International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Semantic Cyber security Monitoring using Machine Learning

[1] Ramavath Vinod Kumar, [2] P.Manikanta, [3] P.Sai Linga, [4] P.Prashanth, [5] S.Paramesh

[1]Assistant Professor in Department of CSE Sri Indu College Of Engineering And Technology
vinodkumar9r@gmail.com
[2,3,4,5] UG Scholars Department of CSE Sri Indu College Of Engineering And Technology

Abstract: Security precautions have been taken in order to preserve both the accessibility and the authenticity of the information included within digital communities. In most cases, information protection methods will restrict persons from accessing, revealing, manipulating,or even erasing facts on both software and hardware technologies. These steps can be used to secure information. According to an assessment that was carried out by industry specialistsin the field of information security, new forms of cyberattack continue to surface in all business operations. After all of the data had been analysed, an assessment of the level of risk revealed that although it is not extremely dangerous in the majority of instances, it isextremely dangerous for valuable data, and the severity of those attacks is prolonged. This was discovered despite the fact that the data had been analysed. In order to identify and protect against a wide variety of cyber-threats, several levels of protection have already beenput into place. These layers of security primarily make use of a processed data feed or alertin order to disclose both predictable and stochastic behaviour. Cyber assaults have been shown to follow deterministic patterns, which indicate that they are neither random nor unbiased over time. Attacks that have been carried out in the past can be used to estimate the likelihood of attacks in the future. In a general sense, the deterministic approaches can be utilised to provide monitoring that is just somewhat correct.

Keywords : *Cyber Security, Network Monitoring, Machine Learning, Cyber Attacks*

I. INTRODUCTION

To raise awareness of growing cyber-threats and new attacks, systems that offer real-time assessment were created. While explaining the concept of a risk system, we introduce a define of software systems that provide real-time visibility into global cyber-attacks risk systems, which provide animated maps that are created from real-time data about attacks on the location. Real-time traffic data is designed to identify the different types of traffic that could be indicative of malicious intent. When it comes to finance, social development, and even our every-day lives, it has become increasingly difficult to deal with communication networks and facts systems. However, the rapid development of the internet offerings and digital communication networks, along with increasing use of records structures, make them vulnerable to one or more kinds of cyber-attacks.

A. *Cyber Security*

In this context, "cyber protection" is synonymous with "asset protection," which is made up of a variety of tools and strategies designed to defend assets (computers, networks, applications, and statistics) from assault, unauthorised access, change, or destruction. Community safety systems include PC (host) protection systems while PC (host) protection systems include community safety systems. No firewall, antivirus software, and intrusion detection device is completely essential, but these items have each, at a minimum, a firewall, antivirus software programme, and an intrusion detection gadget (IDS). Increasingly, computing generation engineering statistics, such as security and privacy, are important issues for computer scientists.

B. *Cyber Attacks*

Intentional use of laptop structures, generation-dependent businesses, and networks are utilised in cyber-attacks. The rise of documents, attachments, and malicious configurations on servers causes websites and applications to be attacked every day. Malicious code is used in an attack as both the attack and the malware are being completed at the same time. Information and identification robbery could jeopardise essential statistics because of it. Precise records about the attacks and beliefs are critical in order to avoid finding the poor results and taking preventative measures. [2] Computer-based attacks, such as denial-of-service (DoS) attacks, botnets, man-in-the-middle attacks, phishing, spear phishing attacks, password attacks, malware attacks, brute force attacks, etc, can be detected by a few unique purpose websites like Denial-of-Service (DoS) Attack, Botnets, Man-In-The-Middle Attack, Phishing, Spear Phishing Attacks, Password Attack, Malware Attack, Brute Force Attack.

C. *Machine Learning in Cyber Security*

Cyber Machine Learning takes on an important role in next-generation cyber security. As the next cyber security products develop, increasingly they incorporate AI and ML technologies. According to educational AI software on huge stores of data from the cyber security, community, or maybe physical facts, the cyber security is intended to get an organization's goal to reveal and avert average behaviour, however, without including a “signature” or sample. Cyber security experts predict that, over time, agencies will integrate machine learning into all levels of cyber security products. This latest development in deep learning and one-of-a-kind, promising technology has an undeniable impact on the overall network community. Current efforts have included numerous large advancements in numerous networking sub disciplines. In the future, a number of issues will be solved. To begin, the strength of the equipment's relationship to algorithms is a crucial undertaking for software. [3]

II. RELATED WORK

The project entitled “EMBER” presents the open dataset with labelled factors in order to enable successful training of predictive analytics and learning gadget models. Since the dataset contains specific penetration levels for the training of fashions and predictions, the dataset will serve as a fertile environment for malware.

