2021 IEEE Conference on Control Technology and Applications (CCTA).
August 8-11, 2021. San Diego, California

2021 IEEE Conference on Control Technology and Applications (CCTA).
August 8-11, 2021. San

A Data-Based Moving Target Defense Framework for Switching Zero-Sum Games

Lijing Zhai, *Student Member, IEEE,* Kyriakos G. Vamvoudakis, *Senior Member, IEEE*

2021 IEEE Conference on Control Technology and Applications (CCTA) | 978-1-6654-3643-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/CCTA48906.2021.9659196

***Abstract*— In this paper, a data-based moving target defense framework for cyber-physical systems evolving with unknown and discrete-time dynamics is proposed. Specifically, we develop a proactive mechanism to increase the attacking surface through entropy-based unpredictability measures, and a reactive mechanism to detect and mitigate sensor/actuator attacks. In order to handle worst-case disturbances, we formulate our problem as a zero-sum game, where the minimizing player is the control input and the maximizing player is the disturbance input. We amalgamate a model-free and data-based approximate dynamic programming technique that learns the saddle-point strategies with a Bellman-based intrusion detection mechanism. Switching rules that asymptotically stabilize the switched system are derived. We validate the effectiveness of our proposed framework through simulation results.**

***Index Terms*— Cyber-physical security, zero-sum games, switched systems, moving target defense, model-free.**

I. Introduction

Cyber-physical systems (CPS) are an ensemble of physical systems and computers. These systems are quite complex and work by deep collaboration and integration of physical, communication and computational components. CPS aim to achieve reliable, safe and dynamic cooperation of the cyber domain with the physical system in real time. The computational component of the CPS monitors and controls the physical system with feedback loops. The growing link between the physical and cyber domain and the creation of more advanced learning-based technologies [1] has led to the evolution of CPS into the next-generation smart CPS (sCPS) [2] but opened more vulnerabilities for potential adversaries. In recent years, securing sCPS against malicious attacks has attracted much attention of researchers [3]–[6].

Adversaries can attack sCPS from several angles from availability and integrity to confidentiality [7]. In practice, high-dimensional sCPS are generally easier for malicious attacks to implement successful attacks due to their static and deterministic properties. As a consequence, changing system parameters or structures to generate time-varying and more unpredictable sCPS is a feasible solution. This technique is known as moving target defense (MTD) [8]. In general, time-varying and less static properties introduced by MTD as a moving target enable defenders to deceive attackers and increase attack surface. Applying MTD to addressing CPS security is a growing research area [9]. Various MTD mechanisms have been proposed, such as switching system parameters among multiple modes, adding an extended system with the same performance to the original plant [10], considering a trade-off between security and usability through a game-theoretic framework [11]. One of the most important issues of applying MTD is the design of switching strategies which add unpredictability and guarantee the stability of switched systems [12]–[15]. Two kinds of switching strategies are commonly considered in the literature: a time strategy with an average dwell time [16], [17] and a state-based strategy [17]–[19]. A common quadratic Lyapunov function approach [18], [20] and a less conservative switched (multiple) quadratic Lyapunov functions approach [21]–[23] are utilized to prove the stability of switched systems. Most of the aforementioned works require partial or full knowledge of system dynamics, offline computations which are intractable for large-scale systems. Apart from information of system dynamics, generated data also play important role in sCPS. Data-based learning techniques for CPS security attract much interest in the research community [24], [25], especially approximate/adaptive dynamic programming (ADP) [26].

Contributions: This work expands our previous work [27] to a zero-sum game structure. The contributions of the present work are threefold. First, we develop a data-based MTD framework combined with ADP to solve zero-sum games in a model-free fashion. Moreover, we propose an intrusion detection mechanism in terms of a Bellman error formulation. Finally, switching rules guaranteeing the stability of switched systems with proactive and reactive mechanisms are derived.

Structure: The remainder of the work is organized as follows. We begin with the switching zero-sum games followed by the description of ADP techniques in Section II. In Section III, an intrusion detection mechanism and a data-based MTD framework are proposed. Section IV presents simulation results and Section V summarizes the entire work.

II. Problem Formulation

A. Model Setup

Consider a discrete-time linear switched system in the following form, $\forall k \in \mathbb{Z}$,

$$x_{k+1} = A_{\sigma(k)}x_k + B_{\sigma(k)}u_k + D_{\sigma(k)}d_k, \tag{1}$$

$$y_k = C_{\sigma(k)}x_k, \tag{2}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^l$, $d_k \in \mathbb{R}^q$ and $y_k \in \mathbb{R}^m$ denote state, control input, disturbance input, measurement output,

L. Zhai and K. G. Vamvoudakis are with the Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA e-mail: (lzhai3@gatech.edu, kyriakos@gatech.edu).

This work was supported in part, by the Department of Energy under grant No. DE-EE0008453, by ONR Minerva under grant No. N00014 − 18 − 1 − 2160, by ARO under grant No. W911NF-19 − 1 − 0270, and by NSF under grant Nos. CPS-1851588, and S&AS-1849198.

respectively. $A_{\sigma(k)} \in \mathbb{R}^{n\times n}$, $B_{\sigma(k)} \in \mathbb{R}^{n\times l}$, $D_{\sigma(k)} \in \mathbb{R}^{n\times q}$, and $C_{\sigma(k)} \in \mathbb{R}^{m\times n}$ are the state, control input, disturbance input, and output matrices, respectively. The switching signal $\sigma(k)$ is a piecewise function of time, representing an active mode selected from $M > 1$ available modes, i.e., $\sigma(k) = i \in \mathcal{I} = \{1, 2, \ldots, M\}$, $k \in \mathbb{Z}$. We assume that every subsystem (1)-(2) is controllable and observable with input u_k and output y_k, $\forall k \in \mathbb{Z}$.

B. Attack Strategies

Attack strategies are described as, $\forall k \in \mathbb{Z}$,

$$u_k^a = f_k(u_k), \tag{3}$$
$$y_k^a = g_k(y_k), \tag{4}$$

where f_k (resp., g_k) is a time-varying function of input (resp., output) signal u_k (resp., y_k). Both f_k and g_k are determined by adversaries. Note u_k (resp., y_k) denotes attack-free input (resp., output) signal while u_k^a (resp., y_k^a) represents attacked input (resp., output) signal. Attacks described by (3)-(4) are referred as "data deception attacks" in the literature [28].

