International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 10, October 2023)

A Survey on Artificial Intelligence Techniques for Android Malware Prediction

Nishant Khare[1], Priyanka Asthana[2]
[1]Research Scholar, [2]Professor, Department of Computer Science and Engineering, LNCT, Bhopal, India
[1]Kharenishant968@gmail.com, [2]asthana.priyanka.1984@gmail.com

***Abstract*-Android overlay is a point that allows one program to draw over other operations by adding a fresh View subcaste on top of the host View. Despite this, bad apps (malware) might make use of this point to target druggies. Previous countermeasures concentrated on limiting the capabilities of overlays at the operating system position while immolating the usability of overlays to combat this trouble; lately, the overlay medium has been mainly streamlined to help a variety of attacks; still, significant adversaries can still circumvent this protection. Malware is still a significant threat to computer security, which is why we need discovery styles that calculate on machine literacy. While these sensors have a lot of implicit, its well- honored that they're susceptible to elusion assaults. The vaticination of Android malware, along with performance advancements, is presented in this exploration.**

***Keywords*— Android, Malware, Artificial Intelligence, Secuiry, Attack, Cyber**

I. INTRODUCTION

The explosion in the number of malware attacks against the Android platform may be attributed to the wide use of the Android operating system in smart phones and other Internet- of- effects bias. Malware is a kind of software that poses a significant threat to the safety of computer systems and the services that are offered by similar systems. For illustration, malware may steal tête-à-tête identifiable information that's kept on mobile bias. Because of this trouble, a mounding ensemble frame called SEDMDroid has been developed to descry malware on Android. To be more specific, it uses arbitrary point subspaces and bootstrapping slice approaches to produce subsets, and also it does star element Analysis (PCA) on each of those subsets. This helps to guarantee that individualities have a wide range of characteristics. Keeping all of the primary factors and using the whole dataset in the training of each base learner is how the correctness of the Multi-Layer Perception model is tested (MLP). The affair of the ensemble members is used to learn the implicit redundant information, and a Support Vector Machine (SVM) is used as the emulsion classifier to give the final vaticination result (1).

Figure 1: Android malware

It's essential to search for vicious software on Android. Authorization brace- grounded discovery systems are veritably promising for use in practical discovery. There are numerous different discovery schemes. Conventional styles, on the other hand, are unfit to contemporaneously satisfy criteria for practical operation in terms of effectiveness, intelligibility, and stability of discovery performance. These conditions are intended to insure that the system can be used effectively. Indeed while the most recent strategy is grounded on distinctions between common pairings of inoffensive programs and vicious software, it isn't stable enough to fulfill the conditions. This is due to the fact that ultramodern malware has a tendency to demand fresh rights in order to act inoffensive operations, which renders the operation of the frequentness useless (10). In recent times, machine literacy (ML) has come a decreasingly popular tool for the discovery of malware across a variety of operating systems, including Android. In order to stay up with the progression of malware, the discovery models frequently need to be retrained on a regular base (for illustration, once a month), using the data that's acquired out in the wild. This, still, opens the door to poisoning attacks, more precisely backdoor assaults, which are designed to baffle the literacy process and give elusion coverts for virulently altered clones of software. To this day, we haven't been suitable to detect any former study that delved this veritably important issue in Android malware discovery (12).

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 10, October 2023)

In recent times, ransomware has surfaced as a significant peril that targets mobile bias, videlicet smart phones. A kind of vicious software known as ransomware locks down a mobile device's operating system and prohibits the proprietor of an infected device from penetrating their data unless a rescue is paid. Attacks using ransomware have rebounded in significant losses for persons and stakeholders each around the world.

Yet, owing to the continually evolving nature of ransomware features, the process of feting them has come much more delicate as ransomware families have mushroomed at an intimidating rate. Since they produce a large number of false cons, traditional malware discovery approaches, similar as statistically grounded preventative styles, are unfit to battle the ever- evolving Ransomware. In point of fact, it's of utmost significance (9) to work on the development of a system that isn't conventional and uses intelligence to cover against ransomware.

