Attack Identification for Cyber-Physical Security in Dynamic Games under Cognitive Hierarchy

Christos N. Mavridis *, **Aris Kanellopoulos** **,
Kyriakos G. Vamvoudakis ***, **John S. Baras** *,
Karl Henrik Johansson **

* *Department of Electrical and Computer Engineering, University of Maryland, College Park, MD*
e-mails: {mavridis,baras}@umd.edu
** *School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Sweden*
e-mail: {arisk,kallej}@kth.se
*** *The Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA*
e-mail: kyriakos@gatech.edu

Abstract: This paper considers the problem of identifying the profiles and capabilities of attackers injecting adversarial inputs to a cyber-physical system. The system in question interacts with attackers of different levels of intelligence, each employing different feedback controllers against the system. Principles of behavioral game theory – specifically the concept of level-k thinking – is employed to construct a database of potential attack vectors. By observing the state trajectories under sequential interactions with different adversaries, the defender adaptively estimates both the number and profiles of the different attack signals using an online deterministic annealing approach. This information is used to dynamically estimate the level of intelligence of the attackers. Simulation results showcase the efficacy of the proposed method.

1. INTRODUCTION

Cyber-physical systems (CPS) are complex systems integrating physical with digital components, whose nature renders them vulnerable to external attacks. The effects of attacks to CPS have become more obvious, due to various high profile instances, such as the Stuxnet virus, a malicious computer worm targeting programmable logic controllers (Farwell and Rohozinski, 2011). The history of these attacks highlights the need for the development of security algorithms that explicitly consider the CPS as a whole, rather than focusing on the software and communication vulnerabilities.

Game theory (Basar and Olsder, 1999) is a potent tool for modeling interactions between selfish and competitive agents. At the core of game theory one can find the notion of the solution concept; the focal policies that describe the expected behavior of the agents according to a set of predefined assumptions on their cognitive capabilities. The most commonly used solution concept in game theory is the Nash Equilibrium (NE), according to which the players of the game are both optimal in their actions and mutually consistent in their beliefs over their opponents' behavior. However, experimental results have shown that human agents rarely act according to a game's NE (Camerer, 2003). Alternatively, approaches to modeling bounded rationality of agents have been proposed since the advent of game theory. Level-k thinking (Crawford and Iriberri, 2007) is a bounded rationality framework that augments the agents with a specific cognitive ability corresponding to the number of iterative best response steps they are able to perform. Level-k thinking players operate under the assumption that their opponents have abilities of level $k-1$, i.e., they are strictly one step less strategic than themselves. Regardless of the intuition, the value of adopting a cognitive hierarchy approach in CPS systems under attack is twofold. First, it greatly reduces the computational complexity of estimating a possibly intractable optimal policy in differential games, and, second, it constructs a database of attacker policies that can be used to identify their intelligence level.

Identification of the attacker's profile and intelligence level is crucial in defending against dynamically changing adversaries that may become more intelligent with time, especially in the context of level-k thinking. In this work, to identify the different deployed attack profiles, the defender observes the state trajectories of the system under sequential interactions with different adversaries, and uses an adaptive recursive least squares filter to estimate the attack policies. Then, an online deterministic annealing learning scheme (Mavridis and Baras, 2023b) is used as a discrete-time dynamical system that estimates a quan-

⋆ This work was supported in part by ONR grant. N00014-17-1-2622, ARO under grant Nos. W911NF-19 − 1 − 0270, by ONR Minerva under grant No. N00014 − 18 − 1 − 2160, and by NSF under grant Nos. SATC-2231651, CAREER CPS-1851588, and CPS-2038589.

tized distribution of the attacks in the space defined by a set of level-k attacker policies. Thus, online deterministic annealing acts as an adaptive partitioning algorithm in the space of attack policies and provides information on the number of different attacks observed, as well as their distribution in the space of level-k policies.

Related work

Iterative bounded rationality methods are investigated in (Chong et al., 2016). Those methods have been studied extensively in the context of autonomous driving. Specifically, the authors in (Li et al., 2016) leverage a bounded rationality approach called cognitive hierarchy in order to train autonomous vehicles against realistic human-driver models. Similar ideas of bounded rationality have been explored in driver modeling in (Albaba and Yildiz, 2020). Furthermore, level-k thinking and its extensions have found use in various security-related problems where the performance of complex systems rests on their ability to successfully defend against the full unpredictability of human attackers. As such, the authors in (Abuzainab et al., 2016) considered the problem of distributed uplink random access for the Internet of Things, while various results have been presented regarding securing CPS (Vamvoudakis and Kokolakis, 2020; Kanellopoulos et al., 2020; Kokolakis et al., 2021, 2020). In frameworks of repeated games, bounded rationality has been introduced in (Dai et al., 2020), where the authors propose a recursive reasoning formalism for Bayesian games with unknown payoff functions. Results on repeated games have also been reported in (Tian et al., 2020) with the extraction of interpretable human behavior models and (Tian et al., 2020), where autonomous agents are trained to beat humans in simple repeated games. Finally, bounded rationality in the form of level-k thinking and cognitive hierarchy is employed in (Fotiadis and Vamvoudakis, 2022) in the context of stochastic games, solved via both recursive and parallel algorithms.

