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

The Security Requirement to Prevent Zero Dynamics Attacks and Perfectly Undetectable Cyber-Attacks in Cyber-Physical Systems

Mahdi Taheri[1], Khashayar Khorasani[1], and Nader Meskin[2]

Abstract— **The impacts of zero dynamics attacks and perfectly undetectable cyber-attacks cannot be observed in outputs of cyber-physical systems (CPS). Adversaries are capable of executing these cyber-attacks and leading the CPS to undesirable trajectories while remaining undetected. In this paper, we introduce and formally define the notion of security effort (SE) as a novel security metric for CPS that determines the minimum number of actuators and sensors that should be secured and kept attack free in order to prevent adversaries from executing zero dynamics attacks, covert attacks, and controllable attacks. Moreover, since zero dynamics attacks, covert attacks, and controllable attacks belong to weakly unobservable and controllable weakly unobservable subspaces of the CPS, conditions under which these subspaces become zero are obtained and investigated. An illustrative numerical simulation is provided to demonstrate the effectiveness of our proposed security measure.**

I. Introduction

In recent years, there has been a growing number of cyber-attacks targeting cyber-physical systems (CPS) [1]. These cyber-attacks inflict damages on critical infrastructures such as power grids and cause financial losses [1]. Addressing the cyber security problems in CPS has attracted a significant amount of attention [2]–[8]. On the other hand, adversaries attempt to make their cyber-attacks stealthy in order to avoid being detected by monitoring systems in the CPS. Therefore, one of the main challenges in addressing cyber-security problems in CPS is to investigate the vulnerability of these systems to stealthy cyber-attacks.

The impact of stealthy cyber-attacks on sensor readings received by CPS operators is minimal. Hence, their detection is extremely challenging. Zero dynamics attacks, covert attacks, and controllable attacks, that are referred to as undetectable cyber-attacks in [4], [5], [7], [9], are among the most damaging stealthy cyber-attacks since they can cause CPS to become unstable while remaining undetected. Moreover, covert attacks and controllable attacks are considered as *perfectly undetectable* cyber-attacks since their impacts on sensor readings are zero [2].

The notion of security index has been defined in the literature as the minimum number of actuators and sensors that should be manipulated by adversaries to perform certain types of stealthy cyber-attacks [2]–[4], [9], [10]. In [4], [9], [10], the security index is defined as the minimum number of actuators and sensors that are needed to execute a zero dynamics attack, a covert attack, or a controllable attack. Furthermore, in [2] and [3], the security index denotes the minimum number of required inputs and outputs to perform a perfectly undetectable cyber-attack.

Moreover, computing the security index is an NP-hard problem [2], [10]. Therefore, in [2] and [3] structural system framework has been utilized to describe the CPS by using graph theory to compute the security index in a generic sense by using computationally efficient algorithms. In [4] and [9], an upper bound of the security index is defined and geometric control theory is utilized to compute the security index over the weakly unobservable and controllable weakly unobservable subspaces of the CPS.

However, the notion of security index considers the CPS from the adversary's point of view. Hence, in certain cases, it may not provide CPS operators with adequate information to prevent zero dynamics attacks, covert attacks, and controllable attacks. Consequently, one needs to study a security measure that studies the CPS from the operator's perspective. In this paper, we define the security effort (SE) as a measure that denotes the minimum number of actuators and sensors that should be secured by CPS operators to prevent adversaries from executing zero dynamics attacks, covert attacks, and controllable attacks.

In particular, we investigate the weakly unobservable and controllable weakly unobservable subspaces of the CPS since zero dynamics attacks, covert attacks, and controllable attacks belong to these subspaces. Consequently, conditions under which weakly unobservable and controllable weakly unobservable subspaces become zero are derived and investigated. Consequently, if these conditions are satisfied, adversaries are not capable of performing zero dynamics attacks, covert attacks, and controllable attacks. Furthermore, SE is defined as the minimum number of inputs and outputs that should be secured such that weakly unobservable and controllable weakly unobservable subspaces of the CPS become zero.

To summarize, the main contributions of this paper are stated as follows:

1) The notion of SE is formally defined as a measure that denotes the minimum number of actuators and sensors that should be secured to prevent adversaries from executing zero dynamics attacks, covert attacks, and controllable attacks.
2) Conditions under which the weakly unobservable subspace of CPS becomes zero are developed and investigated. If these conditions are satisfied, no zero dynamics attacks, covert attacks, and controllable attacks can be performed by the adversaries on the CPS.
3) In order to study perfectly undetectable cyber-attacks, conditions under which the controllable weakly unob-

[1]Mahdi Taheri (m_eri@encs.concordia.ca) and Khashayar Khorasani (kash@ece.concordia.ca) are with the Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada.

[2]Nader Meskin (nader.meskin@qu.edu.qa) is with the Department of Electrical Engineering, Qatar University, Doha, Qatar.

The authors would like to acknowledge the financial support received from NATO under the Emerging Security Challenges Division program. K. Khorasani and N. Meskin would like to acknowledge the support received from NPRP grant number 10-0105-17017 from the Qatar National Research Fund (a member of Qatar Foundation). K. Khorasani would also like to acknowledge the support received from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Department of National Defence (DND) under the Discovery Grant and DND Supplemental Programs. This work was also supported in part by funding from the Innovation for Defence Excellence and Security (IDEaS) program from the Department of National Defence (DND). Any opinions and conclusions in this work are strictly those of the authors and do not reflect the views, positions, or policies of - and are not endorsed by - IDEaS, DND, or the Government of Canada.

servable subspace of CPS becomes zero are investigated. Therefore, under these conditions, adversaries cannot execute perfectly undetectable cyber-attacks, i.e., covert attacks and controllable attacks.

Given that one of the main objectives of CPS operators is to secure their systems against undetectable cyber-attacks, it is imperative that they are made aware of the baseline security requirements in terms of disruption resources to accomplish this goal. This paper introduces and specifies the SE as the metric to study and compute the above requirement. However, the security index does not provide the CPS operators with such information, rather it indicates the minimum number of compromised sensors and actuators necessary for performing undetectable cyber-attacks.

