times n-p}$ é uma matriz Hurwitz.

2.2 Classe de ataques

O modelo de ataques considerados nesse artigo são os *Deception Attacks* (Teixeira et al., 2010) e os *Stealth Attacks* (Ao et al., 2016).

Deception Attacks são capazes de afetar os estados do sistema, podendo alterar o comportamento do mesmo em regime permanente e sua estabilidade, como apresentado em (Teixeira et al., 2010; Ao et al., 2016). *Stealth Attacks* corrompem as medidas do sistema, podendo ser usados para afetar controladores por realimentação de saída (afetando indiretamente os estados do sistema) ou para ocultar *Deception Attacks*. Quando estes ataques são usados juntos, são denominados como Ataques Coordenados (*Coordinated Attacks*).

Baseado nos argumentos apresentados, nota-se que a detecção de ataques coordenados pode se tornar impraticável (Pasqualetti et al., 2015). Como a Hipótese 1.5 é satisfeita, uma condição suficiente é dada pela Hipótese 2:

Hipótese 2: Ataques coordenados não ocorrem, isto é, $||\bar{B}_f|| \cdot ||\bar{D}_f|| = 0$.

Note que esta hipótese implica que B_f ou D_f é uma matriz de zeros. Portanto, o vetor de ataques $\mathbf{f}(\mathbf{x}, t)$ em (1) não ocorre simultaneamente como ataque aos estados e a saída. Além disso, a seguinte Hipótese é satisfeita:

Hipótese 3: A derivada temporal do vetor de ataques possui um limitante superior conhecido:

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

onde $\rho_2(\mathbf{x}, t)$ é uma função positiva.

2.3 Detecção e reconstrução de ataques aos estados

Note que de acordo com a Hipótese 2, $D_f = 0$. Reescrevendo (2), tem-se que

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

Visto que $\mathbf{y}(t)$ e $D_u\mathbf{u}(t)$ são sinais conhecidos, considere o estimador a seguir:

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

Definindo os erros de estimação como $\mathbf{e}_1(t) = \mathbf{x}_1(t) - \hat{\mathbf{x}}_1(t)$ e $\mathbf{e}_2(t) = \mathbf{x}_2(t) - \hat{\mathbf{x}}_2(t)$, segue que:

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

onde $\mathbf{e}_2(t)$ é um sinal mensurável, uma vez que $\mathbf{y}(t) - \hat{\mathbf{y}}(t) = \mathbf{e}_2(t)$. Logo, o erro de estimação $\mathbf{e}_2(t)$ pode ser aplicado ao esquema de monitoramento. Como A_{11} é uma matriz Hurwitz, $\mathbf{e}_1(t)$ pode ser interpretado como um filtro estável para o sinal $\mathbf{d}(\mathbf{x}, t)$. Adicionalmente, assume-se que a seguinte hipótese é satisfeita:

Hipótese 4: O vetor de perturbação/incerteza e sua derivada temporal possuem limitantes superiores conhecidos:

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

onde $\rho_d(\mathbf{x}, t)$ e $\bar{\rho}_d(\mathbf{x}, t)$ são funções limitadas conhecidas.

Baseado nos resultados acima, sabe-se que para qualquer condição inicial limitada de $\mathbf{e}_1(0)$, $\mathbf{e}_1(t)$ permanece limitado para todo $t \geq 0$, visto que:

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

onde α_1, α_2, $||\mathbf{e}_1(0)||$, e ρ são constantes positivas limitadas. Em (Corradini and Cristofaro, 2017), considera-se que o limitante superior ρ é conhecido, embora $||\mathbf{e}_1(0)||$ não esteja disponível. Nesse trabalho, assume-se que tanto $||\mathbf{e}_1(0)||$ como ρ são *desconhecidos*.

Observação *2.* Nota-se por (8) que, embora a convergência exponencial de $\mathbf{e}_1(t)$ seja impedida por $\mathbf{d}(\mathbf{x}, t)$, sua limitação é garantida pela Hipótese 4.

Definindo $\hat{\mathbf{f}}(\mathbf{x}, t) = -B^{-1}\bar{\mathbf{f}}(\mathbf{x}, t)$, o erro de estimação $\mathbf{e}_2$ em (8) pode ser escrito como

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

com

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

Segue de (12) e (8), que:

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

De (13), um limitante superior para $||\dot{\boldsymbol{\Delta}}||$ é dado por:

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

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

Conforme apresentado em (10), o conhecimento de um limitante superior para $||\mathbf{e}_1(t)||$ não é considerado, pois essa hipótese tornaria os resultados de estabilidade e convergência locais, visto que $||\mathbf{e}_1(0)|| \leq \mu$, onde $\mu \in \mathbb{R}$. Visando a contornar essa restrição, nesse trabalho propõe-se a aplicação de um Filtro de Aproximação de Primeira Ordem (*First Order Aproximation Filter* (FOAF)) (Cunha et al., 2003; Hsu et al., 1997):

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

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

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

para algum $c_\eta, \lambda_\eta > 0$ e $\mathbf{e}_\eta(t) = [\mathbf{e}_1^T \ \hat{\eta}_e]^T$. Logo, pode-se mostrar que após um certo tempo finito t_1, $\hat{\eta}_e(t) > ||\mathbf{e}_1(t)||, \forall t > t_1$.

