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

Switching Control for Identification Deception in Cyber-Physical Systems

Christos Mavridis, Aris Kanellopoulos, Henrik Sandberg, and Karl Henrik Johansson

2024 IEEE 63rd Conference on Decision and Control (CDC) | 979-8-3503-1633-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/CDC56724.2024.10886412

***Abstract*— We investigate the problem of deceiving a malicious agent employing an identification method to estimate the closed-loop dynamics of a cyber-physical system. In particular, we propose a moving target defense mechanism that utilizes stochastic switching between linear closed-loop dynamics to drive a linear system identification process of a potential adversary to sub-optimal solutions with non-vanishing error. We provide a statistical analysis of the induced identification error and show that it is not possible for any linear system identification method to reconstruct the average dynamics of a stochastic switched linear system. Finally, we utilize the theory of Markov jump linear systems to guarantee asymptotic stability of the switching system, and formulate the switching control problem as an optimization problem that guarantees stability while taking into account the trade-off between security and switching effort. Simulation results showcase the efficacy of the proposed approach in inducing identification error for the adversary using minimal switching.**

I. Introduction

Cyber-physical systems (CPS) are large-scale, complex systems where physical interfaces are tightly interconnected with communication and computational devices [1]. A plethora of CPS applications is found in different domains, ranging from the health industry [2], the power grid [3] as well as autonomous ground [4] and aerial vehicles [5]. However, the employment of CPS in safety-critical applications is hindered by the large fault surface that their complexity causes, and by the fact that CPS have become a prime target for malicious agents; attackers that are able to compromise the system either via its software or via its physical layer.

The problem of securing CPS has been addressed by various research communities. Initially, computers scientists have focused on developing more appropriate software defenses for embedded critical systems [6], while similar endeavors have been made regarding the shielding of the underlying communication network of the system [7]. More recently, control-theoretic tools have been proposed for analyzing and developing defense frameworks that capture more abstract behaviors of the system or focus on the vulnerabilities in the cyber/physical boundary. These solutions, however, often take a reactive approach to security, aiming to detect manipulated signals and mitigate their effect to the system. On the other hand, Moving Target Defense (MTD) approaches [8], [9] have been proposed as a security framework that seeks to mitigate this asymmetry by proactively and continuously changing the parameters of the system in order to spoof the attacker. In this sense, MTD approaches for CPS systems are inherently connected to switching control techniques, which are based on alternating between different controllers, in an adaptive context, while aiming to achieve stability and optimize an appropriately defined performance metric [9]–[13]. A special class of stochastic switched systems can be modeled as Markov jump linear systems which capture important behavioral properties and have been extensively studied in terms of their stability properties [14], [15].

In this work, we employ MTD principles to develop a stable switching control system, modeled as a Markov jump linear system, that can hinder the system identification process of a potential adversary. We provide a statistical analysis of the induced identification error, and formulate the optimal switching signal control problem as an optimization problem that takes into account the trade-off between induced identification error and switching effort. Finally, simulation results are provided to illustrate the efficacy of the proposed approach in inducing identification error for the adversary using minimal switching.

A. Related Work

Secure control for CPS applications is an active research field. Starting from [16], where the authors stressed the importance of dynamical systems and control in CPS security, different approaches have been developed. In [17], the authors design control sequences that aim to increase the detectability of attacks in control systems, while from a different point of view, the authors of [18] design control signals that are meant to remain private, using tools from homomorphic encryption. Using dynamical games as their main focus, the authors in [19] design control strategies that mitigate stealthy attacks in a receding-horizon setting.

Proactive defense approaches were initially developed by the computer science community, with special focus on computer networks [20]–[22]. In [23] the authors introduce a constantly shifting IPv6 address mechanism to design a secure internet protocol. In [24], a related proactive defense strategy was formulated to deceive an attacker targeting nodes in a wireless network. A more formalized approach to MTD was first proposed in [25] and led to an MTD entropy hypothesis framework that is more generally applicable. An MTD approach, that is closely related our method, was used to enlarge the dimension of the state space in [26] for the purposes of attack detection, rather than proactive defense based on an unpredictability measure. One possible

Division of Decision and Control, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm. emails:{mavridis,arisk,hsan,kallej}@kth.se.

This work was supported in part by the Swedish Foundation for Strategic Research (SSF) grant IPD23-0019, and the Swedish Civil Contingencies Agency CERCES2 project.

mechanism for introducing unpredictability in a dynamical system is via randomized switching devices; whether those are sensors or actuators. This ideas were explored in [9], [10], where the authors developed an entropy-based probabilistic switching rule that rendered the evolution of the system unpredictable without compromising its stability properties. These notions were also investigated in [12], where the authors considered also the effects of switching to the underlying communication network of the CPS.

Finally, the problem of privacy in dynamical systems has been extensively studied in the literature [27]. In addition, switched system identification approaches to counteract MTD defense approaches have recently been developed using more complex identification methods based on the theory of online deterministic annealing [11], [28]–[31].

