2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Game-Theoretic Deception Methods for Perfectly and Bounded Rational Stealthy Attackers

Filippos Fotiadis, Kyriakos G. Vamvoudakis

Abstract— **In this paper, we consider a system that is under the effect of multiple stealthy attackers, whose inputs we design using perfectly and imperfectly rational game-theoretic approaches. The goal of the attackers is to steer the state of the system as far as possible from the origin, so as to disrupt the nominal objective of system regulation. However, to remain stealthy, the attackers must ensure that the total magnitude of their inputs remains below a certain threshold, otherwise they are at risk of being exposed to a detection mechanism that monitors the system. To derive the optimal attack policies for the attackers, we interpret the aforementioned setup as a constrained game, and we solve it in two cases: in the first case, we assume that the attackers are perfectly rational and operate on the Nash equilibrium, which we derive in closed-form; and in the second case, we assume that the attackers are imperfectly rational, and we design two models of bounded rationality as a means to capture their different levels of rationality. Under certain conditions, it is proved that the corresponding bounded rationality models converge to a Nash equilibrium as the levels of rationality increase. Simulations demonstrate the efficiency of the derived attack policies in both the perfectly and the imperfectly rational case.**

I. Introduction

Cyber-physical systems (CPS) are systems of high heterogeneity and complexity, comprising multiple digital and physical components that interact with one another through a variety of communication channels. Due to their ability to incorporate complex structures, CPS can be traced in a large number of real-world settings, such as in the automotive industry [1] and in smart grids [2]. In that respect, they are an enticing target for adversaries who may want to create confusion, disruption, and performance deterioration.

An efficient way for an attacker to interfere with the operation of a CPS is by launching an actuation attack [3], [4], also often referred to as false data-injection or deception attack. This attack introduces perturbations in the CPS' control input through interference with its software, hardware or communication channels, and can thus directly affect the system's performance. It is often designed using tools from game and optimal control theory, so as to maximize damage to the CPS while conforming to constraints that guarantee that the attack remains undetected[1] [5]–[9]. However, when multiple attackers are simultaneously attacking the system, the problem of optimal stealthy attack design becomes more challenging: on the one hand, mere (single-player) optimal control design becomes inapplicable in such cases, owing to the inherent multiplayer nature of the problem; on the other hand, conventional multiplayer game-theoretic solutions, which rely on the notion of the Nash equilibrium, will yield unsatisfactory results if some attackers are imperfectly rational[2]. This is because, if at least one imperfectly rational attacker exists, all optimality guarantees provided by playing the equilibrium no longer hold.

To alleviate the restrictive assumption of the Nash equilibrium, namely that all players participating in the game are perfectly rational, alternative solutions concepts can be drawn from *bounded rationality* theory [10], [11]. The main idea of bounded rationality is that, since some agents may not operate on the Nash equilibrium of the game, it might be beneficial to model these players using behavioral models constructed from human engineering and real-world experiments. Towards this direction, [12] employed a bounded rationality model, known as cognitive hierarchy, to model imperfectly rational attackers launching undetectable sensor attacks on a CPS. Interestingly, it was shown that as the levels of rationality increase, cognitive hierarchy converged to the Nash equilibrium of the game, but all of the analysis was restricted to a static game setup and did not consider actuation attacks. The design of bounded-rational attack strategies was also considered in [13], [14], where optimal actuation attacks were designed against a CPS described by a dynamical system. However, both of these works focused on setups where only one adversary tries to disrupt the operation of the CPS. Clearly, in multiplayer settings, the problem of optimal attack design is still an open one.

Motivated by this gap in the literature, in this paper we consider the problem of designing optimal undetectable actuation attacks against a CPS, in a game-based setting where more than one attacker exists. The goal of each attacker is to maximize the squared norm of the CPS state – so as to disrupt the nominal objective of regulating the state to the origin – while also avoiding being revealed to an attack detection mechanism that monitors the system. The problem is solved using two distinct solution concepts: in the first one, all attackers are assumed to be perfectly rational, hence the solution is a Nash equilibrium that we derive in closed form; and in the second one, the attackers are assumed to be imperfectly rational, and two models of bounded rationality

F. Fotiadis and K. G. Vamvoudakis are with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. Email: {ffotiadis, kyriakos}@gatech.edu.

This work was supported in part, by ARO under grant No. W911NF-19−1−0270, by Minerva under grant No. N00014−18−1−2874, by NSF under grant Nos. CAREER CPS-1851588, S&AS-1849198, and SATC-2231651, and by the Onassis Foundation-Scholarship ID: F ZQ 064 − 1/2020 − 2021.

[1]Unnoticed by a detection mechanism [4] that monitors the system.

[2]An imperfectly rational player is one whose policy is not dictated by the Nash equilibrium.

are employed to identify their behavior. It is proved that these models of bounded rationality, under certain conditions, converge to the game's Nash equilibrium.

Notation: The sets $\mathbb{R}$ and $\mathbb{N}$ denote the real and natural numbers (including zero), respectively, while $\mathbb{N}_+$ denotes the set of non-zero natural numbers. For a vector $x \in \mathbb{R}^n$ and a symmetric matrix $F \in \mathbb{R}^{n\times n}$, we denote $\|x\|_F^2 = \frac{1}{2}x^{\mathrm{T}}Fx$, while $\|x\|_\infty$ will denote the infinity norm of x. If $l \in \{1,\dots,n\}$, then $[x]_l$ will denote the l-th entry of x. For a matrix $Z \in \mathbb{R}^{n\times n}$, $\mathrm{diagv}(Z) \in \mathbb{R}^n$ will denote a vector equivalent to the diagonal of Z. Using the Iverson bracket notation, $\mathrm{sgn}(x) := -[x < 0] + [x > 0]$ is the signum function. We denote as $\mathbb{1}_n \in \mathbb{R}^n$ a vector of ones.

II. Problem Formulation

Consider the continuous-time system:

$$\dot{x}(t) = Ax(t) + B\sum_{i\in\mathcal{N}} a_i(t),\ x(0) = x_0,\ t \geqslant 0, \quad (1)$$

where $x(t) \in \mathbb{R}^n$ denotes the state vector, $a_i(t) \in \mathbb{R}^m$ denotes an actuation attack originating from attacker $i \in \{1,\dots,N\} =: \mathcal{N}$, and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ denote the system's state and input dynamics matrices. For the purposes of this formulation, it is assumed without loss of generality that no defending input $u(t) \in \mathbb{R}^m$ appears in (1). Such an input can be considered part of the matrix A in cases where it moves the poles of the system to the open left half plane, or it can be absorbed in the state $x(t)$ using standard exponential matrix formulas.

