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

Learning, Simulating, and Predicting Adversary Attack Behaviors for Proactive Cyber Defense

Shanchieh Jay Yang[1]*, Computer Engineering, Rochester Institute of Technology*
Aunshul Rege[2]*, Criminal Justice, Temple University*
Michael E. Kuhl[3]*, Industrial and Systems Engineering, Rochester Institute of Technology*

83 Lomb Memorial Drive, Rochester NY, 14623, USA
(Tel) 1-585-475-2987
(Fax) 1-585-475-4084
[1]jay.yang@rit.edu, [2]rege@temple.edu, [3]michael.kuhl@rit.edu

1. Proactive Cyber Defense

In 2016, the cyberthreat landscape showcased advanced attack techniques, escalated attack frequency, and high levels of adversarial sophistication (Kulkarni 2016). Conventional cyberattack management is response driven, with organizations focusing their efforts on detecting Indicators of Compromise, or threats (Kulkarni 2016). This reactive approach has limited efficacy, as it does not capture advanced and sophisticated adversaries, mutating or unknown malware, living-off-the-land techniques, or new variants being deployed (Kulkarni 2016). Furthermore, responding to incidents after the attack has occurred is costly for two reasons. First, the attack has successfully occurred and damage has occurred in the form of data theft, system manipulation, service/functionality disruption, or the like, which is costly to fix (Barnum 2013). Second, during the attack, the adversary may have established several footholds in different parts of the targeted system. Identifying and eradicating these footholds are costly with regards to manpower and time (Barnum 2013).

The average time taken to identify and contain data breaches caused by malicious or criminal attacks was 229 and 82 days, respectively, and cybercrime detection and recovery activities accounted for more than 55 percent of total internal company activity costs in FY 2016 (Ponemon 2016). US organizations had the highest average cost of cybercrime ($17.36 million), with cybercrime costs in Germany and the UK averaging at $7.84 million and $7.21 million, respectively (Ponemon 2016). There is thus an immediate need for a *paradigm shift* in the area of cybersecurity. Security experts are calling for *anticipatory* or *proactive* defense measures that focus on Indicators of Attack that identify adversarial behavior and movement (Barnum 2013, Kulkarni 2016). Doing so requires a timely comprehension and predictive analysis of adversary decision-making capacities, which are currently downplayed in existing research.

Imagine a theoretically grounded system that learns, simulates, and predicts the attack progressions of adversaries of different intents, tactics, capabilities, and preferences. Through limited data accompanied with theoretical explanation, the system learns one or few adversary behaviors with salient features, extrapolates from the learned behaviors to many simulated ones, and generates plausible future activities based on the observed and extrapolated behaviors. The extrapolated behaviors and plausible futures can be key ingredients of proactive cyber defense measures, including providing anticipatory intelligence to human and autonomous agents. The following section provides a brief description of interdisciplinary research directions that address some of the current gaps towards such a system.

2. Theoretically Grounded Learning, Simulation, and Prediction

Criminology Theories: According to Routine Activity Theory (*RAT*), a criminological theory, crime is more likely to occur when three elements converge in space and time: (i) a capable offender, (ii) a suitable victim or target, and (iii) the absence of capable guardianship (Cohen & Felson 1979). RAT offers more about where and when crimes are likely to occur (when the three elements converge) than about why crime is likely to happen (why and how this convergence results in crime) (Wikstrom & Treiber 2016). Furthermore, the interaction of the three RAT elements is dynamic and shifts as the cyberattack progresses (Sutton 2012).

In the criminological discipline, 'crime scripts' provide a "standardized, systematic and comprehensive understanding of the crime commission processes" (Leclerc 2016; Cornish & Clarke 2002). Crime scripts also help identify the decisions, actions, and resources that are needed at each stage for the successful completion of the crime (Leclerc 2016). In the context of cybercrime, as conducted by state actors, cyber criminals, hacktivists, etc., crime scripts are captured by intrusion chains. Barnum's (2013) intrusion chain model (Figure 1) illustrates how adversaries study their targets, break into the targeted system, establish footholds, pivot and move laterally to strengthen their presence, and repeat the process until their objectives are completed.

