2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

A Dynamic Coding Scheme for Preventing Controllable Cyber-Attacks in Cyber-Physical Systems

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

Abstract— **Controllable attacks are considered as perfectly undetectable cyber-attacks that are performed by compromising input communication channels of cyber-physical systems (CPS). They are referred to as perfectly undetectable since they have zero impact on the sensor measurements of the system. In this paper, we investigate conditions under which adversaries are capable of performing controllable cyber-attacks and develop methods for designing these attack signals. Moreover, under certain assumptions, conditions for designing controllable attacks in terms of the Markov parameters of the CPS are derived. In order to analyze the vulnerability of the CPS to controllable attacks from the system operators' point of view, a security metric designated as the security effort (SE) for controllable attacks is formally defined and proposed. The SE for controllable attacks denotes the minimum number of input communication channels that need to be secured to prevent adversaries from executing this type of cyber-attack. Consequently, as a countermeasure, we develop a coding scheme on the input communication channels that increases the minimum number of required input communication channels for performing controllable attacks to its maximum possible value. Consequently, in presence of the proposed coding scheme, adversaries need to compromise all the input communication channels to execute controllable attacks. Therefore, securing only one input channel prevents adversaries from performing controllable cyber-attacks. Finally, an illustrative numerical case study is provided to demonstrate the effectiveness and capabilities of our derived conditions and proposed methodologies.**

I. Introduction

Cyber-physical systems (CPS) have positively affected our today's life through their wide range of applications, such as in transportation systems, water treatment networks, power systems and smart grids, unmanned aerial vehicles (UAV), and industrial process control systems [1]–[6]. On the other hand, CPS are prone to cyber threats as several cases of malicious cyber-attacks have been carried out and reported over the past decade [1], [7]. Therefore, it is of paramount importance to study and investigate the vulnerability of CPS to cyber-attacks and develop countermeasures for these threats.

Adversaries are capable of compromising the integrity of the communicated information in the CPS. These types of cyber-attacks are referred to as deception attacks [3], [8], [9]. In order to perform a deception attack, adversaries need to have access to certain levels of disruption resources, system knowledge, and disclosure resources of the CPS [3]. Given the availability of different disruption resources, disclosure resources, and system knowledge, adversaries can maintain their attacks undetected and perform stealthy cyber-attacks.

In particular, adversaries are capable of performing stealthy deception attacks on actuators as in the controllable attacks [10]–[12] that their impacts cannot be detected in sensor measurements. Controllable attacks can be performed on the CPS that are not left invertible [11]. Furthermore, controllable attacks are in the controllability subspace within the weakly unobservable subspace of the system [11].

The minimum number of required actuators and sensors, i.e., the disruption resources, to execute an undetectable or a perfectly undetectable cyber-attacks such as the zero dynamics attacks, controllable attacks, and covert attacks has been defined as the security index for the CPS [12]–[15]. In the case of undetectable cyber-attacks, if the initial conditions of the CPS are known, the impact of the cyber-attack can still be detected in the sensor measurements [2], [13], but perfectly undetectable attacks leave no impact on the outputs of the CPS [15], [16].

The security index (SI) for perfectly undetectable cyber-attacks in the CPS is defined as the minimum number of actuators and sensors that should be compromised by adversaries to execute a perfectly undetectable cyber-attack [11], [12], [14], [15]. In [12], a control geometric approach is adopted to define subspaces that are related to perfectly undetectable cyber-attacks and find an upper bound for the SI. Moreover, in [14], [15], the structural system framework is employed to find a generic value for the SI.

The SI analyzes the CPS from the adversary's point of view which may not support system operators to determine the requirements for preventing the occurrence of certain cyber-attacks in the CPS. Hence, in [17], [18], geometric structural conditions are derived to find the minimum number of communication channels that should be secured to prevent certain stealthy cyber-attacks. In [17], under various disclosure scenarios, minimal input communication channels to protect for preventing stealthy cyber-attacks is determined. Moreover, in [18], the minimum number of actuators and sensors that should be secured to prevent executing zero dynamics attacks, covert attacks, and controllable attacks is defined as the security effort (SE) for CPS. However, the derived geometric conditions in [17], [18] for determining the communication channels that should be secured are not

[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.

simple methods to check and verify (e.g., in the case of large-scale CPS). Hence, defining SE for controllable attacks and deriving easy to check algebraic conditions for obtaining the SE is one of the objectives of this paper.

Various methodologies have been proposed to either detect stealthy cyber-attacks or prevent adversaries from executing them [5], [19]–[27]. In [20], [21], monitoring systems that utilize auxiliary filters have been developed to detect stealthy cyber-attacks such as the zero dynamics attacks. Furthermore, coding schemes [24], modulation matrices [22], moving target approaches [28], and watermarking schemes [5], [23] have been developed and employed that distort the system knowledge from adversaries point of view and prevent them from executing stealthy cyber-attacks.

In this paper, the input/output (I/O) model of the CPS is utilized to study conditions under which adversaries can perform controllable attacks. Specifically, Markov parameters, elements of the observability matrix, and the characteristic matrices of the CPS are used to investigate conditions for existence of the controllable attacks that can provide one with the required disruption resources, i.e., actuators to be attacked, and system knowledge for performing the above cyber-attacks. Moreover, the implementation of the controllable attacks by means of Markov parameters of the CPS and elements of the observability matrix are studied and developed.

As a countermeasure against the controllable attacks, a dynamic coding scheme is developed and employed. The proposed coding scheme targets and increases the required disruption resources for executing controllable attacks. Hence, in presence of the coding scheme, adversaries need to compromise all the input communication channels and actuators of the CPS to carry out the controllable attacks. Therefore, having only one secure actuator will prevent the adversaries from executing the above stealthy cyber-attacks.

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

1) Under certain assumptions, the conditions under which one can carry out the controllable attacks are obtained. These conditions are derived in terms of the Markov parameters of the CPS, elements of the observability matrix, and characteristic matrices of the system.
2) By utilizing the proposed conditions for existence of controllable attacks, and details on the implementation methodologies of these cyber-attacks are then provided.
3) A dynamic coding scheme is then developed and proposed that under certain conditions can increase the number of actuators that are needed to execute the controllable cyber-attacks to its maximum possible value.

In contrast to [5], [20]–[27] which either rely on distorting the adversary's system knowledge or developing auxiliary filters to detect stealthy cyber-attacks, in this work, we propose a dynamic coding scheme that increases the actuators security index. Therefore, having only one secure input channel will prevent adversaries from executing controllable cyber-attacks. In particular, as opposed to [27] which assumes that only one actuator under cyber-attacks at an instant of time, and studies covert cyber-attacks, and finally develops a static coding scheme, in this work we assume that all actuators except for one are under cyber-attacks, and focus on controllable attacks, and develop dynamic coding scheme as our countermeasure against this type of cyber-attack.

II. Problem Statement and Formulation

We consider the following linear time-invariant CPS:

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

where $x(k) \in \mathbb{R}^n$, $u(k) \in \mathbb{R}^m$, $y(k) \in \mathbb{R}^p$, $\omega(k) \in \mathbb{R}^n$, and $\nu(k) \in \mathbb{R}^p$, denote the state, the control input, the sensor measurement, process noise, and measurement noise respectively. The characteristic matrices (A, B, C) are of appropriate dimensions.