B. Contribution

The contribution of this paper is twofold. We first provide a statistical analysis of the properties of linear system identification and quantify the inherent estimation error that arises when the system to be learned utilizes stochastic switching between different linear systems. In particular, we show that it is impossible for any linear system identification method to reconstruct the average dynamics of a stochastic switched linear system. Based on this result, we make use of the principles of Markov jump linear systems and develop an MTD approach that utilizes switching between predefined, and not necessary stable, linear dynamical systems, to hinder the system identification mechanism of a potential attacker, while maintaining stability in the mean square sense. Finally, we formulate the problem of finding the optimal switching strategy as an optimization problem that balances the trade-off between the induced identification error and the switching effort, and illustrate its efficacy in simulated experiments.

C. Notation

The sets $\mathbb{R}$ and $\mathbb{Z}$ represent the sets of real and integer numbers, respectively, while $\mathbb{Z}_+$ represents the set of non-negative integers. For a real matrix $A \in \mathbb{R}^{n\times m}$, $A^{\mathrm{T}} \in \mathbb{R}^{m\times n}$ denotes its transpose. Unless otherwise specified, $\|A\|$ denotes the Euclidean norm of A. The eigenvalue of A with maximum real value is denoted as $\bar{\lambda}(A)$ and the associated eigenvector $\bar{v}(A)$ such that $A\bar{v}(A) = \bar{\lambda}(A)\bar{v}(A)$. The identity matrix dimensions $n \times n$ is denoted by $\mathbb{I}_n$. Unless otherwise specified, random variables $\mathcal{X} : \Omega \to \mathbb{R}^d$ are defined in a probability space $(\Omega, \mathbb{F}, \mathbb{P})$, the probability of an event is denoted by $\mathbb{P}[\mathcal{X} \in S] := \mathbb{P}[\omega \in \Omega : \mathcal{X}(\omega) \in S]$, and the expectation operator as $\mathbb{E}[\mathcal{X}] = \int_\Omega \mathcal{X} \mathrm{d}\mathbb{P}$. Given two random variables $(\mathcal{X}, \mathcal{Y})$, the expectation $\mathbb{E}[f(\mathcal{X}, \mathcal{Y})]$ is understood as the expectation with respect to the joint probability measure, while $\mathbb{E}[\mathcal{X}|\mathcal{Y}] := \mathbb{E}[\mathcal{X}|\sigma(\mathcal{Y})]$ denotes the expectation of $\mathcal{X}$ conditioned to the σ-field of $\mathcal{Y}$. Stochastic processes $\{\mathcal{X}(k)\}_k$, $k \in \mathbb{Z}_+$, are defined in the filtered probability space $(\Omega, \mathbb{F}, \{\mathcal{F}_n\}_n, \mathbb{P})$, where $\mathcal{F}_n = \sigma(\mathcal{X}(k)|k \le n)$, $k \in \mathbb{Z}_+$, is the natural filtration. Finally, $\mathbb{1}_{[\mathcal{X}\in S]}$ denotes the indicator function of the event $[\mathcal{X} \in S]$.

D. Structure

The paper is structured as follows. In Section II we provide a statistical analysis of linear system identification under different assumptions. In Section III, we develop a Moving Target Defense (MTD) framework based on the theory of MJLS systems and formulate it as an optimization problem for identification deception. Section IV presents simulation results that showcase the efficacy of the approach, and, Section V concludes the paper and discusses potential future research directions.

II. Statistical Analysis of System Identification

In this section we introduce a statistical analysis that will be used to formally justify the ability of an MTD framework to hinder the system identification mechanisms of a potential attacker. To simplify our analysis, we consider a random variable $\mathcal{X} : \Omega \to \mathbb{R}^d$ and a single-output map $f : \mathbb{R}^d \times \mathbb{R}^{d_\rho} \to \mathbb{R}$ that defines the random variable:

$$\mathcal{Y} = f(\mathcal{X}, \rho) + \mathcal{W}, \tag{1}$$

where $\rho \in \mathbb{R}^{d_\rho}$ is a parameter vector, and $\mathcal{W} \in \mathbb{R}$ is a random variable with $\mathbb{E}[\mathcal{W}] = 0$, $\mathbb{E}[\mathcal{W}^2] < \infty$, and $\mathbb{E}[\mathcal{X}\mathcal{W}] = \mathbb{E}[\mathcal{X}]\mathbb{E}[\mathcal{W}] = 0$. The input-output pair of random variables $(\mathcal{X}, \mathcal{Y})$ is used to formulate the identification problem of the map f in (1), which reads as follows:

$$\underset{g\in G}{\text{minimize}} \quad L(g) := \mathbb{E}_{(\mathcal{X},\mathcal{Y})}\left[l(\mathcal{Y}, g(\mathcal{X}))\right], \tag{2}$$

where $l : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ is an appropriate dissimilarity measure, and G is the function space that the attacker has the ability to search for. We will assume that $l(x, y) = \frac{1}{2}\|x - y\|^2$, and that the attacker has the capacity to search only within the space of bounded linear functions, i.e.,

$$G = \left\{g : g(x) = \theta^{\mathrm{T}} x,\ \exists c > 0 : \|\theta^{\mathrm{T}} x\| \le c\|x\|\right\}. \tag{3}$$

