Adaptive Strategic Cyber Defense for Advanced Persistent Threats in Critical Infrastructure Networks

Linan Huang and Quanyan Zhu
\{lh2328, qz494\}@nyu.edu

Department of Electrical and Computer Engineering, New York University
2 MetroTech Center, Brooklyn, NY, 11201, USA

ABSTRACT

Advanced Persistent Threats (APTs) have created new security challenges for critical infrastructures due to their stealthy, dynamic, and adaptive natures. In this work, we aim to lay a game-theoretic foundation by establishing a multi-stage Bayesian game framework to capture incomplete information of deceptive APTs and their multi-stage multi-phase movement. The analysis of the perfect Bayesian Nash equilibrium (PBNE) enables a prediction of attacker’s behaviors and a design of defensive strategies that can deter the adversaries and mitigate the security risks. A conjugate-prior method allows online computation of the belief and reduces Bayesian update into an iterative parameter update. The forwardly updated parameters are assimilated into the backward dynamic programming computation to characterize a computationally tractable and time-consistent equilibrium solution based on the expanded state space. The Tennessee Eastman (TE) process control problem is used as a case study to demonstrate the dynamic game under the information asymmetry and show that APTs tend to be stealthy and deceptive during their transitions in the cyber layer and behave aggressively when reaching the targeted physical plant. The online update of the belief allows the defender to learn the behavior of the attacker and choose strategic defensive actions that can thwart adversarial behaviors and mitigate APTs. Numerical results illustrate the defender’s tradeoff between the immediate reward and the future expectation as well as the attacker’s goal to reach an advantageous system state while making the defender form a positive belief.

1. INTRODUCTION

With the integration of communication networks and information technologies with the critical infrastructures including power grids, transportation systems, and water distribution systems, the direct use of the off-the-shelf technologies has made our infrastructure vulnerable to cyber attacks. One emerging threat is the Advanced Persistent Threats (APTs) which are a class of multi-phase and multi-stage hacking processes \[9\], initiating their infections in cyberinfrastructures yet targeting at specific physical infrastructures such as nuclear power stations and automated factories. Unlike the “spray-and-pray” attacks, APTs as the targeted at-structure vulnerable to cyber attacks. One emerging threat is the Advanced Persistent Threats (APTs) which are a class of multi-phase and multi-stage hacking processes \[9\], initiating their infections in cyberinfrastructures yet targeting at specific physical infrastructures such as nuclear power stations and automated factories. Unlike the “spray-and-pray” attacks, APTs as the targeted attacks, perform reconnaissance and tailor their hacking techniques to the targeted system. As shown in Fig. 1, the APTs’ life cycle includes a sequence of phases and stages such as the initial entry, privilege escalations, and lateral movements. APTs use each stage as a stepping stone for the next one. Since APTs have a specific target at the final stage, they receive no benefits going back to previous stages. Thus, the multi-stage attack graph bears a tree structure without jumps or loops. Unlike the “smash-and-grab” attacks, APTs behave seemingly as legitimate users, wait until the final stage to launch the “critical hit” on their specific targets, and inflict an enormous loss.

The classical intrusion prevention techniques such as the cryptography and the physical isolation can be ineffective for APTs. An APT-type adversary can steal the full cryptographic keys by exploiting zero-day vulnerabilities and techniques such as social engineering. Stuxnet can bridge the air gap between local-area networks with the insertion of infected USB drives. Similarly, the intrusion detection approach \[2\] has proven to be insufficient when APTs acquire knowledge of the system response as well as the detection rule with the help of insiders and the reconnaissance. Moreover, APTs operated by human experts can analyze, learn, and update the knowledge of the system, thus evading detection by stealthy and strategic movements, e.g., scan the port sufficient slow to avoid the alarm and even choose the No Operation (NOP) at some stages. Hence, it is essential to design up-to-date security mechanisms that can mitigate the risks despite the successful infiltration and the strategic response of APTs.

