ISSN No:-2456-2165

A Survey on Security System using Machine Learning and Deep Learning

Smit Parikh
Computer Science and Engineering (Network and security)
MIT-ADT University
Pune, India

Srikar Banka
Computer Science and Engineering (Network and security)
MIT-ADT University
Pune, India

Atrayee Chatterjee
Computer Science and Engineering (Core)
MIT-ADT University
Pune, India

Abstract:- In the era of technical advancements, cybersecurity becomes an important aspect of the computing and networking world. Our systems and networks are more prone to hacking than it was before. Hacking refers to access to systems by unauthorized users in order to steal, change, or destroy information. There are different kinds of threats that can be sent across a network. Therefore, not only the hacking has to be stopped but also it is equally important to understand the origin to stop future attempts. This project fetches the IP address of the system and makes it easier for the cyber-crime department to trace the location of the hacker. However, it uses machine learning and deep learning approaches to learn the system for viruses and remove them and analyze previous history and work and store data according to that. The project aims to identify the threats to the system and resolve them without any harm.

Keywords:- *Cyber-security, Hacking, Machine learning, Deep learning, Virus.*

I. INTRODUCTION

➢ *About Security*

Computer security basically is the protection of computer systems and information from harm, theft, and unauthorized use. It is the process of preventing and detecting unauthorized use of your computer system.

Often people confuse computer security with other related terms like information security and cybersecurity. One way to ascertain the similarities and differences among these terms is by asking what is being secured.

So, Computer security can be defined as the controls that are put in place to provide confidentiality, integrity, and availability for all components of computer systems.

Computer security is mainly concerned with three main areas:

- Confidentiality is ensuring that information is available only to the intended audience
- Integrity is protecting information from being modified by unauthorized parties
- Availability is protecting information being available to unauthorized parties

➢ *About hacking and virus*

Hacking is the process of gaining unauthorized access into a computer system or group of computer systems. This is done through the cracking of passwords and codes which gives access to the systems. Cracking is the term that specifies the method by which the password or code is obtained. The person who undertakes hacking is known as the hacker. Hacking can be done on single systems, a group of systems, an entire LAN network, a website, or a social media site or an email account. The access to a password is obtained by the hacker through password cracking algorithm programs.

It goes without saying that most of the individuals, as well as business associations, use computers and laptops for all their daily needs. Especially for organizations (of any form), it is essential to have a computer network, domain, or website, Wide Area Network (WAN) for a seamless flow of information and business applications. Consequently, these networks are under exposure to high risk of the outside world of hacking and hackers.

A computer virus is a malicious software program loaded onto a user’s computer without the user’s knowledge and performs malicious actions.

A virus is a fragment of code embedded in a legitimate program. Viruses are self-replicating and are designed to infect other programs. They can wreak havoc in a system by modifying or destroying files causing system crashes and program malfunctions. On reaching the target machine, a virus dropper(usually trojan horse) inserts the virus into the system.

➢ *Hacking and virus types*

There are different types of hackers in the world. They are:-

- **White hat: -** White Hat hackers are also known as Ethical Hackers. They never intend to harm a system, rather they try to find out weaknesses in a computer or a network system as a part of penetration testing and vulnerability assessments.

- **Black hat: -** Black Hat hackers, also known as crackers, are those who hack in order to gain unauthorized access to a system and harm it's operations or steal sensitive information.

- **Grey hat: -** Grey hat hackers are a blend of both black hat and white hat hackers. They act without malicious intent but for their fun, they exploit a security weakness in a computer system or network without the owner's permission or knowledge.

There are different types of hacking methods. Some of which are:-DoS and DDoS attacks, Keylogging, Cookie Stealer, Bait and Switch, IoT Attacks, Fake wireless Access Points(WAPs), Phishing, Clickjacking or UI Redressing, Passive Attacks, Social Engineering

There are different types of viruses. They are as follows: - File Virus, Boot sector Virus, Macro Virus, Source code Virus, Polymorphic Virus, Encrypted Virus, Stealth Virus, Multipartite Virus, Armored Virus.

