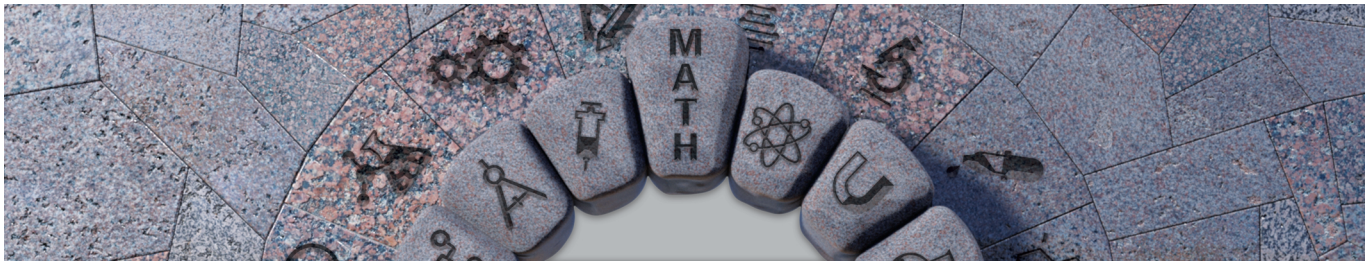

The Mathematics of Cyber Defense

John A. Emanuello and Ahmad Ridley

The speed, complexity, and ubiquity of cyber-attacks has never been more apparent and the far-reaching impacts they have on society demonstrate the critical need for robust security solutions, which can reduce the success of cyber-attackers *when* (not *if*) they compromise critical networks. Current cyber-defense capabilities are static and rules-based, i.e., they require a priori knowledge of the precise attacker tactics that will be employed. But this approach is unsustainable, given that malicious cyber actors rapidly change their approaches and chain their activities in complex and stealthy ways to thwart defenses. These challenges are driving a wide body of research and development of cybersecurity defensive solutions that are enhanced by artificial intelligence (AI), machine learning (ML), and data science. At their core, these approaches involve building mathematical models of cyber systems in order to derive information and devise strategies that enable their protection. However, unlike AI technologies applied in domains such as computer vision, natural language processing, and robotics, the complexities of the cyber domain present unique challenges, which we and others across government, academia, and industry have begun to address.

Before describing our work in greater detail, we present some key cybersecurity concepts. A computer *network*, such as the one at one's university or place of work, is collection of computers and other devices (collectively called *hosts* or *end-points*) that share common resources and infrastructure. Cyber attacks themselves are generally not any one event, rather a collection of events corresponding to the various phases of the attack and are oriented toward a particular goal. For example, cyber attackers, may exploit *vulnerabilities*, or defects, in software, to gain unauthorized access to a host, which may contain data of interest, and exfiltrate that data to their own systems. To detect and respond to attacks, cyber defenders have to examine lots of event data from disparate sources and various modalities, including log data, e.g., data from files that log activity corresponding to processes that occurred on a host; meta data corresponding to host-to-host communications, called *netflow*;[1] and rules-based alerts, e.g., data generated when events violate a set of host/network behavior rules manually created by a cyber defender. Some specialized organizations conduct analysis on malware and share resulting data with clients. Defenders may also rely on

John A. Emanuello is a senior research mathematician at the National Security Agency (NSA). His email address is jaeman2@uwe.nsa.gov.
Ahmad Ridley is a senior research mathematician at the National Security Agency (NSA). His email address is adridle@uwe.nsa.gov.

Communicated by Notices *Associate Editor Emilie Purvine.*

DOI: https://doi.org/10.1090/noti2493

[1] *E.g., a computer accessing the content of a web-page, or a remote desktop connection from your home computer to your work computer. These data include information about the hosts involved in the communication, the duration, bytes exchanged, and the communication protocol.*

publicly-available, technical reports describing cyber incidents or vulnerabilities.

