IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 10, Issue 10, October 2023

DOI: 10.17148/IARJSET.2023.101018

Literature Survey on Android Malware Detection Through ML - Based Analysis

Sairaj Paygude[1], Sonal Sonawane[2], Siddham Tatiya[3], Nakul Sarda[4], Ms. A. Dirgule[5]

Student, Computer Engineering, Sinhgad College of Engineering, Pune, India[1-4]

Assistant Professor, Computer Engineering, Sinhgad College of Engineering, Pune, India[5]

Abstract: The surge in Android-based devices has led to an alarming increase in the spread of malware through mobile applications. This literature survey delves into the realm of Android malware detection, focusing on Machine Learning (ML) approaches with a specific emphasis on the analysis of Android Package (APK) permissions. The study explores the existing body of research that employs ML techniques to scrutinize the permissions requested by Android apps during installation. These permissions are pivotal indicators of an app's behaviour and potential security risks. The survey examines various ML algorithms utilized in the detection process, such as Support Vector Machines (SVM), Naive Bayes, Decision Trees, and K-Nearest Neighbors. Additionally, it reviews the methodologies employed in feature extraction from APK files, including static and dynamic attributes, API calls, and system calls. The survey critically evaluates the performance metrics used in assessing the efficacy of ML-based models, such as accuracy, precision, recall, and F1-score. By consolidating insights from diverse studies, this literature survey provides a comprehensive overview of the state-of-the-art techniques in Android malware detection, fostering a deeper understanding of the challenges and opportunities in securing the Android ecosystem.

Keywords: Machine learning, Android malware, Malware detection, APK, Permissions.

I. INTRODUCTION

The proliferation of Android devices has led to an unprecedented growth in mobile applications, accompanied by a parallel rise in the threat of Android malware. Malicious actors exploit the open nature of the Android ecosystem to deploy a variety of malware, compromising user privacy and security. As a result, there is a critical need for robust and adaptive methods to detect and mitigate these threats. This literature survey delves into the domain of Android malware detection, specifically focusing on leveraging Machine Learning (ML) techniques for the analysis of Android Package (APK) permissions. The permissions requested during installation serve as crucial indicators of potential security risks, and analyzing them through ML approaches offers a proactive means of identifying malicious behaviour, providing a dynamic defense against evolving malware tactics. The survey systematically reviews existing research to identify diverse ML models and methodologies employed in APK permissions analysis, considering algorithms such as Support Vector Machines, Decision Trees, and K-Nearest Neighbors. Additionally, the survey investigates intricacies of feature extraction from APK files, encompassing static and dynamic attributes, API calls, and system calls. Performance metrics like accuracy, precision, recall, and F1-score are examined to evaluate the efficacy of ML-based models, contributing to a deeper understanding of state-of-the-art techniques in Android malware detection and informing future research for effective strategies to safeguard mobile devices.

II. LITERATURE REVIEW

[1] The examination presents a permissions-based malware detection system (PerDRaML), a framework for distinguishing versatile malware in Android applications. It centers around authorization based investigation, utilizing AI to distinguish malignant way of behaving. By examining 10,000 applications, it pinpoints key highlights like authorizations, small sizes, and rates to group applications as harmless or malevolent. The strategy accomplishes high location exactness (up to 89.96%) with models like SVM and Irregular Woods, fundamentally enhancing existing methods. It features the potential for a savvy and proficient arrangement in Android malware discovery. This work expands upon earlier investigations, especially Zhu et al.'s approach, underscoring the significance of insignificant however essential authorization sets for keeping up with high exactness. Future endeavors expect to upgrade datasets and investigate more classifiers to additional improve identification abilities.

[2] The research paper introduces a novel approach for detecting malware in Android applications, addressing the limitations of traditional methods. Traditional detection methods, reliant on signatures or battery consumption

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 10, Issue 10, October 2023

DOI: 10.17148/IARJSET.2023.101018

monitoring, are increasingly proving insufficient against sophisticated malware. In response, the study introduces a groundbreaking solution utilizing a Gated Recurrent Unit (GRU), a form of Recurrent Neural Network (RNN), to analyze static features like Application Programming Interface (API) calls and Permissions in Android applications. The comparative analysis between traditional machine learning and deep learning approaches, conducted on the CICAndMal2017 dataset, highlights the superiority of the deep learning model. The proposed classifier, leveraging the GRU architecture, achieves an impressive 98.2% accuracy and a 98.0% F-measure, surpassing conventional methods. This research significantly advances the field of Android malware detection by emphasizing the efficacy of deep learning techniques in addressing the evolving nature of cyber threats on the Android platform.

