Leveraging Artificial Intelligence and Machine Learning for Enhanced Cybersecurity: A Proposal to Defeat Malware

Emmanuel Boateng
Old Dominion University

Leveraging Artificial Intelligence and Machine Learning for Enhanced Cybersecurity: A Proposal to Defeat Malware

Emmanuel Boateng

Old Dominion University

April 14. 2023

Table of Contents

Abstract

Cybersecurity is very crucial in the digital age in order to safeguard the availability, confidentiality, and integrity of data and systems. Mitigation techniques used in the industry include Multi-factor Authentication (MFA), Incident Response Planning (IRP), Security Information and Event Management (SIEM), and Signature-based and Heuristic Detection.

MFA is employed as an additional layer of protection in several sectors to help prevent unauthorized access to sensitive data. IRP is a plan in place to address cybersecurity problems efficiently and expeditiously. SIEM offers real-time analysis and alerts the system of threats and vulnerabilities. Heuristic-based detection relies on detecting anomalies when it comes to the behavior of files and domains, whereas signature-based detection uses predefined malware codes and known signatures to help identify malware.

Artificial intelligence along with machine learning could enhance cyber detection and response by utilizing a vast amount of data and algorithms to help identify trends, make predictions, and take actions without human supervision. This paper discusses how this proposal can be accomplished and could help defeat malware.

Introduction

In a world where the protection of systems in the digital age is detrimental to both the consumer and producer sector, there needs to be an impregnable method to safeguard existing and future entities. The five main core of cybersecurity consist of application security, network security, cloud security, and internet of things security are all under attack and needs reformed security architecture in order to protect the confidentiality, integrity, and availability of these pillars of cybersecurity. The fundamental principle of modern health ‘Prevention is better than cure’ can also be applied to Cybersecurity. Most security departments and agencies act in the

form of a reactionary concept, where an issue would need to be current in order to reach a resolution.

The Internet Crime Compliant Center (IC3) which is an umbrella under the Federal Bureau of Investigations reported the significant trend of monetary damage caused by the penetration of systems with an estimate of 10.3 billion dollars in the year 2022 alone (Figure 1).