Learning algorithms for attack identification are mainly represented by clustering approaches with prototype-based models (Mavridis and Baras, 2020, 2023b). These methods can be viewed as iterative, consistent (Mavridis and Baras, 2020), interpretable (Mavridis et al., 2022), and topology-preserving competitive-learning neural networks (Uriarte and Martín, 2005). They use a set of representatives to partition the observation space mimicking similar concepts from cognitive psychology and neuroscience. Deterministic annealing methods (Mavridis and Baras, 2023b,a; Rose, 1998) define a specific class of prototype-based algorithms, that offer properties desirable to cyber-physical systems applications, such as adaptive adjustment of the model complexity, robustness and the ability to control the performance-complexity trade-off of the algorithm.

Contributions: The contributions of the present paper are twofold. First, we model a CPS under attack as a dynamical system with adversarial input injection, and present a computationally feasible framework to construct: (i) a database of level-k attack policies, and (ii) the corresponding defend policies for level-k attacks. Second, we employ an adaptive identification algorithm based on online deterministic annealing learning to estimate a quantized distribution of the attacks in the space defined by the set of level-k attacker policies, and we use this information to identify the intelligence level of the attackers.

2. PROBLEM FORMULATION

Consider a CPS under attack by N_d attackers which interact with the system in a serial manner. As such, the evolution of the CPS is described by a linear time-invariant (LTI) system:

$$\dot{x}(t) = Ax(t) + Bu(t) + \sum_i \mathbb{1}_{[\sigma(t)=i]} K_i d_i(t),\ t \geqslant 0, \quad (1)$$

where $x(t) \in \mathbb{R}^n$ is the state of the system with $x(0) = x_0$, $u(t) \in \mathbb{R}^m$ the defender's input to the system and $d_i(t) \in \mathbb{R}^d$ the attacking signal injected by the active attacker $i \in \{1, \dots, N_d\}$ to the system. The activation function $\sigma : [0, \infty) \to \{1, \dots, N_d\}$ is used to model the sequential attacks. Finally, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $K_i \in \mathbb{R}^{n\times k}$, $i \in \{1, \dots, N_d\}$ are the drift, input, and adversarial matrices, respectively.

Remark 1. For ease of exposition, we drop the subscript i in the attacker during the construction of the database of attacks based on level-k thinking theory. □

2.1 Designing Defense Strategies

In designing defense strategies for the system's operator, a typical approach is to consider a zero-sum game between the defender and the active attacker. The analysis of the game rests on the choice of the solution concept employed. The most common and well-understood solution concept for a differential game is the Nash equilibrium (NE), formalized through the following cost function:

$$J(x; u, d) = \frac{1}{2}\int_0^\infty \left(x^{\mathrm{T}} Q x + u^{\mathrm{T}} R u - \gamma^2 \|d\|^2\right) \mathrm{d}t, \quad (2)$$

where for the weight matrices of appropriate dimensions, it holds that $Q \succeq 0$, $R > 0$ and $\gamma \geqslant \gamma^\star > 0$, where $\gamma^\star$ is the attenuation level of the attacker, required to guarantee that the integral is finite.

Assumption 1. The pair (A, B) is controllable and the pair $(A, Q^{\frac{1}{2}})$ is detectable. □

To derive the feedback policies corresponding to the NE the defender minimizes (2) while the attacker maximizes it. Consequently, we can define a function $V^\star : \mathbb{R}^n \to \mathbb{R}$ that quantifies the value of the game given an initial state $x \in \mathbb{R}^n$ subject to (1). The value of the game is then given by

$$V(x) = \min_u \max_d \int_t^\infty \left(x^{\mathrm{T}} Q x + u^{\mathrm{T}} R u - \gamma^2 \|d\|^2\right) \mathrm{d}\tau.$$