In presence of cyber-attacks, the CPS (1) can be expressed by

$$\begin{aligned} x(k+1) =& Ax(k) + B(u(k) + L_{\mathrm{a}} a_{\mathrm{u}}(k)) + \omega(k), \\ y(k) =& Cx(k) + \nu(k), \end{aligned} \tag{2}$$

where $a_{\mathrm{u}}(k) \in \mathbb{R}^{m_{\mathrm{a}}}$ is the actuator attack signal. We assume L_{a} which denotes the input communication channels that are attacked by adversaries is an injective map, i.e., one-to-one, since otherwise, the linearly dependent columns can be removed. Moreover, the actuator attack signature is captured by $B_{\mathrm{a}} = BL_{\mathrm{a}}$. In the our theoretical analysis we assume that $\omega(k) = 0$ and $\nu(k) = 0$, $\forall k \geq 0$ and do not consider the impact of noise. However, in our numerical case study in Section VI we consider both process and sensor noise to illustrate robustness of our proposed methodologies to noise.

Let $y_{\mathrm{o}}(x(0), u(k), a_{\mathrm{u}}(k))$ denote the output of (2) as a function of the initial state $x(0)$, the control input $u(k)$, and the actuator attack signal $a_{\mathrm{u}}(k)$. In the following, by utilizing the notion of $y_{\mathrm{o}}(x(0), u(k), a_{\mathrm{u}}(k))$, we define the "controllable attacks" (studied in [10]–[12], [26]).

Definition 1: The cyber-attack signal $a_{\mathrm{u}}(k) \neq 0$ is designated as a controllable cyber-attack if $y_{\mathrm{o}}(0, 0, a_{\mathrm{u}}(k)) = 0$, for every $k \geq 0$.

It should be emphasized that the cyber-attack in Definition 1 is referred to as the "zero stealthy attack" in [10], "zero state induced attack" in [26], and "controllable attack" in [11] and [12]. However, we have adopted the convention from [11] and [12] since the above cyber-attack is related to a certain controllable subspace of the system (see [12] for more details).

Definition 2 ([29], [30]): The CPS (2) is left invertible with respect to the cyber-attack signal $a_{\mathrm{u}}(k)$ if for all $a_{\mathrm{u}}^1(k), a_{\mathrm{u}}^2(k) \in \mathbb{R}^{m_{\mathrm{a}}}$, having $y_{\mathrm{o}}(0, 0, a_{\mathrm{u}}^1(k)) = y_{\mathrm{o}}(0, 0, a_{\mathrm{u}}^2(k))$ implies $a_{\mathrm{u}}^1(k) = a_{\mathrm{u}}^2(k)$, for every $k \geq 0$.

Remark 1: By considering the linearity of the CPS, it follows from Definition 2 that the CPS is left invertible if and only if $y_{\mathrm{o}}(0, 0, a_{\mathrm{u}}(k)) = 0$ implies that $a_{\mathrm{u}}(k) = 0$. Hence, controllable cyber-attacks in Definition 1 can be executed if and only if the CPS (2) is not left invertible in the sense of Definition 2 (see [10]–[12] for more details).

Below, we have adopted and modified the definitions in [2], [9], [11], [12], [14], [19], [23] in order to provide a formal and unified definition for the *perfectly undetectable* cyber-attacks on actuators of the CPS.

Definition 3: Let $x(0) = x_0 \in \mathbb{R}^n$. An actuator attack that is performed by utilizing $a_{\mathrm{u}}(k) \neq 0$ on the CPS (2)

is designated as *perfectly undetectable* if $y_{\rm o}(x_0, u(k), 0) = y_{\rm o}(x_0, u(k), a_{\rm u}(k))$, $\forall\, k \in \mathbb{N}$.

It should be noted that in Definition 3, we have adopted the notion of "perfectly undetectable attacks" from [14]. The main reason for choosing this designation is that the impact of perfectly undetectable attacks cannot be seen in the output measurements.

Due to linearity of the CPS (2) and according to Definition 3, a cyber-attack is perfectly undetectable if and only if $y_{\rm o}(0, 0, a_{\rm u}(k)) = 0$, $\forall\, k \geq 0$ [14]. Consequently, controllable attacks are perfectly undetectable cyber-attacks.

A. Objectives

Our objectives in this paper are threefold. Our first objective is to study and develop conditions under which adversaries can execute controllable cyber-attacks in the CPS. These conditions are derived in terms of Markov parameters of the CPS and elements of the observability matrix to show the required system knowledge for performing controllable attacks. As for our second objective, we formally define and introduce the notion of security effort (SE) for controllable attacks as a security measure that denotes the minimum number of input communication channels to be secured for preventing adversaries from performing controllable attacks. Finally, our last objective is to develop and study a coding scheme on the input communication channels of the CPS that increases the minimum number of input communication channels that should be compromised to carry out controllable attacks. Consequently, in presence of this coding scheme, adversaries need to attack all input communication channels for executing controllable attacks. Therefore, by utilizing our proposed coding scheme, the CPS operators need to secure only one communication channel to prevent the occurrence of controllable cyber-attacks.

III. Input/Output Model of the CPS and Controllable Cyber-Attacks

In this section, we expand the state-space representation of the CPS in (2) and derive the input/output (I/O) model of the CPS to derive easy to check algebraic conditions for the existence of controllable cyber-attacks.

The I/O model of the CPS (2) over the time window $\{0, 1, \ldots, N-1\}$ for $N \geq n$ can be expressed by

$$Y(N) = \mathcal{O}_{\rm N} x(0) + \mathcal{C}_{\rm N} U(N) + \mathcal{C}_{\rm a} U_{\rm a}(N), \tag{3}$$

where $Y(N) = [y(0)^\top, y(1)^\top, \ldots, y(N-1)^\top]^\top$, $U(N) = [u(0)^\top, u(1)^\top, \ldots, u(N-1)^\top]^\top$, and $U_{\rm a}(N) = [a_{\rm u}(0)^\top, a_{\rm u}(1)^\top, \ldots, a_{\rm u}(N-1)^\top]^\top$ denote the output of the CPS, the vector of inputs, and the vector of actuator cyber-attacks, respectively. Moreover,

$$\mathcal{O}_{\rm N} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{bmatrix}, \mathcal{C}_{\rm N} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ CB & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}B & CA^{N-3}B & \cdots & 0 \end{bmatrix},$$

$$\mathcal{C}_{\rm a} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ CB_{\rm a} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}B_{\rm a} & CA^{N-3}B_{\rm a} & \cdots & 0 \end{bmatrix}. \tag{4}$$

In the following subsection, we show how the knowledge on $\mathcal{O}_{\rm N}$ and $\mathcal{C}_{\rm N}$ is sufficient for adversaries to execute controllable cyber-attacks.