Banin S, Dehghantanha A, Shalaginov A, and Franke K. Highlights the separate survey for malware detection that may include techniques and procedures. To arrive at higher degrees of accuracy and predictions, the methods and algorithms described in this work use highly powerful and advanced methods and algorithms. Sixteen - two students, doyen Sahoo, Chenghao Liu, and Steven C.H. Hoi, and Categorize and overview the components of research that attempt to counter various angles of Malicious URL Detection, such as design of functions and collection of policies.

N. Whitton, Crockett, A. Latham, & Proposed Predicting learning patterns in conversational creative tutoring systems by using a fuzziness rating system on a random sample of previous students' choices is possible. The publication is available online: B. Sun, S. Chen, J. Wang, H. Chen Described a method called noise-detection that is based on AdaBoost called AdaBoost Boosting through which one can decorate AdaBoost's robustness (2016)

E. M. El-Alfy, M.M. Awais, and M. Baig. Covered a new method of studying a feed-before ANN with a single hidden layer and a single output neuron. X, Pan, and Y. Luo are renowned performers. a structural dual vector machine proposing to implement K-nearest neighbour installations with a singular vector machine (KNNSTSVM). Instead of calculating the samples based on their beauty scores, the intra-beauty KNN approach ensures that certain weights are given to the samples so that they may help to embellish the structural facts. To speed up the education system, wasteful constraints are eliminated using the inter-eligance KNN technique. [10]

B. Ottersten, D. Aouada, and A. C. Bahnsen, they introduced a fee-sensitive selection tree which includes rules that rely on one of kind examples. Later, people began to use it (Adler et al., 2002; Mayhew et al., 2001; Cleveland et al., 2002; Atighetchi et al., 2002; and Greenstadt et al., 2002.) When using every okay-manner clustering and manual SVM, take advantage of every overlap between skills with the skills you have selected and tested in the MBM machine and studied in the relevant literature. [12]

The flora of F. U., P. Palmieri, A. Castiglione, and A. Santis When both randomness and burstiness of traffic behaviour are present; the classifier's general overall performance is affected. As long as ICS networks remain unaffected, these issues will have no impact on them. The subject of this study is Restricted Boltzmann Machines (RBM), or more specifically, Discriminative RBM (DRBM). The findings unveiled a novel method. This use of a non-labelled approach is comparable to the method applied with MBM, as there are no previous records on the records of individuals who have visited an anomalous site. [13]

III. OBJECTIVE OF THE PRESENT WORK

The ultimate goal of this work is to discover a gadget that detects network anomalies and cyber-attacks with no more infrastructures on the network, which is capable of locating those issues in a very short period of time by utilising data-mining tools without compromising the overall network. Here, the goal of the study is to look into work.

- A new set of rules for keeping the Network Security under watch will be suggested by the researcher
- Researcher will compare a new set of rules with the existing set of rules to determine how well they perform together.
- The next steps are to be completed in order to achieve objectivity
- Study activities and alerts in order to establish if they are connected/linked to assaults that are ongoing
- Buildings, networks, and applications should be covered.
- Discriminate between possible threats that may exploit, acquire, or take advantage of the vulnerabilities for unauthorised entry
- Examine the company's information technology to catch actual-time or nearly-real time cyber-attacks, security violations, or breaches, as well as symptoms of anomalous or symptomatic sports.
- Provide records and documents.

