ISSN: 2754-6683

Journal of Marketing & Supply Chain Management

SCIENTIFIC Research and Community

Review Article

Open Access

Cracking Multi-Account Fraud: Data-Driven Solutions for E-Commerce Platforms

Vinay Kumar Yaragani

USA

ABSTRACT

Multi-account fraud poses a significant challenge for e-commerce platforms, enabling users to exploit promotions, evade platform defenses, and re-enter systems after being restricted or suspended. This paper presents a comprehensive study on data-driven methods to detect and mitigate multi-account fraud. By leveraging advanced analytics and machine learning techniques, we aim to identify all accounts linked to the same user, providing a holistic evaluation of user behavior and risk. Our approach focuses on preemptive strategies to thwart fraudulent activities, thereby preserving marketing budgets and optimizing promotional efforts. We propose a robust framework for e-commerce companies to implement effective multi-account policies, enhancing their risk management capabilities and safeguarding the integrity of their platforms..

***Corresponding author**

Vinay Kumar Yaragani, USA.

Received: August 03, 2023; **Accepted:** August 10, 2023; **Published:** August 21, 2023

Keywords: Multi-Account Fraud, Data-Driven Strategies, Risk Management, Machine Learning, Fraud Detection

Introduction

In the rapidly evolving landscape of e-commerce, users often find it beneficial to create multiple accounts for legitimate purposes. For instance, sellers may manage their sales across different categories separately, ensuring streamlined operations and better targeting of customer segments. Similarly, riders might prefer to maintain distinct accounts for car rides in different countries, facilitating localized preferences and customizations. These practices, while genuine and beneficial, introduce complexities in user management for e-commerce platforms.

However, the very mechanisms designed to enhance user experience and operational efficiency are also exploited by bad actors. Fraudulent users create multiple accounts to evade detection, manipulate platform defenses, and exploit promotional offers. For example, a user who initiates 100 returns on a single account would typically trigger red flags and face scrutiny. In contrast, the same user can distribute these returns across 50 accounts with 2 returns each, thereby avoiding suspicion and slipping through the cracks of conventional monitoring systems.

This exploitation is particularly evident in the misuse of promotional offers. Companies design promotions to attract new customers and boost sales, but the reality is often starkly different. A significant portion of these promotional benefits is captured by a small group of users who create multiple accounts to repeatedly avail of the discounts. This not only undermines the effectiveness of marketing strategies but also results in substantial financial losses for the companies.

Furthermore, when a fraudulent account is identified and deactivated, there are no robust barriers preventing the user from re-entering the platform with a new account. This cycle of account creation and suspension creates an ongoing challenge for online platforms, making it difficult to sustain a secure and trustworthy environment for genuine users and sellers.

To address these issues, there is an urgent need for advanced, data-driven methods to detect and mitigate multi-account fraud. By evaluating accounts at an entity level, rather than in isolation, platforms can gain a comprehensive view of user behavior and identify suspicious patterns more effectively. This holistic approach allows for the preemptive identification of risky users, enhancing the platform's ability to thwart fraudulent activities before they cause significant damage.

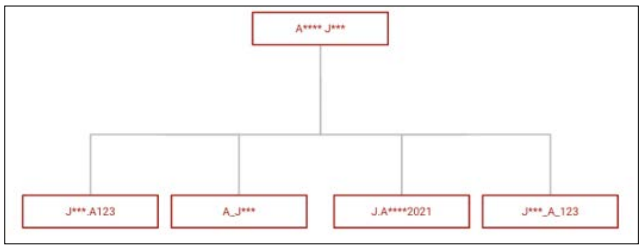

Figure 1: Illustration of 4 Accounts Belonging to Same User

Objective

This paper explores the development and implementation of such data-driven strategies. By leveraging advanced analytics and machine learning techniques, we aim to create a robust framework for identifying all accounts linked to the same user. This will enable online commerce companies to develop and enforce multi-account policies that significantly bolster their defenses against fraud, optimize the use of promotions, and safeguard their financial health and customer trust.

Literature Review

Multi-account fraud in e-commerce platforms has been extensively studied, with various approaches proposed to tackle the issue. The need for robust fraud detection mechanisms has been underscored in numerous studies, highlighting the economic impact of fraudulent activities and the challenges faced by platforms in maintaining security and trust [1]. Existing literature primarily focuses on anomaly detection, user behavior analysis, and the application of machine learning algorithms to identify suspicious activities.