We restrict to the set of bounded linear functions to showcase the properties of the proposed approach which is based on linear systems, for which identification results are prominent and widely used. Under these assumptions, (2) can be written as:

$$\underset{\theta\in\mathbb{R}^d}{\text{minimize}} \quad L(\theta) := \frac{1}{2}\mathbb{E}_{(\mathcal{X},\mathcal{Y})}\left[\|\mathcal{Y} - \theta^{\mathrm{T}}\mathcal{X})\|^2\right], \tag{4}$$

where with a slight abuse of notation we denote the operator $L(g)|_{g(x)=\theta^{\mathrm{T}}x}$ by $L(\theta)$. The identification error $L(\theta)$ in (4) can be decomposed as:

$$L(\theta) = \frac{1}{2}\mathbb{E}\left[\left(\mathcal{Y} - \mathbb{E}[\mathcal{Y}|\mathcal{X}]\right)^2\right] + \frac{1}{2}\mathbb{E}\left[\left(\mathbb{E}[\mathcal{Y}|\mathcal{X}] - \theta^{\mathrm{T}}\mathcal{X}\right)^2\right]. \tag{5}$$

Notice that the first term does not depend on the identification process, and represents the uncertainty of the output given all possible information from the input. Using (1), the first term satisfies $\frac{1}{2}\mathbb{E}\left[\left(\mathcal{Y} - \mathbb{E}[\mathcal{Y}|\mathcal{X}]\right)^2\right] = \frac{1}{2}\mathbb{E}[\mathcal{W}^2]$. The second term of (5) entails both the estimation and the approximation error. The approximation error refers to the deviation from the optimal parameter vector $\theta^* = \arg\min_\theta L(\theta)$. The estimation error stems from attempting to estimate $\mathbb{E}[\mathcal{Y}|\mathcal{X}] =$

$f(\mathcal{X},\rho)$ by a linear function $g(\mathcal{X})=\theta^{\mathrm{T}}\mathcal{X}$. These two types of errors will become clear in what follows. The MTD approach is based on maintaining high estimation error by not allowing the attacker to know the class of functions that $f(\cdot,\rho)$ in (1) belongs to.

To solve (4), we take $\nabla_\theta L(\theta)=0$, which gives:

$$\begin{aligned}\nabla_\theta L(\theta)=\nabla_\theta\frac{1}{2}\mathbb{E}\left[\left(\mathbb{E}\left[\mathcal{Y}|\mathcal{X}\right]-\theta^{*\mathrm{T}}\mathcal{X}\right)^2\right]=0\\ \Leftrightarrow\quad \mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]\theta^*=\mathbb{E}\left[\mathcal{X}\mathcal{Y}\right].\end{aligned}\tag{6}$$

Assuming that the covariance matrix $\mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]$ is invertible, the optimal parameter vector θ^* is given by:

$$\theta^*=\mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]^{-1}\mathbb{E}\left[\mathcal{X}\mathcal{Y}\right].\tag{7}$$

In practice, one approximates θ^* by

$$\hat{\theta}=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y,\tag{8}$$

by collecting sufficiently many data samples such that $X^{\mathrm{T}}X=\frac{1}{n}\sum_{i=1}^n x_i x_i^{\mathrm{T}}\simeq\mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]$, and $X^{\mathrm{T}}Y=\frac{1}{n}\sum_{i=1}^n x_i y_i\simeq\mathbb{E}\left[\mathcal{X}\mathcal{Y}\right]$, where $X\in\mathbb{R}^{n\times d}$, and $Y\in\mathbb{R}^{n\times 1}$ are matrices comprised of n realizations (observations) of $\mathcal{X}$ and $\mathcal{Y}$, respectively. This approach is well known in the literature as the least-squares identification method.

The error $\|\hat{\theta}-\theta^*\|^2$ is the approximation error and stems from the lack of knowledge of the distribution of $(\mathcal{X},\mathcal{Y})$ and is directly affected by the data samples available and the numerical methods used. However, in this work we are mainly interested in the estimation error that results from not knowing the class of functions G in (2), which requires the measure-theoretic analysis presented in this section.

1) Case of Linear Systems: In this case, we make the usual assumption that $f(x,\rho)=\rho^{\mathrm{T}}\mathcal{X}+\mathcal{W}$ which results in the input-output map:

$$\mathcal{Y}=\rho^{\mathrm{T}}\mathcal{X}+\mathcal{W}.\tag{9}$$

Therefore, (6) gives:

$$\mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]\theta^*=\mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]\rho+\mathbb{E}\left[\mathcal{X}\mathcal{W}\right],\tag{10}$$

which yields

$$\theta^*-\rho=\mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]^{-1}\mathbb{E}\left[\mathcal{X}\mathcal{W}\right]=0,\tag{11}$$

since $\mathbb{E}\left[\mathcal{X}\mathcal{W}\right]=\mathbb{E}\left[\mathcal{X}\right]\mathbb{E}\left[\mathcal{W}\right]=0$. In other words, when the map to be identified belongs to G, the estimation error $\|\theta^*-\rho\|^2$ becomes zero, and the identification method needs only to minimize the approximation error.

