Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-40 (2022) 259–264

Optimal Stealth Attacks to Cyber-Physical Systems: Seeking a Compromise between Maximum Damage and Effort ⋆

Luca Faramondi *,** **Gabriele Oliva** * **Roberto Setola** *

* *University Campus Bio-Medico of Rome, Via A. Del Portillo 21, 00128, Rome, Italy.*

** *Corresponding author (e-mail: l.faramondi@unicampus.it)*

Abstract: This paper aims at modeling the optimal behavior for an attacker that has the objective to maliciously manipulate the output of an industrial power plant, which is fed via a public network to a digital twin in order to reconstruct the state. In particular, we consider a scenario where the attacker seeks a tradeoff between two conflicting objectives: dealing the maximum damage in terms of the norm of the estimation error for the observer and keeping the magnitude of the variation of the systems' output to a minimum. In doing so, we assume the observer is equipped with a bad data detector and the attacker must choose the injected signals in a way that guarantees that the bad data detection condition is not triggered.

Keywords: Cyber Attack, Cyber Security, Cyber-Physical Systems, Critical Infrastructures, Operational Technologies

1. INTRODUCTION

The widespread adoption of network-based remote control solutions has exposed industrial systems, such as plants or infrastructures, to unprecedented cyber-security risks that could jeopardize their proper operation. Cyber threats remain one of the key weaknesses for industrial plants and SCADA systems, despite the deployment of countermeasures based on network traffic monitoring for intrusion detection. Several malwares targeting industrial control systems have been developed in the previous ten years, the most notable of which being Stuxnet (2010), Havex (2014), Black Energy (2015), Clash Override (2016), and Triton (2017) (see Setola et al. (2019)). Stuxnet is a piece of malware that can recognize a specific industrial controller and change the rotational speed of some motors. The malware Havex is used as a backend for more elaborate assaults since it can scan the network using genuine OPC protocol functionality to detect network devices and their addresses. Black Energy may gather information about industrial processes by monitoring the Human-Machine Interface (HMI) and extracting pertinent data from the graphical representation of the plant and the operations conducted by SCADA operators. Crash Override can inject legitimate directives into the power network, changing its behavior in order to cause a blackout. Finally, Triton is intended for SIS (Safety Instrumental System), which are specialized industrial control systems, distinct from general-purpose systems, that are meant to avert catastrophic incidents. The most successful attack was on the Colonial pipeline in May 2021, which shut down the 2200 km gasoline and jet fuel pipeline for more than a week Hobbs (2021).

According to Cardenas and Sastry (2008), there are two types of threats that can affect measurement and actuator data by intercepting the communication channel: *block* (intended to reduce data availability) and *deception* (intended to deceive) (which aim is to compromise data integrity). The first, which is usually based on DoS (Denial of Service) attacks, can disrupt the interchange of information between the control center and the Remote Terminal Units (RTUs), causing the plant to degrade or even shut down, according to Gupta and Basar (2010). In Zhang et al. (2015), DoS attacks on state estimators are carried out using jamming strategies in order to increase estimate error, whereas in Li et al. (2015), a game-theoretic framework is offered in order to detect jamming assaults.

On the other hand, an attack that compromises the integrity of the data shared has the potential to have far more serious implications, such as forcing the plant to operate under hazardous conditions. Because some industrial protocols lack authentication and cryptography mechanisms, black attacks are quite straightforward to carry out (e.g. Modbus). Simple attacks rely on a replay attack, in which the attacker gains access to, records, and replays sensor data (Miciolino et al. (2017); Mo and Sinopoli (2012)). However, in the presence of a state estimator, such attacks can be easily identified, as demonstrated by Zhao et al. (2017), which uses an extended Kalman filter to identify bias injection and replay attacks.

In this paper we model the optimal behavior for an attacker that aims to maliciously manipulate the output of an industrial power plant, modeled as a linear system, which is fed via a public network to an observer in order to reconstruct the state. In particular, we consider a scenario

⋆ This work was supported by Italian National Project "DRIVERS: Approccio combinato data-driven ed experience-driven all'analisi del rischio sistemico" funded by INAIL under grant n. BRIC2021-ID03.

10.1016/j.ifacol.2023.01.082

where the attacker seeks a tradeoff between two conflicting objectives: dealing the maximum damage in terms of the norm of the estimation error for the observer and keeping the magnitude of the variation of the systems' output to a minimum. This latter requirement allows to prevent BDD (Bad Data Detection) strategies from identify anomalies and consequently to discovering the attack. Hence, the attacker must choose the injected signals in a way that guarantees that the bad data detection condition is not triggered, while attempting to maximize the damage dealt.