Consequently, the main difference between the SE and the security index lies in their perspectives on analyzing the CPS. Moreover, considering that zero dynamics and perfectly undetectable cyber-attacks belong to weakly unobservable and controllable weakly unobservable subspaces of CPS [7], [9], the SE indicates the minimum required input and output channels that should be secured to make the above subspaces zero. Hence, provided the conditions in this paper on making weakly unobservable and controllable weakly unobservable subspaces are utilized then one can determine the SE for the CPS.

II. Problem Statement and Formulation

A. Model of the CPS

We consider a linear time-invariant (LTI) CPS in the following form:

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

where $x(t) \in \mathbb{R}^n$ is the state, $y(t) \in \mathbb{R}^p$ is the output, and $u(t) \in \mathbb{R}^m$ denotes the control input. The characteristic matrices of the system, i.e., (A, B, C), are of appropriate dimensions. We assume that B is an injective map, i.e., B has full column rank, since otherwise, its linearly dependent columns can be removed.

B. CPS Under Cyber-Attacks

Let $\mathcal{U} = \{u_1, \ldots, u_{\mathrm{m}}\}$ and $\mathcal{Y} = \{y_1, \ldots, y_{\mathrm{p}}\}$ denote the sets of input and output communication channels in the CPS (1) with $|\mathcal{U}| = m$ and $|\mathcal{Y}| = p$, respectively, where $|\cdot|$ denotes the cardinality of a set. Moreover, let $\mathcal{U}_{\mathrm{s}}$ and $\mathcal{Y}_{\mathrm{s}}$ denote the sets of secured input and output channels of the CPS, respectively. Consequently, $\mathcal{U}_{\mathrm{a}} = \mathcal{U}/\mathcal{U}_{\mathrm{s}} = \{u^{\mathrm{a}}_1, \ldots, u^{\mathrm{a}}_{m_{\mathrm{a}}}\}$, with $|\mathcal{U}_{\mathrm{a}}| = m_{\mathrm{a}}$ is the set of attacked inputs and $\mathcal{Y}_{\mathrm{a}} = \mathcal{Y}/\mathcal{Y}_{\mathrm{s}} = \{y^{\mathrm{a}}_1, \ldots, y^{\mathrm{a}}_{p_{\mathrm{a}}}\}$, with $|\mathcal{Y}_{\mathrm{a}}| = p_{\mathrm{a}}$ denotes the set of attacked outputs.

The CPS (1) under cyber-attacks can be expressed by

$$\begin{aligned} \dot{x}(t) =& Ax(t) + B(u(t) + L_{\mathrm{a}} a_{\mathrm{u}}(t)), \\ y(t) =& Cx(t) + D_{\mathrm{a}} a_{\mathrm{y}}(t), \end{aligned} \tag{2}$$

where $a_{\mathrm{u}}(t) \in \mathbb{R}^{m_{\mathrm{a}}}$ is the actuator attack signal and $a_{\mathrm{y}}(t) \in \mathbb{R}^p$ is the sensor attack signal. Moreover, $B_{\mathrm{a}} = BL_{\mathrm{a}}$ and D_{a} are the actuator attack and the sensor attack signatures, respectively. The matrix $D_{\mathrm{a}} = \mathrm{diag}(d_1, d_2, \ldots, d_p) \in \mathbb{R}^{p \times p}$ is diagonal, where $d_r = 1$ if the r-th sensor measurement belongs to the set $\mathcal{Y}_{\mathrm{a}}$ for $r = 1, \ldots, p$, and $d_r = 0$ if $y_r \in \mathcal{Y}_{\mathrm{s}}$. Hence, one has $\mathrm{rank}(D_{\mathrm{a}}) = p_{\mathrm{a}}$. Furthermore, the matrix $L_{\mathrm{a}} = [l_{u^{\mathrm{a}}_1}, \ldots, l_{u^{\mathrm{a}}_{m_{\mathrm{a}}}}] \in \mathbb{R}^{m \times m_{\mathrm{a}}}$ denotes the input channels that are compromised by adversaries, where u^{a}_q-th element of $l_{u^{\mathrm{a}}_q} \in \mathbb{R}^m$ is equal to 1, and the rest of its entries are zero, for $q = 1, \ldots, m_{\mathrm{a}}$. Consequently, L_{a} is an injective operator, i.e., $\mathrm{rank}(L_{\mathrm{a}}) = m_{\mathrm{a}}$.

C. Various Types of Cyber-Attacks

Given the linearity of the CPS (2), and due to the superposition principle, one can separately consider and study the impact of cyber-attacks and control inputs on the CPS. Hence, we eliminate the effects of $u(t)$ from the CPS in the following form:

$$\begin{aligned} \dot{x}(t) =& Ax(t) + B_{\mathrm{a}} a_{\mathrm{u}}(t), \\ y(t) =& Cx(t) + D_{\mathrm{a}} a_{\mathrm{y}}(t). \end{aligned} \tag{3}$$

Let $Y(x(0), a_{\mathrm{u}}(t), a_{\mathrm{y}}(t))$ denote the output of the CPS (3) as a function of the initial condition $x(0)$, the actuator attack signal $a_{\mathrm{u}}(t)$, and the sensor attack signal $a_{\mathrm{y}}(t)$, $\forall t \geq 0$. This paper is concerned with cyber-attacks that their impacts cannot be observed in the output measurements of CPS. In the following, the above types of cyber-attacks are defined according to [4]–[6], [9].

***Definition** 1:* In the CPS (3), the following cyber-attacks are defined:

1) The actuator attack signal $a_{\mathrm{u}}(t) \neq 0$ is a *zero dynamics attack* if $Y(x(0), a_{\mathrm{u}}(t), 0) = 0$, $\forall t \geq 0$, where $x(0) \neq 0$, and adversaries only need to have access to input communication channels.
2) The attack signal $a(t) = [a_{\mathrm{u}}(t)^\top, a_{\mathrm{y}}(t)^\top]^\top \neq 0$ is a *covert attack* if $Y(0, a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) = 0$, $\forall t \geq 0$. In this type of cyber-attacks, adversaries need to have access to both input and output communication channels.
3) The actuator attack signal $a_{\mathrm{u}}(t) \neq 0$ is a *controllable attack* if $Y(0, a_{\mathrm{u}}(t), 0) = 0$, $\forall t \geq 0$, where adversaries need to compromise input communication channels.

