Journal of Engineering Sciences
Vol 12, Issue 04, April/2021
ISSN NO:0377-9254

Artificial Intelligence framework for Cyber Attack Detection and Notifying using Machine Learning Techniques

B. Ramesh[1], J. Vineeth Kumar[2], Vuppala Vyshnavi[2], Mutha Prem[2]
[1]Assistant Professor, [2]UG Scholar, Department of CSE
[1,2]Kommuri Pratap Reddy Institute of Technology, Hyderabad, Telangana

Abstract- Cyber-crime is proliferating everywhere exploiting every kind of vulnerability to the computing environment. Ethical Hackers pay more attention towards assessing vulnerabilities and recommending mitigation methodologies. The development of effective techniques has been an urgent demand in the field of the cyber security community. Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Machine learning for cyber security has become an issue of great importance recently due to the effectiveness of machine learning in cyber security issues. Machine learning techniques have been applied for major challenges in cyber security issues like intrusion detection, malware classification and detection, spam detection and phishing detection. Although machine learning cannot automate a complete cyber security system, it helps to identify cyber security threats more efficiently than other software-oriented methodologies, and thus reduces the burden on security analysts. Hence, efficient adaptive methods like various techniques of machine learning can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs. Our main goal is that the task of finding attacks is fundamentally different from these other applications, making it significantly harder for the intrusion detection community to employ machine learning effectively.

Keywords: Cyber-crime, Machine learning, Cyber-security, Intrusion detection system.

1. INTRODUCTION

Today, political and commercial entities are increasingly engaging in sophisticated cyber-warfare to damage, disrupt, or censor information content in computer networks. In designing network protocols, there is a need to ensure reliability against intrusions of powerful attackers that can even control a fraction of parties in the network. The controlled parties can launch both passive (e.g., eavesdropping, nonparticipation) and active attacks (e.g., jamming, message dropping, corruption, and forging). Intrusion detection is the process of dynamically monitoring events occurring in a computer system or network, analysing them for signs of possible incidents and often interdicting the unauthorized access. This is typically accomplished by automatically collecting information from a variety of systems and network sources, and then analysing the information for possible security problems. Traditional intrusion detection and prevention techniques, like firewalls, access control mechanisms, and encryptions, have several limitations in fully protecting networks and systems from increasingly sophisticated attacks like denial of service. Moreover, most systems built based on such techniques suffer from high false positive and false negative detection rates and the lack of continuously adapting to changing malicious behaviours. In the past decade, however, several Machine Learning (ML) techniques have been applied to the problem of intrusion detection with the hope of improving detection rates and adaptability. These techniques are often used to keep the attack knowledge bases up-to-date and comprehensive. In recent days, cyber-security and protection against numerous cyber-attacks are becoming a burning question. The main reason behind that is the tremendous growth of computer networks and the vast number of relevant applications used by individuals or groups for either personal or commercial use, especially after the acceptance of the Internet of Things (IoT). The cyber-attacks cause severe damage and severe financial losses in large-scale networks. The

existing solutions like hardware and software firewalls, user's authentication, and data encryption methods are not sufficient to meet the challenge of upcoming demand, and unfortunately, not able to protect the computer network's several cyber-threats. These conventional security structures are not sufficient as safeguard due to the faster rigorous evolution of intrusion systems. Firewall only controls every access from network to network, which means prevent access between networks. But it does not provide any signal in case of an internal attack. So, it is obvious to develop accurate defense techniques such as machine learning-based intrusion detection system (IDS) for the system's security In general, an intrusion detection system (IDS) is a system or software that detects infectious activities and violations of policy in a network or system. An IDS identifies the inconsistencies and abnormal behavior on a network during the functioning of daily activities in a network or system used to detect risks or attacks related to network security, like denial-of-service (Dos). An intrusion detection system also helps to locate, decide, and control unauthorized system behaviour such as unauthorized access, or modification and destruction. There are different types of intrusion detection systems based on the user perspective. For instance, they are host-based and network-based IDS.

