2023 International Conference on Computer Communication and Informatics (*ICCCI*), Jan. 23 – 25, 2023, Coimbatore, INDIA

Intrusion Detection System Using Machine Learning

Ajmeera Kiran
Department of Computer Science and Engineering
MLR Institute of Technology
Hyderabad, India
kiranphd.jntuh@gmail.com

S. Wilson Prakash
Assistant Professor,
Department of Computer Science and Engineering
Kalasalingam Academy of Research and Education,
wprakash.s@gmail.com

B Anand Kumar
Department of Computer Science and Engineering
MLR Institute of Technology
Hyderabad, India
anandbiyani.b@mlrinstitutions.ac.in

Likhitha
UG Student
Department of Computer Science and Engineering
MLR Institute of Technology
Hyderabad, India
likhita.v@gmail.com

Tammana Sameeratmaja
UG Student
Department of Computer Science and Engineering
MLR Institute of Technology
Hyderabad, India
sameera.tammana@gmail.com

Ungarala Satya Surya Ram Charan
UG Student
Department of Computer Science and Engineering
MLR Institute of Technology
Hyderabad, India
ramc06766@gmail.com

2023 International Conference on Computer Communication and Informatics (ICCCI) | 979-8-3503-4821-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCI56745.2023.10128363

Abstract— **The use of computers and the internet has spread rapidly over the course of the past few decades. Every day, more and more peopleare coming to rely heavily on the internet. When it comes to the field of information security, the subject of security is one that is becoming an increasingly important focus. It is vital to design a powerful intrusion detection system in order to prevent computer hackers and other intruders from effectively getting into computer networks or systems. This can be accomplished by: (IDS). The danger and attack detection capabilities of the computer system are built into the intrusion detection system. Abuse has occurred and can be used to identify invasions when there is a deviation between a preset pattern of intrusion and an observedpattern of intrusion. An intrusion detection system (IDS) is a piece of hardware (or software) that is used to generate reports for a Management Station as well as monitor network and/or system activities for unethical behaviour or policy violations. In the current study, an approach known as machine learning is suggested as a possible paradigm for the development of a network intrusion detection system. The results of the experiment show that the strategy that was suggested improves the capability of intrusion detection.**

Keywords- Support vector machine, Machine Learning, Network Intrusion Detection System, Host Intrusion Detection System, Intrusion Prevention System, Intrusion Detection System, Host, Network, Intrusion Detection System.

I. INTRODUCTION

Over the past few years, there has been an increase in the usage of computer systems to make the lives of consumers easier and more convenient. When people try to takeadvantage of the amazing capabilities and processing capacity of computer systems, however, security has been one of the most significant problems in the field of computer science. This is because attackers frequently try to break into systems and act maliciously, such as stealing vital information from a corporation, rendering the systems useless, or even destroying the systems. Internal attacks, such as pharming, distributed denial-of-service (DDoS), eavesdropping, and spear-phishing attempts, are often among the most difficult to identify of all well-known attacks. This is due to the fact that firewalls and intrusion detection systems (also known as IDSs) often guard against attacks from the outside. At this time, the majority of systems authenticate users by analysing a login pattern consisting of the user ID and password. As a result of this, we have proposed in this study a security solution that we have dubbed the Internal Intrusion Detection and Protection System (IIDPS) [1]. This solution recognises hostile or malicious behaviour carried out against a system at the System call level. IIDPS uses data mining and forensic profiling techniques in order to mine system call patterns, also known as SC-patterns, which are the longest system call sequences (SC-sequences) that have repeatedly appeared numerous times in a user's log file for the user. SC-patterns can be used to identify malicious activity. The user's computer usage history is used to compile the user's forensic features, which are then defined as a SC-pattern that commonly appears in the user's own submitted SC-sequences but is rarely utilised by other users. This information is gleaned from the user's computer.

