



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 12    **Issue:** III    **Month of publication:** March 2024

**DOI:** <https://doi.org/10.22214/ijraset.2024.59035>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Fraud and Malware Detection of Apps in Google Play

Arun Kumar R<sup>1</sup>, Vijayaselvadurai G<sup>2</sup>, Mohammed Aarif Hussain K<sup>3</sup>, Raghavan R<sup>4</sup>

<sup>1, 2, 3, 4</sup> Dept of Computer Science and Business Systems, Sethu Institute of Technology, Pulloor, Kariapatti – Virudhunagar 626 115

**Abstract:** *Fraudulent Behaviors in Google Play, the Most Popular Android App Market, Fuel Search Rank Abuse and Malware Proliferation. To Identify Malware, Previous Work Has Focused on App Executable and Permission Analysis. In This Paper, We Introduce FairPlay, a Novel System That Discovers and Leverages Traces Left Behind by Fraudsters, to Detect Both Malware and Apps Subjected to Search Rank Fraud. FairPlay Correlates Review Activities and Uniquely Combines Detected Review Relations with Linguistic and Behavioral Signals Gleaned from Google Play App Data (87K Apps, 2.9M Reviews, and 2.4M Reviewers, Collected Over Half a Year), in Order to Identify Suspicious Apps. FairPlay Achieves Over 95% Accuracy in Classifying Gold Standard Datasets of Malware, Fraudulent and Legitimate Apps. We Show That 75% of the Identified Malware Apps Engage in Search Rank Fraud. FairPlay Discovers Hundreds of Fraudulent Apps That Currently Evade Google Bouncer's Detection Technology. FairPlay Also Helped the Discovery of More Than 1,000 Reviews, Reported for 193 Apps, That Reveal a New Type of "Coercive" Review Campaign: Users Are Harassed into Writing Positive Reviews, and Install and Review Other Apps.*

**Keywords:** *Fair Play, malware, fraudulent apps, Google Play, review activities, search rank fraud.*

## I. INTRODUCTION

Shady App Developers Are Resorting to Fraudulent Tactics to Manipulate Chart Rankings on App Stores, Employing Methods Like "Bot Farms" or "Human Water Armies" to Inflate Downloads, Ratings, and Reviews Rapidly. This Practice Raises Significant Concerns in the Mobile App Industry, With Apple Warning of Crackdowns on Ranking Fraud. Detecting Such Fraud Poses Challenges, as It Involves Identifying Local Anomalies Within Leading Sessions of Mobile Apps. to Address This, a Proposed System Focuses on Automatically Detecting Ranking Fraud Using Historical Ranking, Rating, and Review Data. by Analyzing App Behaviors and Extracting Fraud Evidence, Including Ranking Patterns and Anomaly Patterns in Ratings and Reviews, the System Aims to Evaluate the Credibility of Leading Sessions. Evaluation With Real-World App Data Demonstrates the Effectiveness and Scalability of the Proposed Approach in Detecting Ranking Fraud Activities.

## II. PROBLEM IDENTIFICATION

With the increasing popularity of mobile applications, app marketplaces such as Google Play have become crucial platforms for users and developers. However, the prevalence of Search Rank Fraud and Malicious App Installations poses significant threats to the integrity and security of these platforms. Search Rank Fraud involves manipulating the ranking algorithms to artificially boost the visibility of apps, while Malware Detection aims to identify and prevent the distribution of harmful applications. Despite efforts by app marketplaces and researchers, the dynamic nature of fraudulent activities and the evolving tactics employed by malicious actors present ongoing challenges. There is a pressing need for robust and effective methodologies to detect Search Rank Fraud and Malware in GooglePlay.

## III. OBJECTIVES

- 1) *Objective:* Understand the ecosystem of the Android app market on Google Play, involving users and developers with Google accounts.
- 2) *Functionality:* Investigate the components of apps, including executables (APKs), permissions, descriptions, reviews, ratings, and install counts.
- 3) *Versatility:* Examine the motivations and methods of both malicious and fraudulent developers, including tampering with search rankings through fake reviews, ratings, and installs.
- 4) *Purpose:* Explore the impact of search rank fraud on app visibility and installs, benefiting both fraudulent developers seeking revenue and malicious developers seeking to spread malware.



- 1) *Data Retrieval*: Reviews are collected and sent for sentiment analysis.
- 2) *Preprocessing*: Reviews undergo preprocessing, which might involve removing irrelevant information or formatting the text for further analysis.
- 3) *Sentiment Analysis*: Extracts sentiment (positive, negative, or neutral) from the reviews.
- 4) *Keyword Management*: Keywords are identified and potentially filtered based on a defined criteria (e.g., removing generic words or obscenities).
- 5) *Database Integration*: Extracted keywords and sentiment information are stored in a database, possibly alongside the original review text.

## VI. ADVANTAGES

- 1) Aim to improve detection accuracy by reducing false positives and negatives.
- 2) Implement real-time detection to promptly respond to emerging threats.
- 3) Design the system to adapt quickly to new tactics employed by fraudsters and malware developers.
- 4) Ensure regular updates and continuous monitoring to stay ahead of emerging threats.

## VII. CONCLUSION

In conclusion, We have introduced FairPlay, a system to detect both fraudulent and malware Google Play apps. Our 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, we showed FairPlay's ability to discover hundreds of apps that evade Google Play's detection technology, including a new type of coercive fraud attack.

## REFERENCES

- [1] L. Azzopardi, M. Girolami, and K. V. Risjbergen, "Investigating the relationship between language model perplexity and its precision-recall measures," in Proc. 26th Int. Conf. Res. Develop. Inform. Retrieval, 2003, pp. 369–370.
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., pp. 993–1022, 2003.
- [3] Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "A taxi driving fraud detection system," in Proc. IEEE 11th Int. Conf. Data Mining, 2011, pp. 181–190.
- [4] D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclear norm minimization," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 60–68.
- [5] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.
- [6] G. Heinrich, Parameter estimation for text analysis, "Univ. Leipzig, Leipzig, Germany, Tech. Rep., <http://faculty.cs.byu.edu/~ringger/CS601R/papers/Heinrich-GibbsLDA.pdf>, 2008.
- [7] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219–230.
- [8] J. Kivinen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," in Proc. 27th Annu. ACM Symp. Theory Comput., 1995, pp. 209–218.





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)