1.1 State of the Art

Cyber attacks based on measurements or actuator signal modifications are investigated in Fawzi et al. (2014); in particular, the authors characterize the maximum number of attacks that can be detected and corrected, providing a necessary condition for state reconstruction in the event of multiple attacks. Data injection attacks against state estimators are discussed in Mo et al. (2010), in which false signals are injected to threaten the integrity of the Kalman filter used to accomplish state estimation.

The work described in Wu et al. (2018) discusses the construction of an optimal input injection technique based on switching attacks, in which the adversary seeks to maximize the deviation of the plant's state from nominal condition by using appropriate input injection and switching locations. In Teixeira et al. (2012), an attack approach based on zero-dynamics and covert data injection attacks on control systems is explored. In particular, the authors of Teixeira et al. (2012) characterize and analyze the stealthiness properties of attacks on linear time-invariant systems and provide detection strategies based on system structure modification; additionally, necessary and sufficient conditions for the identification of zero-dynamics attacks are provided.

The behavior of a malicious attacker attempting to compromise linear state estimators is investigated in Jovanov and Pajic (2019); in the paper, the authors consider a system with a Kalman filter-based state estimator and residual-based intrusion detectors. They examine the effects of intermittent data integrity improvements, such as the usage of message authentication codes, for such systems. The design of a robust state estimation scheme for continuous-time linear dynamical systems is addressed in Lee et al. (2015); thanks to sensor redundancy, the method correctly estimates the states under sensor attacks and guarantees a bounded estimation error despite measurement noises and process disturbances. An attacker uses the Ren et al. (2018) to increase the error of a multi-agent system's state estimator by introducing noise into a portion of the network's communication channels. In Faramondi et al. (2021b) a scenario is considered where the attacker aims at dealing the largest damage without being caught, but the attacker does not attempt to minimize the magnitude of the injected signals.

Notice that, to test such scheme in a more realistic setting, several data sets have been produced in the literature, e.g., see Laso et al. (2017); Faramondi et al. (2021a) and references therein.

In this paper, we consider a setting where the attacker seeks a tradeoff between two conflicting objectives: dealing the maximum damage in terms of the norm of the estimation error for the observer and keeping the magnitude of the variation of the systems' output to a minimum. In more detail, in this paper we formalize the problem at hand in terms of an optimization formulation and we develop a necessary and a sufficient local optimality condition. Then, we provide an example to numerically demonstrate the effectiveness of such an attack strategy.

2. PRELIMINARIES

2.1 Notation and Definitions

We denote vectors by boldface lowercase letters and matrices with uppercase letters. We refer to the (i, j)-th entry of a matrix A by A_{ij}. We represent by $\mathbf{0}_n$ and $\mathbf{1}_n$ vectors with n entries, all equal to zero and to one, respectively. Moreover, we denote by I_n the $n \times n$ identity matrix. The set of positive real numbers is denoted by $\mathbb{R}_{>0}$.

2.2 Nonconvex Optimization

In this paper, we consider optimization problems in the form

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathbb{R}^d} f(\boldsymbol{x}) \\ &\text{Subject to} \\ &\{g(\boldsymbol{x}) \leq 0, \end{aligned} \tag{1}$$

where f, g are continuous, differentiable and, in general, nonconvex scalar functions. Let us now review the Karush-Kuhn-Tucker (KKT) first order necessary conditions for local optimality; the reader is referred to Floudas (1995) for a comprehensive overview of the topic.

Theorem 1. (KKT First Order Necessary Condition). Let us consider a constrained optimization problem as in Eq. (1) and let the *Lagrangian function* $\mathscr{L}(\cdot)$ be defined as

$$\mathscr{L}(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x}),$$

where $\lambda \in \mathbb{R}$ is the Lagrangian multiplier that corresponds to the constraint $g(\boldsymbol{x}) \leq 0$. If the point $\boldsymbol{x}^* \in \mathbb{R}^d$ is a *local minimum* for the problem then there exist a $\lambda^* \in \mathbb{R}$ such that the following conditions hold true:

$$\nabla_{\boldsymbol{x}} \mathscr{L}(\boldsymbol{x}, \lambda)\Big|_{\boldsymbol{x}=\boldsymbol{x}^*,\, \lambda=\lambda^*} = 0, \tag{2}$$

$$\lambda^* g(\boldsymbol{x}^*) = 0 \tag{3}$$

$$g(\boldsymbol{x}^*) \leq 0 \tag{4}$$

$$\lambda^* \geq 0. \tag{5}$$

We now review the KKT second order sufficient conditions for local optimality (see Floudas (1995) and references therein).

Theorem 2. (KKT Second Order Sufficient Condition).