***Definition** 2 ([2]):* The cyber-attack signal $a(t) = [a_{\mathrm{u}}(t)^\top, a_{\mathrm{y}}(t)^\top]^\top \neq 0$ is designated as *perfectly undetectable* if it satisfies $Y(0, a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) = 0$, $\forall t \geq 0$.

Consequently, as per Definitions 1 and 2, there are two types of perfectly undetectable cyber-attacks. First, if one has $a_{\mathrm{u}}(t) \neq 0$ and $a_{\mathrm{y}}(t) \neq 0$ such that $Y(0, a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) = 0$, $\forall t \geq 0$, this is referred to as a *covert attack* in [5], [6], [11]. Second, if $a_{\mathrm{u}}(t) \neq 0$ and $a_{\mathrm{y}}(t) = 0$ such that $Y(0, a_{\mathrm{u}}(t), 0) = 0$, $\forall t \geq 0$, this is defined as a *controllable attack* in [4], [9], and a *zero stealthy attack* in [7].

However, since the cyber-attack that results in having $Y(0, a_{\mathrm{u}}(t), 0) = 0$ is related to the controllable weakly unobservable subspace of the system (see [4] and [12] for more details), we have adopted the convention from [9] and [4] and refer to this type of perfectly undetectable cyber-attacks as controllable attacks. Moreover, despite the fact that the given zero dynamics attack in Definition 1 is not perfectly undetectable (as per Definition 2), under certain initial conditions, it results in a zero output. Hence, in this paper, in addition to perfectly undetectable cyber-attacks, we also investigate zero dynamics attacks.

D. Objectives

Our objectives in this paper are twofold. Our first objective is to develop and study conditions under which adversaries cannot perform zero dynamics attacks, covert attacks, and controllable attacks that are provided in Definition 1. The latter is achieved by studying conditions under which the largest weakly unobservable and the largest controllable weakly unobservable subspace of the CPS are zero. As for our second objective, we formally define a security

measure that determines the minimum number of input and output communication channels that should be secured in order to prevent adversaries from performing certain cyber-attacks that are provided in Definitions 1 and 2. Moreover, the proposed security measure is studied from a geometric control perspective.

III. Investigation of Weakly Unobservable and Controllable Subspaces

In case of the covert attacks, adversaries design their sensor attack signals such that they cancel out the impact of actuator attacks from sensor readings [11]. Hence, the sensor attack signal is designed in the following from:

$$\begin{aligned} \dot{x}_{\mathrm{a}}(t) =& Ax_{\mathrm{a}}(t) + B_{\mathrm{a}}a_{\mathrm{u}}(t), \\ y_{\mathrm{a}}(t) =& -Cx_{\mathrm{a}}(t), \end{aligned} \tag{4}$$

where $x_{\mathrm{a}}(t) \in \mathbb{R}^n$ and $a_{\mathrm{y}}(t) = y_{\mathrm{a}}(t)$. One can augment the dynamics in (3) and (4) into the following form:

$$\begin{aligned} \dot{\check{x}}(t) =& \check{A}\check{x}(t) + \check{B}_{\mathrm{a}}a_{\mathrm{u}}(t), \\ y(t) =& \check{C}\check{x}(t), \end{aligned} \tag{5}$$

where $\check{x}(t) = [x(t)^\top,\, x_{\mathrm{a}}(t)^\top]^\top$, $y(t) = Cx(t) + D_{\mathrm{a}}y_{\mathrm{a}}(t)$, $\check{A} = \mathrm{diag}(A, A)$, $\check{B}_{\mathrm{a}} = [B_{\mathrm{a}}^\top,\, B_{\mathrm{a}}^\top]^\top$, and $\check{C} = [C,\, -D_{\mathrm{a}}C]$. In terms of the main advantage of the augmented system (5), only the actuator attack signal is an input to the system, and the sensor attack signal $a_{\mathrm{y}}(t)$ is expressed by using the dynamics given by (4). Let $\check{Y}(\check{x}(0), a_{\mathrm{u}}(t))$ represent the output of the augmented system (5) as a function of the initial condition $\check{x}(0)$ and the actuator attack signal $a_{\mathrm{u}}(t)$. In the following, it is shown how one can utilize the augmented system (5) in order to study cyber-attacks on the CPS (3). In particular, it is shown that covert attacks, controllable attacks, and zero dynamics attacks in CPS (3) can be equivalently studied in the augmented system (5).

***Theorem** 1:* In the augmented dynamics (5), one has $\check{Y}(\check{x}(0), a_{\mathrm{u}}(t)) = 0$ if and only if there exists a sensor attack signal $a_{\mathrm{y}}(t) \in \mathbb{R}^p$ and $\check{x}(0) = [x(0)^\top,\, x(0)^\top]^\top$ such that $Y(x(0), a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) = 0$ holds true, $\forall t \geq 0$.

Proof:

Necessary Condition: Suppose $\check{Y}(\check{x}(0), a_{\mathrm{u}}(t)) = 0$ holds and for any $a_{\mathrm{y}}(t) \in \mathbb{R}^p$, one has $Y(x(0), a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) \neq 0$, where $\check{x}(0) = [x(0)^\top,\, x(0)^\top]^\top$. It follows from $\check{Y}(\check{x}(0), a_{\mathrm{u}}(t)) = 0$ that $y(t) = Cx(t) + D_{\mathrm{a}}y_{\mathrm{a}}(t) = 0$, $\forall t \geq 0$. Since $Y(x(0), a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) \neq 0$, from (3), one obtains $y(t) = Cx(t) + D_{\mathrm{a}}a_{\mathrm{y}}(t) \neq 0$. However, considering $a_{\mathrm{y}}(t) = y_{\mathrm{a}}(t)$ results in having $y(t) = Cx(t) + D_{\mathrm{a}}a_{\mathrm{y}}(t) = 0$, which contradicts the assumption.