[3] The research paper addresses the pressing issue of smartphone vulnerability to cyber-attacks through malware applications, posing a threat to data security and personal privacy. By proposing a machine learning approach for Android malware detection, the study focuses on utilizing various static features extracted from Android Application Package (APK) files, including permissions, API calls, services, opcodes, and activities. Notably, the comparative analysis of machine learning models reveals promising results, with Gaussian Process exhibiting the highest efficacy, followed by Random Forest and Decision Trees. While existing literature underscores the significance of mobile cybersecurity and the application of machine learning in threat detection, this study distinguishes itself by advocating for a multi-faceted static feature-based approach. Future research directions, including the integration of dynamic features and exploration of advanced machine learning models, align with the evolving landscape of mobile cybersecurity, presenting opportunities for further refinement and innovation in the field.

[4] The paper introduces a novel approach to malware detection by creating a dataset comprising 16,300 records and 215 features gathered from various malware projects. Employing supervised machine learning classifiers, feature reduction, and ensembling techniques, the study evaluates the dataset's performance metrics such as Accuracy, AUC, FPR, TPR, Precision, and Cohen Kappa Score. Notably, the CatBoost Classifier outperforms others with 93.15% Accuracy, a ROC value of 0.91, and an 81.56% Cohen Kappa Score, indicating its exceptional predictive capability and nearly perfect agreement between classifications and true classes. The study underscores the efficacy of employing machine learning methods for malware detection over conventional approaches, emphasizing the significance of a comprehensive dataset in training models for improved performance.

[5] This paper highlights the escalating threat of Android malware by proposing a fully connected deep learning model for pre-installation detection. With a focus on achieving a high accuracy of 94.65%, the model utilizes a unique humanoid malware detection framework that employs permissions to mirror application behavior. By extracting diverse permissions from various APK files and leveraging deep learning techniques, the model successfully distinguishes between malware and benign applications, reaching an accuracy rate of 94.64%. The paper also contributes insights into the features required by malware and benign applications during and after execution, providing a valuable foundation for future research in Android malware detection.

[6] The research paper delves into the critical realm of Android malware by conducting a comprehensive survey focused on the detection, identification, and categorization of malware families. By reviewing forty research papers, the study emphasizes the dearth of attention given to characterizing malware families despite the escalating volume of malicious applications. The paper categorizes existing literature based on analysis types, features, and methodologies, shedding light on prevalent techniques such as machine learning, graph mining, and statistical analysis. It identifies challenges such as limited benchmark datasets, sample scarcity, and the absence of standardized family naming as significant hurdles in this domain. Proposing future directions, the paper advocates for leveraging advanced artificial intelligence, big data technologies, automated detection, and crowdsourcing to fortify efforts in tackling Android malware family identification.

[7] The study reviews existing detection and analysis methods for Android malicious code and identifies the persistent issue of low accuracy in detecting new malware. In response, the paper proposes the integration of machine learning algorithms and semantic analysis for more effective malware analysis. The proposed system focuses on permission and semantic analysis, utilizing a dataset of permissions for malicious applications for comparison with the permissions extracted from the target application. The system enables users to gauge the extent of malicious permissions within an application and conducts additional analysis through comments. While the proposed system holds promise for applications in security systems and malware detection software, the paper acknowledges limitations related to varying user perspectives on permissions.

[8] The study reviews existing detection and analysis methods for Android malicious code and identifies the persistent issue of low accuracy in detecting new malware. In response, the paper proposes the integration of machine learning algorithms and semantic analysis for more effective malware analysis. The proposed system focuses on permission and

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 10, Issue 10, October 2023

DOI: 10.17148/IARJSET.2023.101018

semantic analysis, utilizing a dataset of permissions for malicious applications for comparison with the permissions extracted from the target application. The system enables users to gauge the extent of malicious permissions within an application and conducts additional analysis through comments. While the proposed system holds promise for applications in security systems and malware detection software, the paper acknowledges limitations related to varying user perspectives on permissions.

[9] This paper presents an innovative solution for addressing the growing threat of Android malware through an automatic detection method using the text semantics of network traffic. Each HTTP flow generated by mobile apps is treated as a text document, allowing natural language processing to extract text-level features. The resulting text semantic features form the basis of a robust malware detection model, outperforming existing approaches with a remarkable 99.15% accuracy in distinguishing between benign and malicious flows. The method proves effective in detecting newly discovered malware, requiring minimal samples for accurate results. Real-world application demonstrates its superiority over popular antivirus scanners, achieving a 54.81% detection rate for malicious apps. The proposed detection system for encrypted traffic in bring-your-own-device enterprise networks, home networks, and 3G/4G mobile networks further showcases the versatility and practical applicability of the proposed malware detection approach, providing a comprehensive solution to the evolving challenges posed by Android malware.

