Rogue Apps Analytics: A Systematic Literature Review

Govinda Roy, MD. Rasidul Haq Shuvo, Fatin Ishraq, Zahiduzzaman Pranta, MD. Shamsur Rahim
Department of Computer Science, American International University – Bangladesh
Govindaroy.ofc94@gmail.com

***Abstract*—Information theft from the smart phones is one of the most concerning incidents throughout the world. And the information thefts are mostly being done by making apps which are referred as Rogue App or fraud app. Due to the large volume of apps on the app sharing platforms, such activities cannot be stopped just by identifying the apps manually. The platforms' administrators need something more reliable, repeatable and something vigilant. Several studies have been conducted till now to identify the rogue apps. In this paper, we are presenting a systematic literature review on the existing approaches of identification, prediction of rogue apps. Here, different types of techniques for identifying the apps are reviewed. Furthermore, this paper identifies research gaps in the existing works and provides future research directions.**

***Index Terms*— Rogue apps, fraud apps, systematic literature review, app store analysis.**

I. INTRODUCTION

Our smartphones contain various types of data. Such as, bank password, phone numbers, Facebook id, Emails etc. Simultaneously, we are installing various apps and checking them for their worthiness. If the worthiness of the app is satisfactory then we keep the app in our device and keep using it until we do not have any use of it anymore. However, one of the most important facts that we try to ignore is the apps that we are installing or updating may steal our valuable information and that valuable information could cost us very badly.

Popular apps can sound fun and convenient but some of these popular apps may carry virus or malwares by which the hackers can easily access to our phones and thus our personal information. A software firm found that, about 75 to 80 percent of top free apps were breached and the number of breaching percentage increases to 97 percent in paid apps[1].

The concept of fraud or rogue app is not a recent topic in the world of smartphones. Fraud apps can be classified into two categories:

- False Feature Rich: During searching, the source has showed that, the app contains "ABC" feature. But after installing there is no sign of those features. In the end, that app is fraud.
- Malware: In this case, the app shows all the features which was shown in the source portal, however it contains malwares that users are not aware of. And through that malware, all the information of that user is being extracted.

In both cases the user cannot do anything but uninstalling the app and move on. So, identifying these apps are very crucial so that, in the distant future these types of apps cannot not harm other devices and the user data.

Though several techniques are proposed to identify fraud apps over the years, however people are still affected by these apps. This means the existing works are not enough to detect these apps and these works may have limitations. As a result, there is a high demand of a Systematic Literature Review (SLR) on the existing works of detecting fraud apps. An SLR on rogue apps analytics will help the researchers and practitioners to get all the required information, limitations and future work direction at one place.

So, this paper is about a literature review on the existing papers and finding out the approaches that has been pulled. By revising the previous works, the future development can be improved. Other researchers have used different methods to detect fraud apps. By understanding the previous papers and the methods mentioned, we have presented research methods, further research scopes as well as limitations.

We have also gone through the Google Play store, Apple Store and Windows app store. From these stores, we have found that Google Play store do not keep any apps which have rating lower than 3. Same goes for Apple app store. Apple app store bans the developers who try to create such apps that will be harmful for the users. However, Windows app store keeps every apps in their Microsoft Store. They also have underrated apps which are most likely to be as we say Rogue App. Means, the app may false feature rich or may take information of the user silently and with those valuable data the hackers can hack into the device and do the worst thing to imagine.

Finally, based on the findings we have provided future research directions to identify fraud or rogue apps with better accuracy.

The rest of paper is structured as follows:

- Section II: The previous works which has been done in this field. The section contains a table which is the summarization of the previous papers.
- Section III: This section provides the research methods and directions also the Research Method flowchart.
- Section IV: Here conclusion and discussion are given with future research directions.

II. RELATED WORKS

The following evaluation criteria are used to compare against each of the works on rogue app analytics in this systematic literature review:

1. App Stores analytics.

2. Detecting fraud apps in those stores.
3. Recommending not to install the risky apps.
4. Data mining techniques to detect the fraud apps from the stores.