We can derive the NE policies by defining the Hamiltonian function associated with the given cost function and dynamics as

$$H(x, u, d, \nabla V(x)) = x^{\mathrm{T}} Q x + u^{\mathrm{T}} R u - \gamma^2 \|d\|^2 + \left(\nabla V(x)\right)^{\mathrm{T}} (Ax + Bu + Kd), \quad (3)$$

where $\nabla V(x) \in \mathbb{R}^n$ is the gradient of the value function. Applying the stationarity conditions to (3), ones gets $\forall x \in \mathbb{R}^n$

$$u^\star(x) = -R^{-1} B^{\mathrm{T}} \nabla V(x), \quad d^\star(x) = \frac{1}{\gamma^2} K^{\mathrm{T}} \nabla V(x),$$

as the NE feedback policies for the defender and the attacker, respectively. Furthermore, it is known that the value function satisfies the Hamilton-Jacobi-Isaacs equation $H(x, u^\star(x), d^\star(x), \nabla V(x)) = 0,\ \forall x \in \mathbb{R}^n$, which, in the linear-quadratic case (when it holds that $V(x) = x^{\mathrm{T}} P x$ for a matrix $P \succ 0$) becomes the Riccati equation:

$$A^{\mathrm{T}} P + PA - PBR^{-1}B^{\mathrm{T}} P + \frac{1}{\gamma^2} PKK^{\mathrm{T}} P + Q = 0. \quad (4)$$

It is known that (4) has a unique positive definite solution under Assumption 1.

2.2 Attack Identification

We assume that the defender can estimate the attacker's control trajectory $X^i := \hat{d}_i(x(t_i - T_{\text{int}} : t_i)) \in S$ using state observations gathered in a time window $W_i := [t_i - T_{\text{int}} : t_i]$ while the system is under attack by a set of active adversaries $i \in \{1, \ldots, N_d\}$. To estimate the different attack profiles observed, we treat X_i as realizations of a random variable $X : \Omega \to S$. Our goal is to estimate the distribution of X. We build upon the notion of Online Deterministic Annealing (ODA) (Mavridis and Baras, 2023b,a), and define a discrete random variable $Q : S \to S$, with a finite domain μ representing different attac profiles. Once the joint probability space of (X, Q) is defined, we successively solve a series of optimization problems:

$$\min_{\mu}\ (1 - \lambda) D(X, Q) - \lambda H(X, Q), \quad (5)$$

parameterized by a Lagrange coefficient $\lambda \in [0, 1]$ controlling the trade-off between minimizing an average distortion measure $D(X, Q)$ (to be defined in Section 4), and maximizing the Shannon entropy $H(X, Q)$.

In Section 4 we show that the optimization problems (5) can be solved using gradient-free stochastic approximation updates, i.e., using a discrete-time dynamical system running at times $\{nT_{\text{int}}\}_{n \in \mathbb{N}}$. Moreover, decreasing the values of λ will lead in a series of bifurcation phenomena when the cardinality of the domain μ of Q increases and the underlying probability distribution of the observations X^i is more closely represented (Mavridis and Baras, 2023b).

3. BOUNDED RATIONALITY IN DIFFERENTIAL GAMES

Level-0 (Anchor) Policy: To construct the reasoning iterations for a player, we must define a "naive" policy of a level-0 player. Thus, we let a level-0 defender act based on the belief that there are no attackers in the environment. Their policy comprises the solution of an optimal feedback problem, corresponding to the value function

$$V_u^0(x) = \min_u \int_t^\infty (x^{\mathrm{T}} Q x + u^{\mathrm{T}} R u) \mathrm{d}\tau,\ \forall x \in \mathbb{R}^n. \quad (6)$$

The optimal feedback policy based on (6) is given as

$$u^0(x) = -R^{-1}B^{\mathrm{T}} \nabla V_u^0(x) = -R^{-1}B^{\mathrm{T}} P_u^0 x,\ \forall x \in \mathbb{R}^n,$$

and is solved based on the assumption that $d(t) = 0,\ \forall t \geqslant 0$ in (1). Moreover, due to the linear-quadratic nature of the problem, the value function has the form $V_u^0(x) = x^{\mathrm{T}} P_u^0 x$, where P_u^0 solves

$$A^{\mathrm{T}} P_u^0 + P_u^0 A + Q - P_u^0 B R^{-1} B^{\mathrm{T}} P_u^0 = 0.$$

Extending the same principles to a level-0 attacker, we model their naive anchor policy as an attack under the belief that the defender is also level-0, i.e., unaware of the attack. Thus, the attacker assumes the use of a feedback defense policy $u(x) = u^0(x),\ \forall x \in \mathbb{R}^n$ and solves an optimal feedback problem according to the value function

$$V_d^0(x) = \max_d \int_t^\infty (x^{\mathrm{T}} Q x + u^{0\mathrm{T}} R u^0 - \gamma^2 \|d\|^2) \mathrm{d}\tau,\ \forall x \in \mathbb{R}^n,$$

subject to,

$$\dot{x}(t) = (A - BR^{-1}B^{\mathrm{T}} P_u^0) x(t) + Kd(t),\ x(0) = x_0,\ t \geqslant 0.$$

Consequently, the level-0 attack becomes

$$d^0(x) = \frac{1}{\gamma^2} K^{\mathrm{T}} P_d^0 x,\ \forall x \in \mathbb{R}^n,$$

where P_d^0 solves the following Riccati equation:

$$\begin{aligned} 0 = {} & (A - BR^{-1}B^{\mathrm{T}} P_u^0)^{\mathrm{T}} P_d^0 + P_d^0 (A - BR^{-1}B^{\mathrm{T}} P_u^0) \\ & + (Q + P_u^0 B R^{-1} B^{\mathrm{T}} P_u^0) + \frac{1}{\gamma^2} P_d^0 K K^{\mathrm{T}} P_d^0. \end{aligned}$$