The attackers are malicious in nature, and their purpose is to disrupt the system's nominal operation as much as possible within their capabilities. In many applications, such a nominal operation is equivalent to regulating the system's state to the origin within some time interval $T > 0$. Accordingly, the purpose of the malicious attackers here is to steer the system's state as far from the origin as possible. This objective can be captured by the maximization, by each attacker $j \in \mathcal{N}$, of:

$$J(a_j;\ a_{-j}) = \|x(T)\|_F^2, \quad (2)$$

where $x : [0,T] \to \mathbb{R}^n$ is subject to the dynamics (1), $F > 0$ is a weighting matrix that scales for possibly different units of measurement between the states, and $a_{-j} = \{a_1,\dots,a_{j-1},a_{j+1},\dots,a_N\}$ denotes the collective input of all attackers in $\mathcal{N}\backslash\{j\}$.

Remark 1. In this work, we will focus on deriving attack policies that are optimal with respect to (2), and are thus defined only over $t \in [0,T]$. Nevertheless, similar to [5], [8], any such attack policies can be implemented in a receding horizon fashion to derive a feedback policy, defined over all $t \geqslant 0$. □

In an entirely defenseless system, the optimal choice for the attackers with respect to the objective function (2) would consist of collectively injecting arbitrarily large inputs to the system, and driving the state substantially far from the origin. In the present framework, however, this is not assumed to be the case: the system is equipped with an attack detection mechanism that monitors variations of the system's state and is able to detect any possible such attempts by the attackers. The attackers must thus design their attacks properly and remain below the detection threshold that would trigger the attack detection mechanism. The following assumption is considered in this regard.

Assumption 1. The attackers remain undetectable if $\|\sum_{i\in\mathcal{N}} a_i(t)\|_\infty \leqslant \Delta,\ \forall t \in [0,T]$, where $\Delta > 0$ is a threshold known by the attackers. □

Remark 2. Certain attack policies with magnitude larger than Δ may still be undetectable in practice. In that respect, Assumption 1 is a sufficient but not necessary condition, stating that undetectability is at least guaranteed when the threshold Δ is not crossed. □

The rest of the paper is focused on deriving attack policies independently for each attacker $j \in \mathcal{N}$, which maximize the objective function (2) while satisfying the undetectability condition of Assumption 1. Since each attacker has its own objective function, this leads to a game-theoretic setup, in which we will consider two cases: in the first case, the attackers will be assumed to be perfectly rational, able to compute and operate on the Nash equilibrium of the game; and in the second case, the attackers will be considered to be imperfectly rational, i.e., not operating on the equilibrium.

Before we proceed, for the purpose of exposition, we define $\mathcal{A} := \{a : [0,T] \to \mathbb{R}^m \mid a \text{ is measurable}\}$ as the space of functions in which the attack policies will belong.

III. The Perfectly Rational Case

In this section, we will assume that the attackers are perfectly rational, implying that each of the attackers can reason perfectly about the policies that the other attackers will choose, and that each attacker is able to compute and operate on the Nash equilibrium of the game. Such an equilibrium is a tuple of policies that maximizes (2) for each attacker, while remaining within the undetectability constraint of Assumption 1.

Definition 1. A collection of attack policies $\{a_1^\star,\ a_2^\star,\dots,a_N^\star\} \in \mathcal{A}^N$ is a Nash equilibrium if for all $j \in \mathcal{N}$

$$J(a_j^\star;\ a_{-j}^\star) \geqslant J(a_j;\ a_{-j}^\star),\ \forall a_j \in \mathcal{A},$$

where $\{a_j^\star;\ a_{-j}^\star\}$ and $\{a_j;\ a_{-j}^\star\}$ satisfy the undetectability constraint of Assumption 1. □

More formally, such an equilibrium $\{a_1^\star,\ a_2^\star,\dots,a_N^\star\} \in \mathcal{A}^N$ is a solution of:

$$\begin{aligned}
\max_{a_j\in\mathcal{A}}\quad & J(a_j;\ a_{-j}) = \|x(T)\|_F^2,\\
\text{s.t.}\quad & \dot{x}(t) = Ax(t) + B\Big(a_j(t) + \sum_{i\in\mathcal{N}\backslash\{j\}} a_i(t)\Big),\\
& \Big\|a_j(t) + \sum_{i\in\mathcal{N}\backslash\{j\}} a_i(t)\Big\|_\infty \leqslant \Delta,\ \forall t \in [0,T],\\
& x(0) = x_0,
\end{aligned} \quad (3)$$

where the maximization takes place for each $j \in \mathcal{N}$, thus yielding a set of coupled optimization problems, i.e., a game. Note that in (3), as is natural in almost all equilibrium-based solution concepts, each player is assumed to know perfectly the action of the other players, hence being *perfectly rational*. To solve (3) for the equilibrium, we will resort to tools from dynamic non-cooperative game theory. The following theorem summarizes this approach.

Theorem 1. *Let $\{a_1^\star,\ a_2^\star, \dots, a_N^\star\} \in \mathcal{A}^N$ be a Nash equilibrium to the game* (3)*. Let $c_j : [0,T] \to \mathbb{R}^m$, $j \in \mathcal{N}$, be such that $\sum_{j\in\mathcal{N}}[c_j(t)]_l = \Delta$ for all $l = 1, \dots, m$. If there exists a root $\rho \in \mathbb{R}^n$ to the equation:*

$$\rho = e^{AT}x_0 + \Delta \int_0^T e^{A(T-\tau)} B \operatorname{sgn}(B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-\tau)} F\rho) \mathrm{d}\tau, \quad (4)$$

such that no entry of $B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho$ vanishes identically over a sub-interval of $t \in [0,T]$, then for one such root ρ:

$$a_j^\star(t) = c_j(t) \cdot \operatorname{sgn}(B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho). \quad (5)$$

Proof. Define the Hamiltonian of each attacker $j \in \mathcal{N}$ as:

$$H_j(x, \lambda_j, a_j;\ a_{-j}) = \lambda_j^{\mathrm{T}} \left(Ax + B\left(a_j + \sum_{i\in\mathcal{N}\setminus\{j\}} a_i \right) \right)$$

where $\lambda_j : [0,T] \to \mathbb{R}^n$ denotes the costate, and $x : [0,T] \to \mathbb{R}^n$ the state trajectory satisfying the dynamics in (3). To compute an equilibrium, it is necessary that the costate satisfies the adjoint equation:

$$\dot{\lambda}_j(t) = -\frac{\partial H_j}{\partial x} = -A^{\mathrm{T}} \lambda_j(t),$$

which, when integrated backwards over $t \in [0,T]$, yields:

$$\lambda_j(t) = e^{A^{\mathrm{T}}(T-t)} \lambda_j(T). \quad (6)$$

Additionally, it is necessary that the costate satisfies the transversality condition:

$$\lambda_j(T) = \frac{\|x(T)\|_F^2}{\partial x(T)} \Longrightarrow \lambda_j(T) = Fx(T). \quad (7)$$

