# Adaptive learning in two-player Stackelberg games with continuous action sets

Guosong Yang, Radha Poovendran, and João P. Hespanha

*Abstract*— **We study a two-player Stackelberg game in which the follower's strategy depends on a parameter vector that is unknown to the leader. An adaptive learning algorithm is designed to simultaneously estimate the unknown parameter and minimize the leader's cost, based on adaptive control techniques and hysteresis switching. The algorithm guarantees that the leader's cost predicted using the parameter estimate becomes indistinguishable from its actual cost in finite time, up to a preselected, arbitrarily small error threshold, and that the first-order necessary condition for optimality holds asymptotically for the predicted cost. If an additional persistent excitation condition holds, then the parameter estimation error can also be bounded by a preselected, arbitrarily small threshold in finite time. The algorithm and convergence results are illustrated via a simple simulation example in the domain of network security.**

## I. Introduction

A modern engineering system often involves multiple self-interested decision makers whose actions have mutual consequences. Examples include communication devices sharing a network with limited capacity, and computer programs sharing a limited computational resource. Game theory provides a systematic framework for modeling cooperation and conflict between these so-called strategic players, and has been widely used in areas such as robust design, resource allocation, and network security [1]–[3].

A fundamental question in game theory is whether the players can converge to a Nash equilibrium—a tuple of strategies for which no one has a unilateral incentive to change—if they play the game repeatedly and adjust their strategies based on historical outcomes. A primary example of such a learning process is fictitious play [4], in which each player believes that its opponents are playing constant mixed strategies in agreement with the empirical distributions of their past actions, and plays the corresponding best response. Another well-known example is the gradient response method [5], in which each player adjusts its strategy using the gradient of its cost function. These learning processes have attracted significant research interests [6], [7].

In this paper, we propose an adaptive learning approach for a hierarchical game model proposed by Stackelberg [8], in which one player (called the *leader*) selects its action first, and then the other player (called the *follower*), informed of the leader's choice, selects its own action. Therefore, a follower's strategy in a Stackelberg game is a function that specifies a response to each leader's possible action.

Stackelberg games provide a natural framework for understanding systems with asymmetrical information, a common feature of many network problems [9], [10]. They are especially useful for modeling security problems, where the defender (leader) is usually unaware of the attack objective ahead of time, whereas the attacker (follower) is able to observe the defense strategy and attack after careful planning. Stackelberg Security Games have been applied to various real-world security domains, and have lead to practical implementations such as the ARMOR program at the Los Angeles International Airport [11].

Asymmetrical information often leads to scenarios with no Nash equilibrium but with a Stackelberg equilibrium, as the conditions for existence of the former is much stronger than those of the latter [1, p. 181]. For these scenarios, learning options like fictitious play and gradient response cannot be applied, and novel approaches are needed to achieve convergence to a Stackelberg equilibrium. Existing results on learning in Stackelberg games are limited to linear and quadratic costs and finite action sets [12]–[14], which are too restrictive for many applications including network security.

This paper studies two-player Stackelberg games with continuous action sets. We consider the scenario where the leader only has partial knowledge of the follower's action set and cost function. As a result, the follower's strategy belongs to a known family of functions parameterized by an unknown parameter vector. Our main contribution is an adaptive learning algorithm that simultaneously estimates the unknown parameter based on the follower's past actions and minimizes the leader's cost, designed based on adaptive control techniques and hysteresis switching. It guarantees that the leader's cost predicted using the parameter estimate becomes indistinguishable from its actual cost in finite time, up to a preselected, arbitrarily small error threshold, and that the first-order necessary condition for optimality holds asymptotically for the predicted cost. If an additional persistent excitation condition holds, then the parameter estimation error can also be bounded by a preselected, arbitrarily small threshold in finite time. Furthermore, we consider the case where the parameterized function known to the leader does not perfectly match the follower's actual strategy, and prove that our adaptive learning algorithm can be adjusted to guarantee the same convergence results for preselected error thresholds larger than the size of the mismatch. The algorithm and convergence results are illustrated via a simple simulation example motivated by link-flooding denial-of-

This work was supported by the ONR MURI grant N00014-16-1-2710.

G. Yang and J. P. Hespanha are with the Center for Control, Dynamical Systems, and Computation, University of California, Santa Barbara, CA 93106, USA. Email: {guosongyang, hespanha}@ucsb.edu.

R. Poovendran is with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA. Email: rp3@uw.edu.

service (DoS) attacks such as the Crossfire attack [15].

*Notations:* Let $\mathbb{R}_+ := [0, \infty)$ and $\mathbb{N} := \{0, 1, \ldots\}$. Let $I_n$ be the identity matrix in $\mathbb{R}^{n \times n}$; the subscript is omitted when the dimension is implicit. Denote by $\|\cdot\|$ the Euclidean norm for vectors and the (induced) Euclidean norm for matrices. For a set $\mathcal{S} \subset \mathbb{R}^n$, denote by $\partial\mathcal{S}$, $\overline{\mathcal{S}}$, and $\overline{\text{conv}}\,\mathcal{S}$ its boundary, closure, and closed convex hull, respectively. A signal $u : [t_0, \infty) \to \mathbb{R}^n$ is of class $\mathcal{L}_\infty$ if $\sup_{t \geq t_0} \|u(t)\|$ is finite. A function is of class $\mathcal{C}^1$ if it is continuously differentiable.

## II. Problem formulation

Consider a two-player game where $\mathcal{U} \subset \mathbb{R}^{n_u}$ and $\mathcal{A} \subset \mathbb{R}^{n_a}$ are the *action sets* of the first and the second players, respectively, and $J : \mathcal{U} \times \mathcal{A} \to \mathbb{R}$ and $H : \mathcal{A} \times \mathcal{U} \to \mathbb{R}$ are the corresponding *cost functions*. We are interested in a hierarchical game model proposed by Stackelberg [8], where the first player (called the *leader*) selects its action $u \in \mathcal{U}$ first, and then the second player (called the *follower*), informed of the leader's choice, selects its action $a \in \mathcal{A}$. The corresponding notion of equilibrium is defined as follows.

**Definition 1** ([1, Def. 4.6])**.** An action $u^* \in \mathcal{U}$ is a *Stackelberg equilibrium action* for the *leader* if

$$\sup_{a \in \beta_a(u^*)} J(u^*, a) = \inf_{u \in \mathcal{U}} \sup_{a \in \beta_a(u)} J(u, a),$$

where $\beta_a(u) \subset \mathcal{A}$ denotes the set of best responses against $u$, that is, $\beta_a(u) := \{a \in \mathcal{A} : H(a, u) = \inf_{a' \in \mathcal{A}} H(a', u)\}$, and $\sup_{a \in \beta_a(u)} J(u, a) = \infty$ if $\beta_a(u) = \emptyset$.