IV. WORK PLAN AND METHODOLOGY

a) *Approach*

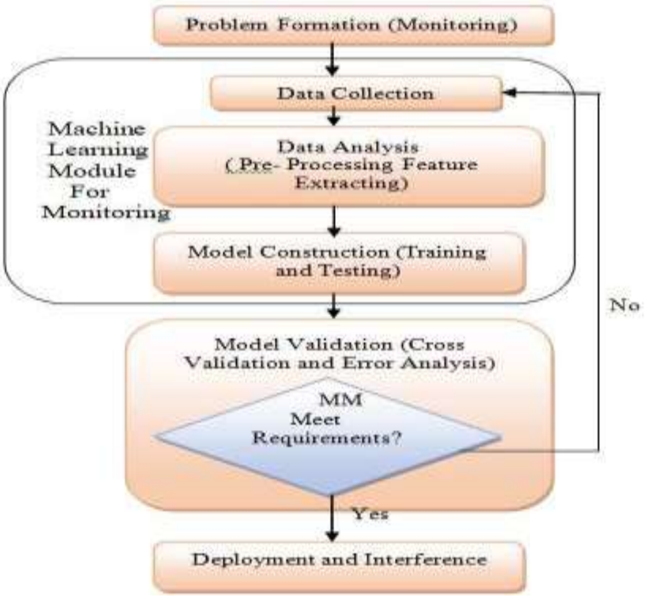
To perform this task, first UCI machine repository data will be gathered to create the training, validation, and test sets after which processing will be done. To obtain the matrix of collected data, it will be processed to the various input formats from the different algorithms, and this will result in a dataset where each instance is a row and each feature column is a feature. In order to avoid biased results, the values in the entire matrix will be standardized so that no feature can have greatly exaggerated values relative to other features. In this case, standardization consists of determining each feature's mean and standard deviation, and then subtracting the former from the latter. Additionally, feature selection is implemented in this step, which is aimed at determining the features of relevance for this problem.

After this, a machine learning algorithm is used to build a model, which is tested and validated using the training, validation, and test sets. This model can be used in the future to find out what network conditions exist.

b) Methodology

In this system architecture, several modules have been constructed to help fulfill objectives. Data Collection, Data Pre-processing, Model Creation, and, finally, Network State Identification are the modules of the system. In the Data Collection Module, the user decides whether or not the network is normal, anomalous, or unknown. The user can also choose how long they want to collect data and on which device, and they can do this by specifying the IP address of the device to be monitored. The training, validation, and testing sets in the second module, Data Pre-processing, are prepared beforehand. Since in a multi-class algorithm both anomalous and non-anomalous data are used in all sets, the user must specify if the data is to be used with a one-class or multi-class classification algorithm. In addition to providing multiple data files, the user can also give the resulting data set a name.

The Model Creation and Performance Measurement module includes a machine learning algorithm for training, saving, and testing a model that is built using the data that is acquired. The user only has to specify which data set matrix (previously created in the Data Pre-processing module) and the desired machine learning algorithm are required, except for the specific machine learning algorithm that has been selected


Problem Formation (Monitoring)
Data Collection
Machine Learning Module For Monitoring
Data Analysis (Pre- Processing Feature Extracting)
Model Construction (Training and Testing)
No
Model Validation (Cross Validation and Error Analysis)
MM Meet Requirements?
Yes
Deployment and Interference

a) ***Anaconda***

V. TOOLS FOR IMPLEMENTATION

Anaconda is commonly used for computational sciences, data science, statistical analysis, and system evaluation. Anaconda 5.0.1 is released on October 1, 2017, in the ultra-modern model. A recently released model 5.0.1 contains a few minor bugs fixes, and includes features such as up-to-date R programming assistance. As with previous versions, all of those capabilities were not included in the authentic 5.0.0 launch. This supervisor includes a Python distribution, a collection of open supply packages, and an environment supervisor. This package deal supervisor also incorporates more than 1,000 R and Python packages.

b) ***Spyder***

An open source Python project known as Spyder provides a powerful medical environment for scientists, engineers, and records analysts. it is a bit of a departure a combination of the development device's advancement optimizing, investigation, development, and monitoring expertise with the programming package's exploration, self-procedural execution, deep look, and brilliant visual results. Additionally, many famous clinical applications are integrated into Spyder, with the option to integrate in more of these via Python integration.

c) ***SIEM Tool***

Additionally, managing the hectic volume of facts gained from pastimes on systems is a significant demanding situation in cyber security. In order to make sense of it, one must derive warning signs of attacks, understand the nature of faults, or supply proof for decision makers. the 'security facts event management' concept was first put forward by Gartner in 2005 (SIEM). Conventional safety tracking machines meet audit and compliance needs, which is why they used it to describe such a machine. On the other hand, as record security has advanced, the needs of the SIEM have also increased.