Baseado em (8), os parâmetros em (15) podem ser definidos como:

$$\lambda_f = \min_j\{-Re(\lambda_j)\} \cdot \gamma_1 \qquad c_f = ||D_1|| + \gamma_2$$

onde $\{\lambda_j\}$ são os autovalores de A_{11} e $0 < \gamma_1 < 1$ e $\gamma_2 > 0$ são constantes de projeto.

Portanto, de acordo com (15) e as Hipóteses 3 e 4, após determinado tempo finito, um limitante superior para $||\dot{\boldsymbol{\Delta}}(t)||$ é dado por:

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

onde $\boldsymbol{\Delta}^\star(t)$ é um sinal conhecido, com a_1, a_2, a_3, e a_4 definidos em (14).

Finalmente, note que (11) é um sistema de primeira ordem incerto cuja a derivada temporal do distúrbio possui um limitante superior conhecido. Para esta classe de sistemas, existem controladores capazes de rejeitar o termo de perturbação enquanto garantem um determinado desempenho pré-estabelecido para o sistema. Aqui, um método popular na comunidade de controle por modos deslizantes será adotado, o Algoritmo Super-Twisting com Ganhos Variáveis (*Variable Gain Super-Twisting Algorithm* (VGSTA)).

2.4 Reconstrução de Ataques aos Estados com o Algoritmo Super-Twisting com Ganhos Variáveis Multivariável

O algoritmo Super-Twisting é uma técnica de modo deslizante de segunda ordem que garante convergência em tempo finito da variável de deslizamento e sua derivada temporal. Esse algoritmo ganhou destaque por não depender da derivada temporal da variável de deslizamento para sua implementação. Além disso, por ser uma técnica baseada em modos deslizantes de ordem superior, esse algoritmo pode atenuar o efeito do *chattering*, presente em técnicas baseadas em modo deslizante de primeira ordem.

Nesse trabalho, o Algoritmo Super-Twisting com Ganhos Variáveis (VGSTA) para sistemas MIMO proposto em (Vidal et al., 2017) é considerado. O algoritmo é dado por:

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

onde

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

Baseado na lei de controle proposta em (18) e o termo de perturbação (12), a dinâmica do sistema (8) em malha fechada é dada por:

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

onde os termos $|\pi_1(t)|\boldsymbol{\phi}_1(\mathbf{e}_2)$ e $|\pi_2(t)|\boldsymbol{\phi}_2(\mathbf{e}_2)$ são usados com intuito de representar a presença do FOAF no sistema em malha fechada, com $|\pi_1(t)|$ e $|\pi_2(t)|$ definidos de modo semelhante à (16).

Note que, para $\mathbf{e}_2(t) \neq \mathbf{0}$, tem-se que

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

onde, de (17), $\varrho_2(\mathbf{e}_1,\mathbf{x},t)$ pode ser definido como $2\boldsymbol{\Delta}^\star(t)$. Além disso, note de (20), que:

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

O resultado principal da estratégia considerada é apresentado a seguir.

Teorema *1.* Considere que as Hipóteses (1)-(4) são satisfeitas. Além disso, considere o sistema em malha fechada (20) com o vetor de estados completo $\boldsymbol{X}(t) = \left[\hat{\eta}(t) \ \boldsymbol{e}_1^T(t) \ \boldsymbol{e}_2^T(t) \ \boldsymbol{z}^T(t)\right]^T$. Então, para qualquer condição inicial $\boldsymbol{X}(t_0) = \left[\hat{\eta}(t_0) \ \boldsymbol{e}_1^T(t_0) \ \boldsymbol{e}_2^T(t_0) \ \boldsymbol{z}^T(t_0)\right]^T$, as trajetórias do sistema são globalmente uniformemente limitadas e a superfície de deslizamento $\boldsymbol{e}_2 = \dot{\boldsymbol{e}}_2 = \mathbf{0}$ é alcançada em tempo finito se os ganhos variáveis são definidos como

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

onde δ, β, e ε são constantes positivas arbitrárias.

Prova. Ver Teorema 1, em (Vidal et al., 2017).