[10] The paper provides an extensive survey of malware detection techniques in the Android environment, examining their strengths and limitations. It highlights the escalating security threats within the Android ecosystem despite the numerous existing detection methods. The conclusion emphasizes persisting challenges such as code obfuscation, absence of source code, and the emerging issue of malware collusion, urging researchers to focus on these concerns. Additionally, the paper offers 15 recommendations to enhance malware detection tools, emphasizing the necessity of standardized testing datasets and advocating for a fusion of static and dynamic analysis for more robust security measures in Android applications. This comprehensive review serves as a roadmap for future research endeavors to combat evolving Android malware effectively.

III. CONCLUSION

In conclusion, the literature survey on Android malware detection through ML-based analysis of APK permissions reveals a landscape marked by continuous advancements and challenges. The pervasive threat of malware on Android devices necessitates robust and adaptive detection mechanisms. ML, particularly through the analysis of APK permissions, emerges as a promising avenue for addressing this concern.

The surveyed literature demonstrate a diverse array of ML techniques applied to the analysis of Android app permissions, showcasing the versatility of models such as Support Vector Machines, Decision Trees, Naive Bayes and K-Nearest Neighbors.

The incorporation of static and dynamic attributes, API calls, and system calls for feature extraction underscores the multidimensional nature of the detection process. Researchers are actively exploring innovative methodologies to enhance the accuracy and efficiency of Android malware detection. Despite these advancements, challenges persist. The dynamic nature of malware and the evolving tactics employed by malicious actors necessitate ongoing research and adaptability in detection strategies. Moreover, the need for large and diverse datasets, ethical considerations in data collection, and the interpretability of ML models remain subjects of active investigation.

As the survey consolidates findings from various studies, it provides a valuable synthesis of current knowledge, laying the groundwork for future research directions. The identified trends and challenges underscore the importance of a collaborative and interdisciplinary approach involving researchers, industry experts, and policymakers. By fostering a deeper understanding of the nuances in Android malware detection, this survey contributes to the ongoing efforts to secure the Android ecosystem and protect users from evolving cybersecurity threats.

REFERENCES

[1]. Akbar, F.; Hussain, M.; Mumtaz, R.; Riaz, Q.; Wahab, A.W.A.; Jung, K.-H. “Permissions Based Detection of Android Malware Using Machine Learning’, Symmetry 2022.
[2]. Omar N. Elayan∗ , Ahmad M. Mustafa, “Android Malware Detection Using Deep Learning”, The 2nd International Workshop on Data-Driven Security (DDSW 2021) March 23 - 26, 2021.
[3]. Ali Al , Djedjiga Mouheb, “Android Malware Detection Using Static Features And Machine Learning”, IEEE 2020.
[4]. Prerna Agrawal, Bhushan Trivedi, “Evaluating Machine Learning Classifiers to detect Android Malware”, IEEE International Conference for Innovation in Technology (INOCON) Bengaluru, India. Nov 6-8, 2020.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 10, Issue 10, October 2023

DOI: 10.17148/IARJSET.2023.101018

[5]. Sandeep HR, "Static Analysis of Android Malware Detection using Deep Learning", Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS), 2019.
[6]. Fahad Alswaina ,Khaled Elleithy, "Android Malware Family Classification and Analysis: Current Status and Future Directions", Electronics 2020.
[7]. Rishab Agrawal, Vishal Shah, Sonam Chavan , Prof. Ganesh Gourshete, Prof. Nahid Shaikh, "Android Malware Detection Using Machine Learning", International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020.
[8]. Ahmed Sabbah Birzeit, Adel Taweel Birzeit, Samer Zein,"Android Malware Detection: A Literature Review", Communications in Computer and Information Science, February 2023.
[9]. Shanshan Wang, Qiben Yan, "Detecting Android Malware Leveraging Text Semantics of Network Flows", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 13, NO. 5, MAY 2018.
[10]. Asma Razgallah, Raphaël Khoury, Sylvain Hallé, Kobra Khanmohammadi, "A survey of malware detection in Android apps: Recommendations and perspectives for future research", Computer Science Review, 2021.