We consider games with perfect but incomplete information, where the leader only has partial knowledge of the follower's action set and cost function, and thus the follower's actual strategy is an unknown function $f^* : \mathcal{U} \to \mathcal{A}$ such that $f^*(u) \in \beta_a(u)$ for all $u \in \mathcal{U}$. However, $f^*$ belongs to a parameterized family of functions $\{u \mapsto f(\theta, u) : \theta \in \Theta\}$ defined by $f : \Theta \times \mathcal{U} \to \mathbb{R}^{n_a}$ and a parameter set $\Theta \subset \mathbb{R}^{n_\theta}$, that is, there is a constant $\theta^* \in \Theta$ such that

$$f^*(u) = f(\theta^*, u) \qquad \forall\, u \in \mathcal{U}. \tag{1}$$

The parameterized function $f$ and the parameter set $\Theta$ are known to the leader, but the actual value of $\theta^*$ is unknown.

In practice, assuming that the follower's actual strategy belongs to a known parameterized family of functions introduces little loss of generality, as it can always be approximated on a compact set up to an arbitrary precision as a finite weighted sum of a preselected class of basis functions. An example of such an approximation is the *radial basis function (RBF)* model [16], in which the leader assumes

$$f(\theta^*, u) = \sum_{j=1}^{n_\theta} \theta_j^* F_j(u) = \sum_{j=1}^{n_\theta} \theta_j^* \phi(\|u - u_j^c\|), \tag{2}$$

where $\theta^* = (\theta_1^*, \ldots, \theta_{n_\theta}^*)$ is the unknown parameter vector, and each $F_j : \mathcal{U} \to \mathbb{R}^{n_a}$ is an RBF centered at $u_j^c$. In the RBF model, the approximation is affine with respect to the unknown parameter, which is also common in many other widely-used approximation models such as *orthogonal polynomials* and *multivariate splines* [16]. This motivates restricting our attention to affine maps $\theta \mapsto f(\theta, u)$. The following assumption captures this and additional regularity conditions that we use to guarantee existence of a Stackelberg equilibrium.

**Assumption 1** (Regularity)**.** The leader's action set $\mathcal{U}$ and the parameter set $\Theta$ are convex and compact, the leader's cost function $J$ is $\mathcal{C}^1$ with locally Lipschitz Jacobian, the parameterized function $f$ is $\mathcal{C}^1$ with locally Lipschitz gradient, and the map $\theta \mapsto f(\theta, u)$ is affine for each fixed $u \in \mathcal{U}$.

Under Assumption 1, existence of a Stackelberg equilibrium follows from standard results, e.g., [1, Th. 4.8]. In particular, these conditions are much weaker than the sufficient conditions for existence of a Nash equilibrium [1, p. 181], which is consistent with our interest in games games with no Nash equilibrium but with a Stackelberg equilibrium.

We denote by $\nabla_u J(u, a)$ and $\nabla_a J(u, a)$ the gradients of the maps $u \mapsto J(u, a)$ and $a \mapsto J(u, a)$, respectively, and by $\nabla_\theta f(u)$ and $\nabla_u f(\theta, u)$ the Jacobian matrices of the maps $\theta \mapsto f(\theta, u)$ and $u \mapsto f(\theta, u)$, respectively. (To be consistent with the definition of Jacobian matrix, we take gradients as row vectors.) In particular, the Jacobian matrix $\nabla_\theta f(u)$ is independent of $\theta$ due to the affine condition in Assumption 1.

Our goal is to adjust the leader's action $u$ to minimize its cost $J(u, a)$, that is, to solve the optimization problem

$$\min_{u \in \mathcal{U}} J(u, f(\theta^*, u)),$$

based on past observations of the follower's action $a = f(\theta^*, u)$ and the leader's cost $J(u, a)$, but without knowing the actual parameter $\theta^*$. Our approach to solve this problem consists of two components:

1) Construct a *parameter estimate* $\theta$ that approaches the actual parameter $\theta^*$.
2) Adjust the leader's action $u$ based on a gradient descent method to minimize its *predicted cost* $\hat{J}(u, \theta) := J(u, f(\theta, u))$, that is, to solve the optimization problem

$$\min_{u \in \mathcal{U}} \hat{J}(u, \theta). \tag{3}$$

In this paper, our analysis and design are formulated using continuous-time dynamics, which is common in the literature of learning in game theory [6].

## III. Estimation and minimization

To specify the adaptive algorithm for estimating the unknown parameter $\theta^* \in \Theta$ and optimizing the leader's action $u \in \mathcal{U}$, we recall the following notions and basic properties from convex analysis; for more details, see, e.g., [17, Sec. 6].

For a closed convex set $\mathcal{C} \subset \mathbb{R}^n$ and a point $v \in \mathbb{R}^n$, we denote by $[v]_\mathcal{C}$ the *projection* of $v$ onto $\mathcal{C}$, that is, $[v]_\mathcal{C} := \arg\min_{w \in \mathcal{C}} \|w - v\|$. Then $[v]_\mathcal{C}$ is unique as the set $\mathcal{C}$ is closed and convex, and $[v]_\mathcal{S} = v$ if $v \in \mathcal{S}$.

For a convex set $\mathcal{S} \subset \mathbb{R}^n$ and a point $x \in \mathcal{S}$, we denote by $T_\mathcal{S}(x)$ the *tangent cone* to $\mathcal{S}$ at $x$, that is,

$$T_\mathcal{S}(x) := \overline{\{h(z - x) : z \in \mathcal{S},\, h > 0\}}, \tag{4}$$

and by $N_{\mathcal{S}}(x)$ the *normal cone* to $\mathcal{S}$ at $x$, that is,

$$N_{\mathcal{S}}(x) := \{v \in \mathbb{R}^n : \forall\, w \in T_{\mathcal{S}}(x),\ v^\top w \le 0\}. \tag{5}$$

