Risk Score Combined Malware Prediction Using Machine Learning Approach

Dr. P.A. Selvaraj[1], Dr.M.Jagadeesan[2], R.Gowri Sankari[3]

[1]Assistant Professor, Department of computer Applications, Kongu Engineering College, Erode-638060, Tamilnadu, India.

[2] Assistant Professor, Department of computer Applications, Kongu Engineering College, Erode-638060,Tamilnadu, India.

[3]PG scholar, Department of computer Applications, Kongu Engineering College, Erode-638060, Tamilnadu, India.

Abstract

The mobile phones became the target for risky and snoopy applications. The android's current risk communication technique depends on users to spot the permissions that an app is requesting. But users are unaware of permissions because it requires some technical knowledge. Therefore, android's protection against malicious app is a risky communication method where any person who wishes to put in an app are going to be warned about permissions, the appliance would involve then the user has got to take the right decision. Therefore, the protection against malware applications should depend upon decisions made by users. the most a part of protection against malware on mobile devices is to alert the users about malware and permit them to require decisions about whether to settle on and install specific apps. Compute risk score that users can apply while choosing applications whether or not they want to use that app or not.

Keywords: Android Devices, Risk Score, NLP, data processing

1. INTRODUCTION

Mobile devices are widely used thanks to its popularity and functionality. Smart phones became trendy for personal and business use in recent years. All details are stored on the devices, which include contact lists, email messages, passwords, private information and access to those information that are stored locally and within the cloud. With the arrival of smart phones, users have their private information with them on their phones. This information ranges from location of the phone and also their bank particulars. While attacks on mobile devices have mainly focused on stealing users personal data contained on the devices. Possible access to tip by unknown person puts users in danger . because the Android platform has developed to require one among the main shares of the smart-phone market, it seems to be the prime target for criminals who are in search of the private data about users. Simultaneously, the safety of the platform has come under scrutiny from security professionals. Malicious software may be a problem for all.

2. LITERATURE SURVEY

The android's basic defense reaction against malicious applications may be a risk communication system which cautions the client about the permissions before the client introduces an application. This method was unsuccessful because it shows the danger information of each application during a "stand alone" manner and during a way that needs focused learning and tons of your time to distil important data [1]. Check the specified properties of risk signals for Android applications with the top goal to get another metric that clients can utilize while choosing applications. Show a good range of techniques to formulate risk scores that specialise in heuristics and also principled machine learning systems. Trial comes on directed utilizing certifiable information sets that show that these techniques can identify malware as dangerous, are simple to work out, and straightforward to take advantage of . There are distinct and large numbers of applications on the Play store where a couple of of them look similar. Moreover various applications are duplicate and also fraud application which contains malware. they'll get information from the android and should damage the android device anytime. Such applications even have client ratings and review which figures out the benign and fraud application. In android application, it's rundown of uses from a spread of classes that it's necessary to spot whether it's fake or not. Mobile vide contact to individual person and fragile data devices are becoming universal, and that they are telephone records, call lists, geolocation, and SMS messages, making their safety a really important test . live store users download and utilize numerous applications. The protection against malware applications depends on users. a crucial a part of malware protection on mobile devices is to reveal about the danger of putting in the appliance to clients and to permit the user to require decisions about whether to pick and install the applications .

Duen Horng Chau--Polonium may be a scalable and helpful method for locating malware. Then evaluate it with the foremost anonym zed file submissions dataset ever available, which spans over 60 terabytes of disc space . the tactic for identifying malware is to locate files with low reputation. With one iteration, this method attained 85.5% of true positive rate (in detecting malware). With more iterations, truth positive rate further improves for a further 2.1%, which may be a significant improvement given the baseline performance is already excellent . Then detail significant design and implementation features of the tactic which enable its successful application on the dataset. Experimental observations are presented on characteristics and patterns within the large billion-node graph. Symantec introduced the new protection model that computes a reputation score for each application that users may encounter, and protects them from files with poor name. Good applications which are employed by many users, from recognized publishers, produce other attributes that contains their legitimacy and good reputation. Bad applications, on the opposite hand, come from mysterious publishers, have appeared on few computers, and produce other attributes that indicate poor reputation. the appliance status is computed by leveraging tens of terabytes of data anonymously contributed by the many users participating within the worldwide Norton Community Watch program.

