Matt Walsh
Professor Davidson
May 2021

Machine Learning in Cybersecurity

As the internet and sophisticated technologies have become highly integrated into modern society, the need for protection of personal data and critical systems has become increasingly urgent. On a personal level, cyber attacks aim to steal personal information in the form of bank accounts, social security numbers, and other sensitive information. On a wider scale, infrastructure systems such as energy, transportation, communication, and other important, widespread systems may be the targets of cyber attacks with the goal of causing massive damage. As IBM reports (2020), the average cost of a data breach or cyber attack on a company is 3.86 million dollars. For threats on the nation level, the consequences can be far more severe. With this large incentive for adversaries, cybersecurity systems must be constantly evolving to keep up with similarly evolving attacks, and machine learning methods may provide promising solutions to many of the issues in the cybersecurity industry.

Ambiguity in the Modern Battleground

The prominence of cyberattacks in international warfare has risen exponentially over the past decade, and the line between peaceful disputes and acts of war has become increasingly difficult to define. The anonymous nature of cyberattacks and the difficulties involved in identifying the motives of attacks makes it extremely difficult for countries to determine the severity of cyber threats. As described by Oliver Fitton (2016), the Russian Annexation of Crimea in 2014 is a great example of this new form of hybrid-warfare that involves a combination of physical and cyber attacks. Denial of service attacks were used to take down

Ukranian government organization websites, though the source of the attacks was not clear. NATO websites were also targeted, with Ukraniam “hacktivist” groups speculated at the perpetrators. Russian citizens had previously been accused of launching cyberattacks on Estonia, taking down websites for government organizations, banks, media sources, and other essential services. This form of hybrid war, in which cyber attacks were used to target government organizations on either side, introduced a unique legislative issue. Because these attacks were largely anonymous and unable to be accurately traced, world peace organizations such as NATO were unable to classify the level of severity of each situation. It is highly difficult to employ legislation to effectively prevent these attacks and hold the attackers accountable when many attackers do not act on behalf of the national governments.

The result of the 2007 attack on Estonia was the Tallin Manual on the International Law Applicable to Cyber Warfare. This document defines the appropriate responses to cyberattacks that are deemed as dangerous or that occur during times of physical conflict. However, the ambiguity in this manual and other legislative measures to prevent cyberattacks leads to a dangerous grey area where lesser cyberattacks become nearly impossible to respond to. Similar to Mutual Assured Destruction in the Cold War where the presence of nuclear weapons only encouraged smaller proxy wars, NATO’s response to cybercrimes has led to an abundance of smaller, more frequent cyberattacks. Adding to the complications of defining and punishing cybercrimes is the fact that many attacks are perpetrated by civilians with no national-level incentives. Countries must constantly defend against these attacks with little jurisdiction to retaliate, since there is often no way of proving whether a cyberattack is coming from a rival country’s government or a civilian group.

DARPA: CHASE Program

In 2017, the Defense Advanced Research Projects Agency (DARPA), received an $18 million budget for its Cyber Hunting at Scale program, known as CHASE. The program "will develop novel algorithms and analysis tools to dynamically collect data from across the network, actively hunt for advanced threats that evade routine security measures, and disseminate protective measures that auto- matically bolster the collective cyber defense posture" over a four year period. The program has five focuses: "threat detection and characterization; informed data planning; global analysis; protective measure generation and dissemination; and infrastructure for evaluation exercises." This program aims to use artificial intelligence and machine learning to enable its defense tactics to constantly learn and evolve as cyber attacks against the United States continue to become more frequent and more sophisticated. Currently, the amount of data collected on cyber attacks is magnitudes greater than human analysts have the capacity to monitor, and adding machines that can process and learn from this data will greatly enhance the cyberdefense systems in place.

Without machine learning or artificial intelligence, any irregularities in network data would need to be passed to a human analyst who would decide if the threat is real and if so, what the best course of action may be. This process is not only incredibly time consuming, but it also leaves substantial room for error. If the analyst does not correctly identify the issue or falsely identifies it as non-malicious, the attack could succeed. According to a BAE Systems cybersecurity analyst, about 80% of the current data collected for analyzing cyber risk is not able to be processed by human analysts. CHASE aims to implement programs that will not only recognize these irregularities, but also effectively identify what should happen next and learn from the petabytes of incoming data to better defend against future attacks. This approach

eliminates the need for such extensive human intervention and thus dispels much of the potential for error. With the help of BAE Systems, DARPA is in the prototype stage of an effective program with the goal of an autonomous system being released to government organizations in the coming years.