Then $T_{\mathcal{S}}(x)$ and $N_{\mathcal{S}}(x)$ are closed and convex, and $T_{\mathcal{S}}(x) = \mathbb{R}^n$ and $N_{\mathcal{S}}(x) = \{0\}$ if $x \in \mathcal{S}\backslash\partial\mathcal{S}$. For all $v \in \mathbb{R}^n$, we have

$$[v]_{T_{\mathcal{S}}(x)} \in T_{\mathcal{S}}(x), \quad v - [v]_{T_{\mathcal{S}}(x)} \in N_{\mathcal{S}}(x) \tag{6}$$

and

$$\big(v - [v]_{T_{\mathcal{S}}(x)}\big)^\top [v]_{T_{\mathcal{S}}(x)} = 0. \tag{7}$$

### A. Parameter estimation

We construct the parameter estimate $\theta$ by comparing past observations of the follower's action $a = f(\theta^*, u)$ and the leader's cost $J(u, a)$ with the corresponding predicted values $f(\theta, u)$ and $\hat{J}(u, \theta) = J(u, f(\theta, u))$. Our goal is to ensure that the norm of the *estimation error* $\|\theta - \theta^*\|$ is monotonically nonincreasing, regardless of how the leader's action $u$ is adjusted. First, we establish a relation between the *observation error*

$$e_{\text{obs}} := \begin{bmatrix} f(\theta, u) - a \\ \hat{J}(u, \theta) - J(u, a) \end{bmatrix}$$

and the estimation error $\theta - \theta^*$.

**Lemma 1.** *The observation error $e_{\text{obs}}$ satisfies*

$$e_{\text{obs}} = K(u, a, \theta)(\theta - \theta^*) \tag{8}$$

*with the gain matrix*

$$K(u, a, \theta) := \begin{bmatrix} I \\ \int_0^1 \nabla_a J(u, \rho f(\theta, u) + (1 - \rho)a)\, \mathrm{d}\rho \end{bmatrix} \nabla_\theta f(u). \tag{9}$$

Following Lemma 1, the observation error $e_{\text{obs}}$ would be zero if the current estimate $\theta$ of the unknown parameter $\theta^*$ was correct. However, in most interesting scenarios, the dimension $n_\theta$ of $\theta^*$ is much larger than the dimension $n_a + 1$ of $e_{\text{obs}}$; thus the gain matrix $K(u, a, \theta)$ cannot be invertible, and $e_{\text{obs}} = 0$ does not imply $\theta = \theta^*$.

We propose the following estimation law to drive the parameter estimate $\theta$ towards the actual parameter $\theta^*$:

$$\dot{\theta} = [-\lambda_e K(u, a, \theta)^\top e_{\text{obs}}]_{T_\Theta(\theta)} \tag{10}$$

with the gain matrix $K(u, a, \theta)$ defined by (9) and the *switching signal* $\lambda_e : \mathbb{R}_+ \to \{0, \lambda_1\}$ defined by

$$\lambda_e(t) := \begin{cases} \lambda_1 & \text{if } \|e_{\text{obs}}(t)\| \ge \varepsilon_{\text{obs}}; \\ \lim\limits_{s \nearrow t} \lambda_e(s) & \text{if } \|e_{\text{obs}}(t)\| \in (\varepsilon_{\text{obs}}/2, \varepsilon_{\text{obs}}); \\ 0 & \text{if } \|e_{\text{obs}}(t)\| \le \varepsilon_{\text{obs}}/2 \end{cases} \tag{11}$$

and $\lambda_e(0) := \lambda_1$ if $\|e_{\text{obs}}(0)\| \in (\varepsilon_{\text{obs}}/2, \varepsilon_{\text{obs}})$, where $\varepsilon_{\text{obs}}, \lambda_1 > 0$ are preselected constants. Several comments are in order: First, the gain matrix $K(u, a, \theta)$ does not depend on the actual parameter $\theta^*$, so (10) can be implemented without knowing $\theta^*$. Second, the projection $[\cdot]_{T_\Theta(\theta)}$ onto the tangent cone $T_\Theta(\theta)$ is used to ensure that $\theta$ remains inside the compact convex set $\Theta$. Finally, the right-continuous, piecewise constant switching signal $\lambda_e$ is designed so that the adaption is on when $\|e_{\text{obs}}\| \ge \varepsilon_{\text{obs}}$ and off when $\|e_{\text{obs}}\| \le \varepsilon_{\text{obs}}/2$, with a hysteresis switching rule that avoids chattering. The key feature of (10) is that the estimation error $\theta - \theta^*$ satisfies

$$\begin{aligned} \frac{\mathrm{d}\|\theta - \theta^*\|^2}{\mathrm{d}t} &= 2(\theta - \theta^*)^\top [-\lambda_e K(u, a, \theta)^\top e_{\text{obs}}]_{T_\Theta(\theta)} \\ &\le 2(\theta - \theta^*)^\top \big(-\lambda_e K(u, a, \theta)^\top e_{\text{obs}}\big) = -2\lambda_e \|e_{\text{obs}}\|^2, \end{aligned}$$

where the inequality follows from (4)–(6). Hence the estimation law (10) with the switching signal (11) guarantees

$$\frac{\mathrm{d}\|\theta - \theta^*\|^2}{\mathrm{d}t} \le -2\lambda_e \|e_{\text{obs}}\|^2 \le 0, \tag{12}$$

which implies that $\|\theta - \theta^*\|$ is monotonically nonincreasing and will not stop approaching zero unless $\|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$. In the convergence results in Section IV, we will show that the adaption of the parameter estimate $\theta$ stops in finite time, and the observation error $e_{\text{obs}}$ satisfies $\|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$ afterward.

### B. Cost minimization

Several options are available to adjust the leader's action $u$, but in this paper our analysis will focus on a gradient descent method, which is fairly robust for a wide range of problems. Our ultimate goal is to minimize the leader's cost $J(u, a) = J(u, f(\theta^*, u))$. However, computing the gradient descent direction of the actual cost requires knowledge of the actual parameter $\theta^*$. Therefore, we minimize instead the leader's estimated cost $\hat{J}(u, \theta)$, which only depends on the parameter estimate $\theta$. This change in objective is justified by the fact that $\|\hat{J}(u, \theta) - J(u, a)\| \le \|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$ holds after finite time, which will be established in Section IV.