Combining (6)-(7), we obtain:

$$\lambda_j(t) = e^{A^{\mathrm{T}}(T-t)} Fx(T). \quad (8)$$

Next, consider the following attack strategies:

$$\hat{a}_j(t) = c_j(t) \cdot \operatorname{sgn}(B^{\mathrm{T}} \lambda_j(t)) = c_j(t) \cdot \operatorname{sgn}(B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} Fx(T)), \quad (9)$$

where $c_j : [0,T] \to \mathbb{R}^m$ are such that $\sum_{j\in\mathcal{N}}[c_j(t)]_l = \Delta$ for all $l = 1, \dots, m$, and the multiplication between c_j and the sgn term in (9) takes place entry-wise. It can be seen that with these strategies, it will hold that:

$$\left(\hat{a}_j + \sum_{i\in\mathcal{N}\setminus\{j\}} \hat{a}_i \right) = \Delta \cdot \operatorname{sgn}(B^{\mathrm{T}} \lambda_j)$$

and $\lambda_j = \lambda_i$ for all $i, j \in \mathcal{N}$. Hence, these attack strategies satisfy the undetectability constraint in (3). In addition, for these strategies, it can be verified by inspection that

$$H_j(x, \lambda_j, \hat{a}_j;\ \hat{a}_{-j}) = \max_{\{a_j, a_{-j}\}\in\mathcal{A}^N} H_j(x, \lambda_j, a_j;\ a_{-j}),$$

and thus also:

$$H_j(x, \lambda_j, \hat{a}_j;\ \hat{a}_{-j}) = \max_{a_j\in\mathcal{A}} H_j(x, \lambda_j, a_j;\ \hat{a}_{-j}),$$

where it is implied that the maximums are taken over the policies that satisfy the undetectability constraint in (3). Therefore, the proposed strategies form a Nash equilibrium over the Hamiltonians H_j, and thus satisfy all of the necessary conditions to qualify for a Nash equilibrium of the game (3).

Finally, we need to rewrite (9) in a causal manner. In this direction, note that with $a_j = \hat{a}_j$:

$$\dot{x}(t) = Ax(t) + \Delta B \operatorname{sgn}(B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} Fx(T)),$$

hence:

$$x(T) = e^{AT}x_0 + \Delta \int_0^T e^{A(T-\tau)} B \operatorname{sgn}(B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-\tau)} Fx(T)) \mathrm{d}\tau. \quad (10)$$

Solving this equation for $x(T)$ and plugging it to (9) yields a causal expression for the attack policies $\hat{a}_j$. However, the above analysis makes sense only if no component of $B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} Fx(T)$ vanishes identically over a sub-interval of $t \in [0,T]$, where $x(T)$ is a root of (10). ∎

While Theorem 1 does provide a closed-form expression for the Nash equilibrium of the attackers' game, it can be seen that this equilibrium is valid only if no entry of the vector $B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho$ vanishes identically over a time sub-interval of $t \in [0,T]$. In the opposite case, the tuple of policies (9) may not necessarily form an equilibrium. This is because the vector $\lambda_j^{\mathrm{T}} B$ in the Hamiltonian, $\forall j \in \mathcal{N}$, will have an entry identically equal to zero, hence the maximization of the Hamiltonian with respect to a_j will provide no information on how to determine the Nash equilibrium. In other words, the game will present a singularity.

The cases where the singularity appears, are those where each attacker may have multiple possible optimal policies to choose from, irrespective of what the other attackers select to do. For example, in the trivial case where $B = 0$ (in which $B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho \equiv 0$), all possible attack policies form a Nash equilibrium. Hence, a controllability assumption will naturally exclude such singular cases, similar to how it does in the minimum-time problem [15]. In this direction, let us decompose the matrix B as:

$$B = [b_1\ b_2\ \dots\ b_m],$$

where each column $b_l \in \mathbb{R}^n$, $l = 1, \dots, m$, corresponds to an actuator of system (1). The following proposition provides conditions that exclude the presence of singularities in the attackers' game, hence guaranteeing that the equilibrium of the game will be given by the tuple of policies (5).

Proposition 1. *Let $(A,\ b_l)$ be a controllable pair for all $l = 1, \dots, m$, and $x_0 \neq 0$. Then, if $\rho \in \mathbb{R}^n$ is a root of* (4)*, no entry of $B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho$ vanishes identically over a time sub-interval of $t \in [0,T]$.*

Proof. Picking up from the proof of Theorem 1, suppose that some entry $l \in \{1, \dots, m\}$ of $\lambda_j^{\mathrm{T}} B$ vanishes identically over

some time sub-interval $[t_1, t_2] \subset [0, T]$. Then, $[\lambda_j^{\mathrm{T}} B]_l = 0$, or equivalently $\lambda_j^{\mathrm{T}} b_l = 0$ for all $t \in [t_1, t_2]$. Therefore, over $[t_1, t_2]$ it must also hold that:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}(\lambda_j^{\mathrm{T}} b_l) = 0 \Longrightarrow \lambda_j^{\mathrm{T}} A b_l = 0,\\
&\qquad\vdots\\
&\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}}(\lambda_j^{\mathrm{T}} b_l) = 0 \Longrightarrow \lambda_j^{\mathrm{T}} A^{n-1} b_l = 0,
\end{aligned}$$

hence $\lambda_j^{\mathrm{T}}[b_l \ A b_l \ A^{n-1} b_l] = 0$. Since (A, b_l) is controllable, this may hold only if $\lambda_j(t) = 0$ for all $t \in [t_1, t_2]$, and using the fact that $\lambda_j(t) = e^{A^{\mathrm{T}}(T-t)} F x(T)$ from (8), we obtain $x(T) = 0$. Plugging this in (10) yields $x_0 = 0$, which is contradicting. Hence, no entry of $\lambda_j^{\mathrm{T}} B$ vanishes identically over some time sub-interval of $[0, T]$. ■

Remark 3. Apart from a controllability condition, Proposition 1 also requires that $x_0 \neq 0$ for (5) to form a Nash equilibrium. This is again due to the fact that, at $x_0 = 0$, multiple optimal attack policies exist. □

IV. The Imperfectly Rational Case

In this section, we will assume that the attackers are imperfectly rational, implying that they do not collectively operate on the Nash equilibrium. This can happen in cases where not every agent knows the decision-making mechanism of one another, or if some agents are not aware of the existence of one another, or generally in cases where some agents are not able to find the Nash equilibrium [11]. In all of these scenarios, it does not make sense for an intelligent attacker to use the Nash equilibrium solution (5) of the perfectly rational case, because this solution no longer guarantees optimality with respect to (2).[3] Rather, it would be more beneficial for the attacker to identify every other attacker using a model of bounded rationality, and consequently employ an imperfectly rational policy [16], [17]. In what follows, we will consider two models of bounded rationality for modeling imperfectly rational attackers, namely level-k thinking and cognitive hierarchy [10], [11], [14].