Observe que quando o modo deslizante é alcançado, verifica-se por (20) que $\mathbf{z}(t) = 0$. Logo, após um tempo finito, segue de (11) e (12) que $\hat{\mathbf{f}}(\mathbf{x},t) = B^{-1}\boldsymbol{\Delta}(\mathbf{e}_1,\mathbf{x},t)$. Além disso, por (12), tem-se que:

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

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

De acordo com (10) e a Hipótese 4, $\mathbf{d}(\mathbf{x},t)$ e $\mathbf{e}_1(t)$ são sinais limitados. Logo, pode-se mostrar que o erro de monitoramento $\left|\left|\hat{\mathbf{f}}(\mathbf{x},t) - \mathbf{f}(\mathbf{x},t)\right|\right|$ é limitado. Observe ainda que o sinal reconstruído pelo monitor de ataques é uma combinação do ataque real, do erro de estimação dos estados não medidos e da perturbação externa. No entanto, se $\mathbf{d}(\mathbf{x},t) \equiv 0$, conclui-se por (8) que $\hat{\mathbf{f}}(\mathbf{x},t)$ converge exponencialmente para $\mathbf{f}(\mathbf{x},t)$. Portanto, o vetor de ataques $\mathbf{f}(\mathbf{x},t)$ é reconstruído exponencialmente.

É válido frisar que a maior parte dos ciber-ataques e cenários de falhas existentes podem ser modelados como entradas aditivas desconhecidas que afetam os estados e as medidas. Apesar de refletirem falhas genuínas de componentes do sistema, tais distúrbios modelam os efeitos de ataques em sistemas ciber-físicos (Pasqualetti et al., 2015). Logo, a abordagem proposta pode ser aplicada em sistemas ciber-físicos e em sistemas tolerantes a falhas.

2.5 Reconstrução de Ataques aos Estados com o Algoritmo de Camada Pré-Definida

Para fins de comparação, o monitor de ataques proposto em (Corradini and Cristofaro, 2017) é apresentado a seguir. Esse método impõe a seguinte desigualdade:

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

Para tal, o seguinte monitor de ataques é proposto:

$$\dot{\hat{f}}(t) = \frac{\gamma k\left(\alpha(\mathbf{x},t)||\mathbf{e}_2|| + \bar{\beta}(\mathbf{x},t)^2 + \kappa(\mathbf{x},t) + \eta\right)}{k\mathbf{e}_2^T B - \epsilon sign(\mathbf{e}_2^T B)}, \text{ se } |\mathbf{e}_2^T B| > \epsilon$$

$$\alpha(\mathbf{x},t) := a_1\rho + a_2\rho_d(t) + a_3\rho_2(\mathbf{x},t) + a_4\bar{\rho}_d(t)$$

$$\bar{\beta}(\mathbf{x},t) := ||A_{21}||\rho + ||B||\rho_1(\mathbf{x},t) + ||D_2||\rho_d(t)$$

$$\kappa(\mathbf{x},t) := ||B||\left(||B||\hat{f}(t)^2 + 2\beta(\mathbf{x},t)|\hat{f}(t)|\right)$$

onde γ, k, ϵ e η são constantes positivas, $\rho_1(\mathbf{x},t)$ é um limitante superior para $||\mathbf{f}(\mathbf{x},t)||$ e ρ, $\rho_2(\mathbf{x},t)$, $\rho_d(t)$, e $\bar{\rho}_d(t)$ são dados por (10), (5), e (9), respectivamente.

A condição imposta em (25) garante que a norma de $\mathbf{e}_2(t)$ decresça até que alcance, em tempo finito, a camada $||\mathbf{e}_2(t)|| < \epsilon$. De agora em diante nesse artigo, esse método será referenciado como *Método 2*. É válido frisar que os autores da técnica apresentada não especificaram $\hat{f}(\mathbf{x},t)$ dentro da camada pré-definida. Aqui, assume-se que $\hat{f}(\mathbf{x},t)$ se mantém constante dentro da camada em questão. Deve-se ressaltar que os resultados de simulação em (Corradini and Cristofaro, 2017) e os apresentados aqui para o Método 2 são similares.

2.6 Exemplo de monitoramento de ataques aos estados: Sistema de Potência WECC

Considere o sistema de potência US WECC apresentado em (Pasqualetti et al., 2015). Após a redução de Kron (Dorfler and Bullo, 2012), um modelo simplificado é obtido em (Corradini and Cristofaro, 2017):

$$\dot{\boldsymbol{\theta}}(t) = \boldsymbol{\omega}(t), \quad \dot{\boldsymbol{\omega}}(t) = \mathbf{L}_s\boldsymbol{\theta}(t) - \boldsymbol{M}_g^{-1}\boldsymbol{D}_g\boldsymbol{\omega}(t) - \boldsymbol{P}_s$$

onde $\boldsymbol{\theta}$, $\boldsymbol{\omega} \in \mathbb{R}^3$ denotam os ângulos e as frequências dos geradores respectivamente, enquanto os termos $\boldsymbol{L}_s = \boldsymbol{M}_g^{-1}\left(\boldsymbol{L}_{gl}\boldsymbol{L}_{ll}^{-1}\boldsymbol{L}_{lg} - \boldsymbol{L}_{gg}\right) \in \mathbb{R}^{3\times3}$ são parâmetros do sistema e $\boldsymbol{P}_s = \boldsymbol{M}_g^{-1}\left(\boldsymbol{L}_{gl}\boldsymbol{L}_{ll}^{-1}\boldsymbol{P}_{\boldsymbol{\theta}} + \boldsymbol{P}_{\boldsymbol{\omega}}\right) \in \mathbb{R}^3$ está relacionado a potência de entrada controlada.

