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

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Fraud and Malware Detection of Apps in Google Play

Arun Kumar R[1], Vijayaselvadurai G[2], Mohammed Aarif Hussain K[3], Raghavan R[4]
[1, 2, 3, 4] *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.

IV. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things r satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration r taken into account for developing the proposed system.

A. Search Rank Fraud and Malware Detection in Google Play

Malicious activities in Google Play, the most widely used Android app store, encourage the manipulation of search ranks and the spread of malware. Previous research has concentrated on programme executable and permission analysis to discover malware. We demonstrate that search rank fraud is used by 75% of the malware applications that have been found. Just Play finds hundreds of phoney applications that are now avoiding the detection methods of Google Bouncer. We are pleased to present Fair Play, a technology that can distinguish between malicious and fraudulent Google Play apps. Our research on a recently submitted longitudinal app dataset has demonstrated that a significant portion of malware is responsible for search rank deception; Fair Play correctly detects both types of activity.

B. Android Malware Detection

Mobile applications are increasingly being used to support critical domains such as health, logistics, and banking, to name a few. These mobile apps, hence, became a target for malware attackers. An-droid is an open-source operating system, which runs apps that can be downloaded from official or third-party app stores. Malware exploits these applications to penetrate mobile devices in different ways for different purposes. To address this, different approaches for malware analysis have been proposed for the detection of malware, ranging from pre-installation to post-installation. This paper presents a literature review of recent malware detection approaches and methods. 21 prominent studies, that report three most common approaches, are identified and reviewed. Challenges, limitations, and research directions are identified and dis-cussed. Findings show most studies focus on malware classification and detection, but lack studies that investigate securing apps and detecting vulnerabilities that malware exploits to stealth into mobile apps and de-vices. They also show that most studies focused on enhancing machine learning models rather than the malware analysis process.

C. A Comparative Analysis of Search Engine Ranking Algorithms

Ranking Algorithm is the most proper way of positioning on a scale. As the information and knowledge on the internet are increasing every day.The search engine's ability to deliver the most appropriate material to the customer. It is more and more challenging without even any assistance in filtering through all of it. However, searching what user requires is extremely difficult. In this research, an effort has been made to compare and analyze the most popular and effective search engines.The keywords were used in uniform resource locator like, title tag, header, or even the keyword's resembles to the actual text.The page rank algorithm computes a perfect judgment of how relevant a webpage is by analyzing the quality and calculating the number of links connected to it. In this study the keyword relevancy and time response were used for search engines and observed the results. It is observed that the google search engine is faster than the bing and youtube, and after all, bing is the best search engine after google. Moreover, youtube is the fastest search engine in terms of video content search. The google results were found more accurate. However, it is better than all of the search engines.

V. BLOCK DIAGRAM & CIRCUIT DIAGRAM

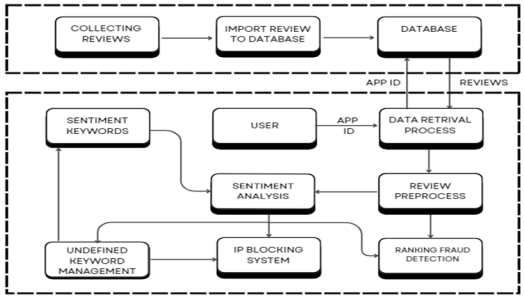


Fig. 1. Block Diagram of system

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

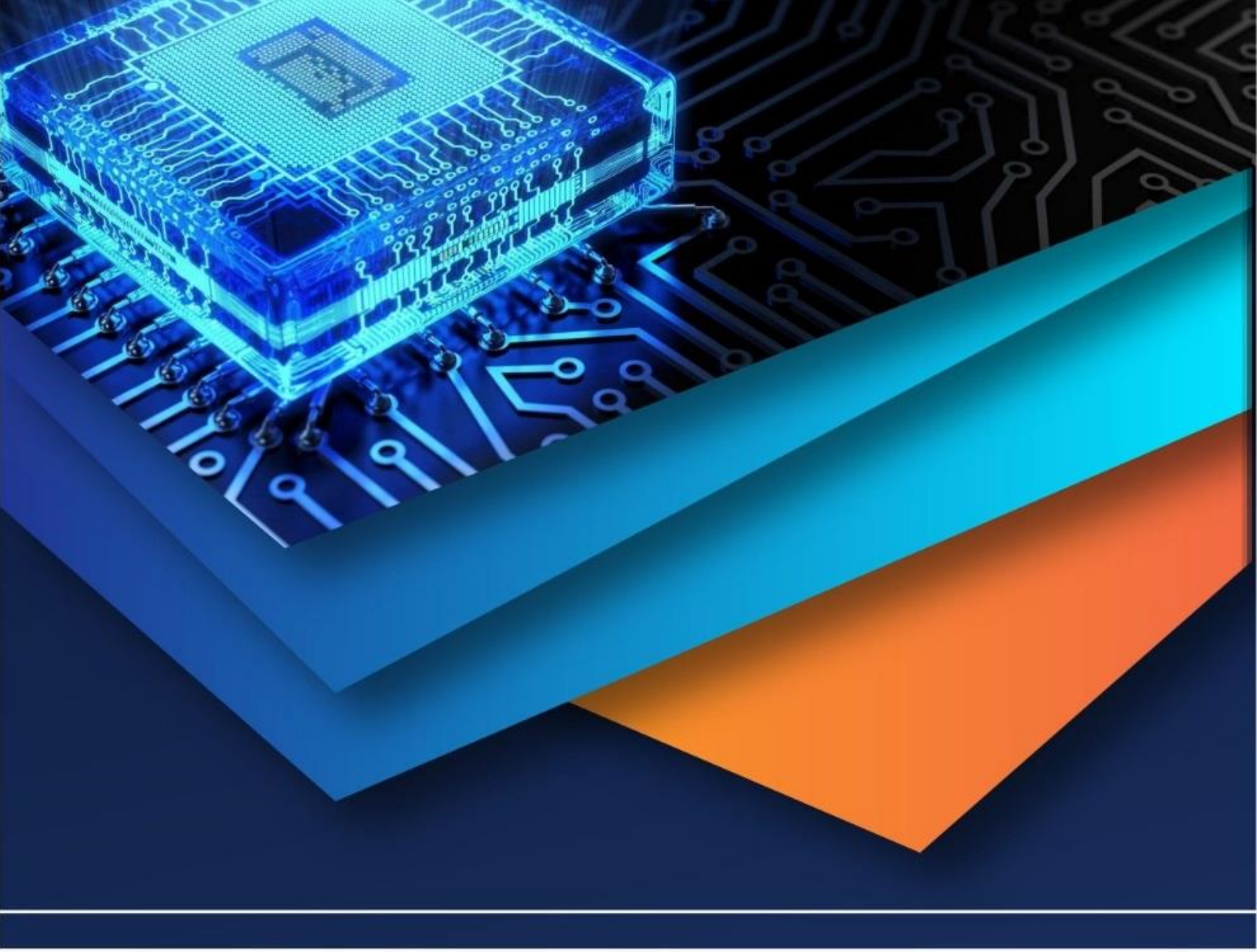
1) Aim to improve detection accuracy by reducing false positives and negatives.
2) Implement real-time detection to promptly respond to emerging threats.
3) Design the system to adapt quickly to new tactics employed by fraudsters and malware developers.
4) Ensure regular updates and continuous monitoring 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 ir precision-recall measures," inProc. 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," inProc. IEEE 11th Int. Conf. Data Mining, 2011, pp. 181–190.
[4] D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclear norm minimization," inProc. 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," inProc. 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