A Scalable Malware Detection Approach through Significant Permission Identification for Android Devices

IJIMSR, Vol. 2, No. 1, (2024) 24.2.1.013

***Mr. K. Praveen Kumar**
Assistant Professor , Department of Computer Science & Engineering,
Sri Vasavi Engineering College (Autonomous),Pedatadepalli, Tadepalligudem, AP-534101.
Email: praveenkumar.cse@srivasaviengg.ac.in

Dr. D. Jaya Kumari
Professor & HOD, Department of Computer Science & Engineering,
Sri Vasavi Engineering College (Autonomous),Pedatadepalli, Tadepalligudem, AP-534101.
Email: hod_cse@srivasaviengg.ac.in

Sowmya Sree Karri
Senior Java Developer, Spire Software Solutions Ltd, Visakhapatnam, Andhra Pradesh.
Email: sowmyasree.dannina@gmail.com

***Corresponding author**

Received 31st January 2024; Accepted 19th April 2024
KEYWORDS: Smartphones, Malicious Software, Android, Machine Learning, Scalable Detection, Mobile security

A Scalable Malware Detection Approach through Significant Permission Identification for Android Devices

Mr. K. Praveen Kumar
Assistant Professor , Department of Computer Science & Engineering,
Sri Vasavi Engineering College (Autonomous),
Pedatadepalli, Tadepalligudem, AP-534101.
Email: praveenkumar.cse@srivasaviengg.ac.in

Dr. D. Jaya Kumari
Professor & HOD, Department of Computer Science & Engineering,
Sri Vasavi Engineering College (Autonomous),
Pedatadepalli, Tadepalligudem, AP-534101.
Email: hod_cse@srivasaviengg.ac.in

Sowmya Sree Karri
Senior Java Developer, Spire Software Solutions Ltd,
Visakhapatnam, Andhra Pradesh.
Email: sowmyasree.dannina@gmail.com

ABSTRACT

The global ubiquity of smartphones has led to the availability of many free apps for gaming, communication, financial, and educational needs. However, hazardous malicious software targeting smartphones has increased as the global adoption of these devices has grown. Malware is growing rapidly, with reports predicting a new Android app every 10 seconds, threatening the mobile ecosystem. Due to Android's versatility, users can install apps from third-party app shops and file-sharing websites, compounding malware outbreaks. The seriousness of this situation requires scalable malware detection. Based on permission usage analysis, this project introduces Significant Permission Identification (SigPID), a novel malware detection technique. SigPID uses a three-tiered permission pruning mechanism to discover the most important permissions for distinguishing benign from malicious apps, unlike standard methods that scan all Android permissions. The system first uses the Random Forest method for machine learning classifications. The study reduces non-sensitive permissions by identifying benign and harmful permission lists. Support Vector Machine classification, K-Nearest Neighbor, and Linear Regression are then used to a fresh dataset. SigPID, written in Python 3.7, is a powerful and scalable Android malware countermeasure. SigPID uses advanced machine learning and large permissions to protect the mobile ecosystem from harmful apps, making it safer and more secure.

KEYWORDS: Deep Learning, Lung Sound, Spectrogram, Alexnet, Googlenet, Performance Metrics

1. INTRODUCTION

SIGPID relies on the MLDP process to detect important permissions, avoiding the need to check every Android permission. The abundance of apps makes it difficult to analyze their many permissions, which slows analysis. To reduce the impact on malware identification and productivity, the authors devised three data

pruning styles to selectively filter permissions that provide little benefit.The initial pruning approach, "Authorization Ranking (with Negative Rate)," carefully evaluates each permission based on its relevance to app activities. The INTERNET permission shows if an app has internet. Given that benign and malicious apps need different permissions, the hypothesis suggests that malignant apps may have similar requirements. This knowledge allows us to focus on permission sets that cause high-threat attack scenarios and are frequently requested by malware samples.