Higher level policies: Once we have defined the level-0 strategies of both players, we introduce a procedure of constructing higher level policies in an iterative manner. Specifically, a player of level-k solves for their best response policy by holding the belief that their opponent is of level-$k-1$. In the case of the level-k defender, the best response feedback policy $u^k(x)$ will be derived via the level-k value function $V_u^k(x)$ defined as

$$V_u^k(x_0) = \min_u \int_0^\infty (x^{\mathrm{T}} Q x + u^{\mathrm{T}} R u - \gamma^2 \|d^{k-1}\|^2) \mathrm{d}\tau,$$

subject to

$$\dot{x}(t) = Ax(t) + Bu(t) + Kd^{k-1}(x(t)),\ x(0) = x_0,\ t \geqslant 0.$$

Assuming that the attacker is of level-$k-1$, the defender lets $d^{k-1}(x) = \frac{1}{\gamma^2} K^{\mathrm{T}} P_d^{k-1} x,\ \forall x \in \mathbb{R}^n$ which yields the level-k best response

$$u^k(x) = -R^{-1}B^{\mathrm{T}} P_u^k x = -L_u^k x,\ \forall x \in \mathbb{R}^n. \quad (7)$$

Similar to the previous levels, matrix P_i^k solves the following level-k Riccati matrix equation:

$$\begin{aligned} 0 = {} & (A + \frac{1}{\gamma^2} K K^{\mathrm{T}} P_d^{k-1})^{\mathrm{T}} P_u^k + P_u^k (A + \frac{1}{\gamma^2} K K^{\mathrm{T}} P_d^{k-1}) \\ & + (Q - \frac{1}{\gamma^2} P_d^{k-1} K K^{\mathrm{T}} P_d^{k-1}) - P_u^k B R^{-1} B^{\mathrm{T}} P_u^k. \end{aligned}$$

Via the same process, the attacker of a level-k rationality, designs their attack based on the best response to a level-k defender. This corresponds to the following value function, defined $\forall x \in \mathbb{R}^n$:

$$V_d^k(x) = \max_d \int_t^\infty (x^{\mathrm{T}} Q x + \left(u^k\right)^{\mathrm{T}} R u^k - \gamma^2 \|d\|^2) \mathrm{d}\tau,$$

subject to

$$\dot{x}(t) = Ax(t) + Bu^k(t) + Kd(t),\ x(0) = x_0,\ t \geqslant 0.$$

Their best response is

$$d^k(x) = \frac{1}{\gamma^2} K^{\mathrm{T}} P_d^k x = -L_d^k x, \quad \forall x \in \mathbb{R}^n. \quad (8)$$

The matrix P_d^k solves the following Riccati equation:

$$\begin{aligned} 0 = {} & (A - BR^{-1}B^{\mathrm{T}} P_u^k)^{\mathrm{T}} P_d^k + P_d^k (A - BR^{-1}B^{\mathrm{T}} P_u^k) \\ & + (Q + P_u^k B R^{-1} B^{\mathrm{T}} P_u^k) + \frac{1}{\gamma^2} P_d^k K K^{\mathrm{T}} P_d^k. \end{aligned}$$

In the following theorem we characterize conditions of existence of level-k policies.

Theorem 1. ((Kanellopoulos and Vamvoudakis, 2019)). Consider the system (1) under the effect of agents with bounded rationality and policies given by (7) and (8) for the defender and the attacker, respectively. The game can be solved up to any level-k as long as the following holds:

$$P_u^k BR^{-1}B^{\mathrm{T}}P_u^k > \max\big(\frac{1}{\gamma^2}P_d^{k-1}KK^{\mathrm{T}}P_d^{k-1}, \frac{1}{\gamma^2}P_d^kKK^{\mathrm{T}}P_d^k\big).$$

The described process enables the construction of a database of policies that a defender can utilize along with level estimation algorithms, such as ODA.

4. ATTACK IDENTIFICATION WITH ONLINE DETERMINISTIC ANNEALING

We assume that the defender has access to the state trajectory vector $x(t),\ t>0$ of system (1) gathered while the system is under attack by a set of active adversaries $i \in \{1,\dots,N_d\}$. We define a single observation at time t_i to be obtained by the trajectory $x(t_i - T_{\text{int}} : t_i)$, where the interaction time $T_{\text{int}} \in \mathbb{R}^+$ is defined a priori, and assumed large enough such that the estimation algorithms (9), (10) presented below are able to achieve practical convergence. This approach can be generalized to dynamically changing time windows.

