International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Detecting Intruders in NetworkUsing Machine Learning

S.Uma.M.E[1], Sharath.R[2], Sujith.S[3], Vasanth.K[4]

[1]Assistant Professor, Computer Science and Engineering, Paavai Engineering college
[2,3,4,5]Student, Computer Science and Engineering, Paavai Engineering college

Abstract:

Malware detection plays a crucial role in cyber-security with the increase in malware growth and advancements in cyber-attacks. Malicious software applications, or malware, are the primary source of many security problems. These intentionally manipulative malicious applications intend to perform unauthorized activities on behalf of their originators on the host machines for various reasons such as stealing advanced technologies and intellectual properties, governmental acts of revenge, and tampering sensitive information, to name a few. Malware detection methods rely on signaturedatabases, including malicious instruction patterns in today's practice. The signature databases are used for matching against a signature generated from a newly encountered executable. Nevertheless, more efficientmitigation methods are needed due to the fast expansion of malicious software on the Internet and their self- modifying abilities, as in polymorphic and metamorphic malware. In this work, it detects Network Intrusion anomalies by using NSL-KDD dataset. The user enters the hacking parameters in the front end. The model predicts the type of attack and gives information about the type of attack to the user. The project is fully responsive and completely based on session and cookies (Client-server protocol). Then we activated our malware security device which forms production for the set of attack. This will help many cyber threads.

Keywords: Intruders detection, NSL-KDD dataset, Client- Server protocol.

1.INTRODUCTION

Since the beginning of computer networks, network intrusion detection systems (NIDS) have existed for a considerable amount of time. The main known interruption recognition framework was created in 1980 by James Anderson, who made a program called "The Gatecrasher Identifier" to screen Unix frameworks for unapproved access.As computer networks became more prevalent in the late 1980s and early 1990s, a number of businesses began developing commercial NIDS products. One of the earliest NIDS items was Sri Global's Organization Interruption Identification Framework (NIDS), which was delivered in 1991. By comparing network traffic to a database of known attack signatures, this system used a signature-based approachto detect network attacks.

Snort and Bro are two of the open-source NIDS projects that were developed in the middle of the 1990s. Martin Roschdeveloped Snort in 1998, and it is still widely used today. VernPaxson created Bro, which is now known as Zeek, in 1995. Itwas made to be more adaptable than conventional NIDS systems. At the beginning of the 2000s, NIDS systems that used machine learning and other cutting-edge methods to detect network attacks became more popular. In 2002, the Pacific Northwest National Laboratory of the US Departmentof Energy created one of these systems. The Adaptive Anomaly Detection System (AADS)

employed machine learning algorithms to learn normal network behavior and identify abnormalities.

Currently, many NIDS systems employ a combination of signature-based and behavior-based methods to detectnetwork attacks. To provide a more comprehensive view of anorganization's security posture, there is also a trend toward integrating NIDS with other security technologies, such as endpoint detection and response (EDR) systems.

The computer was attacked in just 8 hours of installationand in 21 days the computer was attacked 20 times and compromised 40 days after installation. The constant growthof Internet users and the provision of online services such as banking and shopping services, provide hacking criminals with a suitable environment to perform their cybercrimes, which leads to a rise in the expenses that are paid to protect the systems. The international damage cost that has been caused by cyber maliciousness takes the attention of the researchers due to its rapid growth. In 2021, this cost is predicted to be around 6 USD trillion according toCybersecurity Ventures Official Annual Cybercrime Report.Malware is considered the biggest threat to cybersecurity andfalls under several types such as viruses, worms, trojan horses, rootkits, and ransomware since malware causes directharm to the systems or steals their sensitive information. In addition, malware represents the most frequent sort of computer, network, or user attacks to cause damage or steal sensitive information. Over recent years, the number of malicious software has increased by 22.9%, which reflects analarming rise in threats to computer user. The authors of stated that there were around one billion infected files in January 2021. So, this work makes the progressive way to develop the essential factor to find the type of attack and maintain protected the system over any set of anomalies invaders over the network. In simpler term it acts as an Anti-virus for the system.

1.1 OBJECTIVE

[1] To Predict the type of attack happened over it.
[2] Maintain the malware function active to avoid attack.
[3] Prevent the data corrupt by using malware detectionapplication.

