IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A COMPREHENSIVE SURVEY ON MALWARE ATTACKS AND MACHINE LEARNING SOLUTIONS ON WIRELESS NETWORKS

[1]STGY SANDHYA, [2]DR.K.SURESH BABU

[1]Research Scholar, [2]Professor

[1] Department of Computer Science and Engineering,

[1]Jawaharlal Nehru Technological University, Hyderabad, India

Abstract: Malware attacks have become a continuous and evolving threat in the digital age, targeting individuals, enterprises, and governments. The ever-changing nature of the malware renders traditional security solutions ineffective for detecting and mitigating sophisticated attacks. Machine learning (ML) techniques have emerged as powerful tools to counter these challenges, enabling adaptive and intelligent detection systems. In many cases, malware attacks make a clear reflection of a common menace in the digital era, the attackers tend to devise very intricate ways for exploiting the vulnerabilities. Current techniques for detection do not avail good fight against advanced evasion such as polymorphism and zero-day attacks. Machine learning (ML) offers achievable strongholds against such threats, yet a number of challenges and the gaps remain. This paper identifies the gaps and existing methodologies while highlighting potential countermeasures on fresh discoveries from various peer-reviewed journal articles. This article presents a comprehensive survey of malware attacks, their classifications, attack vectors, and impacts. We explain the applicability of ML techniques in fighting malware, while discussing their effectiveness, shortcomings, and future research paths.

***Index Terms* -** Malware, attacks, types of malwares, machine learning, deep learning, malicious, benign, algorithms, android malware.

I. INTRODUCTION

Malware is a collective noun used for all software or programs which is more intrusive and developed by cybercriminals, called as hackers, for the purposes of stealing data, damaging, or destroying computers or computer systems or network by unlawful means and ways. It is evident that with Internet connected devices and advanced intrusion mechanisms the malware attacks have grown exponentially over the past few years. Data of bulk quantities has been exfiltrated during the recent malware attacks. Thousands of new malwares originate each day and provide a variety of threats and the use of modern technologies further complicate the threat environment. The AVG report suggests that around 190,000 malware attacks happened per second in 2023. The following are the most common types of malwares among many available.

Table-1: Common types of Malware

SNO	MALWARE	DEFINITION	IMPACT CAUSED
1	Viruses	Malicious code that attaches to legitimate files and spreads when executed.	Corrupts files, spreads across systems, and may slow down performance
2	Worms	Standalone malware that replicates itself to spread to other systems, often via networks.	Consumes bandwidth, crashes systems, and propagates rapidly.
3	Trojans	Disguised as legitimate software but contains harmful code.	Allows attackers to gain control or steal sensitive data.
4	Ransomware	Encrypts user data and demands payment (ransom) for decryption.	Causes financial and data loss; disrupts business continuity.
5	Spyware	Secretly monitors user activities and collects sensitive information.	Steals personal data like login credentials, credit card details, etc.
6	Adware	Displays unwanted advertisements, often bundled with legitimate software.	Annoying but can lead to exposure to other malware.
7	Keyloggers	Records keystrokes to steal sensitive information like passwords.	Compromises user privacy and security.
8	Bots and Botnets	Devices infected and controlled by attackers to perform tasks like DDoS attacks.	Used for large-scale cyberattacks or spam campaigns.
9	Fileless Malware	Operates in memory without leaving traces on the hard disk.	Evades traditional detection mechanisms.
10	Rootkits	Hides the presence of malicious software by altering system operations.	Allows unauthorized access while remaining undetected.

Classification of Malwares: The Malwares are classified into various types based on certain attributes. Based on the type of environments they attack the malwares are also classified into Mobile malwares which targets Android devices, including smartphones, tablets, and Android-based IoT devices, they get delivered via fake applications ,phishing links and infested apk files. IOT malwares that target Internet of Things applications like industrial sensors, smart home applications by exploiting weak passwords, default settings, open ports or insecure firmware. Network malwares the most disruptive that causes communication failures, steals data by attacking the vulnerabilities of the networks, phishing emails or compromised systems or components in the networks. In addition to the below given there are modern advanced malwares that are called Hybrid Malwares which combines features of multiple types like a Trojan that installs Ransomware , a Rorschach that uses hybrid cryptography to encrypt a part of a file , bots initially appearing as Trojans and once executed acts as worms.

