Moving-horizon false data injection attack design against cyber–physical systems

Yu Zheng [*], Sridhar Babu Mudhangulla, Olugbenga Moses Anubi

Department of Electrical and Computer Engineering, Florida State University, Tallahassee, 32303, FL, USA

ARTICLE INFO

ABSTRACT

Systematic attack design is essential to understanding the vulnerabilities of cyber–physical systems (CPSs), to better design for resiliency. In particular, false data injection attacks (FDIAs) are well-known and have been shown to be capable of bypassing bad data detection (BDD) while causing targeted biases in resulting state estimates. However, their effectiveness against moving horizon estimators (MHE) is not well understood. In fact, this paper shows that conventional FDIAs are generally ineffective against MHE. One of the main reasons is that the moving window renders the static FDIA recursively infeasible. This paper proposes a new attack methodology, moving-horizon FDIA (MH-FDIA), by considering both the performance of historical attacks and the current system's status. Theoretical guarantees for successful attack generation and recursive feasibility are given. Numerical simulations on the IEEE-14 bus system further validate the theoretical claims and show that the proposed MH-FDIA outperforms state-of-the-art counterparts in both stealthiness and effectiveness. In addition, an experiment on a path-tracking control system of an autonomous vehicle shows the feasibility of the MH-FDIA in real-world nonlinear systems.

1. Introduction

The control of CPSs is generally a closed-loop feedback process, in which the physical processes, measured by distributed sensors, are driven by control actions depending on the accurate estimates of the state variables. However, the cyber layers are vulnerable to adversarial attacks, and the closed-loop interaction propagates the effect of the inevitable attacks on the physical processes. Recent examples, such as the malicious attacks on Israel's water supply system (Brentan et al., 2021), the 2015 Ukraine blackout (Liang, Weller, Zhao, Luo, & Dong, 2017) and the recent leakage of the Colonial Pipeline due to cyber-attacks (Reeder & Hall, 2021), indicate that cyberattacks can cause severe consequences for CPS stakeholders. Developing resilient CPSs in an adversarial setting has motivated researchers to study possible attack strategies, such as denial of service (DoS) (Xu, Ma, Trappe, & Zhang, 2006) and deception attacks (Pasqualetti, Dörfler, & Bullo, 2013).

False data injection attack (FDIA), most widely studied deception attack, follows a general attack strategy of maximizing the effectiveness on the system behaviors while maintaining stealthiness (Pasqualetti et al., 2013). Stealthiness measures the potential to bypass the bad data detection (BDD) (Hu, Wang, Han, & Liu, 2018), and effectiveness measures the closeness to the intended degradation of system performance (Zhang & Ye, 2020). Early researchers incorporated the full system model into the maximization program to generate a feasible attack (Hu et al., 2018; Liu, Ning, & Reiter, 2011; Mo & Sinopoli,

2010). This rendered the resulting process computationally inefficient and less pragmatic FDIAs. The authors in Liu et al. (2011) studied the FDIA generation problem against the least-square estimator (LSE) with a residual-based BDD. The feasibility of FDIA against the Kalman filter with χ^2 detector was studied in Mo and Sinopoli (2010). A sufficient and necessary condition for insecure estimation under FDIA was derived for the networked control system in Hu et al. (2018). To develop more pragmatic attack generation strategies, several constraints are incorporated to capture the attacker limitations such as limited access to sensors (Liu et al., 2011), incomplete knowledge of system dynamics (Liu, Bao, Lu, & Li, 2015), and incomplete knowledge of implemented state estimators (Lu & Yang, 2022). Data-driven approaches have also been used to generate FDIA in order to incorporate constrained knowledge of system dynamics. Examples include supervised learning using existing small attack dataset (Mohammadpour-fard, Ghanaatpishe, Mohammadi, Lakshminarayana, & Pechenizkiy, 2020), physics-guided unsupervised learning (Khazraei, Hallyburton, Gao, Wang, & Pajic, 2022; Zheng, Sayghe, & Anubi, 2021). Some recent results have streamlined their approach to generate feasible attack that maintains stealthiness and ignored the effectiveness maximization objectives, largely due to the inevitable nonconvexity which results in computationally ineffective FDIAs at best (Sui, Mo, Marelli, Sun, & Fu, 2020).

In this paper, we consider an additional perspective for the FDIA generation problem: How does the attack history affect the feasibility of the FDIA problem at the current time? In other words, is recursive feasibility essential for FDIA problem? Moving-horizon estimation (MHE) (Rawlings, Mayne, & Diehl, 2017) has been widely used in different control systems and is being deployed in CPSs with linear/nonlinear dynamics, networked and distributed architectures (Zou, Wang, Hu, & Han, 2020). One of the main design focuses of MHE is how to guarantee feasibility over the next window given the solution in the current window. This is called recursive feasibility (Löfberg, 2012; Muske, Rawlings, & Lee, 1993). From this angle, MHE is inherently more resilient than static counterparts since any successful FDIA against MHE must themselves be recursively feasible. In this paper, we refer to estimators designed with a window size of 1 as "static state estimators (SSE)", and the FDIA designed to target those estimators as "static FDIA". One of the main differences between SSE and MHE is in their consideration of recursive feasibility. We show that static FDIAs have a low chance of success against BDD built on MHE. Consequently, we propose a more general framework for moving-horizon FDIA (MH-FDIA) design with consideration of recursive feasibility.

Contribution: As a result of the ineffectiveness of static FDIAs against MHE, this paper studies a novel and more challenging, but more pragmatic FDIA problem against MHE. The paper discusses the limitations of various representative FDIA designs against MHE and proposes an adaptive mechanism for generating provably successful MH-FDIAs with consideration of recursive feasibility. The proposed solution solves the problem directly without any heuristic simplification or transformation. As a result, the proposed framework is general for any combination of MHE and BDD considerations. Theoretical results are validated based on an IEEE 14-bus system and a nonlinear path-tracking control system of a wheeled mobile robot. Notably, to the best knowledge of the authors, this paper is the first study of MH-FDIA with consideration of recursive feasibility.

The remainder of the paper is organized as follows. In Section 2, the notations employed are summarized. In Section 3, a concurrent system model of the CPS, including linear physical model, MHE, and BDD, is given for the subsequent development of the proposed FDIA scheme. In Section 4, an MH-FDIA generation scheme is given as a feasibility problem, and an adaptive algorithm is used to produce all feasible FDIAs. In Section 5, the state-of-the-art MH-FDIA generation algorithm is compared to a directed applied eigenvalue maximization approach. The comparison is done via numerical simulation on an IEEE 14-bus system and an experiment on a wheeled mobile robot. Conclusions follow in Section 6.

2. Preliminary

We use $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{m \times n}$ to denote Euclidean spaces of real scalars, n-dimensional column vectors, and $m \times n$ matrix respectively. Normal-face lower-case letters (*e.g.* $x \in \mathbb{R}$) are used to represent real scalars, bold-face lower-case letters (*e.g.* $\mathbf{x} \in \mathbb{R}^n$) represent vectors, while normal-face upper-case letters (*e.g.* $X \in \mathbb{R}^{m \times n}$) represent matrices. $X^\top$ and $X^\dagger$ denote the transpose and left pseudo-inverse of matrix X respectively (*i.e.* $X^\dagger X = I, (I - XX^\dagger)X = 0$). Let $X \in \mathbb{R}^{m \times n}$ ($m < n$) be a full-ranked matrix, $X^\perp$ denotes the orthogonal complement of X (i.e. $X^\top X^\perp = 0$ and $[X \ X^\perp]$ is full-ranked). Let $\mathcal{T} \subseteq \{1, \dots, m\}$ be a set of indices, then for a matrix $X \in \mathbb{R}^{m \times n}$, $X_\mathcal{T} \in \mathbb{R}^{|\mathcal{T}| \times n}$ is a sub-matrix with the rows of X corresponding to the indices in $\mathcal{T}$. We use $\rho(A)$ to denote the spectral radius of the matrix A.

We use $\mathbf{x}_i$ and $\mathbf{x}(i)$ to denote the vector $\mathbf{x}$ at time instance i and the ith elements of vector $\mathbf{x}$ respectively. $I_T \triangleq [i-T+1, i-T+2, \dots, i]$ denotes a moving-window of fixed size T at time instance i. The subscript T will be omitted if clear from the context. Accordingly, $I_T + 1$(or $I + 1$) $\triangleq [i-T+2, i-T+3, \dots, i+1]$ denotes the T time window from $i-T+2$ to $i+1$, and $I^- \triangleq [i-T+1, i-T+2, \dots, i-1]$ denotes a history window of size $T-1$. $\mathbf{x}_I = [\mathbf{x}_{i-T+1}^\top, \mathbf{x}_{i-T+2}^\top, \dots, \mathbf{x}_i^\top]^\top \in \mathbb{R}^{Tn}$ is a column

vector composed of vector $\mathbf{x}_j \in \mathbb{R}^n$ for all $j \in I$. If a continuous function $\phi : [0, a] \to [0, \infty)$ is monotonously increasing and satisfies $\phi(0) = 0$, then we say ϕ belongs to class κ. The complement of a set S is denoted by $\bar{S}$ (or S^c). The support of a vector $\mathbf{x} \in \mathbb{R}^n$ is defined as $\operatorname{supp}(\mathbf{x}) \triangleq \{i \subseteq \{1, \dots, n\} | \mathbf{x}(i) \neq 0\}$. $\Sigma_k \triangleq \{\mathbf{x} \in \mathbb{R}^n \| \operatorname{supp}(\mathbf{x})| \leq k\}$ denotes the set of k-sparse vectors. $\mathbf{x}_I \in \Sigma_k$ means all component vectors are k-sparse. Given a matrix $X \in \mathbb{R}^{m \times n}$ and a support $\mathcal{T}$, we say $X_\mathcal{T} \in \Sigma_k$ if each row of $X_\mathcal{T}$ belongs to Σ_k. We use the symbol $\otimes$ to denote the Kronecker product, i.e. for $B \in \mathbb{R}^{m \times n}$,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \otimes B = \begin{bmatrix} a_{11} B & a_{12} B \\ a_{21} B & a_{22} B \end{bmatrix} \in \mathbb{R}^{2m \times 2n}.$$

