Volume 4, Issue 4, April 2019 ISSN (online): 2456-0006

International Journal of Science Technology Management and Research

Available online at: www.ijstmr.com

Survey on Malware Detection in Google Play

Bhamre Archana B.
Department of Computer
PSGVPM. D.N.Patel College of Engineering
Shahada,M.H

Chaudhari Rohini K.
Department of Computer
PSGVPM. D.N.Patel College of Engineering
Shahada,M.H

Bagal Nayana G.
Department of Computer
PSGVPM. D.N.Patel College of Engineering
Shahada,M.H

Sonar Krutika V.
Department of Computer
PSGVPM. D.N.Patel College of Engineering
Shahada,M.H

Sonar Pallavi R.
Department of Computer
PSGVPM. D.N.Patel College of Engineering
Shahada,M.H

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. we introduce airplay, a novel system that discovers and leverages traces left behind by fraudsters, to detect both malware and apps subjected to search rank fraud. Fair Play correlates review activities and uniquely combines detected review relations with linguistic and behavioral signals gleaned from Google Play app data (87K apps, 2.9M reviews, and 2.4M reviewers, collected over half a year), in order to identify suspicious apps.***

Keywords: Ranking, Android market, search rank fraud, malware detection.

I. INTRODUCTION

Unlike existing solutions, we build this work on the observation that fraudulent and malicious behaviors leave behind telltale signs on app markets. We uncover these nefarious acts by picking out such trails. For instance, the high cost of setting up valid Google Play accounts forces fraudsters to reuse their accounts across review writing jobs, making them likely to review more apps in common than regular users. Resource constraints can compel fraudsters to post reviews within short time intervals. Legitimate users affected by malware may report unpleasant experiences in their reviews. Increases in the number of requested permissions from one version to the next, which we will call "permission ramps", may indicate benign to malware (Jekyll-Hyde) transitions. Google play first releases its app in 2008.Sincethat it distributes applications to all the Android users. In Google Play Store, it provides services that user can discover the particular application, purchase those applications and install it on their mobile devices. Since Android is open source environment all the detail about the application users can be easily accessed by the application developers through Google play. In Google play1.8 Million mobile applications are available and that is downloaded by over 25 billion users across the world. This leads to greater chance of installing malware to the applications that could affect user's mobile devices. Google play store uses its own security system known as Bouncer system [6] to remove the malicious apps from its store. However, this method is not effective as testing some apps using virus tools many apps are found as malicious which are not detected by Bouncer system [6]. Fraudulent developers use search ranking algorithm to promote therapy to the top while searching. After downloading mobile applications from Google play users are asked to give the ratings and reviews about that particular downloaded applications. However fraudulent developers give fake ratings and reviews about their application promote their application to the top. There are two typical approaches used for detecting malware in Google Play. Thus are Static and Dynamic. The dynamic approach needs apps to be run-in a secure environment to detect its benign. The static approach is not used as the

need to give all types of attacking early stage itself but that is impossible as everyday attackers find the new way to inject malware on applications.

II. LITERATURE SURVEY

In paper [2] the author proposes some of modern machine learning algorithms to detect malware. For that these algorithms are applied to the metadata collected from the Google Play. 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 the collected 25k data from Google Play. Developers update their applications in particular interval of days whereas fake applications could not be updated since its upload of the Google Play. All of these works focused on only linear models Future work may focus on non-liners models.

In paper [5] the author aims to protect the review spanners or spam reviews. The spammer may target only on the specific protect. After that, they gave fake reviews to that particular mobile app by creating the different account to review that account. The author proposes novel based scoring method to detect every single review of the particular product. The author creates highly suspicious as a subset. By using web-based spammer evaluation software the fakeness of the review is calculated. After the completion of the evaluation, the result shows the effective to predict the fake reviews.