Within the time window $W_i := [t_i - T_{\text{int}} : t_i]$, assuming the system dynamics of (1) and linear feedback control laws as in (7), (8) with unknown gain on the attacker's controller L_d, we can create an estimate $\hat{L}_d$ of L_d, by adaptively estimating $\hat{A}$ of the system:

$$\dot{x}(t) = (A - BL_u + K\hat{L}_d)x(t) := \hat{A}x(t),\ t \in W_i.$$

We introduce a stable filter

$$\dot{w}(t) = \Lambda w(t) + x(t),\ w_0,\ t \in W_i \tag{9}$$

such that, asymptotically, $x(t) = (\hat{A} - \Lambda)w(t) = \Theta^\star w(t)$, where the elements of $\Theta^\star \in \mathbb{R}^{n\times n}$ can be estimated by,

$$\begin{aligned}\dot{\Theta}(t) &= -\gamma\big(\Theta(t)w(t) - x(t)\big)w^{\mathrm{T}}(t)P(t)\\ \dot{P}(t) &= -P(t)w(t)w^{\mathrm{T}}(t)P(t),\ t \in W_i\end{aligned} \tag{10}$$

where $\gamma > 0$, and $\Theta_0,\ P_0$ are initial conditions such that P_0 is invertible. It can be shown through standard Lyapunov analysis that (10) asymptotically minimizes the error $\min_\Theta \int_0^\infty \frac{1}{2}\|e(t)\|^2 \mathrm{d}t$, $e(t) := \Theta w - x = (\Theta - \Theta^\star)w$.

Using the adaptive identification scheme in (10), we can estimate $\hat{L}_d^i$ as the attacker's control gain for the observation window W_i. To estimate the different attack profiles observed, we define the observations $X^i := \hat{L}_d^{(i)} \in \mathbb{R}^n$ for each time window W_i, as realizations of a random variable $X : \Omega \to \mathbb{R}^n$ defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In a more general framework, where the linear dynamics assumptions (1) do not hold, the attacker's control trajectory $X^i := \hat{d}_i(x(t_i - T_{\text{int}} : t_i))$ can be used as an observation. Our goal is to to find a set of profiles $\{\mu_i\}_{i=1}^M \in S := \mathbb{R}^n$ that estimate the distribution of X, i.e., the distribution of the observed attacks X_i, expressed as a linear combination of the set $\{L_d^k\}_{k=1}^{\bar{k}}$, computed by the defender using (8).

Following the principles of online deterministic annealing (Mavridis and Baras, 2023b), we define a discrete random variable $Q : S \to \text{ri}(S)$ ($\text{ri}(S)$ represents the relative interior of S) described by the association probabilities $p(\mu_i|x) = \mathbb{P}[Q = \mu_i | X = x]$ that represent the probability of $x \in S$ to belong to the subset $S_i := \{x \in S : i = \arg\min_j d_\phi(x, \mu_j)\}$. Once the joint probability space of (X, Q) is defined, we successively solve a series of optimization problems:

$$\min_{\{\mu_i\}}\ F_\lambda(X,Q) := (1-\lambda)D(X,Q) - \lambda H(X,Q), \tag{11}$$

parameterized by a Lagrange coefficient $\lambda \in [0,1]$ controlling the trade-off between minimizing an average distortion measure $D(X,Q) := \mathbb{E}[d_\phi(X,Q)]$, for an appropriately defined Bregman divergence d_ϕ, and maximizing the Shannon entropy $H(X,Q) := \mathbb{E}[-\log p(X,Q)]$. Bregman divergences are information-theoretic dissimilarity measures that generalize convex metric measures and include the widely used squared Euclidean distance and the Kullback-Leibler divergence, as two notable examples. For more information on Bregman divergence, the reader is referred to (Banerjee et al., 2005; Mavridis and Baras, 2023b). It can be seen that the solution of the optimization problem:

$$F_\lambda^\star(\mu) := \min_{\{p(\mu_i|x)\}} F_\lambda(\mu)\ \text{ s.t. } \sum_i p(\mu_i|x) = 1$$

is given by the Gibbs distributions

$$p^\star(\mu_i|x) = \frac{e^{-\frac{1-\lambda}{\lambda}d(x,\mu_i)}}{\sum_j e^{-\frac{1-\lambda}{\lambda}d(x,\mu_j)}},\ \forall x \in S. \tag{12}$$

In addition, the following theorem proven in Mavridis and Baras (2023b), provides a stochastic approximation algorithm Borkar (2009) to solve the optimization problem $\min_\mu F_\lambda^\star(\mu)$.