Researchers have been trying to predict fraud apps since 2008 [2] also various types of data mining techniques have been used [3] . After several attempts, the researchers have also increased their predictive accuracy to predict the fraud apps. Some of the papers also used some techniques to identify rogue apps which are harmful for the smart device also. In paper [4], the researchers made a machine learning approach for the detection of malware on Android platforms. [5][6][7] are survey which are technical articles and the papers compared, analyzed, commented on the existing malware detection methods. A good fraud and malware detection was shown with the accuracy of 75% using Fair Play [8]. Researchers also demonstrated FRAUDEAGLE [9], which successfully reveals fraud-bots in a large online app review database. There are also some research which ranked the fraud apps in daily app leaderboards [10] and also by using the old record data the apps have been ranked as well [11][12][13] . Risk Ranker [14], also provides a system to detect apps with high, low, medium risks of malware and for the OS kernel. However, a paper published for the people of China, suggesting android users to download the APK files from the highest ranked third party app stores which also provides "Safe-to-download" environment [15].

A great work also has been done for android malware detection and using classification techniques as well [2][16][17][18]. Fraud apps can also be found in miscatogorized apps [19]. Information flow type system for the android platform which are written in Java IFT [20] is also used for the fraud app detection. Detection for virus in Bluetooth and messaging is also important for in order to prevent the fraud app[21]. Wiscom can also help to analyze other secondary Android markets and applying in depth analysis for internal or external events[22].

The analysis also gives us the limitations of the papers which are discussed above. Because these limitations are can be a strength overtime if more research is used in this sector. Research works, mobile applications, android platforms any of these related to this contexts we gone carefully. Table 1 presents the summary of all the information collected during related work study.

Table 1
Review on Research Papers

No	Title	Concept	Criteria					Limitations/Future Works
			1	2	3	4	5	
1.	A Comprehensive Survey of Data Mining-based Fraud Detection Research	This survey covers much more technical articles and is the only one, to the best of our knowledge, which proposes alternative data and solutions from related domains.	√	×	×	√	×	Limitations- -It defines the professional fraudster formalizes the main types and subtypes of only known fraud.
2.	A machine learning approach to anomaly-based detection on android platforms	A machine learning approach for the detection of malware on Android platforms is presented. The detection system monitors and extracts feature from the applications while in execution and uses them to perform in-device detection using a trained K-Nearest Neighbor classifier.	√	×	√	√	√	Limitations- -The system is only made for Android -K-NN model gives 93.75 percent accuracy and 6.25 percent of error. Future Works- -Make 100% accuracy
3.	A Survey on Dynamic Mobile Malware Detection	Compared, analyzed and commented existing mobile malware detection methods proposed in recent years based on the evaluation criteria and measures.	√	√	√	×	×	Limitations- 1. Restriction of computing resources, processing capability and memory storage at mobile devices. 2. Real time detection is not well supported.
4.	Android Malware Detection Using Parallel Machine Learning Classifiers	This paper proposes and investigates a parallel machine learning based classification approach for early detection of Android malware. Using real malware samples and benign applications, a composite classification model is developed from parallel combination of heterogeneous classifiers	×	√	√	√	×	Limitations- 1. More classifiers can be added to detect more unknown malware.
5.	Android App Malware Detection	In this work, the permissions and ape level information from the apps are used as the features to detect malicious applications.	×	×	√	√	×	Limitations- 1.Data set accuracy 2.Study about API Future Work- 1. Improvement of data set.