Assumption 1. For any time instant, at least one mode is attack-free, i.e., there are i safe modes with $i \in \mathbb{N}$ and $1 \leqslant i \leqslant M$, $k \in \mathbb{Z}$. □

Motivated by [29], the accumulated effects of actuator and sensor attacks modeled by (3)-(4) on the output signals into controllers are quantified as $\tilde{y}_k = y_k + v_k$, where v_k quantifies the overall adversarial impacts of actuator and sensor attacks on sensory output and is described as, $\forall k \in \mathbb{Z}$,

$$v_k = C_{\sigma(k)} x_k^a + (\delta_k - \mathbf{I}_m) y_k, \tag{5}$$
$$x_{k+1}^a = A_{\sigma(k)} x_k^a + B_{\sigma(k)}(\rho_k - \mathbf{I}_l) u_k + D_{\sigma(k)} d_k, \tag{6}$$

where x_k^a is the attacked state with $x_0^a = 0$, $\rho_k \in \mathbb{R}^{l\times l}$ and $\delta_k \in \mathbb{R}^{m\times m}$ are determined by attack strategies (3)-(4). $\mathbf{I}_m$ and $\mathbf{I}_l$ are identity matrices with dimensions m and l.

Definition 1. Attacks that have no dynamics and thus no impact on the system are called *ineffective attacks*, i.e., $v_k = 0$ $\forall k \in \mathbb{Z}$. Ineffective attacks imply $\rho_k = \mathbf{I}_l$ and $\delta_k = \mathbf{I}_m$, or $\rho_k \neq \mathbf{I}_l$ and $(\delta_k - \mathbf{I}_m) y_k = -C_{\sigma(k)} x_k^a \neq 0$. □

Definition 2. Attacks that have effect on the system are called *effective attacks*, i.e., $v_k \neq 0$ $\forall k \in \mathbb{Z}$. □

C. Zero-Sum Game Structure

For each mode $i \in \mathcal{I}$, given the system (1)-(2), we define the infinite horizon cost functional as $J = \sum_{j=0}^{\infty} \gamma^j (y_j^{\mathrm{T}} Q_i y_j + u_j^{\mathrm{T}} R_i u_j - \alpha_i^2 ||d_j||^2)$, where $Q_i \in \mathbb{R}^{m\times m} \geq 0$ and $R_i \in \mathbb{R}^{l\times l} > 0$ are symmetric matrices, $\alpha_i \in \mathbb{R}^+$, $0 < \gamma < 1$ is a discount factor. We aim to find the feedback saddle point solution $(u_k^\star, d_k^\star)$, $\forall k \in \mathbb{Z}$ for the following value function $\forall x_k$,

$$V^\star(x_k) = \min_u \max_d \sum_{j=k}^{\infty} \gamma^{j-k} (y_j^{\mathrm{T}} Q_i y_j + u_j^{\mathrm{T}} R_i u_j - \alpha_i^2 ||d_j||^2). \tag{7}$$

This can be considered as a two-player zero-sum game, where u_k is the minimizing player aiming to improve the performance while d_k is the maximizing player trying to adversely affect the system. In the attack-free case, value functions are quadratic in state given by $V_i(x_k) = x_k^{\mathrm{T}} P_i x_k$ $\forall x_k \in \mathbb{R}^n$, with symmetric and positive definite matrix $P_i \in \mathbb{R}^{n\times n}$ satisfying the game algebraic Riccati equation (GARE) [30],

$$0 = \gamma A_i^{\mathrm{T}} P_i A_i - P_i + C_i^{\mathrm{T}} Q_i C_i - \gamma \begin{bmatrix} A_i^{\mathrm{T}} P_i B_i & A_i^{\mathrm{T}} P_i D_i \end{bmatrix} \times \begin{bmatrix} \frac{R_i}{\gamma} + B_i^{\mathrm{T}} P_i B_i & B_i^{\mathrm{T}} P_i D_i \\ D_i^{\mathrm{T}} P_i B_i & D_i^{\mathrm{T}} P_i D_i - \frac{\alpha_i^2 I_q}{\gamma} \end{bmatrix}^{-1} \begin{bmatrix} B_i^{\mathrm{T}} P_i A_i \\ D_i^{\mathrm{T}} P_i A_i \end{bmatrix}. \tag{8}$$

The optimal control policy $u_k^\star$ and the worst-case disturbance policy $d_k^\star$ for the two-player game (7) are given by $u_k^\star = K_i^\star x_k$ and $d_k^\star = G_i^\star x_k$ with,

$$\begin{aligned} K_i^\star = &-[(B_i^{\mathrm{T}} P_i B_i + R_i/\gamma) + B_i^{\mathrm{T}} P_i D_i (\alpha_i^2 I_q/\gamma \\ &- D_i^{\mathrm{T}} P_i D_i)^{-1} D_i^{\mathrm{T}} P_i B_i]^{-1} B_i^{\mathrm{T}} P_i [I_n + D_i (\alpha_i^2 I_q/\gamma \\ &- D_i^{\mathrm{T}} P_i D_i)^{-1} D_i^{\mathrm{T}} P_i] A_i, \end{aligned} \tag{9}$$

$$\begin{aligned} G_i^\star = &[(\alpha_i^2 I_q/\gamma - D_i^{\mathrm{T}} P_i D_i) + D_i^{\mathrm{T}} P_i B_i (B_i^{\mathrm{T}} P_i B_i \\ &+ R_i/\gamma)^{-1} B_i^{\mathrm{T}} P_i D_i]^{-1} D_i^{\mathrm{T}} P_i [I_n - B_i (B_i^{\mathrm{T}} P_i B_i \\ &+ R_i/\gamma)^{-1} B_i^{\mathrm{T}} P_i] A_i. \end{aligned} \tag{10}$$

D. Data-Based Framework

The system dynamics might be known to adversaries but are unknown to the defender. In order to design MTD switching rules, it is necessary to learn some knowledge of subsystems first. We utilize ADP technique to solve the two-player zero-sum game (7) in a data-driven way. Following the works of [30], [31], we briefly present the following algorithm with the assumption that the system order n and the upper bound N of the observability index are known. For each mode $i \in \mathcal{I}$, the ADP learning phase is attack-free.