3. Problem formulation

Consider a nonlinear model used to describe the underlying physical processes of CPS

$$\begin{aligned} \dot{\mathbf{x}} &= f(\mathbf{x}, \mathbf{u}) \\ \mathbf{y} &= g(\mathbf{x}) + \mathbf{v}, \end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^m$ are the internal state variables, sensor measurements, and noise. f and g are assumed lipschitz in $\mathbf{x}$ and $\mathbf{u}$. Given an equilibrium point $(\mathbf{x}_0, \mathbf{u}_0)$ and a fixed time step[1] T_s, a discrete linear time-invariant (LTI) model is used to approximate the dynamics of the physical plant around the equilibrium point:

$$\begin{aligned} \Delta \mathbf{x}_{i+1} &= A' \Delta \mathbf{x}_i + B' \Delta \mathbf{u}_i, \\ \mathbf{y}_i &= C \Delta \mathbf{x}_i + \mathbf{v}_i, \end{aligned} \tag{2}$$

where $\Delta \mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_0$, $\Delta \mathbf{u}_i = \mathbf{u}_i - \mathbf{u}_0$, and

$$A' = \frac{\partial f}{\partial \mathbf{x}}\bigg|_{(\mathbf{x}_0, \mathbf{u}_0)} T_s + I_n, \quad B' = \frac{\partial f}{\partial \mathbf{u}}\bigg|_{(\mathbf{x}_0, \mathbf{u}_0)} T_s, \quad C = \frac{\partial g}{\partial \mathbf{x}}\bigg|_{\mathbf{x}_0}.$$

Consider a stable controller $\Delta \mathbf{u}_i = K \Delta \mathbf{x}_i$, where K is designed properly to achieve $\lim_{i \to \infty} \|\Delta \mathbf{x}_i\| = 0$. Then we study the attack design for the system at the stable equilibrium points. Consequently, the following closed-form dynamical model is used throughout the subsequent development:

$$\begin{aligned} \mathbf{x}_{i+1} &= A\mathbf{x}_i, \\ \mathbf{y}_i &= C\mathbf{x}_i + \mathbf{v}_i, \end{aligned} \tag{3}$$

where $A = A' + B'K$ is stable. For the sake of clean presentation, we make the abuse of $\mathbf{x}_i$ as $\Delta \mathbf{x}_i$ in the model (3). Then the measurement model on the window I is given by

$$\mathbf{y}_I = H\mathbf{x}_i + \mathbf{v}_I, \tag{4}$$

where $H = \begin{bmatrix} CA^{1-T} \\ CA^{2-T} \\ \vdots \\ C \end{bmatrix}$ is a backward observation matrix and $\mathbf{y}_I = [\mathbf{y}_{i-T+1}^\top \ \mathbf{y}_{i-T+2}^\top \ \cdots \ \mathbf{y}_i^\top]^\top \in \mathbb{R}^{mT}$, $\mathbf{v}_I \in \mathbb{R}^{Tm}$ are the horizontal measurement vector and the noise vector, respectively, in the window I. The following widely used assumptions are made for the system model above:

1. The CPS (3) is asymptotically stable: $0 < \rho(A) < 1$.
2. The pair (C, A^{-1}) is observable: $\operatorname{rank}(H) = n$.
3. The noise is bounded: $\|\mathbf{v}_I\|_2 \leq \varepsilon_v$, for a known constant $\varepsilon_v > 0$.

Consequently, we start to discuss the MHE and bad data detector (BDD) used in this paper. Given the measurement model in (4), MHE seeks to obtain an estimate of the state vector $\mathbf{x}_i$ from the most recent T measurements in the window I.

---

[1] The physical processes interact with the cyber components operating in discrete time intervals, so researchers often use discrete model for studies of control and estimation in CPS (Weerakkody, Ozel, Mo, Sinopoli, et al., 2019).

Definition 1 (*Moving-Horizon Estimator, MHE* Allan & Rawlings, 2019; Rao, Rawlings, & Mayne, 2003). A moving-horizon estimator is an operator of the form $\mathcal{D} : \mathbb{R}^{Tm} \rightarrow \mathbb{R}^n$ which returns an estimate of the state vector from T-horizon observation. Moreover, $\mathcal{D}$ is said to be stable if there exists $\tau_0, \varepsilon_0 < \infty$ such that

$$\|\mathcal{D}(\mathbf{y}_I) - \mathbf{x}_i\|_2 \le \tau_0 \|\mathbf{x}_i\|_2 + \varepsilon_0, \tag{5}$$

for all $i > 0$.

Remark 1. For ℓ_2 MHE, $\mathcal{D}_2$ is a linear function of $\mathbf{y}_I$ and given by:

$$\mathcal{D}_2(\mathbf{y}_I) \triangleq \arg\min_{\mathbf{x}} \|\mathbf{y}_I - H\mathbf{x}\|_2 = H^\dagger \mathbf{y}_I. \tag{6}$$

As MHE estimates the states from the measurements of window I, BDD monitors the state estimation process to detect any malicious inputs. It is often designed based on the residual $\|\mathbf{y}_I - H\mathcal{D}(\mathbf{y}_I)\|_p$, $0 < p \le \infty$. $p = 2$ is commonly used (Liu et al., 2011; Mo & Sinopoli, 2010). $p = 1$ and $p = \infty$ have also been used (Kosut, Jia, Thomas, & Tong, 2010). The BDD considered in this paper is based on the ℓ_2 MHE in (6).

Definition 2 (*Bad Data Detector, BDD*). Consider the MHE in (6) and the measurement model in (4). Given a threshold $\delta > 0$, a bad data detector is a discriminator given by

$$\text{BDD}(\mathbf{y}_I) = \begin{cases} 1 & if \ \|\mathbf{y}_I - H\mathcal{D}_2(\mathbf{y}_I)\|_2 > \delta, \\ 0 & otherwise. \end{cases} \tag{7}$$

Next, we discuss the ineffectiveness of static FDIA designs (Liu et al., 2011; Pasqualetti et al., 2013) against the MHE in (6) and the basic BDD in (7).

1. The original FDIA design strategy is to define the attack $\mathbf{e}_i$ on the range space of the measurement matrix C, i.e. $\mathbf{e}_i = C\mathbf{a}_i$, where $\mathbf{a}_i$ is an arbitrary signal. We call this type of FDIA a perfect stealth attack against static ℓ_2 estimator since the detection residual is zero:

$$\|\mathbf{y}_i + \mathbf{e}_i - CC^\dagger(\mathbf{y}_i + \mathbf{e}_i)\|_2 = \|(I - CC^\dagger)C(\mathbf{x}_i + \mathbf{a}_i)\|_2 = 0.$$

Here we ignore the noise $\mathbf{v}_i$ in the model (3). However, under the MHE, $\mathbf{y}_I = H\mathbf{x}_i + (I_T \otimes C)\mathbf{a}_I$, where $\mathbf{a}_I = \begin{bmatrix} \mathbf{a}_{i-T+1}^\top, & \cdots, & \mathbf{a}_i^\top \end{bmatrix}^\top$, then the residual is

$$\|\mathbf{y}_I - H\mathcal{D}_2(\mathbf{y}_I)\|_2 = \|(I - HH^\dagger)\mathbf{y}_I\|_2$$
$$= \|(I - HH^\dagger)(I_T \otimes C)\mathbf{a}_I\|_2$$

Notice the second equality follows from $(I - HH^\dagger)H = 0$. It is clear that $\text{span}(I_T \otimes C) \cap \text{range}(H) \ne \emptyset$ is not guaranteed. In fact, if H has full column rank, the above holds if and only if $A = \beta I$ for some $\beta \in \mathbb{R}$. In other words, static FDIA is effective against MHE if and only if the states of CPS are completely decoupled.

2. Another well-known design is a generalized stealth FDIA (Liu et al., 2011) which is given by the following program:

$$\begin{aligned} \text{Maximize} : \quad & \|\mathcal{D}_2(\mathbf{y}_I) - \mathcal{D}_2(\mathbf{y}_i + \mathbf{e}_i)\|_2, \\ \text{Subject to} : \quad & \|\mathbf{y}_i + \mathbf{e}_i - C\mathcal{D}_2(\mathbf{y}_i + \mathbf{e}_i)\|_2 \le \delta/T. \end{aligned} \tag{8}$$

Considering the system model in (3) without noise, the constraint in (8) is equivalent to $\|(I - CC^\dagger)\mathbf{e}_i\|_2 \le \delta/T$ which does not guarantee that $\|(I - HH^\dagger)\mathbf{e}_I\|_2 \le \delta$.

3. Consider the case where the static design is done using H instead of C, for example, $e_I = H\mathbf{a}_I$. Here the conflict arises when the time window moves to $I+1$. The design has to guarantee that the attack vector is consistent from window to window. This renders the static design infeasible for the moving-horizon estimator.

Consequently, a moving-horizon FDIA design is needed against MHE with consideration for recursive feasibility.

4. Moving-horizon FDIA design

In this section, we start with a definition of successful FDIA, then develop the proposed MH-FDIA. The recursive feasibility of the moving-horizon schemes is also analyzed. The following assumptions on the attacks are widely used (Liu et al., 2011; Mo & Sinopoli, 2010):

1. The attacker has limited access to sensors; $\mathbf{e}_i \in \Sigma_k$ for some $k < m$.
2. The attacker can inject arbitrary values on compromised sensors.
3. The attacker will not abandon compromised sensors or gain access to more sensors at runtime. Thus, the attack support is time-invariant ($\dot{\mathcal{T}} = 0$).

Henceforth, we denote the support of the attack sequence for the time-horizon I by $\mathcal{T}_I = [\mathcal{T}^\top, m + \mathcal{T}^\top, \ldots, (T-1)m + \mathcal{T}^\top]^\top$, and $\mathcal{T}_{I^-}$ is defined accordingly. Next, we formalize the attack performance criteria used in this paper.