Moreover, suppose $\check{Y}(\check{x}(0), a_{\mathrm{u}}(t)) = 0$ and $Y(x(0), a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) = 0$, where $\check{x}(0) = [x(0)^\top,\, x_{\mathrm{a}}(0)^\top]^\top$ such that $x(0) \neq x_{\mathrm{a}}(0)$. According to the definition of the augmented system (5), having $\check{Y}(\check{x}(0), a_{\mathrm{u}}(t)) = 0$ implies that one either has $Cx(t) = -D_{\mathrm{a}}y_{\mathrm{a}}(t) = D_{\mathrm{a}}Cx_{\mathrm{a}}(t)$ for $Cx(t) \neq 0$ or in the other case, $Cx(t) = 0$ and $D_{\mathrm{a}}y_{\mathrm{a}}(t) = D_{\mathrm{a}}Cx_{\mathrm{a}}(t) = 0$. Since in both (4) and (5) the input is $a_{\mathrm{u}}(t)$, one should have $x(0) = x_{\mathrm{a}}(0)$ for either case of $Cx(t) = D_{\mathrm{a}}Cx_{\mathrm{a}}(t) \neq 0$ or $Cx(t) = 0$ and $D_{\mathrm{a}}Cx_{\mathrm{a}}(t) = 0$ to hold, which contradicts the assumption.

Sufficient Condition: Assume that there exists a sensor attack signal $a_{\mathrm{y}}(t) \in \mathbb{R}^p$ such that $Y(x(0), a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) = 0$ and $\check{x}(0) = [x(0)^\top,\, x(0)^\top]^\top$. Moreover, due to the linearity of (5), one obtains

$$\check{Y}(\check{x}(0), a_{\mathrm{u}}(t)) = Y(x(0), a_{\mathrm{u}}(t), 0) - D_{\mathrm{a}}Y(x(0), a_{\mathrm{u}}(t), 0). \tag{6}$$

If $a(t) = [a_{\mathrm{u}}(t)^\top, a_{\mathrm{y}}(t)^\top]^\top$ is either a zero dynamics attack or a controllable attack, as per Definition 1, one has $a_{\mathrm{y}}(t) = 0$ and $Y(x(0), a_{\mathrm{u}}(t), 0) = 0$. Consequently, it follows from (6) that $\check{Y}(\check{x}(0), a_{\mathrm{u}}(t)) = 0$, $\forall t \geq 0$. Also, if $a(t) = [a_{\mathrm{u}}(t)^\top, a_{\mathrm{y}}(t)^\top]^\top$ is a covert attack, according to Definition 1, $Y(0, a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) = 0$ holds. Consequently, due to the definition of (4), one obtains $a_{\mathrm{y}}(t) = y_{\mathrm{a}}(t)$ and $y(t) = Cx(t) + D_{\mathrm{a}}y_{\mathrm{a}}(t) = 0$, $\forall t \geq 0$. This completes the proof of the theorem. ∎

***Remark** 1:* As the main implication of the Theorem 1, $a(t) = [a_{\mathrm{u}}(t)^\top, a_{\mathrm{y}}(t)^\top]^\top$ in the CPS (3) results in $Y(x(0), a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) = 0$ if and only if $\check{Y}(\check{x}(0), a_{\mathrm{u}}(t)) = 0$, where $\check{x}(0) = [x(0)^\top,\, x(0)^\top]^\top$. Hence, if no zero dynamics attacks, covert attacks, and controllable attacks can be executed in the augmented system (5), the above cyber-attacks cannot be performed on the CPS (3) as well.

A. Cyber-Attacks and the Weakly Unobservable Subspace

Considering Theorem 1 and Remark 1, in order to study zero dynamics attacks, covert attacks, and controllable attacks in the CPS (3), one can study these cyber-attacks in the augmented system (5). Hence, in the following, we study and derive conditions under which zero dynamics attacks, covert attacks, and controllable attacks cannot be performed in (5) from a geometric control theory perspective.

***Definition** 3 (Weakly Unobservable Subspace):* Consider the triple $(\check{C}, \check{A}, \check{B}_{\mathrm{a}})$ in (5). A point $\check{x}(0) \in \mathbb{R}^{2n}$ is defined as weakly unobservable if there exists $a_{\mathrm{u}}(t) \neq 0$ such that the output satisfies $y(t) = 0$, $\forall t \geq 0$. The set of all weakly unobservable points is called weakly unobservable subspace and is denoted by $\mathscr{V}$. Moreover, the largest weakly unobservable subspace is denoted by $\mathscr{V}^*$.

Given the Definition 3, and considering the results in [7, Theorem 1] and [9, Lemma 7], if $\mathscr{V}^* = 0$, no zero dynamics attacks, covert attacks, and controllable attacks can be executed in the augmented system (5). In the following, conditions under which $\mathscr{V}^* = 0$ are studied and proposed.

***Definition** 4 ([13]):* Consider $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Z} \subseteq \mathbb{R}^p$ as finite-dimensional inner product vector spaces and the matrix $Q \in \mathbb{R}^{p \times n}$. One has

1) $\mathcal{Z} = Q\mathcal{X} \coloneqq \{z : z = Qx,\, x \in \mathcal{X}\}$.
2) $\mathcal{X} = Q^{-1}\mathcal{Z} \coloneqq \{x : z = Qx,\, z \in \mathcal{Z}\}$.

***Theorem** 2:* Let $\mathrm{Im}(\check{B}_{\mathrm{a}}) \neq 0$. In the augmented system (5), one has $\mathscr{V}^* = 0$ if for any $\mathscr{S} \subseteq \mathrm{Ker}(\check{C})$, one has $\check{A}\mathrm{Ker}(\check{C}) \cap (\mathscr{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$.

Proof: As described in [13, Algorithm 4.1.2] and [12], the largest weakly unobservable subspace of the system (5) can be computed in $2n$ steps by using the following algorithm:

$$\begin{aligned} \mathscr{V}_0 =& \mathrm{Ker}(\check{C}), \\ \mathscr{V}_k =& \mathscr{V}_0 \cap \check{A}^{-1}(\mathscr{V}_{k-1} + \mathrm{Im}(\check{B}_{\mathrm{a}})). \end{aligned} \tag{7}$$

Since for any $\mathscr{S} \subseteq \mathrm{Ker}(\check{C})$, one has $\check{A}\mathrm{Ker}(\check{C}) \cap (\mathscr{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$, any $g \in \mathrm{Ker}(\check{C})$ results in having $\check{A}g = z \notin \mathscr{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})$. Hence, we have $\check{A}^{-1}(\mathscr{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})) \cap \mathrm{Ker}(\check{C}) = 0$, since otherwise, as per Definition 4, there exists $g \in \mathrm{Ker}(\check{C})$ such that $\check{A}g = z \in \mathscr{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})$, which contradicts having $\check{A}\mathrm{Ker}(\check{C}) \cap (\mathscr{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$. Consequently, according to (7), $\mathscr{V}^* = \mathscr{V}_{2n} = 0$ since for any $\mathscr{V}_{2n-1} \subseteq \ker(\check{C})$, we have $\mathscr{V}_0 \cap \check{A}^{-1}(\mathscr{V}_{2n-1} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$. This completes the proof of the theorem. ∎

B. Perfectly Undetectable Cyber-Attacks and the Controllable Weakly Unobservable Subspace

In this subsection, conditions under which in the augmented system (5) perfectly undetectable cyber-attacks, i.e., covert and controllable attacks, cannot be performed are investigated. However, one needs to first study the following definitions.