One way to understand the multi-stage and stealthy nature of the APTs is through dynamic games with incomplete information. The dynamic game frameworks capture the multi-stage movement of the defender and the attacker in networks \[3\][6]. The deceptive and stealthy behaviors of the APTs lead to the information asymmetry where an attacker has his own private information encapsulated by a random variable called type \[3\]. The type characterizes the essence and the objective of the user, i.e., whether the user is legitimate or adversarial, which assets serve as his targets, and how

![Figure 1: APTs start the infection by exploiting network vulnerabilities or the human weakness. They aim to cause physical damages or collect confidential data.](image-url)
much damages he can inflict on the system. The user’s type determines his behaviors if he is rational and aims at maximizing his utility, which makes it possible for the defender to form and update a belief of the type based on the history of user’s behaviors. Since the attacker has to follow the network protocol and move stealthily across the networks by hiding his footprints and evading the detection, it is natural to view the defender as the principal who can design security policies and the attacker as the agent who follows the policies to attain his goal. The strategic behaviors of the defender and the attacker will lead to a perfect Bayesian Nash equilibrium (PBNE) where no one can profit from unilateral deviations at any stage. Achieving a long-term statistic optimal is challenging since the belief updates forwardly yet the PBNE strategy pair is computed backwardly. With the beta-binomial conjugate prior assumption, we manage to unify the coupled forward and backward processes and form the dynamic programming with an expanded state. Tennessee Eastman process is used as a case study to illustrate the theoretical underpinning of our framework for the design of strategic defense to deter the attacks and mitigate the impact of the threats.

Related Work: FlipIt game [7] has analyzed the scenario of the key leakage under APTs so that a system operator and APTs will take over the system alternately. Defenders cannot know the time of the stealthy takeover as well as the current system status unless taking defensive actions. FlipIt game provides high-level guidelines on how to allocate the limited resources to deter the APTs. Our multi-stage Bayesian game framework, however, supports a specification of both adversarial and defensive actions with utilities and enables the equilibrium analysis of the game as the prediction of the attack moves. Signaling game, a two-stage game with the one-sided type, has been applied to study the information asymmetry in cyber deception [8]. However, both players receive a one-shot one-sided utility, which makes it possible for the defender to form and update a belief of the type based on the history of user’s behaviors. The user’s type influences the state. A user with a larger type value indicates a higher threat of APTs through the attack tree or honeypots. The described security policies and the attacker as the agent who follows the policies to attain his goal. The strategic behaviors of the defender and the attacker will lead to a perfect Bayesian Nash equilibrium (PBNE) where no one can profit from unilateral deviations at any stage. Achieving a long-term statistic optimal is challenging since the belief updates forwardly yet the PBNE strategy pair is computed backwardly. With the beta-binomial conjugate prior assumption, we manage to unify the coupled forward and backward processes and form the dynamic programming with an expanded state. Tennessee Eastman process is used as a case study to illustrate the theoretical underpinning of our framework for the design of strategic defense to deter the attacks and mitigate the impact of the threats.

Believe Update: To strategically gauge the user’s type, the defender specifies a belief \( B' : \mathcal{X}^T \mapsto \Delta \Theta \) as a distribution over the type space according to the state \( x' \) at stage \( t \). Likewise, \( B' (\theta | x') \) is the conditional probability density function (PDF) of the type \( \theta \) under the state \( x' \). The prior distribution of the user’s type is known to be \( B_0' \) and the belief of the type updates according to the Bayesian rule with the arrival of the action observation \( a_2' \) drawn from the mixed strategy \( \sigma' \).

\[
B'_{t+1} (\theta | f' (x', a_1', a_2')) = \frac{B' (\theta | x') \sigma_2' (d_1' | x', \theta) \int_{\Theta} B' (\theta | x') \sigma_1' (d_2' | x', \theta) d\theta}{\int_{\Theta} B' (\theta | x') \sigma_1' (d_2' | x', \theta) d\theta}.
\]

