Research article Available online www.ijsrr.org ISSN: 2279–0543

International Journal of Scientific Research and Reviews

Ranking fraudulent and malware detection for mobile apps

D. Divya lakshmi

*[*1]M.Tech Computer Science, Prist University, Thanjavur.*
[2] Associate Professor, Department of CSE, Prist University, Thanjavur.

ABSTRACT

The use of mobile devices including Tablets, Smart watch, and note books are increasing day by day. Android has the major share in the mobile application market. Android mobile applications become an easy target for the attackers because of its open source environment. Also user's ignorance the process of installing and usage of the apps. To identify fake and malware applications, all the previous methods focused on getting permission from the user and executing that particular mobile application. A malware detection framework that discovers and break traces left behind by fraudulent developers, to detect search rank fraud as well as malware in Google Play. The fraud app is detected by aggregating the three pieces of evidence such as ranking based, co-review based and rating based evidence. Finally aggregating all the activities of front running apps, it can be achieve certain accuracy in classifying benign standard datasets of malware, fraudulent and legitimate apps. In addition to that apply incremental learning approach to characterize a large number of data sets. It combined effectively for all the evidences for fraud detection. To exactly locate the ranking fraud, there is a need to mining the active period's namely leading sessions, of mobile Apps.

KEYWORDS: Mobile applications, Malware, Ranking, Rating, Google Play.

***Corresponding author:**
Mrs. D. Divya lakshmi

M.Tech Computer Science,

Prist University, Thanjavur.

Trichy–Thanjavur Road,,

Vallam,, Cafe road, Tamil Nadu 613403

1. INTRODUCTION

A process in which the interesting patterns and the knowledge are extracted from the large dataset is called as data mining. Many techniques are used to discover this kind of knowledge, mostly extracted from the machine learning and statistics. The Highlighted part of these approaches is to find the accurate knowledge from the discover data. In the data mining the tasks performed is depends on what sort of knowledge someone needs to mine. Data mining technique are the result of a time-consuming process of study and product development. In the data mining the type of task performed are Classification, Clustering, Regression, Dependence Modeling, Prediction Regression, and Association. The value of a previously defined goal attribute based on other attributes is often represented by IF-THEN rules are searched knowledge that is able to calculate. We can say the Dependence modeling as a generalization of classification. The goal of dependence modeling is to discover rules that are able to calculate the attribute value, from the values of calculated attributes. There are more than one goal attribute in dependence modeling. Clustering is the process of partitioning the item set in a set of significant sub-classes.

1. METHODS AND MATERIAL

Literature survey

In paper [1] the static method to detect the malware in mobile applications is proposed by author in this paper. Reverse engineering concept of the source code for the suspicious APK files are used in this system. Structured mapping author builds the structure for the classes is used after that. Finally using data flow concept several patterns for the different type of threats has been created and use them to detect the malware in applications. This method is calculated depending upon the number of threading pattern the effectiveness.

In paper [2] the author proposed a new method to detect malware in mobile applications by examining the runtime behavior of that particular application in the mobile environment. The author proposes that unexpected behavior mobile app can vary from one application to other applications. Also, it varies from the environment of that particular application running on different devices. Using Xposed framework user can change the user and system application without modifying the application package (APK).Depend upon that user can set particular conditions to identify the malware in the mobile applications.

In paper [3] the author proposes some of modern machine learning algorithms to detect malware. From the Google Play these algorithms are applied to the metadata collected. While all of the existing methods for detecting algorithm focused on inherent characteristics of the particular mobile app this gives a direct method to detect the applications. For the setup of the experiments 25k

data are collected from Google Play. Fake applications could not be updated whereas developers update their applications in particular interval of days since its upload of the Google Play. All these works focuses on linear models whereas non-liners models may focus on Future work.

In paper [4] to protect the review spanners or spam reviews is the aim of the author. For the particular protection the spammer is used. By creating the different account to review that account they gave fake reviews to that particular mobile app. A novel based scoring method to detect every single review of the particular product is proposed by the author. The author creates highly suspicious as a subset. The fakeness of the review is calculated by using web-based spammer evaluation software. The result shows the effective to predict the fake reviews after the evolution completed.

