isr PRESENTS

Yevgeniy Vorobeychik

Adversarial AI, form Models to Practice

Yevgeniy Vorobeychik is an Associate Professor of Computer Science & Engineering at Washington University in St. Louis. Previously, he was an Assistant Professor of Computer Science at Vanderbilt University. Between 2008 and 2010 he was a post-doctoral research associate at the University of Pennsylvania Computer and Information Science department. He received Ph.D. (2008) and M.S.E. (2004) degrees in Computer Science and Engineering from the University of Michigan, and a B.S. degree in Computer Engineering from Northwestern University. His work focuses on game theoretic modeling of security and privacy, adversarial machine learning, algorithmic and behavioral game theory and incentive design, optimization, agent-based modeling, complex systems, network science, and epidemic control. Dr. Vorobeychik received an NSF CAREER award in 2017, and was invited to give an IJCAI-16 early career spotlight talk. He also received several Best Paper awards, including one of 2017 Best Papers in Health Informatics. He was nominated for the 2008 ACM Doctoral Dissertation Award and received honorable mention for the 2008 IFAAMAS Distinguished Dissertation Award.

Artificial Intelligence (AI) and machine learning (ML) techniques are increasingly common in security applications, such as malware and intrusion detection. However, ML models are often susceptible to evasion attacks, in which an adversary makes changes to the input (such as malware) in order to avoid being detected. Commonly, evasion attacks on detection systems study such attacks through two modes of abstraction: 1) the entire decision pipeline is abstracted into measures of accuracy, false positives, and false negatives, ignoring the issue of how alerts are handled in practice where only a small subset can ever be closely examined, and 2) robustness of ML is studied using simplified feature-space models of attacks, where the attacker changes ML features directly to effect evasion, while minimizing or constraining the magnitude of this change.

In this talk, I will describe our recent studies of these two problems in evasion-robust detection. First, we investigate the effectiveness of "feature-space" approaches to designing robust ML in the face of attacks that can be realized in actual malware (realizable attacks). We demonstrate that in the context of structure-based PDF malware detection, such techniques appear to have limited effectiveness. On the other hand, they are quite effective with content-based detectors. In either case, we show that augmenting the feature space models with conserved features (those that cannot be unilaterally modified without compromising malicious functionality) significantly improves performance. Finally, we show that feature space models can enable generalized robustness when faced with multiple realizable attacks, as compared to classifiers which are tuned to be robust to a specific realizable attack.

In the second part of the talk, I describe our work on deciding which of a large number of alerts to choose for further investigation---often a necessary step in the detection pipeline. Here, we propose a game theoretic model of the interaction between the investigation system and an attacker who chooses attacks to balance the likelihood of detection with attack value (based on its expected consequences). This game gives rise to complex dynamics, with combinatorial action and state spaces for both the investigator and attacker. Our approach for computing an equilibrium, and, consequently, the best investigation policy, in this game is a combination of an actor-critic reinforcement learning with a neural network function approximation for both actor and critic, and a double-oracle method. We use case studies in fraud detection and intrusion detection (the latter based on Suricata) to demonstrate the effectiveness of our approach.

Wednesday, August 21, 2019
2:00pm to 3:00pm
4215 Gates Hillman Center

isr institute for SOFTWARE RESEARCH

School of Computer Science
5000 Forbes Avenue, 5212 Wean Hall
Pittsburgh, PA 15213-3890

Carnegie Mellon University