The authors emphasize the significance of warrants that are rarely requested by malware records, with the belief that these warrants are significant indications for distinguishing between applications that are benign and those that are harmful. The suggested trimming approach identifies warrants that are highly discriminative, which plays an important role in identifying both malicious and benign applications. The method makes use of an authorization rating methodology in order to evaluate warrants according to the fact that they are utilized by both benign and harmful applications. In contrast to earlier tactics, which primarily concentrated on warrants with a high level of threat, the unique strategy, which is known as PRNR, recognizes the significance of warrants with a low level of threat.

M and B are the two matrices that are utilized in the PRNR technique. M is the list of warrants that are utilized by malware records, and B is the list that is utilized by benign records. These matrices can be used to ease the classification of permits as either requested or not requested by certain samples, hence generating a ranking result that is both clear and visible. PRNR takes into consideration the broader context of both the identification of malware and the connectivity of benign applications, in contrast to previous approaches that focused exclusively on warrants as a means of assisting in the detection of malware. A full knowledge of the significance of permissions in discriminating between benign and harmful programs is ensured by this nuanced approach, which ensures that this understanding is comprehensive.

2. RELATED WORK

2.1 Risk Ranker: A Novel Approach for Zero-Day Android Malware Detection

This study examines reactive mobile antiviral software's limitations using known malware samples. They propose a novel method to detect zero-day Android malware, emphasizing the necessity of assessing untrusted app security risks. RiskRanker, a scalable automated algorithm, identifies harmful app activities and prioritizes further research. RiskRanker identified many dangerous apps across varied Android queries, proving its usefulness and scalability.

2.2 Over privilege Detection in Android Operations

This work focuses on using automated testing tools on the Android API to produce an authorization chart for overprivilege detection. Stowaway tests a series of operations and finds overprivilege in one-third of them. The study investigates overprivilege and shows how developers may fail to apply least privilege owing to API attestation.The study emphasizes examining Android activities to ensure developers follow the idea of least privilege. This strategy reduces overprivilege and improves user privacy control. The study emphasizes the importance of strong testing methods and API attestation in the Android ecosystem for secure and privacy-centric app development.

2.3 TaintDroid: Uncovering Implicit Abuse in Third-Party Android Applications

The authors investigate widely-used third-party Android apps using TaintDroid to find implicit exploitation of users' private information. TaintDroid monitors sensitive data, providing consumers and smartphone security providers with important information. The study examines the challenges of centralized third-party app downloaders, emphasizing the need for contextual data usage patterns.TaintDroid's automated categorization of sensitive data helps identify unwanted applications through educated analysis. The study emphasizes the importance of such monitoring tools in improving Android device security by alerting users to potential privacy breaches and enabling preventative measures against misbehaving apps.

2.4 DREBIN: A Static Analysis Approach for Android Malware Detection

DREBIN performs broad static analysis to detect malware patterns in Android operations. Outperforming several related approaches, DREBIN identifies a significant number of malware samples with detailed explanations for each finding. The paper underscores the popularity of Android for third-party operations and the need for effective malware detection to counteract the ease of malware distribution. The study also emphasizes the importance of contextual information for analyzing operations' behavior.

2.5 DroidMiner: Mining Malicious Android App Modalities

DroidMiner introduces a novel static analysis system for mining malicious modalities from a corpus of mobile operations, achieving high detection rates with low false positives. The study emphasizes the importance of analyzing Framework API calls and proposes new ways to capture semantic connections across multiple APIs. Droid Miner's modality-based approach proves valuable for malware discovery, family classification, and behavioral characterization, providing a robust means to identify and categorize suspicious traits within Android operations.

3. METHODOLOGY

Significant Authorization Identification (SIGPID) and the random forest technique are used to extract critical application permissions. This retrieved data is used by supervised learning systems to detect malware. Malware detection must be improved due to the growing number of new malware. The suggested approach prioritizes efficient and direct malware detection based on SIGPID. This addition helps analysts correlate and understand the growing threat landscape.