Let us consider a constrained optimization problem as in Eq. (1) and suppose that there is a point $\boldsymbol{x}^* \in \mathbb{R}^d$, together with a Lagrangian multiplier $\lambda^* \in \mathbb{R}$, satisfying the KKT first order necessary condition. Let $\mathcal{D}$ be defined as the set of all $\boldsymbol{d} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$ such that:

$$\nabla_{\boldsymbol{x}} g(\boldsymbol{x}^*)^T \boldsymbol{d} = 0 \quad \text{if } \lambda^* > 0, \tag{6}$$

$$\nabla_{\boldsymbol{x}} g(\boldsymbol{x}^*)^T \boldsymbol{d} \leq 0 \quad \text{if } \lambda^* = 0 \tag{7}$$

If for all $\boldsymbol{d} \in \mathcal{D}$ it holds

$$\boldsymbol{d}^T \nabla_{\boldsymbol{x}\boldsymbol{x}} \mathscr{L}(\boldsymbol{x}, \lambda)\Big|_{\boldsymbol{x}=\boldsymbol{x}^*,\, \lambda=\lambda^*} \boldsymbol{d} \geq 0$$

then $\boldsymbol{x}^*$ is a local minimum. If the above equation holds as a strict inequality, then $\boldsymbol{x}^*$ is a strict local minimum.

Corollary 1. If the Hessian matrix $\mathscr{L}(\boldsymbol{x}, \lambda)\Big|_{\boldsymbol{x}=\boldsymbol{x}^*,\, \lambda=\lambda^*}$ of the Lagrangian function is positive semi-definite (definite), then $\boldsymbol{x}^*$ is a local minimum (strict local minimum).

3. PROBLEM SETTING

In this section, we start by briefly reviewing the setting introduced in Faramondi et al. (2021b) and we present the proposed optimization formulation. In particular, let

$$\begin{cases} \boldsymbol{x}[k+1] & = A\boldsymbol{x}[k] + B\boldsymbol{u}[k], \\ \boldsymbol{y}[k] & = C\boldsymbol{x}[k] \end{cases}$$

denote the discrete-time system dynamics of a plant, with $\boldsymbol{x}[k] \in \mathbb{R}^n$, $\boldsymbol{u}[k] \in \mathbb{R}^p$, $\boldsymbol{y}[k] \in \mathbb{R}^q$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$ and $C \in \mathbb{R}^{q\times n}$. Moreover, assume that the pair (A, C) is observable. In a networked control scenario, we assume that a control center (e.g., a SCADA system) receives the output $\boldsymbol{y}[k]$ and sends the input $\boldsymbol{u}[k]$ through a communication network, which might be prone to cyber attacks. Let us assume that the control center implements a Luenberger-like observer to reconstruct the state of the plant Luenberger (1971), i.e.,

$$\begin{cases} \boldsymbol{z}[k+1] & = A\boldsymbol{z}[k] + B\boldsymbol{u}[k] + L(\boldsymbol{y}[k] - \boldsymbol{y}_z[k]), \\ \boldsymbol{y}_z[k] & = C\boldsymbol{z}[k]. \end{cases}$$

Moreover, the SCADA is equipped with a BDD to discovery anomalous measurements based on the analysis of the *residue* $\boldsymbol{r}[k]$, i.e.,

$$\boldsymbol{r}[k] = \boldsymbol{y}'[k] - \boldsymbol{y}_z[k]; \tag{8}$$

where $\boldsymbol{y}'[k]$ is the measure received from the field. We consider a scenario where an alarm is triggered whenever the condition $\|\boldsymbol{r}[k]\| > \theta$ is met, $\theta \in \mathbb{R}_{>0}$ being a threshold that is tailored to the particular plant.

Let us now discuss the problem at hand in this paper. Consider an attacker aiming to deal an attack that causes a discrepancy between the real state and the one estimated by the SCADA's observer, without triggering the alarm condition. In particular, while the requirement not to trigger the BDD is represented as a constraint, we assume the attacker has two conflicting objectives: maximizing the discrepancy and minimizing the magnitude of the injected signal. In this view, we assume that the attacker is able to replace the real output value $\boldsymbol{y}[k]$ with a maliciously altered value

$$\boldsymbol{y}'[k] = \boldsymbol{y}[k] + \boldsymbol{y}_a[k].$$

As a consequence of the attack, the dynamics of the observer becomes

$$\begin{cases} \boldsymbol{z}[k+1] & = A\boldsymbol{z}[k] + B\boldsymbol{u}[k] + L(\boldsymbol{y}'[k] - \boldsymbol{y}_z[k]), \\ \boldsymbol{y}_z[k] & = C\boldsymbol{z}[k]. \end{cases}$$

Therefore, the estimation error $\boldsymbol{e}[k] = \boldsymbol{x}[k] - \boldsymbol{z}[k]$ is

$$\boldsymbol{e}[k+1] = \boldsymbol{x}[k+1] - \boldsymbol{z}[k+1] = (A - LC)\boldsymbol{e}[k] - L\boldsymbol{y}_a[k]. \tag{9}$$