A. Level-k Thinking

In level-k thinking, different levels of rationality are recursively defined, where each level $k \in \mathbb{N}$ corresponds to the number of steps of reasoning that an agent of that level performs. In particular, a player of cognitive level k is assumed to reason one step ahead of a player of cognitive level-$(k-1)$, in the sense that they best respond to that player. A more concrete description of the level-k thinking model follows next.

Level-0 *policy*: The level-0 policy, also known as an *anchor* policy, is usually chosen by assuming that a level-0 agent is naive [11]. Accordingly, it is natural to impose here that a level-0 attacker, being completely naive, assumes that the rest of the attackers do not exist, i.e., that their control input is zero [14]. Under this assumption, a level-0 attacker chooses their policy $a^0 \in \mathcal{A}$ by solving the optimal control problem:

$$\begin{aligned}
\max_{a \in \mathcal{A}} \quad & J(a;\ 0) = \|x(T)\|_F^2,\\
\text{s.t.} \quad & \dot{x}(t) = Ax(t) + Ba(t),\\
& \|a(t)\|_\infty \leqslant \Delta,\ \forall t \in [0, T],\\
& x(0) = x_0.
\end{aligned} \tag{11}$$

Level-k *policies*: Unlike the level-0 policies, the level-k policies $a^k \in \mathcal{A}$ for $k \neq 0$ are dependent on the structure of the thinking model. That is, for $k \in \mathbb{N}_+$, level-k attackers suppose that the rest of the attackers are level-$(k-1)$, and employ level-$(k-1)$ policies. Under this assumption, a level-k attacker, $k \neq 0$, chooses their policy a^k by solving the optimal control problem:

$$\begin{aligned}
\max_{a \in \mathcal{A}} \quad & J(a;\ \{a^{k-1}, \ldots, a^{k-1}\}) = \|x(T)\|_F^2,\\
\text{s.t.} \quad & \dot{x}(t) = Ax(t) + B\Big(a(t) + (N-1)a^{k-1}(t)\Big),\\
& \|a(t) + (N-1)a^{k-1}(t)\|_\infty \leqslant \Delta,\ \forall t \in [0, T],\\
& x(0) = x_0.
\end{aligned} \tag{12}$$

Apparently, to compute a level k policy, one needs to compute the level $k-1$ policy first, which in turn requires computing the level $k-2$ policy first, and this reasoning goes all the way down to the level 0 "anchor" policy. For this reason, this model of bounded rationality is also often referred to as *recursive reasoning* [18].

Remark 4. Notably, the optimization problems (11)-(12) are no longer games, because the policies of the rest of the attackers are considered fixed. □

While level-k thinking is a well-known and straightforward procedure in modeling imperfectly rational agents, it can be a poor choice within the present game setup. This is indicated in the following theorem.

Theorem 2. *If there exists a root $\rho \in \mathbb{R}^n$ to the equation* (4) *such that no entry of the vector $B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho$ vanishes identically over a time sub-interval of $t \in [0, T]$, then the optimal level-k policies, for each $k \in \mathbb{N}$, are given by*

$$a^k(t) = \Delta_k \cdot \operatorname{sgn}(B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho) \tag{13}$$

for one such root ρ, where $\Delta_k \in \mathbb{R}$ satisfies the difference equation:

$$\Delta_k = -(N-1)\Delta_{k-1} + \Delta,\ \Delta_0 = \Delta. \tag{14}$$

Proof. The level-0 policy a^0 is the solution of the optimal control problem (11). Following a similar line of analysis to that of Theorem 1, it is straightforward to derive that $a^0(t) = \Delta \cdot \operatorname{sgn}(B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho)$.

The level-k policy a^k, with $k \in \mathbb{N}_+$, is the solution of the optimal control problem (12). Performing the input transformation $z(t) = a(t) + (N-1)a^{k-1}(t)$, this problem

[3] An agent has no incentive to deviate from the Nash equilibrium only if all other agents operate on this equilibrium. If this is not the case, then an incentive to deviate exists.

turns into:

$$\begin{aligned}\max_{z\in\mathcal{A}} \quad & J(z;\,0) = \|x(T)\|_F^2\,,\\ \text{s.t.} \quad & \dot{x}(t) = Ax(t) + Bz(t),\\ & \|z(t)\|_\infty \leqslant \Delta,\ \forall t\in[0,T],\\ & x(0) = x_0,\end{aligned}$$

which is exactly the same as (11), and hence its optimal solution is given by $z^\star(t) = a^0(t) = \Delta\cdot\operatorname{sgn}(B^{\mathrm{T}}e^{A^{\mathrm{T}}(T-t)}F\rho)$, where ρ is a root of (4). Reverting back to the original problem, the level-k policy can be derived as $a^k(t) = a^0(t) - (N-1)a^{k-1}(t)$. Since a^k is a linear combination of a_0 and a^{k-1}, and since $a^0(t) = \Delta\cdot\operatorname{sgn}(B^{\mathrm{T}}e^{A^{\mathrm{T}}(T-t)}F\rho)$, it follows that $a^k(t) = \Delta_k\cdot\operatorname{sgn}(B^{\mathrm{T}}e^{A^{\mathrm{T}}(T-t)}F\rho)$, where $\Delta_k = -(N-1)\Delta_{k-1} + \Delta$ and $\Delta_0 = \Delta$. ∎

Notice that (14) can be viewed as a discrete-time dynamical system in the variable k, which will be strictly stable if $N = 1$, marginally stable if $N = 2$, and unstable if $N \geqslant 3$. This means that if there are $N \geqslant 3$ attackers, the control input of a level-k attacker increases geometrically as the level k increases. Apparently, level-k thinking can behave irregularly in the present framework.

B. Cognitive Hierarchy

One way to alleviate the irregularity of level-k thinking is to consider what is known as *cognitive hierarchy* [10]. In cognitive hierarchy, each agent of level k does not assume that all of the rest of the agents are level $k-1$, as in level-k thinking; rather, it assigns a proportion of them to each level, ranging from level 0 up to $k-1$. Specifically, if f denotes a probability mass function over $\mathbb{N}$, then these proportions can be defined as:

$$p_k(\kappa) = \frac{f(\kappa)}{\sum_{i=0}^{k-1} f(i)},\ \kappa\in\{0,1,\ldots,k-1\}. \tag{15}$$

It is a common choice to select $f(\cdot)$ to represent a Poisson distribution, as it has been shown through experiments that the proportion of agents that are level $k-1$ usually decreases as k becomes larger [10], [19], [20]. Accordingly, with this model, $f(\cdot)$ will be given by:

$$f(\kappa) = \frac{\lambda^\kappa e^{-\lambda}}{\kappa!}, \tag{16}$$

where $\lambda > 0$ denotes the variance and the mean of the Poisson distribution. Having defined these proportions, the level-k policies in the cognitive hierarchy framework are derived as follows.