***Definition** 5 (Strongly Reachable Subspace [12]):* The subspace $\mathscr{W} \subseteq \mathbb{R}^{2n}$ is the strongly reachable subspace of the triple $(\check{C}, \check{A}, \check{B}_{\mathrm{a}})$ in (5) if $\mathscr{V} = \mathscr{W}^{\perp}$ is the weakly unobservable subspace of $(\check{B}_{\mathrm{a}}^{\top}, \check{A}^{\top}, \check{C}^{\top})$. Moreover, $\mathscr{W}^*$ denotes the smallest strongly reachable subspace.

***Definition** 6 (Controllable Weakly Unobservable [12]):* The subspace $\mathscr{R} \subseteq \mathscr{V}^*$ is designated as the controllable weakly unobservable subspace of the triple $(\check{C}, \check{A}, \check{B}_{\mathrm{a}})$ if one has $\mathscr{R} \subseteq \mathscr{W}^*$. Moreover, $\mathscr{R}^* = \mathscr{V}^* \cap \mathscr{W}^*$ denotes the largest controllable weakly unobservable subspace.

***Definition** 7 (Left-Invertibility [12]):* Let $\check{x}(0) = 0$. The augmented system $(\check{C}, \check{A}, \check{B}_{\mathrm{a}})$ in (5) is left-invertible if for any $y(t) = 0$ one has $a_{\mathrm{u}}(t) = 0$, $\forall t \geq 0$.

***Lemma** 1 ([12]):* Let $\check{\Sigma} = (\check{C}, \check{A}, \check{B}_{\mathrm{a}})$ denote the system in (5). The following statements are equivalent:

1) The system $\check{\Sigma}$ is left-invertible.
2) $\mathscr{R}^* = 0$ and $\check{B}_{\mathrm{a}}$ is injective.
3) $\check{\mathscr{V}}^* \cap \mathrm{Im}(\check{B}_{\mathrm{a}}) = 0$ and $\check{B}_{\mathrm{a}}$ is injective.

As shown in [4, Theorem 1], covert attacks and controllable attacks are related to the controllable weakly unobservable subspace of the system $(\check{C}, \check{A}, \check{B}_{\mathrm{a}})$, i.e., the subspace $\mathscr{R}^*$. Hence, since $\check{B}_{\mathrm{a}}$ is an injective map by definition, from Lemma 1 it follows that adversaries are capable of executing perfectly undetectable cyber-attacks if and only if for the triple $(\check{C}, \check{A}, \check{B}_{\mathrm{a}})$ one has $\mathscr{R}^* \neq 0$, or equivalently, the triple $(\check{C}, \check{A}, \check{B}_{\mathrm{a}})$ is not left-invertible.

***Theorem** 3:* The system $\check{\Sigma} = (\check{C}, \check{A}, \check{B}_{\mathrm{a}})$ in (5) is left-invertible if for any $\mathcal{S} \subseteq \mathrm{Ker}(\check{C})$, one has $\check{A}(\mathrm{Im}(\check{B}_{\mathrm{a}}) \cap \mathrm{Ker}(\check{C})) \cap (\mathcal{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$.

Proof: Considering Lemma 1 and since B_{a} has full column rank by definition, one needs to show that having $\check{A}(\mathrm{Im}(\check{B}_{\mathrm{a}}) \cap \mathrm{Ker}(\check{C})) \cap (\mathcal{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$ for every $\mathcal{S} \subseteq \mathrm{Ker}(\check{C})$ results in $\mathscr{V}^* \cap \mathrm{Im}(\check{B}_{\mathrm{a}}) = 0$.

Having $\check{A}(\mathrm{Im}(\check{B}_{\mathrm{a}}) \cap \mathrm{Ker}(\check{C})) \cap (\mathcal{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$ implies that for any $g \in \mathrm{Im}(\check{B}_{\mathrm{a}}) \cap \mathrm{Ker}(\check{C})$, one has $\check{A}g \notin \mathcal{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})$. Hence, for any $\mathcal{S} \subseteq \mathrm{Ker}(\check{C})$, we have $\mathrm{Im}(\check{B}_{\mathrm{a}}) \cap \mathrm{Ker}(\check{C}) \cap \check{A}^{-1}(\mathcal{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$. Since for $\mathscr{V}_{2n-1} \subseteq \ker(\check{C})$, we have $\mathscr{V}_0 \cap \mathrm{Im}(\check{B}_{\mathrm{a}}) \cap \check{A}^{-1}(\mathscr{V}_{2n-1} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$, one obtains

$$\mathscr{V}_{2n} \cap \mathrm{Im}(\check{B}_{\mathrm{a}}) = \mathscr{V}_0 \cap \mathrm{Im}(\check{B}_{\mathrm{a}}) \cap \check{A}^{-1}(\mathscr{V}_{2n-1} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0.$$

This completes the proof of the theorem. ■

IV. Security Effort

In this section, the security effort (SE) is formally defined as a measure that shows the minimum number of input and output communication channels that should be secured by CPS operators and should be kept attack free to prevent adversaries from executing cyber-attacks that are provided in Definitions 1 and 2. Specifically, it is shown how one can study the SE for a given CPS from a geometric control theory perspective.

