Attack-Resilient State Estimation with Intermittent Data Authentication ⋆

Amir Khazraei [a], Miroslav Pajic [a]

[a] *Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA*

Abstract

Network-based attacks on control systems may alter sensor data delivered to the controller, effectively causing degradation in control performance. As a result, having access to accurate state estimates, even in the presence of attacks on sensor measurements, is of critical importance. In this paper, we analyze performance of resilient state estimators (RSEs) when *any* subset of sensors may be compromised by a *stealthy attacker*. Specifically, we consider systems with the well-known l_0-based RSE and two commonly used sound intrusion detectors (IDs). For linear time-invariant plants with bounded noise, we define the notion of perfect attackability (PA) when attacks may result in unbounded estimation errors while remaining undetected by the employed ID (i.e., stealthy). We derive necessary and sufficient PA conditions, showing that a system can be perfectly attackable even if the plant is stable. While PA can be prevented with the use the standard cryptographic mechanisms (e.g., message authentication) that ensure data integrity under network-based attacks, their continuous use imposes significant communication and computational overhead. Consequently, we also study the impact that even intermittent use of data authentication has on RSE performance guarantees in the presence of stealthy attacks. We show that if messages from some of the sensors are even intermittently authenticated, stealthy attacks could not result in unbounded state estimation errors.

Key words: Security of control systems; Cyber-physical systems; Attack-resilient state estimation; Perfect attackability;

1 Introduction

The challenge of securing control systems has recently attracted significant attention due to high profile attacks, such as the attack on Ukrainian power grid [25] and the StuxNet attack [8]. In such incidents, the attacker can affect a physical plant by altering actuation commands or sensory measurements, or affecting execution of the controller. One approach to address this problem has been to exploit a dynamical model of the plant for attack detection and attack-resilient control (e.g., [12, 24, 19, 1, 18, 14, 15, 23, 6]).

For instance, consider the problem of attack-resilient control when measurements from a subset of the plant sensors may be compromised. One line of work employs a widely used (non-resilient) Kalman filter, with a standard residual-based probabilistic detector (e.g., $\mathcal{X}^2$ detector) triggering alarm in the presence of attack [24, 7, 3]. These Kalman filter-based controllers of linear time-invariant (LTI) plants may be vulnerable to stealthy (i.e., undetected) attacks resulting in unbounded state-estimation errors; thus, such systems are referred to as *perfectly attackable* (PA) [24, 7, 3]. Specifically, for LTI systems with Gaussian noise and Kalman filter-based controllers, the notion of perfect attackability (PA) [1] is introduced in [24]. In particular, [24], and [3] for larger classes of intrusion detectors (IDs), show that the system is PA if and only if the plant is unstable and the set of compromised sensors satisfies that no unstable eigenvector lie in the kernel of their observation matrix.

Resilient (i.e., secure) state estimation is another approach to achieve attack-resilient control; here, the objective is to estimate the system state when a subset of the sensors is corrupted [1, 16]. This allows for the use of standard feedback controllers to provide strong control guarantees in the presence of attacks. A common approach is to use a batch-processing resilient state estimator (RSE) to estimate the system state and attack

⋆ This work is sponsored in part by the ONR agreements N00014-17-1-2504 and N00014-20-1-2745, AFOSR award FA9550-19-1-0169, and the NSF CNS-1652544 award. This paper was not presented at any IFAC meeting; preliminary version of some of these results were presented in [5].

Email addresses: amir.khazraei@duke.edu (Amir Khazraei), miroslav.pajic@duke.edu (Miroslav Pajic).

[1] For conciseness, we use PA for *perfect attackability* or *perfectly attackable*, when the meaning is clear from the context.

vectors (e.g., [1, 15, 14, 18]). For LTI systems without noise, the state and attack vectors can be obtained by solving an l_0, or under more restrictive condition l_1, optimization problem [1]. These results are extended to systems with bounded noise [14], showing that the worst case state estimation error is a linear with the noise size; thus, the attacker cannot exploit the noise to introduce unbounded state-estimation errors, unless a sufficiently large number of sensors is corrupted. SMT- and graph-based estimators from [18] and [11] improve computational efficiency of the estimators. However, all these methods employ a common restrictive assumption that the maximal number of corrupted sensors is bounded; at best, less than half of sensors can be compromised. Moreover, to the best of our knowledge, the impact of stealthy attacks on the RSEs has not been considered, either in the general case or under such restrictive assumptions.

However, the assumption that measurements from only a subset of sensors are compromised cannot be justified in the common scenarios where the attacker has access to the network used to transmit data from sensors to the controller. Thus, it is important to analyze impact of such Man-in-the-Middle attacks on performance of the RSEs. A common defense against network-based attacks is the use of cryptographic tools, such as adding Message Authentication Codes (MACs) to measurement messages to guarantee their integrity. Yet, continuous use of security primitives such as MACs, can cause computation and communication overhead, which limits its applicability in resource-constrained control systems [9, 10]. To overcome this, intermittent data authentication can be used for control systems [3]; specifically, LTI systems with Gaussian noise and a Kalman filter-based controller, cannot be PA if message authentication is at least intermittently employed. On the other hand, no such guarantees have been shown in systems with bounded-size noise and RSE-based controllers.

Consequently, in this work, we focus on performance of LTI systems with bounded-size noise, employing an RSE-based controller, under stealthy attacks on an *arbitrary number* of sensors. Specifically, we consider a system with an l_0-based RSE, due to the strongest resiliency guarantees, and one of two previously reported intrusion detectors (IDs) for systems with set-based noise. Due to the batch-processing nature of RSEs, we introduce two notions of PA for such systems – at a single time point and over time, where a stealthy attacker may introduce arbitrarily large estimation errors. Then, we provide necessary and sufficient conditions for both notions of PA. We show that unlike PA in the Kalman filter-based estimators, a system may be PA over time even if the physical plant is not unstable. Furthermore, we show that even intermittent data authentication guarantee can help against such perfect attacks for some types of IDs. Unlike [3], we show that using authentication only once in every bounded time interval ensures bounded estimation errors under any stealthy attack.

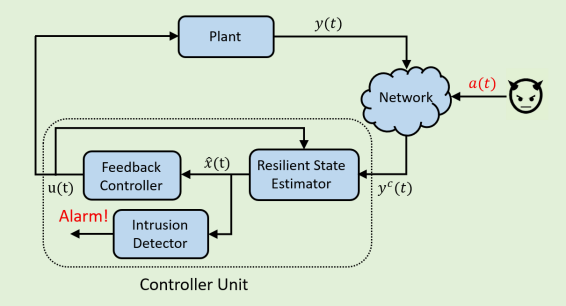


Fig. 1. Control architecture under network-based attacks.

This paper is organized as follows. Section 2 formalizes the problem including the system and attack models. In Section 3, we define the concept of perfectly attackable systems and find the necessary and sufficient conditions for PA. In Section 4, we study effects of intermittent message authentication on performance guarantees under attack. Finally, our results are illustrated in case studies in Section 5, before concluding remarks in Section 6.

Notation. $\mathbb{B}$ and $\mathbb{R}$ denote the set of Boolean and real numbers, respectively, and $\mathbb{I}(.)$ is the indicator function. For a matrix $A, \mathcal{N}(A)$ denotes its null space, A^T its transpose, $A^\dagger$ its Moore-Penrose pseudoinverse, and $||A||$ the l_2 norm of the matrix. For a vector $x \in \mathbb{R}^n$, we denote by $||x||_p$ the p-norm of x; when p is not specified, the 2-norm is implied. In addition, we use x_i to denote the i^{th} element of x, while $\text{supp}(x)$ denotes the indices of nonzero elements of x – i.e., $\text{supp}(x) = \{i \mid i \in \{1, ..., n\}, x_i \neq 0\}$.

Projection vector e_i is the unit vector where a 1 in its i^{th} position is the only nonzero element of the vector. For set $\mathcal{S}$, $|\mathcal{S}|$ denotes the cardinality of the set and $\mathcal{S}^{\complement}$ its complement. $\mathcal{P}_{\mathcal{K}}x$ is the projection from the set $\mathcal{S}$ to set $\mathcal{K}$ ($\mathcal{K} \subseteq \mathcal{S}$) by keeping only elements of x with indices from $\mathcal{K}$; formally, $\mathcal{P}_{\mathcal{K}} = \left[e_{j_1} | ... | e_{j_{|\mathcal{K}|}} \right]^T$, where $\mathcal{K} = \{s_{j_1}, ..., s_{j_{|\mathcal{K}|}}\} \subseteq S$ and $j_1 < j_2 < ... < j_{|\mathcal{K}|}$. If e.g., $\mathcal{S} = \{1, 2, 3, 4\}$ and $\mathcal{K} = \{2, 4\}$, then $\mathcal{P}_{\mathcal{K}}x = \left[x_2 \; x_4 \right]^T$.

2 Problem Description

We start by introducing the system (Fig. 1) and attack model, before formalizing the considered problem.

2.1 System and Attack Model

We now describe each system component from Fig. 1.

Plant Model. We assume that the plant is an observable linear time-invariant (LTI) dynamical system that can be modeled in the standard state-space form as

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + v_P(t), \\ y(t) &= Cx(t) + v_M(t). \end{aligned} \tag{1}$$

Here, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$ denote the state, input and output vectors, respectively. The plant output vec-

tor captures measurements from the set of plant sensors $\mathcal{S} = \{s_1, s_2, ..., s_p\}$.[2] In addition, $v_P \in \mathbb{R}^n$ and $v_M \in \mathbb{R}^p$ are bounded process and measurement noise vectors – i.e., there exist $\delta_{v_P}, \delta_{v_M} \in \mathbb{R}$ such that for all $t \geq 0$,

$$\|v_P(t)\|_2 \leq \delta_{v_P}, \quad \|v_M(t)\|_2 \leq \delta_{v_M}. \tag{2}$$

Note that we make no assumptions about the distributions of the sensor and measurement noise models.

Attack Model. We assume that the attacker was able to compromise information flow from a subset of sensors $\mathcal{K} \subseteq \mathcal{S}$;[3] however, we make no assumption about the set $\mathcal{K}$ (e.g., its size or elements). Hence, the sensor measurements delivered to the controller can be modeled as

$$y^c(t) = y(t) + a(t). \tag{3}$$

Here, $a(t) \in \mathbb{R}^p$ denotes the sparse attack signal injected by the attacker at time t via the compromised information flows (i.e., sensors) from $\mathcal{K}$; hence, $\mathcal{K} = \text{supp}(a(t))$.

We use a commonly adopted threat model (e.g., [3]) where:

(i) the attacker has the full knowledge of the system, its dynamics and design (e.g., controller and ID), as well as the employed security mechanisms – e.g., the times when authentication is used,
(ii) the attacker has the required computation power to calculate suitable attack signals to inject via the set $\mathcal{K}$, while planning ahead as needed,
(iii) the attacker's goal is to design attack signal $a(t)$ such that it always remain *stealthy* (i.e., undetected by the ID), while *maximizing control degradation.*

The notions of *stealthiness* and *control performance degradation* depend on the controller, and thus will be formally defined after the controller design is introduced.