VI. CONCLUSION AND FUTURE SCOPE

A great number of security monitoring systems are available for use in a system. In actuality, machine learning and network security advancements have benefited one another. A records evaluation process in which a decision feature is based on the network's protection level has a history of issues. Dynamic protection monitoring is an important component of system studying. It is essential to have the latest development trends in mind, in order to ensure the maximum level of system mastery. Truly useful systems are very fruitful when it comes to accumulating masses of facts, and as a result there is a pressing need for screening tools to help find possible threats in the community. Together with supervised classification and clustering, the device learning methods have also proven to be useful for network security. On the other hand, cybersecurity specialists know that community safety monitoring is critical, and they can deduce what initiatives humans, procedures, and mindsets are needed to meet those objectives. Network security monitoring is rapidly increasing in both quantity and difficulty.

REFERENCES

[1] X. Ye, J. Zhao, Y. Zhang, and F. Wen. Quantitative vulnerability assessment of cyber security for distribution automation systems. Energies, 8(6):5266–5286,2015.L. Bennett. Cyber security strategy. ITNOW, 54(1):10–11, 2012

[2] J. Blackburn and G. Waters., Optimising Australia’s Response to the Cyber Challenge. Kokoda Foundation,2011.

[3] Anderson HS, Roth P. EMBER ,“An Open Dataset For Training Static PE Malware

Machine Learning Models". arXiv preprint arXiv:1804.04637, April 2018.

[4] Shalaginov A, Banin S, Dehghantanha A, Franke K. ,Machine Learning Aided Static Malware Analysis: A Survey And Tutorial. Cyber Threat Intelligence, pp. 7-45.,2018.

[5] Doyen Sahoo, Chenghao Liu, and Steven C.H. Hoi "A Survey on Malicious URL Detection using Machine Learning" arXiv:1701.07179v2 [cs.LG] ,16 Mar 2017.

[6] Crockett, A. Latham, N. Whitton,"On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees," International Journal of Human-Computer Studies, vol. 97, pp. 98-115, 2017.

[7] B. Sun, S. Chen, J. Wang, H. Chen, "A robust multi-class AdaBoost algorithm for mislabelled noisy data," Knowledge-Based Systems, vol. 102, pp. 87-102,2016.

[8] M. Baig, M .M. Awais, E. M. El-Alfy, "AdaBoost-based artificial neural network learning," Neurocomputing, vol. 16, pp. 22 – 41, 2017.

[9] X. Pan, Y. Luo, Y. Xu, "K-nearest neighbour based structural twin support vector machine," Knowledge-Based Systems, vol. 88, pp. 34- 44, 2015.

[10] A. C. Bahnsen, D. Aouada, B. Ottersten, "Example-dependent cost- sensitive decision trees," Expert Systems with Applications, vol. 42, pp. 6609-6619,2015.

[11] Adler A, Mayhew M, Cleveland J, Atighetchi M & Greenstadt R. " Using machine learning for behaviour-based access control: Scalable anomaly detection on tcp connections and http requests." Proc. Military Communications Conference, MILCOM IEEE,1880–1887,2013.

[12] Fiore U, Palmieri F, Castiglione A & Santis AD, "Network anomaly detection with therestricted boltzmann machine. Neurocomputing" 122(0): pp.13 – 23. Advances in cognitive and ubiquitous computing, 2013.

[13] Bejtlich R ,"The Practice of Network Security Monitoring". No Starch Press,2013.

[14] Nicholson A, Webber S, Dyer S, Patel T & Janicke H ,"SCADA security in the light of cyber-warfare. Computers & Security" 31(4): pp. 418 – 436,2012.

[15] Shon T, Kim Y, Lee C & Moon J, "A machine learning framework for network anomaly detection using SVM and GA. Proc". Information Assurance Workshop, IAW '05 Proceedings from the Sixth Annual IEEE SMC, pp. 176–183,2005.

[16] Tsai CF, Hsu YF, Lin CY & Lin WY "Intrusion detection by machine learning: A review Expert Systems with Applications 36(10): pp. 11994 – 12000,2009.

[17] X. Ye, J. Zhao, Y. Zhang, and F. Wen."Quantitative vulnerability assessment of cyber security for distribution automation systems." Energies, 8(6):5266–5286,2015.

[18] Uma, M., Padmavathi, G. "A survey on various cyber-attacks and their classification" International Journal of Network Security, 15, 5, 390- 396,2013.

[19] M. Alam and S. T. Vuong."An intelligent multi-agent based detection framework for classification of android malware in Active Media Technology", pp. 226–237. Springer,2014.

[20] R. Kumar, G. Poonkuzhali, and P. Sudhakar., Comparative study on email spam classifier using data mining techniques. In The International Multi Conference of Engineers and Computer Scientists, volume 1, pages 14–16, Hong Kong, China,2012.