2.EXISTING SYSTEM

To capture malware, antivirus companies typically employ signature-based detection techniques, which rely on featuresextracted from previously known malware to identify it.However, this method can only identify known malware. Thismethod does not permit the detection of zero-day malware, also known as brand-new malware. Additionally, malware authors employ evasion strategies like encryption and obfuscation to avoid early detection. Protecting systems frommalware is essential now that we know how devastating it can be. For about twenty years, researchers have been looking into how to use machine learning to detect malware. Neural networks, decision trees, support vector machines (SVM), ensemble methods, and numerous other well-known machinelearning algorithms have all been tried by researchers. Late review papers give extensive data on malware recognition methods utilizing AI calculations.

2.1.1 DISADVATAGES

[1] Organizations are challenged to find, train, and retain malware analysis staff. [2] Malware analysis tools lack automation, integration, and accuracy.[3] Malware analysis can become a time-consuming and error-prone manual process across multiple disparate tools and disconnected workflows.

3. PROPOSED SYSTEM

It uses pre-trained datasets and machine learning algorithms to identify invaders over the network for accurate results. Additionally, it provides the means to win and escape this attack. In this work the malware assaults like Trojan, Worms,Ransomware are distinguished and gives the anticipation forthe assault. In a nutshell, it prevents this set of malware attackson the system by acting as an anti-virus.

3.1 ADVANTAGES

- Identify the attackers.
- Counterintelligence and cyber-warfare.
- Executable entry point.
- Detects missing files.
- Detects invalid files.
- Real time Malware Detection System

4. ALGORITHMS USED

A.CONVENTIONAL NEURAL NETWORK

[1] In a typical feedforward neural network, information flowsin one direction, from input nodes through hidden layers to

output nodes.[2] Each node in a layer receives input from theprevious layer and applies a mathematical transformation to produce an output, which is then passed to the next layer.[3] The weights and biases of the network are adjusted during thetraining process to optimize the network's performance on a particular task. [4] Predictive modeling, speech recognition, natural language processing, and image recognition are just afew of the many uses for conventional neural networks.Recurrent neural networks (RNNs) and other more advanced architectures have been developed to overcome these limitations.[5] Recurrent neural networks (RNNs) are used for some updating features of the conventional Neural Network (CNN) using the Deep learning concept BiLSTM.

B. BiLSTM

[1] The acronym BiLSTM refers to bidirectional long-term memory. In a traditional LSTM (Long Short-Term Memory)network, the input sequence is processed in a single direction,from the first element to the last. However, this approach maynot be sufficient for some tasks where information from bothdirections is important. For instance, in natural language processing, the meaning of a word may be influenced by thewords that come before and after it. A BiLSTM network hastwo LSTM layers, one of which processes the input sequencein the forward direction and the other in the reverse direction.[4] The outputs of these two layers are then concatenated to produce the network's final output.[5] By processing the sequence in both directions, the BiLSTMnetwork is able to identify dependencies between elements ofthe sequence that a unidirectional LSTM might miss.

5.PROBLEM DEFINITION

Malware (vindictive programming) is a critical danger to PC frameworks, cell phones, and organizations around the world. Malware can cause different kinds of harm, including taking touchy information, capturing frameworks, andupsetting basic administrations. Since signature-based malware detection systems aren't good at finding new and better malware, more sophisticated methods are needed

morethan ever. By executing themselves on the system, malware code typically aims to violate a system's or device's security policies. Computer system flaws can be exploited by attackersto steal sensitive data, spy on an infected system, or take control of the system. Although malware is typically thoughtof as malicious "files," malicious code fragments typically reside within a file rather than representing the entire file. Typically, feature extraction is the first step in building a model for malware detection. This can be done using static ordynamic analysis, or hybrid analysis in some cases.

6.OVERVIEW OF THE PROJECT

[1] The primary objective is to identify intruders who are employing attacks over the network as a means of settling a high-traffic monetization goal.[2] The primary component ofthis is to make the essential factor to appeal the anti-virus forthe system virtually.[3] To discover and settle down the workdefines to make the user observe what kind of attack they have comes by and how to prevent that.

7. SYSTEM DESIGN

7.1 ARCHITECTURE DIAGRAM

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organizedin a way that supports reasoning about the structures and behaviors of the system. System architecture can comprise system components, the externally visible properties of those components, the relationships (e.g., the behavior) between them. It can provide a plan from which products can be procured, and systems developed, that will work together to implement the overall system. There have been efforts to formalize languages to describe system architecture, collectively these are called architecture description languages(ADLs).

