The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence
Artificial Intelligence for Cyber Security: Technical Report WS-16-03

Socio-Cultural Modeling for Cyber Threat Actors

Paulo Shakarian and Jana Shakarian
Arizona State University
{shak, jshak}@asu.edu

Abstract

In this paper we describe the unique challenges to the important problem of socio-cultural modeling of cyber-threat actors and why they necessitate further advances in artificial intelligence – particularly with regard to interdisciplinary efforts with the social sciences.

Introduction

Cyber security is often referred to as "offense dominant" referring to the notion that the domain generally favors the attacker (Lynn, 2010). The reasoning behind this is simple: a successful defense requires total control over all pathways to a system while a successful attack requires only one. As a result, any given cyber-defense based on the hardening of systems will fall prey to a cyber-attack as perpetrators gain knowledge and resources. Solutions have ranged from sophisticated adaptive defense strategies to offensive cyber-operations directed against malicious hackers. However, these methods have various technical shortcomings – which range from the technical immaturity of adaptive defenses to consequences of aggressive cyber counter-operations which can lead to undesirable effects such as preemptive and preventative cyber war.

A recent trend in the cybersecurity industry has been a move toward "threat intelligence" where various sources of information about potential cyber-attackers are explored with the goal of pre-empting cyber-attacks before they occur. A key source of cyber-threat intelligence lies in the digital communities of the malicious hackers – a collection of sites, markets, chatrooms, and social media channels where information is shared, hackers are recruited, and the latest malware and exploits are bought and sold. While artificial intelligence and machine learning techniques for analyzing communities on the Internet are long-established across specialty areas such as data mining, information retrieval, and web science, we argue that the study of hacker communities combined with the goal of automating the collection and analysis of information about the activity of cyber threat actors produces some very unique challenges. In this position paper, we describe some unique characteristics of cyber threat socio-cultural environments and several challenging modeling problems for which various artificial intelligence techniques can be used to help solve.

Characteristics of Cyber Threat Socio-Cultural Environments

In our group, we have studied hacker communities from a qualitative standpoint (Shakarian, Shakarian, and Ruef 2015). Throughout this research, we have noted several unique characteristics in the online socio-cultural environments frequented by malicious hackers that make these communities distinct from other groups. Some of these characteristics include the following.

- *Bounded anonymity.* Individuals participating in the malicious hacker community online make efforts to hide their identity. Some however seek to maintain a consistent online persona to gain social status in the hacker meritocracy.
- *Participation in high-risk behavior.* Despite recent arrests for individuals associated with darknet markets as well as suspicions of law-enforcement infiltration, many individuals still participate in discussions about illegal activities in darknet forums, though access controls appear to increase. Likewise, individuals participate in hacktivist operations advertised through social media. A recent lab-based behavioral study has explored some of the potential factors that would lead an individual to participate in risky hacktivism activities (Bodford, 2015).
- *High incentives to cheat.* The existence of marketplaces where malicious hackers sell software and exploits to others is an environment where both parties are highly incentivized to cheat. For instance the sale of a

faulty product and violations of exclusive use agreements can be conducted with relative ease.

- *Ability to deceive.* The anonymous nature of these environments combined with the fact that various aspects of a malicious hacker's digital persona can be forged allow for deceptive activities to occur with relative ease.

These characteristics are interesting in several ways. First, from a sociological and behavioral standpoint, the freedom at which individuals in these communities discuss criminal activities as well as share information and code with individuals likely involved with computer related crimes (which itself is also a crime) begs the question how trust is afforded to enable observable social interactions. Second, the characteristics such as anonymity and deception lead to modeling challenges – perhaps requiring consideration of latent attributes. Third, aspects such as cheating may actually constrain models to a degree – hence leading to model simplifications.

Modeling Challenges

In this section, we describe a few major challenges for modeling socio-cultural cyber threat actor communities. Overcoming these challenges will provide new insights into this environment and also aide in higher-level tasks such as predicting cyber-attacks and understanding the development of exploits and malware by this community.

- *Establishment of social status in an anonymous environment.* In order for a malicious hacking community to exist, there must be anonymity, yet actors stand to gain from prestige earned in the hacker meritocracy, such as access to invite-only forums, trust in social interactions in general as opposed to undergoing frequent vetting processes. Modeling the accumulation of this latent quantity with which proxy measurements are challenging in non-anonymous environments – and the level of anonymity itself creates even more difficult challenges. However, in addressing these challenges, we can better identify significant cyber threat actors and associate a greater degree of confidence with their actions. Recently, there has been some initial, descriptive work on this topic (Abbasi et al., 2014).
- *Data-driven modeling of risk taking.* The adoption of risky behavior has gained attention in the computational social science literature using model-based approaches (Roos, Carr, and Nau, 2010). However, instantiating models based on data remains largely an open question. The issue is further complicated by limited data on verified activities – as not all cyber-attacks are reported in the open. The goals in establishing such models for the study of cyber-threats in determining when certain risky behavior will occur is likely to aide in prediction and preventative cyber defense.
- *Emergence and disintegration of trust-based communities.* For darknet marketplaces to thrive, *populations* of individuals have to make decisions to trust both those running the marketplace and many of the vendors. While there are established models for trust among individuals, understanding how the propagation of trust is initiated and spread in anonymous environments – which seem to discourage trust – remains an open question. By addressing this problem, we can better understand when a given cyber-exploit/malware marketplace will become well established.
- *Modeling deception hypotheses.* In order to properly attribute individual activity on the darknet to that seen in public in cases of cyber-attacks or attributing the author of a given malware or exploit, cyber-security analysts consider the "deception hypotheses" – the chance that some or all of the observed evidence can be planted by an adversary. Therefore, for models designed for problems relating to cyber-attribution, we must also consider the deception hypothesis. In some of our ongoing efforts, we are leveraging defeasible logic programming to explicitly consider the deception hypothesis.

Conclusion and Ongoing Efforts

In this paper, we have discussed some of the unique characteristics of the socio-cultural environment for cyber threat actors and associated modeling problems that are of interest to the artificial intelligence community. We are currently exploring these challenges in support of larger cyber-security goals such as attack prediction and critical infrastructure defense.

Acknowledgements. This work was supported by Arizona State University's Global Security Initiative (GSI) as well as the U.S. Office of Naval Research (ONR) NEPTUNE program.

References

Abbasi, A.; Li, W.; V. Benjamin, V.; Hu, S.; Chen, H. (2014) Descriptive Analytics: Examining Expert Hackers in Web Forums. *IEEE Joint International Conference on Intelligence and Security Informatics.*

Bodford, J. (2015) We are Legion: Hacktivism as a Product of Deindividuation, Power, and Social Injustice, Masters Thesis, Arizona State University.

Lynn, W. J. 2010. Defending a New Domain: The Pentagon's Cyberstrategy. *Foreign Affairs,* 89(5).

Roos, P.; Carr, R.; Nau, D. 2010. Evolution of state-dependent risk preferences. *ACM Transactions on Intelligent Systems and Technology* 1(1).

Shakarian, J.; Shakarian, P.; Ruef, A. 2015. Cyber Attacks and Public Embarrassment: A Survey of Some Notable Hacks. *Elsevier SciTechConnect.*