This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3102464, IEEE Transactions on Automatic Control

Bounded Rationality in Byzantine Sensors under Attacks

Aris Kanellopoulos, *Student Member, IEEE,* Kyriakos G. Vamvoudakis, *Senior Member, IEEE*

***Abstract*—In this paper, we investigate the behavior of agents with bounded rationality, attacking a set of stochastic sensors measuring the state of a binary event. The coordination problem between the attackers is formulated as a multi-player non-zero-sum one-shot game. Each attacker aims to maximize the probability that a certain detector will produce an erroneous estimate of the true event, while they remain stealthy. To better predict the outcome of this game, we categorize the players based on the number of strategic thinking steps they will take. Each level-k attacker behaves based on subjective beliefs of the others' behaviors, which are quantified via a Poisson distribution over the lower levels. The expected best responses for the type of each attacker are derived. The limiting conditions as the cognitive levels increase, as well as when the attackers fully coordinate, are shown to converge to the Nash equilibrium.**

***Index Terms*—Bounded reasoning, cognitive hierarchy, Byzantine problem.**

I. Introduction

Due to their complexity and their profound effect on modern infrastructure, e.g., in the automotive [1] and health [2] industries, cyber-physical systems (CPS) must be made safe and secure in their operation. This becomes more pressing in the presence of malicious agents that strive to corrupt, confuse or destabilize the system. It has become clear that software-oriented defense mechanisms, like firewalls, are incapable of securing CPS on their own, and as such, the control community has shifted its focus on developing rigorous mathematical tools that examine all the different layers of a complex system under attack [3]. These problems also emerge in estimation processes, where decisions must be made about the true state of the environment, given potentially corrupted measurements. Game theory offers a unified framework for analyzing the interactions between competing, selfish agents [4]. Due to its flexibility, it has been extensively used in CPS security applications, in order to develop defense and mitigation strategies for a variety of attack scenarios. However, the majority of the aforementioned works on the subject has leveraged relatively simplistic attack models, operating under the assumption of infinite rationality—allowing for the derivation of equilibrium solutions—and by disregarding the heterogeneity of the attackers on a system, as well as the complex interactions that take place among them.

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

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

Developing effective collaborations, when there are more than one agents, is a challenging problem, since there are cognitive limitations, which can be seen as a form of bounded rationality [5]. Several recent experimental and empirical studies [6] suggest that decision makers' initial responses to games often deviate systematically from equilibrium, and that structural non-equilibrium (e.g., cognitive hierarchy) models often out-predict equilibrium. In this sense, some player(s) may be "making mistakes," which is common. Thus, we leverage ideas from cognitive hierarchy as a tool to examine non-equilibrium behavior of boundedly rational players, in the context of the model introduced in [7] that captures a binary sensor system under Byzantine attacks.

Considering game-theoretic approaches to security, in [8], the intrusion detection problem is formulated as a non-cooperative game between an attacker and a defender of infinite intelligence. The formulation utilized in this work was introduced in [7], where the Byzantine attack problem was solved as a zero-sum game between *rational* opponents and closed-form Nash equilibria were computed under certain assumptions. This can be applicable in high-stakes security domains such as infrastructure protection where presumably the adversary will conduct careful surveillance and planning before attacking. However, there are some security domains where adversary may not act perfectly rational due to short planning windows or due to lower stakes associated with the attack. Security strategies generated under the assumption of a perfectly rational adversary might be significantly less effective that ones generated to tackle specific suboptimal attack patterns.

Therefore, addressing the boundedly rationality behavior exhibited by adversaries is a fundamental challenge for applying security games to a wide variety of domains. In the work of [9], the author describes several models of bounded rationality, both for single optimizers, as well as competing players through a combination of limited information about the environment and intractability of the optimal decision in complex scenarios. With the advent of behavioral game theory [5], more sophisticated models of irrational players were introduced. An important approach to bounded rationality in games is based on iterations of best responses. Specifically, in level-k models [6], the authors suggest that a player with k steps of strategic thinking optimally reacts to the assumption that everyone else has conducted $k-1$ steps. The same framework has been applied to a global entry game with private and public information [10]. The authors of [11], introduced cognitive hierarchy, according to which the players are categorized in certain cognitive types

according to the steps of strategic thinking, while each type is equipped with a distribution of beliefs over lower-type players. The method of cognitive hierarchy has been employed to investigate the problem of distributed up-link random access for the Internet of Things in [12]. Cognitive hierarchy was also used alongside reinforcement learning in training autonomous vehicles in [13]. Such approaches have been also developed for dynamic environments; the authors in [14] propose a dynamic model of bounded rationality based on prospect theory in a controls context while in [15] an iterative thinking framework is analyzed from a regret minimization point of view.

In securing a CPS from adversaries, it is important to develop accurate models of prediction, grounded on experimental results regarding human behavior. Specifically, the authors in [16] propose cumulative prospect theory to facilitate unmanned aerial vehicle path planning. In the context of estimation problems, the authors in [17] consider the problem of the interaction between several self-interested sensors under different solution concepts, such as Nash or Stackelberg equilibria. Furthermore, the need for security analysts to consider more complex models of attacker behavior is addressed in [18] where the authors examine experimental data that showcase the appearance of complex relations between attackers with both shared and competing objectives in the security domain. Similarly, in [19], the authors provide a survey that showcases the complexities of attacker behavior with the goal of achieving superior forecasting. Besides discovering, visualizing, and predicting multi-stage attacks, a method such as ours, will provide a framework to understand and profile behaviors of attackers.

Contributions: The contributions of the present work are threefold. Initially, we formulate the decentralized coordinated sensor attack as a non-zero sum (NZS) game between selfish attackers who seek a "balance" between the probability of successful attack and individual stealthiness. Subsequently, we show that the game is guaranteed to have a Nash equilibrium. This game is employed as a model for high-level binary decision processes, such as fault sensors and alarm systems, that are found in safety and mission–critical CPS. Second, we introduce a cognitive hierarchy approach to bounded rationality, by finding closed-form solutions to the internal maximization problem of each level-k attacker. The inclusion of non-equilibrium solution concepts allows us to gain insight into attack policies of selfish agents—either computational ones that are bounded by hardware constraints [20] or human ones bounded by cognitive limits [11]—which do not operate on the Nash equilibrium. To the best of our knowledge, this is the first time that a structured non-equilibrium framework is investigated in this context to analyze and predict attacking behaviors on binary sensors. As such, we derive a more potent prediction model driven by beliefs that may be constructed from previous attack scenarios, thus enabling the defending mechanism to better address realistic attacks. We finally highlight how the game evolves as the attackers increase their cognitive levels, and as they shift the weight of their decision-making process towards successfully attacking, rather than remaining stealthy.