Para fins de comparação entre o monitor proposto e o Método 2, os valores numéricos adotados em (Corradini and Cristofaro, 2017) serão considerados. A matriz de saída é dada por

$$\bar{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Como $p = 2$, até dois ataques podem ser reconstruídos. Portanto, assim como em (Corradini and Cristofaro, 2017), o primeiro e o quarto canal serão monitorados, isto é, $\bar{B}_f = \bar{C}^T$. Após uma transformação linear de coordenadas $\boldsymbol{x} = \boldsymbol{T}\boldsymbol{\zeta}$, o sistema pode ser reescrito em sua forma normal (2). A condição inicial da planta e do estimador são:

$$\boldsymbol{x}(0) = [0.2\ 0.2\ 0\ 0\ 0\ 0] \quad \hat{\boldsymbol{x}}(0) = [0.1\ 0.1\ 0\ 0\ 0\ 0.1]$$

Considera-se que dois ataques senoidais da forma $f(t) = 0.5sen(t)$ corrompem o primeiro e o quarto canal em $t = 6s$ e $t = 7s$, respectivamente. Como apresentado na Subseção 2.5, os limitantes superiores exigidos pelo método 2 foram definidos como $\rho = \rho_1 = \rho_2 = 1$. Note que a estratégia aqui proposta requer apenas o conhecimento de um limitante superior para $||\dot{\boldsymbol{f}}(\boldsymbol{x},t)||$, tornando-se menos restritiva.

Os parâmetros usados em (23) foram sintonizados como $\delta = 1$, $\varepsilon = 0.1$ e $\beta = 5$, enquanto os parâmetros do Método 2 foram definidos como $k = 1.5$, $\gamma = 1.1$, $\eta = 0.1$ e $\epsilon = 0.02$. Os resultados de simulação são apresentados a seguir.

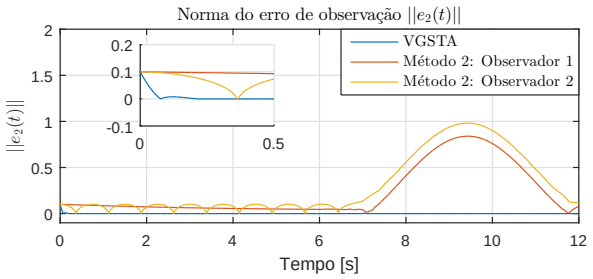

Figura 2. Norma do erro de estimação $||\mathbf{e}_2(t)||$.

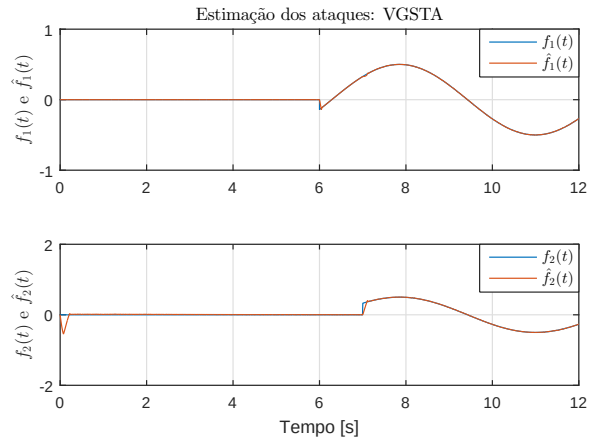

Figura 3. Reconstrução dos ataques via VGSTA.

Como pode-se observar pelas Figuras 2 e 4, o Método 2 garante somente a limitação dos erros de estimação de saída e de monitoramento. Por sua vez, verifica-se pela Figura 3 que a técnica aqui proposta garante a convergência em tempo finito de $\boldsymbol{e}_2(t)$ e a reconstrução exponencial dos ataques. Além disso, a estratégia proposta requer apenas um monitor multivariável e um estimador de estados, enquanto o Método 2 requer um para cada canal monitorado, tornando-se computacionalmente custoso.

2.7 Detecção e Reconstrução de Ataques à Saída

De acordo com a Hipótese 2, sabe-se que durante um ataque aos sensores, não ocorrem ataques aos estados, isto é, $B_f = 0$. Reescrevendo (1), segue que:

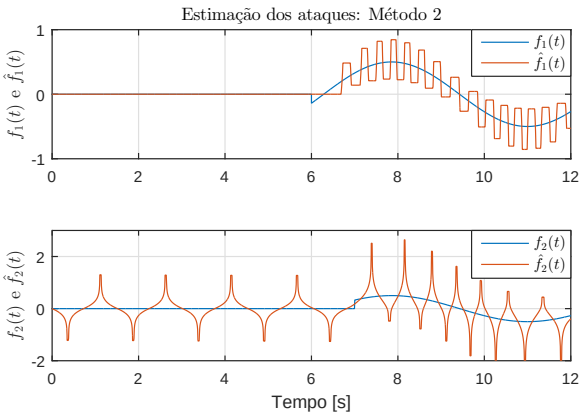

Figura 4. Reconstrução dos ataques via Método 2.