A. Definition of the Security Effort (SE)

The SE is defined as the solution to the following optimization problem:

$$\begin{aligned} SE_{\Sigma} \coloneqq \min_{a_{\mathrm{u}}(\cdot), a_{\mathrm{y}}(\cdot)} \quad & m - \|a_{\mathrm{u}}(t)\|_0 + p - \|a_{\mathrm{y}}(t)\|_0 \\ \text{s.t.} \quad & \dot{x}(t) = Ax(t) + Ba_{\mathrm{u}}(t), \\ & y(t) = Cx(t) + a_{\mathrm{y}}(t), \\ & y(t) \neq 0,\ x(0) \in \mathbb{R}^n, \\ & a(t) \neq 0, \end{aligned} \tag{8}$$

where $a(t) = [a_{\mathrm{u}}(t)^{\top}, a_{\mathrm{y}}(t)^{\top}]^{\top}$.

If conditions in (8) are satisfied, adversaries cannot design a cyber-attack signal $a(t)$ that results in $Y(x(0), a_{\mathrm{u}}(t), a_{\mathrm{y}}(t)) = 0$, $\forall t \geq 0$. In other words, in problem (8), SE_{Σ} denotes the minimum number of actuators and sensors that should be secured so that the weakly unobservable subspace of the CPS in (3) is empty, and consequently, no zero dynamics attacks, covert attacks, and controllable attacks can be initiated and executed.

Considering that in order to perform zero dynamics cyber-attacks and perfectly undetectable cyber-attacks, i.e., covert attacks and controllable attacks, adversaries need to have access to at least one input communication channel and actuator, one has $0 < SE_{\Sigma} \leq m$. This implies that in the worst-case scenario, the CPS operators need to secure all the input communication channels to prevent zero dynamics attacks and perfectly undetectable cyber-attacks. However, similar to the problem of computing the security index in [2], computing SE_{Σ} is an NP-hard problem, which makes it computationally intensive to solve.

It follows from Theorem 1 and Definition 3 that if the weakly unobservable subspace of the augmented system (5) is zero, i.e., $\mathscr{V}^* = 0$, no zero dynamics attacks, covert attacks, and controllable attacks can be executed in both the CPS (3) and the augmented system (5). Consequently, according to Theorem 2, $\mathscr{V}^* = 0$ if for any $\mathscr{S} \subseteq \mathrm{Ker}(\check{C})$, one has $\check{A}\mathrm{Ker}(\check{C}) \cap (\mathscr{S} + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$.

Consequently, an upper bound for the SE in problem (8) can be given in the following form:

$$\begin{aligned} \bar{SE}_{\Sigma} \coloneqq \min_{\mathrm{rank}(B_{\mathrm{a}}),\, \mathrm{rank}(D_{\mathrm{a}})} \quad & m - \mathrm{rank}(B_{\mathrm{a}}) + p - \mathrm{rank}(D_{\mathrm{a}}) \\ \text{s.t.} \quad & \check{A}\mathrm{Ker}(\check{C}) \cap (\mathrm{Ker}(\check{C}) + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0. \end{aligned} \tag{9}$$

Consequently, in Algorithm 1, a pseudo code for finding an upper bound on SE_{Σ} is proposed. Let $S = \{u_1, \ldots, u_{\mathrm{m}}, y_1, \ldots, y_{\mathrm{p}}\}$ denote the set of all actuators and sensors of the CPS as described in Section II-B. In Algorithm 1, by utilizing the binary representation of the elements of S, its power set is created. Consequently, the sufficient condition $\check{A}\mathrm{Ker}(\check{C}) \cap (\mathrm{Ker}(\check{C}) + \mathrm{Im}(\check{B}_{\mathrm{a}})) = 0$ is considered for each subset of the power set to check if $\mathscr{V}^* = 0$ is satisfied and to compute $\bar{SE}_{\Sigma}$ as an upper bound for the SE, i.e., $SE_{\Sigma} \leq \bar{SE}_{\Sigma}$. Moreover, one of the outputs of Algorithm 1 is the set $\bar{S}_{\min}$ which contains the actuators and sensors that should be secured to prevent adversaries from performing zero dynamics and perfectly undetectable cyber-attacks in the CPS, where $|\hat{S}_{\min}| = \bar{SE}_{\Sigma}$.

B. Security Effort (SE) for Perfectly Undetectable Cyber-Attacks

The specified SE in optimization problems (8) and (9) are defined to prevent all zero dynamics attacks, covert attacks, and controllable attacks that belong to the weakly

Algorithm 1 Pseudo code to find an upper bound for SE_Σ

Input: $\check{A} = \text{diag}(A, A)$, $\check{B}_\text{a} = [B_\text{a}^\top, B_\text{a}^\top]^\top$, and $\check{C} = [C, -D_\text{a}C]$, $S = \{u_1, \ldots, u_\text{m}, y_1, \ldots, y_\text{p}\}$
Output: $\bar{SE}_\Sigma$, $\bar{S}_\text{min}$

1: Initialize $\bar{SE}_\text{s} = m + p$
2: Set $l = |S|$, where $|\cdot|$ denotes the cardinality of a set
3: **for** $i = 1 : 2^l - 1$ **do**
4: Create the empty set $\hat{S} = \{\}$
5: **for** $j = 1 : l$ **do**
6: **if** the j-th bit of the binary representation of i is equal to 1 **then**
7: Add j-th member of S to $\hat{S}$
8: **end if**
9: **end for**
10: Secure only actuators and sensors that belong to the set $\hat{S}$, update $\check{B}_\text{a}$ and $\check{C}$ accordingly, and set $\mathcal{Q} = \ker(\check{C})$
11: **if** $\check{A}\mathcal{Q} \cap (\ker(\check{C}) + \text{Im}(\check{B}_\text{a})) = 0$ **and** $|\hat{S}| \leq \bar{SE}_\text{s}$ **then**
12: $\bar{SE}_\text{s} = |\hat{S}|$
13: $\hat{S}^* = \hat{S}$
14: **end if**
15: **end for**
16: $\bar{SE}_\Sigma = \min\{\bar{SE}_\text{s}, m\}$ and $\bar{S}_\text{min} = \hat{S}^*$

unobservable subspace of the CPS. However, as per Definition 1, in contrast to zero dynamics attacks, the execution of covert attacks and controllable attacks does not depend on the initial conditions $x(0)$ of the CPS. Moreover, according to Definition 2, covert attacks and controllable attacks are perfectly undetectable cyber-attacks. Hence, one may only be interested in preventing perfectly undetectable cyber-attacks in the CPS (3). Thus, in the following, SE for perfectly undetectable cyber-attacks is formally defined and investigated.

