ROLE OF DEEP LEARNING IN MALWARE DETECTION

Abstract

Malicious Software (Malware) is a major threat to Computers, networks in the Cyber world. Malware removal is essential to prevent identity theft and document theft by cybercriminals**.** Digital attacks include Malware, ransomware, spyware, Denial-of-service attack, Trojan horse, Man-in-the middle attack, Phishing etc. Malware introduction into electronic gadgets rose along with their widespread use. The correct operation of these devices depends on the elimination of malware from them. These malware issues in cyberspace have a viable remedy in Artificial Intelligence (AI). Deep learning (DL) is a subset of Machine learning (ML), which is a branch of artificial intelligence. This Chapter provides a survey on DL techniques for Malware Detection.

Keywords: Deep Learning, Malware, Machine learning.

Authors

Meenakshi Bhrugubanda
Assistant Professor
Information Technology
Mahatma Gandhi Institute of Technology
bmeenakshi_it@mgit.ac.in

Siva Rama Krishna T
Assistant Professor
Computer Science and Engineering
Jawaharlal Nehru Technological University Kakinada.
srktummalapalli@gmail.com

I. INTRODUCTION

The Internet is a network of networks. The Internet has contributed to the creation of a global community. So, people across the world can share their records, do web shopping, electronic banking, web trading and play game, etc. Because of this far and wide utilization of web certain individuals are utilizing mischievously, for example, bringing malware into the electronic contraptions. Malicious software is referred to as Malware. Cybercriminals employ malicious software to sabotage computers, servers, networks, and other devices to steal data. Examples of Malware include viruses, worms, Trojan viruses, spyware, adware and ransomware, Adware.

1. Virus

A virus is a type of malicious software that tampers with data and software as it spreads from one computer to another. It is capable of stealing employee data. Computers, networks, and other tech-enabled devices can all be impacted by viruses. Computer systems can become infected with viruses through USB devices, file sharing networks, infected websites, and email attachments. ILOVEYOU, Creeper, Elk Cloner are some examples of computer virus [1]

2. Trojan

Any software that poses as a legitimate program in order to deceive users about its true purpose is known as a Trojan horse. Usually, some kind of social engineering is used to propagate it. A few well-known instances of Trojan horse-based real-world cyberattacks are Emotet, Zeus, and Tinba[2].

3. Spyware

Software known as spyware collects information about people or businesses and gives it to cyber criminals. It has the power to ruin people and organizations' privacy and secrecy. Spyware's primary goal is to steal data from individuals or organizations, including credit card numbers, banking records, login passwords, browsing history, e-mail addresses, personnel identification numbers etc.[3][4].

4. Logic Bomb

Logical bomb is a type of malware that activates when specific parameters are satisfied. A certain date, event, etc., could be the condition. Certain circumstances, such pressing a button or key, etc.

5. Worm

One type of malicious software that spreads to other computers and networks by self-replicating is called a worm [5]. Code RedSQL Slammer, Blaster, Witty, Beagle, NetSky, My Doom are some examples of computer worms [6].

6. Rootkit

Malware that grants hackers access to the target device's administrative features is known as a rootkit. Software finds it challenging to recognize rootkits [7]. It stays in the system for quite a while. A rootkit could be made up of one tool or multiple tools working together. Some Rootkit examples are FuTo Rootkit, Mebroot Rootkit, NT RootKit, and Stuxnet Rootkit.

7. Ransomware

Ransomware is one of the largest risks in today's cyberspace. The cybercriminal encrypts the user's or organization's files in order to prevent access [8]. After he pays the ransom, the victim is able to access their files. Examples of ransomware include WannaCry, CryptoLocker, Petya, and Bad Rabbit and others.

8. Adware

The acronym for advertisement-supported software is Adware. Pop-up or pop-under windows will be used to display adverts. These adverts might include harmful software. Adware software is used to spy on users, install malware, steal user information, and other things [9].