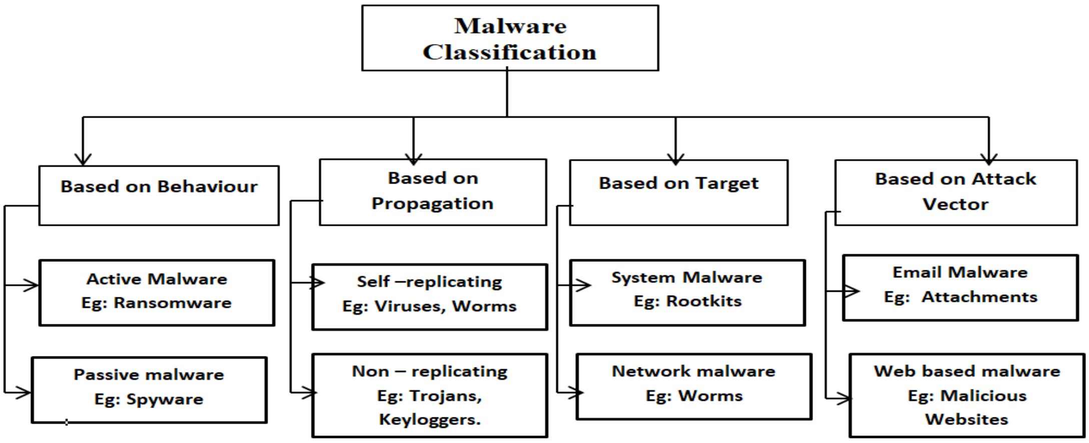
Malware Classification
Based on Behaviour
Based on Propagation
Based on Target
Based on Attack Vector
Active Malware Eg: Ransomware
Passive malware Eg: Spyware
Self –replicating Eg: Viruses, Worms
Non – replicating Eg: Trojans, Keyloggers.
System Malware Eg: Rootkits
Network malware Eg: Worms
Email Malware Eg: Attachments
Web based malware Eg: Malicious Websites

Figure1: Malware classification and examples.

Modern malware attacks comprises of a combination of multiple attacking methods thus making their detection and mitigation a true challenge in real time. Also using modern techniques many of them go undetected through the scan when specific signature based scans are performed. Securing networks from these attacks often become a challenging job as these keep changing their signatures and patterns. This emphasizes on the importance of robust security practices and modern latest and hybrid detection technologies like machine learning.ML is a powerful tool as it allows detection based on patterns rather than signatures. ML models can analyse network traffic, executable files, and user behaviour to predict or detect anomalies.

II. LITERATURE SURVEY

Based on the literature we studied it is evident that Machine learning algorithms can benefit a lot in the process of malware detection and classification. A comparison of various machine learning techniques and the algorithms that come handy in this process are observed as given in the table.

Table – 2: Role of Machine Learning Techniques in Malware Detection

ML Technique	Approach	Algorithm	Applications
Supervised Learning	Trains on labelled datasets of benign and malicious samples	Decision Trees, Support Vector Machines, Random Forests	Signature extraction, binary classification
Unsupervised Learning	Detects anomalies in unlabelled datasets.	K-Means, DBSCAN, Auto-encoders.	Identifying novel malware variants.
Deep Learning	Extracts hierarchical features from raw data.	Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs).	Behavioural analysis, feature extraction.
Reinforcement Learning	Learns optimal actions through trial-and-error	Q-Learning Deep-Q-Network	Dynamic malware response systems.

The process of malware identification, detection and responding to attacks is a very challenging job in the modern day scenario. Detecting and classifying malware is a complex problem that benefits from various types of machine learning algorithms. The efficiency and effectiveness of these algorithms also depends a lot on the type of malware attack, features that are extracted, dataset quality, and operational constraints.