In paper [8] the authors have studied the problem of detecting hybrid shilling attacks on rating data. The proposed approach is based on the semi-supervisedlearning and can be used for trustworthy product recommendation. This paper presents a Hybrid Shilling Attack Detector or HySAD for short, to tackle these problems. In particular, HySAD introduces MCRelief to select effective detection metrics, and Semi supervised Naive Bays (SNBλ) to precisely separate Random-Filler model attackers and Average-Filler model attacker's from normal users.

In paper [4], author proposed novel technique for computing a rank aggregation on the basis of matrix completion to avoid noise and incomplete data. 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 proposed approach with synthetic data and an anecdotal study with Netflix ratings. To find the solutions, they utilized the svp solver for matrix completion. Rank aggregation is combined with the structure of skew-symmetric matrices. Author applied for latest advances in the theory and algorithms of matrix completion to skew-symmetric matrices. Author enhanced existing algorithm for matrix completion to handle skew symmetric data.

III. PROPOSED SYSTEM

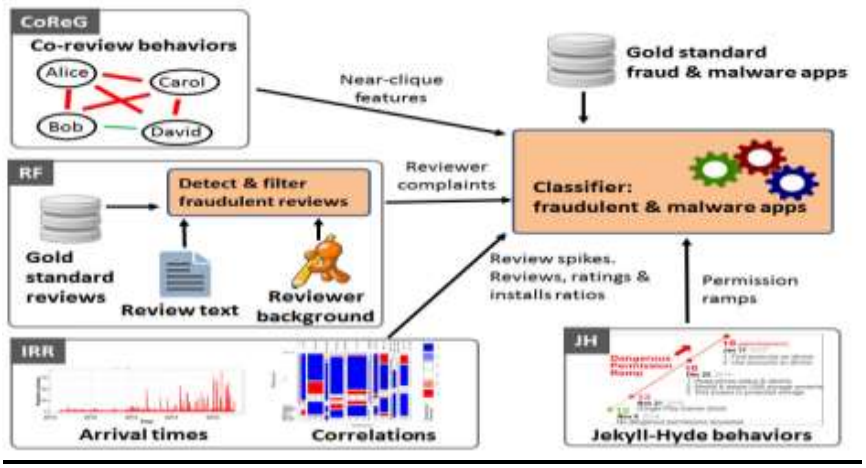

Fig: 1. Architecture Diagram

1) The Co-Review Graph (CoREG) Module:

This module exploits the observation that fraudsters who control many accounts will re-use them across multiple jobs. Its goal is then to detect sub-sets of an app's reviewers that have performed significant common review activities in the past. In the following, we describe the co-review graph concept, formally present the weighted maximal clique enumeration problem, then introduce an efficient heuristic that leverages natural limitations in the behaviors of fraudsters.

2) The Pseudo Clique Finder (PCF) algorithm:
We propose PCF (Pseudo Clique Finder), an algorithm that exploits the observation that fraudsters hired to review an app are like to post those reviews within relatively short time intervals (e.g., days). PCF (see Algorithm 1), takes as input the set of the reviews of an app, organized by days, and a threshold value θ. PCF outputs a set of identified pseudo-cliques with $\geq \theta$, that were formed during contiguous time frames.

3) Reviewer Feedback (RF) Module:
Reviews written by genuine users of malware and fraudulent apps may describe negative experiences. The RF module exploits this observation through a two step approach:

1. Detect and filter out fraudulent reviews, then
2. Identify malware and fraud indicative feedback from the remaining reviews.

4) Inter-Review Relation (IRR) Module:
This module leverages temporal relations between reviews, as well as relations between the review, rating and install counts of apps, to identify suspicious behaviors.