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

Considere agora o filtro passa-baixas proposto em (Ao et al., 2016), que será aplicado à saída de (26):

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

onde $A_f \in \mathbb{R}^{p\times p}$ é uma matriz de projeto, definida de modo que seja Hurwitz. Note que com a inclusão do "novo estado" $\mathbf{x}_f(t)$, o seguinte sistema aumentado é obtido:

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

onde $\boldsymbol{\omega}_1(t) = \boldsymbol{\zeta}(t)$ e $\boldsymbol{\omega}_2(t) = \mathbf{x}_f(t)$. Para esse sistema aumentado o seguinte observador é proposto:

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

Definindo os erros de estimação como $\mathbf{e}_{\omega_1}(t) = \boldsymbol{\omega}_1(t) - \hat{\boldsymbol{\omega}}_1(t)$ e $\mathbf{e}_{\omega_2}(t) = \boldsymbol{\omega}_2(t) - \hat{\boldsymbol{\omega}}_2(t)$, a dinâmica dos erros pode ser escrita como:

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

onde $\mathbf{e}_{\omega_2}$ é um erro mensurável, uma vez que $\mathbf{y}_f(t) - \hat{\mathbf{y}}_f(t) = \mathbf{e}_{\omega_2}(t)$. Comparando (30) à (8) e como pela Hipótese 4, $\mathbf{d}(\mathbf{x},t)$ é um sinal limitado, a limitação de $\mathbf{e}_{\omega_1}(t)$ depende de $\bar{A}$. Portanto, para ataques à saída, a seguinte Hipótese é considerada.

Hipótese 5*:* A matriz $\bar{A}$, apresentada em (1), é Hurwitz.

Observação *3.* Note que com esta hipótese, uma dedução similar à feita em (10) assegura que $\mathbf{e}_{\omega_1}(t)$ é um sinal limitado. Além disso, se $\mathbf{d}(\mathbf{x},t) \equiv \mathbf{0}$, então $\mathbf{e}_{\omega_1}(t)$ converge para zero exponencialmente.

Comparando as dinâmicas de $\mathbf{e}_{\omega_2}(t)$ em (30) e $\mathbf{e}_2(t)$ em (11), nota-se que a aplicação do VGSTA apresentado nesse trabalho para ataques à saída é direta. O esquema para reconstrução de ataques é apresentado na Figura 5

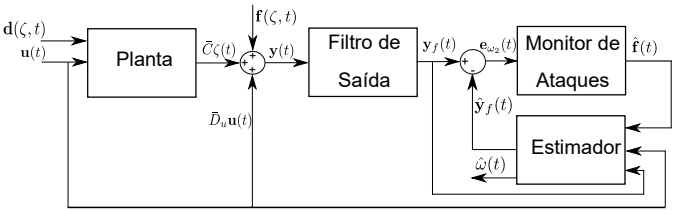

Figura 5. Diagrama de blocos para reconstrução de ataques à saída.

2.8 Compensação de Ataques aos Estados

Após a detecção e reconstrução do ataque, sua estimação pode ser usada para rejeitar seu efeito sobre o desempenho do sistema e garantir que a saída do sistema convirja para algum trajetória $\mathbf{y}_d(t)$ desejada. Logo, o objetivo é garantir que o erro de rastreamento $\boldsymbol{e}_t(t) = \boldsymbol{y}(t) - \ \boldsymbol{y}_d(t)$ se torne nulo após certo tempo finito.

Note que a estratégia baseada em modo deslizante apresentada aqui pode ser aplicada para sistemas com, no mínimo, grau relativo unitário. Portanto, assim como em (Corradini and Cristofaro, 2017), aqui será considerado que $D_u = 0$.

Observação *4.* É válido frisar que essa condição é necessária exclusivamente para fins de compensação de ataques aos estados. O monitoramento dos mesmos não requer essa condição.