Utility Function: The user’s type influences P1’s immediate payoff received at each stage \( t \), i.e., \( \hat{J}_1^T : \mathcal{X}^T \times \mathcal{A}_x^T \times \mathcal{A}_y^T \times \Theta \mapsto \mathbb{R} \). For example, a legitimate user’s access to the sensor benefits the system while a pernicious user’s access can incur a considerable loss. Define \( \sigma_{1^T} := \{ \sigma_1' \in \Sigma_1' \}_{t=1}^T \) as a sequence of policies from \( t' \) to \( T \). The defender has the objective to maximize the cumulative expected utility:

\[
U_1^{1,T} (\sigma_{1^T}, \sigma_{2^T}, T, x') := \sum_{t=1}^T E_{\theta \sim B', \sigma_{1} \sim \sigma_{1^T}, \sigma_{2} \sim \sigma_{2^T}} \hat{J}_1^T (x_1', a_1', a_2', \theta).
\]

and the user’s objective function is

\[
U_2^{1,T} (\sigma_{1^T}, \sigma_{2^T}, T, x', \theta) = \sum_{t=1}^T \sum_{a_1' \in \mathcal{A}_x^T} \sigma_1' (d_1' | x', \theta) \sum_{a_2' \in \mathcal{A}_y^T} \sigma_2' (d_2' | x', \theta) \hat{J}_2'.
\]

Perfect Bayesian Nash Equilibrium: We model the scenario of APTs under the insider threat as a dynamic principal-agent problem where defender P1 as the principal chooses her policy \( \sigma_1' \) first at each stage \( t \). Attacker P2 as the agent perceives \( \sigma_1' \) via insiders, and then chooses his policy \( \sigma_2' \), i.e., maximizes his cumulative expected utility \( U_2^{1,T} \). Since APTs have to follow rules to avoid detection, a sophisticated defender aware of the potential policy leakage under insider threats can acquire the best response of APTs through the attack tree or honeypots. The described security scenario leads to the following definition of perfect Bayesian Nash equilibrium (PBNE) where the defender chooses the most rewarding policy to confront the attacker’s best-response policies.

**Definition 1.** In the two-person multi-stage game with a sequence of beliefs \( B' : \{t', \ldots, T\} \) satisfying the Bayesian update in (1) and the cumulative utility function \( U_2^{1,T} \), the set \( \sigma_2^{T} (\sigma_1^{T}) := \{ \gamma \in \Sigma_2^T : U_2^{1,T} (\sigma_1^{T}, \gamma, x', \theta) \geq U_2^{1,T} (\sigma_1^{T}, \sigma_2', x', \theta), \forall \sigma_2' \in \Sigma_2^T, \forall x' \in \mathcal{X}^T, \theta \in \Theta \} \) is P1’s best-response set to P1’s policy \( \sigma_1^{T} \in \Sigma_1^T \) under state \( x' \) and type \( \theta \).
In the two-person multi-stage Bayesian game with P₁ as the principal, the cumulative utility function \( U^{T'} \), the initial state \( x' \in X' \), the type \( \Theta \), and a sequence of beliefs \( B_{t} \in \{B_1, \ldots, B_T\} \) in \([1]\), a sequence of strategies \( \sigma_i^{T'} \in \Sigma_i^{T'} \) is called a perfect Bayesian Nash equilibrium (PBNE) for the principal, if
\[
U_1^{T':T}(x') := \inf_{\sigma_1^{T':T} \in \Sigma_1^{T':T}} U_1^{T':T}(\sigma_1^{T':T}, x') = \sup_{\sigma_1^{T'}, \sigma_2^{T'} \in \Sigma_1^{T'}, \Sigma_2^{T':T}} U_1^{T':T}(\sigma_1^{T'}, \sigma_2^{T':T}, x').
\]
A strategy \( \sigma_2^{T':T} \in \arg \max_{\sigma_2^{T':T} \in \Sigma_2^{T':T}} U_2^{T':T}(\sigma_1^{T'}, \sigma_2^{T':T}, x', \theta) \) is a PBNE for the agent \( P_2 \).