A. Proposed System Screenshots

Figure 1: Initial Form

Figure 2: APK Extractor Form

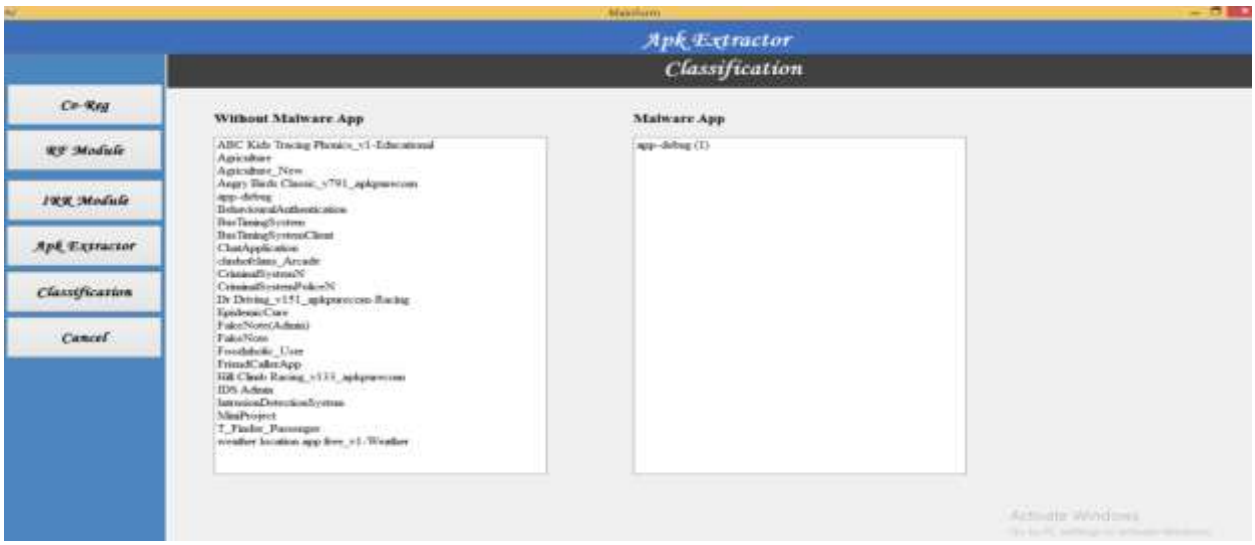

Figure 3: Classification Form

CONCLUSION AND FUTURE WORK

We developed a fraud detection system for mobile Apps. Specifically, we first showed that fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records.

REFERENCES

1. Alaa Salman Imad H. Elhajj Ali ChehabAymanKayss, IEEE Mobile Malware Exposed. International Conference on Knowledge discovery and data mining, KDD'14 pages 978-983.
2. Alfonso Munoz, Ignacio Mart ˜ ´ın, Antonio Guzman, Jose Alberto Hern ´ andez, IEEE Android malware detection from Google Play meta-data: Selection of importantfeatures.2015, pages,245-251.
3. Chia-Mei Chen, Je-Ming Lin, Gu-HsinLai, IEEE Detecting Mobile Application Malicious Behaviors Based on DataFlow of Source Code.2014 International Conference onTrustworthy Systems and their Applications pp 95-109.
4. D. F. Gleich and L.-h.Lim. Rank aggregation via nuclear norm minimization. In Proceedings of the 17th ACMSIGKDD international conference on Knowledge discovery and data mining, KDD '11, pages 60–68, 2011.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.
5. 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 '10, pages 939–948, 2018.
6. N. Jindal and B. Liu, "Opinion spam and analysis," inProc. Int. Conf. Web Search Data Mining, 2008, pp. 219–230.
7. J. Oberheide and C. Miller, "Dissecting the Android Bouncer," presented at the SummerCon2012, New York,NY, USA, 2017.
8. 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'12, pages 204–212, 2016.
9. J.Kivinen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," in Proc. 27th Annu.ACM Symp. Theory Comput. 1995, pp.209–218.
10. N. Spirin and J. Han. Survey on web spam detection: principles and algorithms. SIGKDD Explore. Newsl.13 (2):50–64, May2018.