Definindo $\boldsymbol{\epsilon}_t(t) = \hat{\mathbf{y}}(t) - \mathbf{y}_d(t)$, note que o erro de rastreamento pode ser escrito como

$$\boldsymbol{e}_t(t) = \boldsymbol{e}_2(t) + \boldsymbol{\epsilon}_t(t) \tag{31}$$

onde, por (7), sabe-se que

$$\dot{\boldsymbol{\epsilon}}_t(t) = A_{21}\hat{\mathbf{x}}_1(t) + B_2\mathbf{u}(t) + B\hat{\mathbf{f}}(\mathbf{x},t) + A_{22}\mathbf{y}(t) - \dot{\mathbf{y}}_d(t)$$

Considere agora a seguinte variável de deslizamento

$$\boldsymbol{\sigma}_c(t) = G\boldsymbol{\epsilon}_t(t), \tag{32}$$

onde $G \in \mathbb{R}^{m\times p}$ é uma matriz a ser definida tal que GB_2 seja uma matriz quadrada de posto completo. Observe que para o caso quadrado ($m = p$), se G possuir posto completo, $\boldsymbol{\sigma}_c(t) = \mathbf{0}$ implica que $\boldsymbol{\epsilon}_t(t) = \mathbf{0}$. A dinâmica de $\boldsymbol{\sigma}_c(t)$ é dada por

$$\dot{\boldsymbol{\sigma}}_c(t) = G(\boldsymbol{\Psi}(t) + B_2\mathbf{u}(t))$$

onde $\boldsymbol{\Psi}(t) = A_{21}\hat{x}_1(t) + B\hat{f}(x,t) + A_{22}y(t) - \dot{y}_d(t)$.

Definindo $u(t) = (GB_2)^{-1}(\bar{u} - G\Psi(t))$, a dinâmica de $\dot{\boldsymbol{\sigma}}_c$ em malha fechada pode ser escrita como

$$\dot{\boldsymbol{\sigma}}_c(t) = \bar{u}(t) \tag{33}$$

Comparando (33) e (11), verifica-se que a convergência da variável de deslizamento $\boldsymbol{\sigma}_c(t)$ para zero é garantida em tempo finito com o uso do VGSTA apresentado nesse artigo. Note ainda que, conforme já citado, em sistemas quadrados, a convergência da variável de deslizamento garante a convergência do erro de rastreamento para zero em tempo finito, desde que a matriz G seja não-singular. O esquema de compensação para ataques aos estados é apresentado na Figura 6.

2.9 Exemplo de compensação de ataques aos estados: Sistema de Alimentação de Barramentos IEEE 39

Considere o sistema de energia de barramento IEEE 39 com dez geradores e 39 barramentos descrito em (Mei

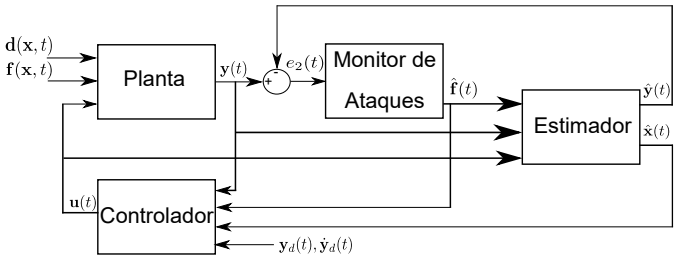

Figura 6. Diagrama de blocos para reconstrução e compensação de ataques aos estados.

et al., 2011; Zimmerman et al., 2010). Usando a técnica de redução de Kron (Dorfler and Bullo, 2012), esse sistema pode ser escrito como

$$\begin{bmatrix}\dot{\boldsymbol{\theta}}(t)\\ \dot{\boldsymbol{\omega}}(t)\end{bmatrix}=\begin{bmatrix}0 & I\\ \boldsymbol{L_s} & -\boldsymbol{M_g}^{-1}\boldsymbol{D_g}\end{bmatrix}\begin{bmatrix}\boldsymbol{\theta}(t)\\ \boldsymbol{\omega}(t)\end{bmatrix}+\begin{bmatrix}0\\ \boldsymbol{I_{10}}\end{bmatrix}\boldsymbol{u}(t)+\bar{B}_f\boldsymbol{f}(\boldsymbol{\xi},t)$$
$$\dot{\boldsymbol{\xi}}(t)=\bar{\boldsymbol{A}}\boldsymbol{\xi}(t)+\bar{\boldsymbol{B}}\boldsymbol{u}(t)+\bar{\boldsymbol{B}}_f\boldsymbol{f}(\boldsymbol{\xi},t) \quad (34)$$

com $\boldsymbol{\xi}(t)=\left[\boldsymbol{\theta}(t)^T\ \boldsymbol{\omega}(t)^T\right]^T$, $\boldsymbol{L_s}=\boldsymbol{M_g}^{-1}\left(\boldsymbol{L_{gl}}\boldsymbol{L_{ll}}^{-1}\boldsymbol{L_{lg}}-\boldsymbol{L_{gg}}\right)$ e $\boldsymbol{u}(t)=\boldsymbol{M_g}^{-1}\left(\boldsymbol{P_\omega}(t)-\boldsymbol{L_{gl}}\boldsymbol{L_{ll}}^{-1}\boldsymbol{P_\theta}(t)\right)$, onde $\boldsymbol{\theta}(t)\in\mathbb{R}^{10}$ e $\boldsymbol{\omega(t)}\in\mathbb{R}^{10}$ denotam os ângulos e as frequências dos rotores dos geradores, respectivamente. O vetor $\boldsymbol{f}(t)$ representa um ataque desconhecido e $\boldsymbol{u}(t)$ a entrada equivalente. A resistência da linha é descrita pela matriz de susceptibilidade da rede, composta por $\boldsymbol{L_{ll}}$, $\boldsymbol{L_{gl}}$, $\boldsymbol{L_{lg}}$ e $\boldsymbol{L_{gg}}$. As matrizes $\boldsymbol{M_g}$ e $\boldsymbol{D_g}$ descrevem a matriz de inércia e os coeficientes de amortecimento, respectivamente. $\boldsymbol{P_\theta}(t)$ é a demanda conhecida de potência reativa dos barramentos e $\boldsymbol{P_\omega}(t)$ é a potência mecânica de entrada para cada gerador. Os valores numéricos para os parâmetros acima são os mesmos considerados em (Corradini and Cristofaro, 2017).