Algorithm 1 Model-Free Two-Player Zero-Sum Games

S1: For each mode $i \in \mathcal{I}$, pick $0 < \gamma < 1$. Let the iteration number $j = 0$, and $\bar{P}_i^{(0)}$ be the null matrix.

S2: *Policy evaluation*: Update $\bar{P}_i^{j+1}$ from $V_i^{j+1}(x_k) = y_k^{\mathrm{T}} Q_i y_k + (u_k^j)^{\mathrm{T}} R_i u_k^j - \alpha_i^2 (d_k^j)^{\mathrm{T}} d_k^j + \gamma V_i^j(x_{k+1})$, where $V_i^{j+1}(x_k) = \bar{z}_{k-N,k-1}^{\mathrm{T}} \bar{P}_i^{j+1} \bar{z}_{k-N,k-1}$, $V_i^j(x_{k+1}) = \bar{z}_{k-N+1,k}^{\mathrm{T}} \bar{P}_i^j \bar{z}_{k-N+1,k}$, measured data at time $k \in \mathbb{Z}$ given by $\bar{z}_{k-N,k-1} = \begin{bmatrix} \bar{d}_{k-N,k-1}^{\mathrm{T}} & \bar{u}_{k-N,k-1}^{\mathrm{T}} & \bar{y}_{k-N,k-1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ with

$$\bar{d}_{k-N,k-1} = \begin{bmatrix} d_{k-N} & d_{k-N+1} & \ldots & d_{k-1} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{qN},$$
$$\bar{u}_{k-N,k-1} = \begin{bmatrix} u_{k-N} & u_{k-N+1} & \ldots & u_{k-1} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{lN},$$
$$\bar{y}_{k-N,k-1} = \begin{bmatrix} y_{k-N} & y_{k-N+1} & \ldots & y_{k-1} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{mN}.$$

S3: *Policy improvement*: Update policies by,

$$\begin{aligned} u_k^{j+1} = &-[(R_i/\gamma + P_{\mathrm{uu}}^{j+1}) + P_{\mathrm{ud}}^{j+1} (\alpha_i^2 I_q/\gamma - P_{\mathrm{dd}}^{j+1})^{-1} \\ &\times P_{\mathrm{du}}^{j+1}]^{-1} [P_{\mathrm{u\bar{d}}}^{j+1} \bar{d}_{k-N+1,k-1} + P_{\mathrm{u\bar{u}}}^{j+1} \bar{u}_{k-N+1,k-1} \\ &+ P_{\mathrm{u\bar{y}}}^{j+1} \bar{y}_{k-N+1,k} + P_{\mathrm{ud}}^{j+1} (\alpha_i^2 I_q/\gamma - P_{\mathrm{dd}}^{j+1})^{-1} \\ &\times (P_{\mathrm{d\bar{d}}}^{j+1} \bar{d}_{k-N+1,k-1} + P_{\mathrm{d\bar{u}}}^{j+1} \bar{u}_{k-N+1,k-1} \end{aligned}$$

$$+ P^{j+1}_{\mathrm{d}\bar{\mathrm{y}}}\bar{y}_{k-N+1,k})], \tag{11}$$

$$\begin{aligned} d_k^{j+1} &= [(\alpha_i^2 I_q/\gamma - P_{\mathrm{dd}}^{j+1}) + P_{\mathrm{du}}^{j+1}(R_i/\gamma + P_{\mathrm{uu}}^{j+1})^{-1} \\ &\times P_{\mathrm{ud}}^{j+1}]^{-1}[P_{\mathrm{d}\bar{\mathrm{d}}}^{j+1}\bar{d}_{k-N+1,k-1} + P_{\mathrm{d}\bar{\mathrm{u}}}^{j+1}\bar{u}_{k-N+1,k-1} \\ &+ P_{\mathrm{d}\bar{\mathrm{y}}}^{j+1}\bar{y}_{k-N+1,k} - P_{\mathrm{du}}^{j+1}(R/\gamma + P_{\mathrm{uu}}^{j+1})^{-1} \\ &\times (P_{\mathrm{u}\bar{\mathrm{d}}}^{j+1}\bar{d}_{k-N+1,k-1} + P_{\mathrm{u}\bar{\mathrm{u}}}^{j+1}\bar{u}_{k-N+1,k-1} \\ &+ P_{\mathrm{u}\bar{\mathrm{y}}}^{j+1}\bar{y}_{k-N+1,k})], \end{aligned} \tag{12}$$

where $P_{\mathrm{d}\bar{\mathrm{d}}} \in \mathbb{R}^{(N-1)q\times(N-1)q}$, $P_{\bar{\mathrm{d}}\mathrm{d}} = P_{\mathrm{d}\bar{\mathrm{d}}}^{\mathrm{T}} \in \mathbb{R}^{(N-1)q\times q}$, $P_{\mathrm{d}\bar{\mathrm{u}}} = P_{\bar{\mathrm{u}}\mathrm{d}}^{\mathrm{T}} \in \mathbb{R}^{(N-1)q\times(N-1)l}$, $P_{\bar{\mathrm{u}}\mathrm{d}} = P_{\mathrm{u}\bar{\mathrm{d}}}^{\mathrm{T}} \in \mathbb{R}^{(N-1)q\times l}$, $P_{\mathrm{d}\bar{\mathrm{y}}} = P_{\bar{\mathrm{y}}\mathrm{d}}^{\mathrm{T}} \in \mathbb{R}^{(N-1)q\times Nm}$, $P_{\mathrm{dd}} \in \mathbb{R}^{q\times q}$, $P_{\mathrm{d}\bar{\mathrm{u}}} = P_{\bar{\mathrm{u}}\mathrm{d}}^{\mathrm{T}} \in \mathbb{R}^{q\times(N-1)l}$, $P_{\mathrm{du}} = P_{\mathrm{ud}}^{\mathrm{T}} \in \mathbb{R}^{q\times l}$, $P_{\mathrm{d}\bar{\mathrm{y}}} = P_{\bar{\mathrm{y}}\mathrm{d}}^{\mathrm{T}} \in \mathbb{R}^{q\times Nm}$, $P_{\bar{\mathrm{u}}\bar{\mathrm{u}}} \in \mathbb{R}^{(N-1)l\times(N-1)l}$, $P_{\bar{\mathrm{u}}\mathrm{u}} = P_{\mathrm{u}\bar{\mathrm{u}}}^{\mathrm{T}} \in \mathbb{R}^{(N-1)l\times l}$, $P_{\bar{\mathrm{u}}\bar{\mathrm{y}}} = P_{\bar{\mathrm{y}}\bar{\mathrm{u}}}^{\mathrm{T}} \in \mathbb{R}^{(N-1)l\times Nm}$, $P_{\mathrm{uu}} \in \mathbb{R}^{l\times l}$, $P_{\mathrm{u}\bar{\mathrm{y}}} = P_{\bar{\mathrm{y}}\mathrm{u}}^{\mathrm{T}} \in \mathbb{R}^{l\times Nm}$ and $P_{\bar{\mathrm{y}}\bar{\mathrm{y}}} \in \mathbb{R}^{Nm\times Nm}$ are obtained by partitioning $\bar{P}_i^{j+1}$ as,