Malicious software elimination from the system is the sole goal of malware detection. It's possible that the malware detector and the malware it's trying to find are on the same computer. Every now and again, it also resides on a different system and searches for peculiar system behavior [10]. Malware removal is essential to protect the organization and system from dangerous malicious software. Malware detection includes traditional methods and AI based methods. Datamining Techniques are also employed to detect malware. ML and DL methods can assist us with distinguishing malware in these electronic contraptions.

II. TRADITIONAL METHODS FOR MALWARE DETECTION

1. Signature-Based Detection

The methods that rely on signatures are based on recognized malware signatures. The four elements of signature-based methods are gathering, tracking, identifying, and screening. Using a database of known signatures, developers employ the signature-based method to scan a required file, compare its contents to the signatures in the database, and identify any viruses if the information matches any known signature [11]. It is capable of efficiently identifying known malware. The harmful software that is being introduced by cybercriminals today is more sophisticated. These new malware cannot be found using the signature-based method. Regular database upgrades can provide a temporary solution because certain viruses can alter the code following a system infection.

2. Checksumming

Cyclic Redundancy Check (CRC) calculation is a step in the Checksumming process. This approach was developed to get around the problem with signature-based detection

producing false positives. Polymorphic harmful advertising are widely used by hackers. The harmful code in these advertising is challenging to find. The virus frequently modifies the body to evade detection using conventional techniques. “Polymorphism is typically achieved by incorporating non-constant keys containing random sets of decryption commands into the main virus code or by modifying the executable virus code. Since a variable code has no signature, alternative techniques must be employed to identify the malicious code” [12].

3. Reduced Masks

The malware detection team can separate the encrypted key and obtain static code by looking at the virus' encrypted code. Later on, the static code identifies the signature or mask.

4. Cryptanalysis

The encrypted viral body is decoded using a series of equations in the cryptanalysis process. This resembles traditional cryptography issues. The keys and decryption algorithm are rebuilt using this way. This technique is then used to the encoded fragment to decode the viral body.

5. Allowlisting

This approach blocks everything that is not on a list of allowed applications that the system keeps up to date [8].

6. Static Analysis

This method analyses the suspicious code or malicious code without having it run on the systemTo determine whether a file is malicious, one can examine file names, hashes, strings such as IP addresses, and file header data. [13].

7. Dynamic Analysis

This method runs and examines the suspicious malware code without harming the system.

8. Honeypot

A virtual trap designed to draw in attackers is a honey pot. It facilitates the security team's identification of vulnerabilities.

Limitations of Traditional Methods

The traditional methods require a great deal of effort and time. Some of the most recent spyware causes them to react less well. Because of this, certain malware samples may even be able to avoid detection and go for long stretches of time without being found. The malware signature database needs to be updated on a regular basis [14].

III. APPLICATIONS OF DEEP LEARNING IN MALWARE DETECTION

Deep Learning: Deep learning uses machine learning methods to teach computers how people naturally learn by doing. A computer model learns to carry out classification tasks directly from images, text, or sound in deep learning. State-of-the-art accuracy can be attained by deep learning models, occasionally surpassing human performance. Neural network architectures with multiple layers and a large quantity of labelled data are used to train models .Applications of deep learning include automated driving, electronics, medical research, aerospace and defense, and industrial automation. The absence of hidden layers in a neural network is what is meant by "deep" in deep learning. Deep networks have numerous hidden layers, as opposed to typical networks that have two or three. Neural networks and large labelled data sets are used in deep learning to train models. When extracting features, human intervention is not required [15].

1. Intrusion Detection Systems

Intrusion Detection Systems (IDS) is a sort of network security that can recognize and sense dangers prior to the loss of services, the granting of unauthorized access, or the loss of data [16**].** Data collection, vectorization, and classification engine make up the three parts of IDS. To gather data from the network, utilize the Data Collection component. The Vectorization component receives the data packets from the Data Collection components. Feature vectors are found here. The classification engine receives the feature vectors as input. Here, it determines if the transformed feature vector satisfies the incursion definition. [17]. Machine Learning algorithms such as Reinforcement learning, KNN, Logistic regression with Genetic Algorithm, Support vector machine and Artificial Neural network are used [18]. But the problems with ML methods are they are creating false alarms. This can be avoided by using Deep Learning techniques such as Convolutional Neural Networks, Autoencoder, RBM, LSTM-RNN, and DBN etc. Deep learning algorithms assist by performing much smarter traffic analysis and providing more precise results.

