ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Visit: www.ijirset.com*

**Vol. 8, Issue 10, October 2019**

# A Study on the Implementation Model for Ranking Fraud Detection towards Mobile Apps

**T. Sahithi[1], T. Sridevi[2]**

B.E Student, Department of CSE, Chaitanya Bharathi Institute of Technology, Hyderabad, India[1]

Associate Professor, Department of CSE, Chaitanya Bharathi Institute of Technology, Hyderabad, India[2]

**ABSTRACT:** Due to the large number of mobile Apps, ranking fraud is the key challenge in front of the mobile App market. Ranking fraud refers to fraudulent or vulnerable activities which have a purpose of bumping up the Apps in the popularity list. While the importance and necessity of preventing ranking fraud has been widely recognized. In the existing system the leading event and leading session of an app is identified from the collected historical records. Then three different types of evidences are collected from the user feedbacks namely ranking based evidence, rating based evidence and review based evidence.As we know that the mobile Apps has grown at vast speed in some years; as for march 2017, there arenearby2.8 million Apps at google play and 2.2 Apps at Apple Apps store. In addition, there are over 400,000 independent app developers all fighting for the attention of the same potential customers. This paper provides a study on the implementation model for ranking fraud detection towards mobile apps.

**KEYWORDS**: Mobile apps, ranking, App recommendation

## I. INTRODUCTION

At the end of April 2013, there are more than 1.6 million Apps at Apple's App store and Google Play. To stimulate the development of mobile Apps, many App stores launched daily App leaderboards, which demonstrate the chart rankings of most popular Apps. Indeed, the App leader- board is one of the most important ways for promoting mobile Apps. A higher rank on the leaderboard usually leads to a huge number of downloads and million dollars in revenue. Therefore, App developers tend to explore various ways such as advertising campaigns to promote their Apps in order to have their Apps ranked as high as possible in such App leaderboards.

However, as a recent trend, instead of relying on traditional marketing solutions, shady App developers resort to some fraudulent means to deliberately boost their Apps and eventually manipulate the chart rankings on an App store. This is usually implemented by using so-called "*bot farms*" or "*human water armies*" to inflate the App downloads, ratings and reviews in a very short time. For example, an article from [4] reported that, when an App was promoted with the help of ranking manipulation, it could be propelled from number 1,800 to the top 25 in Apple's top free leaderboard and more than 50,000-100,000 new users could be acquired within a couple of days. In fact, such ranking fraud raises great concerns to the mobile App industry. For example, Apple has warned of cracking down on App developers who commit ranking fraud [3] in the Apple's App store. In the literature, while there are some related work, such as web ranking spam detection, online review spam detection, and mobile, App recommendation, the problem of detecting ranking fraud for mobile Apps is still under- explored. To fill this crucial void, in this paper, we propose to develop a ranking fraud detection system for mobile Apps. Along this line, we identify several important challenges. First, ranking fraud does not always happen in the whole life cycle of an App, so we need to detect the time when fraud happens. Such challenge can be regarded as detecting the *local* anomaly instead of *global* anomaly of mobile Apps. Second, 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. Finally, due to the dynamic nature of chart rankings, it is not easy to identify and confirm the evidences linked to ranking fraud, which motivates us to discover some implicit fraud patterns of mobile Apps as evidences.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Visit: www.ijirset.com*

**Vol. 8, Issue 10, October 2019**

On daily basis, an app leaderboard can be updated by app store which displays chart rankings of most popular apps, also it is an inspirational thing to make encouraged the development of mobile apps. In fact, for promoting mobile Apps, leader board of apps is the most important way of upgradient in the market. An app should be ranked higher depend upon how its chart of development raise and progressively it can make number of downloads and ultimately high revenue in dollar. There were different ways to advertise Apps promotional drive in order to get top position in App leaderboards the legal one is white hat basis to promote their App to get famous and alternately more number of downloads. But there are also some illegal ways say black hat basis for bumping up the App by using some deceptive means used by corrupt App developers to get famous in some short time period. This technique usually implemented by using so-called "internet bots" or "human water armies" to raise the App downloads, ratings and reviews in a very little time. Some are necessary points that is to restrict fraud, showed as given two constraints. The first constraint is that an app can be rated only once from a user login and the second is implemented with the aid of IP address that limits the number of user login logged per day. Finally, the proposed system will be evaluated with real- world App data which is to be collected from the App Store for a long-time period called historical records. In the existing system, from the collected historical records, the leading event and leading session of an app is identified. There are two main steps for mining leading sessions. First, we need to discover leading events from the App's historical ranking records. Second, we need to merge adjacent leading events for constructing leading sessions. Careful observation shows that the mobile Apps are not always at top most position in leader board. But only in some time period called leading event which is form different leading sessions means ranking fraud particularly occur in this leading session. Then from the user judgmental feedback, three different types of evidences are collected namely ranking based evidence, rating based evidence and review based evidence. As our project based on evidences collected from app data; the one of the mostly judgment by people is rating based evidences which can be used to rate the app while downloading it or we can rate it after seeing its performance. It is most important evidence to judge the app. But as discussed above there are some techniques with help of which the rating can get increases by doing fraud. So, another judged evidence based technique is review based evidence; finding to make the exact specification of app whether it is good or bad app to download. In Review Based Evidences, besides ratings, most of the App stores also allow users to write some textual comments as App reviews. So, people may sure shot about downloading that specific app by reading comments specified in review section and also give their opinion about that app.