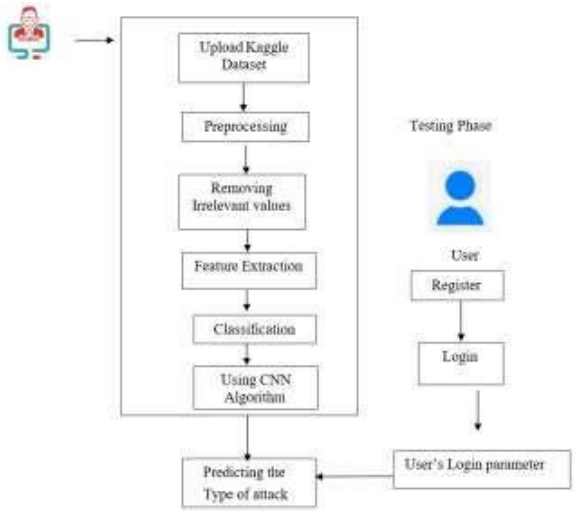

Figure no.6.1 Architecture diagram

Here, the user enters the login parameter, and the system automates to identify the compatibility of which type of attack happens in the system are find and gives the exact set of information about the type of attack happens over the systemare analyzed by using CNN algorithm. And gives the solutionfor the attack are defined.

7.2 DATAFLOW DIAGRAM

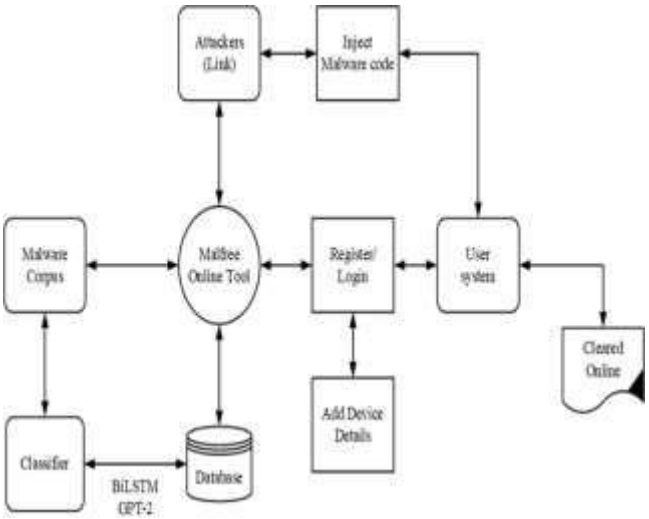

Figure no 6.4 Dataflow Diagram

8. MALWARES AND ITS TYPES

Malware, short for malicious software, is a type of software designed to harm, exploit, or disrupt digital devices, systems,and networks. It is a common tool used by cybercriminals tocompromise the security of computers, servers, mobile devices, and other digital assets.

Here's a list of the common types of malwares and their malicious intent:

Figure no.7.1 Types of Malwares

8.1 MALWARES IMPLEMENTED

8.1.1 TROJANS

A Trojan (or Trojan Horse) disguises itself as legitimate software with the purpose of tricking you into executing malicious software on your computer. Trojans are commonly downloaded through email attachments, website downloads,and instant messages.

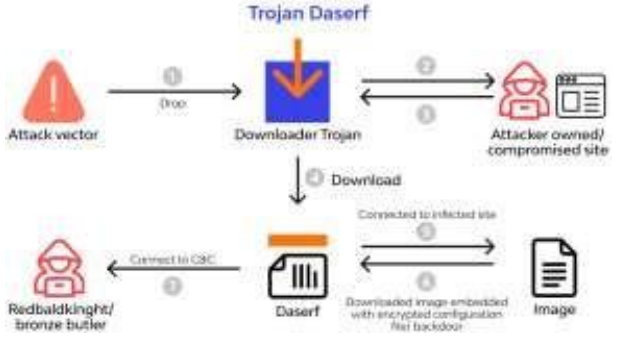

Figure no.7.2 Trojans attack

8.1.2 WORMS

A worm replicates itself by infecting other computers that areon the same network. They're designed to consume bandwidth and interrupt networks.