Level-0 policy: The level-0 policy $\bar{a}_0 \in \mathcal{A}$ in cognitive hierarchy is exactly the same as in level-k thinking. That is, a level 0 attacker ignores the existence of the rest of the attackers, and chooses its policy by solving the optimal control problem (11).

Level-k policies: Different from level k thinking, a level k attacker in cognitive hierarchy, with $k\in\mathbb{N}_+$, does not assume that the rest of the attackers are all level $k-1$. Rather, it assumes that their level ranges from 0 to $k-1$, and specifically, for each $\kappa\in\{0,\ldots,k-1\}$, a proportion $p_k(\kappa)$ of the attackers is level κ. Accordingly, a level-k policy $\bar{a}_k\in\mathcal{A}$ in cognitive hierarchy corresponds to the solution of the optimal control problem:

$$\begin{aligned}\max_{a\in\mathcal{A}} \quad & \|x(T)\|_F^2\,,\\ \text{s.t.} \quad & \dot{x}(t) = Ax(t) + B\bigg(a(t) + \sum_{\kappa=0}^{k-1} p_k(\kappa)(N-1)\bar{a}^\kappa(t)\bigg),\\ & \|a(t) + \sum_{\kappa=0}^{k-1} p_k(\kappa)(N-1)\bar{a}^\kappa(t)\|_\infty \leqslant \Delta,\ \forall t\in[0,T],\\ & x(0) = x_0.\end{aligned}$$

Evidently, like level-k thinking, cognitive hierarchy is also a recursive reasoning model, where the computation of each level k policy requires computing the level $k-1$ policy first. In what follows, we provide the closed-form solution for the level-k policies in cognitive hierarchy.

Theorem 3. *If there exists a root $\rho\in\mathbb{R}^n$ to the equation* (4) *such that no entry of the vector $B^{\mathrm{T}}e^{A^{\mathrm{T}}(T-t)}F\rho$ vanishes identically over a time sub-interval of $t\in[0,T]$, then the optimal level-k policies in cognitive hierarchy, for each $k\in\mathbb{N}$, are given by:*

$$\bar{a}^k(t) = \bar{\Delta}_k\cdot\operatorname{sgn}(B^{\mathrm{T}}e^{A^{\mathrm{T}}(T-t)}F\rho) \tag{17}$$

for one such root ρ, where $\bar{\Delta}_k\in\mathbb{R}$ satisfies the recursion:

$$\bar{\Delta}_k = -\sum_{\kappa=0}^{k-1} p_k(\kappa)(N-1)\bar{\Delta}_\kappa + \Delta,\ \bar{\Delta}_0 = \Delta. \tag{18}$$

Proof. The proof is similar to that of Theorem 2 and it is, thus, omitted. ∎

Next, we prove the initial claim that motivated the use of cognitive hierarchy in lieu of level-k thinking, namely that cognitive hierarchy does not lead to unreasonably large attack input values as $k\to\infty$.

Proposition 2. *The level-k policies $\bar{a}^k\in\mathcal{A}$ in cognitive hierarchy have a well-defined point-wise limit $\bar{a}^\infty\in\mathcal{A}$.*

Proof. Denote $S_k = \sum_{i=0}^{k-1} f(i)$. From (18), for each $k\in\mathbb{N}_+$ we have:

$$\begin{aligned}\bar{\Delta}_k =& -\sum_{\kappa=0}^{k-1} p_k(\kappa)(N-1)\bar{\Delta}_\kappa + \Delta\\ =& -\frac{S_{k-1}}{S_k}\sum_{\kappa=0}^{k-2} p_{k-1}(\kappa)(N-1)\bar{\Delta}_\kappa\\ & -(N-1)p_k(k-1)\bar{\Delta}_{k-1} + \Delta\\ =& \frac{S_{k-1}}{S_k}(\bar{\Delta}_{k-1} - \Delta) - (N-1)p_k(k-1)\bar{\Delta}_{k-1} + \Delta.\end{aligned}$$

Hence, denoting $e_k = \bar{\Delta}_k - \Delta$, we obtain:

$$e_k = \left(\frac{S_{k-1}}{S_k} - (N-1)p_k(k-1)\right)e_{k-1} - (N-1)\Delta p_k(k-1),$$

which simplifies to:

$$e_k = \left(1 - Np_k(k-1)\right)e_{k-1} - (N-1)\Delta p_k(k-1).$$

From the definition of p_k in (15)-(16), there exists $k^\star \in \mathbb{N}$, such that if $k \geqslant k^\star$ then $|1 - Np_k(k-1)| \leqslant 1$. Hence, for $k \geqslant k^\star$:

$$|e_k| - |e_{k-1}| \leqslant (N-1)\Delta p_k(k-1).$$

Summing over $k \geqslant k^\star$, we obtain for any $n \geqslant k^\star$:

$$\begin{aligned}|e_n| - |e_{k^\star-1}| &\leqslant \sum_{k=k^\star}^{n} (N-1)\Delta p_k(k-1) \\ &\leqslant \sum_{k=k^\star}^{n} (N-1)\Delta \frac{f(k-1)}{f(0)} \\ &= \sum_{k=k^\star}^{n} (N-1)\Delta \frac{\lambda^{k-1}}{(k-1)!}.\end{aligned}$$

The summation at the right-hand side converges as $n \to \infty$, hence e_k, and thus $\bar{\Delta}_k$, is a bounded sequence. Next, letting $n, m \in \mathbb{N}_+$ be such that $m \leqslant n$, we have:

$$\begin{aligned}\bar{\Delta}_n &= -\sum_{\kappa=0}^{n-1} p_n(\kappa)(N-1)\bar{\Delta}_\kappa + \Delta \\ &= -\frac{S_m}{S_n}\sum_{\kappa=0}^{m-1} p_m(\kappa)(N-1)\bar{\Delta}_\kappa \\ &\quad -\sum_{\kappa=m}^{n-1} p_n(\kappa)(N-1)\bar{\Delta}_\kappa + \Delta.\end{aligned}$$

Hence

$$\bar{\Delta}_n = \frac{S_m}{S_n}(\bar{\Delta}_m - \Delta) - \sum_{\kappa=m}^{n-1} p_n(\kappa)(N-1)\bar{\Delta}_\kappa + \Delta,$$

which yields the error equation:

$$e_n = \frac{S_m}{S_n} e_m - \sum_{\kappa=m}^{n-1} p_n(\kappa)(N-1)\bar{\Delta}_\kappa.$$

From this equation, one can obtain the inequality

$$|e_n - e_m| \leqslant \left(1 - \frac{S_m}{S_n}\right)|e_m| + \sum_{\kappa=m}^{n-1} p_n(\kappa)(N-1)|\bar{\Delta}_\kappa|.$$