➢ *Existing solution*

There is antivirus software in the existing solutions which help to remove viruses from the system if there are any. It also helps us to protect from the internet like accessing a page which may harm our system and so on.

The major drawback of this system is that different antiviruses detect different viruses and some don't detect the virus.

➢ *Proposed System*

We are proposing a software in which hacking will be prevented by the software and any type of virus in the system would get detected no matter how old or new it is, using Machine learning and Deep learning.

II. LITERATURE REVIEW

In 2019, Smit Parikh [1] has offered that Hacking includes an unknown person taking all your info, and misusing it if they like. Not only does this project stop the hacking process but it also gives the hacker details i.e. the IP address and so on. It makes it easier for the Department of Cybercrime to trace the hacker's location and arrest the person. If our machine has some form of the virus in the network, this program eliminates the virus and also analyzes the path of the virus and, if possible, sends a counter virus to that device, or whatever the path, and generates a backdoor for it and gives us the data of the person/device it was sent to. The basic goal of this project is to minimize cyber-crime and make an approach to improving the security to minimally damage Internet use.

In 2020, Suleiman Y. Yerima [2] has presented a deep learning-based method for the identification of phishing sites with high accuracy. To differentiate legitimate sites from phishing sites, the proposed solution uses convolutional neural networks (CNN) for high precision classification. The models are tested using a sample from actual 6,157 and 4,898 phishing websites. So, the CNN based models have proved to be highly successful in detecting unknown phishing sites, based on the findings of comprehensive experiments. The approach described in this paper contrasts favorably with state-of-the-art phishing website detection based on deep learning.

In 2020, Chinthapalli Sudheer [3] has offered machine learning as a potential solution for improving the wrong positive rate and for increasing SOC analyst productivity. Within this post, in the real organizational sense, a user-centric learning system for the Internet Safety Functional Centre has been built as the SIEM (Data Information and Event Management) program is in place to simplify the various preventive technologies and flag warnings for protection events. In general, the number of alerts is incorrect with the majority and is more than SOC's ability to control all information. So, this article addresses two reader categories. The first category is smart researchers who are unfamiliar with data scientists or computer security fields but who engineers should build machine learning systems for machine health. The second visitor categories are internet security experts with deep experience and skills in information security. Ultimately, the account is used as an example to demonstrate full steps from data collection, label creation, feature engineering, machine learning algorithm, and sample performance evaluations using the computer built in the SOC production of Seyondike.

In 2020, Mohamed Amine Ferrag [4] has presented a survey, a review of deep learning approaches for cybersecurity intrusion detection, the datasets used, and a comparative study. The dataset plays an important role in the detection of intrusion, which is why we identify 35 well-known cyber datasets and group them into seven categories. The seven deep learning models have been evaluated, including recurrent neural networks, functional neural networks, restricted Boltzmann machines, functional-belief networks, convolutional neural networks, deep Boltzmann machines, and deep autoencoders. The output for each model is analyzed in two classification groups (binary and multiclass) within two new actual traffic datasets, namely the CSE-CIC-IDS2018 dataset and the Bot-IoT dataset. Furthermore, the most critical performance metrics, namely, accuracy, false alarm rate, and detection rate is used, to assess the effectiveness of several methods.

In 2020, Ashok Panwar [5] has proposed a routing technique that is a trust-based routing protocol that evaluates the characteristics of intermediate nodes in the network. A weighted trust value for all intermediate routers is determined during node assessment to construct a safe path establishment. The network parameters of nodes are used for calculating weighted confidence i.e. packet drop ratio, RRT, and energy consumption rate. In addition, a trust level is used to identify malicious and legitimate nodes in a network pathway. Hence the trust threshold is used to make decisions to choose a safe and efficient course. Eventually, a comparative analysis is conducted between conventional AODV and proposed trust-based AODV.