Controller Design. The controller employs an RSE whose output is used for standard feedback control, and an ID (Fig. 1). To simplify our notation while describing the RSE, the model (1) can be considered in the form

$$\begin{aligned} x(t+1) &= Ax(t), \\ y(t) = y^c(t) &= Cx(t) + w(t) + a(t); \end{aligned} \tag{4}$$

specifically, we can ignore the contribution of $u(t)$ as it is a known signal (no attacks on actuator are considered in this work) and thus has no effect on resilient state estimation. As shown in [13, 14], the bounds on the size of measurement noise w in (4) can be related to the bounds on the size of process and measurement noise vectors v_P and v_M; i.e., there exists $\delta_w > 0$ such that

$$||w(t)|| \leq \delta_w, \qquad \text{for all } \ t \geq 0. \tag{5}$$

Resilient State Estimator. The goal of an RSE is to reconstruct the system state $x(t)$ from N sensor measurements $\{y(t), ..., y(t+N-1)\}$. We assume that $N = n$; however, the results can be extended to the case $N < n$, or $N > n$. To formally capture RSE requirements, we rewrite the system model from (4) as

$$\mathbf{y}(t) = \mathbf{O}x(t) + \mathbf{a}(t) + \mathbf{w}(t), \tag{6}$$

where $\mathbf{O} = [\mathbf{O}_1^T \mid ... \mid \mathbf{O}_p^T]^T$. For each sensor i and a subset of sensors $\mathcal{K}$, we define the matrices $\mathbf{O}_i$ and $\mathbf{O}_\mathcal{K}$ as

$$\mathbf{O}_\mathcal{K} = \Big[(\mathcal{P}_\mathcal{K} C)^T \ (\mathcal{P}_\mathcal{K} CA)^T \ \dots \ (\mathcal{P}_\mathcal{K} CA^{N-1})^T\Big]^T, \tag{7}$$

with $\mathbf{O}_i = \mathbf{O}_{\{s_i\}}$. Also, each of the block vectors $\mathbf{a}$, $\mathbf{y}$, $\mathbf{w} \in \mathbb{R}^{pN}$, satisfies $\mathbf{a}(t) = [\mathbf{a}_1^T(t) \mid ... \mid \mathbf{a}_p^T(t)]^T$, $\mathbf{y}(t) = [\mathbf{y}_1^T(t) \mid ... \mid \mathbf{y}_p^T(t)]^T$ and $\mathbf{w}(t) = [\mathbf{w}_1^T(t) \mid ... \mid \mathbf{w}_p^T(t)]^T$. Now, for each sensor $i \in \mathcal{S}$, it holds that

$$\mathbf{y}_i(t) = \mathbf{O}_i x(t) + \mathbf{a}_i(t) + \mathbf{w}_i(t) \tag{8}$$

with $\mathbf{a}_i(t) = [a_i(t) \mid a_i(t+1) \mid ... \mid a_i(t+N-1)]^T \in \mathbb{R}^N$ denoting the values injected via i^{th} sensor at time steps $t, ..., t+N-1$, with $\mathbf{a}_i(t) = 0$ if $i \notin \mathcal{K}$. Finally, $\mathbf{y}_i(t) = [y_i(t) \mid y_i(t+1) \mid ... \mid y_i(t+N-1)]^T \in \mathbb{R}^N$ and $\mathbf{w}_i(t) = [w_i(t) \mid w_i(t+1) \mid ... \mid w_i(t+N-1)]^T \in \mathbb{R}^N$ are the values of sensor i measurements and its noise.

In general, the RSE functionality can be captured as [1]

$$\mathcal{E} : \mathbb{R}^{Np} \mapsto \mathbb{R}^n \times \mathbb{R}^{Np} \ \ s.t. \ \ \mathcal{E}(\mathbf{y}(t)) = \big(\hat{x}(t), \hat{\mathbf{a}}(t)\big). \tag{9}$$

Here, $\hat{x}(t)$ and $\hat{\mathbf{a}}(t)$ are the state and attack vectors estimated from the delivered sensor measurements. The estimation error of an RSE is defined as

$$\Delta x(t) = \hat{x}(t) - x(t). \tag{10}$$

A conventional RSE is the l_0-based decoder [14], or its equivalent forms (e.g., [1, 18]), defined as optimization

$$\begin{aligned} \min_{\hat{x}(t), \hat{\mathbf{a}}(t)} & \sum_{i=1}^{p} \mathbb{I}\big(\|\hat{\mathbf{a}}_i(t)\| > 0\big) \\ s.\,t. \ \ & \mathbf{y}(t) = \mathbf{O}\hat{x}(t) + \hat{\mathbf{w}}(t) + \hat{\mathbf{a}}(t) \\ & \hat{\mathbf{w}}(t) \in \Omega. \end{aligned} \tag{11}$$

Here, Ω denotes the feasible set of noise vectors, determined by the noise bounds from (5). The vectors $\hat{\mathbf{w}}(t)$ and $\hat{\mathbf{a}}(t)$ are estimated at time t independently from the estimated vectors at time step $t-1$. Hence, we denote $\hat{\mathbf{w}}(t) = [\hat{\mathbf{w}}_1^T(t) \mid ... \mid \hat{\mathbf{w}}_p^T(t)]^T$, $\hat{\mathbf{a}}(t) = [\hat{\mathbf{a}}_1^T(t) \mid ... \mid \hat{\mathbf{a}}_p^T(t)]^T$, with $\hat{\mathbf{w}}_i(t) = [\hat{w}_i^{(t)}(t) | ... | \hat{w}_i^{(t)}(t+N-1)]^T$ and $\hat{\mathbf{a}}_i(t) =$

[2] To simplify our notation, unless otherwise stated, we will use i instead of s_i to denote the i-th sensor.

[3] To simplify our presentation, we refer to these sensors as compromised since the effects of network-based attack are mathematically equivalent to compromising the sensors [22].

$[\hat{a}_i^{(t)}(t)|...|\hat{a}_i^{(t)}(t+N-1)]^T$, in which $\hat{a}_i^{(t)}(k)$ and $\hat{w}_i^{(t)}(k)$ are the estimated noise and attack vectors at time k, as computed at time t, for $k = t, ..., t+N-1$.

When not more than s sensors are compromised in a $2s$-sparse observable system [17], the estimation error of the RSE (11) is bounded [14]; $2s$-sparse observable depends on the properties of the observability matrix of (A, C).

Intrusion Detector. We consider two ID used to detect the presence of any system anomaly (including attacks):

(1) ID_I: We capture the ID_I functionality in the general form as mapping $\mathcal{D}_I : \mathbb{R}^{Np} \mapsto \mathbb{B}$ defined as

$$\mathcal{D}_I(\hat{\mathbf{a}}(t)) = \mathbb{I}(\|\hat{\mathbf{a}}(t)\| > 0); \tag{12}$$

i.e., if the estimated attack vector is non-zero, ID_I raises alarm. Note that our goal is *not* to identify the exact set of attacked sensors, which would result in a nonzero threshold in (12), as shown in [14].

(2) ID_{II}: We define the ID_{II} as $\mathcal{D}_{II} : \mathbb{R}^{Np+2n} \mapsto \mathbb{B}$ with

$$\begin{aligned}&\mathcal{D}_{II}(\hat{\mathbf{a}}(t), \hat{x}(t), \hat{x}(t-1)) = \\ &\mathbb{I}\big(\|\hat{\mathbf{a}}(t)\| > 0\big) \vee \mathbb{I}\big(||\hat{x}(t) - A\hat{x}(t-1)|| > d\big);\end{aligned} \tag{13}$$

here, $\vee$ is Boolean OR and d is defined by Prop. 1.

We use $\mathcal{D}_I(\hat{\mathbf{a}})$ and $\mathcal{D}_{II}(\hat{\mathbf{a}}, \hat{x})$ instead of $\mathcal{D}_I(\hat{\mathbf{a}}(t))$ and $\mathcal{D}_{II}(\hat{\mathbf{a}}(t), \hat{x}(t), \hat{x}(t-1))$, respectively. We also denote the system (4) with ID_i ($i \in \{\text{I}, \text{II}\}$) as $\Sigma_i(A, C, \delta_w, \mathcal{K})$. Yet, if results hold for both IDs, we remove the subscript i.

Proposition 1 *For the system without attack, it holds that* $||\hat{x}(t) - A\hat{x}(t-1)|| \leq d = 2\sqrt{N}\delta_w||\mathbf{O}^\dagger||_2(1+||A||_2)$.

PROOF. Constraints in (11) at time t and $t-1$ imply

$$\begin{aligned}\mathbf{O}x(t) + \mathbf{w}(t) &= \mathbf{O}\hat{x}(t) + \hat{\mathbf{w}}(t) \\ \mathbf{O}x(t-1) + \mathbf{w}(t-1) &= \mathbf{O}\hat{x}(t-1) + \hat{\mathbf{w}}(t-1)\end{aligned} \tag{14}$$

For $\Delta\mathbf{w}(t) = \mathbf{w}(t) - \hat{\mathbf{w}}(t)$, since (A, C) is observable,

$$\begin{aligned}\hat{x}(t) &= x(t) - \mathbf{O}^\dagger\Delta\mathbf{w}(t) \\ \hat{x}(t-1) &= x(t-1) - \mathbf{O}^\dagger\Delta\mathbf{w}(t-1).\end{aligned} \tag{15}$$

Hence, from (4), $||\hat{x}(t) - A\hat{x}(t-1)|| = ||A\mathbf{O}^\dagger\Delta\mathbf{w}(t-1) - \mathbf{O}^\dagger\Delta\mathbf{w}(t)|| \leq ||\mathbf{O}^\dagger||_2||\Delta\mathbf{w}(t)|| + ||A||_2||\mathbf{O}^\dagger||_2||\Delta\mathbf{w}(t-1)||$. On the other hand, $||\Delta\mathbf{w}(t)||_2 \leq 2\sqrt{N}\delta_w$, which also holds for $\Delta\mathbf{w}(t-1)$, and thus concludes the proof.

2.2 Problem Formulation

In this work, we focus on the following two problems.

Problem 1: Under which conditions, a *stealthy* attacker could introduce arbitrarily large estimation errors (10)?

From (12), (13), the stealthiness conditions for ID_I, ID_{II} are

$$\mathcal{D}_I(\hat{\mathbf{a}}) = 0, \qquad \mathcal{D}_{II}(\hat{\mathbf{a}}, \hat{x}) = 0. \tag{16}$$

Note that if an attack is stealthy from ID_{II} it cannot be detected by ID_I either. Due to the batch-processing nature of the RSE and bounded-size noise, the approach and conditions from [24, 7] cannot be used. Hence, we introduce PA for LTI systems with bounded-size noise.

Problem 2: As we show in next section, for a large class of systems $\Sigma_I(A, C, \delta_w, \mathcal{K})$, an unbounded state estimation error can be inserted by compromising a subset of sensors. Although the use of ID_{II} (i.e., for systems $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$) restricts these conditions, unstable plants are vulnerable to perfect attacks (i.e., stealthy attacks that cause unbounded estimation errors). On the other hand, the use of security mechanisms, such as message authentication, could ensure integrity of the received sensor measurements. Thus, a stealthy attack vector has to satisfy $a_i(t) = 0$ when the measurement of sensor s_i is authenticated at time t, and $a(t) = 0$ if integrity of all sensors is enforced at time t. Since authentication comes with additional computational and communication cost, we study the effects of intermittent data authentication on attack impact. Our goal is to find conditions that the authentication policy (i.e., times when authentication is used) should satisfy so that the systems $\Sigma_I(A, C, \delta_w, \mathcal{K})$, $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ are not PA.