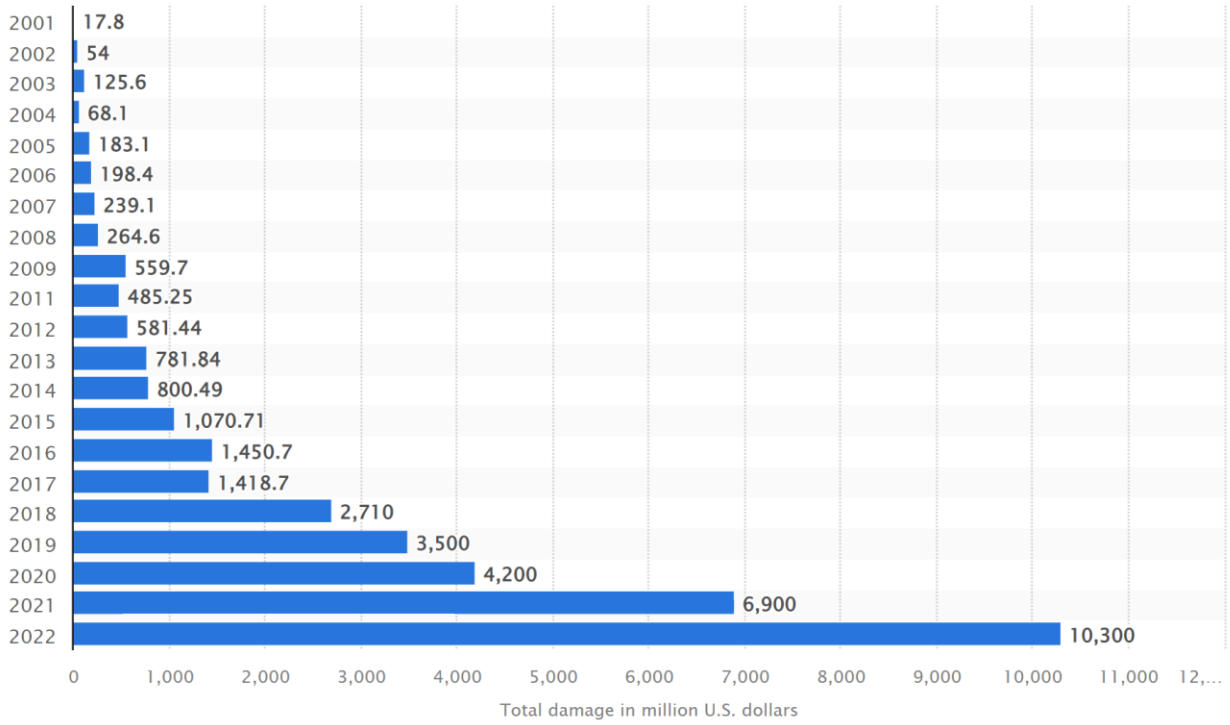

Figure 1

This is a perfect representation of how current systems are archaic and the need for innovation in order to curve this trend. The digitalization of our current reality will see advancements and growth until the end of time, which goes to show that this is not a temporary affair but a more constant and perpetual one.

Multi-Factor Authentication

Multi-Factor Authentication is a modern mitigation effort that provides an extra layer of security in many sectors for the protection of data. It insists on users to have more than a singular form of authentication in order to access data or a system. For financial services, MFA exists to

eliminate the unauthorized access to members bank accounts and critical financial data, and it is ever present where a large number of banks and financial entities require members to exercise MFA in order to access accounts. In Healthcare, personally identifiable information and protected health information are protected data under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) where an attack incident involving this information costs healthcare providers $9.42 million per occasion (Figure 2).