2. LITERATURE SURVEY

An IDS generally has to deal with problems such as large network traffic volumes, highly uneven data distribution, the difficulty to realize decision boundaries between normal and abnormal behaviour, and a requirement for continuous adaptation to a constantly changing environment . In general, the challenge is to efficiently capture and classify various behaviours in a computer network. Strategies for classification of network behaviours are typically divided into two categories: misuse detection and anomaly detection. Misuse detection techniques examine both network and system activity for known instances of misuse using signature matching algorithms. This technique is effective at detecting attacks that are already known. However, novel attacks are often missed giving rise to false negatives. Alerts may be generated by the IDS, but reaction to every alert wastes time and resources leading to instability of the system. To overcome this problem, IDS should not start elimination procedure as soon as the first symptom has been detected but rather it should be patient enough to collect alerts and decide based on the correlation of them. 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 than $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 cyber security protection. Attackers currently make over $1 billion in annual revenue from Ransomware attacks, such as Wannacry and Crypto Wall attacks.

2.1. Improving Cyber Security Assurance Model

Every time a gaggle of auditors are taking part in an IT, 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 cyber-security 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 cyber security 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 cyber security

controls as a part of an overall framework and 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 Tread way Commission (COSO) and world organization for Standardization (ISO).

3. EXISTING WORKS

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 the world network or net. With the facility of Machine Learning, we will advance the cyber security landscape. 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. Most intrusion detection systems are focused on the perimeter attack surface threats, starting with your firewall. That offers protection of your network's north south traffic, but what it doesn't take into account is the lateral spread (east-west) that many network threats today take advantage of as they infiltrate your organization's network and remain there unseen. We know this is true because research has shown that only 20% of discovered threats come from north south monitoring. When AN IDS detects suspicious activity, the violation is typically reported to a security information and event management (SIEM) system where real threats are ultimately determined amid benign traffic abnormalities or other false alarms. However, the longer it takes to distinguish a threat, the more damage can be done. An IDS is immensely helpful for monitoring the network, but their usefulness all depends on what you do with the information that they give you. Because detection tools don't block or resolve potential issues, they are ineffective at adding a layer of security unless you have the right personnel and policy to administer them and act on any threats. An IDS cannot see into encrypted packets, so intruders can use them to slip into the network. An IDS will not register these intrusions until they are deeper into the network, which leaves your systems vulnerable until the intrusion is discovered. This is a huge concern as encryption is becoming more prevalent to keep our data secure. One significant issue with an IDS is that they regularly alert you to false positives. In many cases false positives are more frequent than actual threats. An IDS can be tuned to reduce the number of false positives; however, your engineers will still have to spend time responding to them. If they don't take care to monitor the false positives, real attacks can slip through or be ignored.

4. PROPOSED SYSTEM

Machine Learning algorithms can be used to train and detect if there has been a cyber attack. As soon as the attack is detected, an email notification can be sent to the security engineers or users. Any classification algorithm can be used to categorize if it is a DoS/DDoS attack or not. One example of a classification algorithm is Support Vector Machine (SVM) which is a supervised learning method that analyses data and

recognizes patterns. Since we cannot control when, where or how an attack may come our way, and absolute prevention against these cannot be guaranteed yet, our best shot for now is early detection which will help mitigate the risk of irreparable damage such incidents can cause. Organizations can use existing solutions or build their own to detect cyber attacks at a very early stage to minimize the impact. Any system that requires minimal human intervention would be ideal.

4.1. Problem Modeling

Network admins: The following steps are the functions of Network admin: Intercept network traffic. Read and store the data packets information. Check for alerts regarding the cyber-attacks and network stats.

5. SYSTEM DESIGN

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 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 previously identified and recorded Indicators of Compromise (IOC), which are then used to build models and systems that can monitor, identify and respond 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. 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.

5.1 SYSTEM ANALYSIS

a) System: A system is an orderly group of interdependent components linked together according to a plan to achieve a specific objective. Its main characteristics are organization, interaction, interdependence, integration and a central objective.

b) System Analysis: System analysis and design are the application of the system approach to problem solving generally using computers. To reconstruct a system the analyst must consider its elements output and inputs, processors, controls, feedback and environment.