Definition 3 (*Successful FDIA* Sui et al., 2020; Zheng & Anubi, 2021). Given the estimator-detector pair $(\mathcal{D}_2(\mathbf{y}_I), \text{BDD}(\mathbf{y}_I))$ in (6) and (7) respectively, the attack vector $\mathbf{e}_I \in \Sigma_k$ is said to be (α, ϵ)-successful if the following hold:

$$\|\mathcal{D}_2(\mathbf{y}_I) - \mathcal{D}_2(\mathbf{y}_I + \mathbf{e}_I)\|_2 \ge \alpha, \quad \|\mathbf{y}_I + \mathbf{e}_I - H\mathcal{D}_2(\mathbf{y}_I + \mathbf{e}_I)\|_2 \le \epsilon. \tag{9}$$

In accordance with common usage in the literature (Khazraei et al., 2022; Liu et al., 2011; Mo & Sinopoli, 2010; Pasqualetti et al., 2013; Zheng & Anubi, 2021), we adopt two metrics to evaluate the effectiveness and stealthiness of the attack. To quantify the effectiveness, we use the estimation error $\|\mathcal{D}(\mathbf{y}_I) - \mathcal{D}(\mathbf{y}_I + \mathbf{e}_I)\|_2$. To quantify the stealthiness, we use the estimation residual $\|\mathbf{y}_I + \mathbf{e}_I - H\mathcal{D}(\mathbf{y}_I + \mathbf{e}_I)\|_2$, since the estimation process is the primary target of sensor attacks. If the estimation result could be spoofed without triggering an alarm, the corresponding control action is then manipulated maliciously, causing otherwise stable controller to drive the system to an undesired state.

Remark 2. Without loss of generality, we assume that $\text{supp}(\mathbf{e}_i) \cap \text{supp}(\mathbf{v}_i) = \emptyset$. Whenever $\text{supp}(\mathbf{e}_i) \cap \text{supp}(\mathbf{v}_i) \ne \emptyset$, one could redefine a new attack vector $\mathbf{e}_i + \mathbf{v}_{i_{\text{supp}(\mathbf{e}_i)}}$ and new noise vector $\mathbf{v}_{i_{\overline{\text{supp}(\mathbf{e}_i)}}}$ that will satisfy this assumption.

Remark 3. Given an attack history $\mathbf{e}_{I^-} = [\mathbf{e}_{i-T+1}^\top, \mathbf{e}_{i-T+2}^\top, \ldots, \mathbf{e}_{i-1}^\top]^\top$, the FDIA $\mathbf{e}_i$ at time instance i is said to be (α, ϵ)-successful if $\mathbf{e}_I = [\mathbf{e}_{I^-}^\top, \mathbf{e}_i^\top]^\top$ satisfies the conditions in (9).

Next, since we target an ℓ_2 MHE in (6), the conditions in (9) reduces to

$$\|H^\dagger \mathbf{e}_I\|_2 \ge \alpha, \quad \|(I - HH^\dagger)(\mathbf{y}_I + \mathbf{e}_I)\|_2 \le \epsilon. \tag{10}$$

The second inequality can further be simplified as follows:

$$\|(I - HH^\dagger)(\mathbf{y}_I + \mathbf{e}_I)\|_2 = \|(I - HH^\dagger)(H\mathbf{x}_i + \mathbf{v}_I + \mathbf{e}_I)\|_2$$
$$= \|(I - HH^\dagger)(\mathbf{v}_I + \mathbf{e}_I)\|_2 \le \epsilon.$$

Now, since $\|(I - HH^\dagger)\mathbf{v}_I\|_2 \le \epsilon_v$, using the assumption in Remark 2, the following conditions are sufficient for the (α, ϵ) criteria in (9):

$$\|H^\dagger \mathbf{e}_I\|_2 \ge \alpha, \quad \|(I - HH^\dagger)\mathbf{e}_I\|_2 \le \sqrt{\epsilon^2 - \epsilon_v^2} \triangleq \tilde{\epsilon}. \tag{11}$$

Let H admit the singular value decomposition

$$H = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^\top,$$

where $U_1 \in \mathbb{R}^{Tm \times n}, U_2 \in \mathbb{R}^{Tm \times (Tm-n)}$, $\Sigma \in \mathbb{R}^{n \times n}$ is a diagonal matrix composed of all non-zeros singular values and $V \in \mathbb{R}^{n \times n}$. The next result gives a parameterization of successful MH-FDIAs that will be used for subsequent design.

Proposition 1. *Given* $\mathbf{w}_1 \in \mathbb{R}^n, \mathbf{w}_2 \in \mathbb{R}^{mT-n}$, *the moving-horizon FDIA vector given by*

$$\mathbf{e}_I = U_1 \Sigma \mathbf{w}_1 + U_2 \mathbf{w}_2 \tag{12}$$

is $(\|\mathbf{w}_1\|_2, \|\mathbf{w}_2\|_2)$-*successful against the estimator-detector pair* $(\mathcal{D}(\mathbf{y}_I),$ $\mathrm{BDD}(\mathbf{y}_I))$ *in* (5) *and* (7).

Proof. Since H is full-ranked, Σ is invertible. It follows then that $H^\dagger = V \Sigma^{-1} U_1^\top$ and $I - H H^\dagger = I - U_1 U_1^\top = U_2 U_2^\top$. Then,

$$\begin{aligned}
\|H^\dagger \mathbf{e}_I\|_2 &= \|V \Sigma^{-1} U_1^\top (U_1 \Sigma \mathbf{w}_1 + U_2 \mathbf{w}_2)\|_2 \\
&= \|V \mathbf{w}_1\|_2 = \|\mathbf{w}_1\|_2,
\end{aligned}$$

and

$$\begin{aligned}
\|(I - H H^\dagger) \mathbf{e}_I\|_2 &= \|U_2 U_2^\top (U_1 \Sigma \mathbf{w}_1 + U_2 \mathbf{w}_2)\|_2 \\
&= \|U_2 \mathbf{w}_2\|_2 = \|\mathbf{w}_2\|_2.
\end{aligned}$$

Thus $\mathbf{e}_I$ is $(\|\mathbf{w}_1\|_2, \|\mathbf{w}_2\|_2)$-successful. $\square$

Thus far, we have obtained a structure for a successful FDIA, as defined in (12), for a static window of length T. However, this structure can only ensure static feasibility for the current time window. Next, we derive conditions for recursive feasibility as the observation window moves. At time i, the attackers know their injected historical attacks $\mathbf{e}_{I^-}$, then, according to (12), the goal at the current time i is to find $\mathbf{e}_i \in \mathbb{R}^m$ such that

$$\begin{bmatrix} \mathbf{e}_{I^-} \\ \mathbf{e}_i \end{bmatrix} = U_1 \Sigma \mathbf{w}_1 + U_2 \mathbf{w}_2 \tag{13}$$

for some $\mathbf{w}_1 \in \mathbb{R}^n$ with $\|\mathbf{w}_1\|_2 \geq \alpha$ and $\mathbf{w}_2 \in \mathbb{R}^{Tm-n}$ with $\|\mathbf{w}_2\|_2 \leq \tilde{\epsilon}$. It is challenging to specify a priori the values of α and $\tilde{\epsilon}$ that can guarantee the condition above for all time. Fortunately, for the purpose of this paper, we regard the stealthiness of the resulting attack signal more important than its effectiveness. Thus, we employ the strategy of pre-specifying $\tilde{\epsilon}$, then obtaining a time-varying bias α by searching for $\mathbf{w}_1$ whose norm is as big as we can guarantee while satisfying the constraint $\|\mathbf{w}_2\|_2 \leq \tilde{\epsilon}$. Let

$$\mathbf{w}_1^- = \Sigma^{-1} U_1^\top \begin{bmatrix} \mathbf{e}_{I^-} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{w}_2^- = U_2^\top \begin{bmatrix} \mathbf{e}_{I^-} \\ \mathbf{0} \end{bmatrix}, \tag{14}$$

then $\begin{bmatrix} \mathbf{e}_{I^-} \\ \mathbf{0} \end{bmatrix} = U_1 \Sigma \mathbf{w}_1^- + U_2 \mathbf{w}_2^-$. Consequently, we seek vectors $\mathbf{z}_1 \in \mathbb{R}^n$ and $\mathbf{z}_2 \in \mathbb{R}^{Tm-n}$ that satisfy:

$$\mathrm{supp}(U_1 \Sigma \mathbf{z}_1 + U_2 \mathbf{z}_2) = (T-1)m + \mathcal{T} \triangleq \bar{\mathcal{T}}, \tag{15}$$

$$\text{and } \left\| \mathbf{z}_2 + \mathbf{w}_2^- \right\|_2 \leq \tilde{\epsilon}, \tag{16}$$

where $\bar{\mathcal{T}}$ is the support of $\begin{bmatrix} \mathbf{0} \\ \mathbf{e}_i \end{bmatrix}$. Thus far, we obtain a parameterization of the attack at the current time $\begin{bmatrix} \mathbf{0} \\ \mathbf{e}_i \end{bmatrix} \triangleq U_1 \Sigma \mathbf{z}_1 + U_2 \mathbf{z}_2$. If $\mathbf{z}_1$ and $\mathbf{z}_2$ satisfy the conditions in (15) and (16), then, according to Remark 3, $\mathbf{e}_i$ is $(\alpha, \tilde{\epsilon})$-successful with the historical attacks $\mathbf{e}_{I^-}$. In other words, the attack $\mathbf{e}_i$ maintains the recursive feasibility when the observation window moves from $I-1$ to I.