Theorem 2. ((Mavridis and Baras, 2023b)). Let $\{x_n\}$ be a sequence of independent realizations of X. Then $\mu_i(n)$, defined by the stochastic approximation updates

$$\begin{cases}\rho_i(n+1) &= \rho_i(n) + \alpha(n)\left[\hat{p}(\mu_i|x_n) - \rho_i(n)\right]\\ \sigma_i(n+1) &= \sigma_i(n) + \alpha(n)\left[x_n\hat{p}(\mu_i|x_n) - \sigma_i(n)\right]\end{cases} \tag{13}$$

where $\sum_n \alpha(n) = \infty$, $\sum_n \alpha^2(n) < \infty$, and the quantities $\hat{p}(\mu_i|x_n)$ and $\mu_i(n)$ are recursively updated as follows:

$$\mu_i(n) = \frac{\sigma_i(n)}{\rho_i(n)},\quad \hat{p}(\mu_i|x_n) = \frac{\rho_i(n)e^{-\frac{d(x_n,\mu_i(n))}{T}}}{\sum_i \rho_i(n)e^{-\frac{d(x_n,\mu_i(n))}{T}}} \tag{14}$$

converges almost surely to a locally asymptotically stable solution of the optimization $\min_\mu F_\lambda^\star(\mu)$, as $n \to \infty$.

Using the observations X^i, we sequentially solve (11) for decreasing values of λ. For $\lambda = 1$, the solution to $\min_\mu F_\lambda^\star(\mu)$ yields a unique solution μ_1, i.e., we get $\{\mu_i\}_{i=1}^M = \{\mu_i\}_{i=1}^{M_1} = \mu_1$, where $M_1 = 1$ is the cardinality of the estimated profiles. As explained in (Mavridis and Baras, 2023b), as λ decreases a bifurcation phenomenon takes place, according to which, there exist critical values λ_c when the cardinality M_λ of $\{\mu_i\}_{i=1}^{M_\lambda}$ increases. At the same time, the average distortion term $D(X,Q) := \mathbb{E}[d_\phi(X,Q)]$ decreases, indicating a better representation of the underlying distribution of X by a finite set of profiles $\{\mu_i\}_{i=1}^{M_\lambda}$. This process continues until a certain level $\lambda_{\min}$ is achieved. The algorithmic implementation of this process and details on its parameters are discussed in (Mavridis and Baras, 2023b, 2022). Finally, notice that at $\lambda_{\min}$, the Gibbs probabilities (12) represent the similarity of any point in $S = \mathbb{R}^n$ at the desirable level of detail. We

use (12) to compute the similarity between the estimated profiles $\{\mu_i\}_{i=1}^{M_{\lambda_{\min}}}$ and the set $\{L_d^k\}_{k=1}^{\bar{k}}$, computed by the defender using (8). This yields a set of $\left\{\hat{L}_d^{(i)}\right\}_{i=1}^{M_{\lambda_{\min}}}$ identified attacks expressed as distributions over the set $\{L_d^k\}_{k=1}^{\bar{k}}$, which indicates the level of intelligence of the observed attack profiles.

5. SIMULATION RESULTS

The simulation study was conducted on a linearized vehicle system found in (Guo et al., 2010; Yan et al., 2017), whose dynamical system representation is

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\dot{\phi}\\ \dot{\xi}\\ \phi\\ \xi\end{bmatrix} = \begin{bmatrix}-2.11 & -6.61 & 9.48 & -357.05\\ 73.54 & -61.70 & 11.71 & -757.81\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix}\begin{bmatrix}\dot{\phi}\\ \dot{\xi}\\ \phi\\ \xi\end{bmatrix} + \begin{bmatrix}1.2\\ 10\\ 0\\ 0\end{bmatrix} u + \begin{bmatrix}0\\ 8\\ 0\\ 0\end{bmatrix} d,\ t \geqslant 0,$$

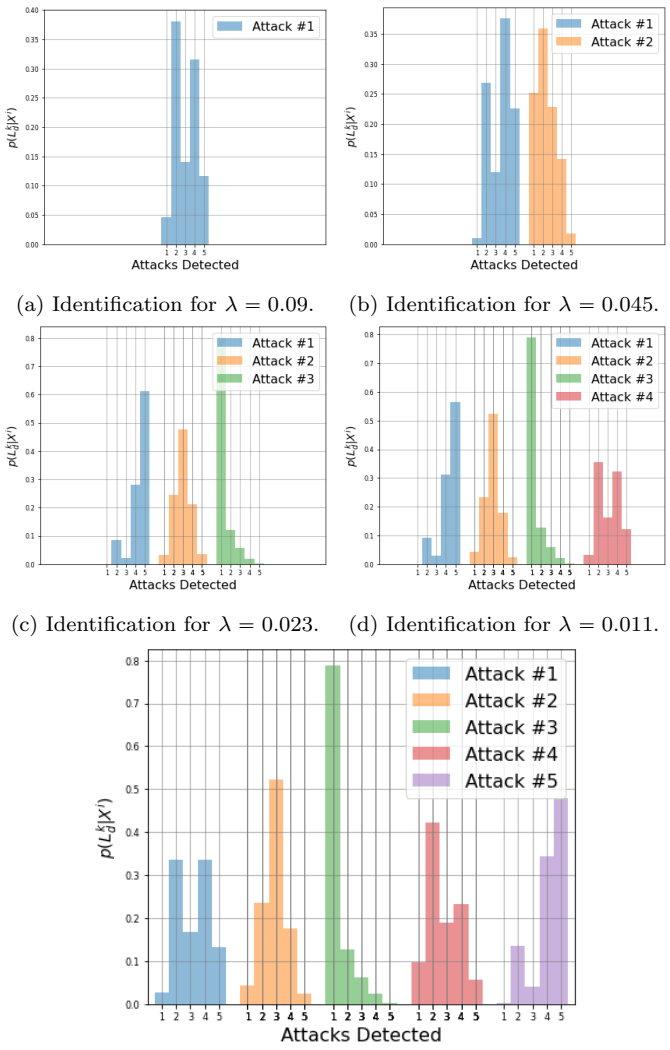
where $\phi, \xi \in \mathbb{R}$ denote the lean rotation and rotation of the front wheel with respect to the rear wheel respectively, and $u, d \in \mathbb{R}$ are the defender's and the attacker's inputs, respectively.