In 2018, S Megira [6] has aimed to study malware using malware samples to better understand how computers and devices can be compromised, the extent of threats they present, and how to defend devices against them. Today's growing usage of the internet and technology cannot be distinguished from cybercrime which could endanger its users. The cyber threat, including malware, attempts to penetrate the offline computer or mobile device or the internet, chat (online) and anyone can be a victim. Cybercriminals also use malware, often known as malicious software, to accomplish their goals by monitoring internet activity, stealing personal information, or blocking computer access. Malware has its own mechanism of protection and can hide from the antivirus or even corrupt the antivirus itself. Malware can be managed by understanding how to function when targeting a computer system.

In 2017, Haris A. Khan [7] has explored the hypothesis of a computer virus threat, and how destructive it can be if executed on a targeted machine. An analysis has been performed from the data derived from various scenarios and lab experiments performed in a simulated setting. Computer virus-related threats to information security will infect computers and other storage devices by copying themselves into a file and other executable programs. These files get the infection and allow attackers to use backdoors to connect to target systems. The findings of this study indicate that proper security implementations and the use of up-to-date patches and antivirus programs in operating systems help users avoid data loss and any viral attack on the system. Nonetheless, this finding may be used for more studies in network security and related fields; this study would also help computer users use the potential steps and strategies to secure their systems and information from any future attacks on their network systems.

In the below table, we have created a table of the above survey papers with their features and challenges/drawbacks if they have.

Year [Citation]	**Methodology**	**Features**	**Challenges**
2019 [1]	Securing System	It has a system which can be secure using the various loophole	The system must be updated manually by the user
2020 [2]	Deep-learning based	It uses convolutional neural networks (CNN) for high precision for identification of phishing sites	The model training process can be improved by automating the search and selection of the key influencing parameters
2020 [3]	Machine-learning based framework	It has a solution for improving the wrong positive rate of the warnings and for increasing SOC analyst productivity	Detection accuracy is less as compared to some of the other approaches
2020 [4]	Deep-learning with datasets	It is used for cybersecurity intrusion detection and also assesses the effectiveness of several methods	No challenges
2020 [5]	Trust-based routing protocol	It uses secure routing and improved routing protocol. It evaluates characteristics of intermediate nodes in the network and hence avoids a wormhole attack	Limited network characteristics have been taken into account
2018 [6]	Reverse Engineering	It is used for malware analysis, detection, and prevention	Absence of a few malware datasets
2017 [7]	Lab analysis	It performs an analysis from the data extracted from different tests of scenarios and labs conducted in a test environment and hence helps in preventing viruses	No challenges