The time derivative of the estimated cost $\hat{J}(u, \theta)$ is

$$\dot{\hat{J}}(u, \theta) = \nabla_u \hat{J}(u, \theta)\, \dot{u} + \nabla_\theta \hat{J}(u, \theta)\, \dot{\theta}, \tag{13}$$

where $\nabla_\theta \hat{J}(u, \theta) = \nabla_a J(u, f(\theta, u)) \nabla_\theta f(u)$ and $\nabla_u \hat{J}(u, \theta) = \nabla_u J(u, f(\theta, u)) + \nabla_a J(u, f(\theta, u)) \nabla_u f(\theta, u)$ are the gradients of the maps $u \mapsto \hat{J}(u, \theta)$ and $\theta \mapsto \hat{J}(u, \theta)$, respectively (note that $\nabla_u J(u, f(\theta, u))$ denotes the gradient of the map $u \mapsto J(u, \hat{a})$ at $\hat{a} = f(\theta, u)$ here). As we will establish that the adaption of $\theta$ stops in finite time, we neglect the second term in (13) and focus exclusively in adjusting $u$ along the gradient descent direction of $u \mapsto \hat{J}(u, \theta)$. This motivates the following minimization law to adjust the leader's action:

$$\dot{u} = \big[-\lambda_2 \nabla_u \hat{J}(u, \theta)^\top\big]_{T_{\mathcal{U}}(u)}, \tag{14}$$

where $\lambda_2 > 0$ is a preselected constant. The projection $[\cdot]_{T_{\mathcal{U}}(u)}$ onto the tangent cone $T_{\mathcal{U}}(u)$ is used to ensure that $u$ remains inside the compact convex set $\mathcal{U}$. Then (7) implies

$$\begin{aligned} \dot{\hat{J}}(u, \theta) &= \nabla_u \hat{J}(u, \theta) \big[-\lambda_2 \nabla_u \hat{J}(u, \theta)^\top\big]_{T_{\mathcal{U}}(u)} + \nabla_\theta \hat{J}(u, \theta)\, \dot{\theta} \\ &= -\big\|\big[-\lambda_2 \nabla_u \hat{J}(u, \theta)^\top\big]_{T_{\mathcal{U}}(u)}\big\|^2 / \lambda_2 + \nabla_\theta \hat{J}(u, \theta)\, \dot{\theta} \\ &= -\|\dot{u}\|^2 / \lambda_2 + \nabla_\theta \hat{J}(u, \theta)\, \dot{\theta}. \end{aligned}$$

Hence the minimization law (14) ensures that if $\dot{\theta} = 0$ then $\dot{\hat{J}}(u, \theta) \le -\|\dot{u}\|^2 / \lambda_2 \le 0$. In the convergence results in Section IV, we will show that the leader's action $u$ approaches the set of points for which the first-order necessary condition for optimality holds for the optimization problem (3).

## IV. Convergence analysis

We now state the main result of this paper:

**Theorem 1.** *Suppose that Assumption 1 holds. Given any threshold $\varepsilon_{\text{obs}} > 0$ in* (11)*, the estimation and minimization algorithm* (10) *and* (14) *with the switching signal* (11) *guarantees the following properties:*

*1) There exists a time $T \geq 0$ such that*

$$\|e_{\text{obs}}(t)\| < \varepsilon_{\text{obs}}, \quad \theta(t) = \theta(T) \qquad \forall\, t \geq T. \tag{15}$$

*2) The first-order necessary condition for optimality holds asymptotically for the optimization problem* (3)*, that is,*

$$\lim_{t\to\infty} \left[-\nabla_u \hat{J}(u(t), \theta(T))^\top\right]_{T_{\mathcal{U}}(u(t))} = 0. \tag{16}$$

Essentially, item 1) guarantees that the parameter estimate $\theta$ converges in finite time to a point which is indistinguishable from the actual parameter $\theta^*$ based on observations of the follower's action $a = f(\theta^*, u)$ and the leader's cost $J(u, a)$, up to an error no larger than the threshold $\varepsilon_{\text{obs}}$. For item 2), the necessity of (16) for optimality is justified by the following result, which is a consequence of [17, Th. 6.12].

**Lemma 2.** *If $\hat{u}^*$ is locally optimal for the optimization problem* (3) *with some fixed $\theta$, then $\left[-\nabla_u \hat{J}(\hat{u}^*, \theta)^\top\right]_{T_{\mathcal{U}}(\hat{u}^*)} = 0$.*

*Proof of Theorem 1.* As the right hand-sides of (10) and (14) are potentially discontinuous due to the projections and switching, the proof of Theorem 1 uses results from differential inclusions theory; see the Appendix for the necessary preliminaries.

First, we establish existence, boundedness, and uniqueness of solutions for the system defined by (10) and (14).

**Lemma 3.** *For each $(\theta_0, u_0) \in \Theta \times \mathcal{U}$, there is a unique Carathéodory solution to the system defined by* (10) *and* (14) *on $\mathbb{R}_+$ with $(\theta(0), u(0)) = (\theta_0, u_0)$, that is, there are unique absolutely continuous functions $\theta : \mathbb{R}_+ \to \mathbb{R}^{n_\theta}$ and $u : \mathbb{R}_+ \to \mathbb{R}^{n_u}$ such that* (10) *and* (14) *hold almost everywhere on $\mathbb{R}_+$ with $(\theta(0), u(0)) = (\theta_0, u_0)$. Moreover, $(\theta(t), u(t)) \in \Theta \times \mathcal{U}$ for all $t \geq 0$, and $\theta, \dot{\theta}, u, \dot{u}, e_{\text{obs}}, \dot{e}_{\text{obs}} \in \mathcal{L}_\infty$.*

Lemma 3 is established by modeling the system defined by (10) and (14) using the project dynamical system (26) in the Appendix with the state $x := (\theta, u)$ and the set $\mathcal{S} := \Theta \times \mathcal{U}$, and then combining the results on hysteresis switching from [18] with the results on existence, boundedness, and uniqueness of solutions for (26) from Lemma 4.

