ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

A Scalable Cyber Attack Detection Using Machine Learning Techniques

S. Rajarajan., S. Manavalan., N. Kumar.,

Assistant Professor, Department of Computer Science and Engineering, St. Anne's College of Engineering and Technology, Panruti, Cuddalore, Tamil Nadu, India

Assistant Professor, Department of Computer Science and Engineering, St. Anne's College of Engineering and Technology, Panruti, Cuddalore, Tamil Nadu, India

Assistant Professor, Department of Computer Science and Engineering, St. Anne's College of Engineering and Technology, Panruti, Cuddalore, Tamil Nadu, India

ABSTRACT: Within the ever-growing and quickly increasing field of cyber security, it is nearly impossible to quantify or justify the explanations why cyber security has such an outsized impact. Permitting malicious threats to run any place, at any time or in any context is a long way from being acceptable, and may cause forceful injury. It particularly applies to the Byzantine web of consumers and using the net and company information that cyber security groups are finding it hard to shield and contain. Cyber security may be a necessary thought for people and families alike, also for businesses, governments, and academic establishments that operate inside the compass of world network or net. With the facility of Machine Learning, we will advance cyber security landscape. Businesses these days are gathering immense amounts of user information. Information is at the heartbeat of any business-critical system you'll be able to think about. This co-jointly includes infrastructure systems that are being implemented these days. Today's high-tech infrastructure, that has network and cyber security systems, is gathering tremendous amounts of data and analytics on almost all the key aspects of mission-critical systems. Whereas people still give the key operational oversight and intelligent insights into today's infrastructure, machine learning and AI are gaining pace and gathering immense momentum in most of the areas of today's systems, whether or not it's positioned on premise or within the cyber security house.

I. INTRODUCTION

Cyber-attacks are increasing within the cyber world. There ought to be some advanced security measures taken to scale back or avoid the amount of cyber-attacks. There are various attacks like D-Dos attacks, Man within the middle, information escape, PROBE, User-To-Root, Remote-ToLocal. These attacks are utilized by the hackers or intruders to realize the unauthorized access to any non-public network, websites, information or perhaps in our personal computers. Therefore, outside or internal hackers use using advanced techniques or finding ways to tickle or break any defense systems to shield the sensitive information, information, money data. Sensible intrusion munitions ought to stop or try and manage varied innovative attacks created or programmed by the hackers Cyber security refers to the science of technologies, processes, and practices designed to shield networks, devices, programs, and information from attacks, damage, or unauthorized access. Cyber security can also be stated because it's security, within the year 2016, witnessed several advancements in machine learning techniques like self-driven cars, linguistic communication process, health sector, and sensible virtual assistant. They need to be used for locating helpful data from varied audit datasets, which are applied to the matter of intrusion detection.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

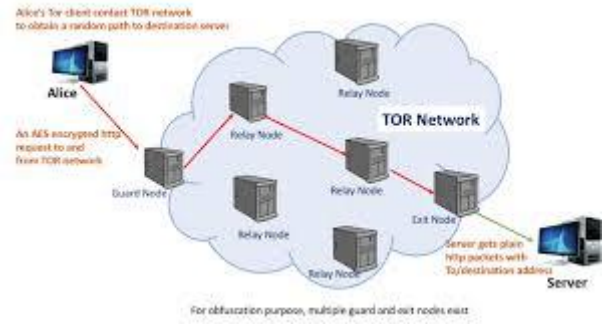

Fig: block diagram

With the assistance of Machine learning technology, we will deploy these ideas in cyber security to boost the protection measures within the intrusion detection system. Initially, we've got to feed the information into the machinelearning model. The model gets trained by the dataset sample and makes it a trained model. Once we feed the dataset sample, future step is to use and apply the machine learning formula.

Machine learning formula plays an important role in rising the protection measures in this intrusion detection system. ML algorithms are classified into 2 types: supervised learning and unsupervised learning. They're differentiated by the information (i.e., input) that they settle for. Supervised learning refers to algorithms that are given a group of labelled training information, with the task of understanding what differentiates the labels. Unsupervised learning refers to algorithms that are given unlabeled training information, with the task of inferring the classes all by itself. Typically, the labelled information is incredibly rare, or even the task of labelled data is itself terribly exhausting and we may not be able or ready to sight if labels actually exist.