Next, let $N = \begin{bmatrix} N_1 \\ N_2 \end{bmatrix}$ be a matrix in the null space of $\begin{bmatrix} U_1 \Sigma & U_2 \end{bmatrix}_{\bar{\mathcal{T}}^c}$ such that

$$\begin{bmatrix} U_{1_{\bar{\mathcal{T}}^c}} \Sigma & U_{2_{\bar{\mathcal{T}}^c}} \end{bmatrix} N = U_{1_{\bar{\mathcal{T}}^c}} \Sigma N_1 + U_{2_{\bar{\mathcal{T}}^c}} N_2 = 0, \tag{17}$$

then the inequality constraint in (16) is equivalent to

$$\|N_2 \mathbf{v} + \mathbf{w}_2^-\|_2 \leq \tilde{\epsilon}, \tag{18}$$

for some vector $\mathbf{v}$ since $\mathbf{z}_2$ must be in the range space of N_2. The next result gives a necessary and sufficient condition for (18).

Proposition 2. *The inequality in* (18) *is feasible if and only if*

$$\|N_2^{\perp\top} \mathbf{w}_2^-\|_2 \leq \tilde{\epsilon}. \tag{19}$$

Proof. Using the identity,[2] $N_2 N_2^\dagger + N_2^\perp N_2^{\perp\top} = I$, it follows that:

$$\begin{aligned}
\|N_2 \mathbf{v} + \mathbf{w}_2^-\|_2^2 &= \|N_2(\mathbf{v} + N_2^\dagger \mathbf{w}_2^-) + N_2^\perp N_2^{\perp\top} \mathbf{w}_2^-\|_2^2 \\
&= \|N_2(\mathbf{v} + N_2^\dagger \mathbf{w}_2^-)\|_2^2 + \|N_2^\perp N_2^{\perp\top} \mathbf{w}_2^-\|_2^2 \\
&= \|N_2(\mathbf{v} + N_2^\dagger \mathbf{w}_2^-)\|_2^2 + \|N_2^{\perp\top} \mathbf{w}_2^-\|_2^2.
\end{aligned}$$

Thus, $\min_{\mathbf{v}} \|N_2 \mathbf{v} + \mathbf{w}_2^-\|_2 = \|N_2^{\perp\top} \mathbf{w}_2^-\|_2$. The result follows by noting that the inequality in (18) is feasible if and only if $\min_{\mathbf{v}} \|N_2 \mathbf{v} + \mathbf{w}_2^-\|_2 \leq \tilde{\epsilon}$. $\square$

Remark 4. The above proposition gives the condition for the existence of a feasible attack vector $\mathbf{e}_i \in \mathbb{R}^m$, at time i, as

$$\left\| N_2^{\perp\top} U_2^T \begin{bmatrix} \mathbf{e}_{I^-} \\ \mathbf{0} \end{bmatrix} \right\|_2 \leq \tilde{\epsilon}. \tag{20}$$

If, at the current instant, this condition is violated, we simply set $\mathbf{e}_i = \mathbf{0}$. Alternatively, the condition in (20) can also be used with $\mathbf{e}_{(I+1)^-}$ to establish recursive feasibility for the next time instant.

For each $\mathbf{v}$ satisfying (18),

$$\alpha(\mathbf{v}) = \|N_1 \mathbf{v} + \mathbf{w}_1^-\|_2 \tag{21}$$

is the corresponding effectiveness (state bias) induced by the attack vector

$$\mathbf{e}_I = \begin{bmatrix} \mathbf{e}_{I^-} \\ \mathbf{e}_i \end{bmatrix} = U_1 \Sigma(\mathbf{w}_1^- + N_1 \mathbf{v}) + U_2(\mathbf{w}_2^- + N_2 \mathbf{v}). \tag{22}$$

Next, we develop an iterative scheme to improve the effectiveness $\alpha(\mathbf{v})$ while satisfying the stealthiness condition in (18). Given a vector $\mathbf{v}_k$ satisfying (18) and a positive constant λ, let

$$\mathbf{d}_k = \begin{cases} \dfrac{N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)}{\|N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2}, & \text{if } \|N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2 \geq \tau, \\ N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-), & \text{otherwise,} \end{cases} \tag{23}$$

$$\mathbf{r}_k = N_2 \mathbf{v}_k + \mathbf{w}_2^-, \quad \hat{\mathbf{d}}_k = \frac{N_2 \mathbf{d}_k}{\|N_2 \mathbf{d}_k\|_2}, \tag{24}$$

$$\lambda_k = \begin{cases} \lambda, & \text{if } \|\mathbf{r}_k + \lambda N_2 \mathbf{d}_k\|_2 \leq \tilde{\epsilon}, \\ \dfrac{-\mathbf{r}_k^\top \hat{\mathbf{d}}_k + \sqrt{(\mathbf{r}_k^\top \hat{\mathbf{d}}_k)^2 - \|\mathbf{r}_k\|_2^2 + \tilde{\epsilon}^2}}{\|N_2 \mathbf{d}_k\|_2}, & \text{otherwise,} \end{cases} \tag{25}$$

then the update law on $\mathbf{v}_k$ is given by

$$\mathbf{v}_{k+1} = \mathbf{v}_k + \lambda_k \mathbf{d}_k, \tag{26}$$

where τ is a pre-defined zero-tolerance value.

Theorem 4.1. *For the update law in* (26), *the followings hold:*

1. $\mathbf{v}_{k+1}$ *satisfies the constraint in* (18);
2. $\alpha(\mathbf{v}_{k+1}) \geq \alpha(\mathbf{v}_k)$.

Proof. (1) According to (26), it follows

$$\begin{aligned}
\|N_2 \mathbf{v}_{k+1} + \mathbf{w}_2^-\|_2^2 &= \|N_2 \mathbf{v}_k + \mathbf{w}_2^- + \lambda_k N_2 \mathbf{d}_k\|_2^2 \\
&= \|\mathbf{r}_k + \lambda_k N_2 \mathbf{d}_k\|_2^2 \\
&= \lambda_k^2 \|N_2 \mathbf{d}_k\|_2^2 + 2\lambda_k \mathbf{r}_k^\top N_2 \mathbf{d}_k + \|\mathbf{r}_k\|_2^2
\end{aligned} \tag{27}$$

According to (25), the two choices for λ_k imply that either $\lambda_k = \lambda$, which implies that

$$\|\mathbf{r}_k + \lambda N_2 \mathbf{d}_k\|_2 \leq \tilde{\epsilon},$$

or $\lambda_k = \dfrac{-\mathbf{r}_k^\top \hat{\mathbf{d}}_k + \sqrt{(\mathbf{r}_k^\top \hat{\mathbf{d}}_k)^2 - \|\mathbf{r}_k\|_2^2 + \tilde{\epsilon}^2}}{\|N_2 \mathbf{d}_k\|_2}$, then substituting it into (27) yields

$$\lambda_k^2 \|N_2 \mathbf{d}_k\|_2^2 + 2\lambda_k \mathbf{r}_k^\top N_2 \mathbf{d}_k + \|\mathbf{r}_k\|_2^2 = \tilde{\epsilon}^2, \tag{28}$$

---

[2] Since N_2 is defined by Eq. (17) one could assume, without loss of generality, that the columns of N_2 are orthogonal. Thus, $N_2^\top = N_2^\dagger$.

which also implies that

$$\|\mathbf{r}_k + \lambda_k N_2 \mathbf{d}_k\|_2 \leq \tilde{\epsilon}.$$

(2) By following (21) and (26),

$$\alpha(\mathbf{v}_{k+1}) = \|N_1 \mathbf{v}_{k+1} + \mathbf{w}_1^-\|_2$$
$$= \|N_1 \mathbf{v}_k + \mathbf{w}_1^- + \lambda_k N_1 \mathbf{d}_k\|_2.$$

If $\|N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2 \geq \tau$, then, according to (23), $\mathbf{d}_k = \frac{N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)}{\|N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2}$. Thus,

$$\alpha(\mathbf{v}_{k+1}) = \left\| N_1 \mathbf{v}_k + \mathbf{w}_1^- + \lambda_k \frac{N_1 N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)}{\|N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2} \right\|_2$$
$$= \left\| \left(I + \frac{\lambda_k N_1 N_1^\top}{\|N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2} \right)(N_1 \mathbf{v}_k + \mathbf{w}_1^-) \right\|_2$$
$$\geq \underline{\sigma}\left(I + \frac{\lambda_k N_1 N_1^\top}{\|N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2} \right) \alpha(\mathbf{v}_k) \geq \alpha(\mathbf{v}_k).$$

If $\|N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2 < \tau$, then, according to (23), $\mathbf{d}_k = N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)$. Thus,

$$\alpha(\mathbf{v}_{k+1}) = \|N_1 \mathbf{v}_k + \mathbf{w}_1^- + \lambda_k N_1 N_1^\top(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2$$
$$= \|(I + \lambda_k N_1 N_1^\top)(N_1 \mathbf{v}_k + \mathbf{w}_1^-)\|_2$$
$$\geq \underline{\sigma}(I + \lambda_k N_1 N_1^\top)\alpha(\mathbf{v}_k)$$
$$\geq \alpha(\mathbf{v}_k).$$

where $\underline{\sigma}(I + \lambda_k N_1 N_1^\top)$ is the smallest singular value of $I + \lambda_k N_1 N_1^\top$ that must be bigger than 1. $\square$

Remark 5. For $T = 1$, the MH-FDIA problem reduces to the classic FDIA problem,

$$\begin{array}{ll} \text{Maximize} & \|N_1 \mathbf{v}\|_2 \\ \text{Subject to} & \|N_2 \mathbf{v}\|_2 \leq \epsilon \end{array} \quad (29)$$

for which numerous results exist in the literature using different relaxation techniques (Liu et al., 2011) or an approximate direct solution to the non-convex problem using learning-based approaches (Khazraei, Hallyburton, Gao, Wang, & Pajic, 2021).

The first conclusion in Theorem 4.1 indicates that (18), which is the constraint for guaranteeing recursive feasibility, can always be met during the iterative procedure. The proposed MH-FDIA scheme is summarized in Algorithm 1.

Algorithm 1 Moving-horizon FDIA

Parameters: $\tau, U_1, U_2, N_1, N_2, \Sigma, \epsilon, M$.