A. Controllable Cyber-Attacks

As stated in Definition 1, under controllable cyber-attacks, one has $y_{\rm o}(0, 0, a_{\rm u}(k)) = 0$, where $a_{\rm u}(k) \neq 0$. Moreover, according to Definition 3, controllable attacks are considered as perfectly undetectable. Let $\mathcal{Y}(x(0), U(N), U_{\rm a}(N))$ designate the output of the I/O model in (3) over the time window $\{0, 1, \ldots, N-1\}$ that is a function of the initial state $x(0)$, the vector of control inputs $U(N)$, and the vector of actuator cyber-attack signals $U_{\rm a}(N)$. In the following, a definition for controllable cyber-attacks in the sense of I/O model (3) is provided.

Definition 4 (Controllable Cyber-Attack in the I/O Model): Let $a_{\rm u}(k) \neq 0$, $\forall k \geq 0$. The attack signal $U_{\rm a}(N)$ in the I/O model of the CPS (3) is designated as a controllable cyber-attack if one has $\mathcal{Y}(0, 0, U_{\rm a}(N)) = 0$, $\forall N \geq 1$, i.e., $\mathcal{C}_{\rm a} U_{\rm a}(N) = 0$.

In the following theorem, we investigate conditions under which controllable cyber-attacks in Definition 4 can be performed on the I/O model of the CPS (3).

Definition 5: Let $\mathcal{I} = \{1, \ldots, m\}$ denote the set of all input channels of the CPS (2). The relative degree of the CPS (2) with respect to the q-th input channel is given by r_q if $CA^i B_q = 0$, for all $i < r_q - 1$ and $CA^{r_q - 1} B_q \neq 0$, for every $q \in \mathcal{I}$, where B_q is the q-th column of B. If for any positive integer i one has $CA^i B_q = 0$, the relative degree with respect the q-th input channel cannot be defined. Moreover, let $\mathcal{I}_{\rm a} = \{a_1, \ldots, a_{m_{\rm a}}\}$ denote the set of attacked input communication channels. Hence, one has $r_{\rm a} = \min\{r_{a_1}, \ldots, r_{a_{m_{\rm a}}}\}$.

Theorem 1: A controllable cyber-attack in the sense of Definition 4 can be executed in the CPS if there exists a nonzero $\hat{a}_0 \in \ker(CA^{r_{\rm a}-1}B_{\rm a})$, such that $\mathrm{Im}(AB_{\rm a}\hat{a}_0) \subseteq \mathrm{Im}(B_{\rm a})$.

Proof: According to Definition 4, in the case of controllable cyber-attacks, the actuator cyber-attack signal $a_{\rm u}(k)$ should be designed such that $\mathcal{C}_{\rm a} U_{\rm a}(N) = 0$, which is equivalent to

$$CA^{r_{\rm a}-1}B_{\rm a}a_{\rm u}(0) = 0, \tag{5a}$$

$$CA^{r_{\rm a}}B_{\rm a}a_{\rm u}(0) + CA^{r_{\rm a}-1}B_{\rm a}a_{\rm u}(1) = 0, \tag{5b}$$

$$CA^{r_{\rm a}+1}B_{\rm a}a_{\rm u}(0) + CA^{r_{\rm a}}B_{\rm a}a_{\rm u}(1) + CA^{r_{\rm a}-1}B_{\rm a}a_{\rm u}(2) = 0, \tag{5c}$$

$$\vdots$$

$$CA^{N-2}B_{\rm a}a_{\rm u}(0) + CA^{N-3}B_{\rm a}a_{\rm u}(1) + CA^{N-4}B_{\rm a}a_{\rm u}(2) + \cdots + CA^{r_{\rm a}-1}B_{\rm a}a_{\rm u}(N-2) = 0. \tag{5d}$$

Suppose that $a_{\rm u}(0) \in \ker(CA^{r_{\rm a}-1}B_{\rm a})$, and there exists $\hat{a}_0 \in \ker(CA^{r_{\rm a}-1}B_{\rm a})$ such that $\mathrm{Im}(AB_{\rm a}\hat{a}_0) \subseteq \mathrm{Im}(B_{\rm a})$. Let $a_{\rm u}(0) = \hat{a}_0$. Let us rewrite the left-hand side of (5b) as $CA^{r_{\rm a}-1}(AB_{\rm a}a_{\rm u}(0) + B_{\rm a}a_{\rm u}(1))$. Consequently, since $\mathrm{Im}(AB_{\rm a}\hat{a}_0) \subseteq \mathrm{Im}(B_{\rm a})$, one can design $a_{\rm u}(1)$ such that

$$AB_{\rm a}\hat{a}_0 + B_{\rm a}a_{\rm u}(1) = B_{\rm a}\hat{a}_0. \tag{6}$$

Hence, by substituting (6) in (5b), one can conclude that there exists the cyber-attack signal $a_{\rm u}(1)$ that satisfies (5b).

The left-hand side of (5c) can be rewritten as $CA^{r_a-1}(A^2B_a\hat{a}_0 + AB_a a_u(1) + B_a a_u(2))$. Considering that $\text{Im}(AB_a\hat{a}_0) \subseteq \text{Im}(B_a)$ and $AB_a\hat{a}_0 + B_a a_u(1) = B_a\hat{a}_0$, $a_u(2)$ can be designed in the following form

$$A(AB_a\hat{a}_0 + B_a a_u(1)) + B_a a_u(2) = B_a\hat{a}_0, \tag{7}$$

which satisfies (5c).

Consequently, it can be shown that since $\text{Im}(AB_a\hat{a}_0) \subseteq \text{Im}(B_a)$, there exists $a_u(j)$ in the j-th equation of (5) that satisfies

$$AB_a\hat{a}_0 + B_a a_u(j) = B_a\hat{a}_0, \tag{8}$$

for $j \geq 2$. Therefore, there exists an $a_u(j)$ that is the solution to the j-th equation in (5). This completes the proof of the theorem. ■

In [26], cyber-attacks that satisfy the condition in Definition 4, i.e., $\mathcal{C}_a U_a(N) = 0$, are defined as "zero state inducing" attacks. Moreover, necessary and sufficient conditions for the existence of this type of cyber-attack based on weakly unobservable and output-nulling reachable subspaces of the system have been provided in [26, Theorem 3]. However, as opposed to [26], the studied conditions in Theorem 1 rely only on A, B_a, and the first Markov parameter of the CPS and are easier to verify and validate. In the following corollary, the implementation of controllable cyber-attacks is investigated.

***Corollary** 1:* Assume that the conditions in Theorem 1 hold and let $a_u(0) \in \ker(CA^{r_a-1}B_a)$. The actuator cyber-attack signal to perform a controllable attack in the sense of Definition 4 can be expressed as

$$a_u(k) = a_u(0)h(k) - B_a^\dagger AB_a a_u(0)h(k-1), \tag{9}$$

for $k \geq 1$, where $h(k) \in \mathbb{R}$ such that $h(0) = 1$ and $h(k)$ for $k \geq 1$ can be any arbitrary function.