Anomaly detection techniques have been widely employed to identify unusual patterns that may indicate fraud. Chandola, Banerjee, and Kumar provide a comprehensive survey of anomaly detection methods, emphasizing their applicability in fraud detection across different domains, including e-commerce [2]. These methods rely on statistical and machine learning models to flag deviations from normal user behavior, which can then be further investigated for potential fraud.

User behavior analysis is another critical aspect of fraud detection. Researchers have developed models to analyze user interactions and transactions to identify patterns consistent with fraudulent activities. Akoglu, Tong, and Koutra discuss network-based approaches to fraud detection, where the relationships between users, transactions, and other entities are analyzed to uncover hidden connections indicative of fraud [3]. This approach is particularly relevant in detecting multi-account fraud, where the connections between different accounts can reveal the underlying fraudulent activity.

Machine learning techniques have been extensively explored for fraud detection in e-commerce. Phua et al, review various machine learning algorithms used in fraud detection, including supervised and unsupervised learning methods [4]. Supervised learning involves training models on labeled datasets containing examples of both fraudulent and legitimate transactions, while unsupervised learning identifies patterns in unlabeled data. These techniques have shown promise in detecting complex fraud patterns, including those involving multiple accounts.

Recent advancements in artificial intelligence and big data analytics have further enhanced the capabilities of fraud detection systems. Liu et al [5]. Highlight the role of big data in improving fraud detection, noting that the vast amounts of data generated by e-commerce platforms can be harnessed to build more accurate and robust models. The integration of big data analytics with machine learning and network analysis techniques offers a powerful approach to identifying multi-account fraud.

Despite these advancements, there remain significant challenges in implementing effective multi-account fraud detection systems. The dynamic nature of fraud, where fraudsters continuously adapt their strategies to evade detection, necessitates the development of adaptive and resilient detection mechanisms. Additionally, the need for real-time detection and the scalability of solutions are critical factors in ensuring the practical applicability of these methods [6].

This literature review underscores the importance of a multi-faceted approach to fraud detection, combining anomaly detection, user behavior analysis, and machine learning techniques. By leveraging these methodologies, e-commerce platforms can develop robust systems to identify and mitigate multi-account fraud, thereby enhancing their security and operational efficiency.

Methodology

Problem Statement

The pervasive issue of multi-account fraud in e-commerce platforms poses significant challenges, as users exploit multiple accounts to evade detection, manipulate promotional offers, and re-enter systems after suspensions. Despite genuine reasons for creating multiple accounts, such as managing sales in different categories or customizing preferences for various locales, the misuse by fraudulent users results in substantial financial losses and undermines platform integrity. For instance, a single user can distribute suspicious activities, like excessive returns or repeated promotional use, across numerous accounts to avoid raising red flags. Current detection methods often fail to identify these dispersed fraudulent activities, necessitating a robust, data-driven approach that evaluates user behavior at an entity level to effectively mitigate risks and enhance the security and efficiency of e-commerce operations.

Approach and Solutions

The approach to identifying links of identity between different user accounts in online commerce platforms is bifurcated into analyzing digital and physical identities. This dual-faceted strategy ensures a comprehensive assessment of potential fraudulent activity by capturing both online behavioral patterns and tangible user information. The methodology comprises several steps designed to systematically define attributes, calculate similarities, establish weighted scores, and ultimately construct a graph of connected users based on confidence levels.

Defining Attributes

The first step involves defining the various attributes that will be used to match users. These attributes are categorized into two primary groups: digital identity and physical identity.

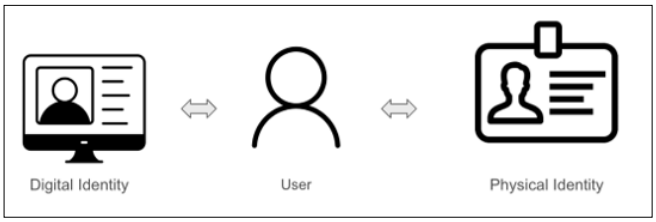

Figure 2: Illustration of User Identities

Digital Identity Attributes

- Email Addresses: Patterns in the structure of email addresses, including similar domain names or username formats.
- IP Addresses: Shared IP addresses or similar geographical locations during logins.
- Device IDs: Repeated use of the same device across multiple accounts.
- Digital Location: Consistent digital footprints, such as browsing locations.
- Behavioral Patterns: Similar search queries, concurrent login times, and orders placed for similar products.