6.	App Miscategorization Detection: A Case Study on Google Play	(I) it is based on a data-driven topic model and automatically suggests the categories appropriate for the app store, and (ii) it can detect miscategorized apps.	√	×	×	×	×	Limitations- 1. Existence of common mechanism.
7.	Cassandra: Towards a Certifying App Store for Android	Cassandra's security analysis soundly detects all potential information leaks, i.e., all flows of information that violate a user's privacy policy.	×	×	√	×	×	Limitations- 1. Covers 211 of 218 Dalvik Instructions. Future Works- 1. Extending the coverage of Dalvik bytecodes. 2. Intending to add support to the remaining instructions.
8.	Collaborative Verification of Information Flow for a High-Assurance App Store	Implemented the information-flow type system for Android apps written in Java and evaluated both its effectiveness at detecting information-flow violations and its usability in practice.	√	×	×	√	×	Limitations- 1. IFT is used in conjunction with complementary techniques that address other security properties. 2. Can analyze at compile time. Future works- 1. Planning to enrich policies. 2. Adding simple and high-level flavor of the specifications.
9.	Data Mining Techniques in Fraud Detection	Bayesian classification model to detect fraud in automobile insurance. Naïve Bayesian visualization is selected to analyze and interpret the classifier predictions.	×	√	×	√	×	Limitations- 1. The paper has been done in an existing fraud system.
10.	Discovery Of Fraud Applications In Mobile Store	Shown how to find whether an application in mobile store is a fraud or genuine one by having look on to large pre-existing data base or in other word we can say that by analyzing parameters or evidences from databases such as rank based evidence, review based evidences, rating and even analyzing some malicious links.	√	×	×	×	×	Limitations- 1. A form of evidence (ranking, rating, review) is required. Without evidence the system cannot detect any fraud apps in the system. Future Works- 1. Improving detection rate.
11.	Discovery of Ranking Fraud for Mobile Applications	Developed a flawless, fraud less and result that shows corrected application accordingly provide ranking; where actually made it happen by searching fraud of applications. They make fraud of App by ranked high the App by methods using such as human water armies and bot farms; where they make fraud by downloading application through various devices and give fake ratings and reviews.	√	×	×	×	×	Limitations- 1. A form of evidence (ranking, rating, review) is required. 2. Result also shows on matching. Future Works-
12.	Fair Play: Fraud and Malware Detection in Google Play	Shown that 75% of the identified malware apps engage in search rank fraud. Fair Play discovers hundreds of fraudulent apps that currently evade Google Bouncer's detection technology, and reveals a new type of attack campaign, where users are harassed into writing positive reviews, and install and review other apps.	×	√	√	×	×	Limitations- 1. The system was only used in Google Play Store. Future Works-
13.	Fraud Detection by Monitoring Customer Behavior and Activities	Preventing the customer from online transaction by using specific technique i.e. based on Data Mining and Artificial Intelligence technique. The risk score is calculated by Bayesian Learning Approach to analyze whether the transaction is genuine or fraudulent based on the two parameters: Customer Spending Behavior and Geographical Locations.	√	×	×	√	×	Limitations- 1. Done on transaction activities. 2. Algorithm used KMEAN Clustering. Future Works- 1. Adding more parameters in addition to spending behavior & geographical location.
14.	Opinion Fraud Detection in Online Reviews by Network Effects	Demonstrated the effectiveness of our framework on synthetic and real datasets; where FRAUDEAGLE successfully reveals fraud-bots in a large online app review database.	×	×	√	×	×	Future Works- 1. Seeding the algorithm with priors inferred from the clues and study the effects of more informed priors on performance.