Proof: Let $h(k) \in \mathbb{R}$ such that $h(0) = 1$. Moreover, as per Theorem 1, consider $a_u(0) \in \ker(CA^{r_a-1}B_a)$ such that $\text{Im}(AB_a a_u(0)) \subseteq \text{Im}(B_a)$. Consequently, $a_u(1)$ can be designed such that

$$AB_a a_u(0)h(0) + B_a a_u(1) = B_a a_u(0)h(1). \tag{10}$$

Given that the left-hand side of (5b) can be rewritten as $CA^{r_a-1}(AB_a a_u(0)h(0)+B_a a_u(1))$, (10) satisfies (5b). Moreover, since B_a is an injective map, $a_u(1)$ can be uniquely derived as $a_u(1) = a_u(0)h(1) - B_a^\dagger AB_a a_u(0)h(0)$.

Similar to $a_u(1)$, one can design $a_u(2)$ to satisfy

$$AB_a a_u(0)h(1) + B_a a_u(2) = B_a a_u(0)h(2). \tag{11}$$

Also, considering (10), the left-hand side of (5c) can be rewritten as $CA^{r_a-1}(AB_a a_u(0)h(1) + B_a a_u(2))$. Thus, the given $a_u(2)$ in (11) satisfies (5c). Hence, one has $a_u(2) = a_u(0)h(2) - B_a^\dagger AB_a a_u(0)h(1)$.

Consequently, $a_u(k)$ can be designed to satisfy

$$AB_a a_u(0)h(k-1) + B_a a_u(k) = B_a a_u(0)h(k), \tag{12}$$

for $k \geq 1$. Moreover, considering (12), the left-hand side of the $(k+1)$-th equation in (5) can be derived as $CA^{r_a-1}(AB_a a_u(0)h(k-1) + B_a a_u(k))$. Given that $a_u(k)$ is designed according to (12), $a_u(k)$ satisfies the $(k+1)$-th equation in (5), for $k \geq 1$. Therefore, from (12) it follows that the controllable cyber-attack signal $a_u(k)$ can be designed according to (9). This completes the proof of the corollary. ■

Theorem 1 can be used to study existence of controllable cyber-attacks in the CPS. Furthermore, one needs to know the first Markov parameter of the CPS, i.e., $CA^{r_a-1}B_a$, and matrices A and B_a to investigate the proposed conditions in Theorem 1. Hence, in the following corollary, under certain conditions, the existence and implementation of controllable cyber-attacks by utilizing only the Markov parameters of the CPS are developed.

***Corollary** 2:* Let us assume that $\ker(CA^{r_a-1}) \bigcap \ker(CA^{r_a}) = 0$. Adversaries can execute a controllable cyber-attack in the CPS according to the Definition 4 if there exist nonzero $\hat{a}_0 \in \ker(CA^{r_a-1}B_a)$ and $\hat{a}_1 \in \mathbb{R}^{m_a}$ that satisfy $CA^{r_a}B_a\hat{a}_0 + CA^{r_a-1}B_a\hat{a}_1 = 0$ and $CA^{r_a+1}B_a\hat{a}_0 + CA^{r_a}B_a\hat{a}_1 = 0$. Moreover, by considering $a_u(0) = \hat{a}_0$, a controllable cyber-attack signal can be expressed as

$$a_u(k) = \hat{a}_0 h(k) + \hat{a}_1 h(k-1), \tag{13}$$

for $k \geq 1$, where $h(k) \in \mathbb{R}$ such that $h(0) = 1$ and $h(k)$ for $k \geq 1$ can be any arbitrary function.

Proof: From $CA^{r_a}B_a\hat{a}_0 + CA^{r_a-1}B_a\hat{a}_1 = 0$, it follows that

$$CA^{r_a-1}(AB_a\hat{a}_0 + B_a\hat{a}_1) = 0. \tag{14}$$

Moreover, having $CA^{r_a+1}B_a\hat{a}_0 + CA^{r_a}B_a\hat{a}_1 = 0$ implies that

$$CA^{r_a}(AB_a\hat{a}_0 + B_a\hat{a}_1) = 0. \tag{15}$$

Consequently, since $\ker(CA^{r_a-1}) \bigcap \ker(CA^{r_a}) = 0$, it follows from (14) and (15) that $AB_a\hat{a}_0 + B_a\hat{a}_1 = 0$, which implies that $\text{Im}(AB_a\hat{a}_0) \subseteq \text{Im}(B_a)$. Hence, conditions in the Theorem 1 for existence of controllable cyber-attacks are satisfied.

From $AB_a\hat{a}_0 + B_a\hat{a}_1 = 0$, it follows that $\hat{a}_1 = -B_a^\dagger AB_a\hat{a}_0$. Hence, as per Corollary 1, one can set $a_u(0) = \hat{a}_0$ and design a controllable cyber attack in the following form:

$$a_u(k) = a_u(0)h(k) - B_a^\dagger AB_a a_u(0)h(k-1),$$

for $k \geq 1$, where $h(0) = 1$ and $h(k) \in \mathbb{R}$ can be any arbitrary function. This completes the proof of the corollary. ■

***Remark** 2:* In order to find $\hat{a}_0$ and $\hat{a}_1$ in the Corollary 2, one needs to solve

$$\begin{bmatrix} CA^{r_a-1}B_a & 0 \\ CA^{r_a}B_a & CA^{r_a-1}B_a \\ CA^{r_a+1}B_a & CA^{r_a}B_a \end{bmatrix} \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{16}$$

for $\hat{a}_0$ and $\hat{a}_1$. It should be noted that if the hypothesis of Corollary 2 holds, (16) can be easily solved by using the `mldivide` MATLAB function.

IV. Security Effort (SE) for Controllable Attacks

In this section, we study security effort (SE) for controllable attacks. The SE denotes the minimum number of input communication channels that should be secured and kept attack-free to prevent adversaries from executing controllable cyber-attacks. It should be noted that the main difference between the SE and the security index (SI) is that in SI, one determines the minimum number of communication channels that are needed to execute undetectable cyber-attacks.

By utilizing Definition 4, the SE for controllable attacks can be defined according to the following problem:

$$SE_{\text{cont}} := \min_{L_{\text{a}}} \quad m - \text{rank}(L_{\text{a}}) \\ \text{s.t.} \quad \mathcal{C}_{\text{a}} U_{\text{a}}(N) \neq 0, \\ \text{for any } U_{\text{a}}(N) \neq 0. \tag{17}$$

In (17), $m - \text{rank}(L_{\text{a}})$ is equal to the number of actuators that are not under cyber-attacks. Hence, in the definition of SE, we are minimizing the number of attack-free input communication channels that prevent adversaries from finding a nonzero $U_{\text{a}}(N)$ that satisfies $\mathcal{C}_{\text{a}} U_{\text{a}}(N) = 0$, i.e., an actuator attack signal $U_{\text{a}}(N)$ that results in a controllable attack in the sense of Definition 4.

Adversaries can design a nonzero $U_{\text{a}}(N)$ that can satisfy $\mathcal{C}_{\text{a}} U_{\text{a}}(N) = 0$, when $\text{rank}(\mathcal{C}_{\text{a}}) < (N-1)m_{\text{a}}$. Hence, the conditions in (17) are satisfied when $\text{rank}(\mathcal{C}_{\text{a}}) \geq (N-1)m_{\text{a}}$. The latter is utilized in the Algorithm 1 for computing the SE for controllable attacks.