Para fins de simulação, considera-se que a saída do sistema é dada por $\boldsymbol{y}(t)=[\boldsymbol{I_{10}}\ \boldsymbol{I_{10}}]\,\boldsymbol{\xi}(t)$ Além disso, os dez últimos estados serão monitorados, isto é, $\bar{\boldsymbol{B}}_{\boldsymbol{f}}=[\boldsymbol{0}\ \boldsymbol{I_{10}}]^T$.

Assume-se que um vetor de ataques aos estados da forma $\boldsymbol{f}(t)=\left[\boldsymbol{0}_{1\times10}\ \ \frac{1}{2}sen(0.2\pi t)\ \ \boldsymbol{0}_{1\times9}\right]^T$ corrompe os estados do sistema a partir de $t=2s$. O objetivo da ação de controle será o rastreamento a seguinte saída de referência $\mathbf{y}_d=[1\ 1\ 2\ \mathbf{0}_{1\times7}]^T$. Observe que neste caso B_2 é uma matriz quadrada e de posto completo. Portanto, de acordo com (32), pode ser definida como $G=\boldsymbol{I_{10}}$.

Duas técnicas foram usadas para o monitoramento desse ataque: VGSTA e Método 2. Para essa simulação, os limitantes superiores necessários para a implementação do Método 2 são definidos como $\rho=2.5$, $\rho_1(\mathbf{x},t)=1.75$, e $\rho_2(\mathbf{x},t)=2$. É válido frisar que a implementação do VGSTA requer apenas o conhecimento de $\rho_2(\mathbf{x},t)$.

A condição inicial para o estimador e para a planta são, respectivamente, $\hat{x}(0)=\mathbf{0}_{20\times1}$ e

$$x(0)=[0.3\cdot\mathbf{1}_{1\times5}\ -0.6\cdot\mathbf{1}_{1\times5}\ -0.5\cdot\mathbf{1}_{1\times5}\ 0.25\cdot\mathbf{1}_{1\times5}]^T$$

Os parâmetros usados no VGSTA foram sintonizados como $\delta=1$, $\varepsilon=0.3$, e $\beta=0.1$, enquanto os parâmetros para o Método 2 foram extraídos de (Corradini and Cristofaro, 2017), a saber. $\kappa=1.1$, $\gamma=1.1$, $\eta=0.2$ e $\epsilon=0.1$. Os resultados são ilustrados a seguir.

Como pode-se observar na Figura 7, o modo deslizante é alcançado em tempo finito usando o VGSTA adotado

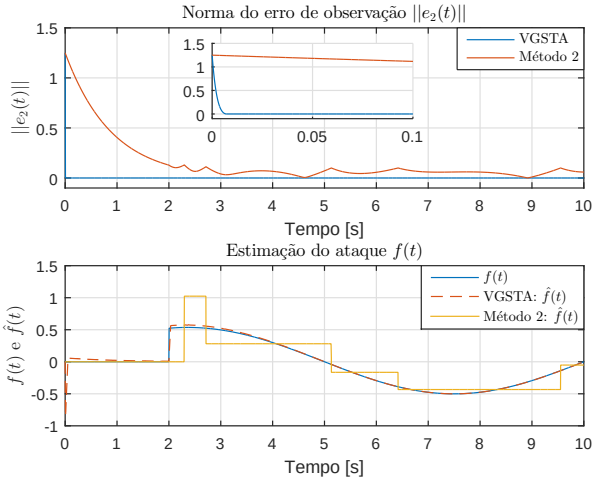

Figura 7. (a) Norma do erro de estimação $||\mathbf{e}_2(t)||$ e (b) Ataque aos estados real $f(t)$ versus estimado $\hat{f}(t)$.

neste trabalho. Além disso, quando o deslizamento ocorre, verifica-se a reconstrução exponencial do ataque. Por sua vez, a estimação do ataque obtida com o Método 2 garante somente a limitação no erro de monitoramento.

Embora a estratégia de compensação não seja dependente da reconstrução exata do ataque, a análise de seus perfis pode auxiliar em estratégias para detecção de ataques futuros. Nesse contexto, verifica-se outra contribuição desse trabalho, visto que o ataque é reconstruído exponencialmente, na ausência de distúrbios externos.