II. RELATED WORK

In the budding stages of building intrusion detection systems, cyber analytics support was studied as a mixture of ML/DM (Machine learning/Data mining). Anomaly-based techniques model the system behavior and traditional network, and helps to spot anomalies as deviations from traditional behavior. Its advantage is that the profiles of traditional activity are custom-made for each system, application or network, thereby creating it tough for attackers to grasp those activities that they can perform undiscovered and in a very clandestine manner. They give the impression of being appealing, owing to their distinctive ability to sight zero-day attacks. Hybrid techniques mix anomaly detection and misuse. They're used to boost detection rates of acknowledged intrusions and reduce the rates of unknown attacks. Again, intelligent intrusion detection systems will solely be engineered if there's accessibility of a good information set. An information set with a large quantity of quality data and the one that mimics real time which will solely facilitate to train the associated check of an intrusion detection system.

dealing with constant and complicated cyberthreats and cyberattacks. As a general warning, organizations should build and develop a cybersecurity culture and awareness so as to defend against cyber criminals. Data Technology like IT and data Security like InfoSec audits that were economical within the past, try to converge into cybersecurity audits to handle cyber threats, cyber risks and cyberattacks that evolve in an aggressive cyber landscape[1]. However, the rise in variety and quality of cyberattacks and therefore the convoluted cyber threat landscape is challenging the running cybersecurity audit models and fitting proof for a brand new cybersecurity audit model. This text reviews the simplest practices and methodologies of world leaders within the cybersecurity assurance and audit arena. By means of the analysis of this approaches and theoretical background, their real scope, strengths and weaknesses are highlighted looking forward at a good and cohesive synthesis. As a result, this text presents an inspired and comprehensive cybersecurity audit model as a proposal to be utilized for conducting cybersecurity audits in organizations and Nation

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

States. The CyberSecurity Audit Model (CSAM) evaluates and validates audit, preventive, rhetorical and detective controls for all structure useful areas [2]. CSAM has been tested, enforced and validated at the side of the Cybersecurity. A research case study is being conducted to validate each model and therefore the findings are revealed consequently.

A completely unique technique to try to feature choices to sight botnets at their section of Command and control (C&C) is conferred. A significant downside is that researchers have proposed options supported in their experience, however there's no technique to judge these options since a number of these options might get a lower detection rate than alternative. To the current aim, we discover the feature set supported connections of botnets at their section of C&C, that maximizes the detection rate of those botnets. A Genetic formula (GA) was accustomed and chosen as the set of options that offers the best detection rate. We tend to use the machine learning formula C4.5, this formula did the classification between connections belonging or not to a botnet. The datasets employed in this paper were extracted from the repositories ISOT and ISCX [3]. Some tests were done to induce the simplest parameters in a GA and the formula C4.5. We tend to co-jointly perform experiments so as to get the simplest set of options for every botnet analyzed (specific), and for every kind of botnet (general) too. The results are shown at the tip of the paper, within which a substantial reduction of features and the higher detection rate than the related work conferred were obtained.
Over the previous couple of years' machine learning has migrated from the laboratory to the forefront of operational systems. Amazon, Google and Facebook use machine learning on a daily basis to boost client experiences, instructed purchases or connect individuals socially with new applications and facilitate personal connections. Machine learning's powerful capability is additionally there for cybersecurity. Cybersecurity is positioned to leverage machine learning to boost malware detection, sorting events, acknowledge breaches and alert organizations to security problems. Machine learning may be accustomed to determine advanced targeting and threats like organization identification, infrastructure vulnerabilities and potential mutually beneficial vulnerabilities and exploits. Machine learning will considerably modify the cybersecurity landscape [5]. Malware by itself will represent as many as three million new samples in a hour. Ancient malware detection and malware analysis is unable to pace with new attacks and variants. New attacks and complicated malware are ready to bypass network and endpoint detection to deliver cyber-attacks at alarming rates. New techniques like machine learning should be leveraged to handle the growing malware downside. This proposition describes the machine learning, that can be accustomed to detect and highlight advanced malware for cyber defense analysts. The results of our initial analysis and a discussion of future analysis to increase machine learning is presented.
With the large growth of laptop networks and content usage of users, there's a necessity for secure and reliable networks. Because it is determined that the various form of network attacks is raised over a amount of your time, it's necessary to create the supply of effective automatic tools so as to spot the attack detection situations. Intrusion Detection System is one among the attack systems that detect intrusions returning from the net. Many approaches were determined within the literature for intrusion detection over the network. Within the recent past, mining techniques were prevailing so as to visualize the intrusion detection [6]. The characteristics of incoming intrusions were known by using the well mined data over the information given within the network. Whenever an identical object is found within the characteristics of the well-mined information then it's declared as an intrusion. Supporting this criterion varied intrusion detection models were developed within the recent analysis and therefore the accuracy is improved. A quick review is carried out over the sooner approaches. The entire approach is divided into information preprocessing approaches and detection approaches. Further, the information preprocessing approaches are divided into Feature extraction and have transformation models that support operating methodology over the options. Similarly, the detection approaches are classified as machine learning and organic process approaches.