2. Mobile Malware Detection

Mobile phones are increasing day-by-day in today's modern world. By 2025, there will be 18.22 billion mobile devices worldwide, an increase of 4.2 billion from the amount in 2020[19].These mobile devices help us in communication, healthcare monitoring, financial transactions, data sharing etc. These mobile gadgets are increasingly used for a variety of purposes, making them more vulnerable to intrusions. The intrusion will introduce harmful malware into these gadgets.The most popular deep learning-based models used to identify malicious software in Android applications include Convolutional neural networks (CNN), gated recurrent neural networks, deep neural networks, bi-directional long short- term memory, long-short term memory, and cubic LSTM[20].

3. IoT Malware Detection

Numerous industries, including agriculture, healthcare, transportation, the military, smart home management, energy management, and others, have found extensive uses for the Internet of Things. By 2025, there will be a 30.9 billion rise in internet of things utilization. Notwithstanding their assurances, there are drawbacks as well. The issue of security has

arisen because of the widespread use of it. IoT security has been implemented through the use of machine learning algorithms. The use of deep learning algorithms has become essential due to the expansion in malware databases. According to research, deep neural networks in particular perform well when it comes to feature extraction and feature detection in the study of IoT malware." As they gained knowledge of the intricate characteristics of IoT malware at various abstraction levels, they are producing excellent outcomes. The higher lawyers get more sophisticated features from the lower layers. These characteristics are taken from the issue domain's visual imagery" [21].

4. Malware Detection in Cloud Infrastructure

Cloud computing is one of the more fascinating computer science paradigms. Pay-and-use, resource pooling, on-demand self-service, efficacious security, etc. are some features of cloud computing. It is a distribution technique that provides easy, on-demand network access to a shared resource pool. In this instance, servers, applications, and data are combined and made available on Internet as a service [22]. Infrastructure as a Service (Iaas), Platform as a Service (Paas) and Software as a Service (Saas) are three popular cloud service models. In Infrastructure as a Service, or IaaS, virtualization is essential. Virtualization allows for the simultaneous operation of several operating systems with various configurations on a single physical computer. In cloud computing, virtual machines are managed by hypervisors. It shares memory, storage, and virtual computer resources. A significant risk to cloud services is malware injection in virtual machines [23]. The other virtual machines could be impacted if malicious malware compromises one of them. Malware can be detected in cloud computing using ML approaches. However, mislabeling is an issue when using those techniques. Malware in cloud infrastructure can be found using deep learning method like Convolutional Neural Network (CNN) [24].

5. Malware Detection in Cyber-physical Systems

Cyber-Physical Systems (CPS) are integrations of computation, networking, and physical processes. CPS are used in Green Buildings, Smart Grid, Medical CPS,Intelligent Transportation System, Humanoid Robots, Smart learning environments, Civil Infrastructure monitoring, Aeronautical applications etc. [25].CPS enables communication between products, machines, and people. A major threat to CPS's security is the introduction of malicious software. These risks have the potential to result in monetary loss, process failure, or even the total cessation of industrial and system operations. Combining deep learning with semi-supervised learning offers some relief from this issue [26].

6. Malware Detection in Autonomous Vehicles

Autonomous vehicles can move people from one place to another without the need for human intervention. They use the Global Positioning System, sophisticated and powerful CPUs, sensors, actuators, artificial intelligence, and other technologies to do this[27][28]. They avoid road dangers and traffic bottlenecks as they make their way to their destination. Malware can, however, infect them, giving attackers access to these vehicles. AV's functionality can be jeopardized as a result. The aforementioned security issue is proposed to be addressed using a Blockchain and Deep Learning Framework. Convolutional neural networks are used in the previously mentioned process to generate accurate results [29].

