2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Model-Unknown Spoofing Attack via False Data Injections

Nachuan Yang, Yuxing Zhong, Yuzhe Li and Ling Shi, *Fellow, IEEE*

***Abstract*— This conference paper studies the spoofing attack via false data injections, where the adversarial attacker aims at misleading a cyber-physical system by distorting its sensor data. Such type of attacks has not been explored in the existing work on false data injection attacks. Besides, the existing research usually assumes an adversary with full knowledge of the target system. In this paper, we consider the case that the attacker does not know the system's parameters. More specifically, we construct an adaptive estimator for the adversary and prove its convergence to the plant's state estimate under adaptive laws. We also show that the convergence of the adaptive estimator is independent of the adversary's strategy. Based on this "separation principle", we propose two false data injection methods to implement online spoofing attacks by solving online linear equations and quadratic programming, respectively, and more can be developed using our proposed adaptive scheme in future research. Finally, we provide a benchmark numerical example of an L-1011 aircraft to illustrate the attack performance.**

I. Introduction

Cyber-physical system has received much attention in the past few years, for its multifarious applications in manufacturing industries, intelligent transportation, and the internet of things (IoT), to name just a few [1]–[3]. Simply speaking, a cyber-physical system is an integration of physical processes and cyber-infrastructures, which is usually monitored and controlled by humans using supervisory control and data acquisition (SCADA) system [3]. Cyber-physical systems are faced with security problems, such as the recent Colonial Pipeline ransomware attack where the computerized equipment managing the pipeline is impacted by attackers [4]. Since the cyber-physical system involves physical processes, its security design must incorporate the physical layer, where system dynamics play an important role. However, this is usually not considered by computer security [5]. This motivates the special focus on the security of cyber-physical systems, especially from the viewpoint of control theory.

A main feature of cyber-security is its diversity caused by various threat models, based on which detection and protection methods can be designed. Typical cyber-threats include denial-of-service (DoS) attacks, false data injection (FDI) attacks, and physical intrusion attacks, which can be summarized by the well-known confidentiality, integrity and availability (CIA) benchmark [6]. The DoS attack aims at degrading the communication channel of cyber-physical systems so that the sensor or controller signals cannot be transmitted. The optimal scheduling of DoS attacks in wireless networked control systems is investigated in [7], where the attack effect on the LQG performance is maximized subject to a resource budget. This problem is further discussed in [8] and [9] where a dynamic attack power allocation model is formulated and solved by Markov decision process (MDP). The DoS attacks are also investigated by [10] and [11] from a viewpoint of game theory. The FDI attack aims at breaking the integrity of cyber-physical systems by injecting fake sensor measurements without being detected [12]. In [13], the FDI attack against state estimation in wireless sensor networks is studied. In [14], a potential class of FDI attacks on electric power market operations is investigated from the viewpoint of economic impact. In [15], the quantitative effect of stealthy integrity attacks on cyber-physical systems is analyzed using a χ^2 failure detector. In [16], a special type of FDI attack called replay attacks is introduced where the cyber-attacker injects an exogenous input while replaying its recorded measurements of system's data. The replay attack and its detector design are further discussed in [17]–[19]. However, in the existing research, the objective of the adversary is usually supposed to degrade the estimation performance, such as increasing the mean squared error [13], which is "unguided". Meanwhile, the adversary is often assumed to have a full knowledge of the target plant. These shortcomings motivate our research.

This paper considers the spoofing attack via false data injections, where the adversary aims at misleading the plant by distorting its sensor data. The contributions of this paper are multi-fold: 1) Different from the existing work on false data injections, we consider a "guided" cyber-attack called spoofing attacks, and assume that the attacker does not know the system's parameters. 2) We construct an adaptive estimator for the adversary and show that its state will converge to the plant's state estimate under adaptive laws. 3) We show that the convergence of the adaptive estimator is independent of attacks, and thereby the "separation principle" holds. 4) We develop two false data injection strategies to implement spoofing attacks by solving online linear equations and quadratic programming, respectively. More can be designed using our adaptive attack scheme in future research.

The remainder of this paper is organized as follows. In Section II, some preliminaries on remote estimation and FDI attack, and the problem formulation are presented. In Section III, we construct an adaptive estimator and prove its convergence under adaptive laws. Then we develop two FDI strategies to implement spoofing attacks. In Section IV, sev-

The work by N. Yang, Y. Zhong and L. Shi is supported by Hong Kong RGC General Research Fund 16211622 (*Corresponding author: Yuzhe Li*).

N. Yang, Y. Zhong and L. Shi are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR (email: nc.yang@connect.ust.hk; yuxing.zhong@connect.ust.hk; eesling@ust.hk).

Y. Li is with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110004, China (email: yuzheli@mail.neu.edu.cn).

eral numerical simulations on an aircraft model are provided. In Section V, the paper is summarized and concluded.