$$\bar{P}_i = \begin{bmatrix} P_{\mathrm{d}\bar{\mathrm{d}}} & P_{\bar{\mathrm{d}}\mathrm{d}} & P_{\mathrm{d}\bar{\mathrm{u}}} & P_{\mathrm{du}} & P_{\mathrm{d}\bar{\mathrm{y}}} \\ P_{\mathrm{d}\bar{\mathrm{d}}} & P_{\mathrm{dd}} & P_{\mathrm{d}\bar{\mathrm{u}}} & P_{\mathrm{du}} & P_{\mathrm{d}\bar{\mathrm{y}}} \\ P_{\bar{\mathrm{u}}\bar{\mathrm{d}}} & P_{\bar{\mathrm{u}}\mathrm{d}} & P_{\bar{\mathrm{u}}\bar{\mathrm{u}}} & P_{\bar{\mathrm{u}}\mathrm{u}} & P_{\bar{\mathrm{u}}\bar{\mathrm{y}}} \\ P_{\mathrm{u}\bar{\mathrm{d}}} & P_{\mathrm{ud}} & P_{\mathrm{u}\bar{\mathrm{u}}} & P_{\mathrm{uu}} & P_{\mathrm{u}\bar{\mathrm{y}}} \\ P_{\bar{\mathrm{y}}\bar{\mathrm{d}}} & P_{\bar{\mathrm{y}}\mathrm{d}} & P_{\bar{\mathrm{y}}\bar{\mathrm{u}}} & P_{\bar{\mathrm{y}}\mathrm{u}} & P_{\bar{\mathrm{y}}\bar{\mathrm{y}}} \end{bmatrix}.$$

S4: If $||\bar{P}_i^{j+1} - \bar{P}_i^{j}||_{\mathrm{F}} < \varsigma$, stop, where ς is a small user-defined threshold and $||\cdot||_{\mathrm{F}}$ is the Frobenius norm. Otherwise, let $j = j + 1$ and go to S2.

III. Proactive and Reactive Defense Framework with Intrusion Detection Mechanism

After learning of $\bar{P}_i, \forall i \in \mathcal{I}$, the saddle-point policies and the value functions for each mode at every time instant are available. Then we develop a proactive and reactive MTD framework that ensures the system operates optimally in the absence of attacks, as well as detects and mitigates attacks in the presence of adversaries with a quantified performance.

A. Intrusion Detection Mechanism

Theorem 1. Given the switched systems (1)-(2) operating based on the switching signals $\sigma(k) = i \in \mathcal{I},\ k \in \mathbb{Z}$, define

$$\begin{aligned} \mathbf{e}_k =& y_k^{\mathrm{T}} Q_{\sigma(k)} y_k + u_k^{\mathrm{T}} R_{\sigma(k)} u_k - \alpha_{\sigma(k)}^2 d_k^{\mathrm{T}} d_k + \gamma V_{\sigma(k)}(x_{k+1}) \\ &- V_{\sigma(k)}(x_k), \end{aligned} \tag{13}$$

with $V_{\sigma(k)}(x_k) = \bar{z}_{k-N,k-1}^{\mathrm{T}} \bar{P}_{\sigma(k)} \bar{z}_{k-N,k-1}$ and $V_{\sigma(k)}(x_{k+1}) = \bar{z}_{k-N+1,k}^{\mathrm{T}} \bar{P}_{\sigma(k)} \bar{z}_{k-N+1,k}$. The system is corrupted by adversaries if and only if Bellman error $\|\mathbf{e}_k\|_2 \geqslant \xi$, with a user-defined threshold $\xi \in \mathbb{R}^+$.

Proof. It follows from (7) that the Bellman equation is $V_{\sigma(k)}(x_k) = y_k^{\mathrm{T}} Q_{\sigma(k)} y_k + u_k^{\mathrm{T}} R_{\sigma(k)} u_k - \alpha_{\sigma(k)}^2 d_k^{\mathrm{T}} d_k + \gamma V_{\sigma(k)}(x_{k+1}), \forall k \in \mathbb{Z}$. Consequently, define the Hamiltonian function as $H_{\sigma(k)}(x_k, u_k, d_k) = y_k^{\mathrm{T}} Q_{\sigma(k)} y_k + u_k^{\mathrm{T}} R_{\sigma(k)} u_k - \alpha_{\sigma(k)}^2 d_k^{\mathrm{T}} d_k + \gamma V_{\sigma(k)}(x_{k+1}) - V_{\sigma(k)}(x_k)$. According to the stationarity conditions for optimality, the optimal control policy (9) and the worst-case disturbance policy (10) should satisfy the discrete-time Hamilton-Jacobi-Isaacs (HJI) equation, i.e., $H_{\sigma(k)}(x_k, u_k^\star, d_k^\star) = 0$. Actuator and sensor attacks following strategies (3)-(4) lead to $H_{\sigma(k)}(\cdot,\cdot,\cdot) \neq 0$, which indicates the presence of attacks. Note that HJI equation is both necessary and sufficient conditions for optimality. But in practice, there are disturbances, noise, inaccurate modeled dynamics in the system. We take such inconsistencies into account by introducing a security threshold $\xi \in \mathbb{R}^+$. ■