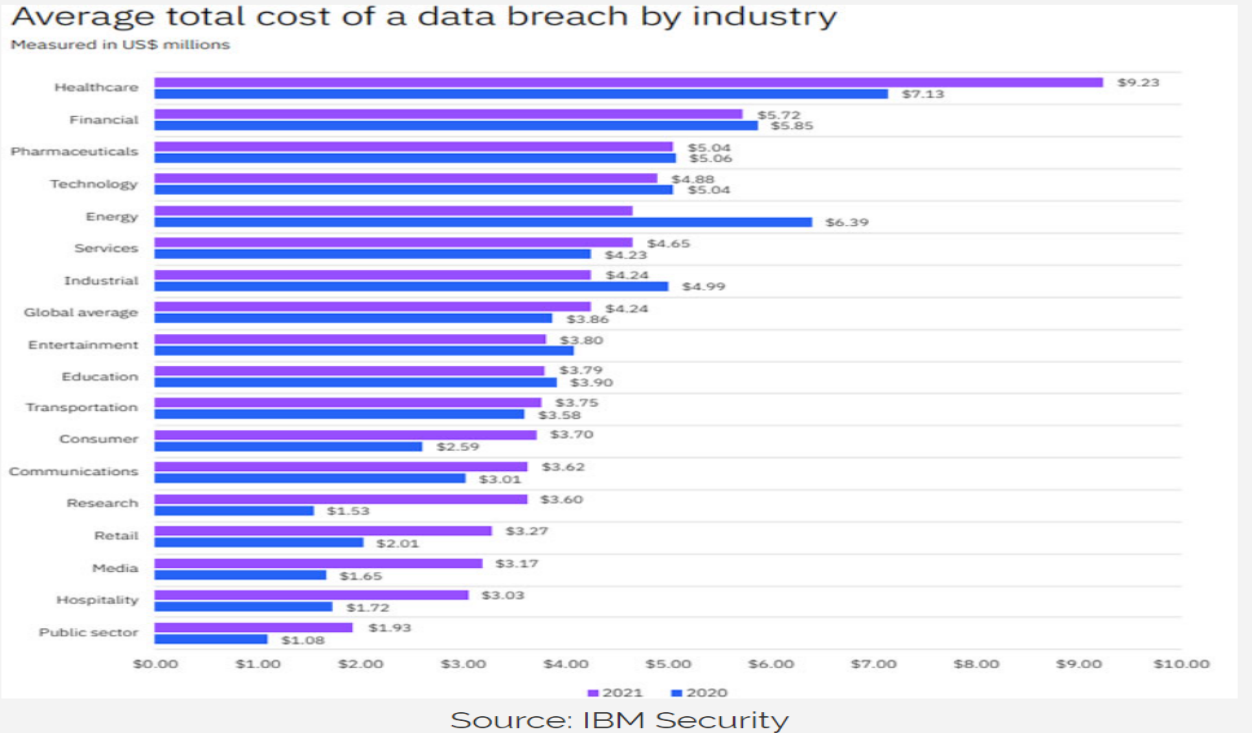

Figure 2

Government agencies also exercise this ideology to shield adversaries from gaining access to sensitive information that could affect the security of an entire nation.

The use of MFA is not a perfect method and some a lot of potential negatives that organizations have to consider when implementing it. There is a significant cost associated with the utilization of MFA where an institution requires additional hardware, software, and maintenance in order to properly achieve some success with MFA. There is also a common misunderstanding that MFA is adequate enough to protect an entire system, which is simply not

true and will lead to a compromised system. In certain sectors where ease of use is detrimental to the success of a company's model, MFA is abandoned due to the fact that it requires several steps in order for users to access accounts or systems.

Incident Response Planning

Any reputable organization has a team in place to confront cybersecurity incidents, and the majority of them have an incident response plan (IRP) with clear procedures that provide a synopsis on how to effective and methodical. An IRP is present to not only diminish the impact of an attack but to respond effectively and promptly. The contents of an IRP involve regulatory compliance where it lists the required items in order to adhere to local or federal mandates in order to avoid penalties and fines. The Cybersecurity and Infrastructure Security Agency is a federal agency that is created to assist in defending and securing the national cyberspace and proposed a documentation that lists the basics of a IRP that consists of having a trained staff, meeting with the CISA regional and law enforcement agency teams, and conducting an attack simulation exercise before a cybersecurity incident even occurs**.**

Incident Response Plan should be the last resort since when exercising this ideology equates to the fact that an incident has already occurred. The principle of cybersecurity is the protection of systems against cyber threats, and many sectors rely so much on IRP instead of applying this energy into preventing incidents from occurring. If an organization spends numerous US dollars on IRP, then it simply has admitted to defeat that their infrastructure is not strong enough or an attack is bound to occur in the near future. There is nothing wrong with preparing for the future, but the most important thing is deciding how to prepare for this said future. Preparing for it through response and not active prevention is already a failed design.

Security Information and Event Management

The combination of both security information management (SIM) and security event management (SEM) to provide real time analysis deriving from security alerts administered by network hardware and applications is the core foundation of Security Information and Event Management (SIEM). The use of intrusion detection/prevention systems, firewalls, and antivirus software are examples of numerous sources that SIEM systems use to analyze and counter security incidents in real time.

Algorithms and correlation rules are present in the execution of SIEM systems and remain the bedrock of the success of this methodology. SIEM systems are honestly a technology-based incident response program that usually specializes in monitoring and management of security incident and tries to isolate a threat after the fact and does not perform well in preventing an incident from occurring.

Signatured-Based and Heuristic Detection

The field of cybersecurity has mastered the act of detection and discovery with the approach of heuristic detection which involves the identification of unknown and new threats that is solely based on behavior instead of recognized signatures. The main behaviors that are monitored through the heuristic approach are changes to system files, communications between domains, and any sign of abnormal activity on a system. The mission of heuristic-based detection is to detect malware that cannot be identifiable with traditional signature-based detection methods. Heuristic detection implements the use of sandboxing; where files and domains are observed in a controlled virtual environment to analyze for behavior that deviates from normality. Heuristic based detection is not a flawless approach and technicians tend to struggle with its constant false positives where legitimate files often tend to get flagged as malicious.

Signature-based detection is an extremely similar approach but instead of depending on the analyzation of behavior it relies on known unique codes or patterns that coexist with malware and viruses in order to identify and respond to known threats. This approach is successful by cross-referencing a database of known malware signatures with network traffic and system files. Obviously, this approach is extremely limited for the reason that without an immense database of known malware signatures then this methodology simply fails. Also, when it comes to zero day attacks this approach is not practical because there are no signatures for malware that has not been detected. There needs to be an occurred event where the malware has been analyzed and processed and then stored in a database signature table in order for the signature-based detection to utilize it and prevent it in the future. Malware can easily evade signature-based detection through a technique that allows malware to change its code when propagating.