3 PA of LTI Systems with Bounded-Noise

The notion of PA is introduced in [24, 7] for systems with a statistical (χ^2) ID and a Kalman-filter implementing continuous (i.e., streamed) processing of sensor measurements. On the other hand, most existing RSEs for systems with bounded noise (e.g., [14, 18, 21, 20]) are based on batch-processing of sensor data – i.e., processing a window of sensor measurements at each time step (mostly even without taking previous computations into account). Thus, the notion of PA needs to differentiate between PA at a single time point vs. PA over a time interval. In this section, we first define these two notions of PA for systems with one of the two IDs, before providing the necessary and sufficient conditions individually.

Definition 1 *System* $\Sigma_I(A, C, \delta_w, \mathcal{K})$ *is perfectly attackable at a single time step if for any* $M > 0$*, there exists a stealthy sequence of attack signals* $\mathbf{a}(t)$ *over* N *time steps (i.e., satisfying* (16)*), for which the RSE estimation error satisfies* $||\Delta x(t)|| > M$*. Such attack vector* $\mathbf{a}(t)$ *is called a perfect attack for the system* $\Sigma_I(A, C, \delta_w, \mathcal{K})$*.*

Definition 1 does not require that the attack is stealthy before or remains stealthy at time steps after t. Such notion of PA for the system $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ is not relevant because the ID_{II}, from (13), validates that the estimated states in every two consecutive steps do not violate plant dynamics. Thus, we characterize a more realistic requirements for stealthy attacks – PA over a time-interval.

Definition 2 *System $\Sigma(A, C, \delta_w, \mathcal{K})$ is perfectly attackable over time if for all $M > 0$ there exists a sequence of attack signals $\mathbf{a}(t), \mathbf{a}(t+1), ...$ and a time point $t' \geq t$ such that for all k, where $k \geq t'$, it holds that $||\Delta x(k)|| > M$, and for all time steps, the estimated attack vectors $\hat{\mathbf{a}}$ satisfies the corresponding stealthiness requirements in* (16).

To simplify our presentation, instead of formally stating that the estimation error may be arbitrarily large, we may say that the estimation error is unbounded.

Remark 1 *If the system $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ is perfectly attackable over time, then the $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is also perfectly attackable over time because the stealthiness condition in (16) for ID_{II} also includes the condition for ID_I.*

Note that PA over time is a stronger notion than PA at a single time point because $\mathcal{D}_I(\hat{\mathbf{a}}(t)))$ should be equal to zero for all time steps. Therefore, the following holds.

Proposition 2 *If the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA over time, then it is also PA at a single time step.*

3.1 PA of $\Sigma_I(A, C, \delta_w, \mathcal{K})$ System

We now capture conditions for PA at a single time.

Theorem 1 *System $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at a single-time step if and only if pair $(A, \mathcal{P}_{\mathcal{K}^\complement} C)$ is not observable.*

PROOF. ($\Rightarrow$) Let us assume that the pair $(A, \mathcal{P}_{\mathcal{K}^\complement} C)$ is observable, while the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at a single time step, which we denote as t. Then, there exists a stealthy attack sequence $\mathbf{a}(t)$ for which the RSE estimated attack vector $\hat{\mathbf{a}}(t) = 0$ and $\|\Delta x(t)\|$ is unbounded.

Consider data from noncompromised sensors in $\mathcal{K}^\complement$; i.e.,

$$\mathcal{P}_{\mathcal{K}^\complement}\mathbf{y}(t) \overset{\text{(i)}}{=} \mathbf{O}_{\mathcal{K}^\complement} x(t) + \mathcal{P}_{\mathcal{K}^\complement}\mathbf{w}(t) \overset{\text{(ii)}}{=} \mathbf{O}_{\mathcal{K}^\complement}\hat{x}(t) + \mathcal{P}_{\mathcal{K}^\complement}\hat{\mathbf{w}}(t),$$

where (i) holds from (6) as the sensors are noncompromised, whereas (ii) holds from (11) since the attack is stealthy (i.e., $\hat{\mathbf{a}}(t) = 0$). Hence, it follows that $\mathbf{O}_{\mathcal{K}^\complement}\Delta x(t) = \mathcal{P}_{\mathcal{K}^\complement}\Delta\mathbf{w}(t)$, where $\Delta\mathbf{w}(t) = \mathbf{w}(t) - \hat{\mathbf{w}}(t)$. Since the matrix $\mathbf{O}_{\mathcal{K}^\complement}$ is full rank, $\Delta x(t) = \left(\mathbf{O}_{\mathcal{K}^\complement}\right)^\dagger \left(\mathcal{P}_{\mathcal{K}^\complement}\Delta\mathbf{w}(t)\right)$, and thus

$$\|\Delta x(t)\| \leq \left\|\left(\mathbf{O}_{\mathcal{K}^\complement}\right)^\dagger\right\| \left(\|\mathcal{P}_{\mathcal{K}^\complement}\Delta\mathbf{w}(t)\| \right). \quad (17)$$

The matrix $\left(\mathbf{O}_{\mathcal{K}^\complement}\right)^\dagger$ has a bounded norm, $\mathbf{w}(t)$ and $\hat{\mathbf{w}}(t)$ are also bounded. Thus, the right side of (17) is bounded, meaning that $\Delta x(t)$ is bounded, which is a contradiction.

($\Leftarrow$) Suppose that the pair $(A, \mathcal{P}_{\mathcal{K}^\complement} C)$ is not observable; thus, there exists a nonzero vector z such that $\mathbf{O}_{\mathcal{K}^\complement} z = 0$. Let us assume that the system is in state $x(t)$ when attack $\mathbf{a}(t) = \left[(\mathcal{P}_{\mathcal{K}}\mathbf{a}(t))^T \ (\mathcal{P}_{\mathcal{K}^\complement}\mathbf{a}(t))^T\right]^T = \mathbf{O}z = \left[(\mathbf{O}_{\mathcal{K}} z)^T \ 0\right]^T$ is applied. Then, from (6) we have that

$$\mathbf{y}(t) = \mathbf{O}x(t) + \mathbf{w}(t) + \mathbf{a}(t) = \mathbf{O}\hat{x}(t) + \hat{\mathbf{w}}(t) + \hat{\mathbf{a}}(t). \quad (18)$$

Consider $\hat{\mathbf{w}}'(t) = \hat{\mathbf{w}}(t)$, $\hat{x}'(t) = \hat{x}(t) + z$ and $\hat{\mathbf{a}}'(t) = 0$. Now, $(\hat{x}'(t), \hat{\mathbf{w}}'(t), \hat{\mathbf{a}}'(t))$ is a feasible point for the RSE optimization problem from (11) that also minimizes the objective to zero. Thus, the output of RSE $(\hat{x}(t), \hat{\mathbf{a}}(t))$ also has to have the same value for the objective function – i.e., $\hat{\mathbf{a}} = \mathbf{0}$, and the attack will not be detected.

Since (A, C) is observable, from (18) and (10), we have $\Delta x(t) = \mathbf{O}^\dagger\Delta\mathbf{w}(t) + \mathbf{O}^\dagger\mathbf{a}(t) = \mathbf{O}^\dagger\Delta\mathbf{w}(t) + z$. As $\Delta\mathbf{w}(t)$ is bounded, and z is any nonzero vector in the null-space of $\mathbf{O}_{\mathcal{K}^\complement}$, it can be chosen with an arbitrarily large norm. Thus, $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at a single time step.

As the plant (A, C) is observable, the next result follows.

Corollary 1 *System $\Sigma_I(A, C, \delta_w, \mathcal{S})$ (i.e., all sensors compromised) is perfectly attackable at a single time step.*

Corollary 2 *If the attack to the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ has the form $\mathbf{a}(t) = \mathbf{O}z(t)$, for some $z \in \mathbb{R}^n$, then $\hat{\mathbf{a}}(t) = 0$ and the RSE error satisfies $\Delta x(t) = \mathbf{O}^\dagger\Delta\mathbf{w}(t) + z$.*

Remark 2 *Although in this paper we consider l_0-based estimators, it is straightforward to show that the results of Theorem 1 are valid for any batch processing estimators like l_1-based estimator or estimators from [18, 17].*

Example 1 *To illustrate PA at time point, consider system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ with $\delta_w = 0$, $\mathcal{K} = \mathcal{S} = \{s_1\}$, $N = 2$, $A = \left[\begin{smallmatrix} .3 & 1 \\ 0 & .5 \end{smallmatrix}\right]$, $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$. The attack vector $\mathbf{a}(t) = \left[a^T(t) \ a^T(t+1)\right]^T = \mathbf{O}z$ results in estimation error $\Delta x(t) = z$ and $\hat{\mathbf{a}}(t) = 0$, for z being any arbitrary nonzero vector; thus, can generate a perfect attack vector at time t.*

We now provide a necessary and sufficient condition that the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA over time.

Theorem 2 *Consider the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ and let us define the matrix $F(\mathcal{K}, N)$ as*

$$F(\mathcal{K}, N) = \left[\mathbf{O}_{\mathcal{K}^\complement}^T \ (\mathcal{P}_{\mathcal{K}} C)^T \ ... \ (\mathcal{P}_{\mathcal{K}} C A^{N-2})^T\right]^T. \quad (19)$$

a) Suppose $F(\mathcal{K}, N)$ is not full rank. Then, the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is perfectly attackable over time if and only if it is perfectly attackable at a single time step.

b) Suppose $F(\mathcal{K}, N)$ is full rank. Then $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA over time if and only if it is PA at a single time step, A is unstable and at least one eigenvector v_i corresponding to an unstable eigenvalue satisfies $v_i \in \mathcal{N}(\mathcal{O}_{\mathcal{K}^\complement})$.

From Theorem 2 it holds that, unlike the notion of PA in systems with probabilistic noise and statistical IDs [24, 7, 3], for systems with bounded noise and l_0-based RSEs, a system can be perfectly attackable over time even if the plant is not unstable. Before proving Theorem 2, first we introduce the following lemmas used in the proof.

Lemma 1 *Consider attack on the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ in the form $\mathbf{a}(t) = \mathbf{O}z(t)$, where it also holds that if $\mathcal{N}(F(\mathcal{K}, N)) = 0$ then $z(t) \notin \mathcal{N}(A)$. If $z(t+1) = Az(t) + \alpha(t)$, where $\alpha(t) \in \mathcal{N}(F(\mathcal{K}, N))$, then $\mathbf{a}(t+1) = \mathbf{O}z(t+1)$ is also a stealthy attack vector for the system.*

PROOF. For $\mathbf{a}(t+1)$ to be a feasible attack vector, we need to show that $\mathcal{P}_{\mathcal{K}^\complement}\mathbf{a}(t+1) = 0$, which is equivalent to

$$\mathbf{O}_{\mathcal{K}^\complement} z(t+1) = \mathbf{O}_{\mathcal{K}^\complement}(Az(t) + \alpha(t)) = 0. \quad (20)$$

As $\alpha(t) \in \mathcal{N}(F(\mathcal{K}, N))$, we have $\mathbf{O}_{\mathcal{K}^\complement}\alpha(t) = 0$. From Cayley-Hamilton theorem and assumption that $\mathbf{O}_{\mathcal{K}^\complement} z(t) = 0$ (since $\mathcal{P}_{\mathcal{K}^\complement}\mathbf{a}(t) = 0$ as sensors from $\mathcal{K}^\complement$ are not compromised), we have $\mathbf{O}_{\mathcal{K}^\complement} Az(t) = 0$; thus, (20) also holds.

