INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 191-200

NONINTRUSIVE SMARTPHONE USER VERIFICATION USING ANONYMIZED MULTIMODAL DATA

Mohana Priya KT[1], Thilagavathi T[2]

[1]Student, The Kavery Engineering College, M. Kalipatti, Tamil Nadu India.

[2]Assistant Professor, The kavery Engineering College, M. Kalipatti, Tamil Nadu India.

ABSTRACT

Smartphone user verification is important as personal daily activities are increasingly conducted on the phone and sensitive information is constantly logged. The commonly adopted user verification methods are typically active, i.e., they require a user's cooperative input of a security token to gain access permission. Though popular, these methods impose heavy burden to smartphone users to memorize, maintain and input the token at a high frequency. To alleviate this imposition onto the users and to provide additional security, we propose a new nonintrusive and continuous mobile user verification framework that can reduce the frequency required for a user to input his/her security token. Using tailored Hidden Markov Models and sequential likelihood ratio test, our verification is built on low-cost, readily available, anonymized, and multimodal smartphone data without additional effort of data collection and risk of privacy leakage. With extensive evaluation, we achieve a high rate of about 94% for detecting illegitimate smartphone uses and a rate of 74% for confirming legitimate uses. In a practical setting, this can translate into 74% of frequency reduction of inputting a security token using an active authentication method with only about 6% risk of miss detection of a random intruder, which is highly desirable.

1. INTRODUCTION

1.1 DIGITAL MEDIA

Google play (formerly android market) is a digital sharing service operated and developed by google. It serve as the official app store for the android operating system, allowing users to browse and download applications developed with the android software development kit (sdk) and published through google. Google play also serve as a digital media store, presenting music, magazines, books, movies, and television programs. It formerly offered Google hardware devices for purchase until the beginning of a separate online hardware retailer, google store, on march 11, 2015.applications are accessible through google play for free of charge or at a cost. They can be downloaded using android device through the play store mobile app or by deploying the application to a gadget from the google play website. Applications exploiting hardware capabilities of a device can be targeted to users of devices with particular hardware components, such as a motion sensor (for motion-dependent games) or a front-facing camera (for online video calling).

1.2 ANDROID APP MARKETS

Commercial achievement of Android app market such as Google Play and the incentive model they offer to popular apps, create them interesting targets for fraudulent and malicious behaviours. Various fraudulent developers dishonestly increase the search rank and fame of their apps (e.g., through fake reviews and bogus installation counts), while malicious developers utilize app markets as a launch pad for their malware. The impulses for such behaviours are for: app popularity surges translate into economic benefits and expedited malware proliferation**.** On daily basis, an app leader board can be updated by app store which display chart rankings of most admired apps, also it is an inspiring thing to make encouraged the growth of mobile apps. In fact, for promoting mobile phone Apps, leader board of apps is the mainly important way of up gradient in the market. An app should be ranked advanced depending upon how its chart of growth raise and progressively it can create number of downloads and ultimately high income. There were dissimilar ways to promote Apps in order to get peak position in App leader boards, the official one is white hat basis to promote their App to get famed and alternately more number of downloads.

2. SCOPE OF THE PROJECT

2.1 MINING SMARTPHONE DATA FOR APP USAGE PREDICTION AND RECOMMENDATION: A SURVEY

H. Cao has proposed Malicious apps hide with-in other normal apps, which makes their detection difficult. Existing mobile anti-virus software are not sufficient in their reactive nature by relying on identified malware samples for signature mining. It describes a proactive technique to spot zero-day Android malware. Without relying on malware samples and their signatures, this scheme is stimulated to evaluate possible security risks posed by means of these untrusted apps. Specifically, an automated system called RiskRanker to scalable have a look at a particular app that exhibits risky behaviour (e.g., launching a root exploit or sending background SMS messages). The output is then used

to make a prioritized list of reduced apps that merit further investigation. We overview in this survey state-of-the-art research on the topic of mining smartphone usage patterns. In particular, we review these studies extensively for two main research streams, namely app usage prediction and app recommendations. Our scope encompasses the data sets used, common phone usage statistics, explicit and implicit feature representation, methodologies, system design considerations, and the currently achieved performances.

2.2 LEARNING HUMAN IDENTITY FROM MOTION PATTERNS

Natalia Neverova1 et al., has proposed in this paper We present a large-scale study exploring the capability of temporal deep neural networks to interpret natural human kinematics and introduce the first method for active biometric authentication with mobile inertial sensors. At Google, we have created a first-of-its-kind data set of human movements, passively collected by 1500 volunteers using their smartphones daily over several months. We compare several neural architectures for efficient learning of temporal multi-modal data representations, propose an optimized shift-invariant dense convolutional mechanism, and incorporate the discriminatively trained dynamic features in a probabilistic generative framework taking into account temporal characteristics. Our results demonstrate that human kinematics convey important information about user identity and can serve as a valuable component of multi-modal authentication systems. From a modeling perspective, this work has demonstrated that temporal architectures are particularly efficient for learning of dynamic features from a large corpus of noisy temporal signals, and that the learned representations can be further incorporated in a generative setting.