Structure: The remainder of the paper is structured as follows. In Section II, we formulate the Byzantine problem. Section III describes the coordination game that is taking place between the decentralized attackers, and introduces competing elements between them. The structure of this game, as well as its equilibrium properties, are analyzed in Section IV. Bounded rationality models are developed by leveraging a cognitive hierarchy approach in Section V. Section VI discusses the convergence properties of the non-equilibrium game. Finally, Section VII concludes and discusses future work.

II. Problem Formulation

A. Binary Hypothesis Testing

Consider a set of n_{s} sensors measuring the value of a single binary event $X \in \mathcal{X} = \{0, 1\}$. In this work, we approach the binary estimation problem from the attackers' perspective in order to understand, analyze, and eventually predict, their behavior. The ultimate goal in this line of research is to develop mechanisms that will act as predictive models in adversarial environments in order to assist in the design of defense mechanisms.

Consider the value X of a binary random variable with Bernoulli distribution

$$P(X = 1) = 1 - P(X = 0) = p_{\text{event}} \in (0, 1). \tag{1}$$

The defender has access to the, potentially corrupted by noise or attacks, measurements of n_{s} sensors, Z_i, $i \in \{1, \ldots, n_{\mathrm{s}}\}$. In the case of sensors reporting strictly binary values and the sensors are homogeneous in their behavior, then it facilitates the analysis to consider the summation of the, potentially corrupted, reported values, $Z = \sum_{k=1}^{n_{\mathrm{s}}} Z_k$.

Assuming now that every sensor has the same probability of reporting the wrong event value due to stochastic faults in the absence of attacks, we have

$$Z = \begin{cases} R, & \text{wp } 1 - p_{\text{at}} \\ S + \sum_{i=1}^{n_{\mathrm{p}}} W_i, & \text{wp } p_{\text{at}}, \end{cases}$$

where, p_{at} is an *a-priori* estimation of the probability that the sensors are attacked and n_{p} is the number of adversarial agents. Also, we have that

$$R \sim \begin{cases} \text{Binom}(n_{\mathrm{s}}, p_{\text{err}}), & X = 0 \\ \text{Binom}(n_{\mathrm{s}}, 1 - p_{\text{err}}), & X = 1, \end{cases}$$

and

$$S \sim \begin{cases} \text{Binom}(n_{\mathrm{s}} - \sum_{i=1}^{n_{\mathrm{p}}} m_i, p_{\text{err}}), & X = 0 \\ \text{Binom}(n_{\mathrm{s}} - \sum_{i=1}^{n_{\mathrm{p}}} m_i, 1 - p_{\text{err}}), & X = 1, \end{cases}$$

where $\text{Binom}(\alpha, p_s)$ denotes the binomial distribution, quantifying the number of successes in a sequence of α experiments, each with p_s probability of success. Furthermore, p_{err} denotes the probability that a single sensor will report erroneously due to stochastic faults and it is a design parameter that quantifies the quality of the sensor and is assumed to be known *a-priori*. The distribution of the values W_i are determined by the attackers as a function of the real state of the world X. Specifically, they shape this distribution according to the

probability measure ω_{j0} when $X = 0$ and ω_{j1} when $X = 1$, each with support $I_j = \{0, \dots, m_j\}$, $\forall j \in \{1, \dots, n_\text{p}\}$, where m_j is the number of sensors that the attacker j has access to. We further assume that the subsets are disjoint, i.e., that different attackers are not able to affect the same sensor. In order to capture the interactions between them, we formulate a NZS game among the various attackers of unknown rationality. While all of them share the common goal of maximizing the probability of error, each of them wishes to remain as stealthy as possible. By assuming selfish attackers, we formulate the following NZS game.

B. Non-Zero Sum Game

In order to fully describe the NZS game, we define the sets contained in the tuple $\mathcal{G} = (\mathcal{O}, \mathcal{J}, \mathcal{S})$. By $\mathcal{O}$, we denote the players participating in the game, by $\mathcal{J}$ the rewards, and by $\mathcal{S}$ the attacking policies. In this scenario, the players are the malicious agents that seek to compromise the estimation process. We will consider the attackers operating in a decentralized fashion, without direct communication among themselves. Such attackers may represent human agents or decentralized bots. Specifically, the set $\mathcal{S} = S_1 \times S_2 \times \cdots \times S_p$ contains a tuple of policies for each player, i.e., $(\omega_{j0}, \omega_{j1}) \in S_j$, $\forall j \in \{1, \dots, n_\text{p}\}$, which determine the values of W_i. Thus, the sets S_j represent the action sets of the players.

Remark 1: Two elements describe the effect that an attacker has to the system. The support I_j quantifies the maximum number of measurements that are accessible to the j-th attacker. His decision-making process is expressed through the shaping of the distributions in S_j assuming that they has access to the real value of the event. This enables him to consider two separate distributions ω_{j0} and ω_{j1} as decision variables. □

We adopt the standard game-theoretic notation for an arbitrary set $\mathcal{Y} = Y_1 \times \cdots Y_\alpha$, such that $\mathcal{Y}_{-j} = Y_1 \times \cdots Y_{j-1} \times Y_{j+1} \times \cdots \times Y_\alpha$, e.g., $(\omega_{-j0}, \omega_{-j1}) \in \mathcal{S}_{-j}$ will denote the profile of strategies for every player apart from j. Also, the notation $s_j = (\omega_{j0}, \omega_{j1})$ will be used interchangeably to denote the strategy of the players. Each player of the NZS game tries to maximize a specific reward $J_j \in \mathcal{J}$, $\forall j \in \{1, \dots, n_\text{p}\}$ as follows,

$$J_j = J_j^\text{err} + J_j^\text{st}, \tag{2}$$

where J_j^err denotes the reward an attacker receives by forcing the estimator to make an erroneous decision. Based on the work of [7], this reward will be quantified by the probability of error, i.e.,

$$J_j^\text{err} = P\bigg(\big(\hat{X} = 0 \mid X = 1\big) \bigcup \big(\hat{X} = 1 \mid X = 0\big)\bigg), \tag{3}$$

with $\hat{X}$ being the decision of the estimator/detector. The reward J_j^st forces the attackers to remain stealthy and introduces a selfish element to the agents. Thus, while the attackers are cooperating in order to increase the probability of error, each of them wants to selfishly remain stealthy.