Notation: The notations used throughout this paper are standard. The transpose of matrix $A \in \mathbb{R}^{m\times n}$ is denoted by A^T. The trace of matrix $A \in \mathbb{R}^{n\times n}$ is denoted by $\mathrm{tr}(A)$. The Euclidean or Frobenius norm is denoted by $\|\cdot\|$. The identity matrix is denoted by I. The zero matrix is denoted by 0. For a matrix $A \in \mathbb{R}^{m\times n}$, its element located at the i-th row and the j-th column is denoted by $[A]_{ij}$. The notation $\mathcal{N}(\mu, \Sigma)$ denotes the normal distribution with mean μ and covariance matrix Σ. The notation $A \geq 0$ (respectively, $A > 0$) means that $[A]_{ij} \geq 0$ (respectively, $[A]_{ij} > 0$) for any i and j. The notation $A \succeq 0$ (respectively, $A \succ 0$) means that matrix A is positive semi-definite (respectively, positive definite). The square root of a matrix is denoted by $\sqrt{Q}$, i.e., $\sqrt{Q}\sqrt{Q} = Q$. Matrices and vectors in this paper are assumed to have compatible dimensions if not explicitly stated.

II. Preliminaries

A. State Estimation

Consider a physical plant that is described by a standard continuous-time linear time-invariant system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + w(t) \\ y(t) &= Cx(t) + v(t) \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and $y(t) \in \mathbb{R}^q$ denote the system's state, control input, and measurement. For physical plant, its variables are all bounded. The variables $w(t) \sim \mathcal{N}(0, Q)$ and $v(t) \sim \mathcal{N}(0, R)$ are white Gaussian process and measurement noises with $Q \succeq 0$ and $R \succ 0$. We assume that (A, C) is observable and $(A, \sqrt{Q})$ is controllable.

We are interested in the state estimation scenario where system (1) is equipped with a standard observer

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L(y(t) - C\hat{x}(t)) \tag{2}$$

where $\hat{x}(t) \in \mathbb{R}^n$ denotes the state estimate and $L \in \mathbb{R}^{n\times q}$ is the observer gain matrix such that $A - LC$ is Hurwitz stable. Moreover, matrix L can be determined by

$$L = XC^T R^{-1} \tag{3}$$

where matrix $X \succ 0$ is the unique solution to a continuous-time algebraic Riccati equation (CARE)

$$AX + XA^T + Q - XC^T R^{-1} CX = 0.$$

The state estimator in Eqn (2) with gain matrix (3) is known as the continuous-time Kalman filter [20], which is optimal in the sense of minimum mean square error (MMSE).

B. False Data Injection Attack

In many real-world cyber-physical systems, such as UAVs, IoT and even artificial satellites, the data are collected by remote sensors and transmitted via wireless channels, which makes them prone to adversarial attacks [21].

Suppose system (1) is remotely sensed and estimated by the estimator in Eqn (2), as shown in Fig. 1. In this paper,

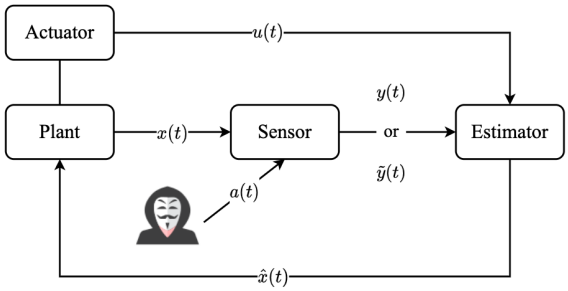

Fig. 1: Diagram of the cyber-physical system (physical plant and remote estimator) and the adversarial attacker.

we mainly consider the false data injection attacks toward the sensor data and the false data is in the form of

$$\tilde{y}(t) = y(t) + a(t) \tag{4}$$

where $\tilde{y}(t) \in \mathbb{R}^q$ is the false sensor data and $a(t) \in \mathbb{R}^q$ is the adversarial data injection, and the amplitude of $a(t)$ is bounded. The adversarial attacker aims at spoofing the system's state estimation (thereby trajectory estimation) by false data injection attacks. For the adversarial model, we make the following assumptions throughout this paper:

1) Plant parameters A, B, C, L, Q, R are unknown.
2) The system's state $x(t)$ is unknown.
3) The estimator's data $\hat{x}(t), u(t), y(t)$ are known.

This assumption is not strong since the estimator's data are remotely transmitted and thereby can be accessed by the adversary via eavesdropping technology [22]. This model-free setup is similar to the recent data-driven control [23].

C. Problem Formulation

The problem investigated in this paper is motivated by the so-called GPS spoofing attack that widely exists in aircraft, vessels and autonomous vehicles [24]. In spoofing attacks, the adversary tries to mislead the target system to believe an unreal state or trajectory by distorting its sensor data.