979-8-3503-4821-7/23/$31.00 ©2023 IEEE

II. LITERATURE SURVEY.

The Internet of Things (IoT) is playing a crucial role in the world of online technology by providing the general public with services that are both speedy and intelligent. The Internet of Things (IoT) and any relevant devices need to be protected in order to fulfil the aim of ensuring security. Based on this research, a new intrusion detection approach was proposed for the purpose of enabling secured communication in wireless environments to provide security for the internet of things (IoT). This research offered a new strategy to feature selection that combines the Conditional Random Field (CRF) and spider monkey optimization in order to locate the most helpful features from the dataset. The CRF is an acronym for the Conditional Random Field (SMO) [2]. The IDS is produced when the process of selecting features and the process of classifying data are combined. In order to cut down on the size of the dataset by employing just the features that are really necessary, it is imperative that these characteristics be prioritised during the feature selection process. During the operation for classifying the data, the reduced dataset needs to be split into two distinct groups, depending on the specific condition. These categories may be "normal" and "abnormal," or they could be "attack." In this particular instance, the CRF is utilised to make the initial selection of the contributing characteristics [3]. After that, the SMO is used to finalise the useable features that were extracted from the reduced features dataset. In addition, the CNN is used to differentiate between attack scenarios and normal conditions within the dataset. Every organization's network infrastructure is always at risk from a wide variety of threats, including infiltration, zero-day, malware, and security breaches. These threats are made possible by the widespread use of the internet and smart devices. As a consequence of this, the traffic on the network requires constant surveillance by an intrusion detection system (NIDS). The findings of this study indicate a novel hybrid approach to the classification of attacks and the detection of intrusions. When it comes to handling high false-positive and low false-negative rates for attack categorization and intrusion detection, the solution that has been offered consists of three parts. In the first step of the process, the dataset is pre-processed by employing the min-max approach as well as the data transformation methodology. Second, the random forest recursive feature removal method is utilised in order to discover the best features that improve the overall performance of the model. After that, we use the Adaptive Neuro Fuzzy System (ANFIS) [4] along with various different kinds of Support Vector Machines (SVM) to classify probe, U2R, R2U, and DDOS attacks in order to detect infiltration. The validation rate of the proposed method was computed with the assistance of Fine Gaussian Support Vector Machines(FGSVM), and it was found to be 99.3 percent for the binary class. As a result of recent attacks on computer networks, data security has become the most important and critical component of all organisational data systems. This has aneffect on the international monetary system. The IntrusionDetection System is the system that is utilised the most in order to address concerns regarding networks (IDS). Because they are used to monitor system performance and give notifications when there is any unexpected behaviour, the administrator of the system needs to respond immediately to these warnings as soon as they are received [5].

This work presented [6] a statistical Nave Bayesian approach that will be used in Intrusion Detection Systems(IDS) systems in a variety of circumstances, such as analysing HTTP service-based traffic and identifying HTTP normal connections and attacks. Specifically, this work was done in response to the need for an improved method for detecting HTTP attacks. However, in order to determine which statistical approach is the most effective and efficient in identifying various types of attacks, a comparison study between them based on performance parameters will be examined. This will allow for the determination of which statistical approach is the most effective and efficient.Intrusion Detection System (IDS) [7, 8, 9] is utilised in this research to introduce and discuss the Multivariate Statistical Analysis (MSA) method through the use of Naive Bayesian Filtering and Multivariate Statistical Analysis (MSA). This plan aims to construct an advanced intrusion detection system (IDS) with improved efficiency to cut down on the number of false alarms that are generated. Cutting down on both false positive and false negative alarms will improve network security and increase the rate at which attacks are detected [10].

An intrusion detection system (IDS) is a piece of software that monitors a network for any malicious or unauthorised activity that could potentially breach security standards pertaining to the system's confidentiality, integrity, and availability. As part of this thesis, we conducted in-depth literature evaluations on the many different types of intrusion detection systems (IDS) [11], anomaly detection techniques, and machine learning algorithms that are available for use in the detection and categorization of intrusions. The construction of a hybrid intrusion detection system (IDS) employing machine learning methods is suggested as a result of this. By implementing appropriate machine learning algorithms into the detection systems that are currently in use, it is possible to achieve improvements in attack detection and categorization. In addition to this, they have made an attempt to compare the various machine learning algorithms by puttingthem to the test in a simulated environment so that they may evaluate how effective each of the algorithms is imperative that these production facilities have adequate protection against cyberattacks. Because of the daily rise in attacks, ensuring the network's safety should be the top priority for anyindustrial control system. Software is an essential component of control systems for industrial environments.

As a result of the vulnerabilities, the exchange of communication must now involve cybersecurity precautions. The security mechanisms that are in place for secure communication need to be improved so that they can keep up with the ever-evolving threats to information security, in addition to the security measures that are in place to combat such threats. Based on the findings of this study, the construction of a versatile and dependable network intrusion detection system should make use of deep learning architectures for the purpose of determining and classifying network attacks (IDS). The focus is on how deep learning or deep neural networks (DNNs) can enable flexible intrusion detection systems (IDS) with the capability to learn and detect known as well as new or zero-day network behavioural features, thereby removing the intruder from the system and reducing the risk of being compromised [12].