The most effective way is to not decide whether to apply either signature-based detection or heuristic based detection, but to adopt both methodologies as it tends to yield better results when it coexists with each other.

Artificial Intelligence and Machine Learning

The solution to eliminating threats can be achieved through the research of implementing artificial intelligence and machine learning to cyber detection and response. Artificial Intelligence (AI) is a field that is growing at a rapid pace where its abilities have not been fully unlocked yet. AI at its core uses a large amount of data along with algorithms to diagnose patterns, create predictions, and perform actions based on that data. The algorithm could either be a predefined set of rules to establish an AI's decision-making ability or could be solely based on machine learning (ML). Machine Learning simply utilizes training algorithms to eventually lead to the automation of improved performance and learning ability based on data. Furthermore,

a computer would have the ability to identify patterns in data, create predictions, and perform decisions without the manual input of a human being.

The proposed approach to a more practical system that will deliberately eliminate malware would be through the use of signature-based and heuristic detection in corporation with an artificial intelligence and machine learning attack system that incorporates the MITRE ATT&CK principles and techniques. The MITRE ATT&CK framework consists of tactics and techniques created for threat hunters but could be used to help eliminate malware. The development of software that exhibits these traits would help curve the trend significantly and will solely reside in the cloud. Therefore, institutions adopting this approach would not require additional hardware, software, or trained staff in order to fully employ this ideology on its systems. The benefits of artificial intelligence and machine learning in cybersecurity far outweigh the drawbacks of it. AI would be able to identify patterns and anomalies far better than the traditional methods mentioned above and will fine tune the process of blocking and quarantining malicious files and traffic.

However, there are still downsides to applying AI in this field and the most significant one is establishing a vast pool of data sets that consists of malicious and non-malicious codes along with anomalies. Preventing false positives would be the most constant issue where a soiled or tampered data set with lead to catastrophic events, so there would need to be an established system that safeguards the data set and makes it impenetrable. Another issue that could arise would be limited transparency; where it would be exceptionally difficult for cybersecurity professionals to analyze issues that relate to the system since eventually through machine learning the system would develop its own algorithms and methodologies.

Currently, The cybersecurity industry is using a form of artificial intelligence to better secure its system and 69% of organizations surveyed count on AI to identify threats and thwart attacks (Figure 3).