Deep Literacy's connection to a variety of tasks has been made possible by the ready vacuity of large data sets and veritably affordable technology. With regard to safety, a number of enterprises have been accepted to move the use of deep literacy from the realm of image recognition or natural language processing into that of contagion discovery. In this exploration, we offer AdMat, a frame that's both simple and effective, with the thing of characterizing Android apps by viewing them as photos (11). App commerce has evolved into a natural and accessible malware distribution route in recent times, despite the fact that they're an essential part of the ultramodern mobile ecosystem. This is because they" advance legality" to dangerous programs. We haven't yet seen an ML- grounded malware discovery result used at request sizes, despite the fact that machine literacy (ML) styles have been intensely delved for automated, robust malware discovery over the last several times. We perform common exploration with T- request, a major Android app store, which provides us with large- scale ground- verity data (8). This allows us to totally probe the issues that are faced in the factual world. Malware designed for Android presents consumers with significant pitfalls, which has led to an increased need for malware discovery. While detecting Android malware in the pall, sequestration leaks and increased communication costs are generally necessary issues. As a result, the on- device Android malware discovery will be the primary emphasis of this study. At the moment, on- device malware sensors are frequently tutored on waiters, and also the knowledge is transferred to mobile bias (e.g., smartphones). In actuality, training that takes place directly on the device is of utmost significance because of the necessity for offline upgrades.

On- device training, still, is delicate to negotiate on mobile bias due to the confined coffers available on mobile bias; this is particularly true for high- complexity malware sensors. We've developed a feather light on- device Android malware discovery that's grounded on the recently described wide learning algorithm (3).

II. Literature Survey

H. Zhu etal., (1) show experimental results on two separate datasets collected by static analysis way to prove the effectiveness of the SEDMDroid. The first one excerpts authorization, sensitive API, covering system event and so on that are extensively used in Android malwares as the features, and SEDMDroid achieves89.07 delicacy in term of thesemulti-level static features. The alternate bone, a public big dataset, excerpts the sensitive data inflow information as the features, and the average delicacy is 94.92. Promising trial results reveal that the proposed system is an effective way to identify Android malware.

D. Li andQ. Li etal., (2) Ensemble literacy generally facilitates countermeasures, while bushwhackers can work this fashion to ameliorate attack effectiveness as well. This motivates us to probe which kind of robustness the ensemble defense or effectiveness the ensemble attack can achieve, particularly when they combat with each other. We therefore propose a new attack approach, named admixture of attacks, by rendering bushwhackers able of multiple generative styles and multiple manipulation sets, to undo a malware illustration without ruining its vicious functionality. This naturally leads to a new externalization of inimical training, which is farther geared to enhancing the ensemble of deep neural networks. We estimate defenses using Android malware sensors against 26 different attacks upon two practical datasets. Experimental results show that the new inimical training significantly enhances the robustness of deep neural networks against a wide range of attacks; ensemble styles promote the robustness when base classifiers are robust enough, and yet ensemble attacks can shirk the enhanced malware sensors effectively, indeed specially downgrading the VirusTotal service.

W. Yuan,Y. Jiang etal.,(3) Our sensor substantially uses one- shot calculation for model training. Hence it can be completely or incrementally trained directly on mobile bias. As far as discovery delicacy is concerned, our sensor outperforms the shallow literacy- grounded models, including support vector machine (SVM) and AdaBoost, and approaches the deep literacy- grounded models multilayer perceptron (MLP) and convolutional neural network (CNN).

Also, our sensor is more robust to inimical exemplifications than the being sensors and its robustness can be further bettered through on- device model retraining. Eventually, its advantages are verified by expansive trials, and its practicality is demonstrated through runtime evaluation on smartphones.

F. Mercaldo andA. Santone etal.,(4) Several ways to overcome the sins of the current hand- grounded discovery approaches espoused by free and marketableanti- malware were proposed by artificial and exploration communities. These ways are substantially supervised machine literacy grounded, taking optimal class balance to induce good prophetic models. In this paper, we propose a system to infer mobile operation meanness by detecting the belonging family, exploiting formal parity checking. We introduce a set of heuristics to reduce the number of mobile operation comparisons and we define a metric reflecting the operation meanness. Real- world trials on 35 Android malware families (ranging from 2010 to 2018) confirm the effectiveness of the proposed system in mobile malware discovery and family identification.

L. Gong,Z. Li etal.,(5) To address these failings, a more realistic approach is to enable early discovery of overlay-grounded malware during the app request review process, so that all the capabilities of overlays can stay unchanged. For this purpose, in this paper we first conduct a large-scale relative study of overlay characteristics in benign and vicious apps, and also apply the Overlay Checker system to automatically descry overlay- grounded malware for one of the world's largest Android app stores. In particular, we've made methodical sweats in point engineering, UI disquisition, emulation armature, and run- time terrain, therefore maintaining high discovery delicacy(97 perfection and 97 recall) and short per- app checkup time(1.7 twinkles) with only two commodity waiters, under an ferocious workload of 10K recently submitted apps per day.