Hao Peng--One of Android's core protection methods against malicious apps may be a risk communication method which, before a user installs an app, warns the user about the permissions the app requires, trusting that the user will make the right judgment. This approach has been shown to be useless because it presents the danger information of each app during a "stand-alone" fashion and during a way that needs an excessive amount of practical awareness and a few time to extort valuable information. Begin the notion of risk scoring and risk ranking for Android apps, to create up risk communication for Android apps, and recognize three desired data for an efficient risk scoring scheme.

Probabilistic generative models are used for risk scoring schemes, and recognize several models, starting from the straightforward Naive Bayes, to advanced hierarchical mixture models. Experimental results administered using real-world datasets reveal that probabilistic general models considerably do better than other existing approaches, which Naive Bayes models provides a promising risk scoring approach.

Probabilistic generative models are used widely during a range of applications in machine learning, computer vision, and computational biology, to model complex facts. The foremost strength is to model functions during a great quantity of unlabeled information. Using these models, it's believed that some parameterized random procedure generates the app data and learn the model parameter supported the knowledge . Then, calculate the probability of every app generated by the model. the danger score are often any function that's inversely associated with the probability, in order that lower probability interprets into a better score.

3. PROPOSED WORK

The proposed framework categorizes the applications in Google play as benign or fraud application using Naïve Bayes classifier. The proposed solution are often employed by both mobile users to form better decision and android markets to filter suspicious applications. during this method malware detection in Google play mainly focuses on similarity matching and behavior profiling to detect suspicious application. This proposed method contains four modules. the primary module is that the Co-Review Graph module. The second module called as Review Feedback module identifies feedback left by genuine reviewers. The third module called as Inter Review Relation module consider the relations between reviews and install counts also as between average rating and install counts. The last module called Jekyll-Hyde is employed to spot the permission requested for applications and it also involves in monitoring the permission until the user removes the appliance . It incorporates Naïve Bayes classification algorithm and therefore the classifier was trained using gold standard data sets.

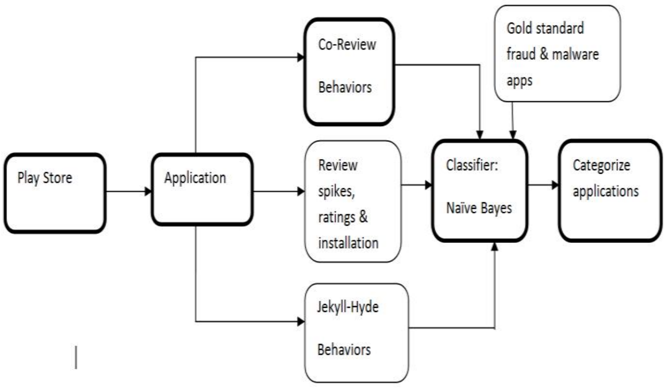


Fig. 1. Flow of proposed work

NAIVE BAYES MODEL

Naive Bayes is one among the techniques which is employed for constructing classifiers: it assigns class labels to problem instances, which was indicated as vectors of feature values, where the category labels are derived from some finite set. it's not just one algorithm for training such classifiers, but a family of algorithms and that they are supported a standard principle: all Naive Bayes classifiers imagine that the worth of a specific feature doesn't depend upon the worth of the other feature for any given the category variable.