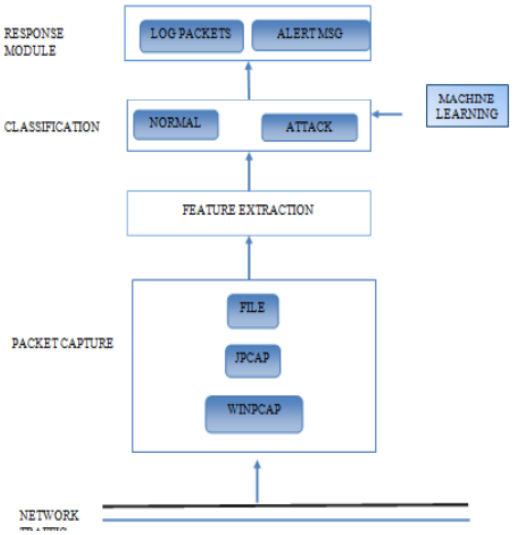

Figure.1. system Architecture

5.2. SYSTEM ARCHITECTURE

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 data sets of security events and identifying the pattern of malicious activities. can monitor, identify and respond 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. 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.

5.2.1 Network Traffic:

Network traffic refers to the amount of data moving across a network at a given point of time. Network data is mostly encapsulated in network packets, which provide the load in the network. Network traffic is the main component for network traffic measurement, network traffic control and simulation. The proper organization of network traffic helps in ensuring the quality of service in a given network. Proper analysis of network traffic provides the organization with the following benefits: Identifying network bottlenecks - There could be users or applications that consume high amounts of bandwidth, thus constituting a major part of the network traffic. Different solutions can be implemented to tackle these. Network security - Unusual amount of traffic in a network is a possible sign of an attack. Network traffic reports provide valuable insights into preventing such attacks. Network engineering - Knowing the usage levels of the network allows future requirements to be analysed.

5.2.2Packet capture

Packet Capture is a networking term for intercepting a data packet that is crossing a specific point in a data network. Once a packet is captured in real-time, it is stored for a period of time so that it can be analysed, and then either be downloaded, archived or discarded. Packets are captured and examined to help diagnose and solve network problems such as: Identifying security threats Troubleshooting undesirable network behaviours Identifying network congestion Identifying data/packet loss Forensic network analysis.

5.2.3. Classification

Classification is another extensively used supervisory machine learning task. In cyber security, spam detection is successfully implemented by ML based classifiers which involves discriminating a given email message as spam or not. The spam filter models are able to separate spam messages from non-spam messages. Machine learning techniques for classification include Logistic Regression, KNearest Neighbours, Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest Classification. Upon the availability of large collection of past data with labels, Deep Learning classification models involving Restricted Boltzmann Machines(RBM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or Long-Short Term Memory (LSTMs) cells for feature extraction followed by a densely connected neural network have become more efficient in solving complex tasks. Applicability of the above supervisory machine learning techniques is conditioned based on the availability of large collections of labeled data.

5.2.4. Response Module

Incident response is a term used to describe the process by which an organization handles a data breach or cyber attack, including the way the organization attempts to manage the consequences of the attack or breach (the "incident"). Ultimately, the goal is to effectively manage the incident so that the damage is limited and both recovery time and costs, as well as collateral damage such as brand reputation, are kept at a minimum and also sends an alert message to security analysts.

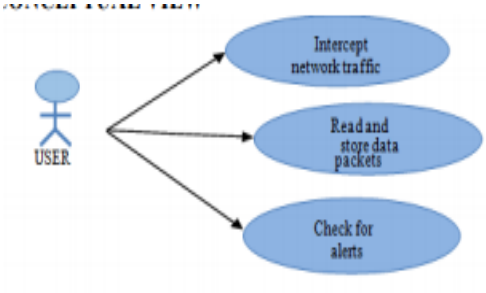

Figure.2. Use Case Diagram

Fig.2. Represents the role of USER is an interaction between a source and a destination. Accordingly the USER look after all the stages (Actors) of the system like intercepting to network and getting the packets from the network ,reading and storing the network and

running the stored data in the machine learning model and then checking for attack alerts.