We now show time consistency for the attacks; i.e., that the corresponding elements of $\mathcal{P}_{\mathcal{K}}a(t+1)$,...,$\mathcal{P}_{\mathcal{K}}a(t+N-1)$ of vectors $\mathcal{P}_{\mathcal{K}}\mathbf{a}(t)$ and $\mathcal{P}_{\mathcal{K}}\mathbf{a}(t+1)$ are equal. We have

$$\mathcal{P}_{\mathcal{K}}\mathbf{a}(t+1) = \Big[(\mathcal{P}_{\mathcal{K}}C)^T \ \dots \ (\mathcal{P}_{\mathcal{K}}CA^{N-1})^T\Big]^T (Az(t) + \alpha(t))$$
$$= \Big[(\mathcal{P}_{\mathcal{K}}CA)^T \ \dots \ (\mathcal{P}_{\mathcal{K}}CA^{N})^T\Big]^T z(t) + \mathbf{O}_{\mathcal{K}}\alpha(t)$$
$$\mathcal{P}_{\mathcal{K}}\mathbf{a}(t) = \Big[(\mathcal{P}_{\mathcal{K}}C)^T \ \dots \ (\mathcal{P}_{\mathcal{K}}CA^{N-1})^T\Big]^T z(t).$$

By comparing the shared elements of vectors $\mathcal{P}_{\mathcal{K}}\mathbf{a}(t)$ and $\mathcal{P}_{\mathcal{K}}\mathbf{a}(t+1)$ we have that they are equal, ending the proof.

Lemma 2 *Let the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$, at two consecutive time steps t and $t+1$ have estimation error $\Delta x(t)$ and $\Delta x(t+1)$, while $\mathcal{D}\big(\hat{\mathbf{a}}(t)\big) = \mathcal{D}\big(\hat{\mathbf{a}}(t+1)\big) = 0$. Then $\Delta x(t+1) = A\Delta x(t) + \alpha(t) + p(t)$, where $\alpha(t) \in \mathcal{N}(F(\mathcal{K}, N))$ and $p(t)$ is a bounded vector.*

PROOF. From $\mathcal{D}_I\big(\hat{\mathbf{a}}(t)\big) = \mathcal{D}_I\big(\hat{\mathbf{a}}(t+1)\big) = 0$, as in (18)

$$\begin{aligned} \mathbf{a}(t) &= \mathbf{O}\Delta x(t) + \Delta\mathbf{w}(t) \\ \mathbf{a}(t+1) &= \mathbf{O}\Delta x(t+1) + \Delta\mathbf{w}(t+1). \end{aligned} \quad (21)$$

Since the sensors from $\mathcal{K}^\complement$ are not compromised, $\mathbf{O}_{\mathcal{K}^\complement}\Delta x(t) = -\mathcal{P}_{\mathcal{K}^\complement}\Delta\mathbf{w}(t)$ holds. Thus, we have that

$$\mathbf{O}_{\mathcal{K}^\complement} A\Delta x(t) = \Big[(\mathcal{P}_{\mathcal{K}^\complement}CA)^T \ \dots \ (\mathcal{P}_{\mathcal{K}^\complement}CA^N)^T\Big]\Delta x(t) =$$
$$\Big[-(\mathcal{P}_{\mathcal{K}^\complement}\Delta w^t(t+1))^T \ \dots \ -(\mathcal{P}_{\mathcal{K}^\complement}\Delta w^t(t+N-1))^T \ \beta^T(t)\Big]^T$$
$$= h(t),$$

where $\beta(t) = \mathcal{P}_{\mathcal{K}^\complement}CA^N\Delta x(t)$ and $\Delta w^t(k) = w(k) - \hat{w}^t(k)$ for $k = t, ..., t+N-1$. Using Cayley-Hamilton theorem, it holds that $A^N = c_0 I + ... + c_{N-1}A^{N-1}$, for some $c_0, ..., c_{N-1} \in \mathbb{R}$. Thus, we get $\beta(t) = \mathcal{P}_{\mathcal{K}^\complement}\big(c_0\Delta w^t(t) + ... + c_{N-1}\Delta w^t(t+N-1)\big)$. On the other hand, we have $\mathbf{O}_{\mathcal{K}^\complement}\Delta x(t+1) = -\mathcal{P}_{\mathcal{K}^\complement}\Delta\mathbf{w}(t+1)$. Hence, it follows that

$$\begin{aligned} \mathbf{O}_{\mathcal{K}^\complement}(\Delta x(t+1) - A\Delta x(t)) &= -\mathcal{P}_{\mathcal{K}^\complement}\Delta\mathbf{w}(t+1) - h(t) \\ &\triangleq r_1(t) \end{aligned} \quad (22)$$

The attack vectors for compromised sensors are

$$\mathcal{P}_{\mathcal{K}}\mathbf{a}(t) = \Big[(\mathcal{P}_{\mathcal{K}}C)^T \ (\mathcal{P}_{\mathcal{K}}CA)^T \ \dots \ (\mathcal{P}_{\mathcal{K}}CA^{N-1})^T\Big]^T\Delta x(t) + \mathcal{P}_{\mathcal{K}}\Delta\mathbf{w}(t)$$

$$\mathcal{P}_{\mathcal{K}}\mathbf{a}(t+1) = \Big[(\mathcal{P}_{\mathcal{K}}C)^T \ (\mathcal{P}_{\mathcal{K}}CA)^T \ \dots \ (\mathcal{P}_{\mathcal{K}}CA^{N-1})^T\Big]^T \times \Delta x(t+1) + \mathcal{P}_{\mathcal{K}}\Delta\mathbf{w}(t+1)$$

Let us define $\mathbf{O}_{\mathcal{K}}^{N-2} = \Big[(\mathcal{P}_{\mathcal{K}}C)^T \ \cdots \ (\mathcal{P}_{\mathcal{K}}CA^{N-2})^T\Big]^T$. Consistency in the overlapping terms of the vectors in above equations, implies that

$$\mathbf{O}_{\mathcal{K}}^{N-2}\Delta x(t+1) + \begin{bmatrix} \mathcal{P}_{\mathcal{K}}\Delta w^{t+1}(t+1) \\ \vdots \\ \mathcal{P}_{\mathcal{K}}\Delta w^{t+1}(t+N-1) \end{bmatrix} =$$
$$\mathbf{O}_{\mathcal{K}}^{N-2}A\Delta x(t) + \begin{bmatrix} \mathcal{P}_{\mathcal{K}}\Delta w^{t}(t+1) \\ \vdots \\ \mathcal{P}_{\mathcal{K}}\Delta w^{t}(t+N-1) \end{bmatrix}.$$

Now, we define

$$r_2(t) \triangleq \begin{bmatrix} \mathcal{P}_{\mathcal{K}}\Delta w^{t}(t+1) \\ \vdots \\ \mathcal{P}_{\mathcal{K}}\Delta w^{t}(t+N-1) \end{bmatrix} - \begin{bmatrix} \mathcal{P}_{\mathcal{K}}\Delta w^{t+1}(t+1) \\ \vdots \\ \mathcal{P}_{\mathcal{K}}\Delta w^{t+1}(t+N-1) \end{bmatrix}.$$

Combining the above equation with (22) results in

$$F(\mathcal{K}, N)(\Delta x(t+1) - A\Delta x(t)) = \begin{bmatrix} r_1(t) \\ r_2(t) \end{bmatrix} \triangleq r(t), \quad (23)$$

where $r(t)$ is bounded since both $r_1(t)$ and $r_2(t)$ are bounded. Thus, the solution of (23) can be captured as

$$\Delta x(t+1) - A\Delta x(t) = \alpha(t) + p(t) \quad (24)$$

where $p(t) \in \mathbb{R}^n$ is any bounded vector that satisfies $F(\mathcal{K}, N)p(t) = r(t)$ and $\alpha(t) \in \mathcal{N}(F(\mathcal{K}, N))$.

Lemma 3 *Suppose that $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at a single time step and $\Delta x(t)$ is bounded while $\hat{\mathbf{a}}(t) = 0$. If $F(\mathcal{K}, N)$*

is full rank, then there exists no attack vector $\mathbf{a}(t+1)$ *such that* $\Delta x(t+1)$ *becomes arbitrarily large while* $\hat{\mathbf{a}}(t+1) = 0$.

PROOF. Assume that we can find attack $\mathbf{a}(t+1)$ such that $\Delta x(t+1)$ becomes arbitrarily large while $\hat{\mathbf{a}}(t+1) = 0$. Since $\Delta x(t)$ is bounded and $\hat{\mathbf{a}}(t) = 0$, it means that $\mathbf{a}(t)$ is also bounded. Also, $\mathcal{P}_{\mathcal{K}^{\complement}} a(t+N) = 0$. Let us define the augmented vectors as $\mathbf{a}_{\mathcal{F}}(t+1) = \Big[a^T(t+1)|...|a^T(t+N-1)|\mathcal{P}_{\mathcal{K}^{\complement}} a^T(t+N)\Big]^T$; similarly $\Delta\mathbf{w}_{\mathcal{F}}(t+1)$. Then, from the constraint of (11) we have that $\Delta x(t+1) = F^{\dagger}(\mathcal{K}, N)\mathbf{a}_{\mathcal{F}}(t+1) - F^{\dagger}(\mathcal{K}, N)\mathbf{w}_{\mathcal{F}}(t+1)$. The right side of the equation is bounded whereas the left may be arbitrarily large, which is a contradiction.

Corollary 3 *If* $F(\mathcal{K}, N)$ *for the system* $\Sigma_I(A, C, \delta_w, \mathcal{K})$ *is full rank, then a stealthy cannot induce an unbounded estimation error in the initial step of the attack.*

PROOF. Before starting attack at time t_0, the estimation error is bounded. Now, based on Lemma 3, if the matrix $F(\mathcal{K}, N)$ is full rank, it will be impossible to have unbounded estimation error $\Delta x(t_0)$ while $\hat{\mathbf{a}}(t_0) = 0$

Lemma 4 *There exists a nonzero attack vector* $\mathbf{a}(t)$ *(i.e.,* $\epsilon < \|\mathbf{a}(t)\|$ *with* $\epsilon > 0$*) such that* $\mathcal{D}(\hat{\mathbf{a}}(i), \hat{x}(i)) = 0$ *for any* $t-(N-1) \leq i \leq t+(N-1)$.

PROOF. The claim should be proven for both $\mathcal{D}_I$ and $\mathcal{D}_{II}$. As the proof for $\mathcal{D}_{II}$ also covers the case for $\mathcal{D}_I$ IDs, due to space constraint, we will focus on $\mathcal{D}_{II}$.

