International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 08 | August -2020 ISSN: 2582-3930

Innovative Cyber Threat Detection Using Machine Learning and Data Mining

Ms Anju,A,
Department of Information Technology, KCG College of Technology, Chennai, India,

Muzammil Sait A,
Department of Information Technology, KCG College of Technology, Chennai, India,

Antony Raj A,
Department of Information Technology, KCG College of Technology, Chennai, India,

Logith Kumar B,
Department of Information Technology, KCG College of Technology, Chennai, India,

Abstract: Security framework is of key note importance as its completely sensitive to breach and when done so leads to ambiguous lose for any typical organization. A security platform employs a variety techniques and mechanisms to detect security related anomalies and threats in a computer network environment. The security platform is "big data" driven and employs machine learning to perform security analytics .On to this scheme our paper presents one such innovation in the field of technology to detect the cyber threats before it becomes impossible to recover from such loss. The technique implemented here is Machine Learning Strategy under unsupervised data model we inhibit Structured Behavioural Model .This model is as suitable as it leads to anomly detection and intrusion detection on the basis of implemented code for the detection process. This paper inhabitively concultates the detection in two main key areas , one such area representing the Mallicious links being clicked by user unknowingly and malware junks the Pc and the other regardively losing the Internet connection (Wireless DDOS Attack) at the at most precision of having all hardware setup perfectly.The part where it outstands at the security inhabitance is Classifying of Cyber Attack using Data mining concepts at the precise conjurance.

I. INTRODUCTION:

Security remains a top concern as the world is becoming increasingly digitalized. With advances in network technologies, such as the Internet, access to cutting edge technology and research findings has never been this easy with research papers being made public daily and the digital world becoming increasingly open sourced. Unfortunately, cutting edge research and breakthroughs in technology are both available to the security analyst and cybercriminals who have various interests in making use of these technologies and information. Research and advances in the field of machine learning has resulted in algorithms and technologies for improving security solutions that help in identifying and decisively dealing with security threats. However, this also makes it possible for cybercriminals to use this knowledge in crafting and launching bigger and more sophisticated attacks. Cybercriminals have a huge advantage in the cyber war since, out of many attempts, they need to be right just once. For security, on the other hand, the desired success rate needs to be 100%. Research shows that in 2017, multiple organizations, business, individuals and applications were victimized by cybercriminals.. Stolen information included sensitive classified intelligence data, financial records, and personally identifiable information. The use of these kinds of information can be catastrophic, especially when it is made publicly available or sold on the black market. Some research statistics with regards to the impact of cyber security to businesses, organizations, and individuals include:

In recent years, cybercrime has been responsible for more that $400 billion in funds stolen and costs to mitigate damages caused by crimes .

It has been predicted that a shortage of over 1.8 million cybersecurity workers will be experienced by 2022 .

It's been predicted that organizations globally will spend at least $100 billion annually on cybersecurity protection .

Attackers currently make over $1 billion in annually revenue from Ransomware attacks, such as Wannacry and CryptoWall attacks

II. PROBLEM DEFINITION:

This paper mainly subdues to define the solution on variant methods to detect and classify the cyber threat that the organisation's Computers may face at circumstances which leads to loss of necessary data .The technological aspects lie on structural behaviour model and some concepts of supervised and semi supervised machine learning algorithms with a classification done through Data mining concepts.

Precisely the paper Conjures around the detection of malicious scripts sent to the user as a social engineering aspect and classifies the attack on the subdual criteria of malicious script present.

III. METHODOLOGY:

The Methodology proposed in this paper consists of different tasks evaluated at the same pace of execution which include Threat detection, Network Risk Scoring , Automotive Routine Security Task and Optimized Human Analysis.

III.I Threat Detection and Classification:

Machine learning algorithms can be implemented in applications to identify and respond to cyber-attacks before they take effect . This is usually achieved using a model developed by analyzing big data sets of security events and identifying the pattern of malicious activities. As a result, when similar activities are detected, they are automatically dealt with. The models' training dataset is typically made up of previous identified and recorded Indicators of Compromise (IOC), which are then used to build models and systems that can monitor, identify and responds to threats in real time. Also, with the availability of IOC datasets, we can use machine learning classification algorithms to identify the various behaviors of malwares in datasets and classify them accordingly. Studies have been made on behavioral-based analysis frameworks that make use of machine learning clustering and classification techniques to analyze the behaviors of thousands of malwares . This makes it possible to use the learned patterns to automate the process of detecting and classifying new malware. This can help security analysts or other automated systems to quickly identify and classify a new type of threat and respond to it accordingly using a data driven decisions. For example, by using a historic dataset containing detailed events of WannaCry ransomware attacks, a machine learning model can learn to identify similar attacks, thereby making it possible to automate the identification and response process of similar attacks. Machine learning techniques have also been used in IP traffic classification which can help automate the process of intrusion detection systems that can be used to identify behavioral patterns as in the case of DDOS attacks. With the increasing number of machine learning techniques, other studies have been focused on analyzing multiple machine learning solutions for intrusion detection systems including single, hybrid and ensemble

III.II Network Risk Scoring:

This refers to the use of quantitative measures to assign risk scores to various sections of a network, thereby helping organizations to prioritize their cyber security resources accordingly with regards to various risk scores. Machine learning can be used to automate this process by analyzing historic cyber-attack datasets and determining which areas of networks were mostly involved in certain types of attacks. Using machine learning is advantageous in the sense that the resulting scores will not only be based on domain knowledge of the networks but most importantly, the scores will be data driven. This score can help quantify the likelihood and impact of an attack with respect to a given network area and can thus help organizations to reduce to risk of being victimized by attacks. Studies have been carried out on the use of machine learning algorithms such K-Nearest Neighbor, Support Vector Machines, and Random Forest algorithms to analyze and cluster network assets based on their connectivity .