Our research at NSA's Laboratory for Advanced Cybersecurity Research centers around building robust AI-enhanced solutions that augment cyber defenders in identifying malicious activity and deciding which remediation to take. The task of defending computer networks from complex cyber-attacks requires a combination of highly specialized skills, possibly unlike tasks from other AI/ML domains such as computer vision, natural language processing, and gaming. Additionally, cyber-defense involves challenges of dynamic, stochastic, and adversarial elements, requiring AI/ML approaches that generalize learning across multiple tasks, time scales, and network environments [4, 5]. As mathematicians on a multi-disciplinary team of computer scientists, data scientists, and behavioral scientists, we use our mathematical intuition and expertise in creative ways to address unique, interesting and challenging problems faced when applying AI/ML to cybersecurity. For example, we might use non-linear function approximation to estimate the quality of cyber responses, stochastic modeling to predict the evolution of host/network behavior, and non-convex optimization to optimize the real-world cost and utility of cyber systems under attack.

This begs the following question: *How does one apply AI/ML to augment defenders in protecting computer networks?* In our work, we decompose the tasks of the AI as follows: (1) detecting and understanding the attack; (2) determining the appropriate mitigating response.

In terms of detecting malicious activity, we have seen the promise of deep learning architectures, which are capable of learning complex patterns present in data, to build mathematical models of cyber behaviors. The nature of cyber also means that we must train these models in an unsupervised fashion, making them anomaly detectors, rather than classifiers. Deep autoencoders (AEs) are multi-layer perceptrons trained in unsupervised fashion to be an approximate identity function. An AE is generally decomposed as a composition of functions $f = g \circ h$, where the image of h is of smaller dimension than the input. As such, the model is a non-linear analogue of principal components analysis (PCA) [2].

We have demonstrated the efficacy of an AE-based anomaly detection scheme in two modalities, so far: host logs and netflow [1, 7]. The components of these logs are largely nominal features, e.g., IP addresses, port numbers, usernames, process names, file paths, etc., which need to be turned into numeric features that are both consistent with deep learning architectures and relevant to the task of detecting abnormal activities. To that end, we applied techniques inspired by Word2Vec[2] that enable us to embed these nominal values in $\mathbb{R}^n$ such that objects which behave similarly, with respect to the logs, are close in $\mathbb{R}^n$ [3, 7]. We concatenate these embeddings to represent a log and use that as input to the AE. Results are promising, but there is still work to be done on this front, including investigating how to enable the AI to chain anomalies from across data modalities and contextualize them with information on previous attack and threat intelligence, for a more complete analysis.

The autonomous detection of cyber threats is only one piece of the desired AI-pipeline. For the autonomous decision-making and response piece of the pipeline, we have investigated traditional AI techniques, such as planning and symbolic reasoning, and more recent ML ones, like reinforcement learning (RL)[6], which involve devising a decision policy which optimizes a utility function.

With the increasing speed, complexity, and ubiquity of cyber-attacks, AI/ML-enabled cyber-defense will prove vital to defending critical networks and infrastructure. The future of cyber-defense will continue to rely heavily on both human and machine. Indeed, the development of robust, autonomous AI/ML agents that, initially, augment the performance of human cyber defenders in a human-machine teaming manner, will, in all likelihood, eventually perform tasks at some level of autonomy, once their detection, reasoning, and response capabilities are trusted by human defenders. When these systems come to fruition, they will have been enabled by the power of mathematics.

References

[1] Andrew Golczynski and John A. Emanuello, *End-to-End Anomaly Detection for Identifying Malicious Cyber Behavior through NLP-Based Log Embeddings*, arXiv:2108.12276v1 (2021).

[2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2016. MR3617773

[3] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, *Efficient estimation of word representations in vector space*, arXiv:1301.3781 (2013).

[4] Andres Molina-Markham, Cory Miniter, Becky Powell, and Ahmad Ridley, *Network environment design for autonomous cyberdefense*, arXiv:2103.07583 (2021).

[5] Andres Molina-Markham, Ransom K. Winder, and Ahmad Ridley, *Network defense is not a game*, arXiv:2104.10262 (2021).

[6] Richard S. Sutton and Andrew G. Barto, *Reinforcement learning: an introduction*, 2nd ed., Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2018. MR3889951

[2] *Word2Vec is a technique devised by Mikolov, et al. which embeds words in a real vector space such that the similarity of two vectors, for example cosine similarity, is a good proxy for the semantic similarity of the corresponding words.* [3]

[7] Vance Wong and John Emanuello, *Robustness of ml-enhanced ids to stealthy adversaries*, arXiv:2104.10742 (2021).

John A. Emanuello

Ahmad Ridley

Credits

Author photos are courtesy of the US Government.