data Security or compliance audit, there'll be consistent phases like designing, shaping objectives and scope, elucidating terms of engagements, conducting the audit, corroboratory proof, evaluating risks, news the audit findings and schedule follow up tasks. Designing a cybersecurity audit isn't totally different than any kind of audit. This however will take a great deal of effort thanks to the quality of the many cybersecurity domains. However, most cyber capabilities aren't reviewed by the inner audits' scope. This specific framework includes risk/compliance management, development life cycle, security program, third-party management, information/asset management, access management, threat/vulnerability management, of implementing cybersecurity controls as a part of an overall framework and

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

strategy, the necessity for assurance which will be achieved by management reviews, cyber risk assessments, information management and protection, risk analytics, crisis management and resiliency, security operation and security awareness and training. Moreover, Deloitte's framework is aligned with trade frameworks just like the National Institute of Standards and Technology (NIST), data Technology Infrastructure Library (ITIL), Committee of Sponsoring Organizations of the Treadway Commission (COSO) and world organization for Standardization (ISO).
In addition, there aren't any metrics to live cybersecurity audits and therefore the cybersecurity audit topic is poorly understood because it transforms extremely quickly. Khan considers that to hide a significant scope for designing a cybersecurity audit, the auditors should embrace all relevant areas of any organization; these areas are client operations, finance, human resources, IT systems and applications, legal, purchasing, regulatory affairs, physical security and every one of the applicable third parties that have relationships with the business.

Several Intrusion detection systems are developed for host systems and networks. However, comparatively few notable works are there for information intrusion detection. One among the latest works revealed was a technique by Chung et al. proposes a misuse detection approach for information intrusion detection. Here frequent information patterns are well-mined and hold on as traditional profiles. The main disadvantage is that it doesn't produce role profiles. The users perform totally different actions supported by their roles. User profiles can't be used as the only criteria. Users can perform actions supported roles and that they will be detected malicious. Lee et al. a proposed true time intrusion detection system supported time signatures. Real time information systems use temporal information objects and their values change with time [7]. Thus whenever time it is updated a device dealing is generated. The temporal information is updated over a certain amount of your time. If a dealing tries to vary the temporal information that has been updated already over that amount, an alarm is raised. But the disadvantage of this technique is that it focuses on updates solely and not role profiles. Hu Panda. uses log files to come up with user profiles. Frequently accessed information and tables and hold on for comparison. The problem with this approach is that, maintenance of knowledge is incredibly tough once the dimensions of information is simply too giant and variety of users conjointly increase dynamically

Today's cyber security threats are too varied and arrive too quick for strictly manual defense. Machine learning in addition provides power and increased speed to tackle enormous volumes of attacks with myriad variations. Nevertheless, the $64000 key to investing AI for cyber protection is to use it with human intelligence, combination of power, speed, skills and judgement. Artificial Intelligence and Machine Learning are often very nice and helpful in detective work in cyber security attacks. The work that human has to do are often through with the assistance of machine learning at a lot of quicker pace and with high accuracy. Implementing numerous Machine-Learning Techniques can facilitate US to discover the cyber security attacks.

