IST-152 Workshop on **Intelligent Autonomous Agents for Cyber Defence and Resilience**

Combining Game Theory and Learning for Dynamic Network Defense Strategies

Branislav Bošanský

Artificial Intelligence Center
Department of Computer Science
Faculty of Electrical Engineering
Czech Technical University in Prague
bosansky@fel.cvut.cz
+420 22435 7581

Christopher Kiekintveld

Department of Computer Science
University of Texas at El Paso
cdkiekintveld@utep.edu
+1 915 747 5564

Securing computer networks is a critical ongoing challenge for both private enterprise and government agencies and defense. Computer networks provide an access to valuable systems that can be damaged (e.g., by disabling a physical defense mechanism), and have valuable information that can be exploited (e.g., in a real-world conflict) or used for profit (e.g., selling a list of stolen credit card information). Attackers exploit the complex, living environment of automated systems and human users performing routine tasks to find gaps in security and conceal their actions. Interactions between defenders and attackers play out in this rich and ever-changing environment, and improving security poses a fundamental question for defense: *How should the defender choose security policies that improve security against sophisticated, adaptive attackers, without violating cost and usability constraints?*

We highlight three characteristics that make this decision problem particularly challenging. (1) Computer security is *adversarial*, with multiple decision-makers making decisions that interact and jointly affect the overall outcome. (2) The interaction is highly *dynamic* (sequential) in nature, with attackers and defenders able to make observations and react to one another over time. (3) There are many *unknown* and changing elements of the environment as well as the capabilities and motivations of the other agents.

The mathematical framework best suited to reasoning about adversarial decision problems is *non-cooperative game theory*. Game theory provides a variety of models for describing interactions, as well as solution concepts for finding good (ideally optimal) strategies, and computational subfields that provide algorithms for computing these solutions. Game theoretic strategies are widely used not only in network security [13, 9] but also in securing critical infrastructures [12] and protecting wildlife [3]. However, many basic game models do not handle the dynamic and unknown elements typical of cybersecurity decisions. We argue the addressing cybersecurity using game theory will require *developing and integrating* novel techniques from both *dynamic game theory* and *machine learning*.

Cybersecurity domains often involve sequential decisions. Each side performs some action, expects a reaction from the system or from the opponent, and reacts accordingly. For example, the attacker may adapt her strategy if one type of the attack fails, or she tries to hide her trace if there is a suspicion of a detection. The attack can thus be much more strategic and can last for a longer period of time, which is often in contrast with the types of games used for modeling physical security. One reason is that gathering information is easier and less risky in a cyber scenario than in an attack on a physical system (e.g., observing the network traffic costs much less than conducting on-site reconnaissance, the chance of being detected by the defender is smaller).

Solving dynamic games (known as *extensive-form games* that model interactions with finite steps or *stochastic games* that model interactions with infinite (or indefinite) number of steps) is a computationally challenging task. Despite the computational challenges, there exists a collection of algorithms that can be used to solve dynamic games [15, 4, 1, 14, 7] and recent results show that state-of-the-art algorithms are able to find super-human strategies in non-trivial games [2, 11, 10].

Game theory can provide optimal reasoning, but only if a model is fully specified and has only so-called "known unknowns." For example, the defender may not have full information about the current situation in the network, but has useful beliefs about the possible states the network could be in, and is able to maintain accurate probabilistic beliefs over these possibilities as new information is observed. However, it is often the case that a complete model is very challenging to specify in network security. There may also be "unknown unknowns" that violate fundamental assumptions, such as an attacker who uses a zero-day exploit that is not known to the defender. In these situations classical game-theoretic models can be cumbersome and of limited value, while methods that focus on adaptation and learning are well-suited to these problems.