I. Input historical attacks $\mathbf{e}_{I^-}$.

II. $\mathbf{w}_1^- = \Sigma^{-1} U_1^\top \begin{bmatrix} \mathbf{e}_{I^-} \\ \mathbf{0} \end{bmatrix}$, $\mathbf{w}_2^- = U_2^\top \begin{bmatrix} \mathbf{e}_{I^-} \\ \mathbf{0} \end{bmatrix}$.

III. Suggest and improve bias update: Set $\mathbf{v}_1 = -N_2^\dagger \mathbf{w}_2^-$

for $k = 1 : M + 1$

  $\mathbf{v}_{k+1} = \mathbf{v}_k + \lambda_k$, where $\lambda_k, \mathbf{d}_k$ are given in (25) and (23).

end

IV. Output current attack vector $\mathbf{e}_i = U_{12} \Sigma(\mathbf{w}_1^- + N_1 \mathbf{v}_M) + U_{22}(\mathbf{w}_2^- + N_2 \mathbf{v}_M)$.

In addition, a conceptual visualization of the Algorithm 1 in $3D$ space is shown in Fig. 1. In this space, the constraint in (18) could be seen as an ellipsoid. The algorithm searches for a direction that increases the effectiveness α at each iterative step. Furthermore, the update law for step size λ in (25) ensures that the algorithm eventually terminates on the boundary at time step M, as proved in (28).
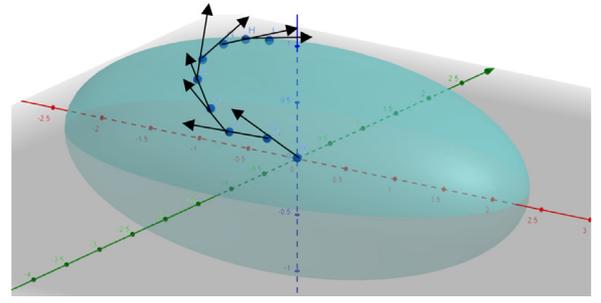
Fig. 1. A conceptual visualization of the Algorithm 1 for a 3 dimensional problem.
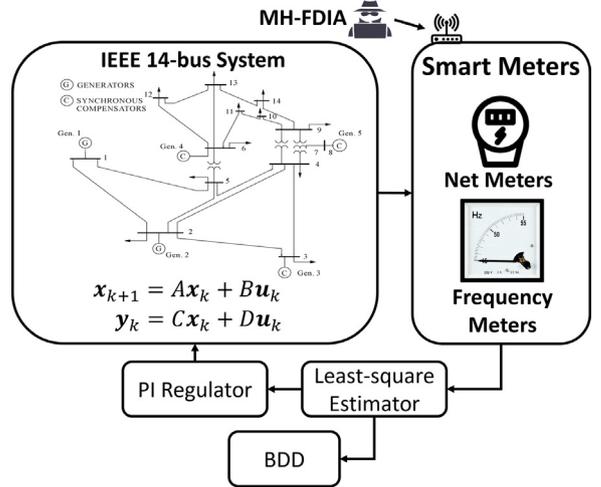
Fig. 2. Schematic description of the simulation of MH-FDIA on the IEEE 14-bus system.

5. Simulation and experiment

In this section, we validate the efficacy of the MH-FDIA using two case studies: a simulation of a linear regulation control system of the IEEE 14-bus network, and an experiment of a nonlinear path-following control system of a wheeled mobile robot (WMR). Firstly, we validate the theoretical performance of the MH-FDIA and analyze the influence of the parameters of Algorithm 1 on the MH-FDIA's performance using the linear bus system simulation. Through the nonlinear WMR experiment, we showcase the implementation of the proposed MH-FDIA on a nonlinear system and demonstrate its seamless performance in real-world scenarios.

5.1. Simulation: IEEE 14-bus system

In this subsection, the MH-FDIA was implemented on an IEEE 14-bus system with a moving-horizon least-square observer and a residual-based BDD, as shown in Fig. 2. The MH-FDIA was compared with conventional FDIA designs in the literature.

A small signal model is constructed by linearizing the generator swing and power flow equations around the operating point (Scholtz, 2004). In order to perform linearization, we assume that voltage is tightly controlled at its nominal value; the angular difference between each bus is small; and conductance is negligible therefore the system is lossless.

By ordering the buses such that the generator nodes appear first, the admittance-weighted Laplacian matrix can be expressed as $L = \begin{bmatrix} L_{gg} & L_{lg} \\ L_{gl} & L_{ll} \end{bmatrix} \in \mathbb{R}^{N \times N}$, where $N = n_g + n_b$ with $n_g = 5, n_b = 14$. Then,
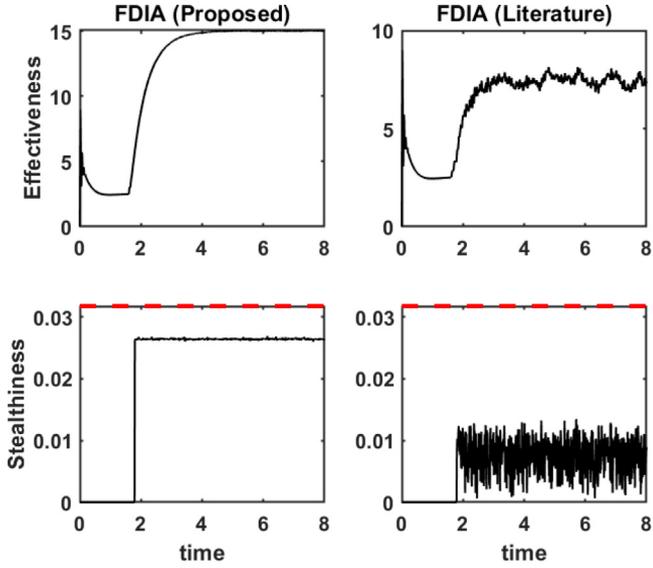
Fig. 3. A comparison of effectiveness and stealthiness of the static ℓ_2 FDIA designed in Liu et al. (2011) and the static form of the proposed MH-FDIA against Luenberger observer. (The effectiveness is given by the resulting state estimation error $\|D_2(\mathbf{y}_I) - D_2(\mathbf{y}_I + \mathbf{e}_I)\|_2$. Red dot line is the threshold of stealthiness, it is 0.0318 for one time step.).

Fig. 4. A comparison of effectiveness and stealthiness of the proposed MH-FDIA and the one in (31) against a moving-horizon least-square observer. (Red dot line is the threshold of stealthiness, it is $0.0318T = 0.6352$ for T time window).

Fig. 5. The effectiveness and stealthiness of the proposed MH-FDIA with respect to different maximum iteration M in Algorithm 1. The black lines are the mean values. The grayed areas show regions of all possible values. The red dot line is the threshold of stealthiness, it is $0.0318T = 0.6352$ for T time window.

by linearizing the dynamical swing equations and algebraic DC power flow equations, we obtain (Zheng & Anubi, 2022)

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & I \\ -M^{-1}(L_{gg} - L_{gl}L_{ll}^{-1}L_{lg}) & -M^{-1}D_g \end{bmatrix} \mathbf{x}$$
$$+ \begin{bmatrix} 0 & 0 \\ M^{-1} & -M^{-1}L_{gl}L_{ll}^{-1} \end{bmatrix} \mathbf{u},$$
$$\mathbf{y} = \begin{bmatrix} 0 & I \\ -P_{node}L_{ll}^{-1}Llg & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 & 0 \\ P_{node}L_{ll}^{-1} & 0 \end{bmatrix} \mathbf{u}, \quad (30)$$
$$\theta = -L_{ll}^{-1}(L_{lg}\delta - P_d),$$

where the state variables $\mathbf{x} = [\delta^\top \ \omega^\top]^\top \in \mathbb{R}^{2n_g}$ contain the generator rotor angles $\delta \in \mathbb{R}^{n_g}$ and the generator frequency $\omega \in \mathbb{R}^{n_g}$. The measurements $\mathbf{y} = [\omega^\top \ P_{net}^\top]^\top \in \mathbb{R}^{n_g + n_b}$ contain the generator frequency $\omega \in \mathbb{R}^{n_g}$ and the net power injected at the buses $P_{net} \in \mathbb{R}^{n_b}$. M is a diagonal matrix of inertial constants for each generator and D_g is a diagonal matrix of damping coefficients. The stealthiness threshold is set as 5% of the maximal nominal measurement value $\epsilon_i = 0.05 \max(\|\mathbf{y}\|_2) = 0.0318$ for each time step, so the stealthiness threshold for T-time horizon is $\epsilon = T\epsilon_i = 0.6352$. The fixed attack support is $\mathcal{T} = \{1, 2, 9, 11, 12, 16, 17\}$ (about 37% of total measurements). The simulation sampling time is set as $T_s = 0.01$ s, and the start time for attack injection is 1.8 s.

Assuming the attacker has the full knowledge of the system model and parameters in (30), we compared the MH-FDIA and the eigenvalue-based FDIA (Liu et al., 2011) against the Luenberger observer, as well as the moving-horizon least-square observer. Firstly, we implemented the MH-FDIA with a time horizon of 1, as shown in (29). Then the MH-FDIA was compared with the static FDIA solved in Liu et al. (2011) against the Luenberger observer. The comparison result is shown in Fig. 3. It is seen that both FDIAs can bypass the detection of the residual-based BDD. However, the proposed MH-FDIA induces a bigger bias in the resulting state estimates, hence is more effective.