Remark 2: Even though they share a common goal, maximizing the probability of erroneous estimation, there has been extensive research that supports the emergence of competitive behavior in the cyber-crime community [18], [21], [22]. It has been found that, either for sociological reasons, e.g. recognition in the community, or due to more practical reasons, e.g. minimizing their own exposure, agents targeting the same system might do so while competing against each other. This behavior is captured by the inclusion of J_j^st. □

To allow for a general framework similar to [7], we denote by $\sigma_0,\ \sigma_1$ the distributions of S when X is 1 or 0 respectively and by $\rho_0,\ \rho_1$ the distributions of R. Furthermore, we introduce the estimator function $f : \mathbb{R} \to [0, 1]$, denoting the stochastic decision of the defender given a measurement $z \in \mathbb{R}$ as

$$\hat{X} = \begin{cases} 1, & \text{w.p. } f(z), \\ 0, & \text{w.p. } 1 - f(z). \end{cases}$$

We note that in this work, we consider the game from the attackers' standpoint. Thus, the estimator function itself is a parameter known to the attackers, as per Kerckhoffs's principle, according to which the defender cannot depend upon the system's obscurity to guarantee security [23]. For examples of estimator strategies in the context of Byzantine binary sensors, the reader is referred to [7]. Consequently, we can express J_j^err as a function of the attackers' policies by utilizing the law of total probability. As a result, we have for J_j^err (note that the derivation of this expression follows from the Appendix of [7] and is omitted due to space restrictions),

$$\begin{aligned} &J_j^\text{err}(\omega_{11}, \omega_{10}, \dots, \omega_{p1}, \omega_{p0}) = p_\text{event} + \\ &p_\text{at}(1 - p_\text{event}) \int_{I_1} \cdots \int_{I_{n_\text{p}}} \int_{\mathbb{R}} f(y + \sum_{i=1}^{n_\text{p}} w_i) \sigma_0(\mathrm{d}y) \prod_{i=1}^{n_\text{p}} \omega_{i0}(\mathrm{d}w_i) \\ &- p_\text{at} p_\text{event} \int_{I_1} \cdots \int_{I_{n_\text{p}}} \int_{\mathbb{R}} f(y + \sum_{i=1}^{n_\text{p}} w_i) \sigma_1(\mathrm{d}y) \prod_{i=1}^{n_\text{p}} \omega_{i1}(\mathrm{d}w_i) \\ &+ (1 - p_\text{at}) P(\neg \mathcal{E}_\text{at}), \end{aligned}$$

where p_at denotes the probability of the system being targeted by adversaries and $P(\neg \mathcal{E}_\text{at})$ the probability of an erroneous estimation when no attackers are involved. Since $P(\neg \mathcal{E}_\text{at})$, and the additive term p_event is not affected by the attackers' decisions, it will be omitted from the reward function.

Following [7], and in order to derive closed-form solutions to the optimization problems of the attackers—which would be impossible over the space of arbitrary measures—we leverage the delta function $\delta(z)$ that yields the decision vectors

$$\omega_{j0}(w) = \sum_{i=0}^{m_j+1} p_{ji} \delta(w - i),\ \omega_{j1}(w) = \sum_{i=0}^{m_j+1} q_{ji} \delta(w - i). \tag{4}$$

Thus, the problem of finding the best attack policy for every player $j, \in \{1, \dots, n_\text{p}\}$, amounts to finding the probabilities $p_{ji},\ q_{ji}$ of the jth attacker attacking i number sensors when the real event is 0 or 1, respectively.

To formulate J_j^st, we argue that an attacker is less exposed, and by extension more stealthy, the less they affects the normal operation of the sensors. To this end, we introduce the probability distributions $\tau_{j0},\ \tau_{j1}$, each with support on I_j, which denote the distributions of W_j when the attacker j does not decide to attack based on the true value of X. We

choose the Kullback-Leibler (KL) divergence as a metric of stealthiness, as

$$J_j^{\text{st}} = -\beta\Big((1-p_{\text{event}})D(\omega_{j0} \mid\mid \tau_{j0}) + p_{\text{event}} D(\omega_{j1} \mid\mid \tau_{j1})\Big), \tag{5}$$

where $D(a \mid\mid b)$ is the KL-divergence between distributions a and b, and β is a weighting parameter. Given that the distributions are expressed according to (4), the KL-divergence is $D(\omega_{j0} \mid\mid \tau_{j0}) = -\sum_{i=0}^{m_j} p_i \ln \frac{p_{ji}}{\tau_{i0}}$.

III. Existence of Nash Equilibria

In this section, we will analyze the behavior of the players given that the Nash equilibrium assumptions are satisfied. Specifically, the policies of the players are:

1. optimal, i.e., given a set of beliefs about the others' policies, every player deterministically follows the policy that maximizes the given reward,
2. consistent, i.e., all the players share the same beliefs about everyone's policy.

The game equilibrium is defined in the sense of Nash, i.e., satisfying the following inequalities $\forall(\omega_{j0}, \omega_{j1}) \in S_j,\ j \in \{1, \dots, n_{\text{p}}\}$,

$$J_j(\omega_{j0}^\star, \omega_{j1}^\star, \omega_{-j0}^\star, \omega_{-j1}^\star) \geqslant J_j(\omega_{j0}, \omega_{j1}, \omega_{-j0}^\star, \omega_{-j1}^\star) \tag{6}$$

where $(\omega_{j0}^\star,\ \omega_{j1}^\star) \in S_j$ denote the optimal policies of player j and $(\omega_{-j0}^\star,\ \omega_{-j1}^\star) \in \mathcal{S}_{-j}$, are the optimal policies of all the other players.

In order to facilitate the derivation of the Nash policies, we formulate the following lemma where, for ease of exposition, we define the decision tuple $s_j = (\omega_{j0}, \omega_{j1})$.