When little is known about the environment for a decision problem, the focus is generally on *exploring* the environment and *learning* from the experience to improve future decisions. A wide variety of machine learning methods have been developed for learning from data, but the subfield of reinforcement learning is particularly relevant since it focuses on learning from direct experience of an environment. A class of learning problems known as *multi-armed bandits* [5] highlight the balance between exploring for new information and exploiting the current information to improve decisions, and many algorithms have been developed for optimizing learning behavior in various contexts. We have recently applied learning methods from the multi-armed bandit literature to the problem of detecting exploits in cybersecurity [6]. However, most learning methods do not account for the adversarial nature of the cybersecurity domain, and do no have a strong model of the underlying dynamic structure to guide learning and decision-making.

We argue that new research is needed to combine the strengths of game theory and machine learning to address the novel challenges that arise in cybersecurity decision-making. The underlying interactions can be effectively modeled as a dynamic game where there is a competition between the defender and the attacker, and most of the possible actions and results for these actions are known. However, there are often significant parts of the model that are initially unknown, or which must react to a changing environment that cannot be fully anticipated. Therefore, the defender must be able to explore for possibilities that the current game-theoretic model of the system does not account for and adapt the model accordingly. We have done preliminary work on combining equilibrium methods with machine learning in the context of border security [8], but much more work is needed to apply these methods in cybersecurity. However, there is great potential since the combination of these theoretical frameworks can provide new

guarantees on the optimality and adaptivity of decision-making in dynamic adversarial settings, ultimately allowing defenders to make better decisions to reduce cybersecurity risks.

References

[1] B. Bošanský, C. Kiekintveld, V. Lisý, and M. Pěchouček. An Exact Double-Oracle Algorithm for Zero-Sum Extensive-Form Games with Imperfect Information. *Journal of Artificial Intelligence Research*, 51:829–866, 2014.

[2] M. Bowling, N. Burch, M. Johanson, and O. Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218):145–149, 2015.

[3] F. Fang, P. Stone, and M. Tambe. When Security Games Go Green: Designing Defender Strategies to Prevent Poaching and Illegal Fishing. In *In Proceedings of 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[4] A. Gilpin, J. Pena, and T. Sandholm. First-Order Algorithm with O(ln(1/epsilon)) Convergence for epsilon-Equilibrium in Two-Person Zero-Sum Games. *Mathematical Programming*, 2011.

[5] J. Gittins, K. Glazebrook, and R. Weber. *Multi-armed bandit allocation indices*. John Wiley & Sons, 2011.

[6] M. P. Gutierrez and C. Kiekintveld. Bandits for cybersecurity: Adaptive intrusion detection using honeypots. In *AAAI Workshop: Artificial Intelligence for Cyber Security*, 2016.

[7] K. Horák, B. Bošanský, and M. Pěchouček. Heuristic Search Value Iteration for One-Sided Partially Observable Stochastic Games. In *In Proceedings of AAAI (to appear)*, 2017.

[8] R. Klíma, V. Lisỳ, and C. Kiekintveld. Combining online learning and equilibrium computation in security games. In *International Conference on Decision and Game Theory for Security*, pages 130–149. Springer, 2015.

[9] A. Laszka, W. Abbas, S. S. Sastry, Y. Vorobeychik, and X. Koutsoukos. Optimal thresholds for intrusion detection systems. In *Proceedings of the Symposium and Bootcamp on the Science of Security*, pages 72–81. ACM, 2016.

[10] M. Moravcik, M. Schmid, N. Burch, V. Lisý, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling. DeepStack: Expert-Level Artificial Intelligence in No-Limit Poker. *Science*, 2017.

[11] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, Jan. 2016.

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

[13] O. Vanek, Z. Yin, M. Jain, B. Bosansky, M. Tambe, and M. Pechoucek. Game-theoretic Resource Allocation for Malicious Packet Detection in Computer Networks. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 902–915, 2012.

[14] J. Čermák, B. Bošanský, K. Durkota, V. Lisý, and C. Kiekintveld. Using Correlated Strategies for Computing Stackelberg Equilibria in Extensive-Form Games. In *Proceedings of AAAI Conference on Artificial Intelligence*, pages 439–445, 2016.

[15] M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione. Regret Minimization in Games with Incomplete Information. *Advances in Neural Information Processing Systems (NIPS)*, 20:1729–1736, 2007.