Automated Penetration Testing in CALDERA

In order to ensure that a defense system can prevent attacks, developers must act as attackers to test their system. By launching attacks and implementing offensive testing, defense developers can learn about and patch any flaws in their system. The most cost efficient way to implement these tests is to use an automated adversary. This adversary will use automated plans for penetration attacks to assess the efficacy of a security system and highlight any attacks that were successfully performed. Miller et al. describe the MITRE ATT&CK framework, which collects data from real-world cyberattacks and provides adversary threat models that represent attacks that are currently seen in "advanced persistent threats." In their 2018 report, the MITRE team details their development and testing of the CALDERA automated adversary emulation system. The goal of this system was not to stop attackers from entering a system by exploiting vulnerabilities, but to prevent any further damage or intrusions once the adversary had already gained access. The goal of an adversary once entering the system is usually either to move laterally within the system or changing permissions to access information or block others from using the system. To do this, CALDERA actions, or attacks, are simulated on a given defense system. At each iteration, different qualities of the system are assumed by the attacker and many attacks are employed. If an attack is successful, it receives a vote that it is the best attack, or most likely to be successful. After all iterations are complete, the attack with the most votes is

executed on the system. This process allows the defense system developers to see which attacks are most successful on their system and where the flaws of the system are. However, there are many difficulties in making this system successful. One difficulty is that the success or potential damage of these attacks is hard to quantify, and the uncertainty in identifying a threat therefore may be high. These uncertainties mean that while an attack may be able to access credentials or information, there is no way to tell if that information is useful to the attacker. Additionally, the set of attacks and propositions that the simulation runs through must always represent the real world space of cyberattacks. The representation of the real world is incredibly difficult as that space is constantly expanding and if the implementation doesn't cover everything in the real world, then the testing procedure will not test for everything. The simulation process also becomes extensively time-consuming as the attack space and the amount of data for each attack increases. While this system can identify the reach of attacks on a system, further research is needed to better identify the goals of attackers.

Summary

Cyber attacks will only become more powerful and more frequent, and without the help of automated systems and machine learning the risks will grow exponentially. The challenges of tracking the source of attacks, reprimanding attackers, and instituting effective legislation means that the best solution to protect systems is one that is proactive rather than reactive. More focus needs to be added to the development of these defense systems through programs and software such as CHASE or CALDERA to ensure that infrastructure systems, enterprises, and other important networks with sensitive data or massive responsibility remain protected from attackers.

References

Brett, M., Duchak, G., Ghosh, A., Sharp, K., Cilluffo, F., Cardash, S., & Center for Cyber and

Homeland Security, The George Washington University. (2017). *DIGITAL THREATS SYMPOSIUM — FALL 2017 — COMPENDIUM OF PROCEEDINGS* (pp. 8-12, Rep.). Center for Cyber and Homeland Security at Auburn University. http://www.jstor.org/stable/resrep21011.6

Doubleday, J. (2017). DARPA budget proposes new cyber, machine-learning projects. Inside the Pentagon, 33(23), 7-7. https://www.jstor.org/stable/90009600

Fitton, O. (2016). Cyber Operations and Gray Zones: Challenges for NATO. Connections, 15(2), 109-119. http://www.jstor.org/stable/26326443

Henry, S., & Brantly, A. (2018). Countering the Cyber Threat. *The Cyber Defense Review, 3*(1), 47-56. http://www.jstor.org/stable/26427375

Kramer, F., & Butler, R. (2019). CYBERSECURITY: CHANGING THE MODEL (pp. 5-20, Rep.). Atlantic Council. http://www.jstor.org/stable/resrep20932.5

Matheson, C. (2019). From Munitions to Malware: A Comparative Analysis of Civilian Targetability in Cyber Conflict. Journal of Law & Cyber Warfare, 7(2), 29-66. https://www.jstor.org/stable/26777971

Miller, D., Alford, R., Applebaum, A., Foster, H., Little, C., & Strom, B. E. (2018, June 28).

Automated Adversary Emulation: A Case for Planning and Acting with Unknowns. The MITRE Corporation. https://www.mitre.org/publications/technical-papers/automated-adversary-emulation-a-case-for-planning-and-acting-with.