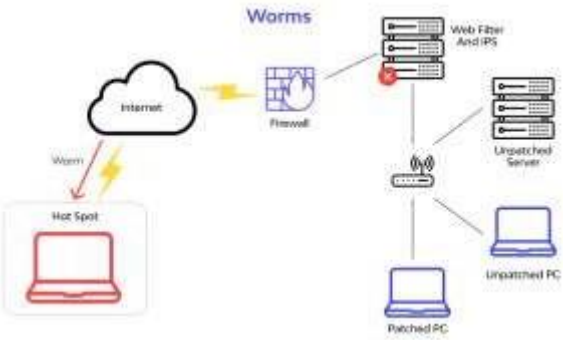

Figure no 7.3 Worms attack.

8.1.3 .RANSOMWARE

Ransomware is designed to encrypt your files and block access to them until a ransom is paid.

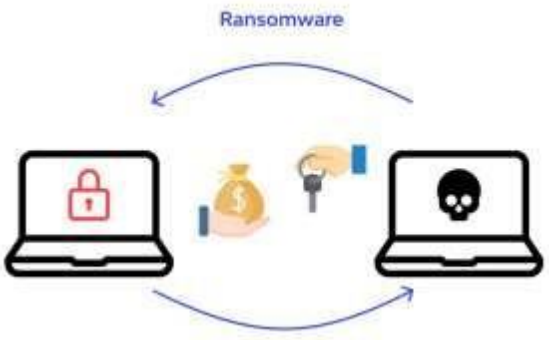

Figure no 7.4 Ransomware attack.

9. WORKING OF THIS MODULE

In this work the user enters the login parameter and register by itself. Once the registration is done the user checks the system and if the system is safe or not. The machine learningalgorithm which works in backend and produces conclusive evidence of live time Malware attack in the system. Which gives the perfect ideas for the user that what type of attack they have come across and what can they do. So the system of our working model prevents the system by keeping the cyber security system implementation. Using TensorFlow and cybersecurity packages our work implemented an Anti- virus to prevent the system from the Malware attack. Whereas the admin side process the dataset implemented in the backend to follows it features preprocesses and established prevents the system.

10. CONCLUSIONS

Malicious software applications, or malware, are the primarysource of many security problems. These intentionally manipulative malicious applications intend to perform the unauthorized activities on behalf of their originators on the host machines for various reasons such as stealing advanced technologies and intellectual properties, governmental acts ofrevenge, and tampering sensitive information, to name a few. This project introduces MalFree, an interactive visualizationplatform for hybrid analysis and diagnosis of malware.

This approach first represents the behavioral properties of themajor malware classes (such as Trojan or backdoor), aiming to capture the common visual signatures of these malicious applications. MalFree implements a web-based prototype fordemonstrating our approach to analyzing 60 malware samplesfrom seven different classes. We focused on opcodes and operands, instead of opcodes only, to develop stacked bidirectional long short-term memory (BiLSTM) models andthe decoder-based transformers generative pretrainedtransformers 2 (GPT-2) models. The resulting accuracy rate 95.4% shows that it is possible to classify malicious and benign assembly codes by GPT-2 with a custom pretrained model. By experimental results, we showed that using byte streams of different formats may contribute to performance improvements.

This also allowed for faster detection of malware classes, permitting a quicker response in anti- malware cybersecurity applications. Overall, the application of this project can help identify malware types faster, preventmalware attack and more accurately than contemporary approaches which can help save time when defending againstmalwares.

10.1FUTURE ENHANCEMENTS

[1] While the deep learning approach is robust and flexible, there are certain steps which can be taken to improve their performance and better classify the data.[2] Integration with other security tools: MalFree can be enhanced to integrate with other security tools, such as firewalls and intrusion detection systems, to provide a more comprehensive cybersecurity solution.[3] Support for multiple operating systems: Currently, MalFree is designed to work with aspecific operating system. Future enhancements can include support for multiple operating systems, such as Windows andLinux, to provide a more comprehensive cybersecurity solution. Integration with threat intelligence feeds:[4] The integration of MalFree with threat intelligence feeds canenhance its capabilities to detect and prevent new and emerging threats.[5] Threat intelligence feeds can provide MalFree with up-to-date information on known threats and vulnerabilities, allowing the system to detect and prevent themin real-time.

