A Secure Control Learning Framework for Cyber-Physical Systems Under Sensor and Actuator Attacks

Yuanqiang Zhou, *Student Member, IEEE*, Kyriakos G. Vamvoudakis, *Senior Member, IEEE*, Wassim M. Haddad, *Fellow, IEEE*, and Zhong-Ping Jiang, *Fellow, IEEE*

***Abstract*—In this article, we develop a learning-based secure control framework for cyber-physical systems in the presence of sensor and actuator attacks. Specifically, we use a bank of observer-based estimators to detect the attacks while introducing a threat-detection level function. Under nominal conditions, the system operates with a nominal-feedback controller with the developed attack monitoring process checking the reliance of the measurements. If there exists an attacker injecting attack signals to a subset of the sensors and/or actuators, then the attack mitigation process is triggered and a two-player, zero-sum differential game is formulated with the defender being the minimizer and the attacker being the maximizer. Next, we solve the underlying joint state estimation and attack mitigation problem and learn the secure control policy using a reinforcement-learning-based algorithm. Finally, two illustrative numerical examples are provided to show the efficacy of the proposed framework.**

***Index Terms*—Attack estimation, cyber-physical security, differential games, mitigation, reinforcement learning (RL).**

I. Introduction

Cyber-physical systems (CPSs) are systems that integrate sensing, control, and actuation components via a communication network. Cyber-physical security has attracted considerable attention in recent years [1]–[7] with a prominent focus on securing against adversarial attacks [8]–[11]. More specifically, recent work has focused on the effects of specific types of attacks, for example, deception and denial-of-service attacks [12]; false data-injection attacks [13], [14]; and replay attacks or covert attacks affecting system stability, performance, and/or state recovery.

Recently, increasing attention has been focused on the detection and identification of attacks [15]–[20], with several notions, such as detectability and identifiability of an attack [13], [19] and sparse detectability and sparse strong detectability [9] introduced in the literature. Chong *et al.* [10] introduced the notion of observability under attacks and pointed out that in order to characterize the resilience of a system against sensor attacks, the system must be observable under attacks, which requires the number of uncorrupted sensors to be larger than twice the number of the attacked sensors. Using [10], an adaptive switching framework was developed in [9] to address the problem of secure state estimation in the presence of sparse sensor attacks. Mo *et al.* [21], [22] formulated a binary random state detection problem as a min–max optimization, where an attacker can manipulate less than half of the sensors. In [11], a satisfiability modulo theory approach was presented to harness the complexity of secure state estimation under sensor attacks.

Further research on cyber-physical security has been done addressing various classes of attacks. In particular, Zhu and Martinez [23] considered replay attackers who maliciously repeated the messages sent from the operator to the actuators, and analyzed the system performance degradation under an attack-resilient receding horizon control law. Yuan *et al.* [24] considered denial-of-service attacks and developed a coupled design framework that incorporated a cyber configuration policy and robust control. Furthermore, considerable attention has focused on false data-injection attacks and several results have been obtained on secure state estimation and attack mitigation strategy under such attacks [13], [18], [25], [26]. An observer-based, event-triggering consensus control problem was investigated in [25] for a class of discrete-time multiagent systems with lossy sensors. The remote state estimation problem was studied in [26] in the presence of resource constraints. For adversarial sensor and actuator attacks, Jin *et al.* [8] developed an adaptive control architecture with guaranteed uniform ultimate boundedness of the closed-loop system when the attack signals are time varying and partial asymptotic stability when the attack signals are time invariant; Kanellopoulos and Vamvoudakis [27] developed a secure control algorithm that consists of a proactive and a reactive

Manuscript received December 31, 2019; revised April 11, 2020; accepted June 25, 2020. Date of publication July 31, 2020; date of current version September 8, 2021. This work was supported in part by NSF under Grant ECCS-1903781 and Grant S&AS-1849198; in part by NATO under Grant SPS G5176; in part by ONR Minerva under Grant N00014-18-1-2160; in part by NSF CAREER under Grant CPS-1851588; and in part by the Air Force Office of Scientific Research under Grant FA9550-20-1-0038. This article was recommended by Associate Editor Y. Shi. *(Corresponding author: Yuanqiang Zhou.)*

Yuanqiang Zhou was with the Tandon School of Engineering, New York University, Brooklyn, NY 11201 USA. He is now with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhouyuanq@gmail.com).

Kyriakos G. Vamvoudakis and Wassim M. Haddad are with the Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: kyriakos@gatech.edu; wm.haddad@aerospace.gatech.edu).

Zhong-Ping Jiang is with the Control and Networks Lab, Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11201 USA (e-mail: zjiang@nyu.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2020.3006871.

Digital Object Identifier 10.1109/TCYB.2020.3006871

defense mechanism, where the proactive mechanism utilizes a stochastic switching structure to alter the parameters of the system, while hindering the attacker's ability to conduct successful reconnaissance to the system and the reactive mechanism detects potentially attacked components, by leveraging online data to compute an integral Bellman error.

To address cyber-physical security in CPS, several machine-learning techniques have been introduced in [28]–[31]. Reinforcement learning (RL) is a machine-learning technique that enables the learning of optimal control actions by interacting with the environment [32], [33]. It refers to a class of methods that design adaptive policies that can learn the solutions in real time to user-prescribed, optimization-based problems [34], [35]. By considering different cyber-physical layers, several game-theoretic methods have been proposed in [36] and [37], which consider the interaction between the defender and the attacker. Without any knowledge of the system model, Vamvoudakis *et al.* [38], Gao *et al.* [39], and Modares *et al.* [40] proposed RL-driven methods to learn the optimal solutions under persistent adversaries. The strategies proposed in [7], [10], and [11] can be adapted to the solution of attack estimation problems. However, most of these results have a static nature; that is, they do not update and adapt the control policy to ensure that the CPS can continue to perform well under attacks. This can affect the safety of the CPS and it can potentially subvert system stability or deteriorate system performance.

Although An and Yang [9], [41]; Chong *et al.* [10]; and Mousavinejad *et al.* [16] proposed several observer-based or filter-based attack monitoring strategies, they have not used a threat-detection level function to characterize each estimator. Moreover, because of the scalar property of the threat-detection level functions, there is no need to know the values nor the structure of such estimators. In this article, we develop a learning-based secure framework for CPSs in the presence of sensor and actuator attacks. Specifically, we use a bank of observer-based estimators to detect the attacks while introducing a threat-detection level function. Furthermore, our work provides a new direction for the secure control problem in the presence of sensor and/or actuator attacks, which is different from [8], [9], and [17] that require the attacked system to run under a secure model by switching to a secure control strategy. Thus, if there is no attack, then the optimization is reduced to the nominal performance function. Once the attacker affects the performance of the system, a two-player, zero-sum differential game is formulated and a joint state estimation and attack mitigation problem is solved using RL. The advantage of using an RL-driven attack mitigation strategy is that the attack signal applied to the system is not required to be adjustable. This article considerably expands our conference work [42] by providing detailed proofs along with additional results on actuator attacks as well as providing additional discussions and several numerical examples.

The remainder of this article is organized as follows. In Section II, we present the problem formulation and provide the necessary mathematical background for this problem. In Section III, we present a real-time attack monitoring strategy and discuss conditions under which it is effective. In Section IV, a learning-based mitigating algorithm is developed using a two-player, zero-sum differential game framework. In Section V, we discuss the practicality of our approach and provide two numerical examples to illustrate the efficacy of the framework. Finally, we draw conclusions and discuss future research directions in Section VI.

The notation used in this article is fairly standard. Specifically, $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, for a matrix $A \in \mathbb{R}^{n\times n}$, we write $A \succ 0$ and $A \succeq 0$ to denote that A is positive definite and positive semidefinite, respectively. For a vector $x \in \mathbb{R}^n$, x^{T} denotes its transpose, $\|x\|$ denotes the Euclidean norm, $\|M\|$ denotes the induced matrix norm for a real matrix $M \in \mathbb{R}^{n\times m}$, and $\|x\|_P^2$ denotes the quadratic form $x^{\mathrm{T}}Px$ for a real symmetric and positive semidefinite matrix P. Furthermore, $\mathrm{Card}(\mathcal{S})$ denotes the cardinality of a set $\mathcal{S}$, $L_2[0,\infty)$ denotes the Banach space of square-integrable Lebesgue measurable functions on $[0,\infty)$; that is, for all $v(\cdot) \in L_2[0,\infty)$, $\int_0^\infty \|v(t)\|^2 \mathrm{d}t < \infty$, and $H^{m,n}(\Omega)$ denotes the Sobolev space over $\Omega \subset \mathbb{R}^n$ consisting of functions in $L_p(\Omega)$, whose weak derivatives of order up to m are also in $L_p(\Omega)$. Finally, $\mathrm{Supp}(\cdot)$ denotes the support of a vector, that is, the number of its nonzero components.

II. Problem Formulation and Preliminaries

A. Modeling Cyber-Physical Systems and Attack

The physical plant is modeled in a continuous-time state-space form given by

$$\dot{x}(t) = Ax(t) + B\tilde{u}(t), \quad x(0) = x_0, \quad t \geq 0 \tag{1}$$

where $x \in \mathbb{R}^n$ is the state vector and $\tilde{u} \in \mathbb{R}^m$ is the control input applied to the system. We assume that the pair (A, B) is controllable and define $B \overset{\triangle}{=} [B_1, \ldots, B_m] \in \mathbb{R}^{n\times m}$. For the physical system (1), we denote the set of sensors by $\mathcal{S} = \{1, \ldots, q\}$ and the set of actuators by $\mathcal{A} = \{1, \ldots, m\}$. Thus, (1) can be rewritten as

$$\dot{x}(t) = Ax(t) + \sum_{i=1}^{m} B_i\tilde{u}_i(t), \quad x(0) = x_0, \quad t \geq 0 \tag{2}$$

$$y_l(t) = x_j(t), \quad j = 1, \ldots, n, \quad l = 1, \ldots, q \tag{3}$$

where $\tilde{u}_i$ is the control signal associated with the ith actuator, and y_j is the measured output associated with the jth sensor. The physical plant operation is supported by a communication network through which the sensor measurements and actuator data are transmitted and correspond to $y(t)$ and $\tilde{u}(t)$, respectively.

Since the communication network may be unreliable, the data exchanged between the plant and the controller may be altered. Here, we will assume that there are no packet losses and delays occurring in the communication network. Instead, we focus on data corruption due to malicious cyber attacks. In the communication network, the system given by (2) and (3) sends the q output signals $y = [y_1^{\mathrm{T}}, \ldots, y_q^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^q$ to a controller and the controller sends the control signals $u = [u_1^{\mathrm{T}}, \ldots, u_m^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^m$ to the m actuators in (2). However, due to the vulnerability of the communication channels, the ith output that is received by the controller is corrupted to

$\tilde{y}_i$ with $\tilde{y} = [\tilde{y}_1^{\mathrm{T}}, \ldots, \tilde{y}_q^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^q$. Furthermore, the ith control input that is received by (1) is corrupted to $\tilde{u}_i$ with $\tilde{u} = [\tilde{u}_1^{\mathrm{T}}, \ldots, \tilde{u}_m^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^m$.

Under nominal conditions, that is, $\tilde{u} = u$ and $\tilde{y} = y$, (2) and (3) are controlled via an optimal-feedback control law

$$u(t) = -K\tilde{y}(t) \tag{4}$$

where $K \in \mathbb{R}^{m\times q}$ is a feedback control gain that is to be determined by optimizing the performance measure

$$J(x_0, u(\cdot)) = \int_0^\infty \left[x^{\mathrm{T}}(\tau)Qx(\tau) + u^{\mathrm{T}}(\tau)Ru(\tau)\right]\mathrm{d}\tau \tag{5}$$

where $Q \succeq 0$ and $R \succ 0$ are symmetric sign-definite weighting matrices with $(A, \sqrt{Q})$ observable. In this case, the optimal-feedback control gain K in (4) can be characterized via a set of coupled Lyapunov and Riccati equations coupled by an oblique projection [43]–[46].

For (2) and (3), we consider a scenario in which the communication layer at a given instant of time $t_0 \geq 0$, system (2) with an unknown initial state x_0 is subjected to a stealthy attack and the attacker is able to manipulate an unknown subset of the sensors $\mathcal{S}_a \subseteq \mathcal{S}$ and actuators $\mathcal{A}_a \subseteq \mathcal{A}$. In particular, for (2) and (3), the sensor attacks are modeled as

$$\tilde{y}_j(t) = y_j(t) + a_{y,j}(t), \quad j \in \mathcal{S}_a \tag{6}$$

where $a_{y,j}(t)$, $t \geq t_0$ represents the attack signal against the jth sensor. We assume $\mathrm{Card}(\mathcal{S}_a) = p \leq q$, which implies that p sensors have been corrupted by the attacker. If $\mathcal{S}_a = \{j_1, \ldots, j_p\} \subseteq \mathcal{S}$ and $a_y \triangleq [a_{y,j_1}^{\mathrm{T}}, \ldots, a_{y,j_p}^{\mathrm{T}}]^{\mathrm{T}}$, then

$$\tilde{y}(t) = y(t) + D_y a_y(t) \tag{7}$$

where D_y is a $q \times p$ matrix that depends on $\mathcal{S}_a$ and whose entries $(j_1, 1), \ldots, (j_p, p)$ are equal to 1 and the remaining entries are equal to 0. Next, we model the actuator attacks as

$$\tilde{u}_i(t) = u_i(t) + a_{u,i}(t), \quad i \in \mathcal{A}_a \tag{8}$$

where $a_{u,i}(t)$, $t \geq t_0$, represents the attack signal against the ith actuator. We assume $\mathrm{Card}(\mathcal{A}_a) = w \leq m$, which implies that w actuators have been corrupted by the attacker. If $\mathcal{A}_a = \{i_1, \ldots, i_w\} \subseteq \mathcal{A}$ and $a_u \triangleq [a_{u,i_1}^{\mathrm{T}}, \ldots, a_{u,i_w}^{\mathrm{T}}]^{\mathrm{T}}$, then

$$\tilde{u}(t) = u(t) + D_u a_u(t) \tag{9}$$

where D_u is an $m \times w$ matrix that depends on $\mathcal{A}_a$ and whose entries $(i_1, 1), \ldots, (i_w, w)$ are equal to 1 and the remaining entries are equal to 0.