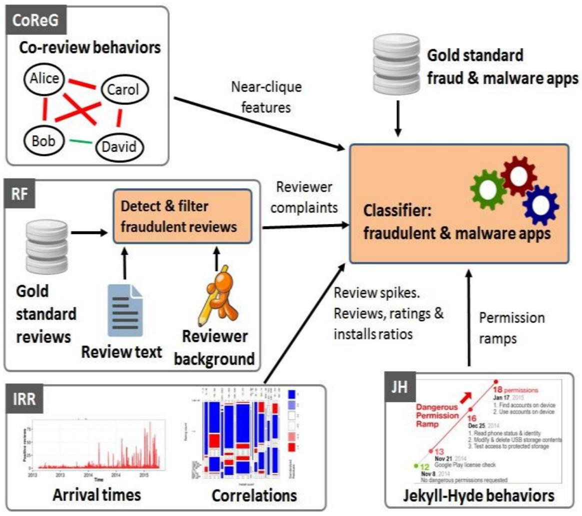
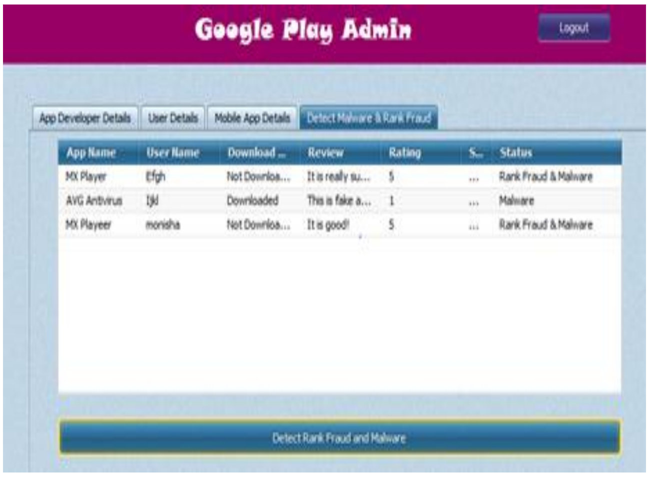

CoReG
Co-review behaviors
Alice
Carol
Bob
David
Near-clique features
Gold standard fraud & malware apps
RF
Detect & filter fraudulent reviews
Gold standard reviews
Review text
Reviewer background
Reviewer complaints
Classifier: fraudulent & malware apps
Review spikes. Reviews, ratings & installs ratios
Permission ramps
IRR
Arrival times
Correlations
JH
Dangerous Permission Ramp
18 permissions
Jan 17, 2015
1. Find accounts on device
2. Use accounts on device
16
Dec 25, 2014
1. Read phone status & identity
2. Modify & delete USB storage contents
3. Test access to protected storage
13
Nov 21, 2014
Google Play license check
12
Nov 8, 2014
No dangerous permissions requested
Jekyll-Hyde behaviors

Fig. 2. Architecture Diagram

NAIVE BAYES CLASSIFICATION ALGORITHM

The Naive Bayes may be a classification algorithm and it's supported Bayes theorem. The word naive from Naive Bayes comes from the very fact that the algorithm takes Bayesian techniques and it ignores dependencies which will exist. This algorithm is a smaller amount computationally intense than other algorithms and thus it's helpful for generating mining models to seek out out relationships between input columns and predictable columns. It trains the classifier so as to spot fraud apps. They're trained using gold standard datasets. Thus Naive Bayes is effective in predicting fraud application in Google play because it requires less training data.



App Name	User Name	Download ...	Review	Rating	S...	Status
MX Player	Efgh	Not Downloa...	It is really su...	5	...	Rank Fraud & Malware
AVG Antivirus	Ijkl	Downloaded	This is fake a...	1	...	Malware
MX Playeer	monisha	Not Downloa...	It is good!	5	...	Rank Fraud & Malware