2.3 GRAPHICAL PASSWORDS IN THE WILD – UNDERSTANDING HOW USERS CHOOSE PICTURES AND PASSWORDS IN IMAGE-BASED AUTHENTICATION SCHEMES

Florian Alt et al., has proposed in this paper Mobile user verification is to authenticate whether a given user is the legitimate user of a smartphone device. Unlike the current methods that commonly require users active cooperation, such as entering a short pin or a one-stroke draw pattern, we propose a new passive verification method that requires minimal imposition of users through modeling users subtle mobility patterns. Specifically, our method computes the statistical ambience features on Wi-Fi and cell tower data from location anonymized data sets and then we customize Hidden Markov Model (HMM) to capture the spatial temporal patterns of each user's mobility behaviors. Our learned model is subsequently validated and applied to verify a test user in a time-evolving manner through sequential likelihood test.

This paper explores the use of image-based passwords in the wild. We released an image-based password app in the Google Play store and collected data from 2318 unique devices over one year. Through investigating aspects, such as the choice of images and passwords, it became apparent, that findings from prior work on the security of other schemes (PINs, lock patterns, etc.) do not easily transfer to image based passwords. Our initial assessment of security shows that there is a need for further research, to make image-based passwords more secure. Our work reveals weaknesses but also opportunities offered by such authentication schemes..

2.4 MOBILITY PROFILING FOR USER VERIFICATION WITH ANONYMIZED LOCATION DATA

Miao Lin et al., has proposed in this paper Mobile user verification is to authenticate whether a given user is the legitimate user of a smartphone device. Unlike the current methods that commonly require users active cooperation, such as entering a short pin or a one-stroke draw pattern, we propose a new passive verification method that requires minimal imposition of users through modeling users subtle mobility patterns. Specifically, our method computes the statistical ambience features on Wi-Fi and cell tower data from location anonymized data sets and then we customize Hidden Markov Model (HMM) to capture the spatial temporal patterns of each user's mobility behaviors. The major purpose of the classification process in is to authenticate users using a classifier. We discuss our authentication model and classifier in this section. Moreover, since there is no systematic study of touch biometric properties so far, we further introduce our discrimination model for studying its biometric properties. The key difference of a discrimination model from an authentication model is that, in a discrimination model, we can have the data of each class for training. Fig. 4 compares these two models and visualizes their difference.

2.5 SENSOR USE AND USEFULNESS: TRADE-OFFS FOR DATA-DRIVEN AUTHENTICATION ON MOBILE DEVICES

Nicolas Micallef et al., has proposed in this paper Modern mobile devices come with an array of sensors that support many interesting applications. However, sensors have different sampling costs (e.g., battery drain) and benefits (e.g., accuracy) under different circumstances. In this work we investigate the trade-off between the cost of using a sensor and the benefit gained from its use, with application to data driven authentication on mobile devices. In this paper, we conducted a cost/benefit analysis of sensors that are commonly found on today's mobile devices. To this end, we computed the sampling costs of sensors and established their usefulness in a data-driven authentication scenario.

While previous work investigated the sensor costs, we are – to the best of our knowledge – the first to investigate this phenomenon in the context of data-driven authentication. Given a large array of sensors that user behaviour modelling techniques can use, we believe it is important to identify the sensors that provide best support for detecting attacks and to establish their sampling costs. Our battery consumption results in Section III indicate that with high sampling rates, light and medium drain users are most impacted by sensor costs, but that rate reductions can significantly reduce consumption. However, detection results under different sampling rates in Section IV showed that sampling rates over a few minutes would not be effective against attacks. When we analysed the sensor data individually we identified that using many sensors at once tempers the extreme values from a single sensor and reduces false positives.

2.6 MOBILE DEVICE USAGE CHARACTERISTICS: THE EFFECT OF CONTEXT AND FORM FACTOR ON LOCKED AND UNLOCKED USAGE

H. Xu et al., has proposed in this paper Current smartphones generally cannot continuously authenticate users during runtime. This poses severe security and privacy threats: A malicious user can manipulate the phone if bypassing the screen lock. To solve this problem, our work adopts a continuous and passive authentication mechanism based on a user's touch operations on the touchscreen. Such a mechanism is suitable for smartphones, as it requires no extra hardware or intrusive user interface. We study how to model multiple types of touch data and perform continuous authentication accordingly. As a first attempt, we also investigate the fundamentals of touch operations as biometrics by justifying their distinctiveness and permanence. A one month experiment is conducted involving over 30 users. This is particularly relevant to ECG signals for the diagnosis of HFD as the first step to treatment and care of patients in general and specifically those with early heart disease to increase their overall survival. This paper outlines a hybrid approach of dual SVM and nonparametric algorithm to spot HFD in ECG signals leading to increase reliability and accuracy of identification and diagnosis of heart failure classes in the early stages using the proposed algorithm. The nonparametric algorithm is used to train SVM and its dual to get two models of SVM. The dual problem gives a different view that is better and sometimes simpler than the original problem. This feature is used to detect Heart failure disease in ECG signals by comparing the outputs of SVM model and those of dual SVM model. Experiments show that the hybrid approach produces good results, is more efficient and increases accuracy of Heart failure disease detection with an acceptable accuracy of 94.97% when compared with other algorithms to which the paper refers to. This is especially noted in patients with multiple diseases who were not initially identified as heart failure.