Table 1:- Literature Review

III. PROPOSED SYSTEM

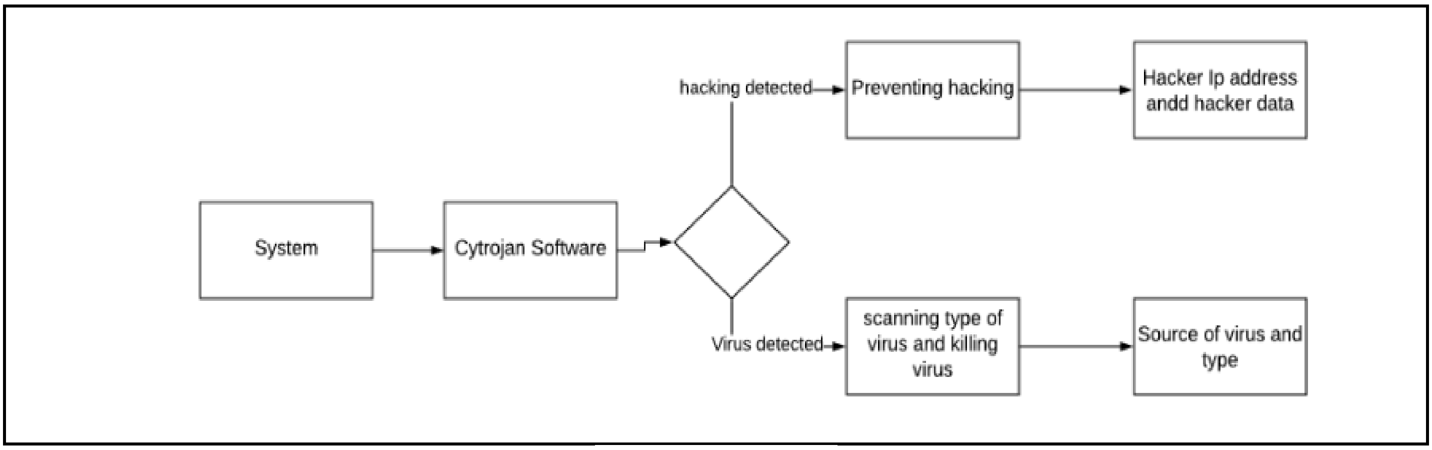

Fig 1:- Block Diagram

- As [1] paper they have mentioned the secure system to prevent hacking and virus but in that one drawback is that they have to keep it updated manually. So, we proposed an easier way to use deep learning and machine learning to automatically update the system on its own.
- In the existing system, if someone tries to hack someone's system then a person of a cyber-security department has to be present physically to stop it. So, while waiting for a security expert we may lose a lot of data. Thus, this method can compromise security. In this, we proposed a software where we don't need any security expert physically to stop the hacker to hack the system.
- Here we will use machine learning and deep learning for systems to prevent future attacks. So that system can learn attacks that happened in history and can prevent upcoming attacks. In figure 2 shows how the system will work.

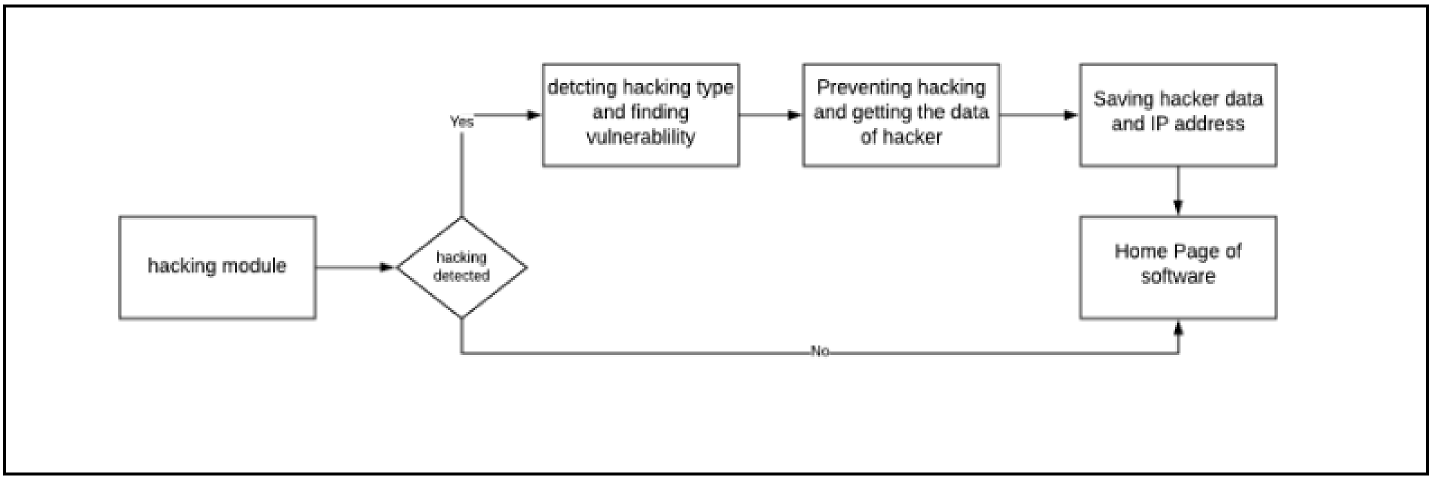
detcting hacking type and finding vulnerablility
Preventing hacking and getting the data of hacker
Saving hacker data and IP address
Yes
hacking module
hacking detected
Home Page of software
No