Second, we establish item 1) of Theorem 1 via arguments along the lines of the proof of Barbalat's lemma [19, Lemma 3.2.6]. We cannot use Barbalat's lemma directly since the switching signal $\lambda_e$ in (10) is not continuous but only piecewise continuous. Following (12), we see that $\|\theta - \theta^*\|^2$ is monotonically nonincreasing. Therefore $\lim_{t\to\infty} \|\theta(t) - \theta^*\|^2$, and thus

$$\lim_{t\to\infty} \int_0^t \lambda_e(s)\|e_{\text{obs}}(s)\|^2 \,\mathrm{d}s, \tag{17}$$

exists and is finite. On the other hand, (10) and (11) imply that (15) holds if there exists a time $T \geq 0$ such that

$$\lambda_e(t) = 0 \qquad \forall\, t \geq T. \tag{18}$$

Assume (18) does not hold for any $T \geq 0$. Then (11) implies that there exists an unbounded increasing sequence $(t_k)_{k\in\mathbb{N}}$ with $t_0 > 0$ such that $\lambda_e(t_k) = \lambda_1$ and $\|e_{\text{obs}}(t_k)\| > \varepsilon_{\text{obs}}/2$ for all $k \in \mathbb{N}$. Now we show that there exists an unbounded sequence $(s_k)_{k\in\mathbb{N}}$ with $s_k \in [t_k - \delta, t_k]$ such that

$$\|e_{\text{obs}}(t)\| > \varepsilon_{\text{obs}}/2, \quad \lambda_e(t) = \lambda_1 \tag{19}$$

for all $k \in \mathbb{N}$ and $t \in [s_k, s_k + \delta)$ with the constant $\delta := \min\{t_0, \varepsilon_{\text{obs}}/(2\sup_{s\geq 0}\|\dot{e}_{\text{obs}}(s)\|)\} > 0$. Indeed, for each $k \in \mathbb{N}$, consider the following two possibilities:

1) If $\|e_{\text{obs}}(t)\| < \varepsilon_{\text{obs}}$ for all $t \in [t_k - \delta, t_k]$, then (11) and $\lambda_e(t_k) = \lambda_1$ imply that (19) holds with $s_k = t_k - \delta$.
2) Otherwise, there exists an $s_k \in [t_k - \delta, t_k]$ such that $\|e_{\text{obs}}(s_k)\| = \varepsilon_{\text{obs}}$, and (19) follows from the definition of $\delta$ and (11).

Following (19), we see that

$$\int_{s_k}^{s_k+\delta} \lambda_e(s)\|e_{\text{obs}}(s)\|^2 \,\mathrm{d}s \geq \frac{\lambda_1 \varepsilon_{\text{obs}}^2 \delta}{4} > 0$$

for the unbounded sequence $(s_k)_{k\in\mathbb{N}}$, which contradicts the property that (17) exists and is finite. Therefore, there exists a time $T \geq 0$ such that (18), and thus (15), holds.

Finally, we establish item 2) of Theorem 1 using the invariance principle for projected gradient descent Proposition 5 in the Appendix. After time $T$, the system (14) becomes

$$\dot{u} = \left[-\lambda_2 \nabla_u \hat{J}(u, \theta(T))^\top\right]_{T_{\mathcal{U}}(u)},$$

which can be modeled using the projected dynamical system (26) in the Appendix with the state $x := u$ and the set $\mathcal{S} := \mathcal{U}$. The corresponding function $g$ in (26) is given by $g(x) := -\lambda_2 \nabla_u \hat{J}(x, \theta(T))^\top$ which satisfies (30) with $V(x) := \lambda_2 \hat{J}(x, \theta(T))$. Then (16) follows from (31) in Proposition 5. □

In Theorem 1, there is no claim that the parameter estimate $\theta$ necessarily converges to the actual parameter $\theta^*$. However, this can be guaranteed if we assume that the following *persistent excitation (PE)* condition holds.

**Assumption 2** (PE)**.** There exist constants $\tau_0, \alpha_0 > 0$ such that the gain matrix $K(u, a, \theta)$ defined by (9) satisfies

$$\int_t^{t+\tau_0} K(s)^\top K(s) \,\mathrm{d}s \geq \alpha_0 I \qquad \forall\, t \geq 0, \tag{20}$$

where we let $K(t) := K(u(t), a(t), \theta(t))$ for brevity.

**Theorem 2.** *Suppose that Assumptions 1 and 2 hold. Then by setting the threshold*

$$\varepsilon_{\text{obs}} := \varepsilon_\theta \sqrt{\alpha_0/\tau_0} \tag{21}$$

*in* (11) *for any given constant $\varepsilon_\theta > 0$, the estimation and minimization algorithm* (10) *and* (14) *with the switching signal* (11) *guarantees the following properties:*

*1) There exists a time $T \geq 0$ such that* (15) *holds, and*

$$\|\theta(T) - \theta^*\| < \varepsilon_\theta. \tag{22}$$

*2) The first-order necessary condition for optimality holds asymptotically for* (3)*, that is,* (16) *holds.*

*Proof.* As (15) and (16) are established in Theorems 1, it remains to prove (22). To this effect, we note that the inequality in (15) implies

$$\int_T^{T+\tau_0} \|e_{\rm obs}(s)\|^2 \,\mathrm{d}s < \varepsilon_{\rm obs}^2 \tau_0 = \alpha_0 \varepsilon_\theta^2,$$

where the equality follows from (21). On the other hand, (8) and the equality in (15) imply

$$\begin{aligned}\int_T^{T+\tau_0} \|e_{\rm obs}(t)\|^2 \,\mathrm{d}t &= \int_T^{T+\tau_0} \|K(t)(\theta(T) - \theta^*)\|^2 \,\mathrm{d}s \\ &\geq \alpha_0 \|\theta(T) - \theta^*\|^2,\end{aligned}$$

where the inequality follows from the PE condition (20). Combining the inequalities above yields (22). □

*Remark* 1. In view of (9), a sufficient condition for (20) is

$$\int_t^{t+\tau_0} \nabla_\theta f(u(s))^\top \nabla_\theta f(u(s)) \,\mathrm{d}s \geq \alpha_0 I \quad \forall\, t \geq 0. \tag{23}$$

The PE condition (23) is more restrictive than (20), but has the advantage that it can be checked without knowing the estimate $\theta$. Moreover, from the proof of Theorem 2, we see that (22) only requires (20) or (23) to hold at $t = T$. Consequently, in practice it suffices to enforce (20) or (23) when $\lambda_e$ in (11) has been set to zero.

## V. Model Mismatch

Up till now we assumed that there was some unknown parameter $\theta^*$ from the known parameter set $\Theta$ such that (1) holds for the follower's actual strategy $f^*$ and the parameterized function $f$ known to the leader. In this section, we consider the case where such perfect matching may not exist, and study the effect of a bounded mismatch between $f^*$ and $f$.

**Assumption 3** (Mismatch)**.** The follower's actual strategy $f^*$ is locally Lipschitz, and there is an unknown parameter $\theta^* \in \Theta$ such that

$$\max_{u \in \mathcal{U}} \|f(\theta^*, u) - f^*(u)\| \leq \varepsilon_f / \kappa$$

with

$$\kappa := \max_{\theta \in \Theta,\, u \in \mathcal{U}} \left\| \begin{bmatrix} I \\ \int_0^1 \nabla_a J(u, \rho f(\theta, u) + (1-\rho) f^*(u)) \,\mathrm{d}\rho \end{bmatrix} \right\|$$

for some known constant $\varepsilon_f > 0$.