Lemma 1: The NZS game described by the reward functions (2), is a potential game with potential given by

$$\begin{aligned}
&\Phi(s_i, s_{-i}) = \\
&(1-p_{\text{err}}) \sum_{i_1=0}^{m_1} \cdots \sum_{i_{n_{\text{p}}}=0}^{m_{n_{\text{p}}}} p_{1i_1} \dots p_{n_{\text{p}} i_{n_{\text{p}}}} \int_{\mathbb{R}} f(z)\sigma_0(z - \sum_{k=1}^{n_{\text{p}}} i_k)\mathrm{d}z \\
&- p_{\text{err}} \sum_{i_1=0}^{m_1} \cdots \sum_{i_{n_{\text{p}}}=0}^{m_{n_{\text{p}}}} q_{1i_1} \dots q_{n_{\text{p}} i_{n_{\text{p}}}} \times \\
&\times \int_{\mathbb{R}} f(z)\sigma_1(z - \sum_{k=1}^{n_{\text{p}}} i_k)\mathrm{d}z + \sum_{k=0}^{n_{\text{p}}} J_k^{\text{st}}.
\end{aligned}$$

Proof. Initially, we consider a game with rewards given by

$$\begin{aligned}
V_j(s_j, s_{-j}) = (1-p_{\text{err}}) &\sum_{i_1=0}^{m_1} \cdots \sum_{i_j=0}^{m_j} \cdots \sum_{i_{n_{\text{p}}}=0}^{m_{n_{\text{p}}}} p_{1i_1} \dots p_{n_{\text{p}} i_{n_{\text{p}}}} \\
&\int_{\mathbb{R}} f(z)\sigma_0(z - \sum_{k=1}^{n_{\text{p}}} i_k)\mathrm{d}z \\
- p_{\text{err}} &\sum_{i_1=0}^{m_1} \cdots \sum_{i_j=0}^{m_j} \cdots \sum_{i_{n_{\text{p}}}=0}^{m_{n_{\text{p}}}} q_{1i_1} \dots q_{n_{\text{p}} i_{n_{\text{p}}}} \\
&\int_{\mathbb{R}} f(z)\sigma_1(z - \sum_{k=1}^{n_{\text{p}}} i_k)\mathrm{d}z.
\end{aligned} \tag{7}$$

Given policies $\tilde{s}_j,\ \forall j$, we can evaluate the cost and subtract from (7) to write

$$\begin{aligned}
&V_j(s_j, s_{-j}) - V_j(\tilde{s}_j, s_{-j}) = \\
&(1-p_{\text{err}}) \sum_{i_1=0}^{m_1} \cdots \sum_{i_j=0}^{m_j} \cdots \sum_{i_{n_{\text{p}}}=0}^{m_{n_{\text{p}}}} (p_{ji_j} - \tilde{p}_{ji_j}) \\
&\qquad p_{1i_1} \dots p_{n_{\text{p}} i_{n_{\text{p}}}} \int_{\mathbb{R}} f(z)\sigma_0(z - \sum_{k=1}^{n_{\text{p}}} i_k)\mathrm{d}z \\
&- p_{\text{err}} \sum_{i_1=0}^{m_1} \cdots \sum_{i_j=0}^{m_j} \cdots \sum_{i_{n_{\text{p}}}=0}^{m_{n_{\text{p}}}} (q_{ji_j} - \tilde{q}_{ji_j}) \\
&\qquad q_{1i_1} \dots q_{n_{\text{p}} i_{n_{\text{p}}}} \int_{\mathbb{R}} f(z)\sigma_1(z - \sum_{k=1}^{n_{\text{p}}} i_k)\mathrm{d}z,
\end{aligned}$$

which is constant $\forall i \in \{1, \dots, n_p\}$.

Consequently, the game defined by V is a potential game, with potential function given by

$$\begin{aligned}
&\Phi_1(s_i, s_{-i}) = \\
&(1-p_{\text{err}}) \sum_{i_1=0}^{m_1} \cdots \sum_{i_{n_{\text{p}}}=0}^{m_{n_{\text{p}}}} p_{1i_1} \dots p_{n_{\text{p}} i_{n_{\text{p}}}} \int_{\mathbb{R}} f(z)\sigma_0(z - \sum_{k=1}^{n_{\text{p}}} i_k)\mathrm{d}z \\
&- p_{\text{err}} \sum_{i_1=0}^{m_1} \cdots \sum_{i_{n_{\text{p}}}=0}^{m_{n_{\text{p}}}} q_{1i_1} \dots q_{n_{\text{p}} i_{n_{\text{p}}}} \int_{\mathbb{R}} f(z)\sigma_1(z - \sum_{k=1}^{n_{\text{p}}} i_k)\mathrm{d}z.
\end{aligned}$$

We now define another game with reward functions given by, $V'(s_i, s_{-i}) = J_i^{\text{st}}(s_i, s_{-i})$. This game is decoupled, since $V'(s_i, s_{-i}) = V'(s_i, \tilde{s}_{-i}),\ \forall i$, whose potential function is $\Phi_2(s_i, s_{-i}) = \sum_i^{n_p} V'(s_i, s_{-i})$.

Consequently, the coordination game between the attackers is a potential game with potential function given by, $\Phi(s_i, s_{-i}) = \Phi_1(s_i, s_{-i}) + \Phi_2(s_i, s_{-i})$, which completes the proof. ■

Corollary 1: The coordination game with rewards given by (2) has a Nash equilibrium.

Proof. The proof follows from [24], given a potential game with strategies in a compact set. ■

Remark 3: Corollary 1 shows existence of a Nash equilibrium. However, computing this strategy profile can be computationally, and cognitively, expensive. Therefore, it is natural to approach the problem from a behavioral game perspective, where the players are modeled as irrational agents. □

IV. Bounded Rationality

In this section, we will introduce the models of non-equilibrium solution concepts that will be used to analyze the sensor attack problem when players of bounded rationality are involved in the process.

A. Cognitive Types and Cognitive Hierarchy

We have shown in Section III, that the consistency with equilibrium profiles depends heavily on the optimality of each individual response, as well as, the mutual subjective belief consistency, i.e., how close each player can approximate (by measuring the cost) the true strategy of every other player.

In this work, we will focus on agents that optimize over their perception of the game, but who are heterogeneous in this perception. Towards this, we utilize the framework of cognitive hierarchy, by following the work of [11]. According to this bounded rationality approach, the players in a game are differentiated via a "cognitive level," which describes both their behavior, and their beliefs regarding the rest of the players. This will lead to heterogeneous strategies depending on the level of thinking of an agent, which is denoted by the appropriate superscript, i.e., for the strategy of a level-k player—himself indexed by j—will be denoted by $s_j^k = (\omega_{j0}^k, \omega_{j0}^k)$.