15.	Ranking Detection and Avoidance Frauds in Mobile Apps Store.	A proposed system for detecting fraud in ranking the apps from the app store. Identify the fraud app based on evidence like old records.	√	√	√	×	×	Limitations- 1. Data is from old rating record. Future Works- 1. Identification of ranking based evidences, rating based evidences and experience-based evidences for detecting ranking fraud might be accomplished. 2. Integrating the application of fraud detection method to make procedure robust.
16.	Ranking Fraud Detection for Mobile Apps using Evidence Aggression Method	A system is proposed for detecting fraud ranking in daily app leaderboards and avoid ranking manipulation. Uses mining technology on old records like old records, old ratings and comments based on ratings. Combined the collecting data to find the fraud.	√	×	×	√	√	Limitations- 1. For only mobile apps. 2. Only used for app ratings.
17.	Risk Ranker: Scalable and Accurate Zero-day Android Malware Detection	Provides a system to Detect those apps with high, low and medium risks of malware and for the OS kernel. Application with encrypted code may carry virus or not relevant with the app, the motivation is detecting those risky apps.	√	×	√	√	×	Limitations- 1. The scheme processes only for the untrusted apps. 2. Very less malwares were detected. Future works- 1. Obfuscation with the opcodes, the semantic pat.
18.	Smart Siren: Virus Detection and Alert for Smartphones	Provides a system to detect virus in Bluetooth and messaging. After detecting the virus, it gives an alert in the smartphones. This paper is motivating for detecting virus while sharing apps using Bluetooth. Protects user's personal data form viruses.	×	√	×	×	×	Limitations- 1. The system is depended on the internet. If the phone does not have internet, then virus detection cannot be done. Future Work- 1. Improving the scalability and resiliency.
19..	Survey on Fraud Ranking Detection in mobile app store	Provides a survey on various existing techniques with the novelties highlighting the need of novel technique to detect fraud mobile apps. This paper is motivated by arising need to detect fraud apps with less time. In proposed system, added recommendation based on the modified ranking.	×	×	×	√	×	Limitation- 1. The paper is survey based. Future Work- 1. Making a detection system that tracks the online campaigning on social media.
20.	Which Android App Store Can be Trusted in China?	Suggesting Android users to download APK files from its corresponding official websites or use the highest ranked third-party app stores; and appealed app stores to ensure all hosting APK files are trustworthy enough to provide a "safe-to-download" environment.	√	×	×	×	×	Limitations- 1. Only assessed the trustworthy level of 25 top downloading apps in 20 popular third-party Android apps. Future Work- 1. Expecting the system to differentiate between two APK files.
21.	Why People Hate Your App — Making Sense of User Feedback in a Mobile App Store.	Discussed how the techniques presented herein can be deployed to help a mobile app market operator such as Google as well as individual app developers and end-users.	√	×	√	×	√	Future Work- 1. Using WisCom to analyze other secondary Android markets as well as other online marketplaces. 2. Applying more in-depth analysis to investigate the different review patterns triggered by various market operations or external events and to what extend these patterns can be used in market prediction.

III. RESEARCH METHOD

The first phase of our method is generating the questions for the proposal. These questions will give us the papers which are related to the questions. The quest ions will be searched in the Search Engine(s) and papers which are related with the topic is collected. Here, Figure 2.1 is Research Method Flowchart of this paper.

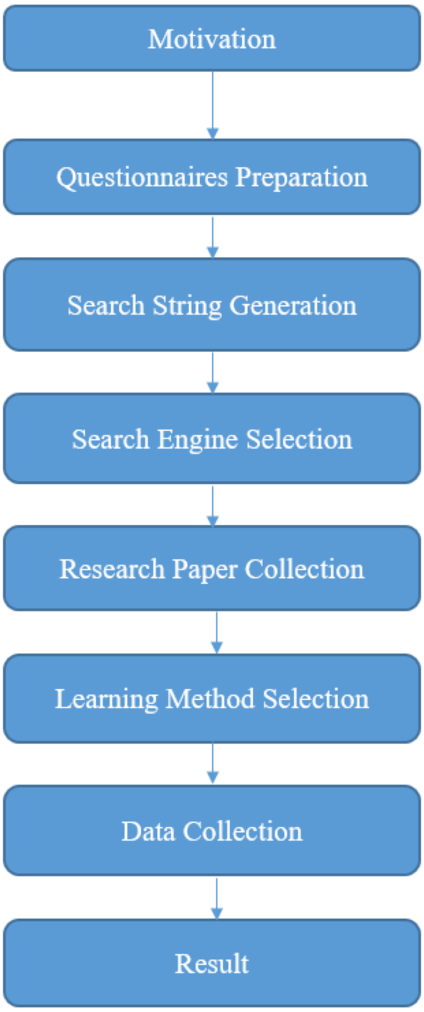

Figure 2.1 Research Method Flowchart

A. Motivation

The motivation of this proposal came from our daily basis usage of smart phone. Since all our personal data and keep them in our smart phone, if these personal data go to the wrong hand loss can be great. And these data can be collected by fraud apps.