K. Liu etal., (6) presents complements the former reviews by surveying a wider range of aspects of the content. This paper presents a comprehensive check of Android malware discovery approaches grounded on machine literacy. We compactly introduce some background on Android operations, including the Android system armature, security mechanisms, and bracket of Android malware. also, taking machine literacy as the focus, we dissect and epitomize the exploration status from crucial perspectives similar as sample accession, data preprocessing, point selection, machine literacy models, algorithms, and the evaluation of discovery effectiveness.

Eventually, we assess the unborn prospects for exploration into Android malware discovery grounded on machine literacy. This review will help academics gain a full picture of Android malware discovery grounded on machine literacy. It could also serve as a base for posterior experimenters to start new work and help to guide exploration in the field more generally.

A. Alzubaidi etal., (7) In recent times, the global pervasiveness of smart phones has urged the development of millions of free and commercially available operations. These operations allow druggies to perform colorful conditioning, similar as communicating, gaming, and completing fiscal and educational tasks. These generally used bias frequently store sensitive private information and, accordingly, have been decreasingly targeted by dangerous vicious software. This paper focuses on the generalities and pitfalls associated with malware, and reviews current approaches and mechanisms used to descry malware with respect to their methodology, associated datasets, and evaluation criteria.

L. Gong et.al.,(8) Our study illustrates that the key to successfully developing similar systems is multiplex, including point selection and encoding, point engineering and exposure, app analysis speed and efficacity, inventor and stoner engagement, as well as ML model elaboration. Failure in any of the below aspects could lead to the "rustic barrel effect" of the whole system. This composition presents our judicious design choices and first- hand deployment gests in erecting a practical ML- powered malware discovery system. It has been functional at T-request, using a single commodity garçon to check 12K apps every day, and has achieved an overall perfection of98.9 percent and recall of98.1 percent with an average per- app checkup time of 0.9 twinkles.

I. Almomani etal., (9) introduces a new methodology for the discovery of Ransomware that's depending on an evolutionary- grounded machine literacy approach. The double flyspeck mass optimization algorithm is employed for tuning the hyper parameters of the bracket algorithm, as well as performing point selection. The support vector machines (SVM) algorithm is used alongside the synthetic nonage oversampling fashion (SMOTE) for bracket. The employed dataset is collected from colorful sources, which consists of 10,153 Android operations, where 500 of them are Ransomware. The performance of the proposed approach SMOTE- tBPSO- SVM achieved graces over traditional machine learning algorithms by having the loftiest scores in terms of perceptivity, particularity, and g-mean.

H. Kato etal., (10). Propose Android malware discovery grounded on a Composition rate (CR) of authorization dyads. We define the CR as a rate of a authorization brace to all dyads in an app. We concentrate on the fact that the CR tends to be small in malware because of gratuitous warrants. To gain features without using the frequentness, we construct databases about the CR. For each app, we calculate similarity scores grounded on the databases. Eventually, eight scores are fed into machine literacy (ML) grounded classifiers as features. By doing this, stable performance can be achieved. Since our features are just eight- dimensional, the proposed scheme takes lower training time and is compatible with other ML grounded schemes. Likewise, our features can quantitatively offer clear information that helps mortal to understand discovery results. Our scheme is suitable for practical use because all the conditions can be met. By using real datasets, our results show that our scheme can descry malware with over to97.3accuracy.Besides, compared with an being scheme, our scheme can reduce the point confines by about 99 with maintaining similar delicacy on recent datasets.

L.N. Vu andS. Jung," AdMat etal., (11) The novelty of our study lies in the construction of an proximity matrix for each operation. These matrices act as “input images” to the Convolutional Neural Network model, allowing it to learn to separate benign and vicious apps, as well as malware families. During the trial, we set up that AdMat was suitable to acclimatize to a variety of training rates and achieve the average discovery rate of98.26 in different malware datasets. In bracket tasks, it also successfully honored over97.00 of different malware families with limited number of training data.

C. Li et al etal., (12) motivated to study the backdoor attack against Android malware sensors. The backdoor is created and fitted into the model stealthily without access to the training data and actuated when an app with the detector is presented. We demonstrate the proposed attack on four typical malware sensors that have been extensively bandied in academia. Our evaluation shows that the proposed backdoor attack achieves up to 99 elusion rate over 750 malware samples. Also, the below successful attack is realised by a small size of triggers (only four features) and a veritably low data poisoning rate (0.3).