In paper [5] the problem of detecting hybrid shilling attacks on rating data is studied by the author. The semi-supervised learning based approach is proposed and can be used for trustworthy product recommendation. A Hybrid Shilling Attack Detector or HySAD for short, to tackle these problems is presented in this paper. Mainly, to select effective detection metrics HySAD introduces MCRelief, and Semi supervised Naive Bayes (SNBλ) to precisely separate Random-Filler model attackers and Average-Filler model attackers from normal users.

In paper [6] for computing a rank aggregation on the basis of matrix completion to avoid noise and incomplete data is studied. Proposed method solves a structured matrix-completion problem over the space of skew-symmetric matrices. The author proves a recovery theorem detailing when proposed approach will work. They also perform a detailed evaluation of an anecdotal study with Netflix ratings and proposed approach with synthetic data. For finding the solutions, they utilized svp solver for matrix completion. The structure of skew-symmetric matrices is combined with Rank aggregation. Latest advances in the theory and algorithm of matrix completion to skew-symmetric matrices is applied by the author. Existing algorithm for matrix completion to handle skew-symmetric data is enhanced by the author.

In paper [7] a survey on Web spam detection is reported by the author, which comprehensively introduces the principles and algorithms in the literature. Indeed, the work of Web ranking spam detection is mainly based on the analysis of ranking principles of search engines, such as Page Rank and query term frequency. Ranking fraud detection for mobile Apps is different by this. They categorize all existing algorithms into three basses they are as follows as the type of information they use: link-based methods, content-based methods, and methods based on non-traditional data such as clicks, user behavior and HTTP sessions. There is a sub categorization of link-based category into five groups based on ideas and principles used: link pruning and reweighting, labels propagation, labels refinement, feature based and graph regularization.

Existing Process

Play Store use the Bouncer system to remove malware. However, out of the 7, 756 Google Play apps we analyzed using Virus Total, 12% (948) were flagged by at least one anti-virus tool and 2% (150) were identified as malware by at least 10 tools. Sarma et al. use risk signals extracted from app permissions, e.g., rare critical permissions (RCP) and rare pairs of critical permissions (RPCP), to train SVM and inform users of the risks vs. benefits tradeoffs of apps. Peng et al. propose a score to measure the risk of apps, based on probabilistic generative models such as Naive Bayes. Yerima et al. also use features extracted from app permissions, API calls and commands extracted from the app executables.

Disadvantages

- Can't detect genuine reviews
- Can't identify fraud users and malware indicators.
- Time taking process with executing app and analysis of code permission methods

Proposed Methodology

It proposes malware detection framework system that effectively detects Play Store fraud and malware. To detect fraud and malware, this paper proposed the incremental learning approach to characterize the dataset. Formulate the notion of review modeling by applying *collaborative filtering*, *relevance feedback* algorithm.Use temporal session of review post times to identify suspicious review spikes received by apps; the application evidence such as rating, ranking and review evidence will be integrated by an unsupervised evidence-aggregation method for evaluating the credibility of leading sessions from mobile Apps. The malware detection framework is scalable and can be extended with other domain generated evidence for ranking fraud detection. When compared to other existing systems this method finds the better mobile app for the end user. Incremental learning approaches effectively characterize all category of app in Play Store. Also based on the review, rating and rank given by the user is also checked. User can review after they download that particular application using their account from app store

Advantages

- Detect fraud ranking in daily App leader board.
- Avoid ranking manipulation.
- Finds the better mobile app for the end user.

- Incremental learning approach effectively characterizes the large amount of app evidence details.
- It provides accurate aggregation when compared to our existing approach.

2. RESULTS AND DISCUSSION

Signature-based detection

Signature-based detection works by scanning the contents of computer files and cross-referencing their contents with the ―code signatures‖ belonging to known viruses. A library of known code signatures is updated and refreshed constantly by the anti-virus software vendor. If a viral signature is detected, the software acts to protect the user's system from damage. Suspected files are typically quarantined and/or encrypted in order to render them inoperable and useless. Clearly there will always be new and emerging viruses with their own unique code signatures. So once again, the anti-virus software vendor works constantly to assess and assimilate new signature based detection data as it becomes available, often in real time so that updates can be pushed out to users immediately and zero-day vulnerabilities can be avoided pattern-marching approach commercial antivirus is an example of signature based malware detection where the scanner scans for a sequence of byte within a program code to identify and report a malicious code. This approach to malware detection adopts a syntactic level of code instructions in order to detect malware by analyzing the code during program compilation. This technique usually covers complete program code and within a short period of time.