Algorithm 1 Pseudo code to find SE_{cont}

Input: (A, B, C), and the set of all inputs $S = \{u_1, \ldots, u_{\text{m}}\}$
Output: SE_{cont}, and SE_{min} which is the set of actuators that should be secured

Initialize $SE_{\text{cont}} = m$, $SE_{\text{min}} = S$, $L_{\text{a}} = I_m$
if $\text{rank}(\mathcal{C}_{\text{N}}) < (N-1)m$ **then**
  Set $l = |S|$, where $|\cdot|$ denotes the cardinality of a set
  for $i = 1 : 2^l - 1$ **do**
    Create the empty set $\hat{S} = \{\}$
    for $j = 1 : l$ **do**
      if the j-th bit of the binary representation of i is equal to 1 **then**
        Add j-th member of S to $\hat{S}$
      end if
    end for
    Compromise actuators that belong to the set $\hat{S}$, and update L_{a} and r_{a} accordingly
    if $\text{rank}(\mathcal{C}_{\text{a}}) \geq (N-1)m_{\text{a}}$ **and** $m - |\hat{S}| \leq SE_{\text{cont}}$ **then**
      $SE_{\text{cont}} = m - |\hat{S}|$
      $SE_{\text{min}} = S - \hat{S}$
    end if
  end for
else
  $SE_{\text{cont}} = 0$
  $SE_{\text{min}} = \{\}$
end if

V. Dynamic Coding Scheme

In this section, a dynamic coding scheme on the input communication channels is developed that, under certain conditions, can be used to prevent adversaries from performing stealthy cyber-attacks such as the controllable attacks on the actuators. The coding scheme is designed such that having only one secure input communication channel will result in preventing adversaries from executing their controllable cyber-attacks.

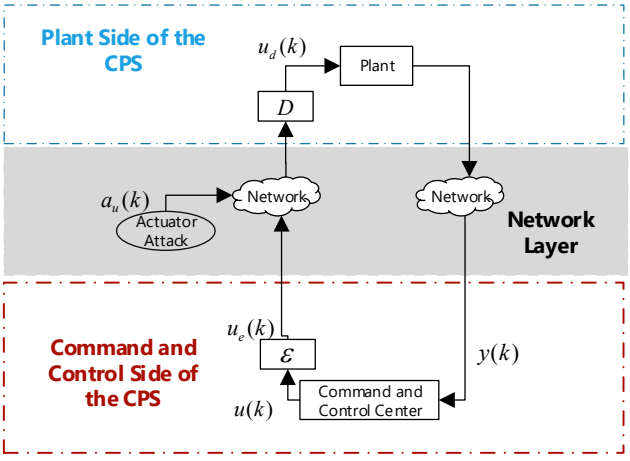

Fig. 1. The architecture of the CPS and the dynamic coding scheme, where $u_{\text{e}}(k)$ is the output of the encoder and $u_{\text{d}}(k)$ is the output of the decoder.

A. CPS Model in Presence of the Dynamic Coding Scheme

An encoder, that is denoted by $\mathcal{E}$, on the command and control (C&C) side and a decoder, denoted by $\mathcal{D}$, on the plant side of the CPS are designed. The CPS along with the encoder $\mathcal{E}$ and the decoder $\mathcal{D}$ are depicted in Fig.1.

The dynamics of the encoder and the decoder on the input communication channels of the CPS are governed by

$$\mathcal{E}: \begin{cases} x_{\text{e}}(k+1) &= A_{\text{e}} x_{\text{e}}(k) + B_{\text{e}} u(k), \\ u_{\text{e}}(k) &= C_{\text{e}} x_{\text{e}}(k) + D_{\text{e}} u(k), \end{cases} \tag{18}$$

$$\mathcal{D}: \begin{cases} x_{\text{d}}(k+1) &= A_{\text{d}} x_{\text{d}}(k) + B_{\text{d}}(u_{\text{e}}(k) + L_{\text{a}} a_{\text{u}}(k)), \\ u_{\text{d}}(k) &= C_{\text{d}} x_{\text{d}}(k) + D_{\text{d}}(u_{\text{e}}(k) + L_{\text{a}} a_{\text{u}}(k)), \end{cases} \tag{19}$$

where $x_{\text{e}}(k), x_{\text{d}}(k) \in \mathbb{R}^{n_e}$ and $u_{\text{e}}(k), u_{\text{d}}(k) \in \mathbb{R}^m$ denote the states and outputs of the encoder $\mathcal{E}$ and the decoder $\mathcal{D}$, respectively. Moreover, one has $x_{\text{e}}(0) = x_{\text{d}}(0) = 0$. The following lemma provides necessary and sufficient conditions under which the decoder $\mathcal{D}$ is the inverse of $\mathcal{E}$ such that once $a_{\text{u}}(k) = 0$, one has $u_{\text{d}}(k) = u(k)$, $\forall k \geq 0$.

***Lemma** 1 ([23]):* Let $a_{\text{u}}(k) = 0$. One has $u_{\text{d}}(k) = u(k)$, $\forall k \geq 0$, if and only if there exists an invertible matrix T that satisfies the following:

$$D_{\text{d}} C_{\text{e}} + C_{\text{d}} T = 0,\ T^{-1} B_{\text{d}} D_{\text{e}} = B_{\text{e}},\ D_{\text{d}} = D_{\text{e}}^{-1},$$
$$T^{-1} A_{\text{d}} T + T^{-1} B_{\text{d}} C_{\text{e}} = T^{-1} A_{\text{d}} T - B_{\text{e}} C_{\text{d}} T = A_{\text{e}}.$$

In presence of $\mathcal{E}$ and $\mathcal{D}$, the dynamics of the CPS (2) can be expressed as

$$x(k+1) = Ax(k) + Bu_{\text{d}}(k) + \omega(k), \\ y(k) = Cx(k) + \nu(k). \tag{20}$$

Consequently, the I/O model of the CPS (20) under noise free conditions is derived in the following form:

$$Y(N) = \mathcal{O}_{\text{N}} x(0) + \mathcal{C}_{\text{N}}(U(N) + \mathcal{C}_{\text{d}} U_{\text{a}}(N)), \tag{21}$$

where $\mathcal{C}_\mathrm{d} = \Gamma_\mathrm{d} \otimes L_\mathrm{a}$ and

$$\Gamma_\mathrm{d} = \begin{bmatrix} D_\mathrm{d} & 0 & \cdots & 0 \\ C_\mathrm{d}B_\mathrm{d} & D_\mathrm{d} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_\mathrm{d}A_\mathrm{d}^{N-2}B_\mathrm{d} & C_\mathrm{d}A_\mathrm{d}^{N-3}B_\mathrm{d} & \cdots & D_\mathrm{d} \end{bmatrix}. \quad (22)$$

Assumption 1: The encoder (18) and the decoder (19) are designed according to Lemma 1. Moreover, adversaries have knowledge on the parameters of Γ_d in (22).