B. Switching Law

As shown in Figure 1, k_l denotes the starting time instant of some active mode while k_{l+1} denotes the ending time instant, $\forall l \in \mathbb{N}$. The mode $\sigma(k_l)$ remains active for $k \in [k_l, k_{l+1})$. At the switching moment, the mode at k_l^- is the same to that during the time interval $[k_{l-1}, k_l)$ while the mode at k_l^+ is the same to that during the time interval $[k_l, k_{l+1})$. We assume that states have no jumps at switching moments $k_l \in \mathbb{Z},\ \forall l \in \mathbb{N}$, i.e., $x_{k_l^-} \equiv x_{k_l^+} \equiv x_{k_l}$.

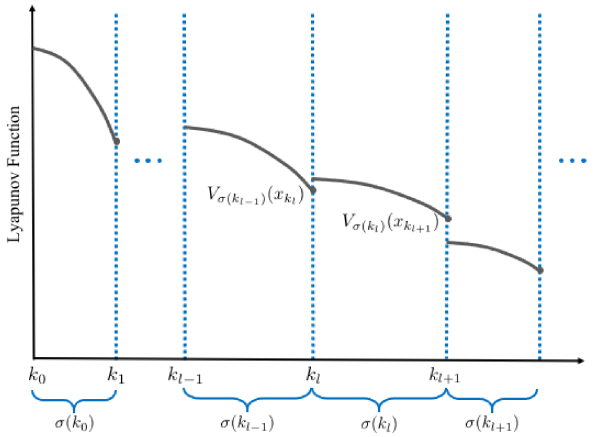

Fig. 1. Illustration of switching conditions in terms of Lyapunov-like functions.

Theorem 2. Given the system (1)-(2) free of data deception attacks with optimal policy gain $K_i^\star$ and worst-case disturbance gain $G_i^\star$, the switched system is stabilized with the following switching condition,

$$V_{\sigma(k_l)}(x_{k_{l+1}}) < V_{\sigma(k_{l-1})}(x_{k_l}), \tag{14}$$

with k_l denoting switching moment, $\forall l \in \mathbb{N}$.

Proof. The following analysis is valid for each mode $i \in \mathcal{I} = \{1, 2, \ldots, M\},\ \forall x_k \in \mathbb{R}^n,\ \forall k \in \mathbb{Z}$. Select the positive and radially unbounded Lyapunov-like functions $V_i(x_k) := x_k^{\mathrm{T}} P_i x_k$ with $P_i > 0$ given by (8). Let $\lambda_{\min}(\cdot)$ (resp., $\lambda_{\max}(\cdot)$) denotes the minimum (resp., maximum) eigenvalue. It follows from Rayleigh-Ritz inequality for symmetric and positive definite matrices that $\lambda_{\min}(P_i)\|x_k\|_2^2 \leqslant x_k^{\mathrm{T}} P_i x_k \leqslant \lambda_{\max}(P_i)\|x_k\|_2^2$. Then we have,

$$\frac{V_i(x_k)}{\lambda_{\max}(P_i)} \leqslant \|x_k\|_2^2. \tag{15}$$

Applying $K_i^\star$ and $G_i^\star$, we get the closed-loop dynamics $x_{k+1} = (A_i + B_i K_i^\star + D_i G_i^\star) x_k = A_{\mathrm{cl},i} x_k$. Note that with the optimal control policy $K_i^\star$ and the worst-case disturbance policy $G_i^\star$, $A_{\mathrm{cl},i}$ is Hurwitz. From (15) we have,

$$\Delta V_i(x_k) \triangleq V_i(x_{k+1}) - V_i(x_k) = x_{k+1}^{\mathrm{T}} P_i x_{k+1} - x_k^{\mathrm{T}} P_i x_k$$

$$= x_k^{\mathrm{T}}(A_{\mathrm{cl},i}^{\mathrm{T}}P_iA_{\mathrm{cl},i} - P_i)x_k \leqslant -\lambda_{\min}(P_i - A_{\mathrm{cl},i}^{\mathrm{T}}P_iA_{\mathrm{cl},i})\|x_k\|_2^2$$
$$\leqslant -\frac{\lambda_{\min}(P_i - A_{\mathrm{cl},i}^{\mathrm{T}}P_iA_{\mathrm{cl},i})}{\lambda_{\max}(P_i)}V_i(x_k).$$

We want the inequality to be valid for all the modes and thus we have $\Delta V_i(x_k) < -\min\limits_{i\in\mathcal{I}} \frac{\lambda_{\min}(P_i - A_{\mathrm{cl},i}^{\mathrm{T}}P_iA_{\mathrm{cl},i})}{\lambda_{\max}(P_i)}V_i(x_k)$. Denote $\delta \triangleq \min\limits_{i\in\mathcal{I}} \frac{\lambda_{\min}(P_i - A_{\mathrm{cl},i}^{\mathrm{T}}P_iA_{\mathrm{cl},i})}{\lambda_{\max}(P_i)}$. Rewrite the above inequality as $\Delta V_i(x_k) < -\delta V_i(x_k)$ with $0 < \delta < 1$. Then,

$$V_i(x_{k+1}) < (1-\delta)V_i(x_k). \tag{16}$$

Since $V_i(x_k) > 0,\ i \in \mathcal{I}$, equation (16) implies the value function strictly decreases. The condition (14) is expressed as $V_{\sigma(k_l)}(x_{k_{l+1}}) = \mu(\sigma(k_l), \sigma(k_{l-1}))V_{\sigma(k_{l-1})}(x_{k_l})$ with $0 < \mu(\sigma(k_l), \sigma(k_{l-1})) < 1$. Let $\mu_l \triangleq \mu(\sigma(k_l), \sigma(k_{l-1}))$, then,