To model the spoofing attack against system (1), we define a false trajectory that is described by an ODE

$$\dot{\tilde{x}}(t) = F\tilde{x}(t) + G\tilde{u}(t) \tag{5}$$

where $\tilde{x}(t) \in \mathbb{R}^n$ is the false state that the adversary wants to mislead the plant to believe. Matrices F, G and the input $\tilde{u}(t)$ and initial state $\tilde{x}(0)$ are prescribed by the adversary. Therefore Eqn (5) describes the unreal state trajectory that the adversary tries to mislead the plant to believe.

In this paper, we are interested in the spoofing attack via false data injections. Although we study the problem mainly from an adversary's point of view, it is helpful for people to understand and better prevent such attacks. The definition of the problem to be investigated is given as follows.

Problem SAFDI (Spoofing Attack via False Data Injections) Consider the false data injection attack in the form of Eqn (4) against the cyber-physical system in Eqn (1). The adversary can access the data $\hat{x}(t), u(t)$ and $y(t)$ by eavesdropping on the wireless channels but cannot know the plant's state $x(t)$ and parameters A, B, C, L, Q, R. Now

design the false data injection attack series $\{a(t), t \geq 0\}$ such that the adversary can mislead the plant to believe the false trajectory in Eqn (5), i.e., $\hat{x}(t) \to \tilde{x}(t)$ as $t \to \infty$.

III. Main Results

The main difficulty of Problem SAFDI stems from that the adversary cannot access the plant's parameters and thereby cannot implement the spoofing attack. In this section, we will introduce adaptive control techniques to solve this problem and develop an adaptive method for spoofing attacks.

Let us consider the adversarial attack against the remote state estimator for system (1). Under the adversarial attack, the state estimator can be rewritten in the form of

$$\dot{\hat{x}}(t) = A_L\hat{x}(t) + Bu(t) + L\tilde{y}(t) \tag{6}$$

where $A_L := A - LC$ is a Hurwitz matrix. Since the plant's and estimator's parameters are unknown to the adversary, we introduce an adaptive estimator in the following form

$$\begin{aligned}\dot{\bar{x}}(t) = \bar{A}_L(0)\bar{x}(t) + [\bar{A}_L(t) - \bar{A}_L(0)]\hat{x}(t) \\ + \bar{B}(t)u(t) + \bar{L}(t)\tilde{y}(t)\end{aligned} \tag{7}$$

where $\bar{x}(t) \in \mathbb{R}^n$ is the adaptive estimator's state, $\bar{A}_L(t) \in \mathbb{R}^{n\times n}$, $\bar{B}(t) \in \mathbb{R}^{n\times m}$ and $\bar{L}(t) \in \mathbb{R}^{n\times q}$ are time-varying matrices, and $\bar{A}_L(0)$ is a Hurwitz stable matrix.

It is worth noticing that the adaptive estimator in Eqn (7) does not include the plant's parameters A, B, C, and L as well as state $x(t)$. We first give the following definition.

Definition 1: The adaptive estimator in Eqn (7) is called stable if its state $\bar{x}(t)$ asymptotically converges to $\hat{x}(t)$.

To design a stable adaptive estimator via adjusting $\bar{A}_L(t)$, $\bar{B}(t)$ and $\bar{L}(t)$, let us define the following error variables

$$e_x(t) := \bar{x}(t) - \hat{x}(t), \tag{8}$$
$$e_A(t) := \bar{A}_L(t) - A_L, \tag{9}$$
$$e_B(t) := \bar{B}(t) - B, \tag{10}$$
$$e_L(t) := \bar{L}(t) - L. \tag{11}$$

By subtracting Eqn (6) from Eqn (7) and reorganizing the equation, the following error dynamics can be obtained

$$\begin{aligned}\dot{e}_x(t) = \bar{A}_L(0)e_x(t) + e_A(t)\hat{x}(t) \\ + e_B(t)u(t) + e_L(t)\tilde{y}(t).\end{aligned} \tag{12}$$

Notice that the adaptive estimator in Eqn (7) is stable if and only if $e_x(t) \to 0$ as $t \to \infty$. Based on the adaptive control techniques, we propose the following adaptive laws

$$\begin{cases}\dot{\bar{A}}_L(t) = -Pe_x(t)\hat{x}(t)^T \\ \dot{\bar{B}}(t) = -Pe_x(t)u(t)^T \\ \dot{\bar{L}}(t) = -Pe_x(t)\tilde{y}(t)^T\end{cases} \tag{13}$$

where $P \succ 0$ is positive definite, and the initial value $\bar{A}_L(0)$ and matrix P are chosen such that

$$\bar{A}_L(0)^T P + P\bar{A}_L(0) \prec 0. \tag{14}$$

The following theorem shows the stability of the adaptive estimator from the viewpoint of Lyapunov theory.

Theorem 1: Given that $\bar{A}_L(t)$, $\bar{B}(t)$ and $\bar{L}(t)$ satisfy the adaptive laws in Eqns (13) and (14), the adaptive estimator in Eqn (7) is stable, i.e. $\bar{x}(t) \to \hat{x}(t)$ as $t \to \infty$.