10.2 FUTURE SCOPE

Nowadays, the attack in the devices over the network is increasing high and high, so making the prevention is more needed. So, this work gives the best way to come out from theattacker's threat over the network and prevents the monetarization threat from the Invaders. This will help the user and the higher sector sections to fall over the wrong andpunishable cybercrimes. Implementing more malware samplewill help the security crises.

This development for the futurewill be makes the major positive side for the user and the system. Many more malwares travel nowadays to reduce andprevent from this we need to apply and implement the aspectual things for the operating system.

The MalFree is used to reduce the capabilities of the feed the windows and Linux.This intelligence to detect the allowance variability and prevent the up-to-date detection over the network.

REFERENCE

1. Caviglione, L.; Choras, M.; Corona, I.; Janicki, A.; Mazurczyk, W.; Pawlicki, M.; Wasielewska, K. Tight Arms Race: Overview of Current Malware Threats and Trends in Their Detection. IEEE Access 2021, 9, 5371–5396. [CrossRef]
2. Morgan, S. Cybercrime Damages $6 Trillion by 2021. 2017. Available online:

https://cybersecurityventures.com/hackerpocalypsecybercrim e-report-2016/ (accessed on 15 July 2021).

3. Cannarile, A.; Dentamaro, V.; Galantucci, S.; Iannacone,A.; Impedovo, D.; Pirlo, G. Comparing Deep Learning andShallow Learning Techniques for API Calls MalwarePrediction: A Study. Appl. Sci. 2022, 12, 1645. [CrossRef]
4. Villalba, L.J.G.; Orozco, A.L.S.; Vivar, A.L.; Vega,E.A.A.; Kim, T.-H. Ransomware Automatic Data AcquisitionTool. IEEE Access 2018, 6, 55043–55051.
5. Urooj, U.; Al-Rimy, B.A.S.; Zainal, A.; Ghaleb, F.A.; Rassam, M.A. Ransomware Detection Using the Dynamic Analysis and Machine Learning: A Survey and Research Directions. Appl. Sci. 2022, 12, 172. [CrossRef]
6. Hansen, S.S.; Larsen, T.M.T.; Stevanovic, M.; Pedersen, J.M. An approach for detection and family classification of malware based on behavioral analysis. In Proceedings of the 2016 International Conference on Computing, Networking and Communications (ICNC), Kauai, HI, USA, 15–18 February 2016; pp. 1–5. [CrossRef]
7. Vignau, B.; Khoury, R.; Halle, S. 10 Years of IoT Malware: A Feature-Based Taxonomy. In Proceedings of the 2019 IEEE 19thInternational Conference on Software Quality, Reliability and Security Companion (QRS-C), Sofia,Bulgaria, 22–26 July 2019; pp. 458–465. [CrossRef]
8. Asam, M.; Hussain, S.J.; Mohatram, M.; Khan, S.H.;Jamal, T.; Zafar, A.; Khan, A.; Ali, M.U.; Zahoora, U. Detection of exceptional malware variants using deep boosted feature spaces and machine learning. Appl. Sci. 2021, 11, 10464. [CrossRef]
9. Sahay, S.K.; Sharma, A.; Rathore, H. Evolution of Malware and Its Detection Techniques. In Advances in Intelligent Systems and Computing; Springer: Singapore,2020; Volume 933, pp. 139–150.
10. Onyedeke obinna Cyrill, Taoufik Elmissaoui, Okoronkwo M.C,, Ihedioha Uchechi .M, Chikodili H.Ugwuishiwu5, Okwume .B. Onyebuchi6 , Signature basednetwork intrusion detection system using feature selection onandroid, 2020.
11. Zakiyabanu S. Malek1, Bhushan Trivedi,'' Userbehavior-based intrusion detection system,2018.
12. Thien Duc Nguyen, Phillip Rieger, Markus Miettinen, Ahmad-Reza Sadeghi, ''Poisoning attacks on federated learning-based iot Intrusion detection system, 2021.
13. Aan Erlansari1, Funny Farady Coastera2, Afief Husamudin 3, ''Early intrusion detection system (ids) using snort and telegram approach, Dec. 2020.
14. M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, ''Wearable 2.0: IoT intrusion detection system,'' IEEE Common., vol. 55, no. 1, pp. 54–61, Jan. 2016.