Based on the stealthiness condition $\mathcal{D}_{II}(\hat{\mathbf{a}}(i)) = 0$ for any $t-(N-1) \leq i \leq t+N-1$, it holds that

$$\mathbf{y}(i) = \mathbf{O}x(i) + \mathbf{w}(i) + \mathbf{a}(i) = \mathbf{O}\hat{x}(i) + \hat{\mathbf{w}}(i). \quad (25)$$

$\hat{\mathbf{w}}(i) = \mathbf{w}(i) + \mathbf{a}(i)$ and $\hat{x}(i) = x(i)$ are a feasible point for constraint (25). In this case, $||\hat{x}(i) - A\hat{x}(i-1)|| = ||x(i) - Ax(i-1)|| = 0$ satisfies the second stealthiness condition of ID_{II} from (13), for any i. Thus, we need to find a nonzero attack $\mathbf{a}(i)$ such that $||\hat{\mathbf{w}}(i)|| \leq \sqrt{N}\delta_w$ is satisfied. If for any $t-(N-1) \leq i \leq t+N-1$ it holds that $||\mathbf{w}(i)|| < \sqrt{N}\delta_w$, then any nonzero attack vector satisfying $||\mathbf{a}(i)|| < \sqrt{N}\delta_w - ||\mathbf{w}(i)||$ is stealthy – note that no other constraint beyond the norm-bound is required. Similarly, if for some i', $||\mathbf{w}(i')|| = \sqrt{N}\delta_w$, then $\mathbf{a}(i) = \gamma\mathbf{w}(i) + \mathbf{a}'(i)$ with any $\mathbf{a}'(i)$ satisfying $||\mathbf{a}'(i)|| \leq (1-|\gamma+1|)\sqrt{N}\delta_w$ and $-2 < \gamma < 0$ is a stealthy nonzero attack vector. (again, $\mathbf{a}'(i')$ in only norm constrained).

Remark 3 *In Definitions 1, 2, we only focus on whether there exists such a sequence of nonzero stealthy attack vectors that results in unbounded estimation errors, and thus, making the system PA – i.e., we do not consider how the attacker attempts to find it.*

PROOF. [Proof of Theorem 2]
a) First, assume that the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA over time. Based on Remark 1, it is also PA at a single time step. Inversely, assume that $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at a single time step. Suppose that the attack starts at time t_0. Thus, $\mathcal{D}_I(\hat{\mathbf{a}}(t)) = 0$ for any $t < t_0 - (N-1)$. The augmented attack vector $\mathbf{a}(t_0-(N-1))$ will be

$$\mathbf{a}(t_0-(N-1)) = \Big[0^T \ (\mathcal{P}_{\mathcal{K}}\mathbf{a}(t_0-(N-1)))^T\Big]^T, \quad (26)$$

$$\text{where} \quad \mathcal{P}_{\mathcal{K}}\mathbf{a}(t_0-(N-1)) = \Big[0 \ \dots \ 0 \ (\mathcal{P}_{\mathcal{K}} a(t_0))^T\Big]^T.$$

As $F(\mathcal{K}, N)$ is not full rank, there exists a nonzero vector $z(t_0-(N-1))$ where $F(\mathcal{K}, N)z(t_0-(N-1)) = 0$ and

$$\begin{aligned}\mathbf{a}(t_0-(N-1)) &= \Big[0 \ \dots \ 0 \ (\mathcal{P}_{\mathcal{K}} a(t_0))^T\Big]^T \\ &= \begin{bmatrix}\mathbf{O}_{\mathcal{K}^{\complement}} \\ \mathbf{O}_{\mathcal{K}}\end{bmatrix} z(t_0-(N-1)) = \mathbf{O}z(t_0-(N-1)).\end{aligned} \quad (27)$$

Here, $z(t_0-(N-1))$ can be chosen arbitrarily large – i.e., $\mathbf{a}(t_0-(N-1))$ is a perfect attack vector. Now, from Lemma 1, the consecutive perfect attack vectors can also be constructed using $\mathbf{a}(t) = \mathbf{O}z(t)$ with $z(t) = A^{t-t_0+(N-1)}z(t_0-(N-1)) + \sum_{i=t_0}^{t-1} A^{t-i-1}\alpha(i)$ for any $t > t_0-(N-1)$, where $\alpha(i) \in \mathcal{N}(F(\mathcal{K}, N))$. Since α can be arbitrarily large, the system will have arbitrarily large estimation error for $t \geq t_0-(N-1)$ while remaining stealthy from ID_I – i.e., $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA over time.

b) ($\Leftarrow$) Suppose that A is unstable and the system is PA at a time step; thus, $\mathbf{O}_{\mathcal{K}^{\complement}}$ is not full rank. From Lemma 4 (and its proof), there exists a nonzero attack vector $\mathbf{a}(t_0)$ such that for any $t_0-(N-1) \leq i \leq t_0$, $\mathcal{D}_I(\hat{\mathbf{a}}(i)) = 0$, as well as $\mathbf{a}(t_0) = \mathbf{O}z(t_0)$ and $\mathbf{O}_{\mathcal{K}^{\complement}}z(t_0) = 0$ (this holds, from the proof of the lemma which only constraints $\mathbf{a}(t_0)$ to have a certain norm bound).

Based on Lemma 1 if $z(t_0+1) = Az(t_0) + \mathcal{N}(F(\mathcal{K}, N))$, it is possible to have $\mathbf{a}(t_0+1) = \mathbf{O}z(t_0+1)$ with $\mathcal{D}_I(\hat{\mathbf{a}}(t_0+1)) = 0$. Since $F(\mathcal{K}, N)$ is full rank, $\mathcal{N}(F(\mathcal{K}, N)) = 0$. By continuing inserting attack vector in the form of $\mathbf{a}(t) = \mathbf{O}z(t)$ for a period of time $[t_0, t]$, we can get $z(t) = A^{t-t_0}z(t_0)$. Now, we consider two cases:

Case I – The unstable eigenvalues of the matrix A are diagonizable. Let us denote by $v_1, ..., v_q$ eigenvectors that correspond to unstable eigenvalues of matrix A, which we sometimes refer to as 'unstable' eigenvectors. From the theorem assumption, one of these eigenvectors $v_i \in \mathcal{N}(\mathcal{O}_{\mathcal{K}^{\complement}})$, $i \in \{1, ..., q\}$. Now, if we consider $z(t_0) = cv_i \neq 0$, we get $\mathbf{O}_{\mathcal{K}^{\complement}}z(t_0) = 0$ where c is chosen so that $\|\mathbf{O}z(t_0)\| = \epsilon$, for some $\epsilon > 0$. Hence, we get $z(t) = A^{t-t_0}z(t_0) = c\lambda_i^{t-t_0}v_i$. Since $|\lambda_i| > 1$, $||z(t)||$ will be unbounded if $t \to \infty$. Therefore, based on the Corollary 2 and Definition 2 the system is PA over time.

Case II – Unstable eigenvalues of A are not diagonizable and we consider generalized eigenvectors. For each independent eigenvector v_i associated with $|\lambda_i| \geq 1$, we index its generalized eigenvector chain with length of q_i as $v_{i+1}, ..., v_{i+q_i}$, where $v_i \in \mathcal{N}(\mathcal{O}_{\mathcal{K}^{\complement}})$. Now, consider $z(t_0) = cv_{i+q_i}$. Similarly to the **Case I**, we get $z(t) = A^{t-t_0}z(t_0) = \sum_{l=0}^{q_i} c\binom{t-t_0}{l}\lambda_i^{t-t_0-l}v_{i+q_i-l}$ [2]. Since $|\lambda_i| \geq 1$, $z(t)$ is unbounded when $t \to \infty$. Hence, the system will be PA over time.

($\Rightarrow$) Let us assume that the system is PA over time and A is stable. From Definition 2, for all $M > 0$ there exists a time step t' such that for any $t \geq t'$, $||\Delta x(t)|| > M$.

Since $F(\mathcal{K}, N)$ is full rank, from Corollary 3, the estimation error is bounded when attack starts at $t_0 + N - 1$, i.e., $||\Delta x(t_0)|| \leq \delta$ for some $\delta > 0$. Now, for the interval $t_0 < t < t'$ from Lemma 2, $\Delta x(t) = A^{t-t_0}\Delta x(t_0) + \sum_{i=t_0}^{t-1} A^{t-i-1}p(i)$. Since the eigenvectors of A span $\mathbb{R}^n$ (here we assume A is diagonizable, yet, the results can be easily extended to the undiagonizable case), we have

$$\begin{aligned}
\|\Delta x(t')\| &= \left\| A^{t'-t_0}\Delta x(t_0) + \sum_{i=t_0}^{t'-1} A^{t'-i-1}p(i) \right\| \\
&= \left\| \sum_{j=1}^{n} d_j \lambda_j^{t'-t_0} v_j + \sum_{i=t_0}^{t'-1}\sum_{j=1}^{n} d'_{i,j}\lambda_j^{t'-i-1}v_j \right\| \\
&\leq |\lambda_{max}|^{t'-t_0} \|\Delta x(t_0)\| + \left\| \sum_{i=t_0}^{t'-1} |\lambda_{max}|^{t'-i-1} p(i) \right\| \\
&\leq \delta + \frac{1}{1-|\lambda_{max}|} p_{max},
\end{aligned}$$

where λ_{max} is the largest-norm eigenvalue and $p_{max} = \max_{t_0 \leq i \leq t'-1} ||p(i)||$. As $|\lambda_{max}| < 1$ (A is stable), for all $t' > t_0$, we have $|\lambda_{max}|^{t'-t_0} < 1$. Thus, $\|\Delta x(t')\|$ is bounded for any $t' > t_0$, contradicting that the system is PA.

Now, assume that none of the unstable eigenvectors of A belong to $\mathcal{N}(\mathcal{O}_{\mathcal{K}^{\complement}})$. Again, we have that $\Delta x(t')$ can be written as $\Delta x(t') = c_1(t')v_1 + ... + c_q(t')v_q + \sigma(t')$, where $c_j(t') = (d_j\lambda_j^{t'-t_0} + \sum_{i=t_0}^{t'-1} d'_{i,j}\lambda_j^{t'-i-1})v_j$ for $j = 1, ..., q$ and $\sigma(t')$ is the expansion of $\Delta x(t')$ over stable eigenvalues (satisfying $\sigma(t') \to 0$ as $t' \to \infty$). As the system is PA over time, at least one of the coefficients $c_j(t')$ should be arbitrarily large as t' increases. Now, since $\hat{\mathbf{a}}(t) = 0$, it follows that $\mathbf{O}_{\mathcal{K}^{\complement}}\Delta x(t') = \mathcal{P}_{\mathcal{K}^{\complement}}\Delta \mathbf{w}(t')$, making $\mathbf{O}_{\mathcal{K}^{\complement}}\Delta x(t')$ bounded. Thus, $\mathbf{O}_{\mathcal{K}^{\complement}}c_j(t')v_j$ is bounded because $v_1, ..., v_n$ span $\mathbb{R}^n$ (the results can be easily extended to the case where A is not-diagonizable) and the other unstable eigenvectors cannot be used to compensate for $c_j(t')v_j$. From $\mathbf{O}_{\mathcal{K}^{\complement}}c_j(t')v_j$ being bounded while $c_j(t')$ is arbitrarily large, it holds that $v_j \in \mathcal{N}(\mathbf{O}_{\mathcal{K}^{\complement}})$, which is a contradiction – i.e., there exists an unstable eigenvector that lies in $\mathcal{N}(\mathbf{O}_{\mathcal{K}^{\complement}})$.