**Dynamic Programming:** Given the type belief at every stage, we use dynamic programming to find the PBNE policies in a backward fashion because of the tree structure and the finite horizon. Define the value function \( V_i^{T'}(x') := U_i^{T'}(\sigma_1^{T':T}, \sigma_2^{T':T}, x') \) and \( V_2^{T'}(x', \theta) := U_2^{T'}(\sigma_1^{T'}, \sigma_2^{T':T}, x', \theta) \) as the optimal utility-to-go for the defender and the user, respectively. We have the following simultaneous equations, i.e.,
\[
V_1^{T'}(x') = \sup_{\sigma_1} E_{\theta \sim B_0 \sigma_1} \left[ V_1^{T'+1}(f(x', d_1', d_2)) + J_1^*(x', d_1', d_2) \right],
\]
\[
V_2^{T'}(x', \theta) = \sup_{\sigma_2} E_{\sigma_1} \left[ V_2^{T'+1}(x', \theta) + J_2^*(x', \theta) \right],
\]
where \( \sigma_1^{T'}, \sigma_2^{T'} \in \{1, 2\} \) is the PBNE policy pair at stage \( t \). The above system equations have to be solved backwardly from stage \( t \) to stage \( 0 \) with the boundary conditions \( V_1^{T'}(x^{T'+1}), V_2^{T'}(x^{T'+1}, \theta) \) at stage \( t = T + 1 \). However, the belief \( B'_2 \) in \([1]\) updates forwardly with the boundary condition \( B_0 \) at initial stage \( t = 0 \). These two equations are coupled, and we need to find the consistent pair of PBNE strategies and beliefs.

**3. COMPUTATION**

In the Bayesian update, the prior probability distribution \( B' \) is called a conjugate prior for the likelihood function \( \sigma_2^T \) if the posterior distribution \( B^{T+1} \) is in the same family as the prior distribution \( B' \). Similar to our previous work \([3]\), the defender divides the action space of the user into \( K + 1 \) time-invariant sets \( j \), i.e.,
\[
\sigma_2^{T'} = \{\sigma_2^T \cap \Theta \} \in \{0, \ldots, K\},\forall i \text{ which are mutual exclusive } \cap \Theta \in \Theta, \forall \theta \in \Theta.
\]
Each set represents a category and each \( d_2' \) uniquely corresponds to one category. Then, we can transform \( \sigma_2^T(d_2', \theta) \), the distribution of \( d_2' \), into a distribution of the corresponding category \( \sigma_2^T(k' \mid \theta) \) with the index \( k' \in \{0, 1, \ldots, K\} \). If we assume \( \sigma_2^T(d_2', \theta) \) to be a binomial distribution with parameter \( q = \theta \) and \( N = K \). The probability mass function (PMF) of category \( k \) is
\[
Pr(k) = \binom{N}{k} \theta^k (1 - \theta)^{N-k}.
\]
The prior belief \( B_0 \) is assumed to be a beta-distribution with hyperparameters \( \alpha^0 \) and \( \beta^0 \). Since binomial and beta-distributions are conjugate, the posterior belief converges to be a beta-distribution with updated hyperparameters \( (\alpha^{T+1'} + k, \beta^{T+1'}) = (\alpha' + k, \beta' + K - k) \), where \( k \) is the category that the user’s action at stage \( t \) falls into. Finally, we transform the belief conditioned on the actions back to the belief conditioned on the corresponding actions using the hard de-aggregation in which actions \( a_1', a_2' \) correspond to the same category \( k' \) share the same belief distribution \( B' \).

**Expanded State in Dynamic Programming:** Since the parameter update \( (\alpha' + k, \beta' + K - k) \) is sufficient to determine the belief update in \([1]\), the original system state \( x' \) and the belief state \( \alpha', \beta' \) compose an expanded state \( y' = \{x', \alpha', \beta'\} \). Since \( \alpha' + \beta' = \alpha^0 + \beta^0 + tk \), we only need one of the two parameters to determine the beta-distribution and the notation \( \theta \sim \beta' \) means that the type is of the beta-distribution with the hyperparameters \( (\alpha' + \beta' + tK - \beta) \).
The process shuts down when the safety constraints are violated such as a high reactor pressure, a high/low separator/stripner liquid level. The control objective is to maintain a desired production in the previous cyber networks, the attacker at the final stage reads of the sensors under his control. Unlike the stealthy transition, i.e., if his type value \( \theta < \theta \), he will choose NOP; otherwise he will choose to attack. The policy \( q^* (y^T, \theta) \) is semi-separating if the threshold \( \theta \in (0, 1) \) and is called a pooling strategy if the threshold \( \theta \not\in (0, 1) \). The defender cannot learn any knowledge about user’s type when he adopts pooling strategies which are independent of his type value.