Similar arguments to those in the proof of Lemma 1 show that the observation error $e_{\rm obs}$ now satisfies

$$e_{\rm obs} = K(u, a, \theta)(\theta - \theta^*) + e_f \tag{24}$$

with the gain matrix $K(u, a, \theta)$ defined by (9) and the *mismatch error*

$$e_f := \begin{bmatrix} I \\ \int_0^1 \nabla_a J(u, \rho f(\theta, u) + (1-\rho)\, a) \,\mathrm{d}\rho \end{bmatrix} (f(\theta^*, u) - a).$$

Then Assumption 3 implies $\|e_f(t)\| \leq \varepsilon_f$ for all $t \geq 0$. The effect of the mismatch error $e_f$ in (24) can be mitigated by keeping the same estimation law (10) while adjusting the definition of the switching signal $\lambda_e$ by

$$\lambda_e(t) := \begin{cases} \lambda_1 & \text{if } \|e_{\rm obs}(t)\| \geq \varepsilon_{\rm obs}; \\ \lim\limits_{s \nearrow t} \lambda_e(s) & \text{if } \|e_{\rm obs}(t)\| \in ((\varepsilon_{\rm obs} + \varepsilon_f)/2, \varepsilon_{\rm obs}); \\ 0 & \text{if } \|e_{\rm obs}(t)\| \leq (\varepsilon_{\rm obs} + \varepsilon_f)/2 \end{cases} \tag{25}$$

and $\lambda_e(0) = \lambda_1$ if $\|e_{\rm obs}(0)\| \in ((\varepsilon_{\rm obs} + \varepsilon_f)/2, \varepsilon_{\rm obs})$, where $\varepsilon_{\rm obs} > \varepsilon_f$ and $\lambda_1 > 0$ are preselected constants.

The following two results extend Theorems 1 and 2 to the current case without perfect matching between the follower's actual strategy $f^*$ and some map $u \mapsto f(\theta^*, u)$.

**Theorem 3.** *Suppose that Assumptions 1 and 3 hold. Given any threshold $\varepsilon_{\rm obs} > \varepsilon_f$ in* (25)*, the estimation and minimization algorithm* (10) *and* (14) *with the switching signal* (25) *guarantees the following properties:*

*1) There exists a time $T \geq 0$ such that* (15) *holds.*

*2) The first-order necessary condition for optimality holds asymptotically for* (3)*, that is,* (16) *holds.*

**Theorem 4.** *Suppose that Assumptions 1–3 hold. Then by setting the threshold $\varepsilon_{\rm obs} := \sqrt{\alpha_0 \varepsilon_\theta^2/(2\tau_0) - \varepsilon_f^2}$ in* (25) *for any given constant $\varepsilon_\theta > 2\varepsilon_f \sqrt{\tau_0/\alpha_0}$, the estimation and minimization algorithm* (10) *and* (14) *with the switching signal* (25) *guarantees the following properties:*

*1) There exists a time $T \geq 0$ such that* (15) *and* (22) *hold.*

*2) The first-order necessary condition for optimality holds asymptotically for* (3)*, that is,* (16) *holds.*

## VI. Simulation Example

We illustrate the estimation and minimization algorithm via a simple example motivated by link-flooding denial-of-service (DoS) attacks such as the Crossfire attack [15].

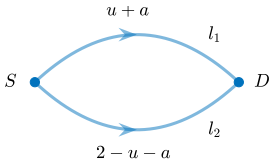



Fig. 1. A simple network with one source $S$, one destination $D$, and two links $l_1$ and $l_2$.

Consider the communication network in Fig. 1, and suppose there is a router (leader) that distributes 1 unit of legitimate traffic between the two links $l_1$ and $l_2$, and an attacker (follower) that disrupts communication by injecting 1 unit of malicious traffic on the two links. Denote by $u, a \in [0, 1]$ the amounts of legitimate and malicious traffic on the link $l_1$, respectively. Then the total traffic on the links $l_1$ and

$l_2$ are given by $f_1 = u + a$ and $f_2 = 2 - u - a$, respectively. We assume that a communication delay is incurred on each link, which is a quadratic function of the corresponding total traffic, that is, $p_1(f_1) := f_1^2$ and $p_2(f_2) := f_2^2$, and that the router aims to minimize the average delay for legitimate traffic, which corresponds to the cost function defined by $J(u,a) := up_1(f_1) + (1-u)p_2(f_2) = 5u^2 + 6ua + a^2 - 8u - 4a + 4$, whereas the attacker aims to disrupt communication by maximizing the average delay for legitimate traffic, which corresponds to the cost function $H(a,u) := -J(u,a)$.

Standard convex analysis shows that the attacker's best response against $u$ is given by

$$\beta_a(u) := \begin{cases} 0 & u \le 1/2; \\ 1 & u > 1/2, \end{cases}$$

and the router's best response against $a$ is given by

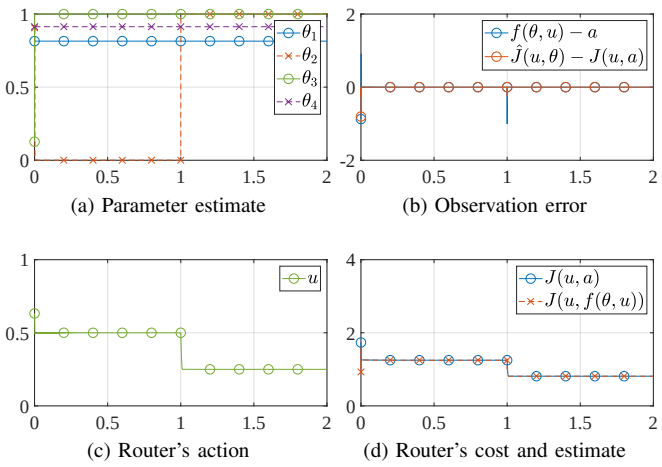
$$\beta_u(a) := (4 - 3a)/5.$$

Then it is straightforward to see that a Nash equilibrium does not exist for this game, but there is a Stackelberg equilibrium action $u^* = 1/2$ for the router.