Remark 4: The emergence of non-equilibrium play in games has been extensively investigated [5]. Since the Byzantine problem can be utilized to model a plethora of different realistic scenarios, this apparent irrationality of the players can be attributed to different causes. Specifically, for human players, it might stem from limited cognitive ability, lack of time or incentive to succeed, while for autonomous agents, the computational complexity of computing Nash equilibria can lead to boundedly rational behavior. □

Level-0: In order to initialize the iterative best-response process, we consider the behavior of a "level-0" attacker, i.e., of an attacker that does not perform any steps of strategic thinking. This policy, denoted "anchor policy", has been shown to affect the decision making process of higher-level thinking players in some cases. Even in the presence of such sensitivity, these is no consensus over what ought to constitute a level-0 policy, with suggested approaches being uniform distributions over the action space, focal points of the game or even complex solutions, such as no-regret policies [25]. In this work, we assume that a level-0 attacker, indexed by j, uniformly randomizes over the number of sensors they can attack, leading to the decision vectors, $\omega_{j0}^0 = \omega_{j1}^0 = \frac{1}{m_j+1}\sum_{i=0}^{m_j}\delta(w-i)$.

Higher-level policies: According to the limited depth of reasoning approaches to bounded rationality, every player has a limit k to the number of strategic that they can perform. Moreover, in cognitive hierarchy, every level assigns a specific belief over the levels of the rest of the players. Following [11], we claim that a level-k player reasons that the rest of the players' levels are distributed according to a Poisson distribution over $l = \{1, \dots, k-1\}$. For example, consider the attacker j, belonging to cognitive level k. The iterative strategic thinking of this player is based on the belief that the probability of another player i belonging to a level $l < k$ is

$$g_k^j(l) = \frac{\frac{\lambda^l e^{-\lambda}}{l!}}{\sum_{h=0}^{k-1}\frac{\lambda^h e^{-\lambda}}{h!}}, \tag{8}$$

where λ is the Poisson parameter that uniquely defines the distribution. Following [11], we will model every player as deterministically optimal. Thus, each level-k thinker acts based on his best-response given his subjective beliefs.

Remark 5: Even though there is a plethora of different bounded rationality solution concepts for games—many of which consider models of subjective beliefs—in this paper we utilize the approach of [11], where those subjective beliefs follow a Poisson distribution. Such distributions—being univariate—are able to capture phenomena of bounded rationality with limited computational complexity. Specifically, by adjusting the mean of the Poisson distribution, one can shift the level over which most agents reside, while keeping the probability that agents are extremely low, or extremely high level, relatively small. Furthermore, those beliefs are assumed by each agent to be true during the design of their strategy. □

B. Iterative-Thinking Process

Level-1: For ease of exposition, we present the analysis of the approach for 3 attackers. All the results will be shown to generalize in a straightforward manner. Initially, let us consider a level-1 attacker that has access to m_1 sensors. Such an agent operates based on the assumption that the rest of the attackers, are level-0, i.e., they uniformly randomize their actions. The reward of this attacker, conditioned by the randomized policies is

$$\begin{aligned} J_1^1(s_1^1, s_{-1}^1) &= (1-p_{\text{err}})\sum_{i=0}^{m_1} p_{1i}\frac{1}{(m_2+1)}\frac{1}{(m_3+1)} \\ &\sum_{j=0}^{m_2}\sum_{k=0}^{m_3}\int_{\mathbb{R}} f(z)\sigma_0(z-i-j-k)\mathrm{d}z - \beta\sum_{i=0}^{m_1} p_{1i}\ln\frac{p_{1i}}{\tau_{i0}} \\ &- p_{\text{err}}\sum_{i=0}^{m_1} q_{1i}\frac{1}{(m_2+1)}\frac{1}{(m_3+1)} \\ &\sum_{j=0}^{m_2}\sum_{k=0}^{m_3}\int_{\mathbb{R}} f(z)\sigma_1(z-i-j-k)\mathrm{d}z - \beta\sum_{i=0}^{m_1} q_{1i}\ln\frac{q_{1i}}{\tau_{i1}}. \end{aligned} \tag{9}$$

Due to the independence of the decision vectors p_i, q_i as well as the independence of their corresponding terms in the reward function, it is possible to analyze them separately. Therefore, the level-1 attacker maximizes deterministically for p_i and q_i.

Let us denote by p_{1i}^1 and q_{1i}^1 the elements of the decision vectors there are derived as best-response strategies to level-0 players. As player 1 performs further strategic thinking steps, he needs to compute every other players' current level strategy as well as perform one more optimization step. This fact highlights the possible computational intractability of Nash equilibrium policies, which are infinite best-response paths for the joint strategy profile of all the players.

Higher-level of thinking: When computing any higher level strategy, agent 1 holds certain beliefs on the frequency of each cognitive level. On a multi-player, NZS game, this can be modeled through the following reward function, for arbitrary number of attackers n_{p} and arbitrarily large thinking iterations as

$$J_1^k(s_1^k, s_{-1}^k) = (1-p_{\text{err}})\sum_{i_1=0}^{m_1} p_{1i_1}^k$$

$$\Bigg(\sum_{h=0}^{k-1} g_k^j(h)\cdots\sum_{h=0}^{k-1} g(h)\sum_{i_2=0}^{m_2} p_{2i_2}^h\cdots\sum_{i_{\mathrm{p}}=0}^{m_{n_{\mathrm{p}}}} p_{\mathrm{p}i_{\mathrm{p}}}^h c_{0i}\Bigg)$$
$$-\beta\sum_{i=0}^{m_1} p_{1i}^2\ln\frac{p_{1i}^2}{\tau_{10}}$$
$$-p_{\mathrm{err}}\sum_{i=0}^{m_i} q_{1i}^k\Bigg(\sum_{h=0}^{k-1} g_k^j(h)\cdots\sum_{h=0}^{k-1} g_k^j(h)\sum_{i_2=0}^{m_2} q_{2i_2}^h\cdots\sum_{i_{\mathrm{p}}=0}^{m_{n_{\mathrm{p}}}} q_{\mathrm{p}i_{\mathrm{p}}}^h c_{1i}\Bigg)$$
$$-\beta\sum_{i=0}^{m_1} q_{1i}^2\ln\frac{q_{1i}^2}{\tau_{11}}, \tag{10}$$

where $c_{0i} = \int_{\mathbb{R}} f(z)\sigma_0(z-i-j-k)\mathrm{d}z$ and $c_{1i} = \int_{\mathbb{R}} f(z)\sigma_1(z-i-j-k)\mathrm{d}z$.