$$V_{\sigma(k_l)}(x_{k_{l+1}}) = \mu_l V_{\sigma(k_{l-1})}(x_{k_l}). \tag{17}$$

Note that the same i is on both sides of inequality (16) and thus it is valid within the same mode. Oppositely, the equality (17) is valid when switching happens, i.e., for two connected modes $\sigma(k_l)$ and $\sigma(k_{l-1})$. We start from the mode at time k_l and work backwards to the initial time k_0. From (16)-(17), we get $V_{\sigma(k_l)}(x_{l+1}) < \mu_l V_{\sigma(k_{l-1})}(x_l) < \mu_{l-1}\mu_l V_{\sigma(k_{l-2})}(x_{l-1}) < \cdots < \mu_1 \ldots \mu_{l-1}\mu_l V_{\sigma(k_0)}(x_{k_1}) < (1-\delta)^{k_1-k_0}\mu_1 \ldots \mu_{l-1}\mu_l V_{\sigma(k_0)}(x_{k_0})$. As $k_{l+1} \to \infty$, $V_{\sigma(k_l)}(x_{l+1}) \to 0$. This completes the proof. ■

Remark 1. Each Lyapunov function is monotonically decreasing within the active mode. With the proposed switching condition (14), the value sequence formed by each Lyapunov function at time instants when the corresponding mode becomes active is monotonically decreasing [32]. □

Switching increases unpredictability and thus enhances the proactive defense capability. The optimal control policy and the worst-case disturbance policy are penalized in the cost functional. Hence, there is a trade-off between optimality represented by cost (7) and unpredictability quantified by information entropy $\mathcal{H}(\mathbf{p}) = -\mathbf{p}^{\mathrm{T}}\log(\mathbf{p})$. We use probability simplex $\mathbf{p} = \{p_1, p_2, \ldots, p_M\}$ satisfying $\|\mathbf{p}\|_1 = \sum_{i=1}^{M}|p_i| = 1$ to represent probabilities of all the available mode to be selected at switching moments.

Lemma 1. Consider the switched system (1)-(2) with optimal control gain $K_i^\star$, worst-case disturbance gain $G_i^\star$ and associated value functional (7). The probability simplex for M is computed by,

$$\min_{\mathbf{p}}(\mathbf{V}^{\star\mathrm{T}}\mathbf{p} - \epsilon\mathcal{H}(\mathbf{p}))$$
$$\text{s.t. } \|\mathbf{p}\|_1 = 1 \text{ and } p_i \geqslant 0, i \in \mathcal{I} = \{1, 2, \ldots, M\}.$$

with solution

$$p_i = e^{[-\frac{V_i^\star}{\epsilon} - 1 - \epsilon\log(e^{-1}\sum_{i=1}^{M} e^{\frac{V_i^\star}{\epsilon}})]}. \tag{18}$$

Proof. We follow the spirit of Theorem 1 in [33] to prove the Lemma. ■

Switching rules for the MTD framework are as follows.

- When the detection mechanism indicates no attacks, select a suitable dwell time τ until meeting the switch condition (14). At the switching instant, select next active mode according to probability simplex $\mathbf{p}$.
- When the detection mechanism indicates attacks, the corrupted mode is taken offline for nonavailability and switch to next active mode according to $\mathbf{p}$.
- Once the system reaches an equilibrium, arbitrary switching is enforced.

Corollary 1. The extra cost for switched systems is expressed as $\Delta J = \mathbf{E}[J] - \min_{i\in\mathcal{I}} V_i^\star = \mathbf{V}^{\star\mathrm{T}}\mathbf{p} - \min_{i\in\mathcal{I}} V_i^\star$, with $V_i^\star = \bar{z}_{1,N}^{\mathrm{T}}\bar{P}_i\bar{z}_{1,N}\ \forall \bar{z}_{1,N} \in \mathbb{R}^{(q+l+m)N}$.

Proof. Without switching, the overall optimal mode with the least cost-to-go is utilized during the operating period. Thus, the cost performance is $\min_{i\in\mathcal{I}} V_i^\star$. When MTD framework is implemented, due to mode switching the system operator utilizes the overall optimal mode less often and other modes are used as well. Thus additional cost is generated. Based on probability theory, the expectation of cost performance under the probability simplex $\mathbf{p}$ is $\mathbf{V}^{\star\mathrm{T}}\mathbf{p}$. ■

Remark 2. Lemma 1 shows that entropy levels denoted by ϵ play a role in the probability simplex $\mathbf{p}$. Larger entropy weight adds more unpredictability to the switched system. Thus frequent mode switching with less usage of the overall optimal mode leads to higher additional cost. □

C. Data-Based MTD

Algorithm 2 Data-Based MTD Framework

Procedure
  Given initial state x_0 and time window N.
  for $i = \{1, 2, \ldots, M\}$
    Given the same x_0, run Algorithm 1 to learn $\bar{P}_i$.
    Compute the optimal cost with the same given x_0 and time window N.
  end for
  Solve for the probability simplex $\mathbf{p}$ using (18).
  At $k = 0$, choose the best mode $\sigma(0) = \arg\max\limits_{i\in\mathcal{I}}(p_i)$.
  Propagate the system by using (1)-(2).
  Compute optimal control input u_k by (11) and worst-case disturbance input d_k by (12).
  Compute detection signals using (13).
  If $\|\mathbf{e}_k\| > \xi$, where ξ is a prescribed threshold,
    Start mode switching based on $\mathbf{p}$. Go to 9.
  End if
  If there exists τ such that $\sigma(\tau) = \sigma(k_l^+)$ and $V_{\sigma(\tau)}(x_\tau) < V_{\sigma(k_l^-)}(x_{k_l})$.
    Start mode switching based on $\mathbf{p}$. Go to 9.
  End if
  If there exists τ, such that $V_{\sigma(\tau)}(x_\tau) < \eta$, where η is a small threshold,
    Start mode switching based on $\mathbf{p}$. Go to 9.
  End if
End procedure