Notice that $\boldsymbol{e}[k+1]$ can be expressed in terms of $\boldsymbol{e}[k]$ and $\boldsymbol{y}_a[k]$ via Eq. (9); similarly, by using Eq. (8) we have that

$$\boldsymbol{r}[k+1] = \boldsymbol{y}'[k+1] - \boldsymbol{y}_z[k+1] = C\boldsymbol{e}[k] + \boldsymbol{y}_a[k].$$

Notably, in typical working conditions, it can be assumed that sufficient time has elapsed so that the transients of the SCADA's observer have vanished; in this case, the attacker may assume that $\boldsymbol{e}[0] = \boldsymbol{0}_n$. Consequently, the attacker is able to calculate $\boldsymbol{e}[k]$ as follows

$$\boldsymbol{e}[k] = -\sum_{m=0}^{k-1} (A - LC)^{k-m-1} L\boldsymbol{y}_a[m]. \tag{10}$$

In Faramondi et al. (2021b), the attacker's problem is formulated expressing $\boldsymbol{y}_a^*[k]$ in terms of a new variable $\boldsymbol{w}[k]$ defined as

$$\boldsymbol{w}[k] = \frac{1}{\theta}\boldsymbol{r}[k+1] = \frac{1}{\theta}\left(C\boldsymbol{e}[k] + \boldsymbol{y}_a[k]\right).$$

hence the condition $\|\boldsymbol{r}[k+1]\| \leq \theta$ is equivalent to $\|\boldsymbol{w}[k]\| \leq 1$. Moreover, we have that

$$\boldsymbol{e}[k+1] = A\boldsymbol{e}[k] - \theta L\boldsymbol{w}$$

and

$$\boldsymbol{y}_a^*[k] = \theta\boldsymbol{w}^*[k] - C\boldsymbol{e}[k], \tag{11}$$

Based on the above simplifications, the attacker problem is defined as follows.

Problem 1. Let $A, B, C, L, \theta, \boldsymbol{u}[k], \boldsymbol{y}[k], \boldsymbol{e}[k]$, $\alpha \geq 0$ be given. Find $\boldsymbol{y}_a^*[k] \in \mathbb{R}^q$ such that $\boldsymbol{y}_a^*[k] = \theta\boldsymbol{w}^*[k] - C\boldsymbol{e}[k]$, where

$$\boldsymbol{w}^*[k] = \underset{\boldsymbol{w}\in\mathbb{R}^q}{\arg\max}\, \frac{1}{2}\left\|A\boldsymbol{e}[k] - \theta L\boldsymbol{w}\right\|^2 - \alpha\|\theta\boldsymbol{w}^*[k] - C\boldsymbol{e}[k]\|^2$$

$$\text{Subject to, } \|\boldsymbol{w}\|^2 \leq 1. \tag{12}$$

Remark 1. The above problem always admits a feasible solution. In fact it can be noted that $\boldsymbol{w} = \boldsymbol{0}_q$ trivially satisfies the constraint.

4. LOCAL OPTIMALITY CONDITIONS

In this section, we develop a necessary local optimality condition and a sufficient local optimality condition for the problem at hand.

Let us first consider a local optimality condition.

Theorem 3. Let us assume that $\lambda_{min}(L^TL) > \alpha$. A necessary condition for $\boldsymbol{w}[k]$ to be a locally optimal solution for Problem 1 is that $\|\boldsymbol{w}^*[k]\| = 1$ and

$$P(\boldsymbol{w}^*[k])\boldsymbol{w}^*[k] = (L^TA - \alpha C)\boldsymbol{e}[k],$$

where

$$P(\boldsymbol{w}) = \theta L^TL + \left(-\alpha\theta + \boldsymbol{w}^TL^TA\boldsymbol{e}[k] - \theta\boldsymbol{w}^TL^TL\boldsymbol{w} - \alpha\boldsymbol{w}^*[k]^TC\boldsymbol{e}[k]\right)I_m. \tag{13}$$

Proof 1. In order to prove the result, we claim that any local maximum $\boldsymbol{w}^*$ must satisfy $\|\boldsymbol{w}^*\| = 1$. In view of a contradiction, let us assume that there is as local maximum $\boldsymbol{w}^*$ with $\|\boldsymbol{w}^*\| < 1$. Since the objective function is convex and $\boldsymbol{w}^*$ is a local maximum, we have that any $\boldsymbol{w}^\dagger$ in an infinitesimal neighborhood of $\boldsymbol{w}^*$ with $\|\boldsymbol{w}^\dagger\| > \|\boldsymbol{w}^*\|$ must have larger objective function than $\boldsymbol{w}^*$. At this point we observe that, for $\boldsymbol{w}^*$ to be a local maximum, any of such points $\boldsymbol{w}^\dagger$ must violate the constraint, i.e., it must