Next, we extended the static FDIA in Liu et al. (2011) to a moving-horizon scheme by solving the following eigenvalue problem:

Maximize$_{\lambda, \mathbf{v}}$ $|\lambda|$

Subject to $\left\| \lambda U_{11} \Sigma \mathbf{v} - \mathbf{e}_{I_{T_{I-}}} \right\|_2^2 \leq \tilde{\epsilon}^2$ $\quad (31)$

where $U_{11} = \begin{bmatrix} I_n & 0_{n \times (Tm-n)} \end{bmatrix} U_{I-}^\top$ corresponds to the direction where $\mathbf{e}_{I-}$ takes effect on the range space of H. Then the resulting FDIA is given by $\mathbf{e}_{i_{T_i}} = \lambda^\star U_{12} \Sigma \mathbf{v}^\star$, where $\lambda^\star, \mathbf{v}^\star$ are the solution of (31), and $U_{12} = \begin{bmatrix} I_n & 0_{n \times (Tm-n)} \end{bmatrix} U_i^\top$ corresponds to the direction where $\mathbf{e}_i$ takes effect on the range space of H.

We implemented the proposed MH-FDIA with $\lambda_0 = 1e{-}4$ and a maximum number of iterations $M = 2000$. The results are shown in Fig. 4. It is seen that the proposed MH-FDIA has a much bigger effectiveness on the system than the state-of-art counterpart. Moreover, the proposed MH-FDIA explores the whole feasible region, while the one in the literature is more conservative. This indicates that the proposed MH-FDIA algorithm generates more powerful attacks.

Next, we performed additional simulations to explore the effect of the parameters of the proposed algorithm. We ran the proposed scheme 50 times with unique values of M ranging from 100 to 20 000. The initial step size was set as $\lambda = 1e{-}4$ and the convergence tolerance is set as $1e{-}4$. 7 attack channels were chosen at random. The results in Fig. 5 show the mean effectiveness and stealthiness, along with
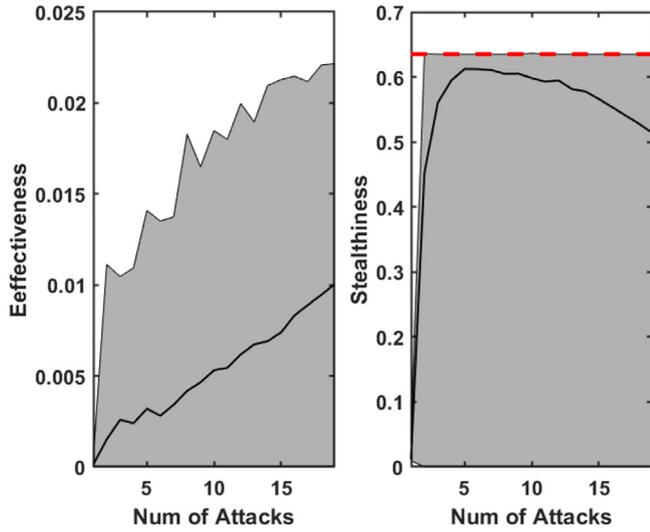
Fig. 6. The effectiveness and stealthiness of the MH-FDIA for different numbers of attacks. The corresponding attack supports are chosen at random. The black lines are the mean values. The grayed areas show the spread of all possible values. The red dot line is the threshold of stealthiness, it is $0.0318T = 0.6352$ for T time window.

Fig. 7. Schematic description of the simulation of MH-FDIA on the path-tracking control system of DDWMR.

their corresponding spread for all 50 simulation cases. It is seen that the mean effectiveness saturates at $M = 2131$ when $\lambda = 1e-4$. The saturation happens sooner ($M = 300$) when we increase the step size to $\lambda = 1e-3$. This shows that the proposed scheme is computationally efficient and one could adjust the parameters based on available computational resources.

Second, we evaluated the performance of the proposed MH-FDIA for different numbers of compromised measurements. 50 simulation experiments were carried out with randomized attack support for each different number of compromised sensors ranging from 1 to 19. We set the initial step size is set as $\lambda = 1e-4$. Fig. 6 shows that the effectiveness increases with the number of compromised measurements. On the other hand, all the maximum stealthiness is below the detection threshold. Moreover, it is seen that the maximum and mean stealthiness are close to the detection threshold, which indicates that the proposed algorithm explores the vulnerability space of the system as much as possible. Furthermore, it is seen that the mean stealthiness increases and then decreases since more explorable channels open up, thus making the attacks more stealthy without sacrificing effectiveness.

5.2. Experiment: Wheeled mobile robot

In this subsection, we implemented the proposed MH-FDIA on a nonlinear path-following control system of a differential-driven wheeled mobile robot (DDWMR). The attacker is assumed to only have model knowledge of vehicle Kinematics.

The non-holonomic kinematics of DDWMR is given by:

$$\begin{bmatrix} \dot{\theta} \\ \dot{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ cos(\theta) & -d\,sin(\theta) \\ sin(\theta) & d\,cos(\theta) \end{bmatrix} \mathbf{u} + \mathbf{w}$$
$$\triangleq f(\mathbf{x}, \mathbf{u}, \mathbf{w}),$$
(32)

where $\mathbf{u} = [v, \; \omega]^\top \in \mathbb{R}^2$ is the kinematic input vector composing of the longitudinal velocity v [m/s] and the yaw rate ω [rad/s], θ [rad] is the heading angle, $\mathbf{z} = [x, \; y]^\top \in \mathbb{R}^2$ is the location vector comprising the x, y-coordinates location of the vehicle, $\mathbf{x} = \begin{bmatrix} \theta & x & y \end{bmatrix}^\top$ is the augmented state vector composing of pose state and location states of the vehicle, $\mathbf{w} \sim \mathcal{N}(0, Q)$ is the process noise, and $d = 0.0562$ m is the distance between the medium point of the axis of wheels and
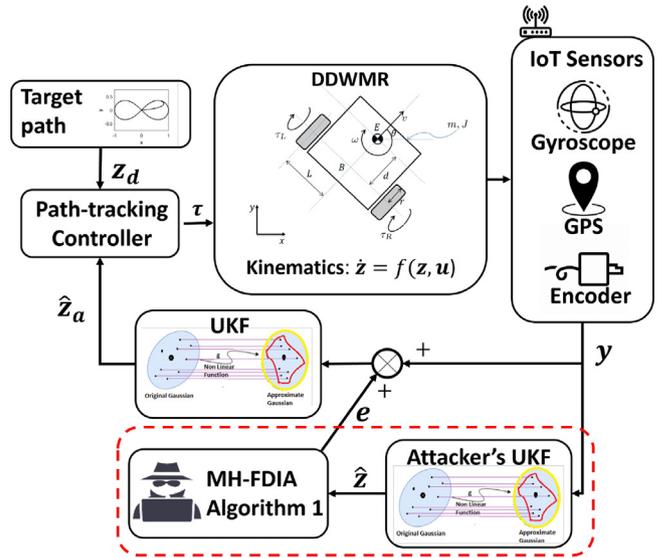
center of mass. Given the desired path $\mathbf{z}_d$, we use a common kinematic controller (Xie, Wang, & Wang, 2018; Zheng & Anubi, 2020):

$$\mathbf{u} = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -\frac{1}{d}sin(\theta) & \frac{1}{d}cos(\theta) \end{bmatrix} \left(\dot{\mathbf{z}}_d + K(\mathbf{z} - \mathbf{z}_d) \right),$$
(33)

where the control gain $K > 0$.

The vehicle is equipped with GPS, IMU, and an encoder, resulting in the measurement model

$$\mathbf{y} = \begin{bmatrix} cos(\theta) & 0 \\ 0 & sin(\theta) \\ 1 & 0 \\ 0 & 1 \\ 1/4r & L/4r \\ 1/4r & -L/4r \end{bmatrix} \cdot \mathbf{z} + \mathbf{v} \triangleq g(\mathbf{x}, \mathbf{v})$$
(34)

where $\mathbf{v} \sim \mathcal{N}(0, R)$. An unscented Kalman filter (UKF) is used to estimate the vehicle's states $\mathbf{x}$. Firstly, following standard unscented transformation (Julier & Uhlmann, 1997), we use $2n + 1$ sigma points to approximate the state $\mathbf{x}$ with assumed mean $\bar{\mathbf{x}}$ and covariance $P_{\mathbf{x}}$ as follows:

$$\chi_0 = \bar{\mathbf{x}}, \quad \chi_i = \bar{\mathbf{x}} + (\sqrt{(\lambda + n)P_{\mathbf{x}}})_i, \quad i = 1, \dots, n,$$
$$\chi_{i+n} = \bar{\mathbf{x}} + (\sqrt{(\lambda + n)P_{\mathbf{x}}})_{i-n}, \quad i = n + 1, \dots, 2n.$$

The corresponding weights for the sigma points are given as $W_0^m = \lambda/(n + \lambda)$, $W_0^c = W_0^m + (1 - \alpha^2 + \beta)$, $W_i = 1/2(n + \lambda)$, and $\lambda = \alpha^2(n + \kappa) - n$ represents how far the sigma points are away from the state, $\kappa \geq 0, \alpha \in (0, 1]$, and $\beta = 2$ is the optimal choice for Gaussian distribution. Assume $\mathbf{x}_{k-1} \sim \mathcal{N}(\bar{\mathbf{x}}_{k-1}, P_{\mathbf{x},k-1})$, the prediction step is given by

$$\mathcal{X}_k = f(\mathcal{X}_{k-1}, \mathbf{u}_k, \mathbf{w}_k), \quad \mathcal{Y}_k = g(\mathcal{X}_k, \mathbf{v}_k),$$
$$\hat{\mathbf{x}}_k^- = \sum_{i=0}^{2n} W_i \mathcal{X}_k, \quad \hat{\mathbf{y}}_k^- = \sum_{i=0}^{2n} W_i \mathcal{Y}_{k,i},$$
$$\hat{P}_{\mathbf{x},k} = \sum_{i=0}^{2n} W_i (\mathcal{X}_k - \hat{\mathbf{x}}_k)(\mathcal{X}_k - \hat{\mathbf{x}}_k)^T + R,$$

Next, given the new measurement $\mathbf{y}_k$, the correction step is as follows