3. SYSTEM DESIGN

3.1 EXISTING SYSTEM

In case of the existing system the fraud is detected after the fraud is done that is, the fraud is detected after the complaint of the Mobile apps holder. And also now a days lot of online purchase are made so we don't know the person how is using the Mobile apps online, we just capture the IP address for verification purpose. So there need a help from the cyber crime to investigate the fraud. In the experiments, we validate the effectiveness of the proposed system, and show the scalability of the detection algorithm as well as some regularity of ranking fraud activities. Due to the huge number of mobile Apps, it is difficult to manually label ranking fraud for each App, so it is important to have a scalable way to automatically detect ranking fraud without using any benchmark information.

DRAWBACKS

- The detection of the fraud use of the Mobile apps is did not easy to found.
- In this system even the original Mobile apps holder is also checked for fraud detection.
- We didn't find the most accurate detection on fraud apps using this technique.
- Therefore, detecting ranking fraud of mobile Apps is actually to detect ranking fraud within leading sessions of mobile Apps.
- Specifically, we first propose a simple yet effective algorithm to identify the leading sessions of each App based on its historical ranking records.

3.2 PROPOSED SYSTEM

In proposed system, we present A Novel Ensemble Learning using PCA. Which does not require fraud signatures and yet is able to detect frauds by considering a Mobile apps holder's spending habit. The details of items purchased in Individual transactions are usually not known to any Fraud Detection System (FDS) running at the bank that issues detecting Mobile apps to the Mobile apps holders. The types of goods that are bought in that transaction are not known to the FDS. It tries to find any anomaly in the transaction based on the spending profile of the Mobile apps holder, activity monitoring. A Java web crawler was developed to download 970 positive reviews and 710 negative reviews randomly.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 191-200

ADVANTAGES

- The detection of the fraud use of the Mobile apps is found much faster that the existing system because we focus by consumer opinions
- In case of the existing system even the original Mobile apps holder is also checked for fraud detection.
- High in classification accuracy by Ensemble process.
- Fast computation and time is less compare than existing work.

3.3 MODULE DESCRIPTION

- **Data Pre-Processing**
- **Co-Review Graph Module**
- **Reviewer Feedback Module**
- **Inter-Review Relation Module**
- **Jekyll-Hyde App Detection Module**

3.3.1. DATA PRE-PROCESSING

- Previous studies revealed that pre-processing of text messages can improve the performance of text classification.

App Feature Identification: As product reviews are about product features the product features are good indicators in classifying the sentiment of product reviews for product review based sentiment classification.

Guaranteed Rating Based Evidences: The ranking based evidences are useful for ranking fraud detection. However, sometimes, it is not sufficient to only use ranking based evidences.

Review Based Evidences: Besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps.

Fraud Apps Blocking: One of the most important perspectives of App ranking fraud. Specifically, before downloading mobile App, users often firstly 5, read its historical reviews to ease their decision making,

3.3.2 CO-REVIEW GRAPH 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. Let the co-review graph of an app, be a graph where nodes correspond to user accounts who reviewed the app, and undirected edges have a weight that indicates the number of apps reviewed in common by the edge's endpoint users.

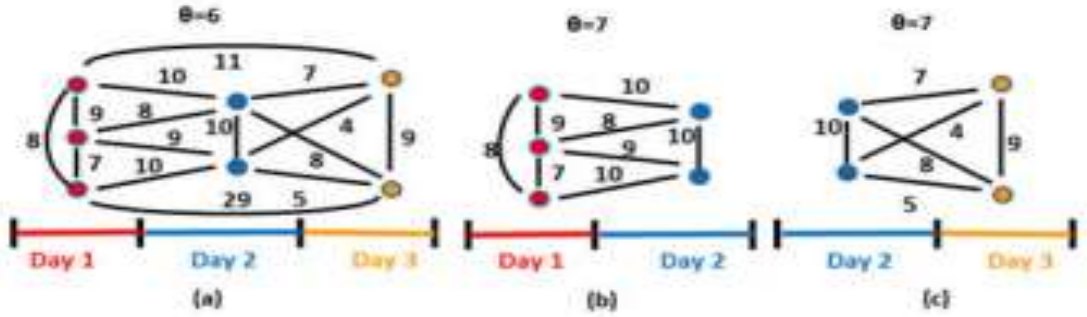

Figure 1 Co-Review Graph of an App

In a Weighted Pseudo-Clique Enumeration Problem for a graph $G = V, E$ and a threshold value θ, say that a vertex sub-set U (and its induced sub-graph G|U|) is a pseudo-clique of G if its weighted density $\rho = \frac{\sum_{e \epsilon E} w(e)}{\binom{n}{2}}$ [29] exceeds θ; $n = |v|$.[1] U is a maximal pseudo-clique if in addition, no other pseudo-clique of G contains U. The weighted pseudo-clique enumeration problem outputs all the vertex sets of V whose induced subgraphs are weighted pseudo-cliques of G.

