**Role of Deep Learning in Malware Detection**

**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**

**Abstract:** Cybersecurity is one of the emerging research areas in the domain of computer science. The practise of defending systems, networks, and programmes from Digital attacks is known as Cybersecurity. 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. 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

# 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.

Malware detection is utilized to safeguard the electronic devices and organization’s network from hazardous malicious software. It includes devices and strategies to perceive, stop, notify and respond to malware threats [1][2]. Malware detection includes traditional methods and AI based methods. ML and DL methods can assist us with distinguishing malware in these electronic contraptions.

# Traditional methods for Malware Detection

* 1. **Signature-based detection**

The methods that rely on signatures are based on recognized malware signatures. A database of known malware signatures is kept up to date by antivirus companies. The software is scanned by antivirus software, which then recognizes the signature and matches it to signatures of known malware. If any known signature is present, it either quarantines or deletes the file. This method makes use of string-based executables, file hashes, and the domains and IP addresses it contacts [2] [8]. The harmful software that is being introduced by cybercriminals today is more sophisticated. These new malware cannot be found using the signature-based method.

**2.2Checksumming**

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” [9].

# 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.

# 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.

# Allowlisting

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

# Static analysis

This method analyses the suspicious code or malicious code without having it run on the system.” File names, hashes, strings such as IP addresses, and file header data can all be evaluated to determine whether a file is malicious” [10].

# Dynamic analysis

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

# Honeypot

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

# 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 [20].Examples of Deep Learning applications are Automated Driving, Aerospace and Defense, Medical Research, Industrial automation, Electronics etc. “The term “deep” usually refers to the number of hidden layers in the neural network. [Traditional neural networks (4:37)](https://www.mathworks.com/videos/getting-started-with-neural-networks-using-matlab-1591081815576.html) only contain 2-3 hidden layers, while deep networks can have as many as 150.Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction”[20].

**3.1 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 [3]. IDS monitors and analyses traffic in real-time while operating throughout the enterprise network. Any packet on the network is inspected for signs of compromise, and any threats or anomalies are notified that are found. IDS notifies human security staff for additional action whenever a breach of the configured security policies—such as a port scanner, ransomware, or malware—is detected [5]. Machine Learning algorithms such as Reinforcement learning, KNN, Logistic regression with Genetic Algorithm, Support vector machine and Artificial Neural network are used [4]. But the problem 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.

**3.2 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[7]. 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. Convolutional Neural Networks, gated recurrent neural networks, deep neural networks, bidirectional long short-term memory, long short-term memory (LSTM) and cubic-LSTM are the most prominent deep learning-based malicious software detection models in Android applications [6].

**3.3 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” [11].

**3.4 Malware Detection in Cloud Infrastructure**

“A delivery model for computing resources in which various servers, applications, data, and other resources are integrated and provided as a service over the internet. Resources are often virtualized, and users typically only pay for the services they use” [13]. 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 [14]. 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, mislabelling is an issue when using those techniques. Malware in cloud infrastructure can be found using deep learning method like Convolutional Neural Network (CNN) [12].

3.5  **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 [16]. 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 [15].

**3.6 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[17][18]. 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 [19].

# 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, we will examine every single Deep learning method applied to malware detection.

REFERENCES

[1]. <https://www.techtarget.com/searchsecurity/definition/malware>

[2].<https://www.cynet.com/malware/4-malware-detection-techniques-and-their-use-in-epp-and-edr/#:~:text=Malware%20detection%20involves%20using%20techniques,%2C%20checksumming%2C%20and%20application%20allowlisting>

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

[4]. 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.

[5]. <https://www.spiceworks.com/it-security/network-security/articles/ids-vs-ips/>

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

[7].[https://www.statista.com/statistics/245501/multiple-mobile-device-ownership-worldwide/#:~:text=In%202021%2C%20the%20number%20of,devices%20compared%20to%202020%20levels](https://www.statista.com/statistics/245501/multiple-mobile-device-ownership-worldwide/" \l ":~:text=In%202021%2C%20the%20number%20of,devices%20compared%20to%202020%20levels).

[8]. <https://www.checkpoint.com/cyber-hub/threat-prevention/what-is-malware/malware-detection-techniques-and-technologies/>

[9]. [Malware Detection: 7 Methods and Security Solutions that Use Them (perception-point.io)](https://perception-point.io/guides/malware/malware-detection-7-methods-and-security-solutions-that-use-them/)

[10]. <https://www.crowdstrike.com/cybersecurity-101/malware/malware-detection/>

[11]. 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>

[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.

[13]. <https://azure.microsoft.com/en-in/resources/cloud-computing-dictionary/>

[14] 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.

[15] 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>

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

*[17]. What is an autonomous car?*, [online] Available: <https://www.synopsys.com/automotive/what-is-autonomous-car.html>.

[18].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.

[19]. 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.

[20]. https://www.mathworks.com/discovery/deep-learning.html