Proof. This theorem can be proved by constructing Lyapunov functions for the error system in Eqn (12). Consider the following Lyapunov candidate function

$$\begin{aligned}V(e_x, e_A, e_B, e_L)(t) = e_x(t)^T P e_x(t) + \operatorname{tr}(e_A(t)^T e_A(t)) \\ + \operatorname{tr}(e_B(t)^T e_B(t)) \\ + \operatorname{tr}(e_L(t)^T e_L(t)).\end{aligned}$$

Since P is a positive definite matrix, we have that

$$V(e_x, e_A, e_B, e_L)(t) \geq 0 \tag{15}$$

where the equality holds if and only if all the error variables are zero, therefore $V(e_x, e_A, e_B, e_L)$ is a well-defined Lyapunov function. Further notice the fact that

$$\begin{aligned}\dot{V}(e_x, e_A, e_B, e_L)(t) &= e_x(t)^T(P\bar{A}_L(0) + \bar{A}_L(0)^T P)e_x(t) \\ &+ 2e_x(t)^T P e_A(t)\hat{x}(t) + 2e_x(t)^T P e_B(t)u(t) \\ &+ 2e_x(t)^T P e_L(t)\tilde{y}(t) + 2\operatorname{tr}(\dot{e}_A(t)^T e_A(t) \\ &+ \dot{e}_B(t)^T e_B(t) + \dot{e}_L(t)^T e_L(t))\end{aligned} \tag{16}$$

Given that $\bar{A}_L(t)$, $\bar{B}(t)$ and $\bar{L}(t)$ satisfy the adaptive laws in Eqn (13), meanwhile notice that

$$\dot{e}_A(t) = \dot{\bar{A}}_L(t) \quad \dot{e}_B(t) = \dot{\bar{B}}(t) \quad \dot{e}_L(t) = \dot{\bar{L}}(t), \tag{17}$$

by combining Eqns (16) and (17) we can obtain that

$$\dot{V}(t) = e_x(t)^T(P\bar{A}_L(0) + \bar{A}_L(0)^T P)e_x(t) \leq 0. \tag{18}$$

By Eqn (14) we know that matrix $P\bar{A}_L(0) + \bar{A}_L(0)^T P$ is strictly negative definite. Then the equality in Eqn (18) holds if and only if $e_x(t) = 0$. By Eqn (18), we have that

$$-\int_{t=0}^{\infty} e_x(t)^T(P\bar{A}_L(0) + \bar{A}_L(0)^T P)e_x(t)dt \leq V(0) \tag{19}$$

and thereby the L_2 norm of $e_x(t)$ is bounded. Since $u(t)$ and $a(t)$ are bounded, $\dot{e}_x(t)$ is also bounded and thus $e_x(t) \to 0$ as $t \to \infty$, i.e., the adaptive estimator is stable. □

Remark 1: The above theorem shows that the adaptive estimator is stable under the adaptive laws. It is worth noticing that its stability is independent of the control input $u(t)$ and the data injection $a(t)$. Therefore the adaptive estimator and the spoofing attack can be implemented separately.

Corollary 1: The error variables $e_x(t)$, $e_A(t)$, $e_B(t)$ and $e_L(t)$ are bounded for any $t \geq 0$. Given the initial condition that $\bar{x}(0) = \hat{x}(0)$, then it holds that

$$\begin{aligned}||\bar{x}(t) - \hat{x}(t)||_P^2 \leq ||\bar{A}_L(0) - A_L||^2 + ||\bar{B}(0) - B||^2 \\ + ||\bar{L}(0) - L||^2\end{aligned} \tag{20}$$

for any $t \geq 0$, where $\|\cdot\|_P$ denotes the weighted norm.

Proof. By the results in Theorem 1, we can obtain that

$$V(e_x, e_A, e_B, e_L)(t) \leq V(e_x, e_A, e_B, e_L)(0) \tag{21}$$

which follows from the fact that $\dot{V}(t) \leq 0$ for any $t \geq 0$. Since the right-hand side is a constant, $V(e_x, e_A, e_B, e_L)(t)$

is always bounded and therefore $e_x(t)$, $e_A(t)$, $e_B(t)$ and $e_L(t)$ are all bounded. If $\bar{x}(0) = \hat{x}(0)$, we have

$$\begin{aligned} e_x(t)^T P e_x(t) &\leq V(e_x, e_A, e_B, e_L)(t) \leq V(\dots)(0) \\ &\leq \|e_A(0)\|^2 + \|e_B(0)\|^2 + \|e_L(0)\|^2 \end{aligned}$$

and therefore Eqn (20) holds for any $t \geq 0$. □

Remark 2: It is worth noticing that the adversary's adaptive estimator, as shown in Eqn (7), has not used the plant's parameters. The above corollary proves an intuitive result, that is, the transient error of the adaptive estimator is smaller when the adversarial attacker's initial guess is more accurate, and finally this error will converge to zero.