hold $\|\boldsymbol{w}^\dagger\| > 1$. However, since we assumed $\|\boldsymbol{w}^*\| < 1$, we have that at least one of such points exists with $\|\boldsymbol{w}^\dagger\| \leq 1$. This is a contradiction; hence, our claim is verified. Let us now elaborate on the KKT first order necessary condition. First of all we observe that the gradient of the constraint, evaluated at a local maximum $\boldsymbol{w}^*$, is equal to $2\boldsymbol{w}^*$, which is nonzero because we established that $\|\boldsymbol{w}^*\| = 1$. Therefore, the problem satisfies the Linear Independence Constraint Qualification (LICQ) condition. Notice that the Lagrangian function for Problem 1 is

$$\mathcal{L}(\boldsymbol{w}, \lambda) = \frac{1}{2}\|A\boldsymbol{e}[k] - \theta L\boldsymbol{w}\|^2 - \frac{\alpha}{2}\|\theta\boldsymbol{w}^*[k] - C\boldsymbol{e}[k]\|^2 + \lambda\left(1 - \|\boldsymbol{w}\|^2\right).$$

According to KKT theory, a necessary condition for $\boldsymbol{w}^*$ to be a local maximum is that it is feasible for Problem 1 (i.e., $\|\boldsymbol{w}^*\| \leq 1$) and there is a $\lambda^* \geq 0$ such that the gradient of $\mathcal{L}(\boldsymbol{w}, \lambda)$ with respect to $\boldsymbol{w}$ (we assume the gradient is a column vector), evaluated at $\boldsymbol{w}^*, \lambda^*$, satisfies

$$\nabla_{\boldsymbol{w}}\mathcal{L}(\boldsymbol{w}, \lambda)\Big|_{\boldsymbol{w}=\boldsymbol{w}^*, \lambda=\lambda^*} = \boldsymbol{0}_m \tag{14}$$

and it holds

$$\lambda^*\left(\|\boldsymbol{w}^*\|^2 - 1\right) = 0. \tag{15}$$

In particular, treating the gradient as a column vector and recalling that

$$\nabla_{\boldsymbol{p}}\boldsymbol{p}^T A\boldsymbol{p} = (A + A^T)\boldsymbol{p}$$

and

$$\nabla_{\boldsymbol{p}}\boldsymbol{p}^T\boldsymbol{q} = \boldsymbol{q},$$

Eq, (14) becomes

$$\theta^2 L^T L\boldsymbol{w}^*[k] - \theta L^T A\boldsymbol{e}[k] - \alpha\theta^2\boldsymbol{w}^*[k] + \alpha\theta C\boldsymbol{e}[k] - 2\lambda^*\boldsymbol{w}^*[k] = \boldsymbol{0}_q. \tag{16}$$

Since $\|\boldsymbol{w}^*\| = 1$, we have that Eq. (15) is satisfied by any $\lambda^* \geq 0$. Let us first assume that $\lambda^* > 0$. By pre-multiplying both sides of Eq. (16) by $\boldsymbol{w}^*[k]^T$, and by noting that $\boldsymbol{w}^*[k]^T\boldsymbol{w}^*[k] = 1$, we get

$$\lambda^* = \frac{\theta^2}{2}\boldsymbol{w}^*[k]^T L^T L\boldsymbol{w}^*[k] - \frac{\theta}{2}\boldsymbol{w}^*[k]^T L^T A\boldsymbol{e}[k] - \alpha\theta^2 + \alpha\theta\boldsymbol{w}^*[k]^T C\boldsymbol{e}[k] \tag{17}$$

At this point we observe that, when $\lambda^* = 0$, Eq. (16) yields

$$\theta(L^T L - \alpha I_n)\boldsymbol{w}^*[k] = (L^T A - \alpha C)\boldsymbol{e}[k];$$

hence, we can express λ^* as in Eq. (17) also in this case. The proof is complete by plugging Eq. (17) into Eq. (16). □

Let us now develop a a sufficient local optimality condition.

Theorem 4. Let $\boldsymbol{w}^*[k]$ be a solution to Problem 3 that satisfies the necessary local optimality conditions from Theorem 1. Moreover, let us define the set

$$\mathcal{D} = \{\boldsymbol{d} \in \mathbb{R}^m \setminus \{\boldsymbol{0}_m\} \,|\, \boldsymbol{w}^*[k]^T\boldsymbol{d} = 0\}. \tag{18}$$

If, for all $\boldsymbol{d} \in \mathcal{D}$, it holds

$$\boldsymbol{d}^T P(\boldsymbol{w}^*[k])\boldsymbol{d} \leq 0 \tag{19}$$

then $\boldsymbol{w}^*[k]$ is a local minimum. If, moreover, Eq. (19) holds as a strict inequality then $\boldsymbol{w}^*[k]$ is a strict local minimum.