Example 2 *Consider the system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ from Example 1; it holds that $F(\mathcal{K}, N) = C$ for $N = 2$. If we assume that attack starts at time zero, then it suffices to have $\mathbf{a}(-1) = \begin{bmatrix} a(-1) & a(0) \end{bmatrix}^T = \mathbf{O}z(-1)$ with $a(-1) = 0$. By solving this equation, we get $z(-1) = \begin{bmatrix} 0 & \eta \end{bmatrix}^T$ where η can be chosen arbitrarily large to impose unbounded estimation error at time -1 (consider that although the attack starts at time 0, delay of the RSE causes unbounded error even at time -1). By choosing $z(t) = Az(t-1)$ for $t \geq 0$, and using $\mathbf{a}(t) = \begin{bmatrix} a(t-1) & a(t) \end{bmatrix}^T = \mathbf{O}z(t)$, the attack vector can be constructed over time. However, if we choose $N = 3$, it is impossible to find the attack vector $\mathbf{a}(-2) = \begin{bmatrix} a(-2) & a(-1) & a(0) \end{bmatrix}^T = \mathbf{O}z(-2)$ with $a(-2) = a(-1) = 0$, and since matrix A is stable, it is impossible to perfectly attack the system over time.*

3.2 Perfect Attackabilty for $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$

As previously described, for the system $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ only PA over time should be considered; we now capture necessary and sufficient conditions.

Theorem 3 *System $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ is PA over time if and only if $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at a single time step, A is unstable and least one eigenvector v_i corresponding to an unstable eigenvalue satisfies $v_i \in \mathcal{N}(\mathcal{O}_{\mathcal{K}^{\complement}})$.*

PROOF. ($\Rightarrow$) Assume that $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ is PA over time. Then from Remark 1, $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at single time step. Hence, we need to show that A is unstable. So, let us assume that A is stable while $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ is PA over time. From Definition 2, $\forall M > 0$ there exists a time point t' such that for all $k \geq t'$, $||\Delta x(k)|| > M$. Now, let us assume that the attack starts at $t_0 + N - 1$. Since $\Delta x(t_0 - 1)$ is bounded, from $\mathcal{D}_{II}(\hat{\mathbf{a}}(t_0), \hat{x}(t_0)) = 0$ there exists $\delta > 0$ such that $||\Delta x(t_0)|| \leq \delta$. Now, for the interval $t_0 < t \leq t'$ we have $\Delta x(t) = A^{t-t_0}\Delta x(t_0) + \sum_{i=t_0}^{t-1} A^{t-i-1}(p(i) + \alpha(i))$ with $||\Delta x(t_0)|| \leq \delta$. On the other hand, by combining the condition $\mathcal{D}_{II}(\hat{\mathbf{a}}, \hat{x}) = 0$ for all time steps $t \geq t_0$ with (4) ($x(t) = Ax(t-1)$) we get

$$\begin{aligned}
&||\hat{x}(t) - A\hat{x}(t-1)|| \\
&= ||\hat{x}(t) - A\hat{x}(t-1) - x(t) + Ax(t-1)|| \\
&= ||\Delta x(t) - A\Delta x(t-1)|| = ||p(t) + \alpha(t)|| \leq d
\end{aligned} \tag{28}$$

Since the eigenvectors of A span the space $\mathbb{R}^n$ (here we assume the matrix A is diagonizable, however, the results can be easily extended to the undiagonizable case), it holds that $\Delta x(t_0) = \alpha_1 v_1 + ... + \alpha_n v_n$, $p(i) = \beta_{i,1}v_1 + ... + \beta_{i,n}v_n$ and $\alpha(i) = \gamma_{i,1}v_1 + ... + \gamma_{i,n}v_n$. Now, we have

$$||\Delta x(t')|| = \left\| A^{t'-t_0}\Delta x(t_0) + \sum_{i=t_0}^{t'-1} A^{t'-i-1}(p(i)+\alpha(i)) \right\|$$
$$= \left\| \sum_{j=1}^{n} \alpha_j \lambda_j^{t'-t_0} v_j + \sum_{i=t_0}^{t'-1}\sum_{j=1}^{n} (\beta_{i,j}+\gamma_{i,j})\lambda_j^{t'-i-1} v_j \right\|$$
$$\leq |\lambda_{max}|^{t'-t_0}\,\|\Delta x(t_0)\| + \left\| \sum_{i=t_0}^{t-1} |\lambda_{max}|^{t'-i-1}(p(i)+\alpha(i)) \right\|$$
$$\leq \delta + \frac{d}{1-|\lambda_{max}|},$$

where λ_{max} is the eigenvalue with the largest absolute value. Based on our assumption, $|\lambda_{max}| < 1$ and for $t > t_0$ we have also $|\lambda_{max}|^{t-t_0} < 1$. Hence, $\|\Delta x(t')\|$ will be bounded for any $t' > t_0$, contradicting our assumption that the system $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ is PA. Finally, proof that at least one unstable eigenvector belongs to $\mathcal{N}(\mathcal{O}_{\mathcal{K}^{\complement}}))$ directly follows the approach for Theorem 2.

($\Leftarrow$) Suppose that matrix A has at least one eigenvalue outside the unit circle. From Lemma 4, there exists a nonzero attack vector $\mathbf{a}(t_0)$ such that for any $t_0 - (N-1) \leq i \leq t_0$, $\mathcal{D}_{II}(\hat{\mathbf{a}}(i)) = 0$. Thus, there exists $\epsilon > 0$ such that $\|\mathbf{a}(t_0)\| = \epsilon$, and similarly to the proof of Theorem 2, we can consider $\mathbf{a}(t_0) = \mathbf{O}z(t_0)$. Since $\mathbf{O}$ is full rank and $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at a single time point, $z(t_0)$ can be any nonzero vector that satisfies $\|\mathbf{O}z(t_0)\| = \epsilon$ with $\mathbf{O}_{\mathcal{K}^{\complement}} z(t_0) = 0$; any such vector $z(t_0)$ may be chosen arbitrarily by the attacker.

From Lemma 1 if $z(t_0+1) = Az(t_0)$, then attack $\mathbf{a}(t_0+1) = \mathbf{O}z(t_0+1)$ results in $\mathcal{D}_I(\hat{\mathbf{a}}(t_0+1)) = 0$. Now, we need to show $||\hat{x}(t_0+1) - A\hat{x}(t_0)|| \leq d$. From Corollary 2

$$||\hat{x}(t_0+1) - A\hat{x}(t_0)|| = ||\Delta x(t_0+1) - A\Delta x(t_0)||$$
$$= ||\mathbf{O}^{\dagger}\Delta\mathbf{w}(t_0+1) + z(t_0+1) - A\mathbf{O}^{\dagger}\Delta\mathbf{w}(t_0) - Az(t_0)||$$
$$= ||\mathbf{O}^{\dagger}\Delta\mathbf{w}(t_0+1) - A\mathbf{O}^{\dagger}\Delta\mathbf{w}(t_0)|| \leq d$$

By continuing with attacks in the form of $\mathbf{a}(t) = \mathbf{O}z(t)$ for a period of time $[t_0, t]$, we get $z(t) = A^{t-t_0}z(t_0)$ while remaining stealthy from ID_{II}. Now, consider two cases:

Case I – The unstable eigenvalues of A are diagonizable. Let us denote by $v_1, ..., v_q$ eigenvectors that correspond to unstable eigenvalues of matrix A. From our assumption, there exists $v_i \in \mathcal{N}(\mathcal{O}_{\mathcal{K}^{\complement}})$, $i \in \{1, ..., q\}$. Now, if we consider $z(t_0) = cv_i \neq 0$, we get $\mathbf{O}_{\mathcal{K}^{\complement}} z(t_0) = 0$, where c is chosen such that $\|\mathbf{O}z(t_0)\| \leq \epsilon$. Thus, $z(t) = A^{t-t_0}z(t_0) = c\lambda_1^{t-t_0}v_i$. Since $|\lambda_i| > 1$, $||z(t)||$ will be unbounded if $t \to \infty$. Hence, from Corollary 2 and Definition 2, the system will be PA over time.

Case II – The unstable eigenvalues of A are not diagonizable and we consider generalized eigenvectors. For each independent eigenvector v_i associated with $|\lambda_i| \geq 1$, we index its generalized eigenvector chain with length q_i as $v_{i+1}, ..., v_{i+q_i}$. Consider $z(t_0) = cv_{i+q_i}$, where $v_i \in \mathcal{N}(\mathcal{O}_{\mathcal{K}^{\complement}})$. Similarly to **Case I**, $z(t) = A^{t-t_0}z(t_0) = \sum_{l=0}^{q_i} c\binom{t-t_0}{l}\lambda_i^{t-t_0-l} v_{i+q_i-1}$ [2]. Since $\|\lambda_i| \geq 1$, $z(t)$ is unbounded as $t \to \infty$, and from Corollary 2 the system is PA over time.

The condition of PA over time for $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ is the same as for $\Sigma_I(A, C, \delta_w, \mathcal{K})$ when $F(\mathcal{K}, N)$ is full rank. When $F(\mathcal{K}, N)$ is rank deficient, we can use $N = n+1$ to make the matrix full rank and get the same PA condition as for $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$. Yet, increasing N would increase computational overhead at each time step, which may be a problem in resource-constrained systems. Instead, one can use $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ (i.e., ID_{II}) that only requires additional comparison, from (13), at each time step.

4 Estimation with Intermittent Authentication

We now study the effects of intermittent data authentication (sometimes refered to as intermittent integrity enforcement [3]) on estimation error of $\Sigma(A, C, \delta_w, \mathcal{K})$.

Definition 3 *The intermittent data authentication policy for i-th sensor ($s_i \in \mathcal{S}$), denoted by (μ_i, L_i) where $\mu_i = \{t_k^i\}_{k=0}^{\infty}$ such that $t_k^i > t_{k-1}^i$ and $L_i = sup(t_k^i - t_{k-1}^i)$, ensures that $a_i(t_k^i) = 0$.*

Intermittent data authentication for sensor i guarantees that the attack injected through the i-th sensor is zero at some specific points (t_k^i), whereas the interval between each of consecutive points is at most L_i time steps. A global intermittent authentication policy is defined if all sensors use same (μ_i, L_i). We now capture conditions that $\Sigma_I(A, C, \delta_w, \mathcal{K})$, satisfying Theorem 1, is not PA.

Theorem 4 *If $\mathcal{I}_i \subseteq \mathcal{S}$, $i \in \{1, ..., N\}$, consider matrix*

$$\mathbf{O}_{\mathcal{I},\mathcal{K}^{\complement}} = \left[(\mathcal{P}_{\mathcal{I}_1 \cup \mathcal{K}^{\complement}} C)^T \ ... (\mathcal{P}_{\mathcal{I}_N \cup \mathcal{K}^{\complement}} C A^{N-1})^T \right]^T. \quad (29)$$

If intermittent data authentication is used at time $t+i$ for each sensor set $\mathcal{I}_i$, $i \in \{0, ..., N-1\}$, then $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is not PA at time t if and only if $\mathbf{O}_{\mathcal{I},\mathcal{K}^{\complement}}$ is full rank.