The suggested system uses multilevel data pruning (MLDP) with PRNR, PMAR, and SPR to achieve this purpose. These methods selectively trim non-significant warrants to improve the system's ability to identify malicious and benign applications.

Proposed system features include:

3.1 Exclusion of SVM Bracket: The system excludes SVM Bracket, preventing probabilistic analysis of benign/suspicious apps in new test data.

3.2 Point Reduction before Malware Identification: Point reduction using unique authorization list values enhances analysis precision.

3.3 Comparison Analysis: SVM and KNN brackets are used to compare the complete and point-reduced authorization lists. The permission landscape is better understood with this comparison study.

Besides SIGPID, the suggested approach prioritizes risky and benign-enabled authorization lists. Systematic point reduction streamlines analysis. Malware detection is strong and comprehensive with SVM and KNN brackets on both the whole authorization lists and the point-reduced dataset. By merging SIGPID with modern techniques like MLDP and SVM-KNN analysis, the proposed system intends to improve malware detection efficiency and accuracy, giving analysts a powerful tool to resist evolving dangerous threats.

4. EXPERIMENT RESULTS AND FINDINGS

In this evolved system, a robust set of features is introduced to enhance the malware detection process, providing a more nuanced and effective approach to discerning between benign and suspicious applications. Key components and improvements include:

4.1 Classification Enrichment:

LR, SVM, and KNN Integration: The system incorporates Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) classifications. This ensures a more comprehensive understanding of the probability of applications being benign or suspicious in new test data. This diverse set of classifiers contributes to a well-rounded analysis.

4.2 Optimized Data Processing:

Point Reduction Strategy: Before malware identification, a careful point reduction approach is done based on unique authorization list values. The analysis is more efficient and accurate with this adjustment.

4.3 Comparative Analysis:

Similarity Checking: A sophisticated feature compares the whole permission list to the feature-reduced permission list. This comparison uses Support Vector Machine classification to reveal overlapping characteristics and improve authorization pattern comprehension.

4.4 Scalability and Performance:

Support Vector Machine Efficiency: The system ensures that Support Vector Machine classification remains robust even when dealing with large dataset sizes. This scalability feature is vital for handling the growing volume of data associated with modern malware landscapes.

By integrating LR, SVM, and KNN classifications, along with a strategic point reduction approach and similarity checking mechanisms, this advanced system provides an enriched set of features for malware detection. The inclusion of Support Vector Machine classification ensures not only accuracy but also scalability, making it well-suited for the challenges posed by large datasets. This comprehensive set of enhancements aims to elevate the system's effectiveness in discerning and mitigating the risks associated with both known and emerging malware threats.

5. CONCLUSION

In conclusion, this study presents a pioneering approach to mobile malware identification by demonstrating the feasibility of reducing permissions count while ensuring high accuracy and effectiveness. The systematic three-level pruning methodology is a key highlight, designed to extract significant permissions and enhance the precision of malware detection. Notably, the comparison between the old and new methods reveals a substantial shift, with the new system analyzing forty-seven permissions for malware apps compared to the twenty-two permissions considered in the old method. This difference stems from

the elimination of non-sensitive permission features, showcasing the system's adaptability and focused approach.

The dynamic adjustment of malware surety by fine-tuning unique percentages in specific permission values adds a layer of flexibility, allowing for tailored security levels based on varying requirements. This adaptive feature contributes to the system's capability to respond dynamically to the evolving landscape of mobile threats.

Future Work:

While this study marks a significant stride in mobile malware detection, several promising directions for future research emerge. The classification investigation conducted in this study is acknowledged as preliminary, opening avenues for more in-depth exploration. Further enhancements can be made by expanding the analysis to incorporate additional permission sets, providing a more comprehensive understanding of the features contributing to effective malware prediction.