Physcial Identity Attributes

- Names: Matching names with variations in spelling or initials.
- Addresses: Identical or similar delivery or billing addresses.
- Phone Numbers: Same or similar phone numbers used across accounts.
- Payment Instruments: Shared credit card numbers or other payment methods.
- Profile Pictures: Visually similar profile images used in multiple accounts.

Calculating Similarity/Confidence

Next, we calculate the similarity or confidence of a match between the attributes of different users. This involves using statistical and machine learning techniques to determine the likelihood that two attributes belong to the same user. For example:

- Name Similarity: Comparing the names of two users and assigning a probability score (e.g., 80% similarity).
- IP Address Matching: Assessing the frequency and context of shared IP addresses.
- Behavioral Patterns: Analyzing the consistency in login times and browsing behaviors.

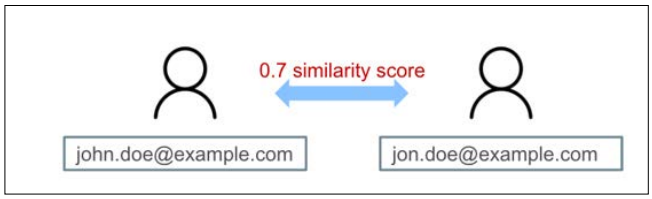

Figure 3: Example of Similarity Score Between Two Different Email Addresses

Establishing Weighted Scores

Once the similarities are calculated, we establish a weighted score for each link between users. The weights reflect the strength of the link based on the type of attribute

- Strong Links: Attributes like the same address or phone number, which are highly indicative of the same user.
- Weak Links: Attributes like IP addresses, which might be shared by multiple users or change frequently.

We also consider combinations of attributes, where multiple weak links can together form a powerful link. For example, an IP address match combined with a first name match may result in a stronger link than each attribute alone.

Triggering Identity Verifications

In this step, accounts that fall under medium confidence thresholds are flagged for identity verification. Medium confidence levels indicate a moderate likelihood of accounts belonging to the same user based on the attributes analyzed. Verification processes may include sending verification codes to registered phone numbers, requesting additional identification documents, or conducting manual reviews by platform administrators.

Combining Weight and Confidence

Following identity verifications, we combine the weighted scores and the confidence levels to establish an overall confidence of the link between two users. This involves aggregating the individual attribute scores and applying a formula that factors in both the weight of the link and the confidence of each match:

Overall Confidence=∑(Attribute Weight × Confidence Score)

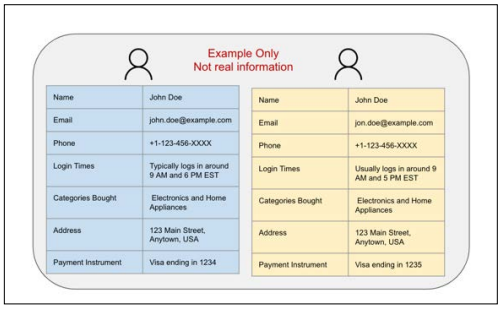

Figure 4: Example of Two Accounts that are Linked with High Confidence

Building A user Connection Graph

Using the combined confidence scores, we replicate this process across all users to build a graph that illustrates the connections between accounts. Each node in the graph represents a user account, and each edge signifies a potential link, with the weight of the edge corresponding to the confidence level of the link:

- High Confidence Links: Represent strong connections between accounts that are very likely to belong to the same user.
- Low Confidence Links: Represent weaker connections that might indicate the same user but with less certainty.