If the router knew that the attacker's cost function was indeed $H$, it could select the Stackelberg equilibrium action $u^* = 1/2$. However, we consider a scenario where it does not and, instead, will use the approach proposed in this paper to construct its optimal action. To this effect, the router assumes that the attacker's strategy satisfies the RBF model (2) with $n_\theta = 4$, the parameter set $\Theta := [0,1]^4$, and the RBF defined by $F_j(u) := \mathbb{1}_{(-1/8,1/8]}(\|u - (2j-1)/8\|)$ for $j = 1, \ldots, 4$.[1] For the specific cost function $H$, the attacker's actual strategy is given by $f(\theta^*, u)$ with $\theta^* = (0,0,1,1)$.

In the simulations shown in Fig. 2 and 3, the constants are set to $\varepsilon_{\text{obs}} = 10^{-3}$, $\lambda_1 = 1$, and $\lambda_2 = 10^{-2}$, and the initial values of the parameter estimate $\theta$ and the router's action $u$ are randomly generated. For the case without enforcing PE in Fig. 2, in the first $10^4$ units of time, the router's action $u$ converges to the optimum $u^* = 1/2$, despite that the parameter estimate $\theta$ does not converge to the actual parameter $\theta^*$. In Fig. 3, we enforce PE by adding some random noise to $u$ for a short interval when the observation error $\|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$. In this case, in the first $10^4$ units of time, the router's action $u$ converges to the optimum $u^* = 1/2$, and the parameter estimate $\theta$ converges to the actual parameter $\theta^*$. In both cases, we also simulate the scenario where after $10^4$ units of time, the attacker starts focusing more on disrupting the link $l_1$ (by using $H(a,u) := -5up_1(f_1) - (1-u)p_2(f_2)$ as the new cost function), so that the new value of the unknown parameter is $\theta^* = (0,1,1,1)$, and the new router's Stackelberg equilibrium action is $u^* = 1/4$. The corresponding simulation results show that our estimation and minimization algorithm is able to identify this switch in the attack, as the router's action converges to the new optimum $u^* = 1/4$ in both Fig. 2 and 3, and the parameter estimate $\theta$ converges to the new parameter $\theta^*$ in Fig. 3.

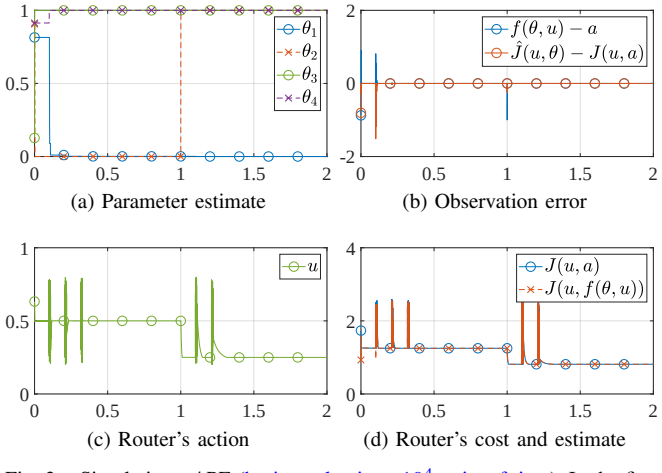
[1]The function $f$ used here actually violates the regularity conditions in Assumption 1 as it is discontinuous in $u$. The continuity requirement of $f$ in Assumption 1 is only needed so that the gradient descent is well-defined and does not lead to chattering. In simulation, these issues can be handled by using generalized subgradients at discontinuities [17] and setting $\dot{u} = 0$ when the right-hand side of (14) becomes very small.



(a) Parameter estimate (b) Observation error

(c) Router's action (d) Router's cost and estimate

Fig. 2. Simulation w/o PE (horizontal axis: $\times 10^4$ units of time). In the first $10^4$ units of time, the router's action $u$ converges to the optimum $u^* = 1/2$; in the second $10^4$ units of time, the attacker's cost function changes, and the router's action $u$ converges to the new optimum $u^* = 1/4$; the parameter estimate $\theta$ does not converge to the actual parameter $\theta^*$ in either case.



(a) Parameter estimate (b) Observation error

(c) Router's action (d) Router's cost and estimate

Fig. 3. Simulation w/ PE (horizontal axis: $\times 10^4$ units of time). In the first $10^4$ units of time, the router's action $u$ converges to the optimum $u^* = 1/2$; in the second $10^4$ units of time, the attacker's cost function changes, and the router's action $u$ converges to the new optimum $u^* = 1/4$; the parameter estimate $\theta$ converges to the actual parameter $\theta^*$ in both cases.

## VII. Future Research Topics

A feature of our learning law (10) is that the norm of the estimation error $\|\theta - \theta^*\|$ is monotonically nonincreasing, and the observation error $e_{\text{obs}}$ is bounded in norm by the preselected, arbitrarily small threshold $\varepsilon_{\text{obs}}$ in finite time, regardless how the leader's action is adjusted. A future research direction is to integrate our learning law with more efficient optimization methods for minimizing the leader's cost. Other future research topics include to relax the affine condition in Assumption 1, and to extend the current results to Stackelberg games on distributed networks.

## APPENDIX
## PROJECTED DYNAMICAL SYSTEMS

Let $\mathcal{S} \subset \mathbb{R}^n$ be a compact convex set, and $g : \mathcal{S} \to \mathbb{R}^n$ a locally Lipschitz function. In this section, we prove existence, boundedness, and uniqueness of solutions for the *projected dynamical system*

$$\dot{x} = [g(x)]_{T_{\mathcal{S}}(x)}. \tag{26}$$

The difficulty in analyzing (26) lies in that fact that its right-hand side is only defined in the domain $\mathcal{S}$ and is potentially discontinuous due to the projection $[\cdot]_{T_{\mathcal{S}}(x)}$. Therefore, we extend (26) to the differential inclusion

$$\dot{x} \in G(x) \tag{27}$$

with the set-valued function $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by

$$G(x) := \bigcap_{\varepsilon > 0} \overline{\operatorname{conv}}\{[g([y]_{\mathcal{S}})]_{T_{\mathcal{S}}([y]_{\mathcal{S}})} : \|y - x\| \le \varepsilon\},$$

which is upper semicontinuous on $\mathbb{R}^n$, and satisfies that $G(x)$ is convex and compact for each $x \in \mathbb{R}^n$.