PROOF. ($\Leftarrow$) Suppose $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at time t. Since for any $i \in 1, ..., N$ intermittent data authentication is used, $\mathcal{P}_{\mathcal{I}_i} a(t+i) = 0$, and thus

$$\mathbf{O}_{\mathcal{I},\mathcal{K}^{\complement}} x(t) + \begin{bmatrix} \mathcal{P}_{\mathcal{I}_1 \cup \mathcal{K}^{\complement}} w(t) \\ \vdots \\ \mathcal{P}_{\mathcal{I}_N \cup \mathcal{K}^{\complement}} w(t+N-1) \end{bmatrix} =$$
$$\mathbf{O}_{\mathcal{I},\mathcal{K}^{\complement}} \hat{x}(t) + \begin{bmatrix} \mathcal{P}_{\mathcal{I}_1 \cup \mathcal{K}^{\complement}} \hat{w}(t) \\ \vdots \\ \mathcal{P}_{\mathcal{I}_N \cup \mathcal{K}^{\complement}} \hat{w}(t+N-1) \end{bmatrix} \Rightarrow$$

$$\mathbf{O}_{\mathcal{I},\mathcal{K}^{\complement}}\Delta x(t) = \\ = \begin{bmatrix} \mathcal{P}_{\mathcal{I}_1\cup\mathcal{K}^{\complement}}\Big(w(t)-\hat{w}(t)\Big) \\ \vdots \\ \mathcal{P}_{\mathcal{I}_N\cup\mathcal{K}^{\complement}}\Big(w(t+N-1)-\hat{w}(t+N-1)\Big) \end{bmatrix} \quad (30)$$

We denote the right side of (30) as $f(t)$. Since $\mathbf{O}_{\mathcal{I},\mathcal{K}^{\complement}}$ is full rank, from (30) we have $\Delta x(t) = \mathbf{O}^{\dagger}_{\mathcal{I},\mathcal{K}^{\complement}} f(t)$; i.e., $\|\Delta x(t)\| = \left\|\mathbf{O}^{\dagger}_{\mathcal{I},\mathcal{K}^{\complement}} f(t)\right\| \leq \left\|\mathbf{O}^{\dagger}_{\mathcal{I},\mathcal{K}^{\complement}}\right\| \|f(t)\| = \bar{c}$. As the actual and estimated noise are bounded, $\|f(t)\|$ and $\bar{c}$ are also bounded, contradicting PA of the system.

($\Rightarrow$) System $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is not PA at time t, and assume that $\mathbf{O}_{\mathcal{I},\mathcal{K}^{\complement}}$ is not full rank. Then, exists a nonzero vector z such that $\mathbf{O}_{\mathcal{I},\mathcal{K}^{\complement}} z = 0$; thus, $\mathbf{O}_{\mathcal{K}^{\complement}} z = 0$ and the pair $(\mathcal{P}_{\mathcal{K}^{\complement}} C, A)$ is not observable, From Theorem 1, $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is PA at t, which is a contradiction.

Theorem 4 provides an intermittent data authentication policy such that the system is not PA at a single time step. To derive conditions of not being PA for all time steps, the condition of Theorem 4 should be satisfied at each time. Our goal is to derive conditions that a system is not PA over time, and we start with the following.

Proposition 3 *Assume $F(\mathcal{S}, N)$ is not full rank. Then system $\Sigma_I(A, C, \delta_w, \mathcal{K})$ is not PA over time for any compromised sensor set $\mathcal{K}$, if the intermittent data authentication policy is used with $L_i = 1$, $\forall i \in \mathcal{F}$, where $\mathcal{F}$ is a sensor subset such that the pair $(A, \mathcal{P}_{\mathcal{F}} C)$ is observable.*

PROOF. As for any $t > 0$, authentication is used in $\{t, ..., t+N-1\}$, $\mathcal{P}_{\mathcal{F}} a(t) = \mathcal{P}_{\mathcal{F}} a(t+1) = ... = \mathcal{P}_{\mathcal{F}} a(t+N-1) = 0$. The corresponding matrix for the authentication policy is $\mathbf{O}_{\mathcal{F}} = \begin{bmatrix} \mathcal{P}_{\mathcal{F}} C^T & \mathcal{P}_{\mathcal{F}}(CA)^T & ... & \mathcal{P}_{\mathcal{F}}(CA^{N-1})^T \end{bmatrix}^T$ Since $\mathbf{O}_{\mathcal{F}}$ is full rank, from Theorem 4, system $\Sigma(A, C, \mathcal{K})$ is not PA at time t. As this holds for any time t, from Definition 2 the system is not PA over time.

From Proposition 3 it follows that if matrix $F(\mathcal{S}, N)$ is not full rank, then we can avoid PA over time by using data authentication at each time step for some specific subset of sensors. Although this may seem conservative, but in the following example, we show that a perfect attack can be achieved by only compromising suitable sensors at a single time step.

Example 3 *Consider again the model from Example 1, and assume that the attack is only injected at time zero. There are two vectors $z(-1)$, $z(0) \in \mathbb{R}^2$ which can satisfy $\mathbf{a}(-1) = \begin{bmatrix} a(-1) \\ a(0) \end{bmatrix} = \mathbf{O} z(-1) = \begin{bmatrix} 1 & 0 \\ .3 & 1 \end{bmatrix} \begin{bmatrix} z_1(-1) \\ z_2(-1) \end{bmatrix}$, $\mathbf{a}(0) = \begin{bmatrix} a(0) \\ a(1) \end{bmatrix} = \mathbf{O} z(0) = \begin{bmatrix} 1 & 0 \\ .3 & 1 \end{bmatrix} \begin{bmatrix} z_1(0) \\ z_2(0) \end{bmatrix}$, with $a(-1) = a(1) = 0$. Solving the above two equations gives $a(0) = z_2(-1) = z_1(0) = -\frac{z_2(0)}{.3}$ and $z_1(-1) = 0$. Using Corollary 2, we get $\Delta x(-1) = \begin{bmatrix} 0 & a(0) \end{bmatrix}^T$ and $\Delta x(0) = \begin{bmatrix} a(0) & -.3a(0) \end{bmatrix}^T$ (which can be chosen arbitrarily large by controlling the scalar $a(0)$), whereas ID_I will not trigger alarm in these two time steps. Consider that $a(0)$ is not included in other time steps; thus, by inserting attack vector only at time zero, the system $\Sigma_I(A, C, 0, s_1)$ can have unbounded estimation error without triggering alarm.*

The above example shows that for $\Sigma_I(A, C, \delta_w, \mathcal{K})$ when $F(\mathcal{S}, N)$ is not full rank, a stealthy attack can result in arbitrarily large estimation error, even by injecting false data only at one time step. Hence, it is essential to use data authentication at all time steps – i.e., non-intermittently. However, as shown below, when $F(\mathcal{S}, N)$ is full rank, $\Sigma_I(A, C, \delta_w, \mathcal{K})$ cannot be PA over time even when only intermittent authentication is used; this holds for $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$ independently of the $F(\mathcal{S}, N)$ rank.

Theorem 5 *Consider two cases: a) $\Sigma_I(A, C, \delta_w, \mathcal{K})$ with full rank $F(\mathcal{S}, N)$; b) $\Sigma_{II}(A, C, \delta_w, \mathcal{K})$. Both (a) and (b) are not PA over time if the intermittent authentication policy is used with $L_i = \mathcal{T}$, $\forall i \in \mathcal{F}$ for a bounded $\mathcal{T}$, where $\mathcal{F}$ is any sensor set such that $(A, \mathcal{P}_{\mathcal{F}} C)$ is observable.*

PROOF. From Lemma 2 and (18), it follows that

$$\begin{aligned} \Delta x(t+1) =& A\Delta x(t) + \alpha(t) + p(t) \\ \mathbf{a}(t) =& \mathbf{O}\Delta x(t) + \Delta \mathbf{w}(t) \end{aligned} \quad (31)$$

for any $t \geq t_0$ if the attacker initiates the attack at time $t_0 + N - 1$. For system (a), $\alpha(t) = 0$ and since $p(t)$ is bounded at all time steps t, thus $p(t) + \alpha(t)$ is bounded. For system (b) the stealthiness condition $||\Delta x(t+1) - A\Delta x(t)|| < d$ causes $||p(t) + \alpha(t)|| < d$. Thus, for both cases $p(t) + \alpha(t)$ is bounded. Assume t_{k_0} is the first time instant that authentication is used after t_0. Then $\forall i \in \mathcal{F}$ we have

$$a_i(t_{k_0}) = a_i(t_{k_0} + \mathcal{T}) = ... = 0 \quad (32)$$

Hence, $\mathcal{P}_{\mathcal{F}} a(t_{k_0}) = \mathcal{P}_{\mathcal{F}} a(t_{k_0} + \mathcal{T}) = ... = 0$. On the other hand, from (31) we get $\mathcal{P}_{\mathcal{F}} a(t) = \mathcal{P}_{\mathcal{F}} C\Delta x(t) + \mathcal{P}_{\mathcal{F}}\Delta w(t)$ for any $t \geq t_0 + N - 1$. Now, consider $\mathcal{P}_{\mathcal{F}} a(t_{k_0} + i\mathcal{T})$ for any $i \geq 0$. Then, for $j \in \{1, ..., N-1\}$ we have

$$\begin{aligned} &\mathcal{P}_{\mathcal{F}} a(t_{k_0} + (i+j)\mathcal{T}) = \mathcal{P}_{\mathcal{F}} C A^{j\mathcal{T}} \Delta x(t_{k_0} + i\mathcal{T}) \\ &+ \sum_{f=t_{k_0}+i\mathcal{T}}^{t_{k_0}+(i+j)\mathcal{T}-1} \mathcal{P}_{\mathcal{F}} C A^{t_{k_0}+(i+j)\mathcal{T}-1-f}(p(f) + \alpha(f)) \\ &+ \mathcal{P}_{\mathcal{F}}\Delta w(t_{k_0} + (i+j)\mathcal{T}) = 0; \end{aligned}$$

for $j = 0$, $\mathcal{P}_{\mathcal{F}} a(t_{k_0} + (i+j)\mathcal{T}) = \mathcal{P}_{\mathcal{F}} C A^{j\mathcal{T}} \Delta x(t_{k_0} + i\mathcal{T})$. By augmenting $\mathcal{P}_{\mathcal{F}} a(t_{k_0} + (i+j)\mathcal{T}), \forall j \in \{0, ..., N-1\}$,

$$\begin{bmatrix} \mathcal{P}_{\mathcal{F}} a(t_{k_0} + i\mathcal{T}) \\ \vdots \\ \mathcal{P}_{\mathcal{F}} a(t_{k_0} + (i+N-1)\mathcal{T}) \end{bmatrix} = 0 \Rightarrow$$

$$\begin{bmatrix} (\mathcal{P}_{\mathcal{F}} C)^T & \dots & (\mathcal{P}_{\mathcal{F}} C A^{(N-1)\mathcal{T}})^T \end{bmatrix}^T \Delta x(t_{k_0} + i\mathcal{T}) =$$

$$\begin{bmatrix} \sum_{f=t_{k_0}+i\mathcal{T}}^{t_{k_0}+(i)\mathcal{T}-1} \mathcal{P}_{\mathcal{F}} C A^{t_{k_0}+(i)\mathcal{T}-1-f} (p(f) + \alpha(f)) \\ \vdots \\ \sum_{f=t_{k_0}+i\mathcal{T}}^{t_{k_0}+(i+N-1)\mathcal{T}-1} \mathcal{P}_{\mathcal{F}} C A^{t_{k_0}+(i+N-1)\mathcal{T}-1-f} (p(f) + \alpha(f)) \end{bmatrix}$$

$$+ \begin{bmatrix} \mathcal{P}_{\mathcal{F}} \Delta w(t_{k_0} + (i)\mathcal{T}) \\ \vdots \\ \mathcal{P}_{\mathcal{F}} \Delta w(t_{k_0} + (i+N-1)\mathcal{T}) \end{bmatrix}$$