Table-3: Comparative Study of Machine Learning Algorithms used for Malware Detection

Algorithm	Strengths	Weaknesses	Best for	FPR	Efficiency
Random Forest (RF)	Robust to noise, interpretable feature importance.	High memory usage for large datasets.	General-purpose malware detection.	Moderate	Moderate
Support Vector Machines (SVM)	Effective for small datasets with well-defined boundaries.	Struggles with large or high-dimensional datasets.	Polymorphic malware.	Low	Moderate
Naive Bayes	Fast and efficient for text-based feature spaces.	Assumes feature independence, which may not hold.	Adware and spam detection.	High	High
k-Nearest Neighbors (k-NN)	Simple implementation, good for anomaly detection.	High computational cost for large datasets.	Network flow analysis.	High	Low
Gradient Boosting (XGBoost)	High accuracy with well-tuned parameters.	Computationally intensive.	Hybrid malware detection.	Low	Low
Deep Neural Networks (DNN)	Capable of detecting complex patterns in large datasets.	Requires significant computational resources.	Behavior-based attacks.	Low	Low
Convolutional Neural Networks (CNN)	Excels in image-based malware detection.	Limited applicability for non-image data.	Malware using obfuscated code.	Low	Moderate

Machine learning and deep learning techniques are employed to identify malware by examining data to recognize patterns and characteristics that differentiate harmful files from harmless ones. Machine learning algorithms acquire knowledge from the data they process, enabling them to infer the attributes of new samples. For instance, these algorithms can scrutinize the strings derived from files to identify patterns that set apart malicious strings from benign ones. In the realm of malware detection, deep learning emphasizes various learning paradigms, types of features, benchmark datasets, and assessment metrics. A common feature utilized in this context is API/System calls. The study shows a comparison of datasets used, the type of approach followed and key findings in this process of malware detection through machine learning.

Table-4: Comparison of Detection mechanism employed, data sets used and Key findings.

Study/Approach	ML Algorithms Used	Targeted Malware Types	Datasets Used	Key Findings
Dynamic Analysis-Based Detection	Random Forest, Decision Trees	Ransomware, Trojans	CICIDS2017	Achieved high detection rates for network-based attacks. Highlighted the need for handling imbalanced datasets.
Static Feature Analysis	SVM, Logistic Regression	Worms, Rootkits	Ember Dataset	Effective for analyzing file binaries. Struggled with new variants of obfuscated malware.
Behavioral Analysis	Recurrent Neural Networks (RNNs), LSTMs	Adware, Spyware	Own proprietary logs	Demonstrated high accuracy in detecting anomalous behaviour. Computationally expensive.
Hybrid Analysis	Deep Neural Networks (DNNs)	Polymorphic Malware	Malheur Dataset	Combines static and dynamic features. Improved accuracy but required significant computational resources.
Network Flow Analysis	k-NN, Naive Bayes	Distributed Denial of Service (DDoS), Botnets	CICFlowMeter	Network-based methods provided early detection for attacks but faced false positives.
Ensemble Models	Gradient Boosting, XGBoost	General Malware	VirusShare Dataset	Improved robustness by combining classifiers. Highlighted the trade-off between precision and recall.

It is relevant to understand the work done in the field to understand the various methods employed their performance metrics. It allows us to understand the existing work done and provide the future scope and challenges that persisit. We have summarized the work done in the field of malware detection and classification using machine learning techniques. A comparative study is shown based on the work done by various researchers on the type of malware detected, the datasets used and their availability followed by the method used and the key findings of their research**.**

Table-5: Summary of the existing work done in malware detection using machine learning.