Fig 2:- Hacking prevention

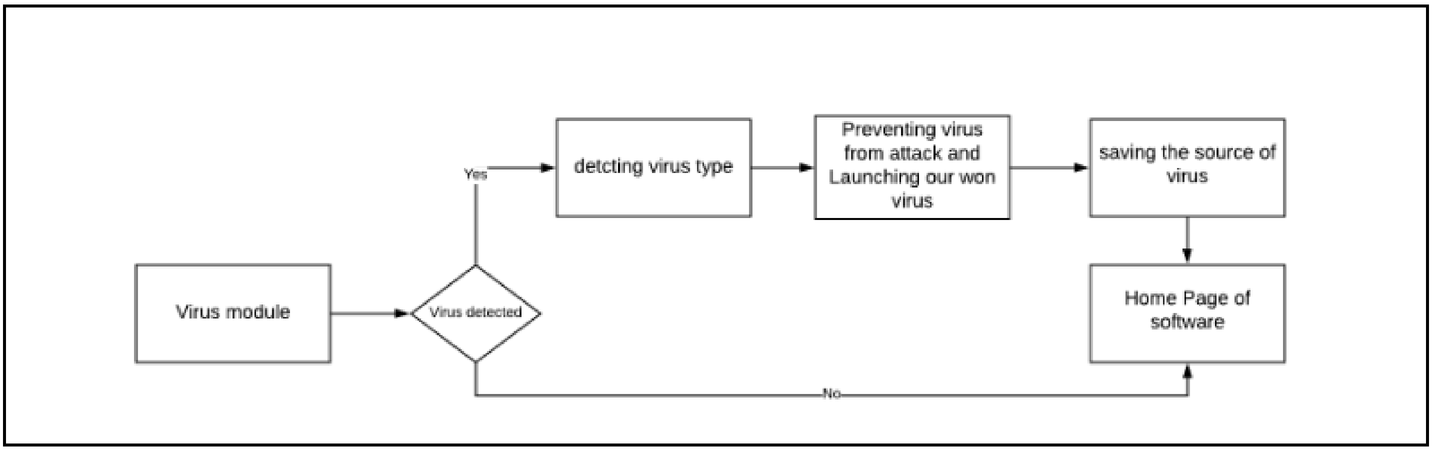
detcting virus type
Preventing virus from attack and Launching our won virus
saving the source of virus
Yes
Virus module
Virus detected
Home Page of software
No

Fig 3:- virus prevention

- In figure 3, We have shown how the hacking module will work. In a hacking module when someone tries to attack a system the software will detect that system is trying to hack.
- After hacking is detected the software will find the vulnerability or by checking the previous history with the help of deep learning and machine learning will try to find the hacking type.
- Once the vulnerability and hacking type are found the software will try to prevent the hacker to stop hacking and also will try to recover the lost data.
- Once the hacking is prevented the software will try to get the IP address using different methods. Once the process is complete all data like hacking type, time of the hacking, IP address of the hacker, etc. will be stored in the database and will return to the home page.
- In figure 4, We have shown how the virus module will work. When the software detects that a virus has been implanted in the system then the software will try to detect a type of virus using the previous history with the help of deep learning and machine learning and will prevent the virus.
- After the system is secure all data like the source of the virus, type of virus, etc. will be stored in the database and will return to the home page.

IV. CONCLUSION

After doing research in more detail on paper [1] and finding drawbacks we have prepared a survey in which we have covered the drawback of paper [1].