Algorithm: THMM Algorithm Pseudo-Code

Input: days, an array of daily reviews, and θ, the weighted threshold density

Output: all Cliques, set of all detected pseudo-cliques

1. for $d := 0\ d < days.size();\ d++$
2. Graph $PC :=$ new $Graph();$
3. $bestNearClique(PC, days[d]);$

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 191-200

```
4.  c := 1; n := PC.size();
5.  for nd := d + 1; d < days.size() & c = 1; d + +
6.  bestNearClique(PC, days[nd]);
7.  c := (PC.size() > n); endfor
8.  if (PC.size() > 2)
9.  allCliques := allCliques.add(PC); fi endfor
10. return
11. function bestNearClique(Graph PC, Set revs)
12. if (PC.size() = 0)
13. for root := 0; root < revs.size(); root + +
14. Graph candClique := new Graph ();
15. candClique.addNode (revs[root].getUser());
16. do candNode := getMaxDensityGain(revs);
17. if (density(candClique ∪ {candNode}) ≥ θ))
18. candClique.addNode(candNode); fi
19. while (candNode ! = null);
20. if (candClique.density() > maxRho)
21. maxRho := candClique.density();
22. PC := candClique; fi endfor
23. else if (PC.size() > 0)
24. do candNode := getMaxDensityGain(revs);
25. if (density(candClique ∪ candNode) ≥ θ))
26. PC.addNode(candNode); fi
27. while (candNode ! = null);
28. return
```

CoReG extracts the following features from the output of THMM:

(i) The number of cliques whose density equals or exceeds θ,
(ii) The maximum, median and standard deviation of the densities of identified pseudo-cliques,
(iii) The maximum, median and standard deviation of the node count of identified pseudo-cliques, normalized by n (the app's review count),
(iv) The total number of nodes of the co-review graph that belong to at least one pseudo-clique, normalized by n.

3.3.3 REVIEWER FEEDBACK 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:

(i) Detect and filter out fraudulent reviews, then
(ii) Identify malware and fraud indicative feedback from the remaining reviews.

Step RF.1: Fraudulent Review Filter.

Certain features can accurately pinpoint genuine and fake reviews.

The trained Naive Bayes classifier is used to determine the statements of R that encode positive and negative sentiments. Extract the following features: (i) the percentage of statements in R that encode positive and negative sentiments respectively, and (ii) the rating of R and its percentile among the reviews written by U.

Step RF.2: Reviewer Feedback Extraction.

Conjecture that (i) since no app is perfect, a "balanced" review that contains both app positive and negative sentiments is more likely to be genuine, and (ii) there should exist a relation between the review's dominating sentiment and its rating. Thus, after filtering out fraudulent reviews, extract feedback from the remaining reviews. For this, use NLTK to extract 5,106 verbs, 7,260 nouns and 13,128 adjectives from the 97,071 reviews collected from the 613 gold standard apps. Removed non ascii characters and stop words, then applied lemmatization and discarded words that appear at most once. Attempted to use stemming, extracting the roots of words, however, it performed poorly. This is due to the fact that reviews often contain (i) shorthands, e.g., "ads", "seeya", "gotcha", "in app", (ii) misspelled words, e.g., "pathytic", "folish", "gredy", "dispear" and even (iii) emphasized misspellings, e.g., "hackkked", "spammer", "spooooky". Thus, ignored stemming.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 191-200

3.3. 4 INTER-REVIEW RELATION 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. In order to compensate for a negative review, an attacker needs to post a significant number of positive reviews.

Claim 1. Let R_A denote the average rating of an app A just before receiving a 1 star review. In order to compensate for the 1 star review, an attacker needs to post at least $\frac{R_A-1}{5-R_A}$ positive reviews.

Proof. Let σ be the sum of all the k reviews received by a before time T. Then,$R_A = \frac{\sigma}{k}$. Let q_r be the number of fraudulent reviews received by A. To compensate for the 1 star review posted at timeT, q_r is minimized when all those reviews are 5 stars. Then have that:$R_A = \frac{\sigma}{k} = \frac{\sigma+1+5q_r}{k+1+q_r}$. The numerator of the last fraction denotes the sum of all the ratings received by A after time T and the denominator is the total number of reviews. Rewriting the last equality, obtain that $q_r = \frac{\sigma-k}{5k-\sigma} = \frac{R_A-1}{5-R_A}$. The last equality follows by dividing both the numerator and denominator by k.