Fig. 3. Fruad and malware apps

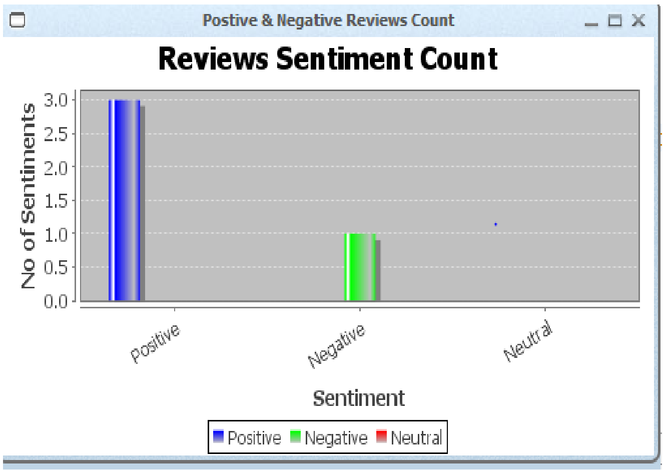


Fig.4. Review count

4. CONCLUSION

Most of the android users are unaware about the permissions requested for application. Thus, mobile devices become target for fraudsters. This made fraudster to interrupt the safety of mobile devices and steal user's tip . Sometimes the applications get updated. Whenever updated versions of applications are installed, it's difficult for users to spot whether the updated version contains malware or not. the prevailing techniques are ineffective in identifying malware as they consider just one factor for predicting malware i.e. reviews or permissions. Therefore, this work mainly focused on user reviews, ratings and permission of every android app. It helps to filter applications and recognize malware and fraud application. The user reviews and ratings also are evaluated to acknowledge fraud reviews. Reviews are evaluated to spot genuine and fraud reviews. Thus, it helps to work out benign and fraud apps. It adapts classification techniques for classifying applications thus improving the safety of android devices.

REFERENCES

[1] Mahmudur Rahman, Mizanur Rahman, Bogdan Carbunar, and Duen Horng Chau, "Search Rank Fraud and Malware Detection in Google Play," in IEEE Transactions On Knowledge and Data Engineering, Vol. 29, No. 6, June 2017.

[2] M. Nauman, S. Khan, and X. Zhang, "Apex: Extending Android Permission Model and Enforcement with User-Defined Runtime Constraints," Proc. Fifth ACM Symp. Information, Computer and Comm. Security, pp. 328-332, 2010.

[3] Christopher S. Gates, Ninghui Li, Senior Member, IEEE, Hao Peng, Bhaskar Sarma, Yuan Qi, Rahul Potharaju, Cristina NitaRotaru, Member, IEEE Computer Society, and Ian Molloy "Generating Summary Risk Scores for Mobile Applications," IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 11, NO. 3, MAY-JUNE 2014.

[4] A.P. Felt, K. Greenwood, and D. Wagner, "The Effectiveness of Application Permissions," Proc. Second USENIX Conf. Web Application Development, (WebApps '11), 2011.

[5] Christopher S. Gates, Jing Chen, Ninghui Li, Senior Member, IEEE, and Robert W. Proctor, "Effective Risk Communication for Android Apps," IEEE Transactions on Dependable and Secure Computing, vol. 11, no. 3, 2014.

[6] T. Vidas, N. Christin, and L.F. Cranor, "Curbing Android Permission Creep," Proc. Workshop Web 2.0 Security and Privacy, vol. 2, 2011.

[7] B.P. Sarma, N. Li, C. Gates, R. Potharaju, C. Nita-Rotaru, and I. Molloy, "Android Permissions: A Perspective Combining Risks and Benefits," Proc. 17th ACM Symp. Access Control Models and Technologies (SACMAT '12), 2012.

[8] M. Grace, Y. Zhou, Q. Zhang, S. Zou, and X. Jiang, "RiskRanker: Scalable and Accurate Zero-Day Android Malware Detection," Proc. 10th Int'l Conf. Mobile Systems, Applications, and Services, (MobiSys '12), pp. 281-294, 2012.

[9] SY Yerima, S Sezer, G McWilliams - IET Information Security, 2014 - ieeexplore.ieee.org "Analysis of Bayesian classificationbased approaches for Android malware detection."