**Lemma 4.** *For each $x_0 \in \mathcal{S}$, there is a unique Carathéodory solution to* (27) *on $\mathbb{R}_+$ with $x(0) = x_0$, that is, there exists a unique absolutely continuous function $x : \mathbb{R}_+ \to \mathbb{R}^n$ with $x(0) = x_0$ such that* (27) *holds almost everywhere on $\mathbb{R}_+$. Moreover, $x(t) \in \mathcal{S}$ for all $t \ge 0$, and $x$ is also a unique Carathéodory solution to* (26) *on $\mathbb{R}_+$ with $x(0) = x_0$.*

The proof of Lemma 4 uses the following properties of the functions $g$ and $G$. As the set $\mathcal{S}$ is compact and the function $g$ is locally Lipschitz, there is a constant $\gamma \ge 0$ such that

$$\|g(x) - g(y)\| \le \gamma \|x - y\| \qquad \forall\, x, y \in \mathcal{S}. \tag{28}$$

Then (5) and (6) imply that the set-valued function $G$ satisfies the *one-sided Lipschitz condition*

$$(v - w)^\top (x - y) \le \gamma \|x - y\|^2 \tag{29}$$

for all $x, y \in \mathcal{S}$, $v \in G(x)$ and $w \in G(y)$.

*Proof of Lemma 4.* Consider an arbitrary $T > 0$. The Lipschitz condition (28), together with the triangle inequality, implies that $\|g(x)\| \le \alpha(1 + \|x\|)$ for all $x \in \mathcal{S}$ with the constant $\alpha := \max_{y \in \mathcal{S}}\{\|g(y)\| + \gamma\|y\|, \gamma\}$. Then similar arguments to those in [20] imply existence and boundedness of a Carathéodory solution to (26) as well as (27) on $[0, T]$. Moreover, as the one-sided Lipschitz condition (29) holds, [21, Cor. 2.4] implies uniqueness of the Carathéodory solution on $[0, T]$. Finally, the proof is completed by noticing that $T > 0$ is arbitrary. □

Next, we establish an invariance principle for the case where $g$ is defined by a gradient descent procedure.

**Proposition 5.** *Suppose that the function $g$ in* (26) *satisfies*

$$g(z) = -\nabla V(z)^\top \qquad \forall\, z \in \mathcal{S} \tag{30}$$

*for some function $V : \mathcal{S} \to \mathbb{R}$. Then every Carathéodory solution $x$ to* (26) *satisfies*

$$\lim_{t \to \infty} [g(x(t))]_{T_{\mathcal{S}}(x(t))} = 0. \tag{31}$$

*Proof.* Based on (4), (6), (7), and (28), we can prove that

$$g(z)^\top w \ge \|[g(z)]_{T_{\mathcal{S}}(z)}\|^2 \qquad \forall\, z \in \mathcal{S}, \forall\, w \in G(z). \tag{32}$$

Then the function $V$ satisfies

$$\nabla V(z)\, w \le -\|[g(z)]_{T_{\mathcal{S}}(z)}\|^2 \qquad \forall\, z \in \mathcal{S}, \forall\, w \in G(z).$$

Note that a Filippov solution to (26) is a Carathéodory solution to (27). Then Lemma 4 and the invariance principle for Filippov solutions [22, Th. 3.2] imply that every Carathéodory solution to (26) approaches the largest invariant set in $\overline{\{z \in \mathcal{S} : \exists\, w \in G(z) \text{ s.t. } \nabla V(z)\, w = 0\}} \subset \{z \in \mathcal{S} : \|[g(z)]_{T_{\mathcal{S}}(z)}\| = 0\}$. Hence (31) holds as $g$ is continuous on the compact set $\mathcal{S}$. □

## REFERENCES


[1] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed. SIAM, 1999.

[2] T. Alpcan and T. Başar, *Network Security: A Decision and Game-Theoretic Approach*. Cambridge University Press, 2010.

[3] J. P. Hespanha, *Noncooperative Game Theory: An Introduction for Engineers and Computer Scientists*. Princeton University Press, 2017.

[4] J. Robinson, "An iterative method of solving a game," *Ann. Math.*, vol. 54, no. 2, pp. 296–301, 1951.

[5] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.

[6] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*. MIT Press, 1998.

[7] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[8] H. von Stackelberg, *Market Structure and Equilibrium*. Springer, 2011.

[9] Y. A. Korilis, A. A. Lazar, and A. Orda, "Achieving network optima using Stackelberg routing strategies," *IEEE/ACM Trans. Netw.*, vol. 5, no. 1, pp. 161–173, 1997.

[10] T. Roughgarden, "Stackelberg scheduling strategies," *SIAM J. Comput.*, vol. 33, no. 2, pp. 332–350, 2004.

[11] M. Tambe, *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*. Cambridge University Press, 2011.

[12] M. Brückner and T. Scheffer, "Stackelberg games for adversarial prediction problems," in *17th ACM Int. Conf. Knowl. Discov. Data Min.*, 2011, pp. 547–555.

[13] J. Marecki, G. Tesauro, and R. Segal, "Playing repeated Stackelberg games with unknown opponents," in *11th Int. Conf. Auton. Agents Multiagent Syst.*, vol. 2, 2012, pp. 821–828.

[14] A. Blum, N. Haghtalab, and A. D. Procaccia, "Learning optimal commitment to overcome insecurity," in *Neural Inf. Process. Syst. 2014*, 2014, pp. 1826–1834.

[15] M. S. Kang, S. B. Lee, and V. D. Gligor, "The Crossfire attack," in *2013 IEEE Symp. Secur. Priv.*, 2013, pp. 127–141.

[16] K.-T. Fang, R. Li, and A. Sudjianto, *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC, 2005.

[17] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*. Springer, 1998.

[18] A. S. Morse, D. Q. Mayne, and G. C. Goodwin, "Applications of hysteresis switching in parameter adaptive control," *IEEE Trans. Automat. Contr.*, vol. 37, no. 9, pp. 1343–1354, 1992.

[19] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Prentice Hall, 1996.

[20] C. Henry, "An existence theorem for a class of differential equations with multivalued right-hand side," *J. Math. Anal. Appl.*, vol. 41, no. 1, pp. 179–186, 1973.

[21] A. Kastner-Maresch, "Implicit Runge–Kutta methods for differential inclusions," *Numer. Funct. Anal. Optim.*, vol. 11, no. 9-10, pp. 937–958, 1990.

[22] D. Shevitz and B. Paden, "Lyapunov stability theory of nonsmooth systems," *IEEE Trans. Automat. Contr.*, vol. 39, no. 9, pp. 1910–1914, 1994.