Next, some standard assumptions are made on system (2) and (3) under the attacks given by (7) and (9).

Assumption 1: System (2) is controllable under $\bar{w}$ attacks and observable under $\bar{s}$ attacks.

Assumption 2: The sensor attack vector $a_y(t)$, $t \geq t_0$, in (7) and the actuator attack vector $a_u(t)$, $t \geq t_0$, in (9) only alter a fixed, albeit unknown, subset of the sensors and the actuators, respectively.

Remark 1: In order to monitor and mitigate the effects of potential attacks, Assumption 1 implies that the attacker is not able to compromise all of the actuators and sensors, that is, $\mathrm{Supp}(a_y(t)) < q$ and $\mathrm{Supp}(a_u(t)) < m$, whereas Assumption 2 implies that $\mathrm{Supp}(a_y(t))$ and $\mathrm{Supp}(a_u(t))$ are constant over any time $t \geq 0$.

B. Some Preliminaries

Following the result from [18] and [47], the actuator attacks given by (9) and the sensor attacks given by (7) will eventually corrupt the output measurements that are received by the controller. Thus, the compromised output measurements are given by

$$\tilde{y}(t) = y(t) + \nu(t) \tag{10}$$

where $\nu(\cdot) \in \mathbb{R}^q$ captures the overall attack signal on the actuators and sensors of the system and is given by

$$\nu(t) = x_a(t) + D_y a_y(t) \tag{11}$$

$$\dot{x}_a(t) = Ax_a(t) + BD_u a_u(t) \tag{12}$$

where $x_a(0) = 0$, $a_y(t) \in \mathbb{R}^p$, $t \geq t_0$, and $a_u(t) \in \mathbb{R}^w$, $t \geq t_0$, are obtained from (7) and (9), respectively. Note that (11) and (12) describe the dynamics of the attacker. For (11) and (12), we consider the following two cases.

1) For $\nu(t) = 0$, $t \geq t_0$, the attack signal is an *ineffective attack*. That is, the attacker does not affect our control objective in minimizing (5). In this case, $a_y = 0$, $a_u = 0$, or $D_y a_y = -x_a \neq 0$, $a_u \neq 0$, and hence, the attacker has no dynamics.
2) For $\nu(t) \neq 0$, $t \geq t_0$, the attack signal is an *effective attack*. That is, the attacker can destabilize the system or significantly deteriorate the system performance.

Using the arguments in [10], [41], [48], and [49] controllability and observability under an attack mode can be obtained, which are intrinsic characteristics of the system [11]. Then, to satisfy Assumption 1, we given the following two lemmas, which provide necessary and sufficient conditions for the attacked system to be controllable and observable under attacks.

Lemma 1 [41]: For every set $\Upsilon \in \{\Upsilon \subset \mathcal{A} : \mathrm{Card}(\Upsilon) = \max\{\bar{w}, m - \bar{w}\}$, assume

$$\mathrm{rank}\left(\begin{bmatrix} B_{-\Upsilon} & AB_{-\Upsilon} & \cdots & A^{n-1}B_{-\Upsilon}\end{bmatrix}\right) = n. \tag{13}$$

Then, system (2) is controllable under $\bar{w}$ attacks, where $B_{-\Upsilon}$ is the matrix obtained from B by setting all the columns indexed by the set Υ to zero.

Lemma 2 [10]: System (2) is observable under $\bar{s}$-attacks if and only if $2\bar{s} < q$ and, for every $\mathcal{J} \subset \mathcal{S}$ with $\mathrm{Card}(\mathcal{J}) \geq q - 2\bar{s}$, the pair $(A, C_{\mathcal{J}})$ is observable, where $C_{\mathcal{J}}$ is a matrix obtained by stacking all the vectors C_i, $i \in \mathcal{J}$, and C_i denotes the vector with the ith component 1 and the other components 0.

In this article, we let $\mathrm{Supp}(\nu) = s \leq q$ and assume that the values of w and s are unknown with known upper bounds denoted by $\bar{w}$ and $\bar{s}$, respectively. Then, the control objective of this article is to check the reliance of the measured outputs in real time using our attack monitoring process and design a suboptimal control policy that guarantees resilience to the sensor and actuator attacks while minimizing system performance given by (5) in the absence of attacks.

III. Attack Monitoring

In this section, we develop a monitoring framework using the attacked outputs from at least $q - 2\bar{s}$ elements. For ease of exposition, we let $\mathcal{J} = \{j_1, \dots, j_l\} \subset \mathcal{S}$ denote a subset of the outputs from (10) with $q - 2\bar{s} \le \mathrm{Card}(\mathcal{J}) = l$ to collect the partial outputs $\tilde{y}_{\mathcal{J}}(t) = [\tilde{y}_{j_1}^{\mathrm{T}}(t), \dots, \tilde{y}_{j_l}^{\mathrm{T}}(t)]^{\mathrm{T}}$, $t \ge 0$, and let $\hat{x}_{\mathcal{J}}(t)$, $t \ge 0$, denote the state estimate that uses the partial outputs $\tilde{y}_{\mathcal{J}}(t), t \ge 0$.

A. Threat-Detection Level Function

First, for $\tilde{y}_{\mathcal{J}}(t)$, $t \ge 0$, where $\mathcal{J} \subset \mathcal{S}$, an observer-based detector is designed as

$$\dot{\hat{x}}_{\mathcal{J}}(t) = A\hat{x}_{\mathcal{J}}(t) + Bu(t) + L_{\mathcal{J}}\big(\tilde{y}_{\mathcal{J}}(t) - \hat{y}_{\mathcal{J}}(t)\big)$$

$$\hat{x}_{\mathcal{J}}(0) = \hat{x}(0), \quad t \ge 0 \tag{14}$$

$$\hat{y}_{\mathcal{J}}(t) = C_{\mathcal{J}}\hat{x}_{\mathcal{J}}(t) \tag{15}$$

where $\hat{x}(0)$ is an estimate of the unknown initial state x_0 and $L_{\mathcal{J}}$ is a gain matrix. Using Lemma 2, the pair $(A, C_{\mathcal{J}})$ is observable, and hence, there exists a matrix $L_{\mathcal{J}}$ such that $A_{\mathcal{J}} \triangleq (A - L_{\mathcal{J}}C_{\mathcal{J}})$ is Hurwitz. For every $\mathcal{J}$ with $\mathrm{Card}(\mathcal{J}) = l \ge q - 2\bar{s}$, we let the control layer of the system operate simultaneously with (14) and (15) and let $\tilde{x}(0) = x_0 - \hat{x}(0)$. Thus, we obtain $\binom{q}{l}$ estimates $\hat{x}_{\mathcal{J}}(t), t \ge 0$, with a common initial condition $\hat{x}(0)$ for all estimators. Defining $\tilde{x}_{\mathcal{J}}(t) \triangleq x(t) - \hat{x}_{\mathcal{J}}(t)$ and using (1), (14) and (15), it follows that:

$$\dot{\tilde{x}}_{\mathcal{J}}(t) = A_{\mathcal{J}}\tilde{x}_{\mathcal{J}}(t) - L_{\mathcal{J}}\nu_{\mathcal{J}}(t), \;\; \tilde{x}_{\mathcal{J}}(0) = \tilde{x}(0), \;\; t \ge 0 \tag{16}$$

where $\nu_{\mathcal{J}}(t)$ denotes the attacked signal corresponding to the $\tilde{y}_{\mathcal{J}}(t)$

Next, define the *threat-detection level* function

$$\Upsilon_{\mathcal{J}}(t) \triangleq r_{\mathcal{J}}^{\mathrm{T}}(t)\Xi r_{\mathcal{J}}(t) \tag{17}$$

for each estimate $\hat{x}_{\mathcal{J}}(t), t \ge 0$, where $\Xi \succ 0$ is a positive-definite weighted matrix and $r_{\mathcal{J}}(t) \in \mathbb{R}^q$ is the residual of the measured outputs $\tilde{y}_{\mathcal{J}}(t)$ characterized by the subset $\mathcal{J} \subset \mathcal{S}$ and given by $r_{\mathcal{J}}(t) \triangleq \tilde{y}(t) - \hat{x}_{\mathcal{J}}(t) = \tilde{x}_{\mathcal{J}}(t) + \nu(t)$, $t \ge 0$. Given an upper bound $\bar{\Upsilon} \in \mathbb{R}_+$ for (17), that is, a threshold for the attack-free case with an unknown initial condition, if $\Upsilon_{\mathcal{J}}(t) \ge \bar{\Upsilon}, t \ge 0$, then a violation of the threat level is triggered, and if $\Upsilon_{\mathcal{J}}(t) < \bar{\Upsilon}, t \ge 0$, then the nominal control policy (4) does not change. In other words, if $\Upsilon_{\mathcal{J}}(t) < \bar{\Upsilon}, t \ge 0$, then the system executes a nominal mode of operation; otherwise, the threat detector triggers an alarm.

Remark 2: Since the system initial condition is unknown, we let $\bar{\Upsilon} \ge \Upsilon_0$, where $\Upsilon_0 = \max_{\mathcal{J}} r_{\mathcal{J}}^{\mathrm{T}}(0)\Xi r_{\mathcal{J}}(0)$ and $r_{\mathcal{J}}(0) = e^{A_{\mathcal{J}}T_0}\tilde{x}(0)$. In this case, $\|G_{r,\nu_{\mathcal{J}}}\|_\infty^2 \le \gamma_1$, where $G_{r,\nu_{\mathcal{J}}}(s) \triangleq -(sI - A_{\mathcal{J}})^{-1}L_{\mathcal{J}} + I$ is the transfer function of the system characterized by (16) and γ_1 is a positive constant. If $\|\nu\|_\infty \le \gamma_2$, indicating that the worst case stealthy attack is upper bounded, then $\Upsilon_0 \le \gamma_1\nu_{\mathcal{J}}^{\mathrm{T}}\nu_{\mathcal{J}} \le \gamma_1\gamma_2^2$. Thus, $\gamma_2 \le \sqrt{(1/\gamma_1)\bar{\Upsilon}}$ implies $\Upsilon_0 \le \bar{\Upsilon}$.

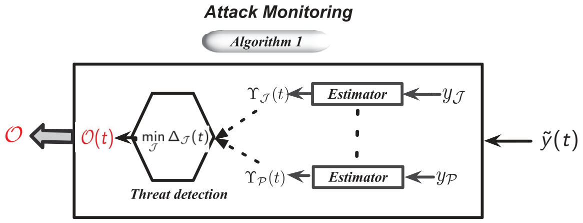


Fig. 1. Obtain the attack-resilient set $\mathcal{O}$ by solving the min–max optimal control problem.

B. Min–Max Optimal Control Problem

Next, we use the threat detection level given by (17) to construct a min–max optimal control problem to sort out a subset of attack-resilient sensors. Consider the following minimization problem:

$$\mathcal{O}(t) = \underset{\{\mathcal{J}:\mathcal{J}\subset\mathcal{S} \text{ and } \mathrm{Card}(\mathcal{J})=l\}}{\arg\min} \Delta_{\mathcal{J}}(t) \tag{18}$$

subject to

$$\Delta_{\mathcal{J}}(t) = \max_{\{\mathcal{P}:\mathcal{P}\subset\mathcal{J} \text{ and } \mathrm{Card}(\mathcal{P})=q-2\bar{s}\le l\}} \big\|\Upsilon_{\mathcal{J}}(t) - \Upsilon_{\mathcal{P}}(t)\big\|.$$

Then, we have the following observations.

1) Given the set $\mathcal{J}$, we have $\Delta_{\mathcal{J}}(t) \ge 0$, $t \ge t_0$. Thus, given all $\binom{q}{l}$ estimates, we can always find a set $\mathcal{J}$ that achieves the smallest threat detection level and is the same as all of its subsets $\mathcal{P} \subset \mathcal{J}$.
2) Although the attack-resilient set $\mathcal{O}(t)$ is a function of the time t, Assumption 2 removes the time dependence on the set $\mathcal{O}(t)$ because Assumption 2 indicates that the set $\mathcal{O}(t)$ in (18), which achieves the minimum at the triggered time, will not change.
3) If more than one subset achieves the minimum, then we determine the attack-resilient set $\mathcal{O}$ by choosing the one with the largest number of sensors.