The SE for perfectly undetectable cyber-attacks in the CPS can be expressed as

$$\begin{aligned} \hat{SE}_\Sigma := \min_{a_\text{u}(\cdot), a_\text{y}(\cdot)} \quad & m - \|a_\text{u}(t)\|_0 + p - \|a_\text{y}(t)\|_0 \\ \text{s.t.} \quad & \dot{x}(t) = Ax(t) + Ba_\text{u}(t), \\ & y(t) = Cx(t) + a_\text{y}(t), \\ & y(t) \neq 0,\ x(0) = 0, \\ & a(t) \neq 0, \end{aligned} \tag{10}$$

The only difference between SE_Σ in (8) and $\hat{SE}_\Sigma$ in (10) is that in (10) one has $x(0) = 0$, which implies that zero dynamics attacks are excluded in computing the $\hat{SE}_\Sigma$. Furthermore, given that $\mathscr{R}^* \subseteq \mathscr{V}^*$, one has $\hat{SE}_\Sigma \leq SE_\Sigma$. Moreover, according to Theorem 1, if conditions in (10) hold, no covert attacks and controllable attacks can be executed in the CPS (3) and the augmented system (5). Also, as per Definition 7, the augmented system (5) is left-invertible. Hence, it can be inferred that the optimization problem (10) determines the minimum number of input and output communication channels that should be secured to make the CPS (3) left-invertible. Similar to the case of SE_Σ in (8), (10) is also an NP-hard problem to be solved.

In order to compute the upper bound of the SE for perfectly undetectable cyber-attacks, which is designated as $\hat{\bar{SE}}_\Sigma$, results in Theorem 3 are utilized. It follows from Theorem 3 that $\check{\Sigma} = (\check{C}, \check{A}, \check{B}_\text{a})$ is left-invertible if for any

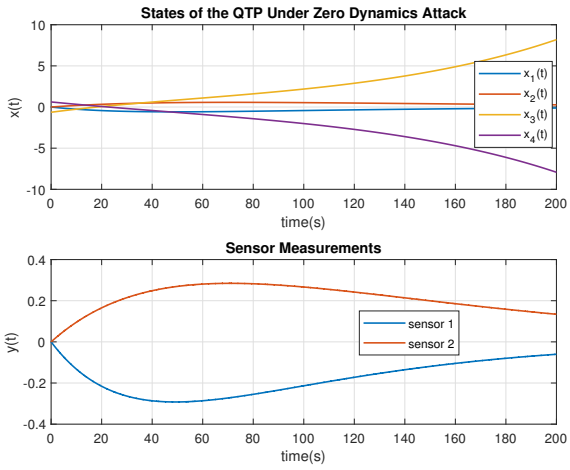

Fig. 1. The QTP under zero dynamics attacks where the states become unbounded while the outputs show an attack-free behavior.

$\mathcal{S} \subseteq \text{Ker}(\check{C})$, one has $\check{A}(\text{Im}(\check{B}_\text{a}) \cap \text{Ker}(\check{C})) \cap (\mathcal{S} + \text{Im}(\check{B}_\text{a})) = 0$. Hence, the problem of computing $\hat{\bar{SE}}_\Sigma$ can be rewritten in the following form:

$$\begin{aligned} \hat{\bar{SE}}_\Sigma := \min_{\text{rank}(B_\text{a}),\, \text{rank}(D_\text{a})} \quad & m - \text{rank}(B_\text{a}) + p - \text{rank}(D_\text{a}) \\ \text{s.t.} \quad & \check{A}(\text{Im}(\check{B}_\text{a}) \cap \text{Ker}(\check{C})) \cap (\text{Ker}(\check{C}) + \text{Im}(\check{B}_\text{a})) = 0. \end{aligned} \tag{11}$$

Therefore, one can modify Algorithm 1 to determine an upper bound for $\hat{SE}_\Sigma$, i.e., $\hat{\bar{SE}}_\Sigma$. In order to compute $\hat{\bar{SE}}_\Sigma$, one needs to set $\mathcal{Q} = \text{Im}(\check{B}_\text{a}) \cap \text{Ker}(\check{C})$ in steps 10 of Algorithm 1. Moreover, the output of the algorithm is $\hat{\bar{SE}}_\Sigma = \min\{\bar{SE}_\text{s}, m\}$.

V. Numerical Case Study

In this case study, we compute the SE by using (9) and (11) for a Quadruple-Tank Process (QTP) with a non-minimum phase zero. The characteristic matrices of the QTP are expressed as follows [15]:

$$A = \begin{bmatrix} -0.0158 & 0 & 0.0256 & 0 \\ 0 & -0.0109 & 0 & 0.0178 \\ 0 & 0 & -0.0256 & 0 \\ 0 & 0 & 0 & -0.0178 \end{bmatrix},$$
$$B = \begin{bmatrix} 0.0482 & 0 \\ 0 & 0.0349 \\ 0 & 0.0775 \\ 0.0559 & 0 \end{bmatrix}, C = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0. & 0.5 & 0 & 0 \end{bmatrix}. \tag{12}$$

The QTP in (12) is left-invertible, which as per Definition 7 it implies that no controllable attack can be performed on it. However, it is vulnerable to zero dynamics attacks and covert attacks. In the case where there exists no secure input and output communication channel, i.e., $B_\text{a} = B$ and $D_\text{a} = I_p$, the adversaries can execute both zero dynamics attacks and covert attacks as shown in Figs. 1 and 2, respectively.

If only one actuator is secured, the adversaries cannot execute zero dynamics attacks, but they are still capable of performing covert attacks. Consequently, securing the first actuator and the first sensor will result in $\mathscr{V}^* = 0$. Hence, the SE for the QTP is $SE_\Sigma = 2$.