Note that $\frac{S_m}{S_n} \to 1$ and $\sum_{\kappa=m}^{n-1} p_n(\kappa) \to 0$ as $n, m \to \infty$. Hence, since e_k and $\bar{\Delta}_k$ are bounded sequences as previously proved, the right-hand side tends to 0 as $n, m \to \infty$, implying that e_k is a Cauchy sequence on $\mathbb{R}$. Therefore, by the completeness of $\mathbb{R}$, e_k converges as $k \to \infty$ to some well-defined real limit, hence $\bar{\Delta}_k$ also converges to some well-defined real limit $\bar{\Delta}_\infty \in \mathbb{R}$. Finally, it follows that for all $t \in [0, T]$, $\lim_{k\to\infty} \bar{a}^k(t) = \bar{a}^\infty(t) = \bar{\Delta}_\infty \mathrm{sgn}(B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho)$, with $\bar{a}^\infty \in \mathcal{A}$, where ρ solves (4). ∎

It should be noted that the limit of the cognitive hierarchy policies $\bar{a}^\infty$ does not necessarily correspond to a Nash equilibrium. This is rather expected: the purpose of cognitive hierarchy is not to approximate the Nash equilibrium, but to generate a set of policies that model bounded rational agents. Still, if the parameter λ of the Poisson distribution is chosen appropriately, it can be proved that the limiting policy $\bar{a}^\infty$ is such that $\{\bar{a}^\infty, \ldots, \bar{a}^\infty\}$ form a Nash equilibrium.

Proposition 3. *If $\lambda = \frac{1}{N-1}$, then $\{\bar{a}^\infty, \ldots, \bar{a}^\infty\}$ form a Nash equilibrium for the game* (3)*.*

Proof. It follows from (18) that $\bar{\Delta}_0 = \Delta$ and $\bar{\Delta}_1 = -(N-2)\Delta$. In addition $p_2(0) = \frac{1}{1+\lambda}$ and $p_2(1) = \frac{\lambda}{1+\lambda}$, hence

$$\bar{\Delta}_2 = \left(-\frac{1}{1+\lambda}(N-1) + \frac{\lambda}{1+\lambda}(N-1)(N-2) + 1\right)\Delta.$$

If $\lambda = \frac{1}{N-1}$, it can be verified that $\bar{\Delta}_2 = \frac{\Delta}{N}$. Next, for any $k \in \mathbb{N}_+$, following the same line of analysis as in the proof of Proposition 2, we can obtain the difference equation:

$$\bar{\Delta}_k = \frac{S_{k-1}}{S_k}(\bar{\Delta}_{k-1} - \Delta) - (N-1)p_k(k-1)\bar{\Delta}_{k-1} + \Delta,$$

which, using the fact that $\frac{S_{k-1}}{S_k} + p_k(k-1) = 1$, is equivalent to

$$\bar{\Delta}_k = \bar{\Delta}_{k-1} - \frac{S_{k-1}}{S_k}\Delta - Np_k(k-1)\bar{\Delta}_{k-1} + \Delta.$$

If $\bar{\Delta}_{k-1} = \frac{\Delta}{N}$ then it follows from this equation that also $\bar{\Delta}_k = \frac{\Delta}{N}$. Therefore, if $\lambda = \frac{1}{N-1}$, then $\bar{\Delta}_k = \frac{\Delta}{N}$ for all $k \geqslant 2$, and thus $\bar{a}^\infty = \frac{\Delta}{N} \cdot \mathrm{sgn}(B^{\mathrm{T}} e^{A^{\mathrm{T}}(T-t)} F\rho)$, where ρ solves (4). Finally, since $\sum_{j\in\mathcal{N}} \frac{\Delta}{N} = \Delta$, it follows from Theorem 1 that $\{\bar{a}^\infty, \ldots, \bar{a}^\infty\}$ form a Nash equilibrium for the game (3). ∎

C. Level of Intelligence Estimation

The bounded rationality models of level-k thinking and cognitive hierarchy provide a database of level-k policies, in the sense that they define a distinct policy for each attacker at each level of rationality $k \in \mathbb{N}$. This database of policies can then be used by an intelligent attacker to estimate the level of rationality of the rest of the attackers, and thus model their behavior. If this estimation is completed successfully, then the attacker can solve its optimization (3) with perfect knowledge of the other attackers' strategies.

To this end, let us define the database of the policies for an attacker, whose level of rationality is $k \in \mathbb{N}_+$, as $\phi := [a^1 \ a^2 \ \ldots \ a^{k-1}]^{\mathrm{T}}$. Let $p_0^l, p_1^l, \ldots, p_{k-1}^l$ be the belief of this level-k attacker that the l-th attacker, $l \in \mathcal{N}$, is level $0, 1, \ldots, k-1$ respectively. Defining $p^l := [p_0^l \ p_1^l \ \ldots \ p_{k-1}^l]^{\mathrm{T}}$, this attacker's belief about the level of the l-th attacker can be updated through the constrained least-squares optimization

$$\min_{p^l \in \mathbb{R}^k} \mathbb{E}_{\kappa,l}\left(\int_t^{t'} \|a^\kappa(\tau) - a_l(\tau)\|_{I_m}^2 \,\mathrm{d}\tau\right) + \left\|p^l - p_b^l\right\|_Q^2, \tag{19}$$

over the probability simplex

$$\mathbb{1}_k^{\mathrm{T}} p^l = 1, \ p^l \geq 0. \tag{20}$$

In the optimization (19), based on the belief of the intelligent attacker about the level of the attacker l, we have $\mathbb{E}_{\kappa,l}\left(\int_t^{t'} \|a^\kappa(\tau) - a_l(\tau)\|_{I_m}^2\right)\mathrm{d}\tau = \sum_{\kappa=0}^{k-1} p_\kappa^l \left(\int_t^{t'} \|a^\kappa(\tau) - a_l(\tau)\|_{I_m}^2\right)\mathrm{d}\tau$. In addition, p_b^l is the initial belief of the intelligent attacker about attacker l, $Q > 0$ forces a bias towards the initial belief, and $T \geqslant t' > t \geqslant 0$

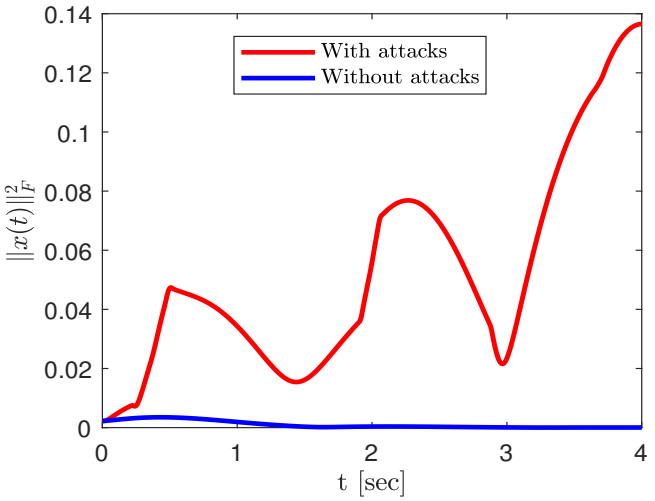

Fig. 1. The evolution of the state norm $\|x(t)\|_F^2$ when the system is affected by attackers operating on the Nash equilibrium, compared to when the system is under no attack.