The detailed framework is shown in Fig. 1. The following proposition shows that the attack-resilient state estimation can be generated using the set $\mathcal{O}$.

Proposition 1: Consider systems (2) and (3) with sensor attacks given by (7) and actuator attacks given by (9), and assume that Assumptions 1 and 2 hold. Then, when the threat detection level is triggered at some time t, we have $\mathcal{O} = \mathcal{O}(t)$, where $\mathcal{O}(t)$ is given by (18), and the attack-resilient state estimation $\hat{x}(\cdot)$ is given by $\hat{x}(t) = \hat{x}_{\mathcal{O}}(t)$, where $\hat{x}_{\mathcal{O}}(t)$, $t \ge 0$, is the attack-resilient state estimate generated by the detector (14) and (15) with the set $\mathcal{O}$.

Proof: Since (1) is $\bar{s}$-attack observable, it follows from Lemma 2 that for every set $\mathcal{J}$ with $\mathrm{Card}(\mathcal{J}) \ge q - 2\bar{s}$, the pair $(A, C_{\mathcal{J}})$ is observable. Thus, using similar arguments as in [10] and [47], it can be shown that $\nu_{\mathcal{O}}(t) = 0, t \ge 0$, and hence, $\hat{x}(0) = \hat{x}_{\mathcal{O}}(0)$ and $\hat{x}(t) = \hat{x}_{\mathcal{O}}(t), t \ge 0$. ■

Theorem 1: Consider the attacked system (2) and (3), and assume that Assumptions 1 and 2 hold. Then, for every unknown initial condition $x_0 \in \mathbb{R}^n$ and input $u(t)$, $t \ge 0$, the estimate of the attack vector $\hat{\nu}(\cdot)$ is given by

$$\hat{\nu}(t) = \tilde{y}(t) - \hat{x}(t) \tag{19}$$

and asymptotically converges to the real attack signal $\nu(t), t \ge 0$.

Proof: See Appendix A. ■

C. Summary of the Attack Monitoring Approach

Our proposed method describing the attack monitoring process is summarized in Algorithm 1.

Algorithm 1 Attack Monitoring Strategy

```
Verify Assumption 1 holds for system (1).
repeat
    Physical layer runs with the nominal control law (4).
    Control layer collects all the outputs and using (14) and (15) generates all the $\binom{q}{l}$ estimators for every set $\mathcal{J}$ with Card($\mathcal{J}$) $= l \geq q - 2\bar{s}$.
    Control layer checks the threat detection level for each set $\mathcal{J}$ in (17).
until One set $\mathcal{J}$ triggers an alarm with $\Upsilon_{\mathcal{J}}(t) \geq \bar{\Upsilon}$ at time $t$.
Determine the attack-resilient set $\mathcal{O}$ using (17)-(18) and reconstruct an estimate of the attack by (19).
```

Remark 3: Note that the estimation algorithm that permits the initial state reconstruction under adversarial attacks is proposed in [10] using a multiple model setup. If the initial condition of system (2) is available, then Algorithm 1 can run with an arbitrary bound. However, in the cases where the initial condition is unknown, the threat detection level is introduced and a threshold-based detection mechanism is developed to sort out a subset of attack-resilient sensors.

IV. Attack Mitigation

After an attacker triggers an alarm using our developed attack monitoring approach, the next step is to develop an attack mitigation framework to mitigate the attacks. In this section, the attack mitigation framework is developed by solving a joint attack-resilient state estimation and attack mitigation problem using an RL-driven, zero-sum differential game with the defender being the minimizer and the attacker being the maximizer.

A. Two-Player, Zero-Sum Differential Game

In this section, a two-player, zero-sum differential game problem is formulated and the solution to this game is obtained. First, the attacked system with (10) can be rewritten as

$$\begin{aligned}\dot{x}(t) &= Ax(t) + B[u(t) - K\nu(t)]\\ &= Ax(t) + Bu(t) - BK\nu(t),\ x(0) = x_0,\ t \geq 0 \quad (20)\end{aligned}$$

where $\nu(\cdot)$ is given by (11) and (12). Using the observer-based detector (14) and (15) for the attack-resilient set $\mathcal{O}$ gives

$$\dot{\hat{x}}_{\mathcal{O}}(t) = A\hat{x}_{\mathcal{O}}(t) + Bu(t) + L_{\mathcal{O}}C_{\mathcal{O}}\tilde{x}_{\mathcal{O}}(t), \quad t \geq 0 \quad (21)$$

where $\hat{x}_{\mathcal{O}}(0) = \hat{x}(0)$. Now, using (20) and (21), with $\tilde{x}_{\mathcal{O}}(t) = x(t) - \hat{x}_{\mathcal{O}}(t), t \geq 0$, it follows that:

$$\begin{aligned}\dot{\tilde{x}}_{\mathcal{O}}(t) &= A\tilde{x}_{\mathcal{O}}(t) - BK\nu(t) - L_{\mathcal{O}}C_{\mathcal{O}}\tilde{x}_{\mathcal{O}}(t)\\ &= (A - L_{\mathcal{O}}C_{\mathcal{O}})\tilde{x}_{\mathcal{O}}(t) - BK\nu(t), \quad t \geq 0 \quad (22)\end{aligned}$$

where $\tilde{x}_{\mathcal{O}}(0) = \tilde{x}(0)$. Then, augmenting the states $\hat{x}_{\mathcal{O}}(t), t \geq 0$, and $\tilde{x}_{\mathcal{O}}(t), t \geq 0$, the augmented state $\xi(t) = [\hat{x}_{\mathcal{O}}^{\mathrm{T}}(t)\ \tilde{x}_{\mathcal{O}}^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{2n}$ satisfies

$$\dot{\xi}(t) = \mathcal{A}\xi(t) + \mathcal{B}u(t) + \mathcal{D}\nu(t), \quad \xi(t_0) = \xi_0, \quad t \geq 0 \quad (23)$$

where

$$\mathcal{A} = \begin{bmatrix} A & L_{\mathcal{O}}C_{\mathcal{O}} \\ 0 & A - L_{\mathcal{O}}C_{\mathcal{O}} \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \quad \mathcal{D} = \begin{bmatrix} 0 \\ -BK \end{bmatrix}$$

and $\xi_0 \triangleq [\hat{x}^{\mathrm{T}}(0)\ \tilde{x}^{\mathrm{T}}(0)]^{\mathrm{T}}$.

Next, augmenting the performance (5) with an attack attenuation level $\gamma \in \mathbb{R}_+$ as a differential game problem yields

$$\begin{aligned}\mathcal{J}(\xi_0, u(\cdot), \nu(\cdot)) = \int_t^{\infty} \Big[\xi^{\mathrm{T}}(\tau)\mathcal{Q}\xi(\tau) + u^{\mathrm{T}}(\tau)Ru(\tau)\\ - \gamma^2\nu^{\mathrm{T}}(\tau)\nu(\tau)\Big]\mathrm{d}\tau, \quad t \geq 0 \quad (24)\end{aligned}$$

where $\mathcal{Q} = \mathrm{diag}[Q, X_{\mathcal{O}}]$, $X_{\mathcal{O}} \in \mathbb{R}^{n\times n}$, $\gamma > 0$, and Q and R are as defined in (5). The attack mitigation problem of finding a secure control policy while optimizing (24) is equivalent to solving the two-player, zero-sum game given by (23) for all $\xi_0 \in \mathbb{R}^{2n}$ and

$$\begin{aligned}V^{\star}(\xi) = \min_{u(\cdot)}\max_{\nu(\cdot)} \int_t^{\infty} \Big[\xi^{\mathrm{T}}(\tau)\mathcal{Q}\xi(\tau) + u^{\mathrm{T}}(\tau)Ru(\tau)\\ - \gamma^2\nu^{\mathrm{T}}(\tau)\nu(\tau)\Big]\mathrm{d}\tau, \quad t \geq 0 \quad (25)\end{aligned}$$

where $V^{\star}(\xi)$ is the optimal value function and $u^{\star}(\xi)$ and $\nu^{\star}(\xi)$ denote the optimal control and attack policies, respectively.

Finally, given system (23) with a performance functional (24), the Hamiltonian function in terms of the secure control policy u and the attack policy ν is defined as

$$\begin{aligned}\mathcal{H}\big(V_{\xi}, u, \nu\big) \triangleq \xi^{\mathrm{T}}\mathcal{Q}\xi + u^{\mathrm{T}}Ru - \gamma^2\nu^{\mathrm{T}}\nu\\ + V_{\xi}(\mathcal{A}\xi + \mathcal{B}u + \mathcal{D}\nu) \quad (26)\end{aligned}$$

where $V_{\xi} = \partial V/\partial\xi$ is the Fréchet derivative of the value function V at ξ. Now, the stationary conditions for optimality $\partial\mathcal{H}/\partial u = 0$ and $\partial\mathcal{H}/\partial\nu = 0$ yield

$$u^{\star}(\xi) = -\frac{1}{2}R^{-1}\mathcal{B}^{\mathrm{T}}V_{\xi}^{\star\mathrm{T}}, \quad \nu^{\star}(\xi) = \frac{1}{2\gamma^2}\mathcal{D}^{\mathrm{T}}V_{\xi}^{\star\mathrm{T}}. \quad (27)$$

Lemma 3: Consider the attack mitigation problem formulated as a two-player, zero-sum game (23) with the cost functional (24) and let $\gamma \in \mathbb{R}_+$. If there exists a positive-semidefinite matrix $Z \in \mathbb{R}^{2n\times 2n}$ satisfying the algebraic bounded real Riccati equation

$$\mathcal{A}^{\mathrm{T}}Z + Z\mathcal{A} + \mathcal{Q} - Z\Big(\mathcal{B}R^{-1}\mathcal{B}^{\mathrm{T}} - \gamma^{-2}\mathcal{D}\mathcal{D}^{\mathrm{T}}\Big)Z = 0 \quad (28)$$

then the control and attack policies $(u^{\star}, \nu^{\star})$ given by (27) generate a saddle point solution in the sense that

$$\mathcal{J}(\xi_0, u^{\star}, \nu) \leq \mathcal{J}(\xi_0, u^{\star}, \nu^{\star}) \leq \mathcal{J}(\xi_0, u, \nu^{\star}) \quad (29)$$

with an optimal value function $V^{\star}(\xi) = \mathcal{J}(\xi_0, u^{\star}, \nu^{\star}) = \xi_0^{\mathrm{T}}Z\xi_0$.

Proof: See [50, Sec. 9.2]. ■

B. RL-Driven Mitigation Algorithm

In this section, we develop an RL-driven mitigation method to approximate the optimal value function and the control and attack policies in (27). To that end, at each iteration i, we define the control input and attack vectors as u^i and ν^i, respectively. Then, (23) with u^i and ν^i can be rewritten as

$$\dot{\xi}(t) = \mathcal{A}\xi(t) + \mathcal{B}u^i(t) + \mathcal{D}\nu^i(t) + \mathcal{B}\big(u(t) - u^i(t)\big) + \mathcal{D}\big(\nu(t) - \nu^i(t)\big), \quad \xi(0) = \xi_0, \quad t \geq 0 \tag{30}$$

where $u^i(t), t \geq 0$ and $\nu^i(t), t \geq 0$, are the control and attack policies to be updated. Here, $\nu(t), t \geq 0$, is not known, and hence, cannot be updated. Thus, to learn the secure control policy, we use $\hat{\nu}(t), t \geq 0$, in (19) to replace the unknown attack signal $\nu(t), t \geq 0$.