IV. CONCLUSION

Recently, it has become commonplace for electronic devices to be infected with malware. The sophistication of fraudsters has reached a point where malware inserted into the aforementioned devices cannot be detected using signature-based methods or machine learning techniques. Deep learning is capable of detecting these malwares thanks to its advanced feature extraction, training, and testing techniques. In the near future, research will be conducted to investigate each and every deep learning technique that can be used to identify malware.

REFERENCES

[1] Z. Yangchun, Y. Zhao and J. Yang, "New Virus Infection Technology and Its Detection," 2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2020, pp. 388-394, doi: 10.1109/ICSESS49938.2020.9237708.

[2] W. Yu, Y. Yalin and R. Haodan, "Research on the Technology of Trojan Horse Detection," 2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA), Xiangtan, China, 2019, pp. 117-119, doi: 10.1109/ICICTA49267.2019.00032.

[3] K. M. E. N. Mallikarajunan, S. R. Preethi, S. Selvalakshmi and N. Nithish, "Detection of Spyware in Software Using Virtual Environment," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 1138-1142, doi: 10.1109/ICOEI.2019.8862547.

[4] M. A. Sheta, M. Zaki, K. A. E. S. E. Hadad and H. A. M., "Anti-spyware Security Design Patterns," 2016 Sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC), Harbin, China, 2016, pp. 465-470, doi: 10.1109/IMCCC.2016.202.

[5] V. S. Koganti, L. K. Galla and N. Nuthalapati, "Internet worms and its detection," 2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), Kumaracoil, India, 2016, pp. 64-73, doi: 10.1109/ICCICCT.2016.7987920.

[6] K. Simkhada, T. Taleb, Y. Waizumi, A. Jamalipour, N. Kato and Y. Nemoto, "An Efficient Signature-Based Approach for Automatic Detection of Internet Worms over Large-Scale Networks," 2006 IEEE International Conference on Communications, Istanbul, Turkey, 2006, pp. 2364-2369, doi: 10.1109/ICC.2006.255123.

[7] Xiongwei Xie and Weichao Wang, "Rootkit detection on virtual machines through deep information extraction at hypervisor-level," 2013 IEEE Conference on Communications and Network Security (CNS), National Harbor, MD, USA, 2013, pp. 498-503, doi: 10.1109/CNS.2013.6682767.

[8] Ekta and U. Bansal, "A Review on Ransomware Attack," 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), Jalandhar, India, 2021, pp. 221-226, doi: 10.1109/ICSCCC51823.2021.9478148

[9] A. S. Sendi and M. Cheriet, "Cloud Computing: A Risk Assessment Model," 2014 IEEE International Conference on Cloud Engineering, Boston, MA, USA, 2014, pp. 147-152, doi: 10.1109/IC2E.2014.17.https://azure.microsoft.com/en-in/resources/cloud-computing-dictionary/

[10] Y. Supriya, G. Kumar, D. Sowjanya, D. Yadav and D. L. Kameshwari, "Malware Detection Techniques: A Survey," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India, 2020, pp. 25-30, doi: 10.1109/PDGC50313.2020.9315764.

[11] Souri, A., Hosseini, R. A state-of-the-art survey of malware detection approaches using data mining techniques. Hum. Cent. Comput. Inf. Sci. 8, 3 (2018). https://doi.org/10.1186/s13673-018-0125-x

[12] Ö. A. Aslan and R. Samet, "A Comprehensive Review on Malware Detection Approaches," in IEEE Access, vol. 8, pp. 6249-6271, 2020, doi: 10.1109/ACCESS.2019.2963724.