### Cyber Transitions:

According to (2), the value functions \( V^T \) from stage \( t = 0 \) under the PBNE can be computed in the same fashion as in (3) if the future expectation \( V^{t+1} \) is assimilated into the direct stage reward \( J^t \) to form an equivalent stage utility. The best-response policy \( R^0 \) is

\[
R^0 (p^T | x^T) = \text{1} \left[ \{ (1 - p^T) r_a (x^T) \} \theta - c^T > 0 \right] = \text{1} \left[ \{ p^T (x^T) \} \theta - c^T > 0 \right],
\]

where \( \theta \in (0, 1) \). The PBNE mixed-strategies are given as follows.

\[
V^T (y^T, \theta) = \max_{q^T (y^T, \theta)} \left[ (1 - p^T (y^T)) (r_a (x^T) - r_a (x^T) \theta) - Q (y^T) \theta \right].
\]

(3)

PBNE mixed-strategies are given as follows.

\[
V^T (y^T, \theta) = \max_{q^T (y^T, \theta)} \left[ (1 - p^T (y^T)) (r_a (x^T) - r_a (x^T) \theta) - Q (y^T) \theta \right].
\]

A

The best-response policy \( R^0 (p^T | x^T) = \text{1} \left[ \{ (1 - p^T) r_a (x^T) \} \theta - c^T > 0 \right] = \text{1} \left[ \{ p^T (x^T) \} \theta - c^T > 0 \right] \), where \( \theta \in (0, 1) \) is called a pooling strategy if the threshold \( \theta \not\in (0, 1) \). The defender cannot learn any knowledge about user’s type when he adopts pooling strategies which are independent of his type value.

### Comparisons and Insights:

As shown in Fig. 3, a high value for the defender is the result of a healthy system state \( x^T \) as well as a belief state where the user is more likely to have a low type value. At the extreme state \( x^T = 1 \) where the reward incurred by the attack is so low that users with any type values choose not to attack. Then the defender does not need to defend \( p^T (y^T) = 0 \) and obtains the maximum utility.

To investigate the effect of the defender’s belief, we fix the system state \( x^T = 3 \) and change the belief state \( (\alpha^T, \beta^T) \) from (9, 1) to (1, 9), which means that the defender grows optimistically that the user is of a low threat level with a high probability. Since players’ value functions are of different scales in terms of the attacking threshold and the probability, we normalize the value functions with respect to their maximum values to illustrate their trends and make them comparable to the threshold and the probability as
shown in Fig. 4. When $\beta^T$ is small, the defender chooses to protect the system with a high probability $p^T \approx 0.67$, which completely deters attackers with any type values because the probability to attack $q^0 \approx 0$ when the attacking threshold $\tilde{\theta}$ is 1.

As the defender trusts more about the user’s legitimacy, the defending probability $p^T(y^T)$ decreases to 0 when $\beta^T = 9$. Since the defender is less likely to defend, the attacker bears a smaller threshold to launch the attack. However, the threshold will not decrease to 0 because the users with type values less than $\theta = 0.33$ (defined as the limiting threshold) cannot receive sufficient rewards from the attack even when the defender chooses NOP. The value of the limiting threshold depends on the expanded state $y^T$, yet it should always be larger than 0 because a user with type $\theta = 0$ has no incentive to attack. The resulting defending policy $p^T$ captures a tradeoff between security and economy and guarantees a high value for defenders at most of the belief states.