Exploiting results from [40], V^i can be found by solving $\mathcal{H}(V^i_\xi, u^i, \nu^i) = 0$ for the control input u^i and the attack ν^i. Then, the learning-based secure control and attack policies are iteratively updated by

$$u^{i+1} = \arg\min_{u} \mathcal{H}\Big(V^i_\xi, u, \nu^{i+1}\Big) = -\frac{1}{2}R^{-1}\mathcal{B}^{\mathrm{T}}V^{i\mathrm{T}}_\xi \tag{31}$$

$$\nu^{i+1} = \arg\max_{\nu} \mathcal{H}\Big(V^i_\xi, u^i, \nu\Big) = \frac{1}{2\gamma^2}\mathcal{D}^{\mathrm{T}}V^{i\mathrm{T}}_\xi. \tag{32}$$

Next, differentiating $V^i(\xi)$ along the solutions of (30) and using (31) and (32) yields

$$\begin{aligned}\dot{V}^i(\xi(t)) &= V^i_\xi(\xi(t))\big(\mathcal{A}\xi(t) + \mathcal{B}u^i(t) + \mathcal{D}\nu^i(t)\big) + V^i_\xi(\xi(t)) \\ &\quad \times \mathcal{B}\big(u(t) - u^i(t)\big) + V^i_\xi(\xi(t))\mathcal{D}\big(\nu(t) - \nu^i(t)\big) \\ &= -\xi^{\mathrm{T}}(t)\mathcal{Q}\xi(t) - u^{i\mathrm{T}}(t)Ru^i(t) + \gamma^2\nu^{i\mathrm{T}}(t)\nu^i(t) \\ &\quad - 2u^{(i+1)\mathrm{T}}(t)R\big(u(t) - u^i(t)\big) \\ &\quad + 2\gamma^2\nu^{(i+1)\mathrm{T}}(t)\big(\nu(t) - \nu^i(t)\big).\end{aligned} \tag{33}$$

Now, integrating (33) over the time interval $[t, t + \delta t]$, where δt is the sampling period, we obtain

$$\begin{aligned}&V^i(\xi(t + \delta t)) - V^i(\xi(t)) \\ &= \int_t^{t+\delta t}\Big[-\xi^{\mathrm{T}}(\tau)\mathcal{Q}\xi(\tau) - u^{i\mathrm{T}}(\tau)Ru^i(\tau) + \gamma^2\nu^{i\mathrm{T}}(\tau)\nu^i(\tau)\Big]\mathrm{d}\tau \\ &\quad - 2\int_t^{t+\delta t}\Big[u^{(i+1)\mathrm{T}}(\tau)R\big(u(\tau) - u^i(\tau)\big) \\ &\qquad - \gamma^2\nu^{(i+1)\mathrm{T}}(\tau)\big(\nu(\tau) - \nu^i(\tau)\big)\Big]\mathrm{d}\tau.\end{aligned} \tag{34}$$

Next, we define N samples satisfying $0 \leq t_j = j\delta t, j = 1, \ldots, N$, and then we use these data samples to iteratively solve (34) for u^{i+1} and ν^{i+1}, which converge to the control and attack policies $u^\star$ and $\nu^\star$ given in (27). For a sufficiently large number of data samples, (34) can be solved using a least-squares method. In particular, to solve (34), we construct three approximators consisting of one critic and two actors as

$$\hat{V}^i(\xi(t)) = \hat{W}^{i\mathrm{T}}_1\phi(\xi(t)) \tag{35}$$

$$\hat{u}^{i+1}(\xi(t)) = \hat{W}^{i\mathrm{T}}_2\varphi(\xi(t)) \tag{36}$$

$$\hat{\nu}^{i+1}(\xi(t)) = \hat{W}^{i\mathrm{T}}_3\psi(\xi(t)) \tag{37}$$

where $\phi(\xi) = [\phi_1(\xi)\ , \ldots,\ \phi_{l_1}(\xi)]^{\mathrm{T}} \in \mathbb{R}^{l_1}$, $\varphi(\xi) = [\varphi_1(\xi)\ , \ldots,\ \varphi_{l_2}(\xi)]^{\mathrm{T}} \in \mathbb{R}^{l_2}$, and $\psi(\xi) = [\psi_1(\xi)\ , \ldots,\ \psi_{l_3}(\xi)]^{\mathrm{T}} \in \mathbb{R}^{l_3}$ are suitable basis functions, $\hat{W}^i_1 \in \mathbb{R}^{l_1}$ is a constant weight vector, $\hat{W}^i_2 = [\hat{W}^i_{2,1}\ , \ldots,\ \hat{W}^i_{2,m}] \in \mathbb{R}^{l_2 \times m}$ and $\hat{W}^i_3 = [\hat{W}^i_{3,1}\ , \ldots,\ \hat{W}^i_{3,q}] \in \mathbb{R}^{l_3 \times q}$ are constant weight matrices, and l_1, l_2, and l_3 are the number of basis functions.

Then, we let $\hat{u}^1 = u$ and $\hat{\nu}^1 = \hat{\nu}$, and given $\hat{u}^i$ and $\hat{\nu}^i$, define $\hat{\zeta}^i \triangleq [\hat{\zeta}^i_1, \ldots, \hat{\zeta}^i_m]^{\mathrm{T}} = u - \hat{u}^i$, $\hat{\varsigma}^i \triangleq [\hat{\varsigma}^i_1, \ldots, \hat{\varsigma}^i_q]^{\mathrm{T}} = \nu - \hat{\nu}^i$. Thus, taking $R = \mathrm{diag}\,[r_1, \ldots, r_m]$ and substituting (35)–(37) into (34) yields

$$\begin{aligned}&\hat{W}^{i\mathrm{T}}_1[\phi(\xi(t + \delta t)) - \phi(\xi(t))] \\ &= \int_t^{t+\delta t}\Big[-\xi^{\mathrm{T}}(\tau)\mathcal{Q}\xi(\tau) - \hat{u}^{i\mathrm{T}}(\tau)R\hat{u}^i(\tau) + \gamma^2\hat{\nu}^{i\mathrm{T}}(\tau)\hat{\nu}^i(\tau)\Big]\mathrm{d}\tau \\ &\quad - 2\sum_{k=1}^{m} r_k\int_t^{t+\delta t}\hat{W}^{i\mathrm{T}}_{2,k}\varphi(\xi(\tau))\hat{\zeta}^i_k\mathrm{d}\tau \\ &\quad + 2\gamma^2\sum_{j=1}^{q}\int_t^{t+\delta t}\hat{W}^{i\mathrm{T}}_{3,j}\psi(\xi(\tau))\hat{\varsigma}^i_j\mathrm{d}\tau + \mu^i(t), \quad t \geq 0\end{aligned} \tag{38}$$

where $\mu^i(t), t \geq 0$, is the Bellman approximation error, $\hat{W}^i_{2,k}$ is the kth column of $\hat{W}^i_2$, and $\hat{W}^i_{3,j}$ is the jth column of $\hat{W}^i_3$. Note that $\mu^i(t), t \geq 0$, is continuous, and if the approximation is performed over a compact set Ω, then $\mu^i(t), t \geq 0$, is bounded [40].

Finally, we rewrite (38) as

$$\pi^i(t) + \mu^i(t) = \hat{W}^{i\mathrm{T}}\theta^i(t) \tag{39}$$

where $\pi^i(t) = \int_t^{t+\delta t}[-\xi^{\mathrm{T}}(\tau)\mathcal{Q}\xi(\tau) - \hat{u}^{i\mathrm{T}}(\tau)R\hat{u}^i(\tau) + \gamma^2\hat{\nu}^{i\mathrm{T}}(\tau)\hat{\nu}^i(\tau)]\mathrm{d}\tau \in \mathbb{R}$, $\hat{W}^i = [\hat{W}^{i\mathrm{T}}_1, \hat{W}^{i\mathrm{T}}_{2,1}\ , \ldots,\ \hat{W}^{i\mathrm{T}}_{2,m}, \hat{W}^{i\mathrm{T}}_{3,1}\ , \ldots,\ \hat{W}^{i\mathrm{T}}_{3,q}]^{\mathrm{T}} \in \mathbb{R}^{l_1+l_2\times m+l_3\times q}$, $\theta^i(t) = [\theta^{i\mathrm{T}}_1(t)\ , \ldots,\ \theta^{i\mathrm{T}}_{1+m+q}(t)]^{\mathrm{T}} \in \mathbb{R}^{l_1+l_2\times m+l_3\times q}$, and $\theta^i_1(t) = \phi(\xi(t + \delta t)) - \phi(\xi(t))$, $\theta^i_k(t) = 2r_k\int_t^{t+\delta t}\varphi(\xi(\tau))\hat{\zeta}^i_k\mathrm{d}\tau$ for $k \in \{2, \ldots, m + 1\}$, and $\theta^i_j(t) = -2\gamma^2\int_t^{t+\delta t}\psi(\xi(\tau))\hat{\varsigma}^i_j\mathrm{d}\tau$ for $j \in \{m+2, \ldots, 1+m+q\}$. Furthermore, we assume that the data samples are collected with $N \gg l_1 + l_2 \times m + l_3 \times q$ (the number of independent entries in $\hat{W}$) points in the state space and $\hat{u}^i$ and $\hat{\nu}^i$ computed over the time interval $[t_j, t_j + \delta t]$, $j = 1, \ldots, N$.

Assumption 3: There exist $\bar{N} \in \mathbb{N}_+$ and $\lambda > 0$ such that for all $N \geq \bar{N}$

$$\Theta^{(i)}\Theta^{(i)\mathrm{T}} \succeq \lambda I_{l_1+l_2\times m+l_3\times q}$$

where $\Theta^{(i)} = [\theta^i(t_1), \ldots, \theta^i(t_N)] \in \mathbb{R}^{(l_1+l_2\times m+l_3\times q)\times N}$.

Using Assumption 3, the weight $\hat{W}^i$ in (35)–(37) that minimizes the approximation error $\sum_{j=1}^{N}\mu^{i\mathrm{T}}(t_j)\mu^i(t_j)$ is given by

$$\hat{W}^i = \Big(\Theta^{(i)}\Theta^{(i)\mathrm{T}}\Big)^{-1}\Theta^{(i)}\Pi^{(i)\mathrm{T}} \tag{40}$$

where $\Pi^{(i)} = [\pi^i(t_1), \ldots, \pi^i(t_N)] \in \mathbb{R}^{1\times N}$.

Lemma 4: Consider the two-player, zero-sum game problem defined in (25) given the dynamics (23). Then, the approximate control and attack policies (31) and (32) converge to the optimal control and attack policies (27) at the rate of $O(1/2^i)$, while the value function in (34) converges to (25) at the rate of $O(1/2^i)$. Furthermore, the closed-loop system (23) and (27) has an asymptotically stable equilibrium point.

Proof: See Appendix B. ■

Theorem 2: Consider the augmented system (23) and assume that Assumption 3 holds. Then, for $\varepsilon > 0$, there

Algorithm 2 Attack Mitigation Strategy

```
Run the attack monitoring process of Algorithm 1 until $\Upsilon_{\mathcal{J}}(t) \geq \bar{\Upsilon}$ is triggered at some time $t$. Select a sufficiently small constant $\varepsilon > 0$ and an integer $N$ satisfying Assumption 3.
repeat
    Data collection: Define the $N$ different samples as $t_j = j\delta t, j = 1, \ldots, N$, obtain $\xi(t_j)$, $\hat{u}(t_j)$, and $\hat{\nu}(t_j)$, and then, form $\Theta^{(i)}$ and $\Pi^{(i)}$.
    Policy search: Compute the solutions to (40).
until $\|\hat{u}^{i+1} - \hat{u}^i\| \leq \varepsilon$.
Apply $u(t) = \hat{u}^i(t)$ to the attacked system (2) instead of (4) and use $\hat{\nu}^i(t)$ as an output amendment to (10).
```

exist integers $i^\star > 0$ and $l^\star > 0$, such that for $i > i^\star$ and $\min\{l_1, l_2, l_3\} > l^\star$

$$\left|\hat{V}^i(\xi) - V^\star(\xi)\right| < \varepsilon$$

$$\left\|\hat{u}^{i+1}(\xi) - u^\star(\xi)\right\| < \varepsilon$$

$$\left\|\hat{\nu}^{i+1}(\xi) - \nu^\star(\xi)\right\| < \varepsilon$$

where ξ belongs to a compact set $\Omega \in \mathbb{R}^{2n}$.

Proof: See Appendix C. ■

C. Summary of the Attack Mitigation Approach

Algorithm 2 gives the proposed attack mitigation method.

Remark 4: Note that in [9] the performance function needs to be redesigned after it removes the tampered-feedback signals. However, in our work, the performance function given by (24) corresponds to an H_∞ norm bound with γ denoting the attack attenuation level. If $\hat{x}_{\mathcal{O}}(0) = x_0$ and $\nu(t) = 0, t \geq 0$, then (24) is reduced to the performance criterion (5) reflecting the absence of adversarial attacks. In addition, we optimize (24) only when we detect the attacks.

Remark 5: Note that (28) gives the condition for selecting γ in (24) and, solving (28), which is nonlinear in Z, requires *a priori* knowledge of the system matrices in (23). Inspired by [40], an RL-driven mitigation method without using the exact knowledge of system dynamics is developed in this section.

Remark 6: Note that the nominal controller (4) associated with the compromised outputs (10) is used as an exploring control policy to collect online data over the time interval $[0, t]$. Based on these data samples, we execute Algorithm 2 to mitigate the attacks.

V. Illustrative Numerical Examples

In this section, two illustrative numerical examples are provided to show the effectiveness of the proposed framework. The first example is an aircraft system in the face of sensor attacks. The second example is a multivehicle system under actuator attacks.

A. Example 1: Aircraft System

Consider the lateral directional dynamics of a transport aircraft system adopted from [54] given by

$$\begin{bmatrix}\dot{\alpha}(t)\\ \dot{\beta}(t)\\ \dot{\delta}_P(t)\\ \dot{\delta}_R(t)\end{bmatrix} = \begin{bmatrix}-0.037 & 0.0123 & 0.00055 & -1.0\\ 0 & 0 & 1.0 & 0\\ -6.37 & 0 & -0.23 & 0.0618\\ 1.25 & 0 & 0.016 & -0.0457\end{bmatrix}\begin{bmatrix}\alpha(t)\\ \beta(t)\\ \delta_P(t)\\ \delta_R(t)\end{bmatrix} + \begin{bmatrix}0.00084 & 0.000236\\ 0 & 0\\ 0.08 & 0.804\\ -0.0862 & -0.0665\end{bmatrix}u(t), \quad t \geq 0$$

$$y_1(t) = \beta(t), \quad C_1 = \begin{bmatrix}0 & 1 & 0 & 0\end{bmatrix}$$

$$y_2(t) = \delta_P(t), \quad C_2 = \begin{bmatrix}0 & 0 & 1 & 0\end{bmatrix}$$

$$y_3(t) = \delta_R(t), \quad C_3 = \begin{bmatrix}0 & 0 & 0 & 1\end{bmatrix}$$

where $\alpha(t)$, $\beta(t)$, $\delta_P(t)$, and $\delta_R(t)$ are the sideslip angle in deg, the roll angle in deg, the roll rate in deg/s, and the yaw rate in deg/s of the aircraft, respectively, and the control input $u(t) = [\tau_R(t), \tau_A(t)]^{\mathrm{T}}$, $t \geq 0$, involves the rudder deflection in deg and the aileron deflection in deg, respectively. Note that the system has $q = 3$ sensors with $\mathcal{S} = \{1, 2, 3\}$ and $m = 2$ actuators with $\mathcal{A} = \{1, 2\}$.