III. CHALLENGES

Android malware vaticination using artificial intelligence (AI) ways faces several challenges. Some of the significant challenges are:-

1. Lack of labeled data one of the significant challenges is the lack of labeled data needed to train machine literacy algorithms. It’s delicate to gain a large dataset of labeled samples for different types of malware.
2. Point engineering rooting useful features from the raw data is essential for erecting accurate machine literacy models. Still, point engineering for Android malware vaticination is challenging due to the dynamic nature of mobile operations and the diversity of Android bias.
3. The number of samples in the malware class is generally much lower than that in the benign class, which can lead to a class imbalance problem. This can affect the delicacy of the machine knowledge models, and can also affect a prejudiced classifier.
4. Attackers can use colorful ways to shirk discovery by AI- grounded malware sensors. For illustration, they can use obfuscation ways or cipher the vicious law to make it delicate to descry.
5. Machine learning algorithms bear significant computational coffers and can put significant outflow on mobile bias, which are generally resource- constrained.
6. Android malware vaticination using AI ways requires access to sensitive stoner data, similar as the list of installed operations and network business. This can raise sequestration enterprises and can lead to legal and ethical issues.

IV. CONCLUSION

Android apps are evolving snappily throughout the mobile ecosystem, but at the same time, a noway - ending flood tide of vicious Android software is also appearing. A multitude of experimenters have examined the issue of detecting malware on Android bias and have proposed several suppositions and approaches, each coming from a unique point of view. The exploration that has been done so far reveals that using machine literacy to identify Android malware is a system that's both successful and promising. Despite this, there are evaluations that have delved a variety of enterprises about Android malware discovery grounded on machine literacy. In the future, develop a vaticination model with a advanced degree of delicacy by making use of an effective machine- literacy bracket strategy

REFERENCES

[1] H. Zhu, Y. Li, R. Li, J. Li, Z. You and H. Song, "SEDMDroid: An Enhanced Stacking Ensemble Framework for Android Malware Detection," in IEEE Transactions on Network Science and Engineering, vol. 8, no. 2, pp. 984-994, 1 April-June 2021, doi: 10.1109/TNSE.2020.2996379.

[2] D. Li and Q. Li, "Adversarial Deep Ensemble: Evasion Attacks and Defenses for Malware Detection," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 3886-3900, 2020, doi: 10.1109/TIFS.2020.3003571.

[3] W. Yuan, Y. Jiang, H. Li and M. Cai, "A Lightweight On-Device Detection Method for Android Malware," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 9, pp. 5600-5611, Sept. 2021, doi: 10.1109/TSMC.2019.2958382.

[4] F. Mercaldo and A. Santone, "Formal Equivalence Checking for Mobile Malware Detection and Family Classification," in IEEE Transactions on Software Engineering, doi: 10.1109/TSE.2021.3067061.

[5] L. Gong, Z. Li, H. Wang, H. Lin, X. Ma and Y. Liu, "Overlay-based Android Malware Detection at Market Scales: Systematically Adapting to the New Technological Landscape," in IEEE Transactions on Mobile Computing, doi:10.1109/TMC.2021.3079433.

[6] K. Liu, S. Xu, G. Xu, M. Zhang, D. Sun and H. Liu, "A Review of Android Malware Detection Approaches Based on Machine Learning," in IEEE Access, vol. 8, pp. 124579-124607, 2020, doi: 10.1109/ACCESS.2020.3006143.

[7] Alzubaidi, "Recent Advances in Android Mobile Malware Detection: A Systematic Literature Review," in IEEE Access, vol. 9,pp. 146318-146349, 2021, doi: 10.1109/ACCESS.2021.3123187.

[8] L. Gong et al., "Systematically Landing Machine Learning onto Market-Scale Mobile Malware Detection," in IEEE Transactions on Parallel and Distributed Systems, vol. 32, no. 7, pp. 1615-1628, 1 July 2021, doi: 10.1109/TPDS.2020.3046092.

[9] Almomani et al., "Android Ransomware Detection Based on a Hybrid Evolutionary Approach in the Context of Highly Imbalanced Data," in IEEE Access, vol. 9, pp. 57674-57691, 2021, doi: 10.1109/ACCESS.2021.3071450.

[10] H. Kato, T. Sasaki and I. Sasase, "Android Malware Detection Based on Composition Ratio of Permission Pairs," in IEEE Access, vol. 9, pp. 130006-130019, 2021, doi: 10.1109/ACCESS.2021.3113711.

[11] L. N. Vu and S. Jung, "AdMat: A CNN-on-Matrix Approach to Android Malware Detection and Classification," in IEEE Access, vol.9,pp.39680-39694, 2021,doi: 10.1109/ACCESS.2021.3063748.

[12] C. Li et al., "Backdoor Attack on Machine Learning Based Android Malware Detectors," in IEEE Transactions on Dependable and Secure Computing, doi: 10.1109/TDSC.2021.3094824.