III. METHODS

Datasets Since the applications for various network security tasks use machine learning methods, large datasets are needed to analyze network flows and distinguish between normal and abnormal traffic. Over the years, several experiments have been conducted to generate network datasets. As shown in Table I, most of the studies using machine learning have tested their work against simulated or real network data. Although a good number of those datasets remain private, primarily due to security concerns, some have become publicly available such as DARPA 98, KDD99, UNSW-NB15, ISCX, CICIDS2017, and N-BaIoT. Although several datasets have been produced, however, the development of realistic IoT and network traffic datasets that include new Botnet scenarios are still few. More importantly, some datasets lack the inclusion of IoT-generated traffic, while others neglect to generate any new features. In some cases, the testbed used was not realistic, while in other cases, the attack scenarios were not diverse enough. For instance, in [12], Meidan et al. created a publicly available IoT dataset named N-BaIoT, and many later studies used this dataset for training and to test their classifier models. While this dataset is relatively large and clean, it is unbalanced, and the ratio of normal data is much lower compared to attack data. Moustafa et al. [9] sought to address the shortcomings by designing the Bot-IoT dataset, which we used for our experiments. The Bot-IoT dataset incorporates legitimate and simulated IoT network traffic along with various types of attacks[14]. The BotIoT database attacks are classified into three types: Probing attacks, DoS, and theft information theft

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

We used the Bot-IoT dataset to evaluate seven well known machine learning classifiers: (K-Nearest Neighbours (KNN), ID3 (Iterative Dichotomiser 3), Random Forest, AdaBoost, Quadratic discriminant analysis (QDA), Multilayer perceptron (MLP), and Naıve Bayes (NB). When choosing these classifiers, the focus is on bringing together popular algorithms with different characteristics. In this context, the algorithms used are briefly examined in the following.

K-Nearest Neighbours(KNN):
KNN is one of the simplest and most effective supervised learning algorithm. It is used for searching through the available dataset to associate new data points with similar existing points [24]. KNN, which provides good performance over multidimensional data and is a fast algorithm during the training phase, is relatively slow in the estimation stage.

Random Forest(RF):
RF is a machine learning approach that uses decision trees. In this method, a "forest" is created by assembling a large number of different decision tree structures that are formed in different ways[28]. This algorithm has many advantages, such as the ability run on huge datasets efficiently, its light weight compared to other methods, and robustness against noise and outliers when compared to single classifiers.

Quadratic discriminant analysis (QDA):
QDA is an ideal algorithm to supervised classification problems. Discriminant analysis is a statistical technique for assigning measured data to one group among many groups. QDA is appropriate to situation where a category is not characterized by much data. In order to be able to apply Quadratic Discriminant Analysis, the number of samples observed must be greater than the number of groups

Feature Extraction:
CICFlowMeter [25] was used to extract flow-based features (in pcap format) from raw network traffic data. CICFlowMeter is a network traffic flow generator distributed by CIC that produces 84 network traffic characteristics. It reads the pcap file and produces a visual document of the features extracted, and also offers a csv file of the dataset. This process was primarily designed to improve classifiers' predictive capabilities by extracting new dataset features.

IV. RESULT

The final results of the implementation (see Table VI) are compared with a study in the literature. For this comparison, the study conducted by Ferrag et al. [14] in 2019 was chosen. The reason for this is that the mentioned work used the same dataset as well as two machine learning methods similar to the ones we used. These similar machine learning algorithms are Random Forest and Naive Bayse. The key difference between our work and theirs is the feature set used. They used the original feature set while we used a new feature set extracted by CICFLOWMETER.