In order to prevent perfectly undetectable cyber-attacks, i.e., covert attacks and controllable attacks, one needs to

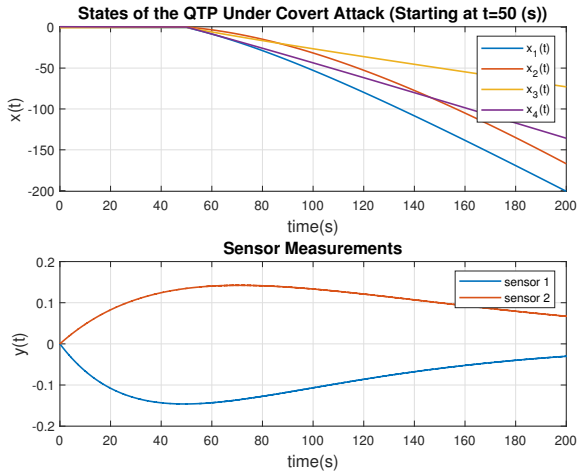

Fig. 2. The QTP under covert attacks where the states become unbounded while the outputs show a normal attack-free behavior.

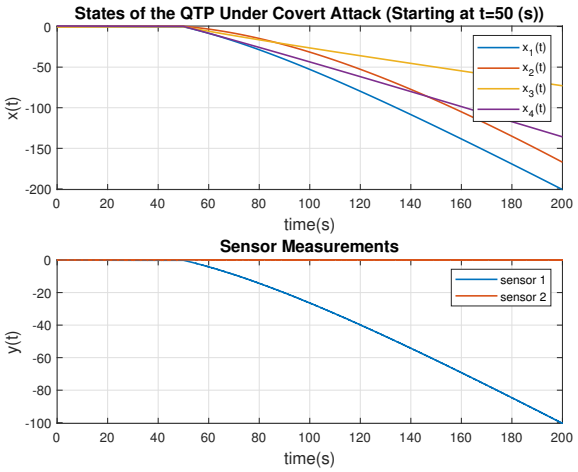

Fig. 3. Preventing adversaries from executing a covert attack in the QTP by securing the first input and the first output communication channel given that the first output remains unbounded and detectable.

compute $\hat{SE}_\Sigma$ given by (11). Consequently, $\hat{SE}_\Sigma = 2$ and securing the first actuator and the first sensor results in having $\mathrm{Im}(\check{B}_\mathrm{a}) \cap \mathscr{V}^* = 0$. Thus, having one secure input and one secure output communication channel prevents adversaries from executing perfectly undetectable cyber-attacks in the QTP. As seen from Fig. 3, once the first actuator and the first sensor are secured, the adversaries cannot perform covert attacks.

As it was mentioned earlier, adversaries need to compromise both input channels to perform zero dynamics attacks in the QTP. Moreover, as for the case of covert attacks, adversaries need to have access to at least 2 input and 1 output communication channels. Hence, if one considers both zero dynamics and perfectly undetectable cyber-attacks, the security index for the QTP is equal to 2. However, the system operators need to secure 1 input and 1 output communication channel to prevent both zero dynamics and perfectly undetectable cyber-attacks in the QTP, i.e., $SE_\Sigma = 2$. Moreover, the security index for only perfectly undetectable cyber-attacks is equal to 3 and the system operators can prevent them by securing 1 input and 1 output communication channel, i.e., $\hat{SE}_\Sigma = 2$. Hence, in this case study, we have $SE_\Sigma = \hat{SE}_\Sigma$ while the security index for undetectable cyber-attacks is 2 and that for perfectly undetectable cyber-attacks is 3.

VI. Conclusion

In this paper, the notion of security effort (SE) is developed and formally specified as a security measure for cyber-physical systems (CPS). The SE metric denotes the minimum number of input and output communication channels that should be secured to prevent adversaries from executing zero dynamics attacks, covert attacks, and controllable attacks. Moreover, it is shown that SE can be specified to prevent only perfectly undetectable cyber-attacks in the CPS, namely covert attacks and controllable attacks. Since zero dynamics attacks, covert attacks, and controllable attacks belong to the weakly unobservable and controllable weakly unobservable subspaces of the CPS, conditions for making these subspaces equal to zero are developed and investigated. Hence, the conditions that are developed to make weakly unobservable and controllable weakly unobservable subspaces equal to zero are utilized to compute SE. However, finding SE is an NP-hard problem. Hence, as for our future work, we are interested in investigating a method to determine the SE in a generic manner that is computationally efficient.

References

[1] H. Sandberg, V. Gupta, and K. H. Johansson, "Secure networked control systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 445–464, 2022.

[2] J. Miloevi, A. Teixeira, K. H. Johansson, and H. Sandberg, "Actuator security indices based on perfect undetectability: Computation, robustness, and sensor placement," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3816–3831, 2020.

[3] S. Gracy, J. Milošević, and H. Sandberg, "Security index based on perfectly undetectable attacks: Graph-theoretic conditions," *Automatica*, vol. 134, p. 109925, 2021.

[4] A. Baniamerian, K. Khorasani, and N. Meskin, "Determination of security index for linear cyber-physical systems subject to malicious cyber attacks," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 4507–4513.

[5] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

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

[7] Z. Zhao, Y. Yang, Y. Li, and R. Liu, "Security analysis for cyber-physical systems under undetectable attacks: A geometric approach," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 11, pp. 4359–4370, 2020.

[8] M. Taheri, K. Khorasani, I. Shames, and N. Meskin, "Data-driven covert-attack strategies and countermeasures for cyber-physical systems," in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 4170–4175.

[9] A. Baniamerian and K. Khorasani, "Security index of linear cyber-physical systems: A geometric perspective," in *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2019, pp. 391–396.

[10] H. Sandberg and A. M. Teixeira, "From control system security indices to attack identifiability," in *2016 Science of Security for Cyber-Physical Systems Workshop (SOSCYPS)*. IEEE, 2016, pp. 1–6.

[11] R. S. Smith, "A decoupled feedback structure for covertly appropriating networked control systems," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 90–95, 2011.

[12] H. L. Trentelman, A. A. Stoorvogel, and M. Hautus, *Control Theory for Linear Systems*. Springer Science & Business Media, 2012.

[13] G. Basile and G. Marro, *Controlled and Conditioned Invariants in Linear System Theory*. Prentice Hall Englewood Cliffs, 1992.

[14] D. E. Knuth, *The art of computer programming, volume 4A: combinatorial algorithms, part 1*. Pearson Education India, 2011.

[15] K. H. Johansson, "The quadruple-tank process: A multivariable laboratory process with an adjustable zero," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 3, pp. 456–465, 2000.