2) Case of Nonlinear Systems: In the general case, (6) gives:

$$\theta^*=\mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]^{-1}\mathbb{E}\left[\mathcal{X}f(\mathcal{X},\rho)\right].\tag{12}$$

An interesting case is when f is given by a linear combination of feature vectors $\phi(x)$ (e.g., an artificial fully-connected neural network model), then (12) becomes

$$\theta^*=\mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]^{-1}\mathbb{E}\left[\mathcal{X}\phi(\mathcal{X})\right]\rho.\tag{13}$$

Note that $\|\theta^*-\rho\|^2>0$ so there is a quantifiable non-vanishing estimation error.

3) Case of Switched Linear systems: In this case, we assume that:

$$\begin{cases}\mathcal{Y} &= \rho_1^{\mathrm{T}}\mathcal{X}+\mathcal{W}, \quad \text{if } \mathcal{Z}=1\\ \vdots & \vdots\\ \mathcal{Y} &= \rho_k^{\mathrm{T}}\mathcal{X}+\mathcal{W}, \quad \text{if } \mathcal{Z}=K\end{cases},\tag{14}$$

where the random variable $\mathcal{Z}\in\{1,\dots,K\}$ is independent from $\mathcal{X},\mathcal{W}$. Then (6) gives:

$$\theta^*=\mathbb{E}\left[\mathcal{X}\mathcal{X}^{\mathrm{T}}\right]^{-1}\mathbb{E}\left[\mathbb{E}\left[\mathcal{X}\sum_{i=1}^K\mathbb{1}_{[\mathcal{Z}=i]}\mathcal{X}^{\mathrm{T}}\rho_i\middle|\mathcal{Z}\right]\right],\tag{15}$$

which can be simplified to

$$\theta^*=\sum_{i=1}^k\mathbb{P}\left[\mathcal{Z}=i\right]\rho_i,\tag{16}$$

i.e., the identification method estimates the average parameter vector of the system.

Remark 1: Equation (16) implies that the error $\|\theta^*-\rho(t)\|^2>0$ is non-vanishing at any given point in time t, where $\rho(t)=\sum_{i=1}^k\mathbb{1}_{[\mathcal{Z}=i]}\rho_i$. This means that the estimation error of a potential attacker that makes use of recursive identification methods will not converge, potentially forcing the attacker to restart the identification attempt with more complex methods.

Remark 1 provides the main motivation behind the use of an MTD framework for system identification deception which will be discussed in Remark 2 of Section III-B.

III. Problem Formulation

In this section we design an MTD-based Markov jump linear system that can hinder the system identification process of a potential adversary, and formulate the optimal switching signal control problem as an optimization problem that takes into account the trade-off between induced identification error and switching effort.

A. Moving Target Defense and Switching Control

Consider a discrete-time linear time-invariant system which can be controlled by one of $K\in\mathbb{N}$ actuating modes:

$$\begin{aligned}x(k+1)&=Ax(k)+Bu_{\sigma(k)}(k),\\ x(0)&=x_0,\ \sigma(0)=\sigma_0,\ k\in\mathbb{Z}_+,\end{aligned}\tag{17}$$

where $x\in\mathbb{R}^n$ is the state vector, $u_i\in\mathbb{R}^m$, $i=1,\dots,K$, are control vector associated with the $i-$th actuator, $\sigma:\mathbb{Z}_+\to\{1,\dots,K\}$ is a random process that decides the mode of the system, $A\in\mathbb{R}^{n\times n}$ describe the open-loop dynamics, $B\in\mathbb{R}^{n\times m}$, and $k\in\mathbb{Z}_+$ denotes the time instance.

We consider a set of predefined linear feedback controllers $u_i(x)=K_i x,\ \forall i\in\{1,\dots,K\}$ is available to the system before its operation, with $K_i\in\mathbb{R}^{m\times n}$ being the feedback gain matrix. The actuator/controller pair utilized is decided based on the switching signal $\sigma(k)$, leading to the closed-loop switched system:

$$\begin{aligned}x(k+1)&=(A-BK_{\sigma(k)})x(k)\\ &=\Gamma_{\sigma(k)}x(k),\\ x(0)&=x_0,\ \sigma(0)=\sigma_0,\ k\in\mathbb{Z}_+,\end{aligned}\tag{18}$$

In the case that the switching sequence $\{\sigma(k)\}_k$ is a Markov process, system (18) becomes a Markov Jump Linear System (MJLS). The framework of MJLS allows for the analytic study of the convergence of such stochastic systems and will provide a means to construct the proposed MTD approach for identification deception.

B. Markov Jump Linear Systems

A Markov jump linear system is a stochastic system that can be described by the dynamics:

$$\begin{aligned} x(k+1) &= \Gamma_{\sigma(k)} x(k), \\ x(0) = x_0,\ \sigma(0) &= \sigma_0,\ k \in \mathbb{Z}_+, \end{aligned} \tag{19}$$

where $x(k) \in \mathbb{R}^n$ is the state vector, $\sigma(k) \in \{1, \dots, K\}$ is a random variable that indicates the mode of the system, $\{\Gamma_i\}_{i=1}^K$, $\Gamma_i \in \mathbb{R}^{n\times n}$ are the matrices that define the dynamics for each mode of the system, and $k \in \mathbb{Z}_+$ denotes the time instance. In particular, $\{\sigma(k)\}_k$ is assumed to be a Markov process with transition probability matrix $P \in \mathbb{R}^{K\times K}$, where $p_{ij} = \mathbb{P}\left[\sigma(k+1) = j | \sigma(k) = i\right]$, and $\sum_j p_{ij} = 1$.