Finally, we investigate the multi-stage effect and the PBNE strategy pair for the long-term maximum utilities. To simplify the computation, we choose $c_0^1 = 2, c_1^1 = 0$, i.e., the expense of privilege escalations is more than lateral movements for the user and $v_1^0 = 0.1, c_1^1 > 2.39$. Then, the optimal policy for both players is to choose NOP for all three expanded states at stage $t = 1$. Therefore, although the attacker prefers to achieve a more advantageous system state $x^{t=1}$, aggressive behaviors $\alpha_1^1$ at stage $t = 1$ can decrease the defender’s trust and result in a less favorable belief state $(\alpha^{t=1}, \beta^{t+1})$. Because of the petty stage utility assumption, it is more beneficial for the attacker to remain stealthy at the intermediate stage and deceive the defender into a false belief. At the initial stage $t = 0$, $P_1$ chooses $\alpha_1^0 = 1$ with probability 0.20. $P_2$ chooses $\alpha_1^0 = 1$ when his type is larger than 0.95 and chooses $\alpha_1^0 = 0$ otherwise. The values are $V^0_1 = 59.97$ and $V^0_2 = -10 + 13.57 \cdot 1_{\{\theta > 0.95\}}$.

Therefore, it illustrates that the attacker of a large type value will take the risk of behaving aggressively to reach a desirable system state in the next stage because he would obtain higher rewards once the attack succeeds. As a countermeasure, $P_1$ chooses to defend yet only with a small probability.

5. CONCLUSION

In this work, we have explored a multistage incomplete information Bayesian game framework for designing proactive and adaptive defensive strategies for critical infrastructure networks with the presence of Advanced Persistent Threats (APTs). This framework well captures the multi-stage and multi-phase structure of APTs and their strategic nature to move stealthily within the network.

With the information asymmetry between the attacker and the system, the defender needs to form a belief dynamically on the type of the user using observable footprints. To enable the online computation of the belief, we have used conjugate priors to reduce Bayesian updates into parameter updates. This approach leads to a computationally tractable extended-state dynamic programming criterion that yields an equilibrium solution consistent with the forward belief update and backward induction. Finally, we have used Tennessee Eastman process as a case study to demonstrate the proposed framework. The numerical experiments have shown that our framework has significantly improved the security of critical infrastructures by strategically deterring the attacker and mitigating the APTs.

6. REFERENCES

[1] Batheilt, A., Ricker, N. L., and Jelali, M. Revision of the tennessee eastman process model. IFAC-PapersOnLine 48, 8 (2015), 309–314.
[2] Coppolino, L., D’Antonio, S., Romano, L., and Spagnuolo, G. An intrusion detection system for critical information infrastructures using wireless sensor network technologies. In Critical Infrastructure (CRI), 2010 5th International Conference on (2010), IEEE, pp. 1–8.
[3] Harsanyi, J. C. Games with incomplete information played by bayesian players, i–iii part i. the basic model. Management science 14, 3 (1967), 159–182.
[4] Huang, L., Chen, J., and Zhu, Q. A large-scale markov game approach to dynamic protection of interdependent infrastructure networks. In International Conference on Decision and Game Theory for Security (2017), Springer, pp. 357–376.
[5] Huang, L., and Zhu, Q. Analysis and computation of adaptive defense strategies against advanced persistent threats for cyber-physical systems. In International Conference on Decision and Game Theory for Security (2018).
[6] Manshaei, M. H., Zhu, Q., Alpcan, T., Bacsar, T., and Hubaux, J.-P. Game theory meets network security and privacy, ACM Computing Surveys (CSUR) 45, 3 (2013), 25.
[7] Van Dijk, M., Juels, A., Oprea, A., and Rivest, R. L. Flipit: The game of stealthy takeover. Journal of Cryptology 26, 4 (2013), 655–713.
[8] Zhang, T., and Zhu, Q. Strategic defense against deceptive civilian gps spoofing of unmanned aerial vehicles. In International Conference on Decision and Game Theory for Security (2017), Springer, pp. 213–233.
[9] Zhu, Q., and Rass, S. On multi-phase and multi-stage game-theoretic modeling of advanced persistent threats. IEEE Access 6 (2018), 13958–13971.