In what follows, we will develop an adaptive false data injection strategy to solve Problem SAFDI. As we already show that, the error $e_x(t) = \bar{x}(t) - \hat{x}(t)$ will always converge to zero. Hence the adversarial attack can be designed to regulate the adaptive estimator's state $\bar{x}(t)$ to asymptotically converge to the false state $\tilde{x}(t)$ in Eqn (5). In this case, the plant's state estimate $\hat{x}(t)$ will also converge to the unreal trajectory $\tilde{x}(t)$ and thereby the plant will be spoofed. The following corollary gives an attack strategy in simple cases.

Corollary 2: Given that $\bar{A}_L(0) = F$ is Hurwitz stable and the adaptive laws in Eqn (13) hold, if $a(t)$ satisfies

$$[\bar{A}_L(t) - \bar{A}_L(0)]\hat{x}(t) + \bar{B}(t)u(t) + \bar{L}(t)\tilde{y}(t) = G\tilde{u}(t)$$

as $t \to \infty$, then the adaptive estimator is stable and the state estimate $\hat{x}(t) \to \tilde{x}(t)$ as $t \to \infty$, i.e., the plant is spoofed.

Proof. Substituting the above equation into Eqn (7), the adaptive estimator's dynamics become

$$\dot{\bar{x}}(t) = \bar{A}_L(0)\bar{x}(t) + G\tilde{u}(t) \tag{22}$$

as $t \to \infty$. Since matrix $\bar{A}_L(0) = F$ is Hurwitz stable and the adaptive laws hold, the adaptive estimator is stable and $\hat{x}(t) \to \bar{x}(t)$ as $t \to \infty$. Combining Eqns (22) and (5), we obtain the error dynamics as $\dot{e}(t) = Fe(t)$ with $e(t) = \bar{x}(t) - \tilde{x}(t)$, and therefore $\bar{x}(t) \to \tilde{x}(t)$ as $t \to \infty$. Finally we can conclude that $\hat{x}(t)$ asymptotically converges to $\tilde{x}(t)$ and thereby the plant will be spoofed. □

The above method involves solving online linear equations, which can only be applied when $q \geq n$ since otherwise the conditions may be infeasible. To include more general plant dynamics, we further propose a receding horizon approach to design the false data injection attacks.

Following the idea of receding horizon control (RHC) [25], the false data injection strategy can be designed over a finite horizon $T > 0$. By *Theorem 1* and the adaptive laws in Eqn (13), $\bar{x}(t) \to \hat{x}(t)$ as $t \to \infty$ and the changes of $\bar{A}_L(t), \bar{B}(t), \bar{L}(t)$ will also asymptotically converge to zero. Therefore we approximate the dynamics in Eqn (2) by

$$\dot{\bar{x}}(t) = \bar{A}_L(kT)\bar{x}(t) + \bar{B}(kT)u(t) + \bar{L}(kT)\tilde{y}(t) \tag{23}$$

for $t \in [kT, (k+1)T)$, where the parameters are only updated according to Eqn (13) at time instants kT for $k \geq 0$. For numerical implementations, the above dynamics can be discretized over intervals such as $[t, t+1]$ as

$$\bar{x}(t+1) = (I + \bar{A}_L(kT))\bar{x}(t) + \bar{B}(kT)u(t) + \bar{L}(kT)\tilde{y}(t)$$

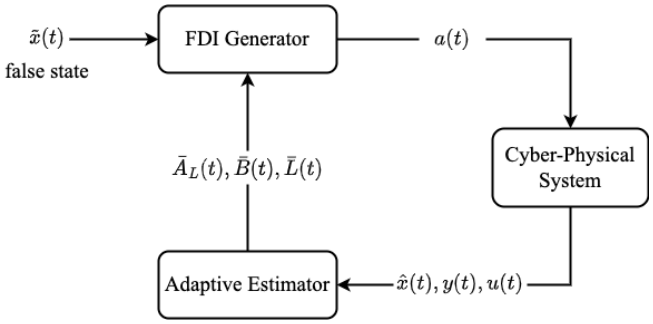

Fig. 2: Structure of the proposed SAFDI algorithm, where the attack is onlinely computed by the FDI generator.

for $t \in [kT, (k+1)T - 1)$. Sometimes the plant's control inputs cannot be predicted by the adversary. In this case, we assume $u(t) = u(kT)$ over the interval $[kT, (k+1)T - 1)$.