In the following, we will constraint our focus to the stability of (19) and the properties of the dynamics of its average state. For a more complete analysis of discrete-time MJLS systems the readers are referred to [14], [15]. Notice that the sequence $\{x(k)\}_k$ of the stochastic system (19) is not a Markov process, but $\{(x(k), \sigma(k))\}_k$ is. To study the stability of (19), we define the first and second moments of the state x as follows:

$$\begin{aligned} \mu(k) &\coloneqq \mathbb{E}\left[x(k)\right] = \sum_{i=1}^K \mathbb{E}\left[x(k)\mathbb{1}_{[\sigma(k)=i]}\right] \coloneqq \sum_{i=1}^K q_i(k), \\ \Sigma(k) &\coloneqq \mathbb{E}\left[x(k)x^{\mathrm{T}}(k)\right] = \sum_{i=1}^K \mathbb{E}\left[x(k)x^{\mathrm{T}}(k)\mathbb{1}_{[\sigma(k)=i]}\right] \coloneqq \sum_{i=1}^K Q_i(k). \end{aligned} \tag{20}$$

It is easy to show that the dynamics for q_i, Q_i, are given by:

$$\begin{cases} q_i(k+1) &= \sum_{j=1}^K p_{ji}\Gamma_j q_j(k), \\ Q_i(k+1) &= \sum_{j=1}^K p_{ji}\Gamma_j Q_j(k)\Gamma_j^{\mathrm{T}}. \end{cases} \tag{21}$$

In addition, we can define linear dynamics for the augmented vectors $q(k) = [q_1^{\mathrm{T}}(k) \dots q_K^{\mathrm{T}}(k)]^{\mathrm{T}} \in \mathbb{R}^{Kn}$, and $Q(k) = [\mathrm{vec}\left(Q_1(k)\right)^{\mathrm{T}} \dots \mathrm{vec}\left(Q_K(k)\right)^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{Kn^2}$, as follows:

$$\begin{cases} q(k+1) &= Mq(k),\ M = \left(P^{\mathrm{T}} \otimes \mathbb{I}_n\right) \mathrm{diag}\left(\{\Gamma_i\}_i\right) \\ Q(k+1) &= TQ(k),\ T = \left(P^{\mathrm{T}} \otimes \mathbb{I}_{n^2}\right) \mathrm{diag}\left(\left\{\Gamma_i^{\mathrm{T}} \otimes \Gamma_i\right\}_i\right). \end{cases} \tag{22}$$

Based on the dynamics (22), we can define and test the stability of system (19) as follows.

Definition 1 ([14]): The MJLS system (19) is Mean Square Stable (MSS), if there is a pair (μ, Σ) such that, for any initial state $x(0) = x_0$ and initial distribution $\sigma(0) = \sigma_0$, it holds that $\mu(k) \to \mu$, and $\Sigma(k) \to \Sigma$.

Theorem 1 ([14]): If the system $Q(k+1) = TQ(k)$ in (21) is stable, then the MJLS system (19) is mean square stable. In addition, if $\mu = 0$, $\Sigma = 0$, then $x(k) \to 0$ almost surely.

Theorem 1 provides a useful condition to test the stability of the system (19). Moreover, it also implies that a collection of unstable systems can be combined through appropriately defined markov jumps to result in a stable system [14]. Although this is a useful result in control applications, we are mainly interested in the average system that a potential attacker can estimate by observations of the MJLS system (19). Notice that, even though the dynamics of the augmented vector $q(k)$ in (22) are linear, the same does not hold for the average state $\mu(k)$. In particular,

$$\begin{aligned} \mu(k+1) &= \sum_{i=1}^K q_i(k+1) = \sum_{i=1}^K \sum_{j=1}^K p_{ji}\Gamma_j q_j(k) \\ &= \sum_{j=1}^K \Gamma_j q_j(k) = [\Gamma_1 \dots \Gamma_K] q(k). \end{aligned} \tag{23}$$

In other words, the average state $\mu(k)$ is linear with respect to the augmented vector $q(k)$, but cannot be described by linear time invariant dynamics, i.e., there exists no matrix $F \in \mathbb{R}^{n\times n}$ such that $\mu(k+1) = F\mu(k)$.

Remark 2: Equation (16) implies that an identification method with input-output observations of the system (18) would recover, in the best case scenario, the following dynamics:

$$m(k+1) = \left(\sum_{i=1}^K \pi_i \Gamma_i\right) m(k),\ m(0) = x_0, \tag{24}$$

where $\pi = [\pi_1 \dots \pi_K]$ is the stationary distribution of the Markov process $\sigma(k)$, i.e., $\pi P = \pi$. However, as shown in (23), the average state of system (18) is not described by linear dynamics, resulting in the deception of a potential attacker that tries to identify the system.

The observations made in Remark 2 constitute the main motivation behind the proposed MTD approach. In fact, as will be shown in Section IV, system (24), which is the one that the attacker can identify, is usually not a good approximation of the actual dynamics (23) of the average state of system (18). In many cases, (24) can even be unstable while the actual stochastic system (18) is mean square stable according to Definition 1 (i.e., the average system (23) is stable).