Now, since $\begin{bmatrix} (\mathcal{P}_{\mathcal{F}} C)^T & (\mathcal{P}_{\mathcal{F}} C A^{\mathcal{T}})^T & \dots & (\mathcal{P}_{\mathcal{F}} C A^{(N-1)\mathcal{T}})^T \end{bmatrix}^T$ is full rank and the right side of the above equation is bounded, we have $\Delta x(t_{k_0} + i\mathcal{T})$ is bounded for any $i \geq 0$. On the other hand, from (31) and the fact that $\alpha(t)$ and $p(t)$ are bounded for $t \geq t_0$, we can conclude that $\Delta x(t)$ is bounded for any $i\mathcal{T} \leq t \leq (i+1)\mathcal{T}$ for any $i \geq 0$.

5 Numerical Results

We illustrate our results on a realistic case study – Vehicle Trajectory Following (VTF). Specifically, we show how the attacker can perfectly attack the system when the necessary conditions are satisfied and how intermittent data authentication effectively prevents such attacks. We consider the model from [4], discretized with sampling time .01 s; i.e., $A = \left[\begin{smallmatrix} 1 & .01 \\ 0 & 1 \end{smallmatrix}\right], B = \left[\begin{smallmatrix} .0001 \\ .01 \end{smallmatrix}\right], C = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{smallmatrix}\right]^T$. We assume that all sensors are compromised – i.e., $\mathcal{K} = \mathcal{S} = \{s_1, s_2, s_3\}$. Therefore, the system is PA over time as A is also unstable. For simulation, we also assume that each element of sensor and system noise comes from uniform distribution $v_P, v_M \sim U(-.05, .05)$. Moreover, $N = 2$ and the maximum possible estimation error when the system is not under attack is obtained as $||\Delta x|| \leq ||\mathbf{O}^{\dagger}||||\Delta \mathbf{w}|| = .0789$ from (15).

Fig. 2 shows the evolution of the l_2 norm of estimation error in different scenarios. In Fig. 2a, $||\Delta x(t)||$ is shown when the system is not under attack, whereas in Fig 2b the system is under a perfect attack. In Fig 2c, we considered two different data authentication policies μ_{10} and μ_{100}; meaning $L = 10$ and $L = 100$, respectively, while the system is under by stealthy attack. As shown, when data authentication is used, the system is not PA over time and the estimation error remains bounded, and very low for less than 10% of authenticated measurements; as the period of authentication increases, a stealthy attack can achieve higher maximum estimation error.

Finally, we considered resilient state estimation within the VTF – trajectory tracking; Fig 3 shows 60 seconds simulation. As shown, if a data authentication policy is used with $L = 10$ (i.e., 10% of authenticated messages), we obtain suitable control performance even under stealthy attack. If authentication is not used, a stealthy attack can force the system from the desired path.

6 Conclusion

In this work, we considered the problem of resilient state estimation for LTI systems with bounded noise, when a subset of sensors are under attack. We defined two notions of perfect attackability (PA) – at a time point and over time – where stealthy attacks can cause an arbitrarily large estimation errors, and derived necessary and sufficient conditions for PA. We showed that, unlike the Kalman filter-based observers, batch processing-based resilient state estimators (RSE), such as l_0-based RSE, may be perfectly attackable even if the plant is not unstable. Furthermore, we studied the effects of intermittent data authentication on attack-induced estimation error. We showed that it is sufficient to even intermittently use data authentication, once every bounded time period, to ensure that a system is not perfectly attackable.

References

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

[2] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU press, 2012.

[3] I. Jovanov and M. Pajic. Relaxing integrity requirements for attack-resilient cyber-physical systems. *IEEE Transactions on Automatic Control*, 64(12): 4843–4858, 2019. ISSN 2334-3303.

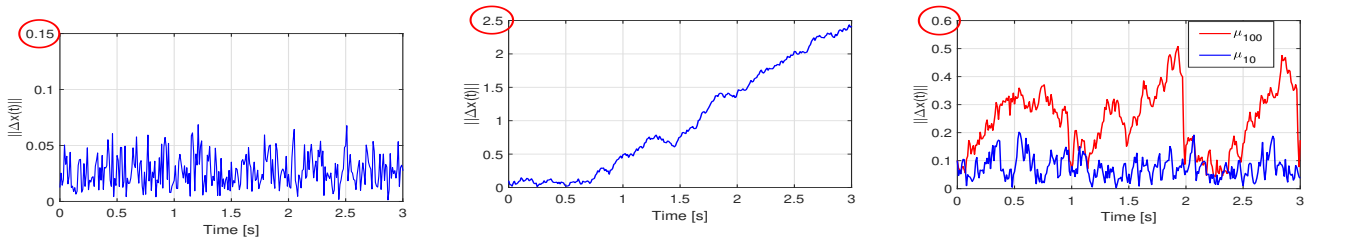
[4] A. J. Kerns, D. P. Shepard, J. A. Bhatti, and T. E. Humphreys. Unmanned aircraft capture and control via gps spoofing. *Journal of Field Robotics*, 31 (4):617–636, 2014.

[5] A. Khazraei and M. Pajic. Perfect attackability of linear dynamical systems with bounded noise. In *2020 American Control Conference (ACC)*, 2020.

[6] A. Khazraei, H. Kebriaei, and F. R. Salmasi. A new watermarking approach for replay attack detection in lqg systems. In *56th IEEE Annual Conf. on Decision and Control (CDC)*, pages 5143–5148, 2017.

[7] C. Kwon, W. Liu, and I. Hwang. Analysis and design of stealthy cyber attacks on unmanned aerial systems. *Journal of Aerospace Information Systems*, 11(8):525–539, 2014.

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

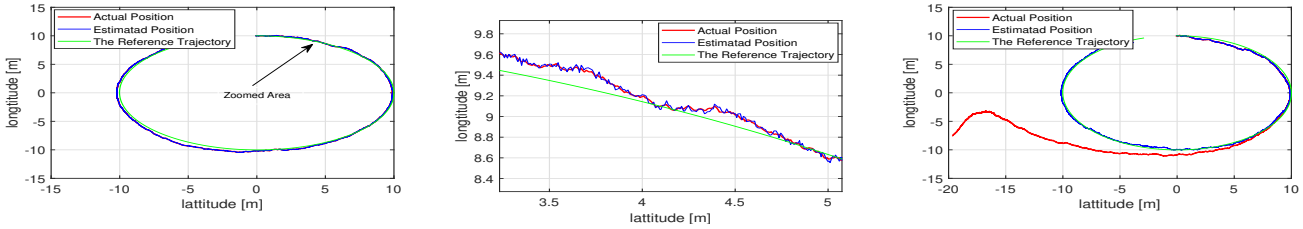


(a) System without attack

(b) Under a perfect attack, without authentication

(c) With two different integrity enforcement policies, with periods $L = 10, 100$

Fig. 2. Evolution of the estimation error norm for the VTF system; note the highlighted (circled) different error ranges.



(a) State estimates under stealthy attack with integrity enforcement policy $L = 10$.

(b) Zoomed area of (a)

(c) State estimates under stealthy attack without integrity enforcement

Fig. 3. Simultaneous trajectory tracking and state estimation for the VTF system. At each time step, system state is estimated and used by the controller to track a circle trajectory. The duration of the simulation is 60 s and the attack starts at $t = 20$ s.

[9] V. Lesi, I. Jovanov, and M. Pajic. Security-aware scheduling of embedded control tasks. *ACM Trans. Embed. Comput. Syst.*, 16(5s):188:1–188:21, 2017.

[10] V. Lesi, I. Jovanov, and M. Pajic. Integrating security in resource-constrained cyber-physical systems. *ACM Trans. on Cyber-Physical Systems*, 2020. `https://arxiv.org/abs/1811.03538`.

[11] X. Luo, M. Pajic, and M. M. Zavlanos. A scalable and optimal graph-search method for secure state estimation. *arXiv preprint arXiv:1903.10620*, 2019.

[12] Y. Mo and B. Sinopoli. Secure control against replay attacks. In *47th Annual Allerton Conference on Communication, Control, and Computing*, pages 911–918. IEEE, 2009.

[13] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas. Robustness of attack-resilient state estimators. In *ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, pages 163–174, April 2014.

[14] M. Pajic, I. Lee, and G. J. Pappas. Attack-resilient state estimation for noisy dynamical systems. *IEEE Transactions on Control of Network Systems*, 4(1): 82–92, 2017.

[15] M. Pajic, J. Weimer, N. Bezzo, O. Sokolsky, G. J. Pappas, and I. Lee. Design and implementation of attack-resilient cyberphysical systems: With a focus on attack-resilient state estimators. *IEEE Control Systems Magazine*, 37(2):66–81, 2017.

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

[17] Y. Shoukry and P. Tabuada. Event-triggered state observers for sparse sensor noise/attacks. *IEEE Trans. on Aut. Control*, 61(8):2079–2091, 2016.

[18] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada. Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach. *IEEE Transactions on Automatic Control*, 62(10):4917–4932, 2017.

[19] R. S. Smith. Covert misappropriation of networked control systems: Presenting a feedback structure. *IEEE Control Systems Magazine*, 35(1): 82–92, 2015.

[20] S. Sundaram and C. N. Hadjicostis. Distributed function calculation via linear iterative strategies in the presence of malicious agents. *IEEE Transactions on Automatic Control*, 56(7):1495–1508, 2011.

[21] S. Sundaram, M. Pajic, C. Hadjicostis, R. Mangharam, and G. Pappas. The Wireless Control Network: Monitoring for Malicious Behavior. In *49th IEEE Conference on Decision and Control (CDC)*, pages 5979–5984, Dec 2010.

[22] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson. Attack models and scenarios for networked control systems. In *1st ACM Int. Conf. on High Confidence Netw. Systems*, pages 55–64, 2012.

[23] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135–148, 2015.

[24] Y. Mo and B. Sinopoli. False data injection attacks in control systems. In *First workshop on Secure Control Systems*, pages 1–6, 2010.

[25] K. Zetter. Inside the cunning, unprecedented hack of ukraine's power grid. *Wired*, 2016.