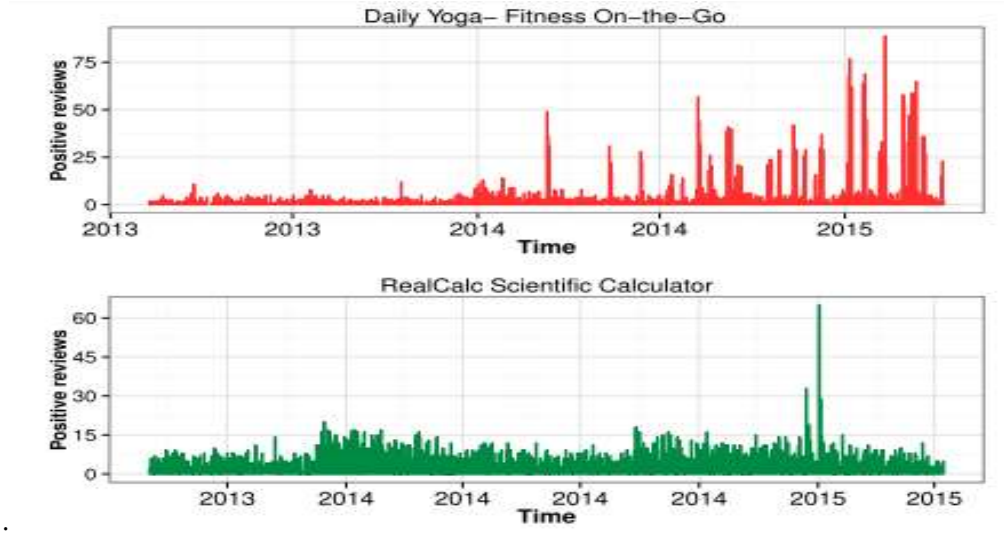

Figure 2 Timelines of reviews for two apps.

Use the Pearson's x2 test to investigate relationships between the install count and the rating count, as well as between the install count and the average app rating of the 87 K new apps, at the end of the collection interval. Then group the rating count in buckets of the same size as Google Play's install count buckets. The mosaic plot of the relationships between ratings and install counts, $p = 0.0008924$, thus conclude dependence between the rating and install counts. The standardized residuals identify the cells (rectangles) that contribute the most to the x^2 test. The most significant rating: install ratio is 1:100.

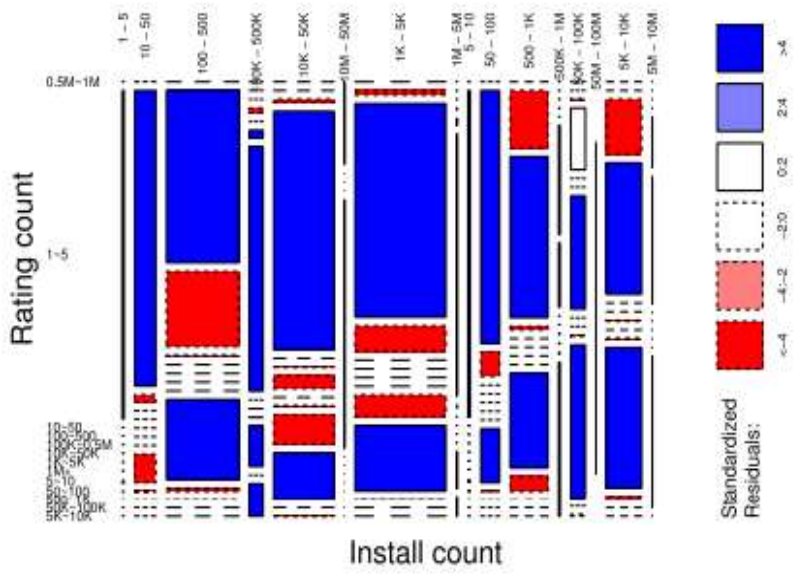

Figure 3 Mosaic plot of install versus rating count.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 191-200

In the mosaic plot of the app install count versus the average app rating, rectangular cells correspond to apps that have a certain install count range (x axis) and average rating range (y axis). It shows that few popular apps, i.e., with more than 1,000 installs, have below 3 stars, or above 4.5 stars.

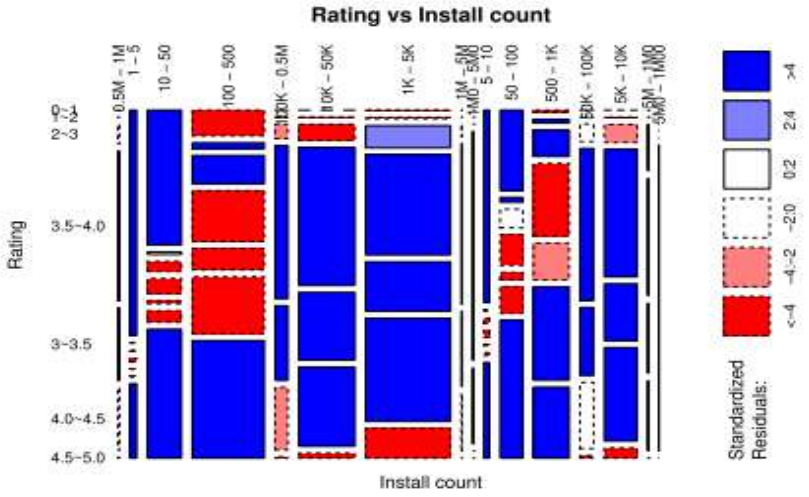

Figure 4 Mosaic plots between install count and app rating

Extract temporal features: the number of days with detected spikes and the maximum amplitude of a spike. Also extract the following:

- The ratio of installs to ratings as two features, I_1/Rt_1 and I_2/Rt_2
- The ratio of installs to reviews, as I_1/Rv_1 and I_2/Rv_2. (I_1, I_2) and denotes the install count interval of an app, (Rt_1, Rt_2) its rating interval and (Rv_1, Rv_2) its (genuine) review interval.