## II. LITERATURESURVEY

In paper [4] they developed a ranking fraud detection system for mobile apps. They observed that mobile apps are not always ranked high in leaderboard but only in some leading events which collectively form different leading sessions. Fraudulent apps often have different ranking patterns in each leading session compared with normal apps. Since an app's ranking can be affected by app developer's reputation and marketing campaigns like limited time discount they also collected fraud evidences based on apps rating and review history. The evidences were extracted by modeling apps' ranking, rating and review behaviors through statistical hypotheses tests. Unsupervised evidence- aggregation method was used to integrate the three types of evidences. The proposed system was evaluated system with real-world app data collected from the Apple's App store for a long period, of about two years. Experimental results obtained showed the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity of ranking fraud activities.

In paper [5], they proposed a sequential approach based on Hidden Markov Model(HMM) to model the heterogeneous popularity information for mobile apps including chart ranking, user rating and review topics extracted from user reviews. It also aims to provide a comprehensive modeling of popularity information for mobile app services. The popularity information is modelled in terms of different transitions of different popularity states, which indicate the latent semantics and relationships of popularity observations. Popularity information include chart rankings, user

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

ratings and user reviews. The PHMM model can be used to build other applications like trend based mobile app recommendation, rating and review spam detection and ranking fraud detection.

In paper [6], The proposed model in this system first preprocess the reviews and tokenizes the words. It then converts words into its root word and removes those words which are not required. Scores are given to the words extracted followed by which scores are given to emoticons. All possible emoticons are stored in a dictionary along with its score. The overall score is used to classify if the reviews are positive or negative. Evidences were then aggregated, and results obtained told if the app had a ranking fraud or not.

In paper [7], they studied about malware and search rank fraud subjects in Google Play. They this by discovering and leveraging traces left behind by fraudsters. To detect both malware and apps subjected to search rank fraud they built a system called FairPlay. Co-review graphs were used to model reviewing relations between users. Temporal dimensions of review post times were used to identify suspicious review spikes received by apps. Apps with unbalanced review, rating and install counts as well as apps with permission request ramps were identified. Instead of analyzing app executables, linguistic and behavioral information was used to detect genuine reviews from which user identified fraud and malware were extracted.

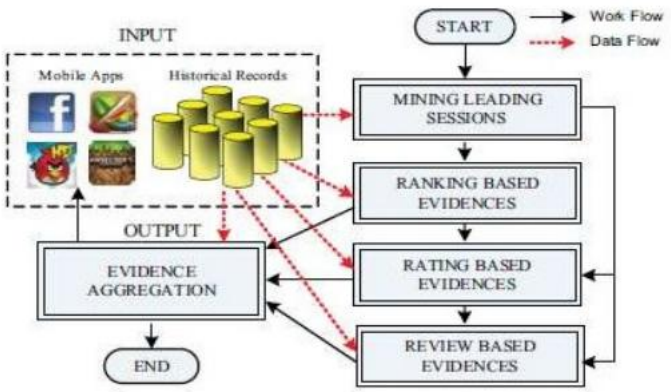


**Figure 1: Model for ranking fraud detection for mobile apps**

In paper [5], they performed a detailed temporal analysis on two datasets that they collected from the Google Play Store, and other newly released apps. They have tried to shed light on the dynamics of Google Play. They developed iMarket, a prototype app market crawling system to collect data. They developed IP blocking system which blocks the reviews from same IP after standard count of reviews from that IP address. The longitudinal study of Google Play app metadata provides unique information that is not available through the standard approach of capturing data. Features extracted from longitudinal app analysis like permission price, update and download count changes provide insights into fraudulent app promotion and malware indicator behavior.