Specification-based malware detection

Specification based detection makes use of certain rule set of what is considered as normal in order to decide the maliciousness of the program violating the predefined rule set. Thus programs violating the rule set are considered as malicious program. In specification based malware detection, where a detection algorithm that addresses the deficiency of pattern-matching was developed. This algorithm incorporates instruction semantics to detect malware instances. The approach is highly resilience to common obfuscation techniques. It used template T to describe the malicious behaviors of a malware, which are sequence of instructions represented by variables and symbolic constants. The limitation of this approach is that the attribute of a program cannot be accurately specified. Specification-based detection is the derivate of anomaly based detection.

Behavioral-based Detection

The behavior-based malware detection system is composed of several applications, which together provide the resources and mechanisms needed to detect malware on the Android platform. Each program has its own specific functionality and purpose in the system and the combination of all of them creates the Behavior-Based malware detection system. The Android data mining script sand applications mentioned in are the responsible for collecting data from Android applications and the script running on the server will be the responsible for parsing and storing all collected data. Furthermore, the script will be responsible for creating the system call vectors for the k-means clustering algorithm.

Cloud Based Malware Detection

Google Play applications are scanned for malware. Google uses a service named Bouncer to automatically scan applications on the Google Play Store for malware. As soon as an application is uploaded, Bouncer checks it and compares it to other known malware, Trojans, and spyware. Every application is run in a simulated environment to see if it will behave maliciously on an actual device. The applications behaviors compared to the behavior of previous malicious apps to look for red flags. New developer accounts are particularly scrutinized –this is to prevent repeat offenders from creating new accounts Google Play can remotely uninstall applications: If you've installed an app that is later found to be malicious, Google has the ability to remotely uninstall this application from your phone when it's pulled from Google Play. Google announced an exciting security feature called the "application verification service" to protect against harmful Android applications. As stated in a recent Google+ post by a member of the Google Android team, "Now, with Jelly Bean Android 4.2 devices that have Google Play installed have the option of using Google as an application verifier. We will check for potentially harmful applications no matter where you reinstalling them from"**.**

Conclusion

In this survey paper developed a fraud detection system for mobile Apps. specially first showed that fraud happened in leading sessions and provided a method for mining leading sessions for each App from its past ranking records. An identified that for the detection of the rank ranking, rating, and review based evidence are considered. Moreover proposed an optimization based aggregation method to integrate all the evidence for evaluating the credibility of leading sessions from mobile Apps. A exclusive perspective of this approach is that all the evidence can be modeled by statistical hypothesis tests, thus it is easy to be extended with other evidence from domain knowledge to detect ranking fraud. Finally validate the proposed system with extensive experiments

on real-world App data collected from the Apple's App Store. Experimental result showed the effectiveness of the proposed approach.

Future Enhancement

In the future plan to study more effective fraud evidence and analyze the latent relationship among rating, review, and rankings.

3. REFERENCES

1. Alaa Salman Imad H. Elhajj Ali Chehab Ayman Kayss, IEEE Mobile Malware Exposed.International Conference on Knowledge discovery and data mining, KDD'14 pages 978-983.
2. Y.T. Yu, M.F. Lau, "A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions", Journal of Systems and Software, 2005, in press.
3. N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Data Mining, 2008; 219–230.
4. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting product review spammers using rating behaviors. In Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM ' 2010; 10: 939–948,
5. D. F. Gleich and L.-h. Lim. Rank aggregation via nuclear norm minimization. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ' 2011; 11: 60–68
6. K.Shi and K.Ali. Getjar mobile application recommendations with very sparse datasets. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ' 2012; 12: 204–212
7. J. Oberheide and C. Miller, "Dissecting the Android Bouncer," presented at the SummerCon2012, New York, NY, USA,2012.
8. N. Spirin and J. Han. Survey on web spam detection: principles and algorithms. SIGKDD Explor. Newsl., May 2012; 13 (2):50–64
9. Chia-Mei Chen, Je-Ming Lin, Gu-Hsin Lai, IEEE Detecting Mobile Application Malicious Behaviors Based on Data Flow of Source Code. International Conference on Trustworthy Systems and their Applications 2014; 95-109.
10. Alfonso Munoz, Ignacio Mart ˜ ´ın, Antonio Guzman, Jos ´ e Alberto Hern ´ and ez, IEEE Android malware detection from Google Play meta-data: Selection of important features.2015; 245-251.