In what follows, we develop a SAFDI algorithm based on RHC, which can be described by an online optimization

$$\min_{a(t), t=kT, kT+1, \dots, (k+1)T-1} J(\bar{x}(t), \tilde{y}(t))$$

subject to

$$\begin{aligned} &\bar{x}(t+1) = (I + \bar{A}_L(kT))\bar{x}(t) + \bar{B}(kT)u(t) + \bar{L}(kT)\tilde{y}(t) \\ &\bar{x}(t) \in \mathcal{X},\ \tilde{y}(t) \in \mathcal{Y},\ t = kT, kT+1, \dots, (k+1)T-1 \end{aligned}$$

with the following updates on the parameters

$$\begin{cases} \bar{A}_L((k+1)T) = \bar{A}_L(kT) - TPe_x(kT)\hat{x}(kT)^T \\ \bar{B}((k+1)T) = \bar{B}(kT) - TPe_x(kT)u(kT)^T \\ \bar{L}((k+1)T) = \bar{L}(kT) - TPe_x(kT)\tilde{y}(kT)^T \end{cases}$$

which is a direct application of forward Euler method. It is worth noticing that both the objective function J and sets $\mathcal{X}, \mathcal{Y}$ can be used to constrain the attacks. For example,

$$J(\cdot) = \sum_{t=kT}^{(k+1)T-1} \|\bar{x}(t) - \tilde{x}(t)\|^2 + w\|\Delta\tilde{y}(t)\|^2 \tag{24}$$

where $\Delta\tilde{y}(t) := \tilde{y}(t+T) - \tilde{y}(t)$ and w can be used to tune the smoothness of the spoofing attack. In this case, the SAFDI algorithm is reduced to quadratic programming that can be computed by numerical solvers such as SeDuMi [26].

Remark 3: The above SAFDI algorithm is similar to the setup of RHC and the main difference is the update of time-varying parameters, which follows from the discretization of adaptive laws. Due to the existence of numerical approximations, the algorithm's theoretical performance is difficult to analyze and instead will be illustrated by simulations.

Both the above RHC method and the method introduced in *Corollary 2* are based on a "separation principle", i.e., the stability of adaptive estimator (7) is independent of the false data injections $a(t)$, and more algorithms can be developed using this methodology. Meanwhile, it is also worth noticing the difference between SAFDI and system identification. In a SAFDI algorithm, it is not necessary for the adversary to identify the plant's parameters, which is also unpractical due to the persistent excitation condition [27].

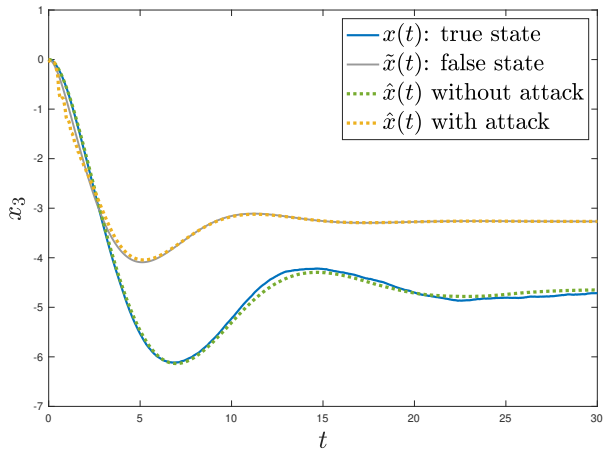

Fig. 3: System dynamics under the SAFDI in *Corollary 1*

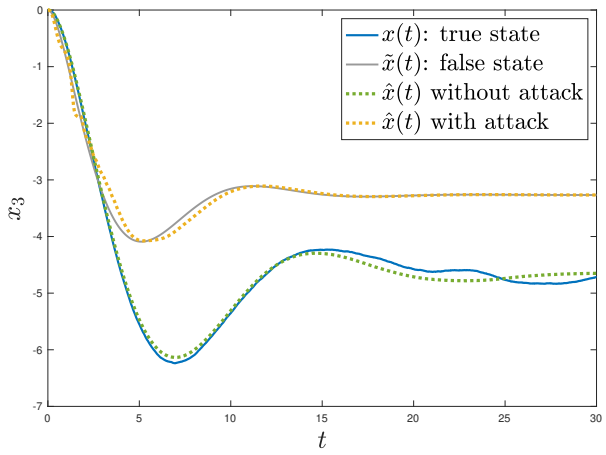

Fig. 4: System dynamics under RHC-based SAFDI ($w = 1$)

IV. Simulations

In this section, we will evaluate the performance of our proposed SAFDI algorithms through the benchmark example (AC47) in COMPleib 1.0 [28]. This example consists of a model of the lateral dynamics for an L-1011 aircraft that can be described by system (1) with the following parameters

$$A = \begin{bmatrix} -2.98 & 0.93 & 0 & -0.034 \\ -0.99 & -0.21 & 0.035 & -0.0011 \\ 0 & 0 & 0 & 1 \\ 0.39 & -5.555 & 0 & -1.89 \end{bmatrix} B = \begin{bmatrix} -0.032 \\ 0 \\ 0 \\ -1.6 \end{bmatrix}$$

and here we consider two cases:

- full-state measurement $C = I$
- partial-state measurement $C = \begin{bmatrix} 0 & I \end{bmatrix}$.

Throughout this example, we assume that the adversary is mainly interested in spoofing the estimate for state $x_3(t)$.