No contexto de compensação de ataques, nenhum sinal deve ser reconstruído, visto que o mesmo já foi estimado pelo monitor de ataques $\mathbf{f}(\mathbf{x},t)$. Portanto outras técnicas baseadas em modos deslizantes podem ser usadas em (33).

Para fins de comparação, o controlador proposto em (Corradini and Cristofaro, 2017), $\bar{\mathbf{u}}(\boldsymbol{\sigma}_\mathbf{c},t)=-\eta_c\frac{\boldsymbol{\sigma}_\mathbf{c}}{||\boldsymbol{\sigma}_\mathbf{c}||}$, será usado na compensação do ataque monitorado por ambas as estratégias, isto é, com o monitor proposto aqui e com o Método 2. Definindo $\eta_c=1$, os resultados relacionados a compensação do ataque são ilustrados a seguir.

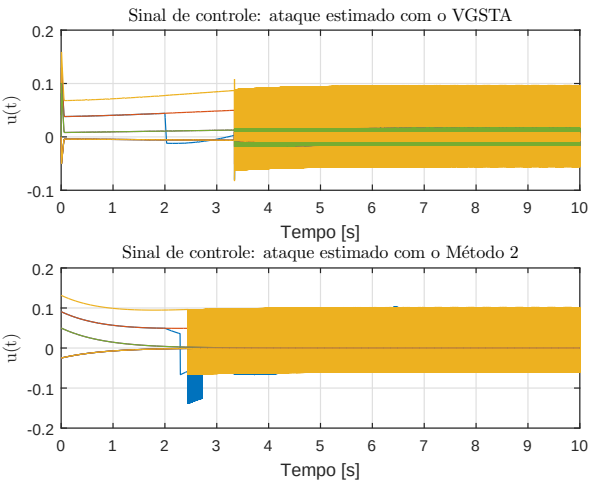

Figura 8. Sinal de controle $\mathbf{u}(t)$.

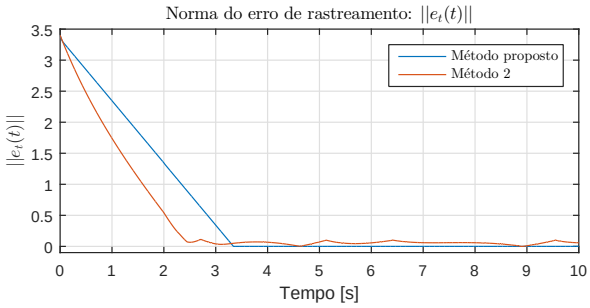

Figura 9. Norma do erro de rastreamento.

Pela Figura 8, nota-se que embora o mesmo controlador baseado em modos deslizantes de primeira ordem (*First Order Sliding Mode* (FOSM)) tenha sido aplicado em ambos os casos, há uma diferença entre os resultados obtidos com estratégias de monitoramento diferentes. Essa distinção se deve pois embora o mesmo $\bar{\mathbf{u}}(t)$ em (33) seja usado em ambos os casos, os termos cancelados em $\mathbf{u}(t)$ são diferentes de acordo a técnica de reconstrução utilizada.

Finalmente, observa-se pela Figura 9 que embora o controlador do Método 2 garanta que ϵ_t convirja para zero em tempo finito, o erro $\mathbf{e}_t(t)$ pode não convergir, visto que o Método 2 garante somente a limitação de $\mathbf{e}_2(t)$. Por sua vez, a técnica proposta aqui garante convergência em tempo finito de ambos os erros. Note ainda que o monitor proposto pode ser combinado com um controlador baseado em modos deslizantes de ordem superior (*Higher Order Sliding Modes* (HOSM)) como o próprio VGSTA, afim de se obter um lei de controle menos propensa ao *chattering* ao custo de um transitório com maior esforço de controle.

3. CONCLUSÃO

Nesse artigo, uma estratégia para o monitoramento de ataques aos estados e a saída foi proposta, permitindo que os ataques sejam reconstruídos com o uso de um único monitor de ataques multivariável. Além disso, uma técnica de modos deslizantes foi usada para garantir a convergência em tempo finito para as variáveis de deslizamento e reconstrução exponencial dos ataques, em casos livres de perturbações externas/incertezas no modelo. É válido frisar que, com o uso de um Filtro de Aproximação de Primeira Ordem, o conhecimento de um limitante superior para os estados não-medidos tornou-se desnecessário, tornando os resultados de estabilidade e convergência globais.

REFERÊNCIAS

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.

Lee, C.H., Chen, B.K., Chen, N.M., and Liu, C.W. (2010). Lessons learned from the blackout accident at a nuclear power plant in taiwan. *IEEE transactions on power delivery*, 25(4), 2726–2733.

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.

Slay, J. and Miller, M. (2007). Lessons learned from the maroochy water breach. In *International conference on critical infrastructure protection*, 73–82. Springer.

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.