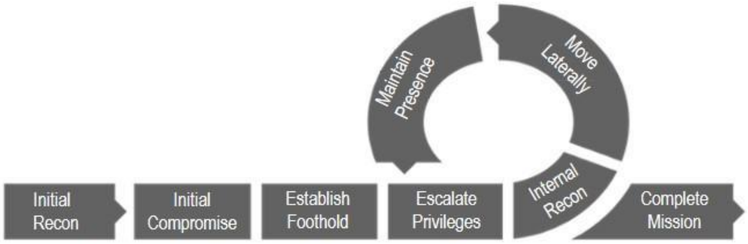
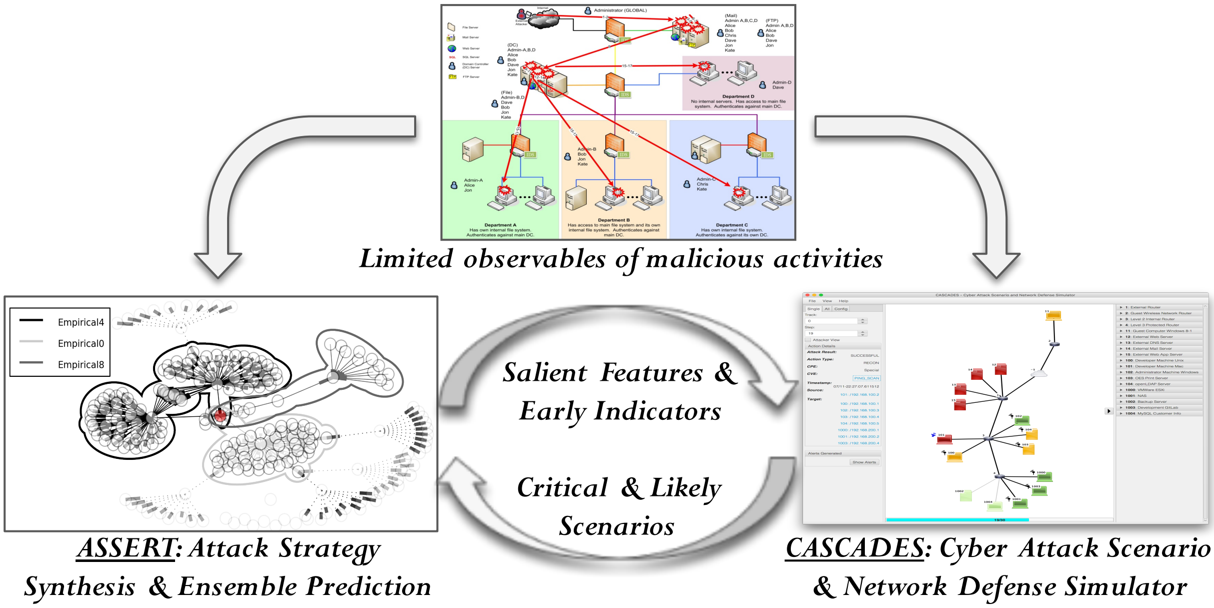


Figure 1: FireEye's Attack Life Cycle (Barnum 2013)

Existing criminological theories can use additional refinement to capture the dynamics of cybercrime, to examine the continually changing interaction between offender, target, and guardian (OTG) along cyberattack trajectories, and to explain how variations in OTG impact dynamic adversarial attack trajectories. The development of an enhanced theory will require interdisciplinary study where computational techniques offer insights and validation of attack behaviors learned and extrapolated from real-world attacks (albeit limited). A preliminary study has shown that as the crime process unfolds, the interaction between the offender, victim/target, and guardianship fluctuates, which determines whether the crime will stop or progress/regress to the next/previous stage (Rege 2016, Rege 2017). Understanding the dynamics and the decision-making capacity of adversaries will be critical to developing sound computational learning, simulation, and prediction systems.

Computational Techniques: Independent of the criminology theories, several probabilistic models have been developed to represent the interdependencies between system vulnerabilities and observables of malicious activities (Qin 2004, Fava 2008, Noel 2009, Yang 2014). This set of works infer the probabilistic dependencies using machine learning through observed malicious activities and/or based on specific properties of system exploits. While the learned models might reflect, implicitly, the adversarial behavior and be used to predict attack actions, they are limited in criminological/behavioral grounding and lack the ability to explain why and how attackers make specific movements. In addition, it is unlikely one will have ample data to comprehensively learn about cyber adversaries given the vast and fast-changing attack landscape and tactics. This calls for a new solution where limited data can be used to learn salient features and extrapolate to additional attack scenarios representing a broader spectrum of evolving adversarial behaviors.

Figure 2 shows a framework where observables of cyberattacks are fed to both *ASSERT* – an ensemble learning system that continuously creates and refines hypothesized attack models by integrating Dynamic Bayesian Network (DBN), Clustering, and Generative Adversarial Network (GAN), and *CASCADES* – a simulator that generates attack scenarios utilizing Monte-Carlo and Importance Sampling over the attack action space subject to attacker capability, opportunity, intent, and preference (Moskal 2014, Krall 2016, Moskal 2017). The two systems also feed to each other to enhance the learning process through simulated data and to provide salient features that can be used to guide simulation.



Limited observables of malicious activities
Empirical4
Empirical0
Empirical8
Salient Features &
Early Indicators
Critical & Likely
Scenarios
ASSERT: Attack Strategy
Synthesis & Ensemble Prediction
CASCADES: Cyber Attack Scenario
& Network Defense Simulator

Figure 1: A framework that takes limited data to synthesize hypothesized attack models and to simulate critical attack scenarios.

Because cyberattack data is limited and evolving without ground truth of agent behaviors, ASSERT must extract features through a carefully crafted GAN, use these features to create DBN-based attack models, and evaluate the quality of observable-model pairing using the concept of clustering. Current works have shown success in dynamically creating attack models (Strapp 2014) and refine them with a cluster validity index (Saxton rev). Attack models resulting from this process may also enhance the learning process in GAN. Meanwhile, referencing the features and models from ASSERT as well as a dynamic criminological theory, CASCADES may generate simulated attack scenarios along with observables to complement the limited real-world data.