Proof 2. Let us specify the KKT second order sufficient local optimality conditions for Problem 3. To this end, we observe that

$$\nabla_{\boldsymbol{w}}\left(\|\boldsymbol{w}\|^2 - 1\right) = 2\boldsymbol{w}$$

and

$$\nabla_{\boldsymbol{w}\boldsymbol{w}}\mathscr{L}(\boldsymbol{w}, \lambda)\Big|_{\boldsymbol{w}=\boldsymbol{w}^*[k],\, \lambda=\lambda^*} = -2\theta P(\boldsymbol{w}^*[k]).$$

From Theorem 3, we know that any local minimum $\boldsymbol{w}^*[k]$ satisfies $\|\boldsymbol{w}^*[k]\| = 1$. Moreover, by construction, we have that $\lambda^* > 0$. Therefore, the set $\mathcal{D}$ can be expressed as in Eq. (18). Hence, according to the KKT second order sufficient local optimality conditions, if Eq. (19) holds true for all $\boldsymbol{d} \in \mathcal{D}$, then $\boldsymbol{w}^*[k]$ is a local minimum. □

Corollary 2. Let the hypotheses of Theorem 4 hold true. If

$$\theta\|L\|^2 - \alpha\theta + \boldsymbol{w}^{*T}[k]\left(L^T A - \alpha C\right)\boldsymbol{e}[k] - \theta\|L\boldsymbol{w}^*[k]\|^2 \leq 0 \tag{20}$$

then $\boldsymbol{w}^*[k]$ is a local minimum for Problem 3. If, moreover, Eq. (20) holds as a strict inequality then $\boldsymbol{w}^*[k]$ is a strict local minimum for Problem 3.

Proof 3. We point out that if $P(\boldsymbol{w}^*[k])$ is negative semi-definite, then Eq. (19) holds true for all $\boldsymbol{d} \in \mathbb{R}^m$, thus including all $\boldsymbol{d} \in \mathbb{D}$. Notice that $P(\boldsymbol{w}^*[k])$ is negative semi-definite if and only if

$$\theta\underbrace{\lambda_{\max}(L^T L)}_{\|L\|^2} - \alpha\theta + \boldsymbol{w}^{*T}[k]L^T A\boldsymbol{e}[k] - \theta\underbrace{\boldsymbol{w}^{*T}[k]L^T L\boldsymbol{w}^*[k]}_{\|L\boldsymbol{w}^*[k]\|^2} - \alpha\boldsymbol{w}^*[k]^T C\boldsymbol{e}[k] \leq 0 \tag{21}$$

Therefore, the condition in Eq. (21) is equivalent to Eq. (20). The proof is complete. □

5. EXAMPLE

This section aims to discuss an example that numerically shows the effectiveness of an attack based on the aforementioned optimal strategy. Let us consider a plant described by the following matrices

$$A = \begin{bmatrix} -2.3161 & -4.5903 & 19.9062 & -12.4361 & -6.5553 \\ 5.1571 & 8.5670 & -33.7278 & 21.3389 & 11.0387 \\ 0.7337 & 0.9996 & -3.4100 & 2.7150 & 1.4910 \\ -0.6511 & -1.1228 & 5.3945 & -2.5439 & -1.7871 \\ 1.4864 & 2.1363 & -9.2198 & 5.7798 & 3.9031 \end{bmatrix},$$

$$B = [0.2\ 0\ 0.4\ 0\ 0]^T,$$

$$C = \begin{bmatrix} 0 & 0.1 & 2 & 0 & 0 \\ -2.11 & 0 & 0 & 4 & -1 \end{bmatrix}.$$

Notice that the system is observable; hence, the SCADA system is able to reconstruct the state via a Luenberger-type observer. In particular, the observer's gain matrix L is

$$L = \begin{bmatrix} 3.8679 & -3.5594 \\ -6.6531 & 6.1420 \\ -0.8546 & 0.7866 \\ 1.1139 & -1.0368 \\ -1.7814 & 1.6388 \end{bmatrix}.$$

Finally, we choose as threshold for the BDD the value $\theta = 0.02$ for the attack-triggering condition based on the norm of the residue. In particular, notice that, to avoid numerical issues resulting in the trespassing of the threshold, we set $\theta = 0.02 - 10^{-5}$.