III.III Automotive Routine Security Task and Optimized Human Analysis.:

Machine learning can be used to automate repetitive tasks carried out by security analysts during security activities. This can be done through analyzing records/reports of past actions taken by security analysts to successfully identify and respond to certain attacks and using this knowledge to build a model that can identify similar attacks and respond accordingly without human intervention. Though it is difficult to automate the full security process and replace the human security analyst, there are some aspects of the analysis that machine learning can automate including malware detection, network log analysis, vulnerability assessments, such as network risk analysis. By incorporating machine learning in the security work flow, 'man and machine' can join forces and accomplish things at a degree of speed and quality that will have been otherwise impossible. With the exponential growth of artificial intelligence, we see an increasing number of tasks being automated. It is tempting to think that artificial intelligence will increase automation, and certain tasks that are currently performed by humans will be taken over by machines. This might be true in some cases, however there are numerous cases where the combination of artificial intelligence and human intelligence produce far better results than each will produce by itself. It is for this reason that we are currently seeing the rise of artificial intelligence companies with a focus on not only creating AI product for automating tasks, but creating products that enhancing and complement the productivity of human analysts. A well-known example of such a company is

Palantir, which creates products that make it easy for analysts to aggregate and make use of massive volumes of data. In other to enhance security analysts activities, studies have been carried out on the use of machine learning algorithms, such as genetic algorithms and decision tries to create applications that generate rules for classifying network connections . Other approaches go far as to implement a cognitive architecture to create an automated cyber defense decision-making system with expert-level ability inspired by how humans reason and learn . Cybersecurity analysts typically have to spend time responding to multiple events, which sometimes include false positives, which mostly turn out to be a waste of their time. Studies have been done to show that machine learning classifiers can be trained on alert data to identify and distinguish between false positives and true positives, thereby making it possible to create an automated system that will alert the analyst only on scenarios that include true positives .

IV. CONCLUSION

The paper thereby provides up in conculsion of different approaches used in detection of Cyber threats at an advencious detailed manner, thereby it brings up a major eyen on perspective of security as a of greater concern. Clearly it can be seen that machinelearning is a powerful tool that can be used for automating complex defense and offense cyber activities.With the recent advancements in machine learning including adversarial machine learning so as to constantly be on the lookout to make use of potential AI related security applications.This paper can act as basis for future research that can focus on analyzing existing security solutions and the various challenges of leveraging machine learning to develop and deploy scalable cybersecurity systems in production environments.

REFERENCES

[1] S. Larson. 10 biggest hacks of 2017. 2017, December 20. Retrieved: November 3, 2018, from https://money.cnn.com/2017/12/18/technology/biggest-cyberattacksof-the-year/index.html

[2] T. Rimo and M. Walth, "McAfee and CSIS: Stopping Cybercrime Can Positively Impact World Economies", McAfee, June 9, 2014.

[3] "2017 Global Information Security Workforce Study", Frost and Sullivan, May 2017.

[4] Worldwide Revenue for Security Technology Forecast to Surpass $100 Billion in 2020, According to the New IDC Worldwide Semiannual Security Spending Guide. 2016, October 12. Retrieved: September 21, 2018, from https://www.businesswire.com/news/home/20161012005102/en/Worldwide-Revenue-Security-Technology-Forecast-Surpass-100.

[5] A. Cuthbertson. Ransomware attacks have risen 250 percent in 2017, hitting the U.S. hardest. 2017, May 28. Retrieved: September 21, 2018, from http://www.newsweek.com/ransomware-attacks-rise-2502017-us-wannacry-614034.

[6] B. M. Cooper. Resiliency and Recovery Offset Cybersecurity Detection Limits. 2015, January 16. Retrieved: September 21, 2018, from https://www.afcea.org/content/resiliency-and-recovery-offsetcybersecurity-detection-limits.

[7] S. Dolev and S. Lodha, "Cyber Security Cryptography and Machine Learning", In Proceedings of the First International Conference, CSCML 2017, Beer-Sheva, Israel, June 29-30, 2017.

[8] S. Dolev and S. Lodha, "Cyber Security Cryptography and Machine Learning", In Proceedings of the First International Conference, CSCML 2017, Beer-Sheva, Israel, June 29-30, 2017.

[9] About. (n.d.). Retrieved: November 03, 2018, from http://www.palantir.com/.

[10] G. A. Wang, M. Chau, and H. Chen. Intelligence and Security Informatics: 12th Pacific Asia Workshop, PAISI 2017, Jeju Island, South Korea, May 23, 2017, Proceedings. Cham, Switzerland: Springer.