The simulated system is under attack sequentially by multiple adversaries, each launching an attack signal based on a random feedback policy that is generated as a noisy mixture of multiple level-k attacks. This randomness highlights the ability of the algorithm to capture arbitrary attack signals of agents that may change their levels during the game or even play sub-optimal instead of best responses. The components of the control gains $\left\{L_d^{(i)} \in \mathbb{R}^n\right\}_{i=1}^{\bar{k}}$, for $\bar{k} = 5$ generated attacks are shown in Table 1. The computation of $\bar{k} = 5$ level-k policies is done by the defender offline.

While the system is under attack, the defender has access to the state of (1), estimates the control gain of the attacker $\hat{L}_d$ in time windows of length T_{int} with (10), and runs the attack identification algorithm (13), (14) starting with initial temperature $\lambda = 0.9$ and stopping temperature $\lambda = 0.005$. The evolution of the estimated attack gains is shown in Tables 2, 3, and 4. The identification algorithm is able to identify both the exact number and the average values of the original attack gains using only online observations. Finally, the similarity between the identified attacks and the pre-computed level-k attack profiles is shown in Fig. 1. The information of Fig. 1e reveals the intelligent levels of the different attackers, i.e., there exists one attacker (Attacker #3) that operates mainly on level-0, one (Attacker #2) that operates mainly on level-2, and so on.

6. CONCLUSION AND FUTURE WORK

We have addressed the problem of identifying the cognitive ability level of agents attacking a CPS. Leveraging a level-k thinking model we have constructed a database of policies that correspond to specific numbers of strategic thinking steps. Subsequently, we let the CPS gather trajectory data generated under the effect of different attackers over specified time intervals. Using a least-squares adaptation process, the operator of the CPS can derive the attack inputs which are then employed by an ODA algorithm to identify the levels of the attackers. The ODA algorithm is shown to successfully identify the distribution of levels and the number of strategic steps observed.



(a) Identification for $\lambda = 0.09$. (b) Identification for $\lambda = 0.045$.

(c) Identification for $\lambda = 0.023$. (d) Identification for $\lambda = 0.011$.

(e) Identification for $\lambda = 0.006$.

Fig. 1. Evolution of the attack identification process for decreasing temperature levels λ. The similarity values between the identified attacks and the pre-computed level-k attack profiles are depicted.

Attack 1	-0.04± 0.1	0.08± 0.1	-0.19± 0.1	0.74± 0.1
Attack 2	-0.10± 0.2	0.11± 0.2	-0.72± 0.2	0.95± 0.2
Attack 3	-0.35± 0.2	0.38± 0.2	-0.70± 0.2	0.87± 0.2
Attack 4	-0.16± 0.3	0.18± 0.3	-0.86± 0.3	1.16± 0.3
Attack 5	-0.08± 0.2	0.08± 0.2	-1.11± 0.2	1.31± 0.2

Table 1. Generated Attacks.

Attack 1	-0.11	0.13	-0.53	0.89

Table 2. Identified Attacks for $\lambda = 0.09$.

Attack 1	-0.14	0.16	-0.70	0.92
Attack 2	-0.14	0.16	-0.95	1.23
Attack 3	-0.04	0.08	-0.19	0.72

Table 3. Identified Attacks for $\lambda = 0.023$.

Attack 1	-0.04	0.08	-0.19	0.75
Attack 2	-0.09	0.10	-0.68	0.89
Attack 3	-0.35	0.38	-0.69	0.86
Attack 4	-0.13	0.15	-0.78	1.05
Attack 5	-0.11	0.12	-1.05	1.29

Table 4. Identified Attacks for $\lambda = 0.006$.