It is possible to find a closed form solution for the strategy of an attacker of level-k by solving the corresponding maximization problem.

Theorem 1: Consider an attacker, arbitrarily indexed as 1 of cognitive level-k, with a reward function given by (10). The best-response solution, expressed through the probabilities p_{1i} and q_{1i}, when the observed real event is $X=0$ and $X=1$ respectively, is given by

$$p_{1i} = \frac{\tau_{10}e^{\hat{c}_{0i}^k}}{\sum_{j=1}^{n_{\mathrm{p}}}\tau_{10}e^{\hat{c}_{0i}^k}}, \tag{11}$$

where

$$\hat{c}_{0i}^k = \sum_{h=0}^{k-1} g_k^j(h)\cdots\sum_{h=0}^{k-1} g_k^j(h)\sum_{i_2=0}^{m_2} p_{2i_2}^h\cdots\sum_{i_{\mathrm{p}}=0}^{m_{n_{\mathrm{p}}}} p_{\mathrm{p}i_{\mathrm{p}}}^h c_{0i}^k,$$

and

$$q_{1i} = \frac{\tau_{11}e^{\hat{c}_{1i}^k}}{\sum_{j=1}^{n_{\mathrm{p}}}\tau_{11}e^{\hat{c}_{1i}^k}}, \tag{12}$$

where

$$\hat{c}_{1i}^k = \sum_{h=0}^{k-1} g_k^j(h)\cdots\sum_{h=0}^{k-1} g_k^j(h)\sum_{i_2=0}^{m_2} q_{2i_2}^h\cdots\sum_{i_{\mathrm{p}}=0}^{m_{n_{\mathrm{p}}}} q_{\mathrm{p}i_{\mathrm{p}}}^h c_{1i}.$$

Proof. Initially, we formulate only the optimization over p_i to write

$$\max_{p_{1i}^k}\sum_{i=0}^{m_1} p_{1i}^k\hat{c}_{0i}^k - \beta\sum_{i=0}^{m_1} p_{1i}^2\ln\frac{p_{1i}^2}{\tau_{10}}$$
$$\text{subject to: }\sum_{i=0}^{m_1} p_{1i}^k = 1 \text{ and } p_{1i}^k \geqslant 0.$$

Thus, we can define the Lagrangian of the optimization in terms of the multipliers μ, and ν as

$$L = \sum_{i=0}^{m_1} p_{1i}^k\hat{c}_{01}^k - \beta\sum_{i=0}^{m_1} p_{1i}^2\ln\frac{p_{1i}^2}{\tau_{10}} + \mu\Bigg(\sum_{i=0}^{m_1} p_{1i}^k - 1\Bigg) + \sum_{i=0}^{m_1}\nu_i p_i.$$

We apply the Karush-Kuhn-Tucker conditions, which lead to a set of m_1+1 equations

$$\frac{\partial L}{\partial p_{1i}^k} = 0 \Rightarrow c_{01}^k - \beta(1-\ln\tau_{10}+\ln p_{1i}^2) + \mu + \nu_i = 0, \tag{13}$$

as well as the complementarity condition $\sum_{i=0}^{m_i}\nu_i p_i^k = 0$.

Since the term $\ln(p_i^k)$ will be undefined if any $p_i = 0$, we can determine that the optimal solution will be an interior point of the simplex. As a result $\nu_i = 0$, $\forall i$, and solving (13) for p_i^k yields, $p_i^k = \tau_{10}e^{\frac{\hat{c}_{01}^k}{\beta}}e^{\frac{\lambda}{\beta}-1}$. Since it holds that $\sum_{j=1}^{m_1} p_j^1 = 1$, we get, $p_i^1 = \frac{\tau_{10}e^{\frac{\hat{c}_{01i}^k}{\beta}}}{\sum_{j=0}^{m_1}\tau_{10}e^{\frac{\hat{c}_{01j}^k}{\beta}}}$. ∎

V. Discussion and Validation

A. Convergence to the Nash equilibrium

We can leverage now the special structure of the game to show that the cognitive hierarchy approach converges to the equilibrium as all the agents increase their cognitive abilities.

Lemma 2: For every potential game with infinite, compact action spaces, every improvement path, i.e., a sequence of profiles in which one player optimizes at each step, terminates arbitrarily close to an equilibrium.

Proof. The proof follows from [24]. ∎

Theorem 2: Consider the coordinated attack game played by attackers of bounded rationality. As all the attackers increase their cognitive level-k, the strategy profile converges to the Nash equilibrium.

Proof. The single parameter used in the Poisson distribution of the intelligent levels, i.e., λ, is the average level of all the players. Therefore, if $k\to\infty$ for all the players, then $\lambda\to\infty$. This leads to all the players having mutually consistent beliefs over everyone's strategy, i.e., $\exists\bar{g}(h),\ h\in\mathcal{N}$, such that, $\lim_{k\to\infty} g_k^j(h) = \bar{g}(h),\ \forall j\in\{1,\ldots,n_p\}$. Coupled with the deterministic optimality assumption, the iteration converges to a Nash equilibrium due to Lemma 2 and the properties of cognitive hierarchy [26]. ∎

B. Cooperation Between Attackers

In this section, we will analyze the behavior of the coordinating attackers when there is no cost of attacking/stealthiness. We note that in this case, the reward function for each attacker will be $V_j(s_j, s_{-j})$, given by (7). In such a scenario, it is possible to explicitly compute the Nash equilibrium of the game for agents of unbounded cognitive capabilities, given that the decision making is known and all the players are aware of the underlying game.