Author/s	Type of Malware Operated on	Method used	Dataset	Machine learning Methods Employed	Key findings and metrics
[1] Ananya, A, Aswathy	Android Malware detection	SysDroid: a dynamic machine learning method using system call traces to detect malware.	Android Malware Dataset (AMD)	Logistic Regression, CART, Random Forest, XGBoost and Deep Neural Networks	A new feature selection mechanism SAILS is proposed and used and accuracies ranging between 95 and 99% for dropout rate and learning rate in the range 0.1–0.8 and 0.001–0.2
[2] Aamir M, Iqbal MW	Android Malware classification and detection	AMDDLmodel a deep learning technique	Drebin dataset	convolutional neural network.	Used innovative deep learning for Android malware detection, enhancing accuracy and practical user security through inventive feature engineering.
[3] Nigel Cesario, Daniel Lewis	Ransomware Detection	t-SNE and SVM employed Op-code based detection	A new dataset combining ransomware samples from virusshare and Benign samples from public sources.	Support Vector Machines	use of t-SNE for dimensionality reduction and SVM for classification improvements in detection accuracy, precision, and F1-score compared to baseline models, while demonstrating the importance of feature reduction in optimizing performance
[4] **S.** Poornima, R. Mahalakshmi	Android Malware classification and detection	MAD-NET, an automated hybrid analysis framework for Android Malware Detection.	A new data private dataset is created from CICAndmal2017	K-Nearest Neighbor (K-NN), Decision Tree (DT), machine support vectors (SVM), Random Forests	A Deep Belief Network (DBN) is employed and with the new method has accuracy of 99.83%, and 8.6% recall.

				(RF), Nave Bayes (NB), and logistic regression (LR)	
[5] Al-Fawa'reh, M., Abu-Khalaf	Malware Botnet detection in IoT networks.	MalBoT-DRL, a robust malware botnet detector	Two representative datasets, MedBIoT and N-BaIoT,	Deep reinforcement learning	Uses damped incremental statistics method exceptional average detection rates of 99.80% and 99.40% in the early and late detection phases
[6] Md. Alamgir Hossain Md. Saiful Islam	Botnet detection	strategy to enhance botnet identification by hybrid feature selection and ensemble-based machine learning approach	botnet identification N-BaIoT dataset CTU-13 dataset	Ensemble methods such as Extra Trees Classifier, XGB Classifier, Random Forest Classifiers.	Uses a combination of of Hybrid Feature Selection methods—Categorical Analysis, Mutual Information, and Principal Component Analysis and achieves over 99% True Positive Rates and an unprecedented False Positive Rate close to 0.00%, thereby setting a new precedent for reliability in botnet detection.
[7] M. Alani. Ali Ismail	Mobile Adware detection.	A Machine learning-based system - AdStop	CIC-AAGM2017	Random forest Decision tree Gaussian naïve Bayes	The method had an accuracy of 98.02% with a false positive rate of 2% and a false negative rate of 1.9%.
[8] Khalid O Ullah	Fileless malware attacks.	Analyze memory dumps and using machine learning models and classifiers for testing the activity of fileless malwares.	A new dataset is created containing PowerShell, WMI, Macros, and VB scripts for attacks.	Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM)	Random Forest achieved the highest accuracy of 93.3% with a TPR of 87.5% at an FPR of 0% on the unseen test set.
[9] Ala, Mughaid Ruba	Trojan Horse detection in IoT	A novel Mmachine learning approach using an extensive dataset	Trojan detection dataset consists of 177,482 instances across 86 attributes used.	Naive Bayes (NB), J48 trees, and IBk	up to 99.63% accuracy using Naive Bayes on a large IoTspecific dataset,
[10] Mostafa Dorrah Asmaa	Advanced Network Malware Detection	A new approach using machine learning and PySpark for effective analysis of network traffic patterns.	Network malware data set from kaggle.	Logistic Regression, Random Forest, Decision Tree, and Naive Bayes.	A high recall rate is achieved by minimizing false negatives and improving true positive rates.
[11] Chinchalkar Somkunwar	Keylogger Detection	A new system that combines Dendritic Cell Algorithms (DCA) and Machine Learning Algorithms (MLA) is implemented.	A new data set is constructed of typing speeds at an average typing speed of 60 as the first experimental parameter.	SVM, NavieBayes	The system achieved a keylogger detection accuracy is 99.8 while enhancing process effectiveness and system security.
[12] Vinaya kumar M. Alazab	Malware Detection	A novel ScaleMalNet system using deep learning is implemented.	2 Datasets Ember and MalConv for ML and 2 new Datasets	Deep neural network (DNN) Convolutional neural network (CNN)	The designed highly scalable framework called ScaleMalNet detected, classify and categorize zero-day malwares.

			collected using VirtualBox5		

III CHALLENGES AND LIMITATIONS

The key challenges in malware detection include the unavailability of up to date datasets, the obfuscation nature of malwares to morph to new variants bypassing traditional signature based detections and scans, scalability of the approaches to real time data, Identification of apt features and need for some modern and hybrid feature selection methods and some cases lack of readiness to face any kind of Zero day attacks.