Future work will focus on generalizing the problem of adversarial intelligence identification in general static and repeated non-zero sum games. Furthermore, we will investigate methods of constructing mixed policies for a defender based on the expected behaviors of the attackers in their environment.

REFERENCES

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

Albaba, B.M. and Yildiz, Y. (2020). Driver modeling through deep reinforcement learning and behavioral game theory. *arXiv preprint arXiv:2003.11071*.

Banerjee, A., Merugu, S., Dhillon, I.S., and Ghosh, J. (2005). Clustering with bregman divergences. *Journal of machine learning research*, 6(Oct), 1705–1749.

Basar, T. and Olsder, G.J. (1999). *Dynamic noncooperative game theory*, volume 23. Siam.

Borkar, V.S. (2009). *Stochastic approximation: a dynamical systems viewpoint*, volume 48. Springer.

Camerer, C.F. (2003). Behavioral game theory: Plausible formal models that predict accurately. *Behavioral and Brain Sciences*, 26(02), 157–158.

Chong, J.K., Ho, T.H., and Camerer, C. (2016). A generalized cognitive hierarchy model of games. *Games and Economic Behavior*, 99, 257–274.

Crawford, V.P. and Iriberri, N. (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica*, 75(6), 1721–1770.

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

Farwell, J.P. and Rohozinski, R. (2011). Stuxnet and the future of cyber war. *Survival*, 53(1), 23–40.

Fotiadis, F. and Vamvoudakis, K.G. (2022). Recursive reasoning with reduced complexity and intermittency for nonequilibrium learning in stochastic games. *IEEE Transactions on Neural Networks and Learning Systems*.

Guo, L., Liao, Q., Wei, S., and Huang, Y. (2010). A kind of bicycle robot dynamic modeling and nonlinear control. In *The 2010 IEEE International Conference on Information and Automation*, 1613–1617. IEEE.

Kanellopoulos, A., Fotiadis, F., Vamvoudakis, K.G., and Gupta, V. (2020). A meta-learning and bounded rationality framework for repeated games in adversarial environments. In *2020 59th IEEE Conference on Decision and Control (CDC)*, 1640–1645. IEEE.

Kanellopoulos, A. and Vamvoudakis, K.G. (2019). Nonequilibrium dynamic games and cyber–physical security: A cognitive hierarchy approach. *Systems & Control Letters*, 125, 59–66.

Kokolakis, N.M.T., Kanellopoulos, A., and Vamvoudakis, K.G. (2020). Bounded rational unmanned aerial vehicle coordination for adversarial target tracking. In *2020 American Control Conference (ACC)*, 2508–2513. IEEE.

Kokolakis, N.M.T., Kanellopoulos, A., and Vamvoudakis, K.G. (2021). Bounded rationality in differential games: A reinforcement learning-based approach. In *Handbook of Reinforcement Learning and Control*, 467–489. Springer.

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

Mavridis, C. and Baras, J. (2022). Multi-resolution online deterministic annealing: A hierarchical and progressive learning architecture. *arXiv preprint arXiv:2212.08189*.

Mavridis, C. and Baras, J. (2023a). Annealing optimization for progressive learning with stochastic approximation. *IEEE Transactions on Automatic Control*.

Mavridis, C., Noorani, E., and Baras, J.S. (2022). Risk sensitivity and entropy regularization in prototype-based learning. In *2022 30th Mediterranean Conference on Control and Automation (MED)*, 194–199. IEEE.

Mavridis, C.N. and Baras, J.S. (2020). Convergence of stochastic vector quantization and learning vector quantization with bregman divergences. *IFAC-PapersOnLine*, 53(2).

Mavridis, C.N. and Baras, J.S. (2023b). Online deterministic annealing for classification and clustering. *IEEE Transactions on Neural Networks and Learning Systems*.

Rose, K. (1998). Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11), 2210–2239.

Tian, R., Li, N., Kolmanovsky, I., and Girard, A. (2020). Beating humans in a penny-matching game by leveraging cognitive hierarchy theory and bayesian learning. In *2020 American Control Conference (ACC)*, 4652–4657.

Tian, R., Sun, L., and Tomizuka, M. (2020). Bounded risk-sensitive markov game and its inverse reward learning problem. *arXiv preprint arXiv:2009.01495*.

Uriarte, E.A. and Martín, F.D. (2005). Topology preservation in som. *International journal of applied mathematics and computer sciences*, 1(1), 19–22.

Vamvoudakis, K.G. and Kokolakis, N.M.T. (2020). Synchronous reinforcement learning-based control for cognitive autonomy. *Foundations and Trends® in Systems and Control*, 8(1–2).

Yan, Y., Antsaklis, P., and Gupta, V. (2017). A resilient design for cyber physical systems under attack. In *American Control Conference (ACC), 2017*, 4418–4423. IEEE.