We consider an attack lasting $k_{\max} = 1000$ time steps and we assume that the system's output is modified by injecting a signal $\boldsymbol{y}_a[k]$, which is chosen by solving Problem 1 at each time instant. In particular, in order

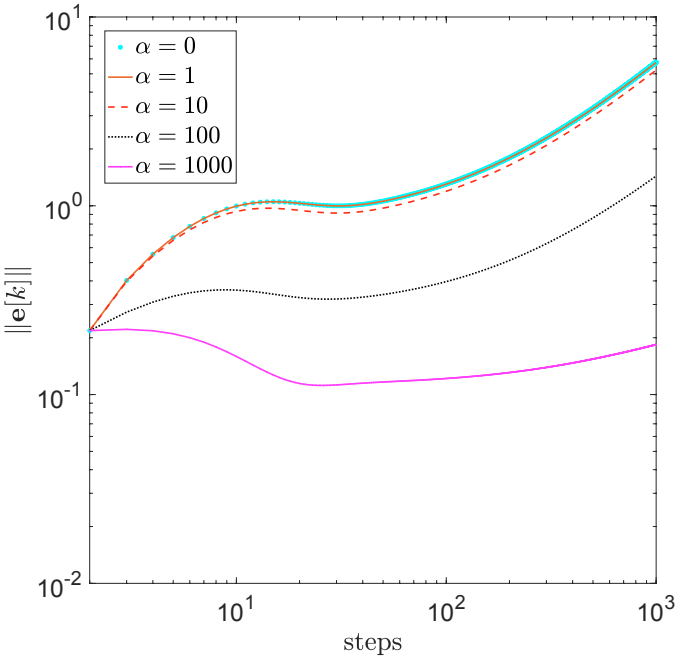

Fig. 1. Norm of the estimation error as a result of the attack, for different choices of the parameter α.

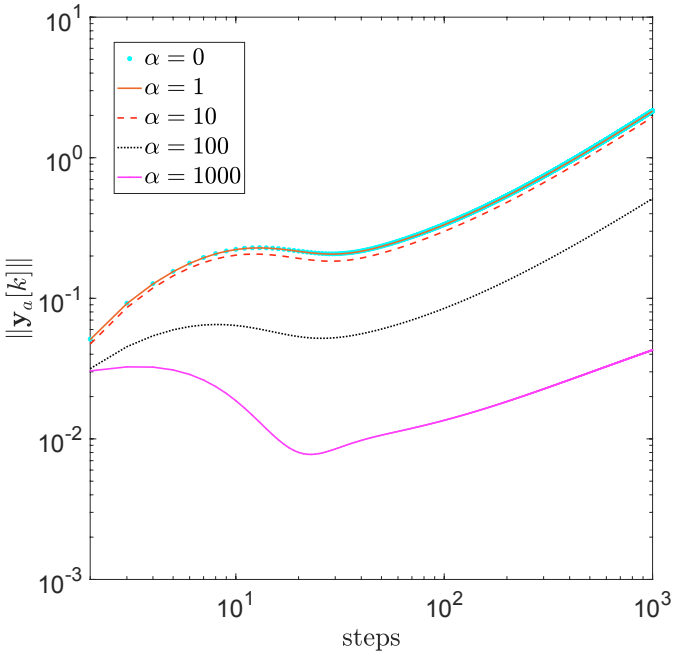

Fig. 2. Norm of the injected signal as a result of the attack, for different choices of the parameter α.

to find a locally optimal solution, in this paper we resort to the `fmincon` function in Matlab^{TM}.

Figure 1 shows the norm of the estimation error as a result of the attack, for different choices of the parameter α. Notably, we observe that the attack strategy is quite successful, in that $\|\boldsymbol{e}[k]\|$ grows with time, meaning that the estimator is getting less and less aware of the real ongoing situation. Moreover, we observe that as α grows the objective of minimizing the norm of the injected signal becomes dominant and the effect of the attack is reduced in terms of $\|\boldsymbol{e}[k]\|$, but the norm of the injected signal is also reduced (Figure 2); however, for small values of α the result is close to the one for $\alpha = 0$, meaning that a tradeoff featuring a limited interest for the attacker in minimizing the injected signal magnitude is effective in a way that is comparable with not caring about this magnitude. Finally, we observe that in all cases the BDD condition is never triggered (Figure 3).

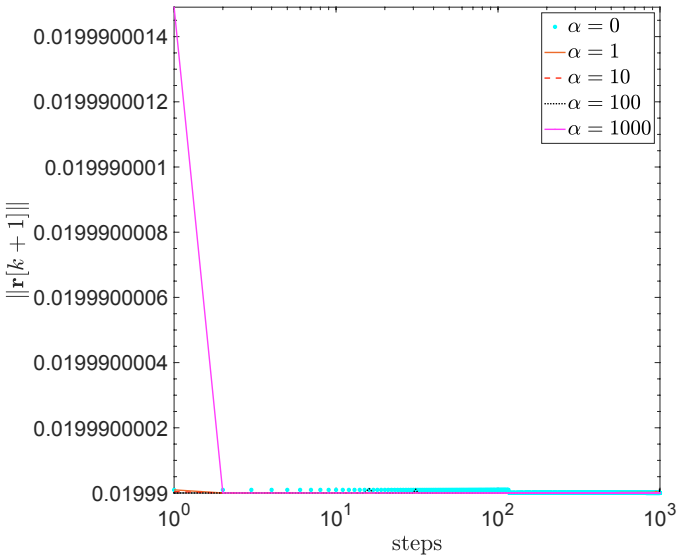

Fig. 3. Norm of the residue as a result of the attack, for different choices of the parameter α. Note that the condition $\|\boldsymbol{r}[k+1]\| > \theta$ that triggers the BDD is never verified.