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \hat{\mathbf{y}}_k^-), \quad P_{\mathbf{x},k} = \hat{P}_{\mathbf{x},k} - \mathbf{K}_k \hat{P}_{\mathbf{y},k} \mathbf{K}_k^T,$$
(35)

where the Kalman gain is $\mathbf{K}_k = \hat{P}_{\mathbf{zy}}\hat{P}_{\mathbf{y},k}^{-1}$ with

$$\hat{P}_{\mathbf{y},k} = \sum_{i=0}^{2n} W_i(\mathcal{Y}_{k,i} - \hat{\mathbf{y}}_k^-)(\mathcal{Y}_{k,i} - \hat{\mathbf{y}}_k^-)^T + Q,$$

$$\hat{P}_{\mathbf{zy}} = \sum_{i=0}^{2n} W_i(\mathcal{X}_{k,i} - \hat{\mathbf{x}}_k^-)(\mathcal{Y}_{k,i} - \hat{\mathbf{y}}_k^-)^T.$$

To establish an MH-FDIA generator, we linearized the models of the DDWMR around the equilibrium points $\mathbf{x}_{eq} = [\theta_0 \quad x_0 \quad y_0]^\top$, discretized it using Euler's approximation with a sampling time $T_s = 0.01$ s, and iterated forward $T_f = 20$ samples:

$$\mathbf{y}_{I_{T_f}} = H\mathbf{x}_{i-T_f+1} + G\mathbf{u}_{I_{T_f}} \tag{36}$$

where

$$H = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T_f} \end{bmatrix}, G = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ CB & 0 & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{T_f-1}B & CA^{T_f-2}B & \cdots & CB \end{bmatrix},$$

with

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ B = \begin{bmatrix} 0 & 1 \\ cos(\theta_0) & -d\,sin(\theta_0) \\ sin(\theta_0) & d\,cos(\theta_0) \end{bmatrix}T_s,$$

$$C = \begin{bmatrix} -x_0 sin(\theta_0) & cos(\theta_0) & 0 \\ y_0 cos(\theta_0) & 0 & sin(\theta_0) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1/4r & L/4r \\ 0 & 1/4r & -L/4r \end{bmatrix}.$$

Since each state of the kinematic model in (33) is an equilibrium point with zero velocity, the attacker must construct their own UKF to obtain the current location states of the vehicle $\mathbf{x}_{eq}$, as shown in Fig. 7. Thereafter, the attacker can implement Algorithm 1 with the linearized model in (36).

The MH-FDIA was injected through the third and fourth measurement channels, $\mathcal{T}_i = \{3, 4\}$. The stealthiness threshold is set as $\epsilon_i = 0.05$ for each time step, then $\epsilon = T_f \epsilon_i = 1$ for an observation window of length T_f. We use the distance from the nominal estimation residual range to measure the stealthiness. The parameters of Algorithm 1 are chosen as $\lambda_0 = 1e-4$ and $M = 2000$.

The experiment platform setup is depicted in Fig. 8. The DDMWR is constructed using the Yujin 2-wheel Kobuki robot with onboard Rasberry Pi computer and odometry sensors including two-wheel encoders, GPS, and a 3-axis gyroscope. We utilized the Kobuki ROS package[3] for the dynamic control of the DDWMR, sensor data collection, and odometry. The path-tracking kinematic controller, UKF and MH-FDIA were all implemented in Matlab/Simulink. The wireless communication channels between Simulink and Rasberry Pi were established based on ROS Melodic.

Three different path-following tasks were undertaken, namely a straight line, a circular path, and a figure-8 path. In the straight line case, the MH-FDIA was initiated at 6s. Prior to 6 s, the DDWMR was controlled to follow the desired straight line. Following the attack injection, the vehicle gradually deviated from the track, as shown in Fig. 9. However, the estimation residual was below the threshold ϵ so the attacks were not detected, as shown in Fig. 10. Similarly, in Figs. 11 and 13, it is seen that the DDWMR was controlled to follow the desired circular path and figure-8 path but run off the track after MH-FDIA was injected. However, the MH-FDIAs were not detected in the two experiments according to Figs. 12 and 14. It should be noted

---

[3] https://github.com/yujinrobot/kobuki.
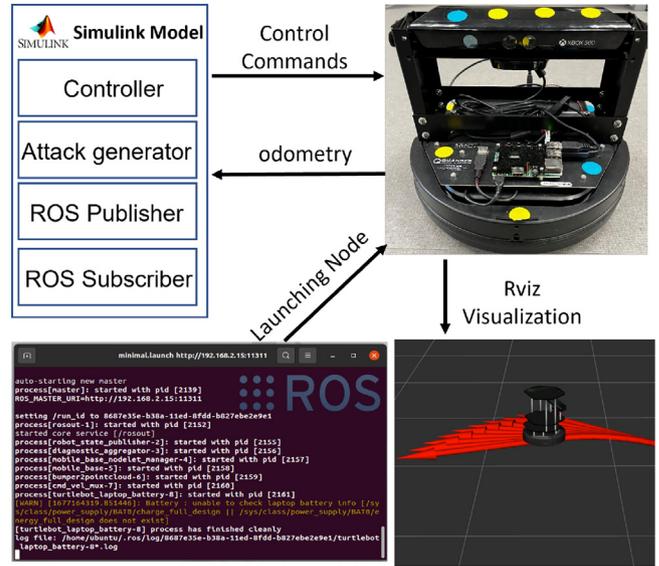
Fig. 8. Experiment platform setup (Simulink + ROS Melodic + Rasberry Pi + DDMWR + Odometry Sensors).
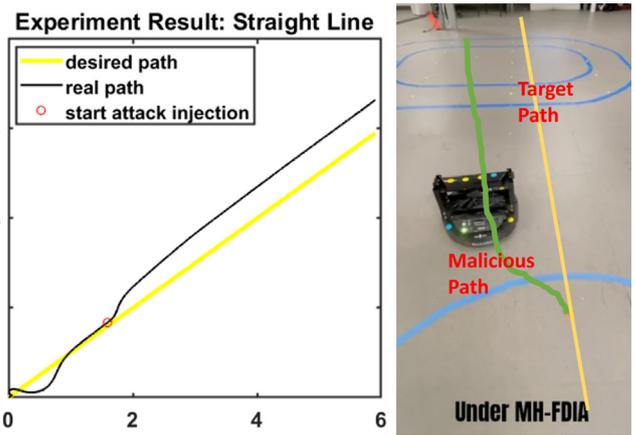
Fig. 9. The path of DDWMR with MH-FDIA in the straight line following test. (Initial pose and location: $\mathbf{x}_0 = [\pi/4 \quad 0 \quad 0]$).
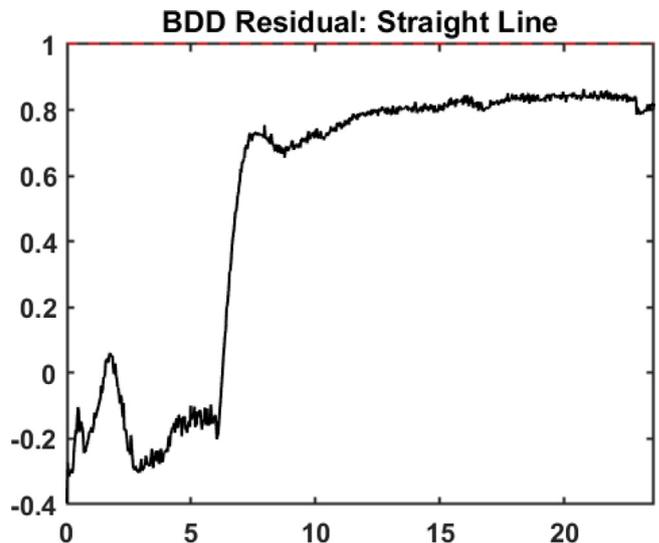
Fig. 10. The BDD residual for the straight line case (1 is the threshold, attack injection started at 6s).
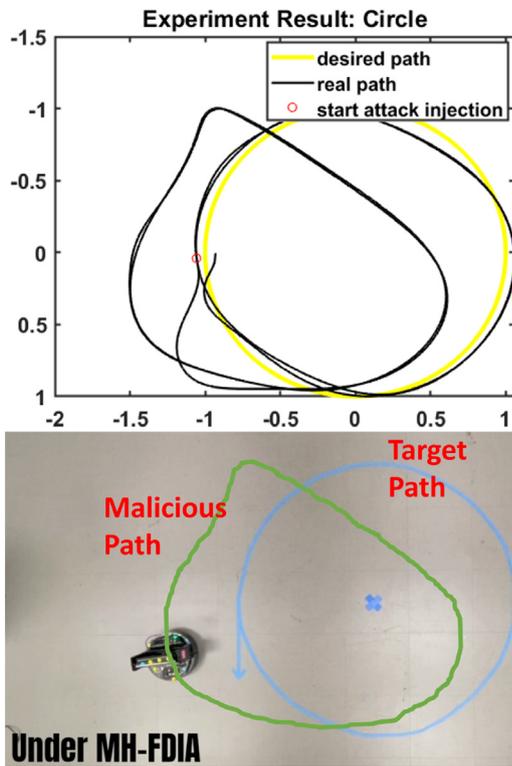
Fig. 11. The path of DDWMR with MH-FDIA in the circular path following test. (Initial pose and location: $\mathbf{x}_0 = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}$).
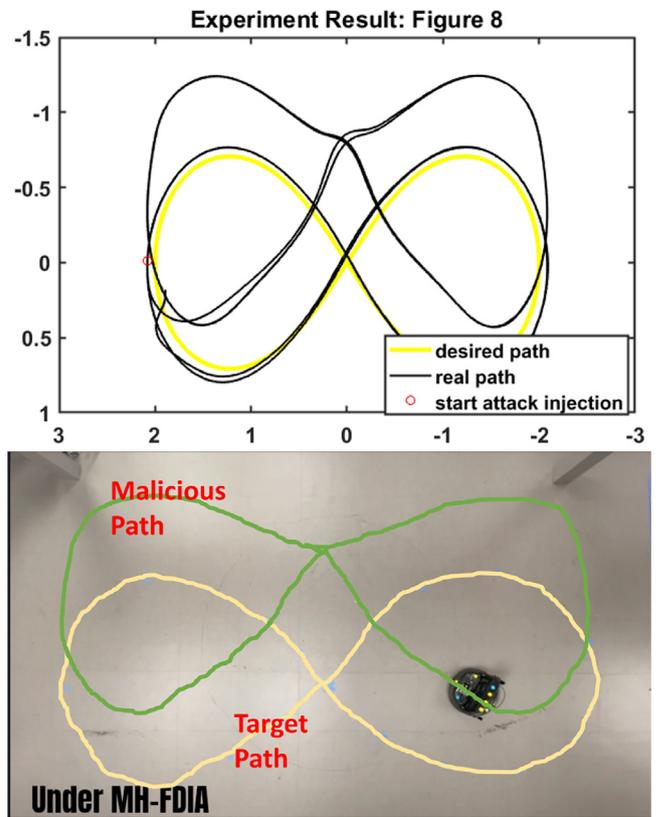
Fig. 13. The path of DDWMR with MH-FDIA in Figure-8 path following test. (Initial pose and location: $\mathbf{x}_0 = \begin{bmatrix} -\pi/2 & 2 & 0 \end{bmatrix}$).
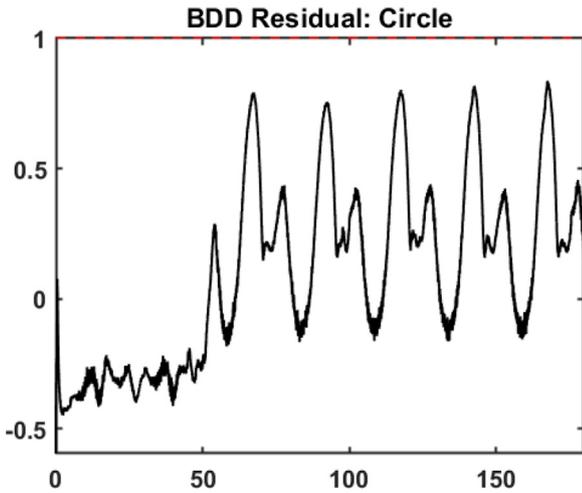
Fig. 12. The BDD residual for the circle case (1 is the threshold, attack injection started at 50s).
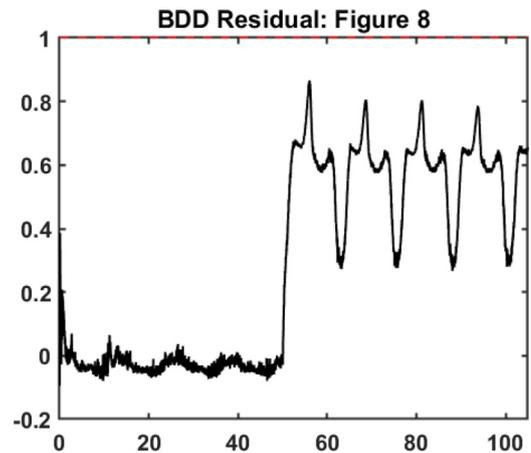
Fig. 14. The BDD residual for Figure-8 case (1 is the threshold, attack injection started at 50s).