The study found that Linear Regression, SVM, and KNN classification forecast better. These methods could be refined and optimized in future research. Mobile dangers evolve, requiring advanced and adaptive detection systems and better categorization methods.This study advances mobile malware identification and sets the foundation for future research. A revised three-level pruning method and sophisticated classification algorithms create the groundwork for mobile security advancements.

REFERENCES

[1] M. Grace, Y. Zhou, Q. Zhang, S. Zou, and X. Jiang, “RiskRanker: Scalable and accurate zero-day android malware detection,” in Proc. 10th Int. Conf. Mobile Syst., Appl., Services, 2012, pp. 281–294.

[2] A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner, “Android permissions demystified,” in Proc. 18th ACM Conf. Comput. Commun. Security, 2011, pp. 627–638.

[3] W. Enck et al., “Taint Droid: An information-flow tracking system for real time privacy monitoring on smartphones,” ACM Trans. Comput. Syst., vol. 32, no. 2, 2014, Art. no. 5.

[4] D. Arp, M. Spreitzenbarth, M. H¨ubner, H. Gascon, K. Rieck, and C. Siemens, “DREBIN: Effective and explainable detection of android malware in your pocket,” presented at Annu. Symp. Netw. Distrib. Syst. Security, 2014.

[5] C. Yang, Z. Xu, G. Gu, V. Yegneswaran, and P. Porras, “DroidMiner: Automated mining and characterization of fine-grained malicious behaviors in android applications,” in Proc. Eur. Symp. Res. Comput. Security, 2014, pp. 163–182.

[6] [Online]. Available: http://www.gartner.com/it/page.jsp?id=1848514.

[7] [Online]. Available: http:// www.android.com/market/.

[8] [Online]. Available: http:// www.amazon.com/mobile-apps/b?ie=UTF8&node=2350149011.

[9] [Online]. Available: http:// www.apple.com/pr/library/2010/ 01/05 appstore.html, January 2010.

[10] [Online]. Available: http:// www.slashgear.com/iphone-spyware-debated-as-app-library-phones-home-1752491/, August 17, 2009.

[11] W. Enck, P. Gilbert, B. gon Chun, L. P. Cox, J. Jung, P. McDaniel, and A. Sheth. "TaintDroid: An information-flow tracking system for realtime privacy monitoringonsmartphones," in Proc. of USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2010, pp. 393–407.

[12] Y. Zhou, Z. Wang, W. Zhou, and X. Jiang. "Hey, you, get off of my market: Detecting malicious apps in official and alternative android markets," in Proc. of Network and Distributed System Security Symposium (NDSS), 2012.

[13] L.-K. Yan and H. Yin. "Droidscope: Seamlessly reconstructing OS and Dalvik semantic views for dynamic android malware analysis," in Proc. of USENIX Security Symposium, 2012.

[14] W. Enck, M. Ongtang, and P. D. McDaniel. "On lightweight mobile phone application certification," in Proc. of ACM Conference on Computer and Communications Security (CCS), 2009, pp. 235–245.

[15] A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner. "Android permissions demystified," in Proc. of ACM Conference on Computer and Communications Security (CCS), 2011, pp. 627–638.

[16] M. Grace, Y. Zhou, Q. Zhang, S. Zou, and X. Jiang. "Riskranker: scalable and accurate zero-day android malware detection," in Proc. of International Conference on Mobile Systems, Applications, and Services (MOBISYS), 2012, pp. 281–294.

[17] V. Rastogi, Y. Chen, and W. Enck. "Appsplayground: Automatic security analysis of smartphone applications," in Proc. ACM Conference on Data and Application Security and Privacy (CODASPY), 2013.

[18] F. Tchakounté and F. Hayata, "Supervised learning based detection of malware on Android," in Mobile Security and Privacy. Amsterdam, The Netherlands: Elsevier, 2017, pp. 101–154.