3.3.5 JEKYLL-HYDE APP DETECTION MODULE

The following shows the distribution of the total number of permissions requested by malware, fraudulent and legitimate apps. Surprisingly, not only malware and fraudulent apps but also legitimate apps request large numbers of permissions. In addition, Android's API level 22 labels 47 permissions as "dangerous". It compares the distributions of the number of dangerous permissions requested by the gold standard malware, fraudulent and benign apps. The most popular dangerous permissions among these apps are "modify or delete the contents of the USB storage", "read phone status and identity", "find accounts on the device", and "access precise location". Perhaps surprisingly, most legitimate (69 percent), malware (76 percent) and fraudulent apps (61 percent) request between 1 and 5 dangerous permissions.

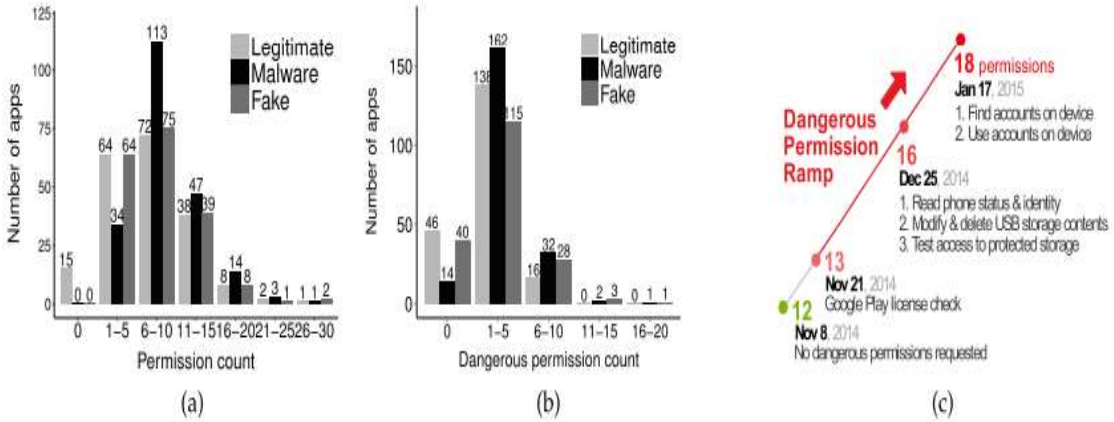

Figure 5 (a) Distribution of total permissions (b) Compares the number of dangerous permissions. (c) Dangerous permissions ramp.

Extract the following features:

- The total number of permissions requested by the app,
- Its number of dangerous permissions,
- The app's number of dangerous permission ramps, and
- Its total number of dangerous permissions added over all the ramps