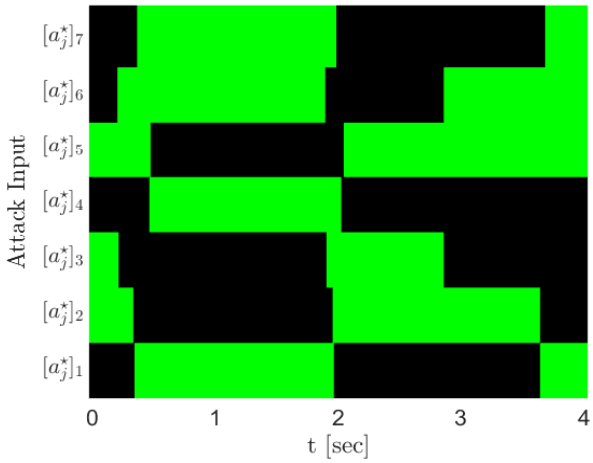

Fig. 2. The evolution of each entry of the attack inputs, when the attackers operate on the Nash equilibrium. Green color indicates the corresponding attack entry takes the positive value $\frac{\Delta}{3}$, while black color indicates it takes the negative value $-\frac{\Delta}{3}$.

are sampling instances. In a moving horizon formulation, one may set p_b^l equal to the probability p^l computed in a previous iteration of the receding horizon framework. The constrained optimization (19)-(20) is equivalent to

$$\begin{aligned} \min_{p^l \in \mathbb{R}^k} \quad & \frac{1}{2} p^{l\mathrm{T}} Q p^l + \left(-Q p_b^l - \Phi \mathrm{A}_l + \frac{1}{2} \Phi\Phi \right)^{\mathrm{T}} p^l \\ \text{s.t.} \quad & \mathbb{1}_k^{\mathrm{T}} p^l = 1, \; p^l \succeq 0, \end{aligned} \tag{21}$$

where $\Phi\Phi := \int_t^{t'} \mathrm{diagv}\left(\phi(\tau)\phi^{\mathrm{T}}(\tau)\right) \mathrm{d}\tau$ and $\Phi\mathrm{A}_l := \int_t^{t'} \phi(\tau) a_l(\tau) \mathrm{d}\tau$. Note that (21) is a convex quadratic program, hence it can be solved in polynomial time by using interior point methods [21]. Projected gradient methods may also be used to increase efficiency, with the projection operator onto the probability simplex defined as in [22].

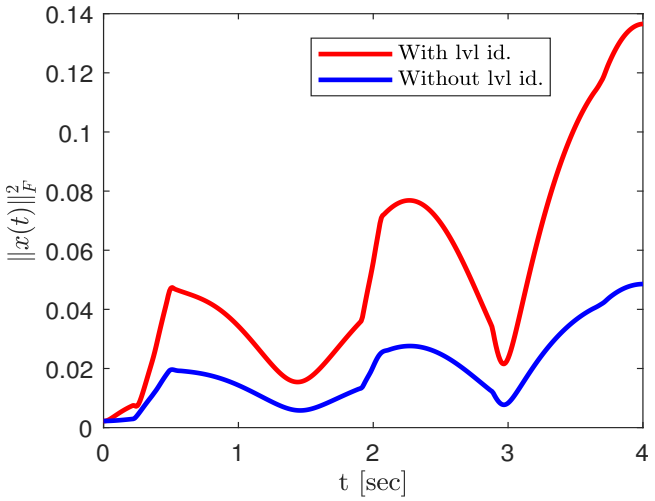

Fig. 3. The evolution of the state norm $\|x(t)\|_F^2$ when the system is affected by bounded rational attackers of cognitive level $1, 2$ and 4. The red line indicates the case where the intelligent level 4 attacker uses its beliefs about the other attackers' levels and issues a corrective input to the system, whereas the blue line indicates the case where the attacker commits to its level 4 policy.

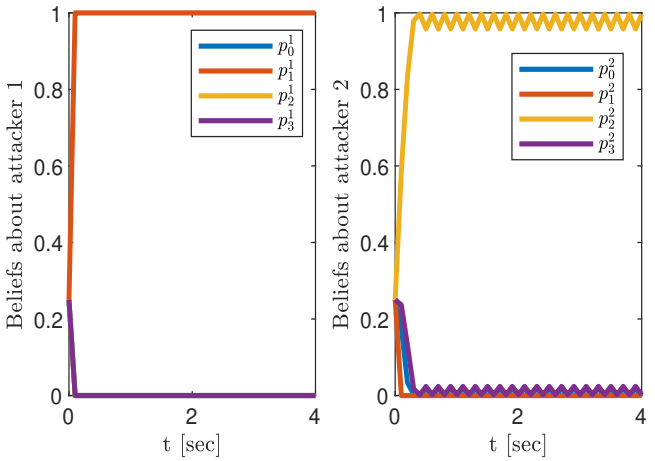

Fig. 4. The evolution of the beliefs of the intelligent attacker about the cognitive level of the other two attackers.

V. Simulations

We perform simulations on the Aero-Data Model in Research Environment (ADMIRE) benchmark aircraft [23], with system matrices given by:

$$A = \begin{bmatrix} -1.0649 & 0.0034 & -0.0000 & 0.9728 & 0.0000 \\ 0.0000 & -0.2492 & 0.0656 & -0.0000 & -0.9879 \\ 0.0000 & -22.5462 & -2.0457 & -0.0000 & 0.5432 \\ 8.1633 & -0.0057 & -0.0000 & -1.0478 & 0.0000 \\ 0.0000 & 1.7970 & -0.1096 & 0.0000 & -0.4357 \end{bmatrix},$$

$$B = \begin{bmatrix} -0.0062 & -0.0072 & 1.2456 & 2.7172 & -0.7497 \\ -0.0062 & 0.0072 & -1.2456 & 2.7172 & 0.7497 \\ -0.0709 & 0.0039 & -10.6058 & -2.4724 & -0.4923 \\ -0.1172 & 0.0188 & -9.2345 & -4.0101 & -1.1415 \\ -0.1172 & -0.0188 & 9.2345 & -4.0101 & 1.1415 \\ -0.0709 & -0.0039 & 10.6058 & -2.4724 & 0.4923 \\ 0.0003 & 0.0003 & 5.3223 & 0.0108 & -3.7367 \end{bmatrix}^{\mathrm{T}},$$

and with initial condition $x_0 = 0.001 \cdot \begin{bmatrix} 1 & -1 & 1 & -1 & 1 \end{bmatrix}^{\mathrm{T}}$. The system is assumed to be under the effect of a stabilizing gain $K \in \mathbb{R}^{7\times 5}$ by a defender, which redefines the system's plant matrix A into the Hurwitz matrix $A-BK$. In addition, it is under the effect of $N = 3$ attackers' inputs, whose purpose is to disrupt the system's stabilization by maximizing the cost (2). Note that this system satisfies the conditions of Proposition 1.