B. Questionnaires Preparation

The main questions addressed in this study were:

- Does this research/system analyze different app stores' apps?
- Does this research/system detects fraud apps in these stores?
- Does this research/system recommend not to install risky apps?
- Has any data mining/machine learning technique been used to detect fraud apps from the store?
- If the system already exists, is the system/research reliable?

C. Search String Generation

From the online research portals which hold huge collection of papers we downloaded the papers. During the search process we used these keywords:

a. Data mining techniques to identify fraud apps
b. Fraud app detection
c. Rogue app detection
d. Identification of fraud apps
e. Predicting fraud apps
f. Detecting fraud apps in android
g. Fraud app recommendation
h. Detecting fraud apps in smart devices
i. Ranking of fraud apps

D. Research Paper Collection

First we have collected all the available papers using the keywords above in the "Search String Generation" section. Then we collected those which are titled with "Data mining", "Fraud apps", "Rogue Apps" and papers which are related with our work. The papers which were not titled and yet do not match with our field of work were excluded. So, the final papers contain different methods which are relative to our work and some of them have predictive accuracy of detecting the fraud apps, fraud apps detection in smart devices.

So, the papers are collected based on the criteria are:

- Papers discussing on algorithms/methods/system to detect fraud apps
- Papers discussing on algorithms/methods/system to detect rogue apps
- Papers discussing on algorithms/methods/system to predict fraud apps
- Papers discussing algorithms/methods/system to detect fraud apps in smart devices
- Papers discussing on the ranking of fraud apps

And the papers which do not answer to the research questions were excluded and papers which show redundancy were also excluded.

E. Data Collection

The data were collected manually. Only relevant papers were considered in the learning method to collect data based on the mentioned criteria. Prediction accuracy, fraud apps detection, limitations and feature works were mainly focused here.

IV. CONCLUSION& DISCUSSION

After collecting 102 papers we have finalized to 21 papers which are necessary for our research work. All of these papers have only one goal and is detecting fraud apps in the app stores. And increasing the predictive accuracy rate is also part of the works. [6][18][9][14][15] papers have promising future works if are made possible then the outcome of analysis will surely change in the future. And [3][14][8] papers sound very promising for this research. Our research work has found systems which also predict fraud app as well. But the accuracy rate of the papers are not satisfying. Our future work will be to increase that rate even higher.

We provide the following scopes and future developments, suggestions that can be performed in the future:

- Making the predictive accuracy rate higher
- Real time detection support
- Improvement of dataset
- Using other data mining techniques
- Improving scalability and resiliency
- Analyzing all the app stores

The authors believe that fraud app is a dangerous and if the personal information is stolen or is used to blackmail people then loss would be great. Interrupting with private life of the users is against the law. The fraud app predictive accuracy will increase if other techniques we used. So there is space for improvements.

REFERENCES

[1] cbsnews, "Beware downloading some apps or risk "being spied on," *cbsnews*, 2016. [Online]. Available: https://www.cbsnews.com/news/mobile-phone-apps-malware-risks-how-to-prevent-hacking-breach/.

[2] R. Bhowmik, "Data Mining Techniques in Fraud Detection," *Proc. Conf. Digit. Forensics, Secur. Law*, vol. 3, no. 2, pp. 57–72, 2008.

[3] P. Singh and M. Singh, "Fraud Detection by Monitoring Customer Behavior and Activities," *Int. J. Comput. Appl.*, vol. 111, no. 11, pp. 23–32, 2015.