Using Lemmas 1 and 2, it follows that Assumption 1 holds with $\bar{w} = 0$ and $\bar{s} = 1$. The initial values for the state variables are randomly selected around the origin and are assumed to be unknown. With the weighting matrices set to $Q = I_4$ and $R = I_2$, the nominal output-feedback control gain is obtained by using the recursive algorithm developed in [45] is given by

$$K = \begin{bmatrix}-0.36 & -1.53 & -7.61\\ 1.27 & 3.54 & 5.06\end{bmatrix}.$$

Next, suppose that over the time interval $t = 10$ s and $t = 25$ s, an attacker has access to sensor two and launches an attack signal $a_{y,2}(t) = 1.2\cos(0.8t) + 4.2\cos(2t)\sin(\delta_P(t))$, $10 \leq t \leq 25$. In this case, $\mathcal{S}_a = \{2\}$, $\mathcal{A}_a = \varnothing$, $D_y = [0,\ 1,\ 0]^{\mathrm{T}}$, $D_u = 0_{2\times 1}$, $a_y = a_{y,2}$, and $a_u = 0$. By (11) and (12), the aircraft system is subjected to the adversarial attack $\nu(t) = [0, a_y, 0]^{\mathrm{T}}$, $10 \leq t \leq 25$. Although we do not know which sensor is attacked, there are three real-time monitors running from $t = 0$ and driven by the subsets of the sensors, namely, $\mathcal{J}_1$ using $\{y_2(t), y_3(t)\}$, $\mathcal{J}_2$ using $\{y_1(t), y_3(t)\}$, and $\mathcal{J}_3$ using $\{y_1(t), y_2(t)\}$. In this case

$$L_{\mathcal{J}_1} = \begin{bmatrix}-0.0196 & 1.2424 & 0.6878 & -0.0375\\ -0.5839 & 8.1866 & -2.6322 & 1.1285\end{bmatrix}$$

$$L_{\mathcal{J}_2} = \begin{bmatrix}0.0179 & 0.8919 & 0.0335 & 0.0171\\ -0.8129 & -4.8449 & -4.0095 & 0.9244\end{bmatrix}$$

$$L_{\mathcal{J}_3} = \begin{bmatrix}1.3843 & 0.6445 & -8.9198 & 1.4413\\ 0.1023 & 0.9999 & 1.1718 & -0.2391\end{bmatrix}.$$

Since the initial values are unknown, we set $\hat{x}(0) = \hat{x}_{\mathcal{J}_1}(0) = \hat{x}_{\mathcal{J}_2}(0) = \hat{x}_{\mathcal{J}_3}(0) = [0.0055, 0.2767, 0.1829, 0.2399]^{\mathrm{T}}$ and set $\bar{\Upsilon} = 2.2377$ and $\Xi = I_2$.

Using Algorithm 1, the threat detection levels of the three monitors are shown in Fig. 2. It can be seen from Fig. 2 that the attack-resilient set is $\mathcal{O} = \mathcal{J}_2$, which illustrates the efficacy of our attack monitoring strategy, and the attack mitigation process is triggered at $t = 10.1$ s, which reflects the sensitivity of

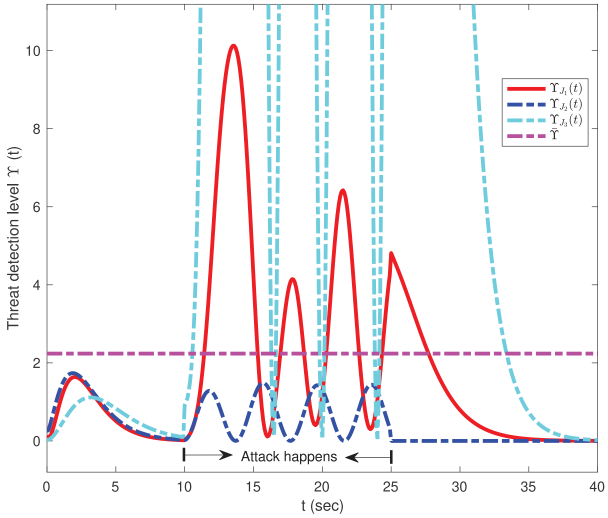


Fig. 2. Threat detection level for the three monitors under the attacks.

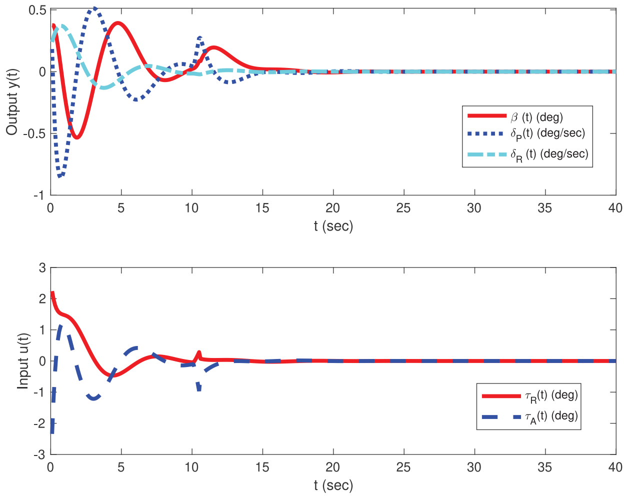


Fig. 3. System performance in the face of adversarial attacks and with Algorithm 2 engaged.

threat detection function. Then, Algorithm 2 is initialized with $\gamma = 10$ and $X_{\mathcal{O}} = I_4$. Here, for the critic given by (35) we consider second-order and fourth-order polynomials of the states and the two actors (36) and (37) are characterized by first-order and third-order polynomials of the states. With $\delta t = 0.01$ and $\varepsilon = 10^{-8}$, the learning process is executed and then the controller is updated starting at $t = 10.1$ s and continuing to the end of the simulation. The system performance in the face of sensor attacks and with Algorithm 2 engaged is depicted in Fig. 3. It can be seen from Figs. 2 and 3 that when the detector triggers an alarm, the secure control learning framework ensures that the attacked system converges to zero.

B. Example 2: Multivehicle System

Consider a network of three vehicles moving in a plane with the dynamics of each described by a single integrator as [55]

$$\dot{\zeta}_i(t) = u_i(t), \quad \zeta_i(0) = \zeta_{i0}, \quad t \geq 0, \quad i = 1, 2, 3$$

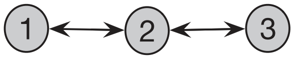


Fig. 4. Communication network for three vehicles.

where $\zeta_i \in \mathbb{R}$ is the state of the ith vehicle. To achieve asymptotic consensus, we assume that there is a nearest-neighbor interconnection topology between the three vehicles given in Fig. 4 and the signals representing the exchange of relative information have the form

$$z_i(t) = \sum_{j \in \mathcal{N}_i} \big(\zeta_i(t) - \zeta_j(t)\big), \quad i = 1, 2, 3$$

where $\mathcal{N}_1 = \{2\}$, $\mathcal{N}_2 = \{1, 3\}$, and $\mathcal{N}_3 = \{2\}$, and each vehicle is controlled by a control law with the form [56]

$$u_i = -K_i \begin{bmatrix} \zeta_i(t) \\ z_i(t) \end{bmatrix} \triangleq -\begin{bmatrix} \mathcal{K}_i & \Lambda_i \mathcal{K}_i \end{bmatrix} \begin{bmatrix} \zeta_i(t) \\ z_i(t) \end{bmatrix}$$

where $K_i \in \mathbb{R}^{1\times 2}$, $\mathcal{K}_i \in \mathbb{R}$, and $\Lambda_i \in \mathbb{R}$, $i = 1, 2, 3$.

Finally, defining $x \triangleq [\zeta_1, \zeta_2, \zeta_3]^{\mathrm{T}}$ and $u \triangleq [u_1, u_2, u_3]^{\mathrm{T}}$, the multivehicle system is represented by

$$\dot{x}(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} u(t), \quad x(0) = x_0, \quad t \geq 0$$
$$y_1(t) = x_1(t), \quad C_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$
$$y_2(t) = x_2(t), \quad C_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$
$$y_3(t) = x_3(t), \quad C_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$

where $x_0 = [\zeta_{10}, \zeta_{20}, \zeta_{30}]^{\mathrm{T}}$, with the controller (4) given by

$$u(t) = -\begin{bmatrix} \mathcal{K}_1 & \Lambda_1\mathcal{K}_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathcal{K}_2 & \Lambda_2\mathcal{K}_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathcal{K}_3 & \Lambda_3\mathcal{K}_3 \end{bmatrix} \begin{bmatrix} \zeta_1(t) \\ z_1(t) \\ \zeta_2(t) \\ z_2(t) \\ \zeta_3(t) \\ z_3(t) \end{bmatrix}$$
$$= -\begin{bmatrix} (1+\Lambda_1)\mathcal{K}_1 & -\Lambda_1\mathcal{K}_1 & 0 \\ -\Lambda_2\mathcal{K}_2 & (1+2\Lambda_2)\mathcal{K}_2 & -\Lambda_2\mathcal{K}_2 \\ 0 & -\Lambda_3\mathcal{K}_3 & (1+\Lambda_3)\mathcal{K}_3 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix}.$$

Using Lemmas 1 and 2, it follows that Assumption 1 holds with $\bar{w} = 1$ and $\bar{s} = 0$. The weighting matrices in (5) are set as $Q = I_3 \otimes Q_1 + \mathcal{L} \otimes Q_2$ and $R = I_3 \otimes R_1$, where $Q_1 = 10$, $Q_2 = 25$, $R_1 = 1$, $\otimes$ denotes the Kronecker product, and $\mathcal{L}$ is the Laplacian matrix for the network. Now, solving a standard LQR problem with performance given by (5) yields

$$\mathcal{K}_i = 3.1623, \quad \Lambda_i = 0.6736, \quad i = 1, 2, 3.$$

Next, suppose that over the time interval $t = 1$ s and $t = 2.5$ s, an attacker has access to the signal $z_1(t)$ that is transmitted to vehicle 1 and launches an attack signal $a_{u,1}(t) = -4.2\cos(10t)$, $1 \leq t \leq 2.5$, which corrupts the control signal $u_1(t)$. In this case, $\mathcal{S}_a = \varnothing$, $\mathcal{A}_a = \{1\}$, $D_y = 0_{3\times 1}$, $D_u = [1, \ 0, \ 0]^{\mathrm{T}}$, $a_y = 0$, and $a_u = a_{u,1}$. The multivehicle system is subjected to the adversarial attack $v(t)$ satisfying $\dot{v}(t) = BD_u a_u(t)$, $1 \leq t \leq 2.5$. Along with the system, there are three real-time monitors running from $t = 0$ and driven by the subsets of the sensors, namely,

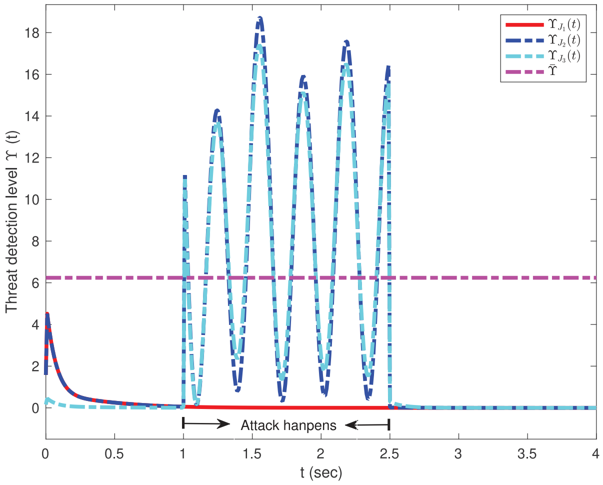


Fig. 5. Threat detection level for the three monitors under the sensor attacks.

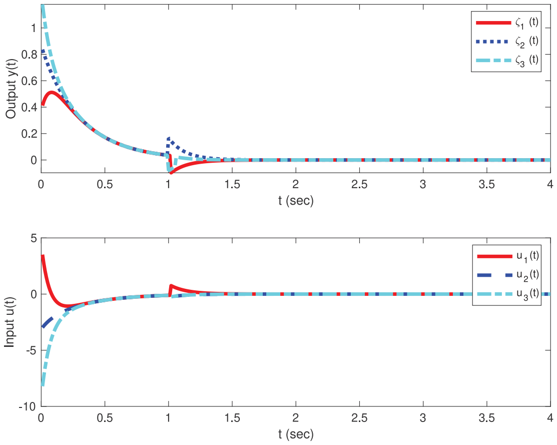


Fig. 6. System performance in the face of adversarial attacks and with Algorithm 2 engaged.