IV FUTURE DIRECTTIONS

The future trends show great scope of improvement in some areas as daily thousands of new malwares arises requiring modern robust methods to identify and mitigate them. Using edge computing to reduce bandwidth by employing lightweight ML models. Develop machine learning models that adapt new threats automatically. Using explainable AI and threat intelligence to mitigate malwares. There is also a need for and engaging in adversarial machine learning methods.

V CONCLUSUONS

Malware attacks remain a significant threat in the digital landscape, with evolving techniques challenging traditional defences. Machine learning offers a promising solution, enabling adaptive and scalable security systems. This survey highlights the potential and challenges of ML in malware detection and outlines future research directions to enhance cybersecurity resilience. This study points to the gaps in the current approaches of malware detection through. Such a framework proposes to develop resilient and efficient detection systems with current malware threats reducing capabilities through addressing dataset biases, improving feature extraction methods, and engaging in adversarial training.

REFERENCES

[1] Ananya, A., Aswathy, A., Amal, T.R. et al. SysDroid: a dynamic ML-based android malware analyzer using system call traces. Cluster Comput 23, 2789–2808 (2020).

[2] Aamir M, Iqbal MW, Nosheen M, Ashraf MU, Shaf A, Almarhabi KA, Alghamdi AM, Bahaddad AA. AMDDLmodel: Android smartphones malware detection using deep learning model. PLoS One. 2024 Jan 19;19(1):e0296722..

[3] Nigel Cesario, Daniel Lewis, Carmine Rosales, et al. Ransomware Detection Using Opcode Sequences and Machine Learning: A Novel Approach with t-SNE and Support Vector Machines. TechRxiv. October 22, 2024.

[4] S. Poornima, R. Mahalakshmi.2024.Automated malware detection using machine learning and deep learning approaches for android applications,Measurement: Sensors,Volume 32,100955.

[5] Al-Fawa'reh, M., Abu-Khalaf, J., Szewczyk, P., & Kang, J. J. (2023). MalBoT-DRL: Malware botnet detection using deep reinforcement learning in IoT networks. IEEE Internet of Things Journal, 11(6), 9610-9629.

[6] Hossain, Md. Alamgir & Islam, Md. (2023). A novel hybrid feature selection and ensemble-based machine learning approach for botnet detection. Scientific Reports. 13. 1-28. 10.1038/s41598-023-48230-1.

[7] Mohammed M. Alani. Ali Ismail Awad. 2022.. AdStop: Efficient flow-based mobile adware detection using machine learning,Computers & Security,Volume 117.

[8] Khalid, O., Ullah, S., Ahmad, T., Saeed, S., Alabbad, D. A., Aslam, M., Buriro, A., & Ahmad, R. (2023). An Insight into the Machine-Learning-Based Fileless Malware Detection. Sensors, 23(2), 612.

[9] Ala, Mughaid., Ruba, Ibrahim., Mahmoud, AlJamal., Issa, Al-Aiash. (2024). Detection of Trojan Horse in the Internet of Things: Comparative Evaluation of Machine Learning Approaches. 35-41.

[10] Mostafa, Dorrah., Asmaa, ElMaghraby., Abdallah, ElSaadany., Mohamed, Atta., Ahmed, Ashraf., Yousef, Adel., Tawfik, Yasser., Mostafa, Fathi., Ibrahim, Abdelbaky. (2024). Advanced Network Malware Detection: Integrating PySpark and Machine Learning Techniques. 131-136.

[11] Chinchalkar, S.P., Somkunwar, R.K. (2024). An innovative keylogger detection system using machine learning algorithms and dendritic cell algorithm. Revue d'Intelligence Artificielle, Vol. 38, No. 1, pp. 269-275.

[12] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran and S. Venkatraman.2019.Robust Intelligent Malware Detection Using Deep Learning," in IEEE Access, vol. 7, pp. 46717-46738,.