IV. Simulation

We consider a second-order discrete-time linear system with $Q = \mathbf{I}$ and $R = \mathbf{I}$. Select $\alpha = 3$. Five modes are considered: $A_1 = A_3 = A_4 = \begin{bmatrix} 1.1 & -0.3 \\ 1 & 0 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & -0.5 \\ 0.7 & 0 \end{bmatrix}$, $A_5 = \begin{bmatrix} 1.1 & -0.2 \\ 1 & 0 \end{bmatrix}$, $B_1 = B_3 = [1\ 0]^{\mathrm{T}}$, $B_2 = [0.9\ 0.1]^{\mathrm{T}}$, $B_4 = [1.1\ 0]^{\mathrm{T}}$, $B_5 = [1\ 0.1]^{\mathrm{T}}$, $D_1 = [0\ 1]^{\mathrm{T}}$, $D_2 = [0.1\ 1.2]^{\mathrm{T}}$, $D_3 = [0.1\ 0.9]^{\mathrm{T}}$, $D_4 = [-0.3\ 0.6]^{\mathrm{T}}$, $D_5 = [0.2\ 1.1]^{\mathrm{T}}$, $C_1 = C_3 = C_5 = \begin{bmatrix} 1 & -0.8 \end{bmatrix}$, $C_2 = \begin{bmatrix} 1.1 & -0.6 \end{bmatrix}$, $C_4 = \begin{bmatrix} 0.9 & -0.8 \end{bmatrix}$. Select $\gamma = 0.2$. The operating time is 200 seconds. Assume malicious attacks start from 50s and end 100s. Algorithm 1 is implemented to learn $\bar{P}$ for each mode. The evolution of $\bar{P}$, states, input and output for mode 1 is shown in Figures 2 and 3. For the remaining modes, evolution plots are similar. Then Algorithm 2 is implemented for a period of 200s. Assume from 50s to 100s there exist both actuator and sensor attacks.

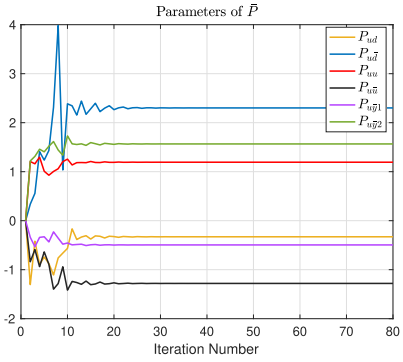

Fig. 2. Evolution of $\bar{P}$ by Algorithm 1 (mode 1). Converged $\bar{P}$ is learned.

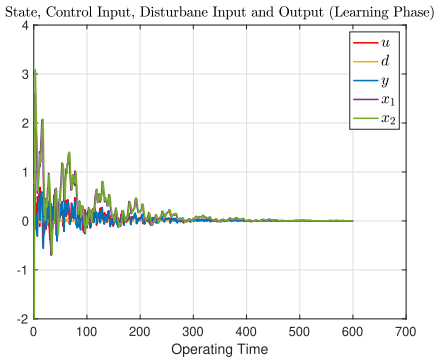

Fig. 3. Evolution of states, input, and output by Algorithm 1 (mode 1). States and output signals converge to zero asymptotically.

The evolution of system states, control input, disturbance input and sensor output is shown in Figure 4. As the compromised mode is taken offline with the proposed reactive defense mechanism, attack effects are mitigated and system remains normal. Figure 5 validates the effectiveness of the detection mechanism. At the transient phase, Bellman error is nonzero though there are no attacks in the system. This is

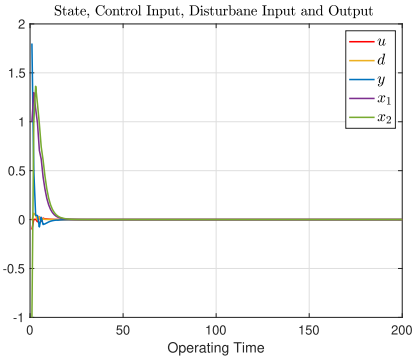

Fig. 4. System evolution under actuator and sensor attacks from 50s to 100s. Compromised modes are taken offline.

due to lack of enough data. Figure 6 depicts the evolution of switching signals. The worst-case attack scenario is assumed for simulation purpose, i.e., from 50s to 100s every active mode is attacked. This means that mode switching happens at every time instant. As shown in Figure 4, fast switching does not cause system fluctuations because of reactive defense mechanism which ensures that the active actuator and sensor modes are safe. Figure 7 presents the relative cost increase due to mode switching with respect to entropy levels. It can be seen that increasing the entropy weight ϵ leads to higher cost increase. This quantifies the performance loss as derived in Corollary 1. As the entropy weight ϵ increases, the unpredictability is also increased and thus mode switching happens more frequently. Since the system picks the less optimal mode more often, a higher cost increase is caused.

V. Conclusions and Future Work

This article proposes a data-based MTD framework with Bellman-based intrusion detection mechanism for switching zero-sum games under attacks. The derived switching rules guarantee the asymptotic stability of switched systems. Simulation results are presented to validate the effectiveness. Future work is to investigate the framework under stochastic exogenous disturbances and measurement noise, as well as the selection of a proper detection threshold.

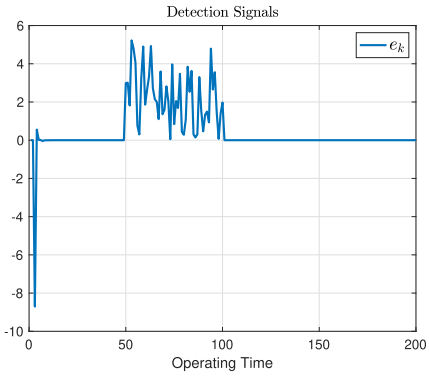

Fig. 5. Evolution of Bellman error signals under attacks from 50s to 100s.

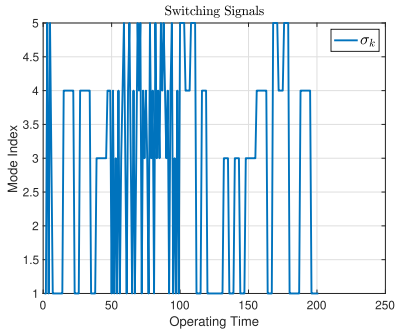

Fig. 6. Evolution of switching signals. The proactive defense mechanism is applied for the attack-free case while the reactive defense mechanism is applied for the attacked case.