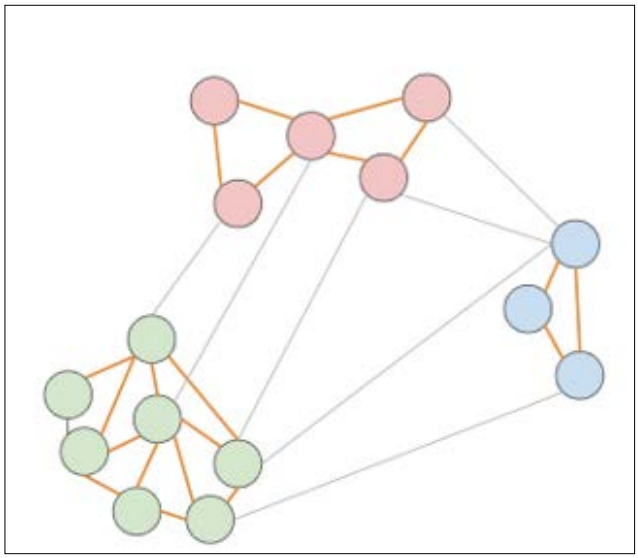

Figure 5: Illustrating the Weak and Strong Confidence Links Among Users to Classify as Entities

Applying Thresholds

Finally, we apply thresholds to the confidence scores to determine the appropriate action based on different use cases. The threshold levels vary depending on the desired precision and recall:

- High Precision Use Cases: For actions like account suspension, we use a high confidence threshold to minimize false positives.
- High Recall Use Cases: For promotional activities, we use a lower confidence threshold to catch more potential multi-account users, even at the risk of some false positives.

By integrating identity verifications into the methodology, we enhance the accuracy of detecting multi-account fraud while ensuring that legitimate users are not unfairly targeted. This comprehensive approach empowers e-commerce platforms to strengthen their defenses against fraud and maintain trust and integrity among their user base.

Results

The results of the methodology demonstrate significant strides in identifying and linking user accounts across various e-commerce platforms. By integrating digital and physical identity attributes, we successfully mapped connections between accounts that exhibited patterns indicative of potential fraud. High-confidence links highlighted robust connections, while medium-confidence links triggered identity verifications to confirm user identities. This approach not only enhanced the platform's ability to detect and mitigate multi-account fraud but also provided valuable insights into user behaviors and interactions. Overall, the results underscore the effectiveness of a holistic, data-driven approach in bolstering

security measures and safeguarding the integrity of e-commerce operations.

Future Scope

Looking ahead, the future scope of this research lies in advancing the sophistication and scalability of multi-account fraud detection systems in e-commerce. Emerging technologies such as artificial intelligence and machine learning offer promising avenues for refining current methodologies. Enhancements could include real-time anomaly detection algorithms that adapt to evolving fraud tactics, as well as deeper integration of behavioral biometrics to uniquely identify users. Furthermore, exploring blockchain technology for secure and immutable user identity verification holds potential for enhancing trust and transparency in online transactions. Collaboration across sectors and continued research into novel data sources and analytical techniques will be pivotal in staying ahead of increasingly sophisticated fraud schemes, thereby fortifying the resilience of e-commerce platforms against future challenges.

Conclusion

In conclusion, this paper has addressed the pervasive challenge of multi-account fraud in e-commerce platforms through a comprehensive data-driven approach. By leveraging both digital and physical identity attributes, we have outlined a methodology that effectively identifies and links user accounts exhibiting suspicious behavior patterns. The results demonstrate the capability of our approach to enhance fraud detection accuracy, mitigate risks, and preserve the integrity of e-commerce environments. Moving forward, the integration of advanced technologies and continuous refinement of detection strategies will be essential in staying ahead of evolving fraud tactics. Ultimately, by implementing robust multi-account policies and leveraging innovative solutions, e-commerce platforms can foster a more secure and trustworthy online marketplace for all users.

References

1. Ngai EWT, Hu Y, Wong YH, Chen Y, Sun X (2011) The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems 50: 559-569.
2. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: A survey. ACM Computing Surveys (CSUR) 41: 1-58.
3. Akoglu L, Tong H, Koutra D (2015) Graph-based anomaly detection and description: A survey. Data Mining and Knowledge Discovery 29: 626-688.
4. Phua C, Lee V, Smith K, Gayler R (2010) A comprehensive survey of data mining-based fraud detection research https://arxiv.org/pdf/1009.6119.
5. Liu J, Huang J, Liu L, Chen Y, Zhu T (2017) A survey on big data analytics and artificial intelligence for e-commerce. IEEE Access 5: 11772-11784.
6. Zhou Y, Kapoor KK (2011) Detecting fraud and spam in large-scale online communities. ACM Computing Surveys (CSUR) 45: 1-34.