C. Optimization Problem

In this section we define an optimization problem to choose the optimal transition probability matrix for the MJLS system (18) for the proposed MTD scheme with predefined control matrices $\{K_i\}_{i=1}^K$. As we showed in Section III-A, the stability of (18) can be evaluated by whether or not the matrix $T = \left(\left(P^{\mathrm{T}} \otimes \mathbb{I}_{n^2}\right) \mathrm{diag}\left(\left\{\Gamma_i^{\mathrm{T}} \otimes \Gamma_i\right\}_i\right)\right)$ is Schur, which does not imply that Γ_i, $i = 1, \dots, K$, are Schur. The mean square stability of (18) will be the only performance criterion considered in this work for the defending system. On the other hand, in Section II we showed that the identification method of the attacker can identify the dynamics of system (24), which can even be unstable. Therefore, the MTD approach will focus on maximizing the long-term average the states of system (24), as stated in Problem 1:

Problem 1:

$$\underset{P}{\text{maximize}} \quad \frac{1}{2}\sum_{k=1}^{\infty}\|m(k)\|^2 \tag{25a}$$

$$\text{s.t.} \quad \sum_i p_{ji} = 1 \tag{25b}$$

$$m(k+1) = \left(\sum_{i=1}^{K}\pi_i\Gamma_i\right)m(k),\ m(0) = x_0 \tag{25c}$$

$$\bar{\lambda}\left(\left(P^{\mathrm{T}}\otimes\mathbb{I}_{n^2}\right)\operatorname{diag}\left(\left\{\Gamma_i^{\mathrm{T}}\otimes\Gamma_i\right\}_i\right)\right) < 1 \tag{25d}$$

$$\pi P = \pi \tag{25e}$$

In order to numerically solve Problem 1, we simplify (25a) to maximizing the eigenvalue of the matrix dynamics in (25c) with the largest real part. In addition, to control the switching effort of the system, we make use of the parameterization of $P = P(\tau)$ by a vector $\tau \in \mathbb{R}^{K(K-1)}$, such that each row i can be written as $P_i = \left[\tau_{i1}\dots\tau_{i(i-1)}(1-\sum_{j\neq i}\tau_{ij})\tau_{i(i+1)}\dots\tau_{iK}\right]$. In particular, a design choice to minimize the switching probability, which roughly correlates with less frequent switching over time, can be implemented by the additional objective $\min_\tau \|\tau\|^2$, since the diagonal elements $p_{ii} = (1-\sum_{j\neq i}\tau_{ij})$ of P are being maximized, which implies higher probability of not switching to a different mode. As a result, in this work, we numerically solve Problem 2 as given below:

Problem 2:

$$\underset{\tau}{\text{maximize}} \quad (1-T)\ \bar{\lambda}\left(\sum_{i=1}^{K}\pi_i\Gamma_i\right) - T\ \|\tau\|^2 \tag{26a}$$

$$\text{s.t.} \quad \bar{\lambda}\left(\left(P^{\mathrm{T}}(\tau)\otimes\mathbb{I}_{n^2}\right)\operatorname{diag}\left(\left\{\Gamma_i^{\mathrm{T}}\otimes\Gamma_i\right\}_i\right)\right) < 1 \tag{26b}$$

$$(1-\sum_{j\neq i}\tau_{ij}) \geq 0 \tag{26c}$$

$$\pi P(\tau) = \pi \tag{26d}$$

where $T \in [0,1]$ is a design parameter.

IV. Simulation Results

We evaluate the efficacy of the proposed MTD approach in (26) in a Markov jump linear system of the form (18) with $x(k) \in \mathbb{R}^2$, $x(0) = [1, 0.5]^{\mathrm{T}}$, $u(k) = -K_{\sigma(k)}x(k)$, $\sigma(k) \in [1,2]$ with $\sigma(0) = 1$, and $\Gamma_i = (A - BK_i)$ given by:

$$\begin{cases} x(k+1) &= \begin{bmatrix} 0.5 & -0.5 \\ -0.2 & 0.5 \end{bmatrix} x(k) + w(k), \text{ if } \sigma(k) = 1 \\ x(k+1) &= \begin{bmatrix} 0.9 & -0.1 \\ -0.5 & 0.1 \end{bmatrix} x(k) + w(k), \text{ if } \sigma(k) = 2 \end{cases} \tag{27}$$

The added noise term $w(k)$ has first and second order statistics $\mathbb{E}[w(k)] = 0$, and $\mathbb{E}\left[w^2(k)\right] = 1$. System (27) has two stable modes ($K = 2$) and the and the switching signal $\sigma(k)$ is a Markov process with:

$$P(\tau_{12}, \tau_{21}) = \begin{bmatrix} 1-\tau_{12} & \tau_{12} \\ \tau_{21} & 1-\tau_{21} \end{bmatrix} \tag{28}$$

The parameters $\tau = (\tau_{12}, \tau_{21}) = (0.69, 0.17)$ are computed by numerically solving Problem 2 in a the discretized space $\{0, \Delta\tau, \dots, 1-\Delta\tau, 1\}^2$ with binning resolution $\Delta\tau = 0.01$. The Lagrange parameter T was predefined as $T = 0.9$, which puts more weight in minimizing the switching effort of the controller, compared to destabilizing the behavior of the attacker. This behavior was chosen to showcase that the identification process of the attacker can be hindered even with minimal switching. Figure 1 displays one realization of the trajectory of system (27) against the reconstruction of the trajectory by the attacker through offline least squares system identification (equation (8)) with known initial conditions. The switching signal is also displayed and confirms that even minimal switching can hinder a linear identification method over the actual dynamics.