6. CONCLUSIONS

This paper presents an approach for designing an optimal attack on a plant, where the attacker is able to substitute the real outputs of a system with altered values, resulting in huge disparities in the SCADA's estimated state without being detected. In particular, we assume the attacker seeks a tradeoff between the desire to deal the largest damage in terms of estimation error and of the aim to keep the magnitude of the injected signals to a minimum. We define the attacker approach as a non-concave maximization problem, for which we provide local optimality conditions.

The proposed example numerically demonstrates the ability of the proposed attack strategy. Future work will focus on relaxing the attacker's perfect information hypothesis and addressing nonlinear systems with noise.

REFERENCES

Cardenas, A. and Sastry (2008). Secure control: Towards survivable cyber-physical systems. In *Proc. 28th Int. Conf. Distrib. Comput. Workshops*, 495–500.

Faramondi, L., Flammini, F., Guarino, S., and Setola, R. (2021a). A hardware-in-the-loop water distribution testbed dataset for cyber-physical security testing. *IEEE Access*, 9, 122385–122396.

Faramondi, L., Oliva, G., and Setola, R. (2021b). Optimal man-in-the-middle covert attack. In *16th international conference on Critical Information Infrastructure Security (CRITIS2021)*. To Appear.

Fawzi, H., Tabuada, P., and Diggavi, S. (2014). Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic control*, 59(6), 1454–1467.

Floudas, C.A. (1995). *Nonlinear and mixed-integer optimization: fundamentals and applications*. Oxford University Press.

Gupta, L. and Basar (2010). Optimal control in the presence of an intelligent jammer with limited actions. In *49th IEEE Conference on Decision and Control (CDC)*, 1096–1101. IEEE.

Hobbs, A. (2021). The colonial pipeline hack: Exposing vulnerabilities in us cybersecurity.

Jovanov, I. and Pajic, M. (2019). Relaxing integrity requirements for attack-resilient cyber-physical systems. *IEEE Transactions on Automatic Control*, 64(12), 4843–4858.

Laso, P.M., Brosset, D., and Puentes, J. (2017). Dataset of anomalies and malicious acts in a cyber-physical subsystem. *Data in brief*, 14, 186–191.

Lee, C., Shim, H., and Eun, Y. (2015). Secure and robust state estimation under sensor attacks, measurement noises, and process disturbances: Observer-based combinatorial approach. In *2015 European Control Conference (ECC)*, 1872–1877. IEEE.

Li, Y., Shi, L., Cheng, P., Chen, J., and Quevedo, D.E. (2015). Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach. *IEEE Transactions on Automatic Control*, 60(10), 2831–2836.

Luenberger, D. (1971). An introduction to observers. *IEEE Transactions on automatic control*, 16(6), 596–602.

Miciolino, S., Bernieri, G., Panzieri, S., Pascucci, F., and Polycarpou, M. (2017). Fault diagnosis and network anomaly detection in water infrastructures. *IEEE Design & Test*, 34(4), 44–51.

Mo and Sinopoli (2012). Integrity attacks on cyber-physical systems. In *Proc. 1st Int. Conf. High Confidence Netw. Syst*, 47–54.

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

Ren, X., Wu, J., Dey, S., and Shi, L. (2018). Attack allocation on remote state estimation in multi-systems: Structural results and asymptotic solution. *Automatica*, 87, 184–194.

Setola, R., Faramondi, L., Salzano, E., and Cozzani, V. (2019). An overview of cyber attack to industrial control system. *Chemical Engineering Transactions*, 77, 907–912.

Teixeira, A., Shames, I., Sandberg, H., and Johansson, K.H. (2012). Revealing stealthy attacks in control systems. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 1806–1813. IEEE.

Wu, G., Sun, J., and Chen, J. (2018). Optimal data injection attacks in cyber-physical systems. *IEEE transactions on cybernetics*, 48(12), 3302–3312.

Zhang, H., Cheng, P., Shi, L., and Chen, J. (2015). Optimal denial-of-service attack scheduling with energy constraint. *IEEE Transactions on Automatic Control*, 60(11), 3023–3028.

Zhao, J., Mili, L., and Abdelhadi, A. (2017). Robust dynamic state estimator to outliers and cyber attacks. In *2017 IEEE Power & Energy Society General Meeting*, 1–5. IEEE.