III. PROPOSED SYSTEM

Learning by machine offers a wide variety of useful applications, many of which are used on a daily basis. It seems that in the not-too-distant future, machine learning will dominate the entire world. As a consequence of this, we came up with the concept that methods of machine learning might be utilised to find a solution to the problem of recognising new attacks or zero-day attacks, which is a challenge that is confronted by technologically advanced businesses in today's world. The suggested system has the capability of identifying a user's forensic features by inspecting the associated security controls (SCs). This helps to improve the accuracy of attack detection and efficiently thwart insider attacks. The algorithm known as Naive Bayes, the algorithm known as Support Vector Machine, and the method known as Random Forest. were the three machine learning classifiers utilised here.

A. Proposed Architecture:

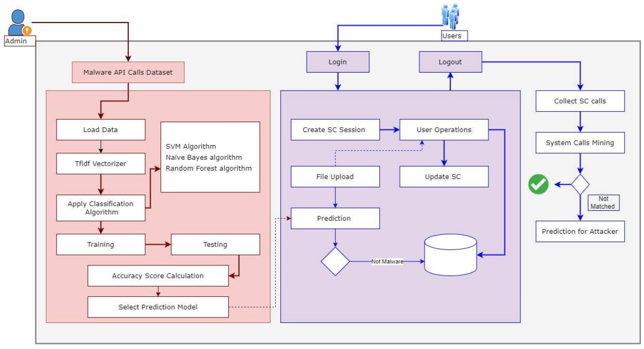

Figure 1: Proposed Architechture

The Architecture has not one but two distinct interfaces, known respectively as the admin and the user. The user will enter his or her login credentials into the system in order to log in. Following the user's successful login, a session will be created; this session will track the user's operations by making system calls. After that, the database entry for system calls is modified. Whenever it's necessary, the admin willcheck the user SC patterns. There are three algorithms that are used to pick the prediction model, evaluate the accuracy, and determine which algorithm has the maximum accuracy. System calls are gathered after the user logs out, and then mined to look for signs of malicious activity and identify the perpetrator of the assault.

B. Use Case Diagram:

The system's role and scope, as well as the requirements forthe system, are outlined in the use case diagram that was just shown.

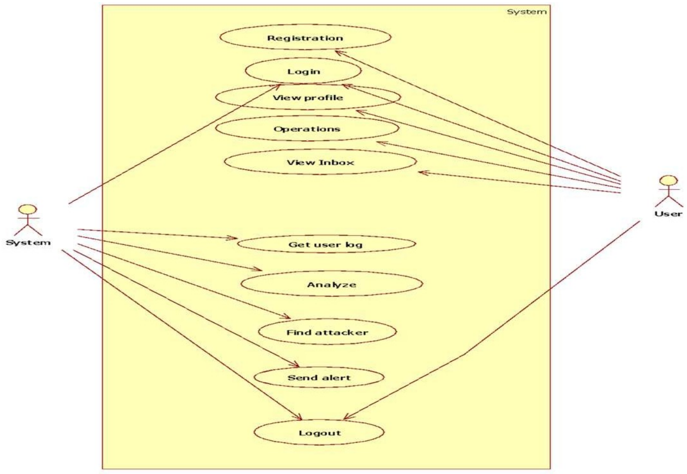

Figure 2.Usaecase Diagram for Proposed Architecture

Our use case diagram depicts two primary actors, 1. Admin 2. End User

Admin:

- The principal user of our system is admin. The administrator can check the user's SC-Patterns.
- The administrator can keep track of attack information such as the type of OS used, attack time and data, operating, attacker information, and attack intensity.
- Admin can get the user's log data, and then analyse to find the attacker.
- Admin uses his own credentials to login and logout.

User:

- In this scenario, User refers to a coworker who is part of a group of people who are employed at an establishment.
- Users can access the system by using their own login information in order to do so. After checking in, users have access to a variety of features, including the ability to browse and send files, upload and download content, and change their profiles if they so desire.
- When the user is being attacked by another player, the user has the option of receiving a notification in the form of an alert. The application will identify the party that is attempting to break in depending on the behaviourS.