3. Transformative Impact

Cyber defense must be proactive, utilizing anticipatory intelligence that enables actionable resilience. A theoretical grounded learning, simulation, and prediction system will be a key to enhance the intelligence of human and autonomous agents. A novel ensemble of advances in machine learning and simulation that incorporates the dynamics of cyber adversary decision making process will be at the frontline to bring forth this new era of cyber defense.

References

Barnum, S. (2013). Standardizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX™). MITRE Corporation, Retrieved July 2016. Online at http://www.mitre.org/sites/default/files/publications/stix.pdf

Cohen, L. E., & M. Felson (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. American Sociological Review, 44, 588–608.

Cornish, D. B., & R.V. Clarke (2002). Analyzing Organized Crimes. In A. Piquero & S. G. Tibbetts (Eds.), Rational choice and criminal Behavior: Recent research and future challenges. New York: Routledge.

Fava, D. S. (2008), S. R. Byers, and S. J. Yang. Projecting cyberattacks through variable-length Markov models. IEEE Transactions on Information Forensics and Security, 3(3):359–369, September 2008.

Krall A. (2016), M. E. Kuhl, S. J. Yang, and S. Moskal, "Estimating the likelihood of Cyber Attack Penetration using Rare-event Simulation," in Proceedings of 2016 IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2016), December 6-9, Athens, Greece.

Kulkarni, A. (2016). "The Five Core Components of Proactive Cybersecurity". Retrieved December 5, 2016. Online at http://www.techzone360.com/topics/techzone/articles/2016/12/05/427743- five-core-components-proactive-cybersecurity.htm

Leclerc, B. (2016). "Crime Scripts" In Wortley, R., & Townsley, M. (Eds.), Environmental criminology and crime analysis. Routledge.

Moskal, S. (2014), B. Wheeler, D. Kreider, M. E. Kuhl, and S. J. Yang, "Context Model Fusion for Multistage Network Attack Simulation," in Proceedings of IEEE Military Communications Conference (MILCOM'14), Baltimore, MD, October 6-8, 2014.

Moskal, S. (2017), S. J. Yang, and M. Kuhl, "Cyber Threat Assessments via Attack Scenario Simulation over Integrated Adversary and Network Modeling Approach," accepted to appear in Journal of Defense Modeling and Simulation.

Noel, S. (2009) and S. Jajodia. Advanced vulnerability analysis and intrusion detection through predictive attack graphs. Critical Issues in C4I, AFCEA Solutions Series. International Journal of Command and Control, 2009.

Ponemon Institute. (2016). "2016 Cost of Data Breach Study: Global Analysis". Retrieved January 10, 2017. Online at https://ssl.www8.hp.com/ww/en/secure/pdf/4aa5-5207enw.pdf

Qin, X. (2004) and W. Lee, "Attack plan recognition and prediction using causal networks," in Proceedings of 20th Annual Computer Security Applications Conference, pages 370–379. IEEE, December 6-10, 2004, Tucson, AZ.

Rege, A. (2016), “Incorporating the Human Element in Anticipatory and Dynamic Cyber Defense,” in Proceedings of the 4th International Conference on Cybercrime and Computer Forensics (ICCCF). IEEE, June 12-14, 2016, Vancouver, Canada.

Rege, A., B. Singer, N. Masceri, & Q. Heath (2017), “Measuring Cyber Intrusion Chains, Adaptive Adversarial Behavior, and Group Dynamics,” in Proceedings of the 12th International Conference on Cyber Warfare and Security. March 2-3, 2017, Dayton, OH.

Saxton, J., S. J. Yang, & A. Okutan (rev), “Dynamic Model Generation and Classification of Network Attacks,” under review.

Strapp, S. & S. J. Yang (2014), “Segmenting Large-scale Cyber Attacks for Online Behavior Model Generation,” in Proceedings of 2014 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction (SBP14), April 1-4, 2014, Washington, DC.

Sutton, M. (2012). Contingency Makes or Breaks the Thief: Introducing the Perception Contingency Process Hypothesis. Retrieved February 5, 2017. Online at https://www.bestthinking.com/articles/science/social_sciences/sociology/contingency-makes-or-breaks-the-thief-introducing-the-perception-contingency-process-hypothesis

Yang, S. J. (2014), H. Du, J. Holsopple, and M. Sudit, “Attack Projection for Predictive Cyber Situation Awareness,” in A. Kott, R. Erbacher, and C. Wang (Eds.), Cyber Defense and Situational Awareness, Springer, pp. 239-261, 2014.

Wikström, P. O. H., & K. Treiber (2015), “Situational Theory: The Importance of Interactions and Action Mechanisms in the Explanation of Crime,” in A. R. Piquero (Ed), The Handbook of Criminological Theory, John Wiley & Sons, Inc.
.