6. SIMULATION RESULTS

Figure.3. Email

Figure.4. Login Page

Figure.5. Statistics

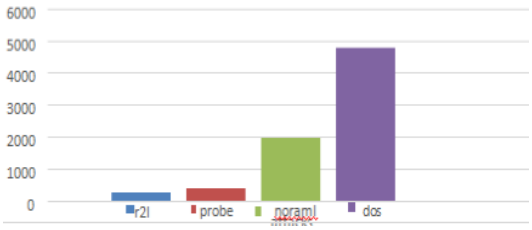

Figure.6. Graphical analysis of attacks types versus number of packets got attacked during the process

7. SUMMARY

Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Hence, efficient adaptive methods like various techniques of machine learning can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs.

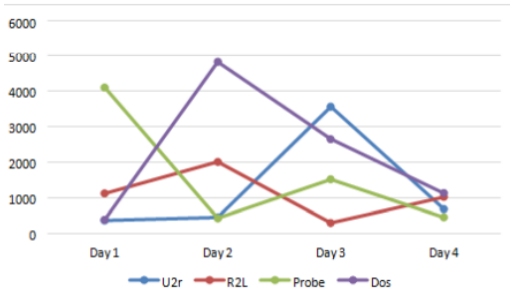

Figure.7. Graphical analysis of attacks occurred on each day of a week/ weekly statistical analysis of attack

We reviewed several influential algorithms for intrusion detection based on various machine learning techniques. Characteristics of ML techniques makes it possible to design IDS that have high detection rates and low false positive rates while the system quickly adapts itself to changing malicious behaviors. IDS using many Machine Learning Techniques like Random Forest, Decision tree and logistic regression to perform better in various metrics. The IDS should provide the most effective solutions based on the requirements. One thing is sure, any company failing to adopt these techniques now or in the immediate future risk compromising data or worse servers.

REFERENCES

[1]. DipankarDasgupta. Immunity-based intrusion detection system: A general frame-work. In Proceedings of the 22nd National Information Systems Security Confer-ence (NISSC). Arlington, Virginia, USA, 1999.

[2]. Jonatan Gomez and DipankarDasgupta. Evolving fuzzy classi_ers for intrusion detection. In Proceedings of the 2002 IEEE Workshop on Information Assurance,West Point, NY, USA, 2002.

[3]. Steven A. Hofmeyr, Stephanie Forrest, and Anil Somayaji. Intrusion detection using sequences of system calls. Journal of Computer Security, 6(3):151{180, August 1998.

[4]. Peter Mell Karen Scarfone. Guide to intrusion detection and prevention systems (idps). National Institute of Standards and Technology, NIST SP - 800-94, 2007.
[5]. Jungwon Kim, Peter J. Bentley, Uwe Aickelin, Julie Greensmith, Gianni Tedesco, and Jamie Twycross. Immune system approaches to intrusion detection { a review. Natural Computing, 6(4):413{466, December 2007.
[6]. Gupta BB, Tewari A, Jain AK, Agrawal DP. Fighting against phishing attacks: state of the art and future challenges. Neural Comput Appl. 2017;28(12):3629–54.
[7]. Av-test institute, germany, https://www.av-test.org/en/statistics/malware/. Accessed 20 Oct 2019.
[8]. Ibm security report, https://www.ibm.com/security/data-breach. Accessed on 20 Oct 2019.
[9]. Fischer EA. Cybersecurity issues and challenges: In brief. Congressional Research Service (2014)
[10]. Juniper research. https://www.juniperresearch.com/. Accessed on 20 Oct 2019.
[11]. Papastergiou S, Mouratidis H, Kalogeraki E-M. Cyber security incident handling, warning and response system for the european critical information infrastructures (cybersane). In: International Conference on Engineering Applications of Neural Networks, p. 476–487 (2019). New York: Springer