[4] J. Abah, W. O.V, A. M.B, A. U.M, and A. O.S, "A Machine Learning Approach to Anomaly-Based Detection on Android Platforms," *Int. J. Netw. Secur. Its Appl.*, vol. 7, no. 6, pp. 15–35, 2015.

[5] P. Yan and Z. Yan, "A survey on dynamic mobile malware detection," *Softw. Qual. J.*, no. October, pp. 1–29, 2017.

[6] S. Wang, "A comprehensive survey of data mining-based accounting-fraud detection research," *2010 Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2010*, vol. 1, pp. 50–53, 2010.

[7] M. A. Khan, "Survey on Fraud Ranking Detection in mobile app store," pp. 9–12.

[8] M. Rahman, M. Rahman, B. Carbunar, and D. H. Chau, "FairPlay: Fraud and Malware Detection in Google Play," no. 948, 2017.

[9] L. Akoglu, C. Faloutsos, R. Chandy, and C. Faloutsos, "Opinion Fraud Detection in Online Reviews by Network Effects," *Proceeding 7th Int. AAAI Conf. Weblogs Soc. Media*, pp. 2–11, 2013.

[10] M. P. U. Gayke and P. S. B. Thakare, "Ranking Fraud Detection for Mobile Apps using Evidence Aggregation Method," vol. 4, no. 3, pp. 58–64, 2016.

[11] "Ranking Detection and Avoidance Frauds in Mobile Apps," pp. 413–418.

[12] Shruthi and Prof.Anand s uppar, "Discovery of Fraud Applications in Mobile Stores," *Int. J. Adv. Sci. Res. Eng.*, vol. 3, no. 4, pp. 51–56, 2017.

[13] H. Zhu, H. Xiong, Y. Ge, and E. Chen, "Discovery of ranking fraud for mobile apps," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 1, pp. 74–87, 2015.

[14] M. Grace, Y. Zhou, Q. Zhang, S. Zou, and X. Jiang, "RiskRanker: Scalable and Accurate Zero-day Android Malware Detection," *10th Int. Conf. Mob. Syst. Appl. Serv.*, pp. 281–294, 2012.

[15] Y. Y. Ng, H. Zhou, Z. Ji, H. Luo, and Y. Dong, "Which android app store can be trusted in China?," *Proc. - Int. Comput. Softw. Appl. Conf.*, pp. 509–518, 2014.

[16] V. Grampurohit, "Android App Malware Detection," no. July, p. 64, 2016.

[17] S. Y. Yerima, S. Sezer, and I. Muttik, "Android Malware Detection Using Parallel Machine Learning Classifiers," *2014 Eighth Int. Conf. Next Gener. Mob. Apps, Serv. Technol.*, no. Ngmast, pp. 37–42, 2014.

[18] S. Lortz, H. Mantel, A. Starostin, T. Bähr, D. Schneider, and A. Weber, "Cassandra: Towards a Certifying App Store for Android," *Proc. 4th ACM Work. Secur. Priv. Smartphones Mob. Devices - SPSM '14*, pp. 93–104, 2014.

[19] D. Surian, S. Seneviratne, A. Seneviratne, and S. Chawla, "App Miscategorization Detection: A Case Study on Google Play," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 8, pp. 1591–1604, 2017.

[20] M. D. Ernst *et al.*, "Collaborative Verification of Information Flow for a High-Assurance App Store," *Proc. 2014 ACM SIGSAC Conf. Comput. Commun. Secur. - CCS '14*, no. 3, pp. 1092–1104, 2014.

[21] J. Cheng, S. H. Y. Wong, H. Yang, and S. Lu, "SmartSiren: Virus Detection and Alert for Smartphones," *Proc. 5th Int. Conf. Mob. Syst. Appl. Serv. - MobiSys '07*, p. 258, 2007.

[22] B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh, "Why people hate your app," *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '13*, p. 1276, 2013.