We initially simulate the system in the case where all of the attackers are perfectly rational, employing the control law (5), with $c_j(t) = \frac{\Delta}{3}$ for all $j = 1, 2, 3$. The weighting matrix of the cost (2) is chosen as $F = I_5$, the detection threshold is $\Delta = 0.005$, and the optimization horizon is $T = 4$ sec. With these parameters, the root of (4) is found as:

$$\rho = \begin{bmatrix} 9.9 \cdot 10^{-6} & -0.0081 & 0.1360 & 2.62 \cdot 10^{-5} & -0.0087 \end{bmatrix}^{\mathrm{T}}.$$

The results are then shown in Figures 1-2. It can be particularly seen from Figure 1 that, when the attackers operate on the Nash equilibrium, the state norm is indeed maximized at the end of the optimization horizon, and is much larger than when the system is under no attack. In addition, Figure 2, which illustrates the value of the attack inputs, verifies Proposition 1 since no entry of the attack inputs becomes identically zero on any time sub-interval of $t \in [0, 4]$.

Next, we simulate the system in the case where the attackers are imperfectly rational. Specifically, attacker 1 is level 1, attacker 2 is level 2, and attacker 3 is level 4 according to the cognitive hierarchy model, while $\lambda = 0.8$ is the Poisson parameter of cognitive hierarchy. To model the lower level attackers, the intelligent attacker 3 solves the optimization (21) with $Q = 10^{-5}I$ in a receding horizon to updates its beliefs p^1_κ and p^2_κ, $\kappa = 0, \ldots, 3$. Based on these beliefs, the attacker then issues a corrective attack input to the system, so that the total attack input imitates the (optimal) total input of the equilibrium case. It can be seen from Figure 4 that the intelligent attacker has successfully identified the levels of the other two attackers. In addition, Figure 3 shows that the corrective feedback is beneficial when compared to the case where this feedback was not issued.

VI. Conclusion

We considered a system under the effect of multiple attackers, whose objective was to drive the system's state far from the origin while remaining stealthy. To obtain optimal policies for them, two cases were considered: in the first, the attackers were assumed to be perfectly rational, operating on the Nash equilibrium of the game; and in the second, they were assumed to be imperfectly rational, and they were modeled using a bounded rationality framework. Simulations verified the efficacy of the derived attack strategies in both cases.

Future work includes extending the present setup to cases where the attack inputs can affect the system's sensor measurements.

References

[1] J. Kim, H. Kim, K. Lakshmanan, and R. Rajkumar, "Parallel scheduling for cyber-physical systems: Analysis and case study on a self-driving car," in *Proceedings of the ACM/IEEE 4th international conference on cyber-physical systems*, pp. 31–40, 2013.

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

[3] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of cps security," *Annual reviews in control*, vol. 47, pp. 394–411, 2019.

[4] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE transactions on automatic control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[5] J. P. Hespanha and S. D. Bopardikar, "Output-feedback linear quadratic robust control under actuation and deception attacks," in *2019 American Control Conference (ACC)*, pp. 489–496, IEEE, 2019.

[6] S. C. Anand and A. M. Teixeira, "Stealthy cyber-attack design using dynamic programming," in *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 3474–3479, IEEE, 2021.

[7] V. Sawant and R. Wisniewski, "Evaluating criticality of nodes in consensus network under false data injection attack," *IEEE Control Systems Letters*, vol. 7, pp. 1435–1440, 2023.

[8] F. Fotiadis and K. G. Vamvoudakis, "Concurrent receding horizon control and estimation against stealthy attacks," *IEEE Transactions on Automatic Control*, vol. 68, no. 6, pp. 3712–3719, 2023.

[9] C. Wu, X. Li, W. Pan, J. Liu, and L. Wu, "Zero-sum game-based optimal secure control under actuator attacks," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3773–3780, 2020.

[10] C. F. Camerer, T.-H. Ho, and J. K. Chong, "Behavioural game theory: thinking, learning and teaching," in *Advances in understanding strategic behaviour*, pp. 120–180, Springer, 2004.

[11] T. Strzalecki, "Depth of reasoning and higher order beliefs," *Journal of Economic Behavior & Organization*, vol. 108, pp. 108–122, 2014.

[12] A. Kanellopoulos and K. G. Vamvoudakis, "Bounded rationality in byzantine sensors under attacks," *IEEE Transactions on Automatic Control*, vol. 67, no. 7, pp. 3606–3613, 2021.

[13] S. C. Anand, A. M. Teixeira, and A. Ahlén, "Risk assessment of stealthy attacks on uncertain control systems," *arXiv preprint arXiv:2106.07071*, 2021.

[14] A. Kanellopoulos and K. G. Vamvoudakis, "Non-equilibrium dynamic games and cyber–physical security: A cognitive hierarchy approach," *Systems & Control Letters*, vol. 125, pp. 59–66, 2019.

[15] D. Liberzon, *Calculus of variations and optimal control theory: a concise introduction*. Princeton university press, 2011.

[16] A. Sanjab and W. Saad, "On bounded rationality in cyber-physical systems security: Game-theoretic analysis with application to smart grid protection," in *2016 Joint Workshop on Cyber-Physical Security and Resilience in Smart Grids (CPSR-SG)*, pp. 1–6, IEEE, 2016.

[17] N. Li, D. W. Oyler, M. Zhang, Y. Yildiz, I. Kolmanovsky, and A. R. Girard, "Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems," *IEEE Transactions on control systems technology*, vol. 26, no. 5, pp. 1782–1797, 2017.

[18] Z. Dai, Y. Chen, B. K. H. Low, P. Jaillet, and T.-H. Ho, "R2-b2: Recursive reasoning-based bayesian optimization for no-regret learning in games," in *International Conference on Machine Learning*, pp. 2291–2301, PMLR, 2020.

[19] C. F. Camerer, "Behavioral game theory: Psychological limits on strategic cognition," in *International Journal of Psychology*, vol. 51, pp. 11–11, ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD 2-4 PARK SQUARE, MILTON PARK ..., 2016.

[20] C. F. Camerer, "Behavioral game theory: Plausible formal models that predict accurately," *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 157–158, 2003.

[21] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[22] W. Wang and M. A. Carreira-Perpinán, "Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application," *arXiv preprint arXiv:1309.1541*, 2013.

[23] X. Yu and J. Jiang, "Hybrid fault-tolerant flight control system design against partial actuator failures," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 4, pp. 871–886, 2011.