It should be noted that in Assumption 1, we consider the worst case scenario for the CPS operators where adversaries have access to the parameters of the decoder through Γ_d. As can be seen from (21), due to existence of the coding scheme, the impact of actuator cyber-attack signals shows up as $\mathcal{C}_\mathrm{N}\mathcal{C}_\mathrm{d}U_\mathrm{a}(N)$ in the sensor measurements. It should be noted that since Γ_d is by definition an invertible matrix, the dimension of $\ker(\mathcal{C}_\mathrm{N})$ is equal to that of $\ker(\mathcal{C}_\mathrm{N}\Gamma_\mathrm{d})$. According to Definition 4, $\ker(\mathcal{C}_\mathrm{N})$ determines the existence of controllable cyber-attacks. Hence, having Γ_d in (21) does not result in introducing additional controllable cyber-attacks that can be executed in the CPS.

Definition 6 (Cyber-Attacks and the Coding Scheme): In the I/O model of the CPS (21), let $U_\mathrm{a}(N) = \tilde{\mathcal{C}}_\mathrm{d}\tilde{U}_\mathrm{a}(N)$, where $\tilde{\mathcal{C}}_\mathrm{d}$ is designed such that $\mathcal{C}_\mathrm{N}\mathcal{C}_\mathrm{d}\tilde{\mathcal{C}}_\mathrm{d} = \mathcal{C}_\mathrm{a}$. Consequently, the actuator cyber-attack is a controllable attack if $\tilde{U}_\mathrm{a}(N)$ is designed according to Definition 4 such that $\mathcal{C}_\mathrm{N}\mathcal{C}_\mathrm{d}\tilde{\mathcal{C}}_\mathrm{d}\tilde{U}_\mathrm{a}(N) = 0$. Moreover, if one cannot execute the controllable cyber-attacks in the CPS (20), the CPS is considered to be secure against this type of cyber-attack.

Definition 7: Let $\mathcal{I}_\mathrm{a} = \{a_1, \ldots, a_{m_\mathrm{a}}\}$ denote the set of compromised input communication channels. The relative degree of the CPS (20) with respect to the q-th attacked input channel is r_q^d if $CA^iB(D_\mathrm{d}L_\mathrm{a})_q = 0$ for all $i < r_q^\mathrm{d} - 1$ and $CA^{r_q^\mathrm{d}-1}B(D_\mathrm{d}L_\mathrm{a})_q \neq 0$, for every $q \in \mathcal{I}_\mathrm{a}$, where $(D_\mathrm{d}L_\mathrm{a})_q$ denotes the q-th column of $D_\mathrm{d}L_\mathrm{a}$. Moreover, $r_\mathrm{a}^\mathrm{d} = \min\{r_{a_1}^\mathrm{d}, \ldots, r_{a_m}^\mathrm{d}\}$.

As stated in Definition 6, the adversary's objective is to cancel out the impact of the dynamic coding scheme by designing the actuator cyber-attack signals and maintaining the cyber-attack perfectly undetectable. Hence, the design conditions for $\mathcal{E}$ and $\mathcal{D}$ under which adversaries cannot evade the coding scheme to maintain their cyber-attacks undetected are discussed in the next subsection.

The following assumption holds throughout this section.

Assumption 2: In the CPS (20), there exists at least one secure input communication channel, i.e., $\mathrm{rank}(L_\mathrm{a}) < m$.

In order to secure certain communication channels as per Assumption 2, system operators can utilize encryption/decryption and authentication methods on a given channel as have been investigated in [31], [32].

B. Designing the Dynamic Coding Scheme for Securing the CPS Against Controllable Cyber-Attacks

As per Definition 6, in order to execute the controllable cyber-attacks, adversaries need to first eliminate the impact of the coding scheme by designing $\tilde{\mathcal{C}}_\mathrm{d}$. Hence, our objective is to design and develop the coding scheme such that having only one secured input channel will prevent adversaries from having $\mathcal{C}_\mathrm{N}\mathcal{C}_\mathrm{d}\tilde{\mathcal{C}}_\mathrm{d} = \mathcal{C}_\mathrm{a}$. If the latter objective is achieved, the impact of the actuator cyber-attacks will always show up in the sensor measurements and cannot be eliminated by adversaries.

Theorem 2: Under Assumption 2, for any set of compromised and attacked input channels $\mathcal{I}_\mathrm{a}$, i.e., any L_a, adversaries cannot perform controllable cyber-attacks in the sense of Definition 6 if $C_\mathrm{d}B_\mathrm{d}$ is a full rank matrix and $\mathrm{Im}((C_\mathrm{d}B_\mathrm{d})_q) \nsubseteq \mathrm{Im}((D_\mathrm{d})_q)$, for $q = 1, \ldots, m$, where $(C_\mathrm{d}B_\mathrm{d})_q$ and $(D_\mathrm{d})_q$ are the q-th columns of $C_\mathrm{d}B_\mathrm{d}$ and D_d, respectively.

Proof: In case of controllable cyber-attacks, at $k = r_\mathrm{a}^\mathrm{d}$ and $k = r_\mathrm{a}^\mathrm{d} + 1$ one has

$$CA^{r_\mathrm{a}^\mathrm{d}-1}BD_\mathrm{d}L_\mathrm{a}a_\mathrm{u}(0) = 0, \quad (23a)$$

$$CA^{r_\mathrm{a}^\mathrm{d}}BD_\mathrm{d}L_\mathrm{a}a_\mathrm{u}(0) + CA^{r_\mathrm{a}^\mathrm{d}-1}BC_\mathrm{d}B_\mathrm{d}L_\mathrm{a}a_\mathrm{u}(0) + CA^{r_\mathrm{a}^\mathrm{d}-1}BD_\mathrm{d}L_\mathrm{a}a_\mathrm{u}(1) = 0. \quad (23b)$$

The actuator attack signal can be recast as $a_\mathrm{u}(1) = a_\mathrm{u}^\mathrm{c}(1) + a_\mathrm{u}^\mathrm{d}(1)$, where $a_\mathrm{u}^\mathrm{c}(1)$ is designed according to the Corollary 1 for the triple $(C, A, BD_\mathrm{d}L_\mathrm{a})$, and $a_\mathrm{u}^\mathrm{d}(1)$ is designed to cancel out the impact of the coding scheme such that it satisfies

$$CA^{r_\mathrm{a}^\mathrm{d}-1}BC_\mathrm{d}B_\mathrm{d}L_\mathrm{a}a_\mathrm{u}(0) + CA^{r_\mathrm{a}^\mathrm{d}-1}BD_\mathrm{d}L_\mathrm{a}a_\mathrm{u}^\mathrm{d}(1) = 0. \quad (24)$$

The condition (24) is satisfied if

$$C_\mathrm{d}B_\mathrm{d}L_\mathrm{a}a_\mathrm{u}(0) + D_\mathrm{d}L_\mathrm{a}a_\mathrm{u}^d(1) = \zeta\hat{a}_0, \quad (25)$$

where ζ is a scalar and $\hat{a}_0 \in \ker(CA^{r_\mathrm{a}^\mathrm{d}-1}B)$. If $\zeta \neq 0$, $C_\mathrm{d}B_\mathrm{d}L_\mathrm{a}a_\mathrm{u}(0) + D_\mathrm{d}L_\mathrm{a}a_\mathrm{u}^d(1)$ will show up in the next instances of the output, i.e., $k \geq r_\mathrm{a}^\mathrm{d} + 2$. Hence, adversaries may try to design $a_\mathrm{u}^d(1)$ to satisfy (25) for $\zeta = 0$.