Theorem 3: Consider the problem of attack coordination with reward functions given by (7). Let us assume that,

$$\int_{\mathbb{R}} f(z)\sigma_0(z-a)\mathrm{d}z \leqslant \int_{\mathbb{R}} f(z)\sigma_0(z-b)\mathrm{d}z \Leftrightarrow a\leqslant b, \tag{14}$$

$$-\int_{\mathbb{R}} f(z)\sigma_1(z-a)\mathrm{d}z \leqslant -\int_{\mathbb{R}} f(z)\sigma_1(z-b)\mathrm{d}z \Leftrightarrow a\leqslant b, \tag{15}$$

then, the Nash equilibrium profile of strategies for this game is given $\forall j\in\{1,\ldots,n_{\mathrm{p}}\}$ by

$$p_{ji} = \begin{cases} 1, \text{ when } i = m_j, \\ 0, \text{ otherwise,}\end{cases} \tag{16}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3102464, IEEE Transactions on Automatic Control

and

$$q_{ji} = \begin{cases} 1, \text{ when } i = 0, \\ 0, \text{ otherwise.} \end{cases} \tag{17}$$

Proof. It can be seen that the game of attack coordination without the stealthiness reward is a game of identical interests, since every player maximizes over the same reward function (7). Consequently, by assumption (14) it holds that

$$\int_{\mathbb{R}} f(z)\sigma_0(z - i - \sum_{j\in\mathbb{I}_{-i}} m_j)\mathrm{d}z \leqslant \int_{\mathbb{R}} f(z)\sigma_0(z - m_i - \sum_{j\in\mathbb{I}_{-i}} m_j)\mathrm{d}z, \tag{18}$$

and

$$-\int_{\mathbb{R}} f(z)\sigma_1(z - i)\mathrm{d}z \leqslant -\int_{\mathbb{R}} f(z)\sigma_0(z)\mathrm{d}z, \tag{19}$$

for all attackers i.

Utilizing now the inequalities (18) and (19), we can find a bound of the reward function (7), as, $V_j(s_i^\star, s_{-i}^\star) \geqslant V_j(s_i, s_{-i}^\star),\ \forall j \in \{1, \ldots, n_\mathrm{p}\}$, which, by definition, constitutes a Nash equilibrium and completes the proof. ■

C. Simulation Results

In order to highlight the behavior shift as different parameters change, we utilize a set of $n_\mathrm{s} = 100$ sensors, where $n_\mathrm{p} = 3$ attackers have access to $m_1 = m_2 = m_3 = 10$ sensors each. The cost function is constructed according to (10) where the estimator strategy—known to the attackers—is a probabilistic majority voting mechanism as described in [7]. The anchor strategy of each attacker is a uniform distribution over their available actions while the probability of erroneous measurement is homogeneous throughout the sensors and is chosen to be $p_\mathrm{err} = 0.1$.

Figure 1 shows the different optimal actions for an attacker as their intelligence level increases. For this scenario, the distribution of intelligence levels led to a Poisson parameter value of $\lambda = 5$ and the attackers were considered relatively selfless, with the weight on stealthiness chose to be $\beta = 5 \times 10^{-3}$. It is worth noting that due to the high level of intelligence amongst the attackers, the optimal strategies converge to the fully cooperative Nash equilibrium. Figure 2, shows that the selfish component of an attacker's cost function—in our case the one corresponding to their stealthiness—dramatically shifts their behavior. We observe that the limiting behavior converges to full coordination as β goes to 0, a phenomenon that is also evident in Figure 3 where we present the KL divergence of the optimal policy of an attacker to the Nash equilibrium corresponding to perfect coordination. As the weight β decreases, the attackers tend to cooperate more.

The importance of accurate attacker modeling becomes apparent by examining Figure 4, which highlights a crucial phenomenon arising in NZS games. In this scenario, we present the accumulated cost of a high-level thinker, operating in a environments of different levels of intelligence. Although the increased level of this player shifts their strategy towards the Nash equilibrium, it can be seen that their optimality is

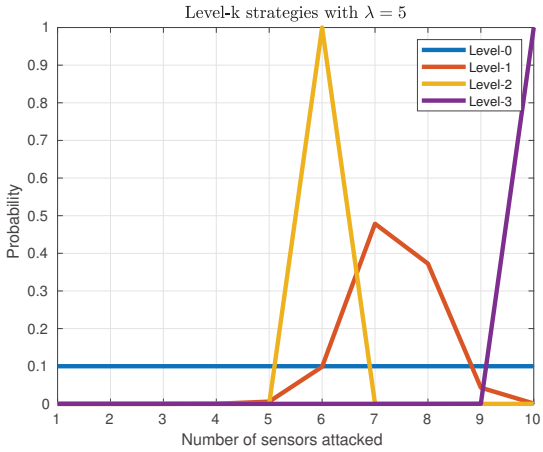

Fig. 1. The decision probabilities as the levels of an attacker increase for $\lambda = 5$. The strategy converges to the Nash equilibrium, and for small β, this is close to the probability of attacking all the available sensors.

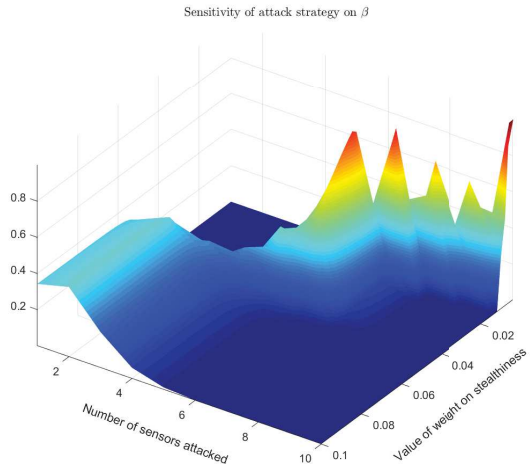

Fig. 2. The decision probabilities as the weight on stealthiness changes. It can be seen that the probability converges to a Dirac distribution on the maximum number of compromised sensors.

decreased when they "overthink" by believing that the rest of the players are of high-level as well, as is shown in the blue bars. This showcases a scenario where the high intelligence of a player may create erroneous beliefs due to an overestimation of the capabilities of their opponents. An important implication for a security scenario in this case is that Nash strategies might be inadequate in protecting a system, even against attackers of limited capabilities due to the "overthinking" effect of the equilibrium solutions.