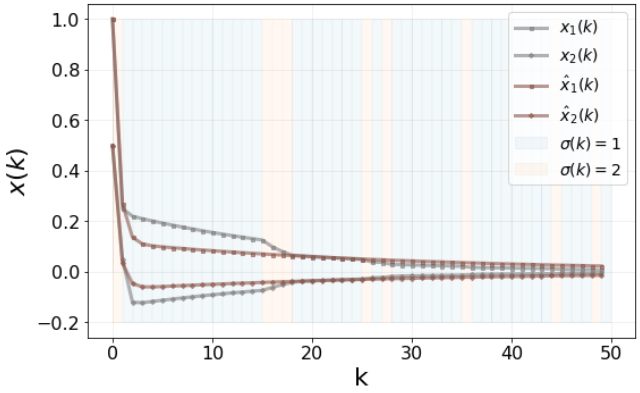

Fig. 1: A realization of the trajectory of system (27) against the reconstruction of the trajectory by the attacker through linear system identification with known initial conditions. The switching signal is also displayed. Notice that even minimal switching can hinder the identification of the dynamics.

Figure 2 displays $N = 50$ realizations of the trajectory of (27), as well as the trajectory of the average state given by (23). In contrast, the attacker is expected to be able to identify a linear system of the form (24). Indeed, system (24) has eigenvalues $\Lambda = (0.9226, 0.0773)$, and the attacker, through all $N = 50$ realizations of system (27), identifies a linear system with eigenvalues $\left\{\hat{\Lambda}_i\right\}_{i=1}^{N}$ with mean value $\mu_\Lambda = (0.9294, 0.1938)$, and standard deviation $\sigma_\Lambda^2 = (0.1087, 0.1031)$.

V. Conclusion and Future Work

In this work, we designed a probabilistic switching scheme for a linear dynamical system. By considering the existence of an agent that tries to learn the dynamics of the system through the use of a batch least squares process, our goal was to impede the correct convergence of his learning method. Thus, by exploiting the gap between the average dynamics

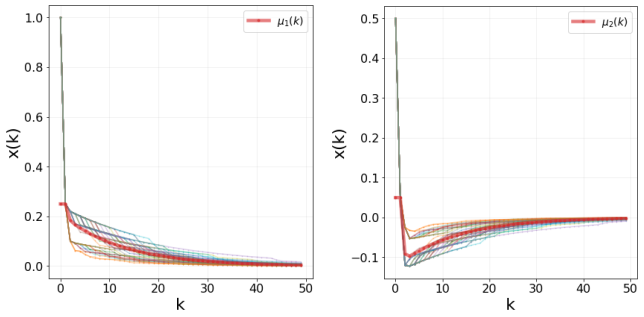

Fig. 2: The trajectory of the average state of system (23) against $N = 50$ realizations of the trajectory of system (27).

induced by the stochastic switching and the average behavior inferred by the learner, we designed an optimization problem that maximizes the trajectory norm of the attacker's learned system while simultaneously guaranteeing that the resulting Markov Jump process that describes the evolution of the real system remains asymptotically stable. The efficacy of the approach was showcased by simulation studies on a 2-dimensional dynamical system with 2 modes of operation.

Future research endeavors include the investigation of efficient numerical solutions for Problem 2, a more comprehensive study of the properties of the switching mechanism as a solution to a stochastic optimal control or a relaxed control problem, and the extension of the proposed approach to continuous-time dynamics. In addition, we will investigate switched system identification approaches to counteract the proposed MTD defense approach using stochastic switched system identification methods based on the theory of Online Deterministic Annealing (ODA) [11], [28]–[30].

REFERENCES

[1] R. R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Proceedings of the 47th design automation conference*. ACM, 2010, pp. 731–736.

[2] Y. Yuehong, Y. Zeng, X. Chen, and Y. Fan, "The internet of things in healthcare: An overview," *Journal of Industrial Information Integration*, vol. 1, pp. 3–13, 2016.

[3] M. H. Cintuglu, O. A. Mohammed, K. Akkaya, and A. S. Uluagac, "A survey on smart grid cyber-physical system testbeds," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 446–464, 2016.

[4] S. Lintelman, K. Sampigethaya, M. Li, R. Poovendran, and R. Robinson, "High assurance aerospace cps & implications for the automotive industry," in *Proceedings of the National Workshop on High Confidence Automotive Cyber-Physical Systems (CPS). Washington, DC*, 2008, pp. 18–20.

[5] H. Wang, H. Zhao, J. Zhang, D. Ma, J. Li, and J. Wei, "Survey on unmanned aerial vehicle networks: A cyber physical system perspective," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1027–1070, 2019.