IV. ANALYSIS

Dataset: Technology is rapidly evolving on a daily basis, and a great number of inventions and technological breakthroughs are being produced in order to defend computer systems from any attacks caused by network intrusions. Research on network infiltration was typically how the accessible data set from the KDD Cup 1999 was put to use. This data set is used in a variety of different methods for the purpose of machine learning. Nevertheless, there were a lot of issues with this dataset. The fact that the KDD dataset contains a significant number of records that are identical to others is, to begin, one of its primary drawbacks. 78% of the records in the training set have been duplicated, and approximately 75% of the total number of records in the testing dataset has been duplicated; as a result, our findings translate to biassed learning methods. The NSL KDD dataset, which now makes use of the dataset for use in machine learning algorithms, is accessible as a public data set for the network intrusion detection system. Even if there may potentially be a new version of the KDD Cup 99 dataset, this dataset is used in the NSL KDD dataset. There is no duplication of data in the new NSL KDD Test and Train dataset, which was created by combining only the most relevant information from the original KDD dataset. As a direct consequence of this, the findings of the research evaluation have been deemed to be a standard dataset that is consistent across all studies. The training dataset is comprised of four distinct kinds of assaults, as a general rule. First is the Denial of Service Attack (DoS), then comes the Probe Attack, then comes what is known as the User to Root Attack (U2R), and last comes the Root to Local Root Attack (R2L). This, in turn, is made up of more than 21 different assaults.

V. CONCLUSION

As a consequence of this, within this work, we suggest asecurity system that we refer to as the Internal Intrusion Detection and Protection System (IIDPS). This system is capable of identifying hostile behaviour that is aimed towardsa system at the SC level. The IIDPS mines system call patterns, also known as SC-patterns, which are defined as the longest system call sequences (System Call-sequences) that have repeatedly appeared numerous times in a user's log file for the user. These are the operations carried out by the user, such as sending a file, updating a file, or viewing a file, and they are validated by an administrator. The user's computer usage history is used to compile the user's forensic features, which are then defined as a SC-pattern that commonly appearsin the user's own submitted SC-sequences but is rarely utilised by other users. This information is gleaned from the user's computer.

REFERENCES

[1]. D.P. Gaikwad and Ravindra C. Thool. (2015). Intrusion detect ion system using bagging with part ial decision tree base classifier. Procedia Computer Science 49 (pp. 92-98). Elsevier.) .

[2]. A. Yulianto, P. Sukarno, and N. A. Suwastika, "Improving AdaBoost-based Intrusion Detection System (IDS) Performance onCIC IDS 2017 Dataset," J. Phys. Conf. Ser., vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012018.

[3]. A. H. L. and A. A. G. Iman Sharafaldin, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," Proc. 4th Int. Conf. Inf. Syst. Secur. Priv., no. Cic, pp. 108–116, 2018.

[4]. V. Hajisalem and S. Babaie, "A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection," Comput. Networks, vol. 136, pp. 37–50, 2018.

[5]. A. A. Aburomman and M. B. I. Reaz, "A survey of intrusion detection systems based on ensemble and hybrid classifiers," Comput. Secur., vol. 65, pp. 135–152, 2017.

[6]. Su-Yun Wua, E. Y. "Data mining-based intrusion detectors", ELSEVIER, 2009.

[7]. Akhilesh Kumar Shrivas, Amit Kumar Dewangan. An Ensemble Model for Classification of Attacks with Feature Selection based on KDD99 and NSL-KDD Data Set. International Journal of computer applications. Volume 99, No.15, 2014.

[8]. Yung-Tsung Hou, Y. C.-S.-M, "Malicious web content detection by machine learning", ELSEVIER, 2010.

[9]. P. Maniriho, "Detecting Intrusions in Computer Network Traffic with Machine Learning Approaches Detecting Intrusions inComputer Network Traffic with Machine Learning Approaches," no. April, 2020, doi: 10.22266/ijies2020.0630.39.

[10]. K. M. Sudar, P. Nagaraj, P. Deepalakshmi and P. Chinnasamy, "Analysis of Intruder Detection in Big Data Analytics," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402402.

[11]. K. M. Sudar, M. Beulah, P. Deepalakshmi, P. Nagaraj and P. Chinnasamy, "Detection of Distributed Denial of Service Attacks in SDN using Machine learning techniques," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402517.

[12]. Praveena, V., Vijayaraj, A., Chinnasamy, P., Ali, I., Alroobaea, R. et al. (2022). Optimal Deep Reinforcement Learning for Intrusion Detection in UAVs. CMC-Computers, Materials & Continua, 70(2), 2639–2653.