VI. Conclusion and Future Work

In this work, we developed a framework to predict the way decentralized attacking agents with bounded cognitive capabilities attempt to compromise a set of faulty binary sensors. To this end, we leveraged results from behavioral game theory. Through a cognitive hierarchy approach, we analyzed the expected responses of the attackers with different intelligent levels via the solutions of iterated maximization

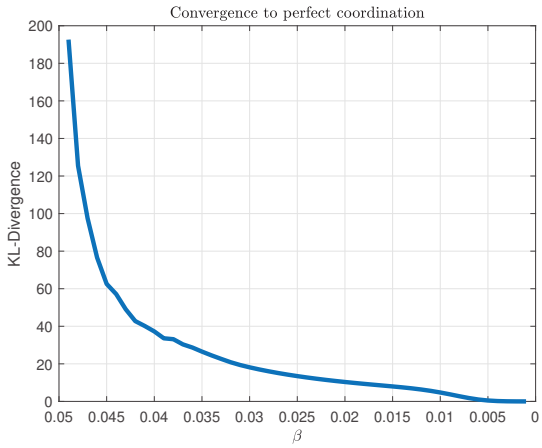

Fig. 3. The KL divergence of the attack strategies from perfect coordination, as their selfishness decreases.

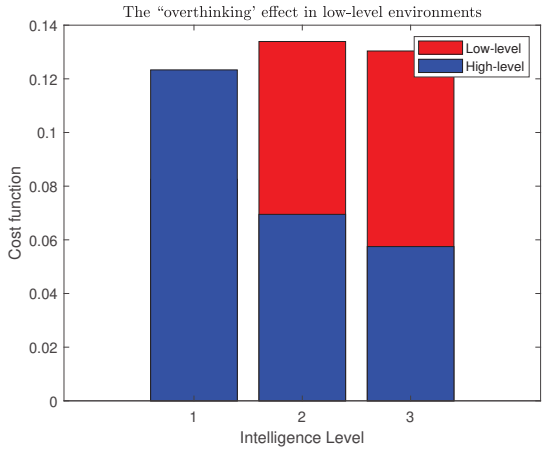

Fig. 4. The cost function of a highly intelligent player operating in environments where they believe that the attackers are of high-level and of low-level, respectively. It can be seen that if the player "overthinks," it is possible for them to play more complex strategies and acquire suboptimal rewards.

problems. Results from potential game theory were utilized to prove convergence of the attack strategies to the Nash equilibrium as the adversaries increase their intelligence level and cooperate to inflict the most damage to the system.

Future efforts will focus on extending the framework to repeated plays, allowing for the attackers to learn the cognitive abilities of each other, as well as in prediction models for sets of heterogeneous sensors with real-valued measurements. Accurate attack prediction models will be used to inform the optimal detection rule for the CPS in a realistic adversarial environment, in which the binary measurements can be derived as alarm signals by redundant fault–detection mechanisms. We will also further theoretically investigate the "overthinking" problem that emerges in bounded rationality contexts. Finally, the computational complexity of the derived expressions with respect to the problem parameters will be quantified in order to investigate connections between complexity and bounded rationality.

REFERENCES

[1] J. Kim, H. Kim, K. Lakshmanan, and R. 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*. ACM, 2013, pp. 31–40.

[2] I. Lee and O. Sokolsky, "Medical cyber physical systems," in *Design Automation Conference (DAC), 2010 47th ACM/IEEE*. IEEE, 2010, pp. 743–748.

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

[4] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.

[5] C. F. Camerer, *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2011.

[6] V. P. Crawford and N. Iriberri, "Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions?" *Econometrica*, vol. 75, no. 6, pp. 1721–1770, 2007.

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

[8] T. Alpcan and T. Basar, "A game theoretic approach to decision and analysis in network intrusion detection," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 3. IEEE, 2003, pp. 2595–2600.

[9] H. A. Simon, "Theories of bounded rationality," *Decision and organization*, vol. 1, no. 1, pp. 161–176, 1972.

[10] T. Kneeland, "Coordination under limited depth of reasoning," *Games and Economic Behavior*, vol. 96, pp. 49–64, 2016.

[11] C. F. Camerer, T.-H. Ho, and J.-K. Chong, "A cognitive hierarchy model of games," *The Quarterly Journal of Economics*, vol. 119, no. 3, pp. 861–898, 2004.

[12] N. Abuzainab, W. Saad, and H. V. Poor, "Cognitive hierarchy theory for heterogeneous uplink multiple access in the internet of things," in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 1252–1256.

[13] N. Li, D. Oyler, M. Zhang, Y. Yildiz, A. Girard, and I. Kolmanovsky, "Hierarchical reasoning game theory based approach for evaluation and testing of autonomous vehicle control systems," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 727–733.

[14] Y. Cheng and C. Langbort, "Hopf bifurcation in the informational nudging of boundedly rational decision makers," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 5761–5766.

[15] Z. Dai, Y. Chen, K. H. Low, P. Jaillet, and T.-H. Ho, "R2-b2: Recursive reasoning-based bayesian optimization for no-regret learning in games," *arXiv preprint arXiv:2006.16679*, 2020.

[16] A. Sanjab, W. Saad, and T. Başar, "A game of drones: Cyber-physical security of time-critical uav applications with cumulative prospect theory perceptions and valuations," *arXiv preprint arXiv:1902.03506*, 2019.

[17] F. Farokhi, A. M. Teixeira, and C. Langbort, "Estimation with strategic sensors," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 724–739, 2016.

[18] Z. Xu and J. Zhuang, "A study on a sequential one-defender-n-attacker game," *Risk Analysis*, vol. 39, no. 6, pp. 1414–1432, 2019.

[19] E. Doynikova, E. Novikova, and I. Kotenko, "Attacker behaviour forecasting using methods of intelligent data analysis: A comparative review and prospects," *Information*, vol. 11, no. 3, p. 168, 2020.

[20] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a nash equilibrium," *SIAM Journal on Computing*, vol. 39, no. 1, pp. 195–259, 2009.

[21] K. Hausken, "A strategic analysis of information sharing among cyber hackers," *JISTEM-Journal of Information Systems and Technology Management*, vol. 12, no. 2, pp. 245–270, 2015.

[22] S. Laube and R. Böhme, "Strategic aspects of cyber risk information sharing," *ACM Computing Surveys (CSUR)*, vol. 50, no. 5, p. 77, 2017.

[23] C. E. Shannon, "Communication theory of secrecy systems," *Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[24] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[25] J.-K. Chong, T.-H. Ho, and C. Camerer, "A generalized cognitive hierarchy model of games," *Games and Economic Behavior*, vol. 99, pp. 257–274, 2016.

[26] C. Camerer, T. Ho, and J.-K. Chong, "A cognitive hierarchy theory of one-shot games and experimental analysis," *Available at SSRN 411061*, 2003.