The noise covariance is simply set as $Q = 0.01I$ and $R = 0.001I$. We consider a constant input $u(t) = 1$ for $t \geq 0$. The reference system's parameters in Eqn (5) can be chosen based on practical experience and here are set as

$$F = \begin{bmatrix} -3 & 1 & 0 & -0.1 \\ -1 & -0.2 & 0.1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & -6 & 0 & -2 \end{bmatrix} \quad G = \begin{bmatrix} -0.1 \\ 0 \\ 0 \\ -2 \end{bmatrix}$$

and we simply set $\tilde{x}(0) = 0$ and $\tilde{u}(t) = 1.2$ for $t \geq 0$. As mentioned in Section II, the adversary cannot access the values of the physical plant's parameters. The spoofing attack is designed based on the eavesdropped data only.

We first evaluate *Corollary 2* for the full-state measurement case. In this case, the adversary designs the spoofing attack by solving online linear equations. In the numerical test, we implement it using MATLAB solver $linsolve()$, and the simulated trajectories and the false data injections are shown in Fig. 2. It shows that the adversary successfully mislead the plant's state estimate to a prescribed false trajectory, provided that the plant's parameters are unknown.

Then we consider the partial-state measurement case. In this case, the method in *Corollary 2* fails to work due to the

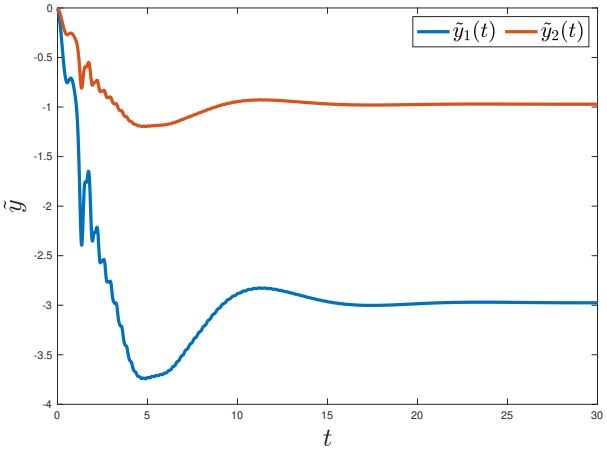

Fig. 5: False sensor data by RHC-based SAFDI ($w = 1$)

singularity of linear equations, and instead, we implement the RHC-based SAFDI algorithm, where the adversary generates false data injections by solving online optimization. Since x_3 is the target, the objective function can be defined as

$$J = \sum_{t=kT}^{(k+1)T-1} \|\bar{x}_3(t) - \tilde{x}_3(t)\|^2 + w\|\Delta\tilde{y}(t)\|^2$$

where we set $w = 1$ and $T = 0.2$. In the numerical test, we implement the SAFDI algorithm using MATLAB solver $quadprog()$, and the simulated trajectories and the false sensor data are shown in Fig. 5 and Fig. 6. It shows that even if the adversary can only eavesdrop on the partial-state measurements, the spoofing attack can still be implemented by using the RHC-based SAFDI algorithm.

Compared with *Corollary 2*, the RHC-based algorithm is more flexible and can achieve the desired performance by choosing appropriate objective functions or constraints. For example, if we set $w = 10$, the generated false sensor data will be more smooth and thus "realistic", as shown in Fig. 7. However, as shown in Fig. 6, the plant's state estimate will converge to the false trajectory with a slower speed.

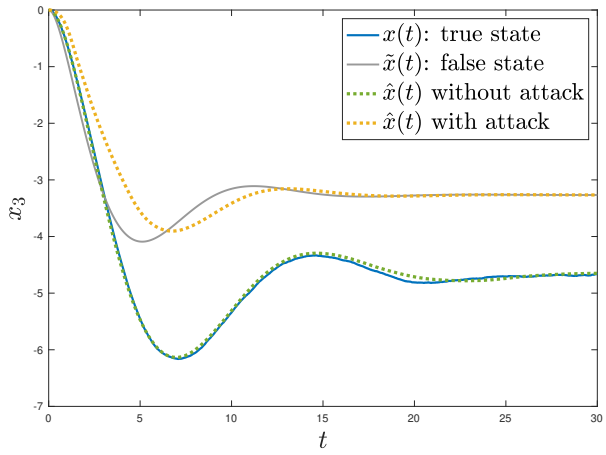

Fig. 6: System dynamics under RHC-based SAFDI ($w = 10$)

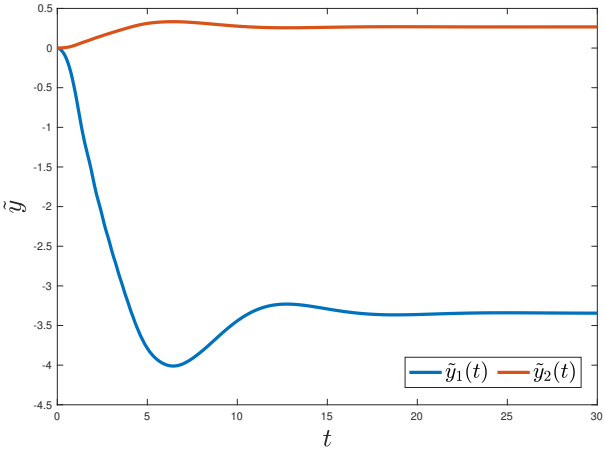

Fig. 7: False sensor data by RHC-based SAFDI ($w = 10$)