There exists $a_\mathrm{u}^\mathrm{d}(1)$ that can satisfy (25) for $\zeta = 0$ if $\mathrm{Im}(C_\mathrm{d}B_\mathrm{d}L_\mathrm{a}a_\mathrm{u}(0)) \subseteq \mathrm{Im}(D_\mathrm{d}L_\mathrm{a})$. Since $C_\mathrm{d}B_\mathrm{d}$ is a square matrix, having a full rank $C_\mathrm{d}B_\mathrm{d}$ such that $\mathrm{Im}((C_\mathrm{d}B_\mathrm{d})_q) \nsubseteq \mathrm{Im}((D_\mathrm{d})_q)$, for $q = 1, \ldots, m$, implies that the q-th column of $C_\mathrm{d}B_\mathrm{d}$ is a basis of $\mathbb{R}^m$ which is different from all the columns of D_d. The latter implies that all the columns of D_d should be accessible by the adversaries, i.e., $L_\mathrm{a} = I_m$, to have $\mathrm{Im}(C_\mathrm{d}B_\mathrm{d}L_\mathrm{a}) \subseteq \mathrm{Im}(D_\mathrm{d}L_\mathrm{a})$. Hence, under Assumption 2, if $\mathrm{Im}((C_\mathrm{d}B_\mathrm{d})_q) \nsubseteq \mathrm{Im}((D_\mathrm{d})_q)$, for $q = 1, \ldots, m$, having any rank deficient L_a results in $\mathrm{Im}(C_\mathrm{d}B_\mathrm{d}L_\mathrm{a}a_\mathrm{u}(0)) \subseteq \mathrm{Im}(C_\mathrm{d}B_\mathrm{d}L_\mathrm{a}) \nsubseteq \mathrm{Im}(D_\mathrm{d}L_\mathrm{a})$. This completes the proof of the theorem. ■

As the main implication of the proposed dynamic coding scheme in Theorem 2, the number of actuators that should be attacked to perform controllable cyber-attacks in the CPS (20) is now equal to m. Hence, even if adversaries have full knowledge of the dynamics of the CPS (1), the encoder $\mathcal{E}$ in (18), and the decoder $\mathcal{D}$ given by (19) (as considered in Assumption 1), the adversaries still need to compromise all the input channels of the CPS to execute controllable cyber-attacks.

VI. Numerical Case Study: Flight Control System of a Fighter Aircraft

The case study considered is concerned with controllable cyber-attacks in the flight control system of a small single-engine fighter aircraft. We obtain the dynamics of the aircraft from [33], [34]. We consider the controllable cyber-attacks that are described in Definition 4. The characteristic matrices

of the linearized aircraft system with the sampling period of $T_s = 0.5$ (s) are given by [33], [34]

$$A = \begin{bmatrix} 1.0214 & 0.0054 & 0.0003 & 0.4176 & -0.0013 \\ 0 & 0.6307 & 0.0821 & 0 & -0.3792 \\ 0 & -3.4485 & 0.3779 & 0 & 1.1569 \\ 1.1199 & 0.0024 & 0.0001 & 1.0374 & -0.0003 \\ 0 & 0.3802 & -0.0156 & 0 & 0.8062 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.1823 & -0.1798 & -0.1795 & 0.0008 \\ 0 & -0.0639 & 0.0639 & 0.1397 \\ 0 & -1.5840 & 1.5840 & 0.2936 \\ 0.8075 & -0.6456 & -0.6456 & 0.0013 \\ 0 & -0.1005 & 0.1005 & -0.4114 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

The matrix B has a full column rank and is an injective map. In this case study, the first 3 actuators of the system are compromised by adversaries, i.e., $\mathcal{I}_a = \{1, 2, 3\}$, and the input channel 4 is attack free. Hence, one has $\text{rank}(L_a) = 3$, and the actuator cyber-attack signature is $B_a = [(B)_1, (B)_2, (B)_3]$, where $(B)_q$ denotes the q-th column of B. Since $C(B)_q \neq 0$, for every $q = 1, \ldots, 4$, each actuator of the system $\Sigma = (C, A, B)$ yields a relative degree equal to 1 which implies that $r_a = 1$.

The basis of the null space of CB_a is $\hat{a}_0 = [-0.8124, -0.4122, -0.4122]^\top$. Given that $\ker(C) \bigcap \ker(CA) = 0$, by utilizing Remark 2, there exists $\hat{a}_1 = [-0.3764, -0.3349, -0.3349]^\top$ that satisfies $CAB_a\hat{a}_0 + CB_a\hat{a}_1 = 0$ and $CA^2B_a\hat{a}_0 + CAB_a\hat{a}_1 = 0$. Consequently, according to Corollary 2 and Definition 2, adversaries are capable of performing controllable cyber-attacks in the sense of Definition 4 and the flight control system Σ is not left invertible. We set $a_u(0) = \hat{a}_0$ and as per Corollary 2 , we design $h(k) = (k+1)^2$ and the actuator attack signal in the following form:

$$a_u(k) = \hat{a}_0 h(k) + \hat{a}_1 h(k-1), \tag{26}$$

for $k \geq 1$. As shown in Fig. 2, the sensor measurements of the flight control system in presence of the controllable attacks and noise are close to zero, while the state of the system is growing unbound.

In order to make the flight control system Σ secure against controllable cyber-attacks in the sense of Definition 6, we design an encoder $\mathcal{E}$ and a decoder $\mathcal{D}$ with their dynamics given by (18) and (19), respectively. The decoder $\mathcal{D}$ and the encoder $\mathcal{E}$ are designed to satisfy the conditions in Lemma 1 such that $A_d = I_4$, $C_d = -C_e = I_4$, $D_d = D_e^{-1} = I_4$, $A_e = A_d - B_eC_d$, and

$$B_e = B_d = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}.$$

Since the q-th column of C_dB_d is different from the q-th column of $D_d = I_4$, for $q = 1, \ldots, 4$, it follows from Theorem 2 that in presence of the dynamic coding scheme, adversaries will not be able to execute controllable cyber-attacks on the flight control system Σ. Moreover, as depicted in Fig. 3, in presence of the proposed dynamic coding scheme, the impact of the controllable cyber-attack given by (26) can now be observed and detected in the sensor measurements due to their unbounded behavior.