that the MH-FDIA subtly misled the vehicle to a repeatable wrong path rather than inducing large state estimation errors. This guarantees that the MH-FDIA is still undetectable in real-world scenarios. Also, the misled wrong path has a similar shape to the target. This is due to the consideration of historical influence on the current attack injection.

6. Conclusion

In this paper, we demonstrated the importance of historical influence on FDIA design. A systematic MH-FDIA design framework is proposed. Based on a formal definition of successful FDIA, the MH-FDIA

design is given against ℓ_2 MHE and BDD, and shown to be (α, ϵ)-successful. Moreover, an adaptive algorithm is proposed to search for the most successful FDIAs while preserving recursive feasibility.

Developing a MH-FDIA generation algorithm that does not require pre-defined attack support will make the attack generation problem more challenging due to nonconvexity. Developing such algorithm is still an open problem. Additionally, a rigorous quantitative analysis of the MH-FDIA focusing on system's stability is also an interesting question that needs to be addressed. Furthermore, it is also valuable to investigate the applicability of the proposed method in other control systems, such as fuzzy systems and neural network-based systems. Last, to relax the reliance on model knowledge, sample-based techniques

could be utilized within the proposed moving-horizon attack generation framework.

Declaration of competing interest

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.conengprac.2023.105552.

References

Allan, D. A., & Rawlings, J. B. (2019). Moving horizon estimation. In *Handbook of model predictive control* (pp. 99–124). Springer.

Brentan, B., Rezende, P., Barros, D., Meirelles, G., Luvizotto, E., & Izquierdo, J. (2021). Cyber-attack detection in water distribution systems based on blind sources separation technique. *Water*, *13*(6), 795.

Hu, L., Wang, Z., Han, Q.-L., & Liu, X. (2018). State estimation under false data injection attacks: Security analysis and system protection. *Automatica*, *87*, 176–183.

Julier, S. J., & Uhlmann, J. K. (1997). New extension of the Kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI, Vol. 3068* (pp. 182–193). International Society for Optics and Photonics.

Khazraei, A., Hallyburton, S., Gao, Q., Wang, Y., & Pajic, M. (2021). Learning-based vulnerability analysis of cyber-physical systems. arXiv preprint arXiv:2103.06271.

Khazraei, A., Hallyburton, S., Gao, Q., Wang, Y., & Pajic, M. (2022). Learning-based vulnerability analysis of cyber-physical systems. In *2022 ACM/IEEE 13th International conference on cyber-physical systems* (pp. 259–269). IEEE.

Kosut, O., Jia, L., Thomas, R. J., & Tong, L. (2010). Limiting false data attacks on power system state estimation. In *44th Annual conference on information sciences and systems*. IEEE.

Liang, G., Weller, S. R., Zhao, J., Luo, F., & Dong, Z. Y. (2017). The 2015 Ukraine blackout: Implications for false data injection attacks. *IEEE Transactions on Power Systems*, *32*(4).

Liu, X., Bao, Z., Lu, D., & Li, Z. (2015). Modeling of local false data injection attacks with reduced network information. *IEEE Transactions on Smart Grid*, *6*(4), 1686–1696.

Liu, Y., Ning, P., & Reiter, M. K. (2011). False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security*, *14*(1), 1–33.

Löfberg, J. (2012). Oops! I cannot do it again: Testing for recursive feasibility in MPC. *Automatica*, *48*(3), 550–555.

Lu, A.-Y., & Yang, G.-H. (2022). False data injection attacks against state estimation without knowledge of estimators. *IEEE Transactions on Automatic Control*.

Mo, Y., & Sinopoli, B. (2010). False data injection attacks in control systems. In *Preprints of the 1st workshop on secure control systems* (pp. 1–6).

Mohammadpourfard, M., Ghanaatpishe, F., Mohammadi, M., Lakshminarayana, S., & Pechenizkiy, M. (2020). Generation of false data injection attacks using conditional generative adversarial networks. In *2020 IEEE PES Innovative smart grid technologies europe (ISGT-Europe)* (pp. 41–45). IEEE.

Muske, K. R., Rawlings, J. B., & Lee, J. H. (1993). Receding horizon recursive state estimation. In *American control conference* (pp. 900–904). IEEE.

Pasqualetti, F., Dörfler, F., & Bullo, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, *58*(11), 2715–2729.

Rao, C. V., Rawlings, J. B., & Mayne, D. Q. (2003). Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, *48*(2), 246–258.

Rawlings, J. B., Mayne, D. Q., & Diehl, M. (2017). *Model predictive control: Theory, computation, and design, Vol. 2*. Nob Hill Publishing Madison, WI.

Reeder, J. R., & Hall, C. T. (2021). Cybersecurity's pearl harbor moment: Lessons learned from the colonial pipeline ransomware attack.

Scholtz, E. (2004). *Observer-based monitors and distributed wave controllers for electromechanical disturbances in power systems* (Ph.D. thesis), Massachusetts Institute of Technology.

Sui, T., Mo, Y., Marelli, D., Sun, X., & Fu, M. (2020). The vulnerability of cyber-physical system under stealthy attacks. *IEEE Transactions on Automatic Control*, *66*(2), 637–650.

Weerakkody, S., Ozel, O., Mo, Y., Sinopoli, B., et al. (2019). Resilient control in cyber-physical systems: Countering uncertainty, constraints, and adversarial behavior. *Foundations and Trends® in Systems and Control*, *7*(1–2), 1–252.

Xie, D., Wang, S., & Wang, Y. (2018). Trajectory tracking control of differential drive mobile robot based on improved kinematics controller algorithm. In *2018 Chinese automation congress* (pp. 2675–2680). IEEE.

Xu, W., Ma, K., Trappe, W., & Zhang, Y. (2006). Jamming sensor networks: attack and defense strategies. *IEEE Network*, *20*(3), 41–47.

Zhang, T.-Y., & Ye, D. (2020). False data injection attacks with complete stealthiness in cyber–physical systems: A self-generated approach. *Automatica*, *120*, Article 109117.

Zheng, Y., & Anubi, O. M. (2020). Attack-resilient observer pruning for path-tracking control of wheeled mobile robot. In *Dynamic systems and control conference, Vol. 84287*. American Society of Mechanical Engineers, Article V002T33A001.

Zheng, Y., & Anubi, O. M. (2021). Attack-resilient weighted L1 observer with prior pruning. In *American control conference* (pp. 340–345).

Zheng, Y., & Anubi, O. M. (2022). Resilient observer design for cyber-physical systems with data-driven measurement pruning. In *Security and resilience in cyber-physical systems: Detection, estimation and control* (pp. 85–117). Springer.

Zheng, Y., Sayghe, A., & Anubi, O. (2021). Algorithm design for resilient cyber-physical systems using an automated attack generative model. TechRxiv.

Zou, L., Wang, Z., Hu, J., & Han, Q.-L. (2020). Moving horizon estimation meets multi-sensor information fusion: Development, opportunities and challenges. *Information Fusion*, *60*, 1–10.