V. Conclusion

In this paper, we considered the spoofing attack via false data injections, where the adversary is distorting the sensor data but does not know the plant's parameters. To mislead the plant to believe a false trajectory, we proposed an adaptive estimator whose state will asymptotically converge to the plant's state estimate under adaptive laws. We showed that the convergence of the adaptive estimator is independent of the attack sequence, and thereby the "separation principle" holds. Then we developed two SAFDI algorithms to implement online spoofing attacks, which are based on solving online linear equations and quadratic programming, respectively. Finally, several numerical simulations were provided to illustrate the performance of our proposed scheme. Future work may focus on extending our proposed adaptive scheme to design more general and advanced attacks.

References

[1] S. Kim and S. Park, "CPS based manufacturing system optimization," *Procedia Computer Science*, vol. 122, pp. 518–524, 2017.

[2] K. Evers, R. Oram, S. El-Tawab, M. H. Heydari, and B. B. Park, "Security measurement on a cloud-based cyber-physical system used for intelligent transportation," in *Proceedings of IEEE Conference on Vehicular Electronics and Safety*. IEEE, 2017, pp. 97–102.

[3] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber–physical systems," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 13–28, 2011.

[4] J. R. Reeder and C. T. Hall, "Cybersecurity's pearl harbor moment: Lessons learned from the colonial pipeline ransomware attack," 2021.

[5] K. Daimi, G. Francia, L. Ertaul, L. H. Encinas, and E. El-sheikh, *Computer and Network Security Essentials*. Springer, 2018.

[6] M. S. Chong, H. Sandberg, and A. M. Teixeira, "A tutorial introduction to security and privacy for cyber-physical systems," in *Proceedings of European Control Conference*. IEEE, 2019, pp. 968–978.

[7] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal DoS attack scheduling in wireless networked control system," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 843–852, 2015.

[8] H. Zhang, Y. Qi, J. Wu, L. Fu, and L. He, "DoS attack energy management against remote state estimation," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 383–394, 2016.

[9] J. Qin, M. Li, J. Wang, L. Shi, Y. Kang, and W. X. Zheng, "Optimal denial-of-service attack energy management against state estimation over an SINR-based network," *Automatica*, vol. 119, p. 109090, 2020.

[10] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, "Jamming attacks on remote state estimation in cyber-physical systems: a game-theoretic approach," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2831–2836, 2015.

[11] Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "SINR-based DoS attack on remote state estimation: a game-theoretic approach," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 3, pp. 632–642, 2016.

[12] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber–physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2011.

[13] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Proceedings of IEEE Conference on Decision and Control*. IEEE, 2010, pp. 5967–5972.

[14] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, 2011.

[15] Y. Mo and B. Sinopoli, "On the performance degradation of cyber-physical systems under stealthy integrity attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2618–2624, 2015.

[16] ——, "Secure control against replay attacks," in *Proceedings of 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2009, pp. 911–918.

[17] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *Proceedings of 52nd IEEE Conference on Decision and Control*. IEEE, 2013, pp. 1854–1859.

[18] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015.

[19] C. Fang, Y. Qi, P. Cheng, and W. X. Zheng, "Optimal periodic watermarking schedule for replay attack detection in cyber–physical systems," *Automatica*, vol. 112, p. 108698, 2020.

[20] F. L. Lewis, L. Xie, and D. Popa, *Optimal and Robust Estimation: with an Introduction to Stochastic Control Theory*. CRC Press, 2017.

[21] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of CPS security," *Annual Reviews in Control*, vol. 47, pp. 394–411, 2019.

[22] X. Sun, D. W. K. Ng, Z. Ding, Y. Xu, and Z. Zhong, "Physical layer security in UAV systems: challenges and opportunities," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 40–47, 2019.

[23] C. De Persis and P. Tesi, "Formulas for data-driven control: stabilization, optimality, and robustness," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 909–924, 2019.

[24] A. Janofsky, "How to defend against GPS spoofing attacks," *Wall Street Journal*, 2018.

[25] F. Borrelli, A. Bemporad, and M. Morari, *Predictive Control for Linear and Hybrid Systems*. Cambridge University Press, 2017.

[26] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[27] L. Ljung, *System Identification*. Springer, 1998.

[28] F. Leibfritz and W. Lipinski, "Description of the benchmark examples in COMPleib 1.0," *Dept. Math., Univ. Trier, Trier, Germany, Tech. Rep*, vol. 32, 2003.