$\mathcal{J}_1$ using $\{y_2(t), y_3(t)\}$, $\mathcal{J}_2$ using $\{y_1(t), y_3(t)\}$, and $\mathcal{J}_3$ using $\{y_1(t), y_2(t))\}$. In this case, the gains $L_{\mathcal{J}_1}, L_{\mathcal{J}_2}$, and $L_{\mathcal{J}_3}$ are designed so that $A_{\mathcal{J}_1}, A_{\mathcal{J}_2}$, and $A_{\mathcal{J}_3}$ are Hurwitz and having eigenvalues $\{-1.4465, -2.3460, -5.6460\}$. Since the initial values are unknown, we set $\hat{x}(0) = \hat{x}_{\mathcal{J}_1}(0) = \hat{x}_{\mathcal{J}_2}(0) = \hat{x}_{\mathcal{J}_3}(0) = [0.4, 0.8, 1.2]^{\mathrm{T}}$ and set $\bar{\Upsilon} = 6.2393$ and $\Xi = I_2$.

Using Algorithm 1, the threat detection levels of the three monitors are shown in Fig. 5. It can be seen from Fig. 5 that the attack-resilient set is $\mathcal{O} = \mathcal{J}_1$ and the attack mitigation process is triggered at $t = 1.01$ s. Then, Algorithm 2 is initialized with $\gamma = 13$ and $X_{\mathcal{O}} = I_3$, and the same basis functions as Example 1 are used for the critic (35) and the two actors (36) and (37). The learning process is executed with $\varepsilon = 10^{-8}$ and then the controller is updated starting at $t = 1.01$ s and continuing to the end of the simulation. The system performance in the face of sensor attacks and with Algorithm 2 engaged is depicted in Fig. 6. This reflects the efficacy of the secure control learning framework.

VI. Conclusion

In this article, we developed a learning-based secure control framework for CPS in the presence of sensor and actuator attacks. The attack mitigation problem is addressed using a secure estimation approach and a game-theoretic architecture is provided for solving the underlying joint state estimation and attack mitigation problems. The implementation algorithm is based on an RL-driven attack mitigating architecture. Future research will focus on exploiting the possibility of the defender and the attacker adapting to their respective control and attack policies.

Appendix A
Proof of Theorem 1

Using Proposition (1), one has $\hat{x}(t) = \hat{x}_{\mathcal{O}}(t)$, $t \geq 0$, where $\mathcal{O}$ is given by (18). From (10), we have $\nu(t) = \tilde{y}(t) - Cx(t)$, $t \geq 0$, and thus, (19) can be obtained. Finally, we have

$$\dot{\tilde{x}}_{\mathcal{J}}(t) = A_{\mathcal{J}}\tilde{x}_{\mathcal{J}}(t) - L_{\mathcal{J}}\nu_{\mathcal{J}}(t), \quad t \geq 0 \tag{41}$$

where $\tilde{x}_{\mathcal{J}}(0) = \tilde{x}(0)$.

Now, note that $A_{\mathcal{J}}$ is a Hurwitz matrix and all the observers are initialized at common initial condition $\hat{x}(0)$. Thus, using [10], there exist positive constants $\kappa > 0$ and $\alpha > 0$ such that the solution to (41) satisfies

$$\|x(t) - \hat{x}(t)\| \leq \kappa e^{-\alpha t}\|\tilde{x}(0)\|, \quad t \geq 0.$$

Finally, defining $e_\nu(t) = \hat{\nu}(t) - \nu(t), t \geq 0$, and using (2) and (19), it follows that:

$$\begin{aligned}\|e_\nu(t)\| &= \|\tilde{y}(t) - C\hat{x}(t) - \nu(t)\| \\ &= \|y(t) - C\hat{x}(t)\| \\ &\leq \|x(t) - C\hat{x}(t)\| \\ &\leq \kappa e^{-\alpha t}\|C\|\|\tilde{x}(0)\|, \quad t \geq 0.\end{aligned}$$

Thus, we obtain the asymptotic convergence of the estimated attack $\hat{\nu}(t)$ to the real attack signal $\nu(t), t \geq 0$. This completes the proof.

Appendix B
Proof of Lemma 4

First, given u^i and ν^i, we solve (31)–(33) for V^i, u^{i+1}, and ν^{i+1}. Since dynamics (23) are linear, we define $V^i = \xi^{\mathrm{T}}P^i\xi$ and using (28), we obtain

$$\begin{aligned}T(P^i) = \mathcal{A}^{\mathrm{T}}P^i + P^i\mathcal{A} + \mathcal{Q} - P^i\mathcal{B}R^{-1}\mathcal{B}^{\mathrm{T}}P^i \\ + \gamma^{-2}P^i\mathcal{D}\mathcal{D}^{\mathrm{T}}P^i = 0\end{aligned} \tag{42}$$

with $T(Z) = 0$. Now, let $\mathcal{T} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$

$$\mathcal{T}(P) = P - \left(T'(P)\right)^{-1}T(P). \tag{43}$$

Then, (31) and (32) are equivalent to the Newton iteration [52]

$$P^{i+1} = \mathcal{T}\left(P^i\right) \tag{44}$$

with $u^{i+1} = -(1/2)R^{-1}\mathcal{B}^{\mathrm{T}}P^{i+1}\xi$, $\nu^{i+1} = (1/2)\gamma^{-2}\mathcal{D}^{\mathrm{T}}P^{i+1}\xi$ [40].

Next, we let $c_0 \geq \|(T'(P^0))^{-1}\|$, $\eta \geq \|(T'(P^0))^{-1}T(P^0)\|$ and define $h = c_0K\eta \leq (1/2)$, where K satisfies $\|T'(P_1) - T'(P_2)\| \leq K\|P_1 - P_2\|$. Using Kantorovtich's theorem [53] and the result from [52], we obtain

$$\left|V^i - V^\star\right| \leq \frac{\eta}{h}\frac{\left(1 - \sqrt{1-2h}\right)^{2^i}}{2^i} \leq \frac{\eta}{h}\frac{1}{2^i} \tag{45}$$

for $i = 0, 1, 2, \ldots$ This implies that the value function in (34) converges to (25) at the rate of $O(1/2^i)$. It follows that the updated control and attack sequences u^{i+1} and ν^{i+1} converge to the optimal control and attack policies $u^\star$ and $\nu^\star$ given by (27) at the same rate. Furthermore, it follows from Lemma 3 that, with the optimal control and attack policies (27), the closed-loop system (23) has an asymptotically stable equilibrium point. This completes the proof.

Appendix C
Proof of Theorem 2

Let $\hat{u}^i$ and $\hat{\nu}^i$ be given and let $\tilde{V}^i(\xi)$ be the solution to

$$0 = \xi^{\mathrm{T}} \mathcal{Q} \xi + \hat{u}^{i\mathrm{T}} R \hat{u}^i - \gamma^2 \hat{\nu}^{i\mathrm{T}} \hat{\nu}^i + \tilde{V}^i_\xi \big(\mathcal{A}\xi + \mathcal{B}\hat{u}^i + \mathcal{D}\hat{\nu}^i \big) \quad (46)$$

with $\tilde{V}^i(0) = 0$. Furthermore, let

$$\tilde{u}^{i+1}(\xi) = -\frac{1}{2} R^{-1} \mathcal{B}^{\mathrm{T}} \tilde{V}^{i\mathrm{T}}_\xi \quad (47)$$

$$\tilde{\nu}^{i+1}(\xi) = \frac{1}{2\gamma^2} \mathcal{D}^{\mathrm{T}} \tilde{V}^{i\mathrm{T}}_\xi . \quad (48)$$

The following lemma is needed to complete the proof.

Lemma 5: For every $i > 0$ and all $\xi \in \Omega$

$$\lim_{l_1, l_2, l_3 \to \infty} \hat{V}^i(\xi) = \tilde{V}^i(\xi)$$
$$\lim_{l_1, l_2, l_3 \to \infty} \hat{u}^{i+1}(\xi) = \tilde{u}^{i+1}(\xi)$$
$$\lim_{l_1, l_2, l_3 \to \infty} \hat{\nu}^{i+1}(\xi) = \tilde{\nu}^{i+1}(\xi).$$

Proof: Using (46)–(48), it follows that:

$$\begin{aligned} &\tilde{V}^i(\xi(t+\delta t)) - \tilde{V}^i(\xi(t)) \\ &\quad = \int_t^{t+\delta t} \Big[-\xi^{\mathrm{T}}(\tau) \mathcal{Q} \xi(\tau) - \hat{u}^{i\mathrm{T}}(\tau) R \hat{u}^i(\tau) + \gamma^2 \hat{\nu}^{i\mathrm{T}}(\tau) \hat{\nu}^i(\tau) \Big] \mathrm{d}\tau \\ &\qquad - 2 \int_t^{t+\delta t} \Big[\tilde{u}^{(i+1)\mathrm{T}}(\tau) R \hat{\zeta}^i(\tau) - \gamma^2 \tilde{\nu}^{(i+1)\mathrm{T}}(\tau) \hat{\varsigma}^i(\tau) \Big] \mathrm{d}\tau. (49) \end{aligned}$$

Next, we expand the three sets of basis functions in (35)–(37). As shown in [35] and [51], given an arbitrary $\epsilon > 0$, there exists $l_{10} > 0$ such that for $l_1^\star > \max\{l_1, l_{10}\}$

$$\left| \tilde{V}^i(\xi(t)) - \tilde{W}_1^{i\mathrm{T}} \bar{\phi}(\xi(t)) \right| \le \frac{\epsilon}{2} \quad (50)$$

where $\tilde{W}_1^i = [\hat{W}_1^{i\mathrm{T}}, \tilde{W}_{j_1}^{i\mathrm{T}}]^{\mathrm{T}}$ denotes a constant weight vector, $\bar{\phi}(\xi) = [\phi_1(\xi) \ , \ldots , \ \phi_{l_1}(\xi), \bar{\phi}_{l_1+1}(\xi) \ , \ldots , \ \bar{\phi}_{l_1^\star}(\xi)]^{\mathrm{T}}$ denotes a linearly independent and continuous function vector, $\tilde{W}_{j_1}^i \in \mathbb{R}^{l_1^\star - l_1 + 1}$ is an expanded weighting vector, and $\{\bar{\phi}_j(\cdot)\}_{j=l_1+1}^{l_1^\star}$ denote an expanded basis functions for $\{\phi_j(\cdot)\}_{j=1}^{l_1}$.

Analogously, there exists $l_{20} > 0$ such that for $l_2^\star > \max\{l_2, l_{20}\}$

$$\left\| \tilde{u}^{i+1}(\xi(t)) - \tilde{W}_2^{i\mathrm{T}} \bar{\varphi}(\xi(t)) \right\| \le \frac{\epsilon}{2} \quad (51)$$

where $\tilde{W}_2^i = [\tilde{W}_{2,1}^i, \ldots, \tilde{W}_{2,m}^i]$ denotes a constant weight matrix with $\tilde{W}_{2,j}^i = [\hat{W}_{2,j}^{i\mathrm{T}}, \tilde{W}_{2,j_{l_2}}^{i\mathrm{T}}]^{\mathrm{T}}$, $j = 1 \ , \ldots , \ m$, $\bar{\varphi}(\xi) = [\varphi_1(\xi) \ , \ldots , \ \varphi_{l_2}(\xi), \bar{\varphi}_{l_2+1}(\xi) \ , \ldots , \ \bar{\varphi}_{l_2^\star}(\xi)]^{\mathrm{T}}$ denotes a linearly independent and continuous function vector, $\tilde{W}_{2,j_{l_2}}^i \in \mathbb{R}^{(l_2^\star - l_2 + 1) \times m}$ is an expanded weighting matrix, and $\{\bar{\varphi}_j(\cdot)\}_{j=l_2+1}^{l_2^\star}$ denote an expanded basis functions for $\{\varphi_j(\cdot)\}_{j=1}^{l_2}$.

Finally, there exists $l_{30} > 0$ such that for $l_3^\star > \max\{l_3, l_{30}\}$

$$\left\| \tilde{\nu}^{i+1}(\xi(t)) - \tilde{W}_3^{i\mathrm{T}} \bar{\psi}(\xi(t)) \right\| \le \frac{\epsilon}{2} \quad (52)$$

where $\tilde{W}_3^i = [\tilde{W}_{3,1}^i, \ldots, \tilde{W}_{3,q}^i]$ denotes a constant weight matrix with $\tilde{W}_{3,j}^i = [\hat{W}_{3,j}^{i\mathrm{T}}, \tilde{W}_{3,j_{l_3}}^{i\mathrm{T}}]^{\mathrm{T}}$, $j = 1, \ldots, q$, $\bar{\psi}(\xi) = [\psi_1(\xi), \ldots, \psi_{l_3}(\xi), \bar{\psi}_{l_3+1}(\xi), \ldots, \bar{\psi}_{l_3^\star}(\xi)]^{\mathrm{T}}$ denotes a linearly independent and continuous function vector, $\tilde{W}_{3,j_{l_3}}^i \in \mathbb{R}^{(l_3^\star - l_3 + 1) \times q}$ is an expanded weighting matrix, and $\{\bar{\psi}_j(\cdot)\}_{j=l_3+1}^{l_3^\star}$ denote an expanded basis functions for $\{\psi_j(\cdot)\}_{j=1}^{l_3}$.