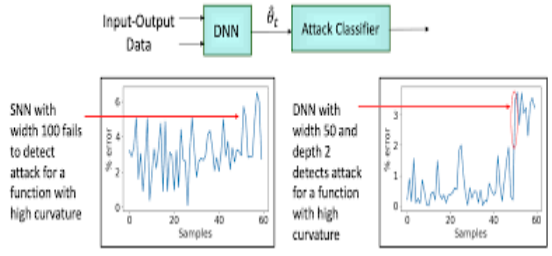

Fig2: Result Analysis

The detection rate (Recall) was determined as the main evaluation criterion. Table VI shows the comparison of the results obtained from the two studies. When the results are examined, it can be seen that the Random Forest algorithm

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

used in our study is higher than that used in [14], and the same thing can be seen for most attack types with the NB algorithm. So, we can see that the new features used in our work increased the performance of both algorithms.

V. CONCLUSION

This paper represents a review of Machine Learning and DL unit methods for network security domain. The literature paper, that has largely targeted on the last four years, introduces the most recent applications of ML and DL unit within the field of intrusion detection systems. Sadly, the foremost effective methodology of intrusion detection has not nevertheless been established, and therefore the analysis remains occurring. Every approach for implementing an intrusion detection system has its own pros and cons, a degree apparent from the discussion conducted for comparisons among the varied ways. Thus, it's tough to settle on a specific methodology to implement an intrusion detection system over the others. Datasets for network intrusion detection are important resources for coaching and testing systems. The Machine Learning and DL methods don't work while not the representative information, and getting such a dataset is tough and long. However, there are several issues with the accessible existing public dataset, like uneven information, or out-of-date content and therefore the likes are similar. These issues have mostly restricted the event of analysis during this explicit space. Network info updates in no time, that brings to the ML and DL model coaching with larger problem. The Model has to be retrained semi-permanent/long-termed and quickly. Thus progressive learning and long learning are the long run focus within the study of this field within the future.

REFERENCES

[1] J. Deogirikar and A. Vidhate, "Security attacks in iot: A survey," International Conference on I-SMAC (I-SMAC), pp. 32–37, 2017.
[2] T. Bodstrom and T. H ¨ am¨ al¨ ainen, "State of the art literature review ¨ on network anomaly detection with deep learning," Internet of Things, Smart Spaces, and Next Generation Networks and Systems, pp. 64–76, 2018.
[3] I. Arnaldo, A. Cuesta-Infante, A. Arun, M. Lam, C. Bassias, and K. Veeramachaneni, "Learning representations for log data in cybersecurity," International Conference on Cyber Security Cryptography and Machine Learning, pp. 250–268, 2017.
[4] M. Du, F. Li, G. Zheng, and V. Srikumar, "Deeplog: Anomaly detection and diagnosis from system logs through deep learning," Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1285–1298, 2017.
[5] B. J. Radford, B. D. Richardson, and S. E. Davis, "Sequence aggregation rules for anomaly detection in computer network traffic," arXiv preprint arXiv:1805.03735, 2018.
[6] I. Lambert and M. Glenn, "Security analytics: Using deep learning to detect cyber attacks," 2017.
[7] M. Stevanovic and J. M. Pedersen, "Detecting bots using multi-level traffic analysis." IJCSA, vol. 1, no. 1, pp. 182–209, 2016.
[8] H. Sedjelmaci, S. M. Senouci, and M. Al-Bahri, "A lightweight anomaly detection technique for low-resource iot devices: A game-theoretic methodology," IEEE International Conference on Communications (ICC), pp. 1–6, 2016.
[9] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset," Future Generation Computer Systems, vol. 100, pp. 779–796, 2019.
[10] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: an ensemble of autoencoders for online network intrusion detection," arXiv preprint arXiv:1802.09089, 2018.
[11] X. Yuan, C. Li, and X. Li, "Deepdefense: identifying ddos attack via deep learning," IEEE International Conference on Smart Computing (SMARTCOMP), pp. 1–8, 2017.
[12] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, and Y. Elovici, "N-baiot—network-based detection of iot botnet attacks using deep autoencoders," IEEE Pervasive Computing, vol. 17, no. 3, pp. 12–22, 2018.