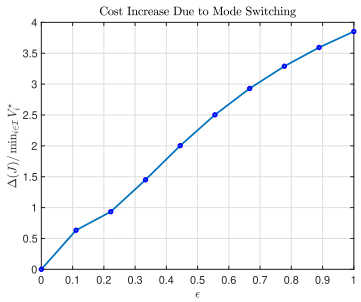

Fig. 7. Relative cost increases due to mode switching with respect to entropy levels. Higher entropy weights lead to larger cost increase.

REFERENCES

[1] D. Vrabie, K. G. Vamvoudakis, and F. L. Lewis, *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*. IET Press, 2012.

[2] F. C. Delicato, A. Al-Anbuky, I. Kevin, and K. Wang, "Smart cyber–physical systems: Toward pervasive intelligence systems," 2020.

[3] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber–physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2011.

[4] A. Sanjab, W. Saad, and T. Başar, "Prospect theory for enhanced cyber-physical security of drone delivery systems: A network interdiction game," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[5] S. R. Etesami and T. Başar, "Dynamic games in cyber-physical security: An overview," *Dynamic Games and Applications*, vol. 9, no. 4, pp. 884–913, 2019.

[6] H. Jafarnejadsani, H. Lee, N. Hovakimyan, and P. Voulgaris, "A multirate adaptive control for mimo systems with application to cyber-physical security," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 6620–6625.

[7] J. Wei, "A data-driven cyber-physical detection and defense strategy against data integrity attacks in smart grid systems," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2015, pp. 667–671.

[8] 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.

[9] H. Okhravi, T. Hobson, D. Bigelow, and W. Streilein, "Finding focus in the blur of moving-target techniques," *IEEE Security & Privacy*, vol. 12, no. 2, pp. 16–26, 2013.

[10] S. Weerakkody and B. Sinopoli, "A moving target approach for identifying malicious sensors in control systems," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2016, pp. 1149–1156.

[11] Q. Zhu and T. Başar, "Game-theoretic approach to feedback-driven multi-stage moving target defense," in *International Conference on Decision and Game Theory for Security*. Springer, 2013, pp. 246–263.

[12] W. Zhang, A. Abate, J. Hu, and M. P. Vitus, "Exponential stabilization of discrete-time switched linear systems," *Automatica*, vol. 45, no. 11, pp. 2526–2536, 2009.

[13] W. Zhang, J. Hu, and A. Abate, "On the value functions of the discrete-time switched lqr problem," *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2669–2674, 2009.

[14] D. Antunes and W. M. Heemels, "Linear quadratic regulation of switched systems using informed policies," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2675–2688, 2016.

[15] J. Zhao, M. Gan, and G. Chen, "Optimal control of discrete-time switched linear systems," *Journal of the Franklin Institute*, 2020.

[16] G. Zhai, B. Hu, K. Yasuda, and A. N. Michel, "Stability analysis of switched systems with stable and unstable subsystems: an average dwell time approach," *International Journal of Systems Science*, vol. 32, no. 8, pp. 1055–1061, 2001.

[17] J. C. Geromel and P. Colaneri, "Stability and stabilization of discrete time switched systems," *International Journal of Control*, vol. 79, no. 07, pp. 719–728, 2006.

[18] G. Zhai, "Quadratic stabilizability of discrete-time switched systems via state and output feedback," in *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No. 01CH37228)*, vol. 3. IEEE, 2001, pp. 2165–2166.

[19] H. Liu, "Finite-time stability for switched linear system based on state-dependent switching strategy," in *2014 International Conference on Mechatronics and Control (ICMC)*. IEEE, 2014, pp. 112–115.

[20] G. Zhai and X. Xu, "A unified approach to analysis of switched linear descriptor systems under arbitrary switching," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*. IEEE, 2009, pp. 3897–3902.

[21] X. Liu and X. Zhao, "Stability analysis of discrete-time switched systems: a switched homogeneous lyapunov function method," *International Journal of Control*, vol. 89, no. 2, pp. 297–305, 2016.

[22] G. Zhai, I. Matsune, J. Imae, and T. Kobayashi, "A note on multiple lyapunov functions and stability condition for switched and hybrid systems," in *2007 IEEE International Conference on Control Applications*. IEEE, 2007, pp. 226–231.

[23] J. Daafouz, P. Riedinger, and C. Iung, "Stability analysis and control synthesis for switched systems: a switched lyapunov function approach," *IEEE transactions on automatic control*, vol. 47, no. 11, pp. 1883–1887, 2002.

[24] P. J. Werbos *et al.*, "Approximate dynamic programming for real-time control and neural modeling," *Handbook of intelligent control: Neural, fuzzy, and adaptive approaches*, vol. 15, pp. 493–525, 1992.

[25] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems*, vol. 32, no. 6, pp. 76–105, 2012.

[26] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.

[27] L. Zhai and K. G. Vamvoudakis, "Data-based and secure switched cyber–physical systems," *Systems & Control Letters*, vol. 148, p. 104826.

[28] 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.

[29] Y. Zhou, K. G. Vamvoudakis, W. M. Haddad, and Z.-P. Jiang, "A secure control learning framework for cyber-physical systems under sensor and actuator attacks," *IEEE Transactions on Cybernetics*, 2020.

[30] L. Zhai and K. G. Vamvoudakis, "A data-based private learning framework for enhanced security against replay attacks in cyber-physical systems," *International Journal of Robust and Nonlinear Control*, 2020.

[31] B. Kiumarsi, F. L. Lewis, M.-B. Naghibi-Sistani, and A. Karimpour, "Optimal tracking control of unknown discrete-time linear systems using input-output measured data," *IEEE transactions on cybernetics*, vol. 45, no. 12, pp. 2770–2779, 2015.

[32] D. Chatterjee and D. Liberzon, "Stability analysis of deterministic and stochastic switched systems via a comparison principle and multiple lyapunov functions," *SIAM Journal on Control and Optimization*, vol. 45, no. 1, pp. 174–206, 2006.

[33] 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.