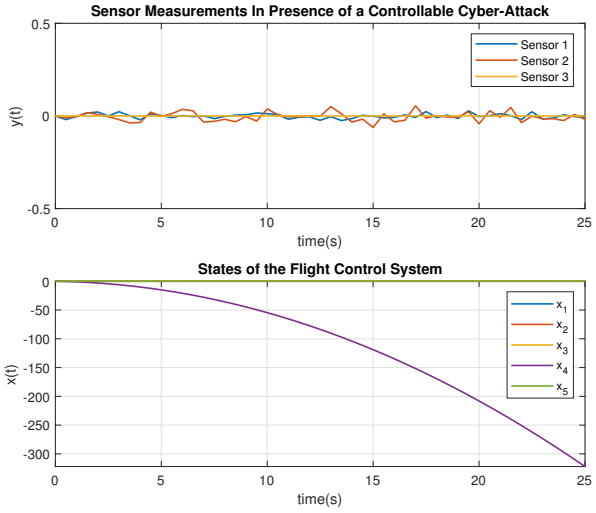

Fig. 2. Controllable cyber-attacks in the flight control system.

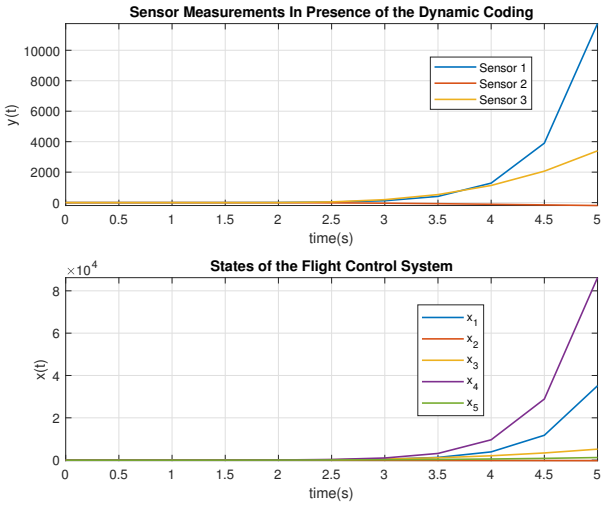

Fig. 3. Impact of the dynamic coding scheme in securing the flight control system against controllable cyber-attacks.

VII. Conclusion

In this paper, we have addressed three main objectives related to controllable cyber-attacks in cyber-physical systems (CPS). First, conditions under which adversaries can execute such cyber-attacks by utilizing the Markov parameters of the CPS and elements of the observability matrix were investigated and developed. Moreover, a novel security metric, termed security effort (SE) for controllable attacks was introduced. SE denotes the minimum number of input communication channels that must be secured to prevent adversaries from executing controllable attacks. Consequently, as a countermeasure, we proposed and studied a coding scheme on the CPS input communication channels that increases the

minimum number of compromised channels necessary for performing controllable attacks. With this coding scheme in place, adversaries need to target all input communication channels to execute controllable attacks. Therefore, CPS operators need to secure only one input communication channel to effectively prevent the occurrence of controllable cyber-attacks. Extending our results to other types of cyber-attacks such as zero dynamics attacks is a topic of our future studies.

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] 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, Nov 2013.

[3] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, no. Supplement C, pp. 135 – 148, 2015.

[4] X. Li, Z. Wang, C. Zhang, D. Du, and M. Fei, "A novel dynamic watermarking-based EKF detection method for fdias in smart grid," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 7, pp. 1319–1322, 2022.

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

[6] W. Duo, M. Zhou, and A. Abusorrah, "A survey of cyber attacks on cyber physical systems: Recent advances and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 5, pp. 784–800, 2022.

[7] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: Risk assessment, detection, and response," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '11. New York, NY, USA: ACM, 2011, pp. 355–366.

[8] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *2008 The 28th International Conference on Distributed Computing Systems Workshops*. IEEE, 2008, pp. 495–500.

[9] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st international conference on High Confidence Networked Systems*, 2012, pp. 55–64.

[10] 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.

[11] 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)*, April 2019, pp. 391–396.

[12] 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)*. IEEE, 2019, pp. 4507–4513.

[13] 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)*, 2016, pp. 1–6.

[14] J. Milošević, 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.

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

[16] S. Weerakkody, X. Liu, S. H. Son, and B. Sinopoli, "A graph-theoretic characterization of perfect attackability for secure design of distributed control systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 60–70, 2017.

[17] K. Zhang, A. Kasis, M. M. Polycarpou, and T. Parisini, "Structural analysis and design for security against actuator stealthy attacks in uncertain systems," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 8051–8056.

[18] M. Taheri, K. Khorasani, and N. Meskin, "The security requirement to prevent zero dynamics attacks and perfectly undetectable cyber-attacks in cyber-physical systems," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 7067–7072.

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

[20] A. Baniamerian, K. Khorasani, and N. Meskin, "Monitoring and detection of malicious adversarial zero dynamics attacks in cyber-physical systems," in *2020 IEEE Conference on Control Technology and Applications (CCTA)*, 2020, pp. 726–731.

[21] M. Taheri, K. Khorasani, I. Shames, and N. Meskin, "Cyberattack and machine-induced fault detection and isolation methodologies for cyber-physical systems," *IEEE Transactions on Control Systems Technology*, vol. 32, no. 2, pp. 502–517, 2024.

[22] A. Hoehn and P. Zhang, "Detection of covert attacks and zero dynamics attacks in cyber-physical systems," in *2016 American Control Conference (ACC)*. IEEE, 2016, pp. 302–307.

[23] R. M. Ferrari and A. M. Teixeira, "A switching multiplicative watermarking scheme for detection of stealthy cyber-attacks," *IEEE Transactions on Automatic Control*, vol. 66, no. 6, pp. 2558–2573, 2020.

[24] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 106–117, 2016.

[25] S. Fang, K. H. Johansson, M. Skoglund, H. Sandberg, and H. Ishii, "Two-way coding in control systems under injection attacks: from attack detection to attack correction," in *Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems*, 2019, pp. 141–150.

[26] Y. Chen, S. Kar, and J. M. Moura, "Dynamic attack detection in cyber-physical systems with side initial state information," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4618–4624, 2016.

[27] 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)*, 2021, pp. 4170–4175.

[28] S. Weerakkody and B. Sinopoli, "Detecting integrity attacks on control systems using a moving target approach," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 5820–5826.

[29] H. L. Trentelman, A. A. Stoorvogel, and M. Hautus, *Control theory for linear systems*. Springer Science & Business Media, 2012.

[30] M. Sain and J. Massey, "Invertibility of linear time-invariant dynamical systems," *IEEE Transactions on Automatic Control*, vol. 14, no. 2, pp. 141–149, 1969.

[31] R. C. Merkle, "Secure communications over insecure channels," *Communications of the ACM*, vol. 21, no. 4, pp. 294–299, 1978.

[32] R. F. Schaefer, H. Boche, and H. V. Poor, "Secure communication under channel uncertainty and adversarial attacks," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1796–1813, 2015.

[33] O. Härkegård and S. T. Glad, "Resolving actuator redundancy—optimal control vs. control allocation," *Automatica*, vol. 41, no. 1, pp. 137–144, 2005.

[34] B. Boussaid, C. Aubrun, J. Jiang, and M. N. Abdelkrim, "Ftc approach with actuator saturation avoidance based on reference management," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 17, pp. 2724–2740, 2014.