This is different from ranking fraud detection for mobile Apps. They categorize all existing algorithms into three categories based on the type of information they use: content-based methods, link-based methods, and methods based on non-traditional data such as user behavior, clicks, HTTP sessions. In turn, there is a sub categorization of link-based category into five groups based on ideas and principles used: labels propagation, link pruning and reweighting, labels refinement, graph regularization, and feature based. The second category is focused on detecting online review spam. For example, [4] have identified several representative behaviors of review spammers and model these behaviors to detect the spammers. This paper aims to detect users generating spam reviews or review spammers. They identify several characteristic behaviors of review spammers and model these behaviors so as to detect the spammers. In

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

particular, authors seek to model the following behaviors. First, spammers may target specific products or product groups in order to maximize their impact. Second, they tend to deviate from the other reviewers in their ratings of products. They propose scoring methods to measure the degree of spam for each reviewer and apply them on an Amazon review dataset. Authors then select a subset of highly suspicious reviewers for further scrutiny by user evaluators with the help of a web based spammer evaluation software specially developed for user evaluation experiments. [5] have studied the problem of detecting hybrid shilling attacks on rating data. The proposed approach is based on the semi-supervised learning 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 Bayes (SNBλ) to precisely separate Random-Filler model attackers and Average-Filler model attackers from normal users. [6] have studied the problem of singleton review spam detection. Specifically, they solved this problem by detecting the co-anomaly patterns in multiple review based time series. Although some of above approaches can be used for anomaly detection from historical rating and review records, they are not able to extract fraud evidences for a given time period (i.e., leading session). Finally, the third category includes the studies on mobile App recommendation. For example, [7] developed a mobile App recommender system, named Appjoy, which is based on user‟s App usage records to build a preference matrix instead of using explicit user ratings.

## III. PROPOSED SYSTEM

There were so many investigations run on fraud finding areas such as for web ranking spam detection, online review spam detection and portable App awareness. Here we develop some fraud finding activity over fraud Apps, which are made by fraud App developers for leading their App ranking in leaderboard. Firstly, we take an account over time when fraud happens so that we can easily make sure about the fraudness of the App. So, detection of such App is conduct by making leading sessions of small leading events that shown when the Apps are in phases of achievement. Those are rising phase, maintaining phase and recession phase where we detect App ranking behavior from historical ranking records. Such a demonstration done over small start, we have called local anomaly detection strategy. When these phases run over App historical records the liable Apps can maintain their level of ranking constantly over long period, but the fraud Apps found fluctuation over that time and find fraud. But some App developers run the wrong way to make the other Apps developer to downfall by ranking wrongly to their App. So, there are more fraud finding evidences such as rating and review based evidences. The users who are newly logging to the app stores, they decide based on the existing ranking, rating, reviews for the individual apps.User do not understand about fake Apps and may download it. As review, also may be fake but we have conduct one semantic analysis test here on reviews by conducting test as Natural language processing to make the overall comments judge and get some final sure comment to make trust over the App by customer who downloaded it.

## IV. IMPLEMENTATION

### Identifying Leading Sessions

Ranking fraud usually happens in leading sessions. 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. Then, with the analysis of Apps' ranking 'behaviors, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps.

Mining Leading Sessions: There are two main steps for mining leading sessions. First, we need to discover leading events from the App's historical, ranking records. Second, we need to merge adjacent leading events for constructing leading sessions.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Visit: www.ijirset.com*

**Vol. 8, Issue 10, October 2019**

**Ranking Based Evidences**

A leading session is composed of several leading events. Therefore, we should first analyze the basic characteristics of leading events for extracting fraud evidences. By analyzing the Apps' historical ranking records, we observe that Apps' ranking behaviors in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase. Specifically, in each leading event, an App's ranking first increases to a peak position in the leader board (i.e., rising phase), then keeps such peak position for a period (i.e., maintaining phase), and finally decreases till the end of the event (i. e., recession phase).

**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. Specifically, after an App has been published, it can be rated by any user who downloaded it. Indeed, user rating is one of the most important features of App advertisement. An App which has higher rating may attract more users to download and can also be ranked higher in the leader board. Thus, rating manipulation is also an important perspective of ranking fraud. Intuitively, if an App has ranking fraud in a leading session s, the ratings during the time period of s may have anomaly patterns compared with its historical ratings, which can be used for constructing rating 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. Indeed, review manipulation is one of the most important perspectives of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users often firstly 5, read its historical reviews to ease their decision making, and a mobile App contains more positive reviews may attract more users to download. Therefore, imposters often post fake reviews in the leading sessions of a specific App in order to inflate the App downloads, and thus propel the App's years, the problem of detecting the local anomaly of reviews in the leading sessions and capturing them as evidences for ranking fraud detection are still under-explored.