[13] ”Malware Detection” [Online].Available: https:/ Malware Detection: 10 Techniques - CrowdStrike/.[Aceessed:21-Oct-2023]

[14] Yanfang Ye, Tao Li, Donald Adjeroh, and S. Sitharama Iyengar. 2017. A survey on malware detection using data mining techniques. ACM Comput. Surv. 50, 3, Article 41 (June 2017), 40 pages. DOI: http://dx.doi.org/10.1145/3073559

[15] ”What is Deep Learning?” [Online]. Available: https://www.mathworks.com/deep-learning.html/.[Accessed:26-Oct-2023]

[16] H. Dhillon, ‘‘Building effective network security frameworks using deep transfer learning techniques,’’ M.S. thesis, Dept. Comput. Sci., Western Univ., London, ON, Canada, 2021

[17] G. Karatas, O. Demir and O. Koray Sahingoz, "Deep Learning in Intrusion Detection Systems," 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), Ankara, Turkey, 2018, pp. 113-116, doi: 10.1109/IBIGDELFT.2018.8625278.

[18] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi and R. Ahmad, "Machine Learning and Deep Learning Approaches for CyberSecurity: A Review," in IEEE Access, vol. 10, pp. 19572-19585, 2022, doi: 10.1109/ACCESS.2022.3151248.

[19] ”Forecast number of mobile devices worldwide from 2020 to 2025(in billions)* [Online].Available:https://www.statista.com/statistics/245501/multiple-mobile-device-ownership-worldwide/ [Accessed:21-Oct-2023]

[20] Elliot Mbunge, BenhildahMuchemwa, John Batani, NobuhleMbuyisa,”A review of deep learning models to detect malware in Android applications”,Cyber Security and Applications,Vol.1, ,2023,100014,ISSN 2772-9184,

[21] Asam, M., Khan, S.H., Akbar, A. et al. IoT malware detection architecture using a novel channel boosted and squeezed CNN. Sci Rep 12, 15498 (2022). https://doi.org/10.1038/s41598-022-18936-9

[22] A. S. Sendi and M. Cheriet, "Cloud Computing: A Risk Assessment Model," 2014 IEEE International Conference on Cloud Engineering, Boston, MA, USA, 2014, pp. 147-152, doi: 10.1109/IC2E.2014.17.https://azure.microsoft.com/en-in/resources/cloud-computing-dictionary/

[23] N. Gruschka and M. Jensen, "Attack Surfaces: A Taxonomy for Attacks on Cloud Services," 2010 IEEE 3rd International Conference on Cloud Computing, Miami, FL, USA, 2010, pp. 276-279, doi: 10.1109/CLOUD.2010.23.

[24] [12]M. Abdelsalam, R. Krishnan, Y. Huang and R. Sandhu, "Malware Detection in Cloud Infrastructures Using Convolutional Neural Networks," 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), San Francisco, CA, USA, 2018, pp. 162-169, doi: 10.1109/CLOUD.2018.00028.

[25] Bhrugubanda, M. A review on applications of Cyber Physical Systems information. Int. J. Innov. Sci. Eng. Technol. 2015, 728–730.

[26] Sharmeen, Shaila; Huda, Shamsul; Abawajy, Jemal (2019). Identifying malware on cyber physical systems by incorporating semi-supervised approach and deep learning. Deakin University. Conference contribution. https://hdl.handle.net/10536/DRO/DU:30130717

[27] *[17]*. T. Zhang, "Toward Automated Vehicle Teleoperation: Vision, Opportunities, and Challenges," in IEEE Internet of Things Journal, vol. 7, no. 12, pp. 11347-11354, Dec. 2020, doi: 10.1109/JIOT.2020.3028766..

[28] D. Reebadiya, T. Rathod, R. Gupta, S. Tanwar and N. Kumar, "Blockchain-based secure and intelligent sensing for autonomous vehicles activity tracking beyond 5g networks", Peer-to-Peer Networking and Applications, vol. 14, 09 2021.

[29] D. Patel et al., "Deep Learning and Blockchain-based Framework to Detect Malware in Autonomous Vehicles," 2022 International Wireless Communications and Mobile Computing (IWCMC), Dubrovnik, Croatia, 2022, pp. 278-283, doi: 10.1109/IWCMC55113.2022.9824186.