Next, define $\Upsilon_\phi(\xi) \triangleq [\bar{\phi}_{l_1+1}(\xi), \ldots, \bar{\phi}_{l_1^\star}(\xi)]^{\mathrm{T}}$, $\Upsilon_\varphi(\xi) \triangleq [\bar{\varphi}_{l_2+1}(\xi), \ldots, \bar{\varphi}_{l_2^\star}(\xi)]^{\mathrm{T}}$, and $\Upsilon_\psi(\xi) \triangleq [\bar{\psi}_{l_2+1}(\xi), \ldots, \bar{\psi}_{l_2^\star}(\xi)]^{\mathrm{T}}$ and note that

$$\tilde{W}_1^{i\mathrm{T}} \bar{\phi}(\xi(t)) = \hat{W}_1^{i\mathrm{T}} \phi(\xi(t)) + \tilde{W}_{j_1}^i \Upsilon_\phi(\xi(t))$$
$$\tilde{W}_2^{i\mathrm{T}} \bar{\varphi}(\xi(t)) = \hat{W}_2^{i\mathrm{T}} \varphi(\xi(t)) + \tilde{W}_{j_2}^i \Upsilon_\varphi(\xi(t))$$
$$\tilde{W}_3^{i\mathrm{T}} \bar{\psi}(\xi(t)) = \hat{W}_3^{i\mathrm{T}} \psi(\xi(t) + \tilde{W}_{j_3}^i \Upsilon_\psi(\xi(t))$$

where $\tilde{W}_{j_2}^i = [\hat{W}_{2,1_{l_2}}^{i\mathrm{T}}, \ldots, \tilde{W}_{2,m_{l_2}}^{i\mathrm{T}}]^{\mathrm{T}}$ and $\tilde{W}_{j_3}^i = [\hat{W}_{3,1_{l_2}}^{i\mathrm{T}}, \ldots, \tilde{W}_{3,q_{l_3}}^{i\mathrm{T}}]^{\mathrm{T}}$. Now, using (35),–(37), and subtracting (38) from (49), it follows that, for all samples t_j

$$\begin{aligned} \mu^i(t_j) = &-\tilde{W}_{j_1}^i \big[\Upsilon_\phi(\xi(t_j + \delta t)) - \Upsilon_\phi(\xi(t_j)) \big] \\ &- 2 \sum_{k=1}^{m} r_k \int_{t_j}^{t_j+\delta t} \tilde{W}_{j_2}^i \Upsilon_\varphi(\xi(\tau)) \hat{\zeta}_k^i \mathrm{d}\tau \\ &+ 2\gamma^2 \sum_{l=1}^{r} \int_{t_j}^{t_j+\delta t} \tilde{W}_{j_3}^i \Upsilon_\psi(\xi(\tau)) \hat{\varsigma}_l^i \mathrm{d}\tau \end{aligned} \quad (53)$$

where $j = 1, \ldots, N$.

Next, the weights $\hat{W}^i$ in (39) are obtained using a least-square method, and by Assumption 3, it follows that:

$$\sum_{j=1}^{N} \mu^{i\mathrm{T}}(t_j) \mu^i(t_j) \le \frac{\epsilon}{N(4 + 2m\Xi_2 + 2q\Xi_3)} \quad (54)$$

where $\Xi_2 = 2\delta t \max_{k=1,\ldots,m}\{r_k \hat{\zeta}_k^i\} > 0$ and $\Xi_3 = 2\gamma^2 \delta t \max_{l=1,\ldots,q}\{\hat{\varsigma}_l^i\} > 0$. Thus, by continuity of the expanded basis function $\Upsilon_\phi(\xi)$, it follows that, for all $\xi(t_k) \in \Omega$, $k = 1, 2, \ldots, N$, and large enough N

$$\begin{aligned} \left| \tilde{W}_{j_1}^i \Upsilon_\phi(\xi(t)) \right| &\le \max_{k=1,\ldots,N} \left| \tilde{W}_{j_1}^i \Upsilon_\phi(\xi(t_k)) \right| \\ &\le \frac{2\epsilon}{(4 + 2m\Xi_2 + 2q\Xi_3)} \le \frac{\epsilon}{2}. \end{aligned} \quad (55)$$

Hence, for $l_1^\star > \max\{l_1, l_{10}\}$ and $\xi \in \Omega$

$$\begin{aligned} \left| \hat{V}^i(\xi) - \tilde{V}^i(\xi) \right| &\le \left| \tilde{V}^i(\xi) - \tilde{W}_1^{i\mathrm{T}} \bar{\phi}(\xi) \right| + \left| \tilde{W}_{j_1}^i \Upsilon_\phi(\xi) \right| \\ &\le \frac{\epsilon}{2} + \frac{\epsilon}{2} \le \epsilon. \end{aligned} \quad (56)$$

Similarly, it can be shown that for all $\xi \in \Omega$

$$\begin{aligned}\left\|\hat{u}^{i+1}(\xi) - \tilde{u}^{i+1}(\xi)\right\| &\le \left\|\tilde{u}^{i+1}(\xi) - \tilde{W}_2^{i\mathrm{T}}\bar{\varphi}(\xi)\right\| + \left\|\tilde{W}_{j2}^{i}\Upsilon_{\varphi}(\xi)\right\| \\ &\le \frac{\epsilon}{2} + \frac{\epsilon}{2} \le \epsilon \qquad (57)\\ \left\|\hat{\nu}^{i+1}(\xi) - \tilde{\nu}^{i+1}(\xi)\right\| &\le \left\|\tilde{\nu}^{i+1}(\xi) - \tilde{W}_3^{i\mathrm{T}}\bar{\psi}(\xi)\right\| + \left\|\tilde{W}_{j3}^{i}\Upsilon_{\psi}(\xi)\right\| \\ &\le \frac{\epsilon}{2} + \frac{\epsilon}{2} \le \epsilon \qquad (58)\end{aligned}$$

which proves the lemma. ■

Now, the proof of Theorem 2 follows by mathematical induction. Specifically:

1) For $i = 0$, we have $\tilde{V}^0(\xi) = V(\xi)$, and hence, it follows that $\tilde{u}^1(\xi) = u^1(\xi)$ and $\tilde{\nu}^1(\xi) = \nu^1(\xi)$. Now, convergence follows from Lemmas 4 and 5.
2) Suppose that for some $i > 0$ and all $\xi \in \Omega$, $\lim_{l_1,l_2,l_3\to\infty} \hat{V}^{i-1}(\xi) = V^{i-1}(\xi)$, $\lim_{l_1,l_2,l_3\to\infty} \hat{u}^i(\xi) = u^i(\xi)$, and $\lim_{l_1,l_2,l_3\to\infty} \hat{\nu}^i(\xi) = \nu^i(\xi)$.

Then, given $\hat{V}^{i-1}(\xi)$, $\hat{u}^i(\xi)$, and $\hat{\nu}^i(\xi)$, consider the relationship between $\hat{V}^i(\xi)$ and $V^i(\xi)$, the relationship between $\hat{u}^{i+1}(\xi)$ and $u^{i+1}(\xi)$, and the relationship between $\hat{\nu}^{i+1}(\xi)$ and $\nu^{i+1}(\xi)$. First, using $\hat{u}^i(\xi)$ and $\hat{\nu}^i(\xi)$ to solve (46), $\tilde{V}^i(\xi)$ is obtained, and using (34) and (49), it follows that:

$$\begin{aligned}&\left|V^i(\xi) - \tilde{V}^i(\xi)\right| \\ &\quad\le \left|\int_t^{\infty}\left(u^{i\mathrm{T}}(\xi)Ru^i(\xi) - \hat{u}^{i\mathrm{T}}(\xi)R\hat{u}^i(\xi)\right)\mathrm{d}\tau\right| \\ &\qquad+\ \gamma^2\left|\int_t^{\infty}\left(\nu^{i\mathrm{T}}(\xi)R\nu^i(\xi) - \hat{\nu}^{i\mathrm{T}}(\xi)R\hat{\nu}^i(\xi)\right)\mathrm{d}\tau\right| \\ &\qquad+\ 2\left|\int_t^{\infty}\left(u^{(i+1)\mathrm{T}}(\xi)R\zeta^i - \tilde{u}^{(i+1)\mathrm{T}}(\xi)R\hat{\zeta}^i\right)\mathrm{d}\tau\right| \\ &\qquad+\ 2\gamma^2\left|\int_t^{\infty}\left(\nu^{(i+1)\mathrm{T}}(\xi)\varsigma^i - \tilde{\nu}^{(i+1)\mathrm{T}}(\xi)\hat{\varsigma}^i\right)\mathrm{d}\tau\right|. \qquad (59)\end{aligned}$$

Next, using induction, we have

$$\begin{aligned}\lim_{l_1,l_2,l_3\to\infty}\left|\int_t^{\infty}\left(u^{i\mathrm{T}}(\xi)Ru^i(\xi) - \hat{u}^{i\mathrm{T}}(\xi)R\hat{u}^i(\xi)\right)\mathrm{d}\tau\right| &= 0\\ \lim_{l_1,l_2,l_3\to\infty}\left|\int_t^{\infty}\left(\nu^{i\mathrm{T}}(\xi)R\nu^i(\xi) - \hat{\nu}^{i\mathrm{T}}(\xi)R\hat{\nu}^i(\xi)\right)\mathrm{d}\tau\right| &= 0.\end{aligned}$$

Now, since $\lim_{l_1,l_2,l_3\to\infty} \hat{u}^i(\xi) = u^i(\xi)$ and $\lim_{l_1,l_2,l_3\to\infty} \hat{\nu}^i(\xi) = \nu^i(\xi)$, and using (31), (32), (47), and (48), it follows that for all $\xi \in \Omega$, $\lim_{l_1,l_2,l_3\to\infty} \|u^{i+1}(\xi) - \tilde{u}^{i+1}(\xi)\| = 0$, $\lim_{l_1,l_2,l_3\to\infty} \|\nu^{i+1}(\xi) - \tilde{\nu}^{i+1}(\xi)\| = 0$. Finally, we obtain

$$\lim_{l_1,l_2,l_3\to\infty}\left|V^i(\xi) - \tilde{V}^i(\xi)\right| = 0. \qquad (60)$$

In addition, since

$$\left|\hat{V}^i(\xi) - V^i(\xi)\right| \le \left|\hat{V}^i(\xi) - \tilde{V}^i(\xi)\right| + \left|V^i(\xi) - \tilde{V}^i(\xi)\right| \qquad (61)$$

and, by Lemma 5, $\lim_{l_1,l_2,l_3\to\infty} |\hat{V}^i(\xi) - \tilde{V}^i(\xi)| = 0$, it follows that:

$$\lim_{l_1,l_2,l_3\to\infty} \hat{V}^i(\xi) = V^i(\xi)$$

Now, it follows from Lemma 4 that $\lim_{l_1,l_2,l_3\to\infty} V^i(\xi) = V^{\star}(\xi)$, which verifies the first result of Lemma 5.

Next, using similar arguments used to prove $\lim_{l_1,l_2,l_3\to\infty} \hat{V}^i(\xi) = V^i(\xi)$ and Lemma 5, and using mathematical induction, it follows that:

$$\begin{aligned}\lim_{l_1,l_2,l_3\to\infty} \hat{u}^{i+1}(\xi) &= u^{i+1}(\xi)\\ \lim_{l_1,l_2,l_3\to\infty} \hat{\nu}^{i+1}(\xi) &= \nu^{i+1}(\xi).\end{aligned}$$

Then, using Lemma 4, it follows that $\lim_{l_1,l_2,l_3\to\infty} u^{i+1}(\xi) = u^{\star}(\xi)$, $\lim_{l_1,l_2,l_3\to\infty} \nu^{i+1}(\xi) = \nu^{\star}(\xi)$. Thus, we obtain that for all $\xi \in \Omega$, $\lim_{l_1,l_2,l_3\to\infty} \hat{u}^{i+1}(\xi) = u^{\star}(\xi)$ and $\lim_{l_1,l_2,l_3\to\infty} \hat{\nu}^{i+1}(\xi) = \nu^{\star}(\xi)$. This completes the proof.

References

[1] F. Pasqualetti, F. Dorfler, and F. Bullo, "Control-theoretic methods for cyber-physical security: Geometric principles for optimal cross-layer resilient control systems," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 110–127, Feb. 2015.

[2] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 46–65, Feb. 2015.

[3] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 24–45, Feb. 2015.

[4] D. Senejohnny, P. Tesi, and C. De Persis, "A jamming-resilient algorithm for self-triggered network coordination," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 981–990, Sep. 2018.

[5] S. Amin, G. A. Schwartz, and S. S. Sastry, "Security of interdependent and identical networked control systems," *Automatica*, vol. 49, no. 1, pp. 186–192, 2013.

[6] S. Feng and P. Tesi, "Resilient control under denial-of-service: Robust design," *Automatica*, vol. 79, pp. 42–51, May 2017.

[7] Y. Chen, S. Kar, and J. M. F. Moura, "Dynamic attack detection in cyber-physical systems with side initial state information," *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4618–4624, Sep. 2017.