## V. CONCLUSION

Specifically, we first showed that ranking fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. Then, we identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. Then, we identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. Moreover, we pro- posed an optimization based aggregation method to integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps. An unique perspective of this approach is that all the evidences can be modeled by statistical hypothesis tests, thus it is easy to be extended with other evidences from domain knowledge to detect ranking fraud. This paper provided a study on the implementation model for ranking fraud detection towards mobile apps

## REFERENCES

1. Raju Dara, Pasupuleti Anil sai , International Journal of Modern Computer Science and Applications (IJMCSA) , Dept.of CSE, JNTUH, "Leading Session Mining and Evidence Based Approach for Ranking Fraud Detection of Mobile Applications", Volume No.-5, Issue No.-1, January, 2017
2. Pingale, L. Kuhile, P. Phapale, P. Sapkal, S. Jaiswal. " Fraud detection & prevention of mobile apps using optimal aggregation method." International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE). 2016; 6(3), 491- 497.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Visit: www.ijirset.com*

**Vol. 8, Issue 10, October 2019**

3. 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, 2010.
4. Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li. Supervised rank aggregation. In Proceedings of the 16th international conference on World Wide Web, WWW '07, pages 481–490, 2007.
5. Malyadri. K, "An Overview towards the Different Types of Security Attacks", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 8, August 2014
6. Malyadri. K, "Security Threats, Security Vulnerabilities and Advance Network Security Policies", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 2, Issue 9, September 2013
7. Malyadri. K, "Need for Key Management in Cloud and Comparison of Various Encryption Algorithm", International Journal of Scientific Research in Computer Science, Engineering and Information Technology , volume 1, issue 1, July-August 2016
8. Malyadri. K, "Integration of Appropriate Analytic tools towards Mobile Technology Development", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 6, Issue 6, June 2018
9. Malyadri. K,"A STUDY ON EXPERIENCES ANDLIMITATIONS OF MOBILE COMMUNICATION", Alochana Chakra Journal, Volume VI, Issue VIII, August 2017
10. Malyadri. K,"CHALLENGES CONCERNING MOBILEDEVELOPMENT AND MODEL-DRIVEN DEVELOPMENT OF MOBILE APPS", Airo International Research Journal, volume XVI, Nov 2018
11. Malyadri, N. Surya Teja, "Related technologies and the role of mobile app development life cycle", International Journal of Research and Applications, Volume 5, Issue 17, Jan-Mar 2018.
12. Rakesh Rojanala, "A STUDY ON BIG DATA-ORIENTED STREAM DATA MINING AND STREAMED DATA MANAGEMENT", The International journal of analytical and experimental modal analysis, Volume X, Issue III, March 2018
13. Rakesh Rojanala, "Generic Working of an Artificial Neuron and Its Output Mathematical Representation", "International Journal of Innovative Research in Science, Engineering and Technology", Vol. 8, Issue 1, January 2019
14. Yeshwanth Valaboju, "A LITERATURE REVIEW ON NEURAL NETWORK ARCHITECTURES", Journal of Interdisciplinary Cycle Research, Volume VII, Issue II, December 2015
15. Yeshwanth Valaboju, "Routing and Vulnerabilities in MANETS", "International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)", Volume 1, Issue 2, December 2018
16. ANUMANDLA MOUNIKA, "THREATS, OPPORTUNITIES OF THE CLOUD AND PROVISION OF APPLICATION SERVICES", JASC: Journal of Applied Science and Computations, Volume 2, Issue 1, Jan-June 2015
17. ANUMANDLA MOUNIKA, "DATA SECURITY IN THE CLOUD", The International journal of analytical and experimental modal analysis, Volume 1, Issue 4, July-December-2012
18. ANUMANDLA MOUNIKA, "CLOUD COMPUTING INFRASTRUCTURE AND CLOUD ADOPTION CHALLENGES", Journal of Interdisciplinary Cycle Research, Volume VI, Issue II, July-December 2014
19. ANUMANDLA MOUNIKA, "AN OVERVIEW ON THE ARCHITECTURAL COMPONENTS OF CLOUD", International Journal of Research, Volume 6, Issue 12, December 2017
20. ANUMANDLA MOUNIKA, "PROCESS OF MIGRATING INTO A CLOUD AND ISSUES IN CLOUD COMPUTING", Journal of Interdisciplinary Cycle Research, Volume 2, Issue 1, January-June-2010
21. A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh. Spotting opinion spammers using behavioral footprints. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '13, 2013.
22. Ntoulas, M. Najork, M. Manasse, and D. Fetterly. Detecting spam web pages through content analysis. In Proceedings of the 15th international conference on World Wide Web, WWW '06, pages 83–92, 2006.