[6] C. Spensky, A. Machiry, M. Busch, K. Leach, R. Housley, C. Kruegel, and G. Vigna, "Trust. io: protecting physical interfaces on cyber-physical systems," in *2020 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2020, pp. 1–9.

[7] S. Kim, K.-J. Park, and C. Lu, "A survey on network security for cyber–physical systems: From threats to resilient design," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1534–1573, 2022.

[8] S. Jajodia, A. K. Ghosh, V. Swarup, C. Wang, and X. S. Wang, *Moving target defense: creating asymmetric uncertainty for cyber threats*. Springer Science & Business Media, 2011, vol. 54.

[9] A. Kanellopoulos and K. G. Vamvoudakis, "A moving target defense control framework for cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 1029–1043, 2019.

[10] ——, "Entropy-based proactive and reactive cyber-physical security," *Proactive and Dynamic Network Defense*, pp. 59–83, 2019.

[11] C. N. Mavridis, A. Kanellopoulos, K. Vamvoudakis, J. S. Baras, and K. H. Johansson, "Attack identification for cyber-physical security in dynamic games under cognitive hierarchy," *IFAC-PapersOnLine*, 2023.

[12] M. Segovia-Ferreira, J. Rubio-Hernan, R. Cavalli, and J. Garcia-Alfaro, "Switched-based resilient control of cyber-physical systems," *IEEE Access*, vol. 8, pp. 212 194–212 208, 2020.

[13] J. Giraldo, A. Cardenas, and R. G. Sanfelice, "A moving target defense to detect stealthy attacks in cyber-physical systems," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 391–396.

[14] O. L. V. Costa, M. D. Fragoso, and R. P. Marques, *Discrete-time Markov jump linear systems*. Springer Science & Business Media, 2005.

[15] P. Bolzern, P. Colaneri, and G. De Nicolao, "Stochastic stability of positive Markov jump linear systems," *Automatica*, vol. 50, no. 4, pp. 1181–1187, 2014.

[16] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *Distributed Computing Systems Workshops, 2008. ICDCS'08. 28th International Conference on*. IEEE, 2008, pp. 495–500.

[17] M. Hosseinzadeh and B. Sinopoli, "Active attack detection and control in constrained cyber-physical systems under prevented actuation attack," in *2021 American Control Conference (ACC)*. IEEE, 2021, pp. 3242–3247.

[18] A. B. Alexandru, M. Morari, and G. J. Pappas, "Cloud-based MPC with encrypted data," in *2018 IEEE conference on decision and control (CDC)*. IEEE, 2018, pp. 5014–5019.

[19] F. Fotiadis and K. G. Vamvoudakis, "Concurrent receding horizon control and estimation against stealthy attacks," *IEEE Transactions on Automatic Control*, 2022.

[20] S. Jajodia, A. Ghosh, V. Subrahmanian, V. Swarup, C. Wang, and X. Wang, *Moving Target Defense II: Application of Game Theory and Adversarial Modeling*, ser. Advances in Information Security. Springer New York, 2012. [Online]. Available: https://books.google.com/books?id=yFzKRGJatCIC

[21] V. Casola, A. De Benedictis, and M. Albanese, "A multi-layer moving target defense approach for protecting resource-constrained distributed devices," in *Integration of Reusable Systems*. Springer, 2014, pp. 299–324.

[22] J. H. Jafarian, E. Al-Shaer, and Q. Duan, "Openflow random host mutation: transparent moving target defense using software defined networking," in *Proceedings of the first workshop on Hot topics in software defined networks*. ACM, 2012, pp. 127–132.

[23] M. Dunlop, S. Groat, W. Urbanski, R. Marchany, and J. Tront, "Mt6d: A moving target IPv6 defense," in *Military Communications Conference, 2011-Milcom 2011*. IEEE, 2011, pp. 1321–1326.

[24] Z. Lu, C. Wang, and M. Wei, "A proactive and deceptive perspective for role detection and concealment in wireless networks," in *Cyber Deception*. Springer, 2016, pp. 97–114.

[25] R. Zhuang, S. A. DeLoach, and X. Ou, "Towards a theory of moving target defense," in *Proceedings of the First ACM Workshop on Moving Target Defense*. ACM, 2014, pp. 31–40.

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

[27] R. Alisic, M. Molinari, P. E. Paré, and H. Sandberg, "Maximizing privacy in MIMO cyber-physical systems using the Chapman-Robbins bound," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 6272–6277.

[28] C. Mavridis and J. S. Baras, "Identification of piecewise affine systems with online deterministic annealing," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023.

[29] ——, "Annealing optimization for progressive learning with stochastic approximation," *IEEE Transactions on Automatic Control*, vol. 68, no. 5, pp. 2862–2874, 2023.

[30] C. N. Mavridis, A. Kanellopoulos, J. S. Baras, and K. H. Johansson, "State-space piece-wise affine system identification with online deterministic annealing," in *IEEE European Control Conference (ECC)*, 2024.

[31] C. Mavridis and K. H. Johansson, "Real-time hybrid system identification with online deterministic annealing," *arXiv preprint arXiv:2408.01730*, 2024.