This paper has portrayed the survey related to system security with help of ML and Deep Learning here we have surveyed the Paper-based on the existing solution and proposed system based on [1] paper and also on the basis of their drawbacks, we proposed a better solution using ML and deep learning to secure the system from new and old attacks.

REFERENCES

[1]. A software to prevent hacking and virus (Cytrojan) by Smit Parikh, Shruti Agrawal and Simran Singh in 2019

[2]. High Accuracy Phishing Detection Based on Convolutional Neural Networks by Suleiman Y. Yerima and Mohammed K. Alzaylaee in 2020.

[3]. User-Centric Machine Learning Framework for Cyber Security Operation Center by Chinthapalli Sudheer and M S Venugopala Rao in 2020.

[4]. Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study by Mohamed Amine Ferrag, Leandros Maglaras, Sotiris Moschoyiannis, and Helge Janicke in 2020.

[5]. A Trust-Based Approach for Avoidance of Wormhole Attack in MANET by Ashok Panwar, Bhavana Panwar, D. Srinivasa Rao, and G. Sriram in 2020.

[6]. Malware Analysis and Detection Using Reverse Engineering Technique by S Megira, A R Pangesti, and F W Wibowo in 2018.

[7]. Computer Virus and Protection Methods Using Lab Analysis by Haris A. Khan, Ali Syed, Azeem Mohammad, and Malka N. Halgamuge in 2017.

[8]. Malware Analysis and Vulnerability Detection Using Machine Learning by Farrukh A. Khan, Muhammad Faisal Amjad, Yin Zhang, and Hammad Afzal in 2020.

[9]. Using a Subtractive Center Behavioral Model to Detect Malware by Ömer Aslan, Refik Samet, and Ömer Özgür Tanrıöver in 2020.

[10]. Phishing Trends and Intelligence Report: The Growing Social Engineering Threat by Phishlabs in 2019.

[11]. A survey of network-based intrusion detection data sets by M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho in 2019.

[12]. Application of deep learning to cybersecurity by S. Mahdavifar and A. A. Ghorbani in 2019.

[13]. Timeline of computer viruses by S. Spencer in 2019.

[14]. MALDC: a depth detection method for malware based on behavior chains by H. Zhang, W. Zhang, Z. Lv, A. K. Sangaiah, T. Huang, and N. Chilamkurti in 2019.

[15]. MalDy: portable, data-driven malware detection using natural language processing and machine learning techniques on behavioral analysis reports by E. B. Karbab and M. Debbabi in 2019.

[16]. Classifying Phishing Email Using Machine Learning and Deep Learning by S. Bagui, D. Nandi, S. Bagui, and R. J. White in 2019.

[17]. Metamorphic malicious code behavior detection using probabilistic inference methods by C. Choi, C. Esposito, M. Lee, and J. Choi in 2019.

[18]. McAfee mobile threat report Q1 by R. Samani and G. Davis in 2019.

[19]. Deep learning-based multi-channel intelligent attack detection for data security by F. Jiang, Y. Fu, B. B. Gupta, F. Lou, S. Rho, F. Meng, and Z. Tian in 2018.

[20]. A state of the art survey on polymorphic malware analysis and detection techniques by E. Masabo, K. S. Kaawaase, J. Sansa-Otim, J. Ngubiri, and D. Hanyurwimfura in 2018.

[21]. Internet Security Threat Report by Symantec in 2018.

[22]. Cloud-based cyber-physical intrusion detection for vehicles using deep learning by G. Loukas, T. Vuong, R. Heartfield, G. Sakellari, Y. Yoon, and D. Gan in 2017.

[23]. Investigation of possibilities to detect malware using existing tools by O. Aslan and R. Samet in 2017.

[24]. Intelligent Phishing URL detection using association rule mining by S. C. Jeeva and E. B. Rajsingh in 2016.

[25]. A novel intrusion detection method using deep neural network for in-vehicle network security by M. J. Kang and J. W. Kang in 2016.