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 191-200

e-ISSN : 2583-1062

Impact Factor : 5.725

4. SYSTEM DESIGN

4.1 System Architecture

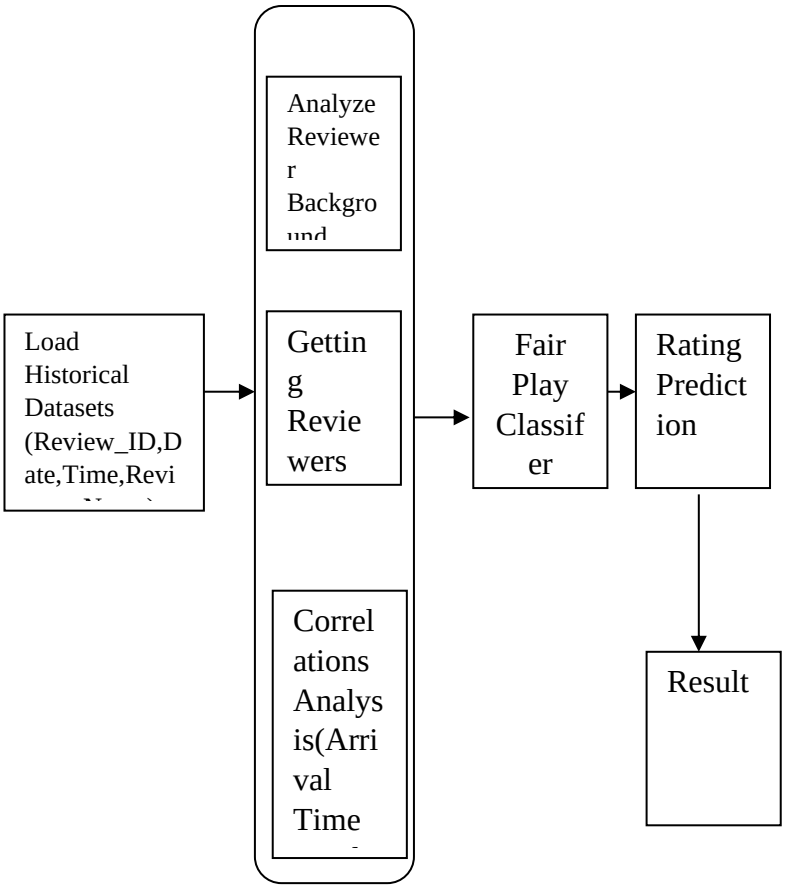

Fig 6: System Architecture

4.2 Data Flow Diagram

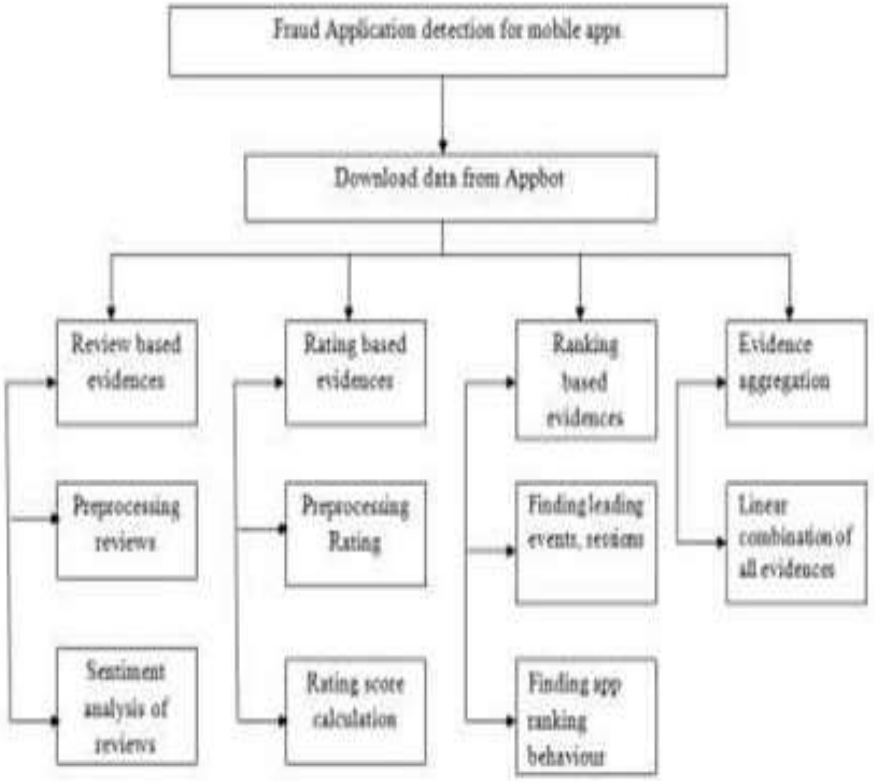

Fig 7: Data Flow Diagram

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 191-200

4.3 System Flow Diagram

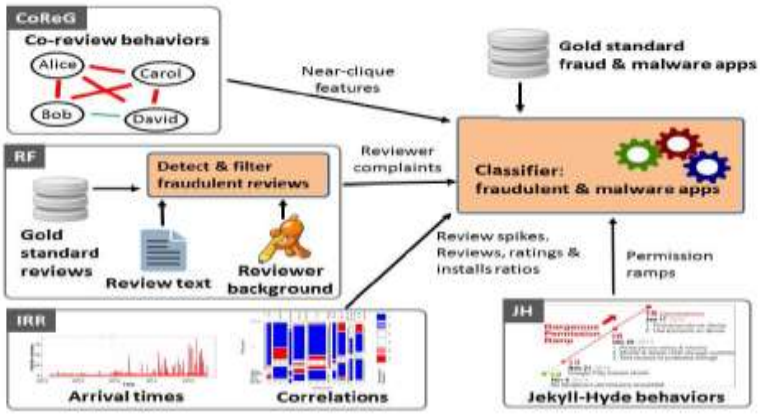

Fig 8 : System Flow Diagram

5. EXPERIMENTAL SETUP AND PROCEDURE

Reviewer Based Features: The Expertise Of U For App A, Defined As The Number Of Reviews U Wrote For Apps That Are "Similar" To A, As Listed By Google Play. The Bias Of U Towards A: The Number Of Reviews Written By U For Other Apps Developed By A's Developer. In Addition, Extract The Total Money Paid By U On Apps It Has Reviewed And The Number Of Apps That U Has Liked, And The Number Of Google+ Followers Of U. Text Based Features: The Nltk Library And The Naive Bayes Classifier, Trained On Two Datasets: (I) 1,041 Sentences Extracted From Randomly Selected 350 Positive And 410 Negative Google Play Reviews, And (Ii) 10,663 Sentences Extracted From 700 Positive And 700 Negative Imdb Movie Reviews. 10-Fold Cross Validation Of The Naive Bayes Classifier Over These Datasets Reveals A False Negative Rate Of 16.1 Percent And A False Positive Rate Of 19.65 Percent, For An Overall Accuracy Of 81.74 Percent. Use The Resulting Words To Manually Identify Lists Of Words Indicative Of Malware, Fraudulent And Benign Behaviours. Malware Indicator Word List Contains 31 Words (E.G., Risk, Hack, Corrupt, Spam, Malware, Fake, Fraud, Blacklist, Ads). The Fraud Indicator Word List Contains 112 Words (E.G., Cheat, Hideous, Complain, Wasted, Crash) And The Benign Indicator Word List Contains 105 Words. Apps Can Request A Group Of Permissions And Gain Implicit Access Also To Dangerous Permissions. Upon Manual Inspection Of Several Apps, There Identified A New Type Of Malicious Intent Possibly Perpetrated By Deceptive App Developers: Apps That Seek To Attract Users With Minimal Permissions, But Later Request Dangerous Permissions. The User May Be Unwilling To Uninstall The App "Just" To Reject A Few New Permissions. Here Call These Jekyll-Hyde Apps. In Addition It Shows The Dangerous Permissions Added During Different Version Updates Of One Gold Standard Malware App.