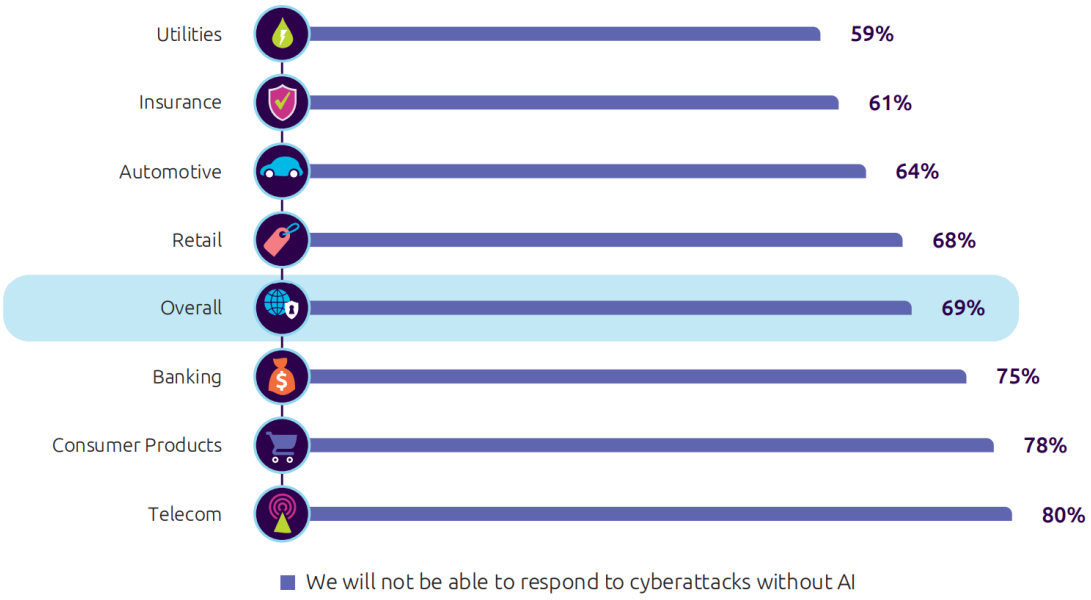

Figure 3

The development of AI in the field of cybersecurity needs to be accelerated, because malware writers are currently using the same method to create more efficient programs and cybersecurity professionals cannot afford to be behind the curve.

Conclusion

Cybersecurity is a necessity in today's digitalized world and there needs to further innovation to strengthen the field due to the dependence of digital systems. There are current mitigations in place to help protect systems and eliminate threats, but these methodologies are starting to become antiquated. Industries have an over-reliance on incident response planning and tend to become more reactionary and preventative, which is a design that will eventually lead to failure. The cost associated with multi-factor authentication influences an organization's decision on whether to add the extra layer of security or to just not adopt it.

Security Information and Event Management (SIEM) helps provide real-time analysis of security threats however it does not prevent security incidents from occurring. The use of signature-based and heuristic detection methods helps pinpoint unknown and new threats that already prevailed. The field of artificial intelligence has been rapidly advancing in recent years with an endless potential that could eventually eliminate malware completely. The combination of an immense detection system along with an attack system created through the implementation of machine learning would assist in winning the fight against malware and adversaries.

References

HIPAA Journal. (2022, December 5). *The average cost of a healthcare data breach is now $9.42 million*. HIPAA Journal. Retrieved March 1, 2023, from https://www.hipaajournal.com/average-cost-of-a-healthcare-data-breach-9-42-million-2021/

Centers for Disease Control and Prevention. (2022, June 27). *Health Insurance Portability and accountability act of 1996 (HIPAA)*. Centers for Disease Control and Prevention. Retrieved March 3, 2023, from https://www.cdc.gov/phlp/publications/topic/hipaa.html#:~:text=The%20Health%20Insurance%20Portability%20and,the%20patient's%20consent%20or%20knowledge.

Published by Ani Petrosyan, & 13, M. (2023, March 13). *Cyber crime: Reported damage to the IC3 2022*. Statista. Retrieved March 28, 2023, from https://www.statista.com/statistics/267132/total-damage-caused-by-by-cyber-crime-in-the-us/

Published by Ani Petrosyan, & 7, J. (2022, July 7). *Cyber crime incidents by industry and organization size 2021*. Statista. Retrieved March 21, 2023, from https://www.statista.com/statistics/194246/cyber-crime-incidents-victim-industry-size/

FBI. (2016, May 3). *Cyber crime*. FBI. Retrieved March 17, 2023, from https://www.fbi.gov/investigate/cyber

Incident response plan (IRP) basics - CISA. (n.d.). Retrieved March 5, 2023, from https://www.cisa.gov/sites/default/files/publications/Incident-Response-Plan-Basics_508c.pdf

Canfora, G., Di Sorbo, A., & Mercaldo, F. (n.d.). *Obfuscation techniques against signature-based detection: A case study*. Retrieved March 10, 2023, from https://www.researchgate.net/profile/Corrado-Aaron-Visaggio/publication/303459304_Obfuscation_Techniques_against_Signature-Based_Detection_A_Case_Study/links/5744241708ae298602f0fe55/Obfuscation-Techniques-against-Signature-Based-Detection-A-Case-Study.pdf

Capgemini Research Institute. (n.d.). *Reinventing cybersecurity with Artificial Intelligence*. Retrieved March 13, 2023, from https://www.capgemini.com/wp-content/uploads/2019/07/AI-in-Cybersecurity_Report_20190711_V06.pdf?o=12387

Belani, G. (n.d.). *Artificial Intelligence in cybersecurity: IEEE CS*. Artificial Intelligence in Cybersecurity | IEEE CS. Retrieved March 11, 2023, from https://www.computer.org/publications/tech-news/trends/the-use-of-artificial-intelligence-in-cybersecurity