[8] X. Jin, W. M. Haddad, and T. Yucelen, "An adaptive control architecture for mitigating sensor and actuator attacks in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 6058–6064, Nov. 2017.

[9] L. An and G. H. Yang, "Secure state estimation against sparse sensor attacks with adaptive switching mechanism," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2596–2603, Aug. 2018.

[10] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *Proc. Amer. Control Conf.*, 2015, pp. 2439–2444.

[11] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 4917–4932, Oct. 2017.

[12] C. De Persis and P. Tesi, "Input-to-state stabilizing control under denial-of-service," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 2930–2944, Nov. 2015.

[13] C. Z. Bai, F. Pasqualetti, and V. Gupta, "Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs," *Automatica*, vol. 82, pp. 251–260, Aug. 2017.

[14] K. G. Vamvoudakis and J. P. Hespanha, "Cooperative *Q*-learning for rejection of persistent adversarial inputs in networked linear quadratic systems," *IEEE Trans. Autom. Control*, vol. 63, no. 4, pp. 1018–1031, Apr. 2018.

[15] H. Li, Z. Chen, L. Wu, H. K. Lam, and H. Du, "Event-triggered fault detection of nonlinear networked systems," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1041–1052, Apr. 2017.

[16] E. Mousavinejad, F. Yang, Q. L. Han, and L. Vlacic, "A novel cyber attack detection method in networked control systems," *IEEE Trans. Cybern.*, vol. 48, no. 11, pp. 3254–3264, Nov. 2018.

[17] Y. Zhou, V. Lakamraju, I. Koren, and C. M. Krishna, "Software-based failure detection and recovery in programmable network interfaces," *IEEE Trans. Parallel Distrib. Syst.*, vol. 18, no. 11, pp. 1539–1550, Nov. 2007.

[18] G. Wu, J. Sun, and J. Chen, "Optimal data injection attacks in cyber-physical systems," *IEEE Trans. Cybern.*, vol. 48, no. 12, pp. 3302–3312, Dec. 2018.

[19] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[20] K. G. Vamvoudakis, J. P. Hespanha, B. Sinopoli, and Y. Mo, "Detection in adversarial environments," *IEEE Trans. Autom. Control*, vol. 59, no. 12, pp. 3209–3223, Dec. 2014.

[21] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1396–1407, Jul. 2014.

[22] Y. Mo, J. P. Hespanha, and B. Sinopoli, "Resilient detection in the presence of integrity attacks," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 31–43, Jan. 2015.

[23] M. Zhu and S. Martinez, "On the performance analysis of resilient networked control systems under replay attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 804–808, Mar. 2014.

[24] Y. Yuan, Q. Zhu, F. Sun, Q. Wang, and T. Basar, "Resilient control of cyber-physical systems against denial-of-service attacks," in *Proc. 6th Int. Symp. Resilient Control Syst. (ISRCS)*, 2013, pp. 54–59.

[25] D. Ding, Z. Wang, D. W. Ho, and G. Wei, "Observer-based event-triggering consensus control for multiagent systems with lossy sensors and cyber-attacks," *IEEE Trans. Cybern.*, vol. 47, no. 8, pp. 1936–1947, Aug. 2017.

[26] F. Li and Y. Tang, "False data injection attack for cyber-physical systems with resource constraint," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 729–738, Feb. 2020.

[27] A. Kanellopoulos and K. G. Vamvoudakis, "A moving target defense control framework for cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 1029–1043, Mar. 2020.

[28] A. Mandlekar, Y. Zhu, A. Garg, L. Fei-Fei, and S. Savarese, "Adversarially robust policy learning: Active construction of physically-plausible perturbations," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 3932–3939.

[29] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against machine learning," in *Proc. ACM Asia Conf. Comput. Commun. Security*, 2017, pp. 506–519.

[30] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1773–1786, Aug. 2016.

[31] S. Dua and X. Du, *Data Mining and Machine Learning in Cybersecurity*, Boca Raton, FL, USA: CRC, 2016.

[32] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, "Optimal and autonomous control using reinforcement learning: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2042–2062, Jun. 2018.

[33] Y. Jiang and Z. P. Jiang, *Robust Adaptive Dynamic Programming*. Hoboken, NJ, USA: Wiley, 2017.

[34] T. Bian, Y. Jiang, and Z. P. Jiang, "Adaptive dynamic programming and optimal control of nonlinear nonaffine systems," *Automatica*, vol. 50, no. 10, pp. 2624–2632, 2014.

[35] 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 Syst. Mag.*, vol. 32, no. 6, pp. 76–105, Dec. 2012.

[36] A. Fielder, E. Panaousis, P. Malacaria, C. Hankin, and F. Smeraldi, "Game theory meets information security management," in *Proc. IFIP Int. Inf. Security Conf.*, 2014, pp. 15–29.

[37] Y. Han, T. Alpcan, J. Chan, C. Leckie, and B. I. Rubinstein, "A game theoretical approach to defend against co-resident attacks in cloud computing: Preventing co-residence using semi-supervised learning," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 3, pp. 556–570, Mar. 2016.

[38] K. G. Vamvoudakis, F. L. Lewis, and G. R. Hudas, "Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality," *Automatica*, vol. 48, no. 8, pp. 1598–1611, 2012.

[39] W. Gao, Y. Jiang, Z. P. Jiang, and T. Chai, "Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming," *Automatica*, vol. 72, pp. 37–45, Oct. 2016.

[40] H. Modares, F. L. Lewis, and Z. P. Jiang, "H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2550–2562, Oct. 2015.

[41] L. An and G. H. Yang, "LQ secure control for cyber-physical systems against sparse sensor and actuator attacks," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 2, pp. 833–841, Jun. 2019.

[42] Y. Zhou, K. G. Vamvoudakis, W. M. Haddad, and Z. P. Jiang, "A secure control learning framework for cyber-physical systems under sensor attacks," in *Proc. Amer. Control Conf.*, Philadelphia, PA, USA, Jul. 2019, pp. 4280–4285.

[43] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. Hoboken, NJ, USA: Wiley, 2012.

[44] J. B. Hoagg, W. M. Haddad, and D. S. Bernstein, *Linear-Quadratic Control: Theory and Methods for Continuous-Time Systems*. Princeton, NJ, USA: Princeton Univ. Press.

[45] W. Levine and M. Athans, "On the determination of the optimal constant output feedback gains for linear multivariable systems," *IEEE Trans. Autom. Control*, vol. 15, no. 1, pp. 44–48, Feb. 1970.

[46] W. M. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. Princeton, NJ, USA: Princeton Univ. Press, 2008.

[47] J. P. Hespanha, *Linear Systems Theory*. Princeton, NJ, USA: Princeton Univ. Press. 2009.

[48] J. M. Hendrickx, K. H. Johansson, R. M. Jungers, H. Sandberg, and K. C. Sou, "Efficient computations of a security index for false data attacks in power networks," *IEEE Trans. Autom. Control*, vol. 59, no. 12, pp. 3194–3208, Dec. 2014.

[49] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[50] T. Basar and P. Bernhard, *H-Infinity Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. New York, NY, USA: Springer, 2008.

[51] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.

[52] H. N. Wu and B. Luo, "Neural network based online simultaneous policy update algorithm for solving the HJI equation in nonlinear H_∞ control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 12, pp, pp. 1884–1895, Dec. 2012.

[53] R. A. Tapia, "The Kantorovich theorem for Newton's method," *Amer. Math. Month.*, vol. 78, no. 4, pp. 389–392, 1971.

[54] S. Choi and H. Sirisena, "Computation of optimal output feedback gains for linear multivariable systems," *IEEE Trans. Autom. Control*, vol. 19, no. 3, pp. 257–258, Jun. 1974.

[55] W. Ren and R. W. Beard, *Distributed Consensus in Multi-Vehicle Cooperative Control*. London, U.K.: Springer, 2008.

[56] P. Deshpande, P. P. Menon, C. Edwards, and I. Postlethwaite, "A distributed control law with guaranteed LQR cost for identical dynamically coupled linear systems," in *Proc. Amer. Control Conf.*, 2011, pp. 5342–5347.

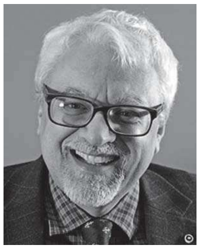
Yuanqiang Zhou (Student Member, IEEE) received the B.S. degree in mathematics and applied mathematics and the M.S. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree in control science and engineering with the Department of Automation, Shanghai Jiao Tong University, Shanghai, China.

From 2017 to 2019, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, New York University, Brooklyn, NY, USA. His research interests include cyber-physical security, networked control systems, distributed optimization, and model predictive control.

Kyriakos G. Vamvoudakis (Senior Member, IEEE) was born in Athens, Greece. He received the Diploma (five-year degree, equivalent to a Master of Science) (Highest Hons.) in electronic and computer engineering from the Technical University of Crete, Chania, Greece, in 2006, and the M.S. and Ph.D. degrees in electrical engineering from the University of Texas at Arlington, Arlington, TX, USA, in 2008 and 2011, respectively, under the guidance of F. L. Lewis.

From 2011 to 2012, he was working as an Adjunct Professor and the Faculty Research Associate with the University of Texas at Arlington and with Automation and Robotics Research Institute. From 2012 to 2016, he was a Project Research Scientist with the Center for Control, Dynamical Systems and Computation, University of California at Santa Barbara, Santa Barbara, CA, USA. He was an Assistant Professor with the Kevin T. Crofton Department of Aerospace and Ocean Engineering, Virginia Tech, Blacksburg, VA, USA, until 2018. He currently serves as an Assistant Professor with the Daniel Guggenheim School of Aerospace Engineering, Georgia Tech, Atlanta, GA, USA. He holds a secondary appointment with the School of Electrical and Computer Engineering. His research interests include reinforcement learning, control theory, cyber-physical security, bounded rationality, and safe/assured autonomy.

Dr. Vamvoudakis is a recipient of the ARO YIP Award in 2019, the NSF CAREER Award in 2018, and of several international awards, including the International Neural Network Society Young Investigator Award in 2016, the Best Paper Award for Autonomous/Unmanned Vehicles at the 27th Army Science Conference in 2010, the Best Presentation Award at the World Congress of Computational Intelligence in 2010, and the Best Researcher Award from the Automation and Robotics Research Institute in 2011. He is listed in Who's Who in the World, Who's Who in Science and Engineering, and Who's Who in America. He has also served on various international program committees and has organized special sessions, workshops, and tutorials for several international conferences. He is a member of Tau Beta Pi, Eta Kappa Nu, and Golden Key honor societies. He is currently a member of the Technical Committee on Intelligent Control of the IEEE Control Systems Society; the Technical Committee on Adaptive Dynamic Programming and Reinforcement Learning of the IEEE Computational Intelligence Society; and the IEEE Control Systems Society Conference Editorial Board, and an Associate Editor of *Automatica*; *IEEE Computational Intelligence Magazine*; the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS; *Neurocomputing*; the *Journal of Optimization Theory and Applications*; IEEE CONTROL SYSTEMS LETTERS; *Frontiers in Control Engineering-Adaptive*; and *Robust and Fault Tolerant Control*. He is also a registered Electrical/Computer Engineer (PE), a member of the Technical Chamber of Greece, and a Senior Member of AIAA.



Wassim M. Haddad (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Florida Tech, Melbourne, FL, USA, in 1983, 1984, and 1987, respectively.

Since 1994, he has been with the School of Aerospace Engineering, Georgia Tech, Atlanta, GA, USA, where he holds the rank of a Professor, the David Lewis Chair of dynamical systems and control, and the Chair of flight mechanics and control discipline. He also holds a joint Professor appointment with the School of Electrical and Computer Engineering, Georgia Tech. He is the Co-Founder, the Chairman of the Board, and the Chief Scientific Advisor of Autonomous Healthcare, Inc., Hoboken, NJ, USA. He has made numerous contributions to the development of nonlinear control theory and its application to aerospace, electrical, and biomedical engineering. His transdisciplinary research in systems and control is documented in over 650 archival journal and conference publications, and eight books in the areas of science, mathematics, medicine, and engineering.

Dr. Haddad was a recipient of the AIAA Pendray Aerospace Literature Award in 2014. He is an NSF Presidential Faculty Fellow and a member of the Academy of Nonlinear Sciences.

Zhong-Ping Jiang (Fellow, IEEE) received the M.Sc. degree in statistics from the University of Paris XI, Orsay, France, in 1989, and the Ph.D. degree in automatic control and mathematics from the École des Mines de Paris (currently, ParisTech-Mines), Paris, France, in 1993, under the direction of Prof. L. Praly.

He is currently a Professor of electrical and computer engineering with the Tandon School of Engineering, New York University, Brooklyn, NY, USA. His main research interests include stability theory, robust/adaptive/distributed nonlinear control, robust adaptive dynamic programming, learning-based control and their applications to information, mechanical, and biological systems. He has written five books and has authored/coauthored over 450 peer-reviewed journal and conference papers in the above areas.

Prof. Jiang has served as the deputy editor-in-chief, a senior editor, and an associate editor for numerous journals. He is a fellow of IFAC and CAA and is among the Clarivate Analytics Highly Cited Researchers.