Table 1: Experimental Table For Test Cases

S.NO	TEST CASE	INPUT	EXPECTED OUTPUT	OUTPUT	PASS/FAIL
1	Check username /password /role	Username/password/role	Login/failed	Login valid	Pass
2	Developer-Upload application	Upload Application	APK Files Upload	Upload APK	Pass
3	User-Download Application	Download Application	APK File Download/reviews/view Malware APPS	Download APK/View Malware APP	Pass
4	Google-play-Admin	APP Developer Details/user/Details/Mobile User Details/Detect Malware	View developer and user details/detect Malware	Detect and list malware application	Pass
5	User	Download APK Files	Write text reviews	Reviews written	Pass

6. RESULTS AND DISCUSSION

Adversaries who want to increase the rating of an app, i.e., cancel out previously received negative reviews, will need to post an increasing, significant number of positive reviews. Such a "compensatory" behaviour is likely to lead to suspiciously high numbers of positive reviews. Detect such behaviours by identifying outliers in the number of daily positive reviews received by an app. The next graph shows the timelines and suspicious spikes of positive reviews for 2 apps from the fraudulent app dataset. Identify days with spikes of positive reviews as those whose number of positive reviews exceeds the upper outer fence of the box-and-whisker plot built over the app's numbers of daily positive reviews. After a recent Google Play policy change, Google Play organizes app permissions into groups of related permissions.

7. CONCLUSION

The proposed method Fair play is a system to detect both fraudulent and malware Google Play apps. Experiments on a newly contributed longitudinal app dataset have shown that a high percentage of malware is involved in search rank fraud; both are accurately identified by FairPlay. In addition, it showed FairPlay's ability to discover hundreds of apps that evade Google Play's detection technology, including a new type of coercive fraud attack. We introduced another nonintrusive client confirmation structure, which is based upon the promptly accessible multi-dimensional cell phone use information with ease perceptions on cell tower associations, Wi-Fi, application use, battery level and charging practices. Utilizing customized HMM to consolidate various perceptions, our proposed system consistently incorporates multimodal anonymized cell phone information successions into a solitary probabilistic model. With successive probability proportion test, we build our model on the anonymized information to mitigate the danger of spilling client private data when the client's models are shared online for incorporated organization and security administration on the cloud. Contrasting and the universal dynamic client confirmation techniques that require client's contribution of PIN, one stroke example and biometrics, for example, face and unique mark, our nonintrusive strategy has the huge preferred position of having zero burden for cell phone clients. It can likewise supplement the omnipresent dynamic check strategies to improve the compromise among convenience and security of cell phone client verification. This thus decreases the likelihood that clients bargain their security tokens in their dynamic confirmation.

8. REFERENCES

[1] H. Cao and M. Lin, "Mining smartphone data for app usage prediction and recommendation: A survey," Pervasive and Mobile Computing, vol. 37, pp. 1–22, 2017.

[2] N. Neverova, C. Wolf, G. Lacey, L. Fridman, D. Chandra, B. Barbello, and G. Taylor, "Learning human identify from motion patterns," IEEE Access, vol. 4, pp. 1810–1820, 2016.

[3] F. Alt, S. Schnessgass, A. S. Shirazi, M. Hassib, and A. Bulling, "Graphical passwords in the wild ? understanding how users choose pictures and passwords in image-based authentication schemes," in MobileHCI '15, 2015.

[4] M. Lin, H. Cao, V. Zheng, K. C. Chang, and S. Krishnaswamy, "Mobility profiling for user verification with anonymized location data," in Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, 2015, pp. 960–966.

[5] N. Micallef, H. G. Kayacik, M. Just, L. Baillie, and D. Aspinall, "Sensor use and usefulness: Trade-offs for data-driven authentication on mobile devices," in Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on, 2015, pp. 189–197.

[6] D. Hintze, R. D. Findling, S. Scholz, and R. Mayrhofer, "Mobile device usage characteristics: The effect of context and form factor on locked and unlocked usage," in Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia, ser. MoMM '14, 2014, pp. 105–114

[7] H. Xu, Y. Zhou, and M. R. Lyu, "Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones," ser. Symposium On Usable Privacy and Security (SOUPS 2014), 2014, pp. 187–198.

[8] N. Zheng, K. Bai, H. Huang, and H. Wang, "You are how you touch: User verification on smartphones via tapping behaviors," ser. IEEE 22nd Int. Conf. on Network Protocols (ICNP), 2014, pp. 221–232.

[9] H. Lu, J. Huang, T. Saha, and L. Nachman, "Unobtrusive gait verification for mobile phones," ser. ISWC'14, 2014, pp. 91–98.

[10] C. Bo, L. Zhang, T. Jung, J. Han, X. Y. Li, and Y. Wang, "Continuous user identification via touch and movement behavioral biometrics," ser. 2014 IEEE 33rd International Performance Computing and Communications Conference (IPCCC), 2014, pp. 1–8.