International Journal of Information and Cybersecurity

Int. J. Inf. Cybersec.-2023

AI-based Strategies in Combating Ad Fraud in Digital Advertising: Implementations, and Expected Outcomes

Shobhit Agrawal
Sr. Software Engineer – Meta (Facebook)
https://orcid.org/0009-0000-4957-5575

Swapna Nadakuditi
Sr. IT BSA, Florida blue, Jacksonville FL
https://orcid.org/0009-0005-2188-5340

Abstract

The digital advertising industry faces significant challenges due to ad fraud, which encompasses various deceptive practices such as click fraud, domain spoofing, ad injection, pixel stuffing, forced redirect ads, and SDK spoofing. These fraudulent activities lead to financial losses for advertisers and undermine the effectiveness of their campaigns. This research aims to investigate the application of artificial intelligence (AI) techniques to combat ad fraud and discuss five AI-based strategies, their implementations, and potential outcomes. The proposed strategies include: 1) anomaly detection and behavioral analysis, 2) domain verification and network analysis, 3) real-time monitoring and ad content analysis, 4) SDK analysis and app attribution modeling, and 5) collaborative filtering and industry collaboration. Each strategy uses AI algorithms and machine learning models to identify and mitigate fraudulent activities in different aspects of the digital advertising ecosystem. The implementation of these strategies involves training AI models to detect anomalies in ad traffic patterns, analyze user behavior, verify domain authenticity, monitor ad content, and accurately attribute app installs. Anomaly detection and behavioral analysis utilize machine learning to identify suspicious patterns and deviations from normal user engagement. AI-powered techniques are used in domain verification and network analysis to detect disparities that indicate domain spoofing. Real-time monitoring and ad content analysis use AI to scan for malicious ad placements and fraudulent content. SDK analysis and app attribution modeling leverage AI to identify abnormal SDK interactions and discrepancies in install reporting. In collaborative filtering and industry collaboration, stakeholders share data and ideas to improve collective fraud detection skills. The expected outcomes of implementing these AI-based strategies include proactively identifying and mitigating fraudulent activities, avoiding payments for

low-quality traffic, ensuring ad placement on legitimate websites, preventing malware infections or privacy breaches, optimizing app marketing campaigns, and strengthening defenses against evolving ad fraud tactics.

Keywords: *ad fraud, artificial intelligence, anomaly detection, behavioral analysis, domain spoofing, ad injection, pixel stuffing, forced redirect ads, SDK spoofing, collaborative filtering, industry collaboration, digital advertising*

Introduction

The web-based online advertising industry has experienced significant growth, creating numerous opportunities for advertisers to generate leads, raise brand awareness, and engage in electronic commerce. In the digital marketplace, various online activities, such as page views, form submissions, clicks, downloads, and purchases, frequently involve financial transactions between advertisers, ad networks, and website publishers. These interactions are to the functioning of the online advertising ecosystem, enabling businesses to reach their target audiences and drive revenue growth.

The financial nature of these web-based actions has also attracted criminals who seek to exploit the system for their own benefit. Fraudsters have identified opportunities to manipulate the online advertising marketplace and illegally divert funds into their own pockets. They employ a range of deceptive techniques, including the use of crimeware, to defraud advertisers, ad networks, and publishers, resulting in significant financial losses and undermining the integrity of the industry.

Ad income from an online advertising campaign is often displayed and accounted for in one of three key categories:

I. Impression-Based

Advertisers frequently pay search engines or online magazines a fixed price for every 1000 banner ads displayed, known as CPM, or *cost per mille*. The display of an ad is referred to as an ad impression or simply an impression. The term impression also describes the ad creative itself that is shown to the user. In theory, charging and accounting by CPM is straightforward, as each web server log entry for a banner ad image represents a single impression (*1*, *2*). However, in practice, CPM advertising can be quite complex, with multiple web servers involved in the process of delivering a banner to an end user (*3*).

To ensure accurate impression tracking, web page caching is disabled through the use of HTTP response headers, and HTTP proxies must respect these headers. This adds an additional complexity to the implementation of CPM-based advertising. Measuring the effectiveness and results of a CPM-based advertising campaign can be challenging. This is due to the greater emphasis on branding campaigns and offline sales, which are more difficult to attribute directly to online ad impressions. The presence of impression spam, which refers to fraudulent or invalid ad impressions, can further complicate the accurate measurement of campaign performance.

Impression spam can take various forms, such as automated bots generating fake ad impressions or the use of hidden or stacked ads that are not visible to the user. These fraudulent practices can inflate impression numbers and distort campaign metrics, making it difficult for advertisers to assess the true impact of their CPM-based advertising efforts. As a result, advertisers and publishers must implement robust monitoring and

verification systems to detect and prevent impression spam, ensuring the integrity of their advertising campaigns and the accuracy of their performance data.

II. Click-Based

In Click-based advertising model, advertisers pay per click (PPC) for their search result ads, with the amount paid per click referred to as the cost per click (CPC) (*4*, *5*). Google's approach to ad placement differed from Overture's in a significant way. Instead of solely considering advertisers' bids, Google also factored in an ad's click-through rate (CTR) (*6*, *7*). Google's ad slots on search results pages are not simply auctioned off to the highest bidder; instead, an ad slot is allocated to the advertiser whose bid, multiplied by the predicted CTR of the ad (as calculated by the ad network), yields the highest value. This means that ad placement is not solely determined by an advertiser's willingness to pay but also by the "*quality*" of the ad, as measured by user engagement through clicks and other factors (*8*).

The inclusion of CTR in the ad placement algorithm ensures that users receive more relevant ads, as the frequency of user clicks serves as an implicit vote of the ad's relevance to the user's query. This user-driven feedback mechanism helps to improve the overall user experience by presenting ads that are more likely to be of interest to the user. Secondly, by better targeting advertisements based on relevance and user engagement, Google can increase its revenue. The combination of advertiser bids and ad quality allows Google to optimize ad placement for maximum revenue generation while simultaneously providing value to users through more pertinent advertising content.

In the pay-per-action (PPA) online advertising model, advertisers pay a cost-per-action (CPA), where an "*action*" is defined as a user reaching a specific "landing" page on the advertiser's site or engaging in a commercial transaction. While cost-per-click (CPC) advertising can be considered a special case of CPA, where the "*action*" is the user clicking on an ad, the term CPA typically refers to a more involved action than a simple click and usually implies that the advertiser pays based on a completed commercial transaction.

CPA-based advertising might be less vulnerable to click fraud compared to CPC-based advertising. The reasoning behind this is that fraudsters may need to engage in actual commercial transactions to successfully deceive advertisers, which could be more challenging and costly than simply generating fraudulent clicks. If the cost of inducing a commercial transaction is relatively low, CPA-based advertising may be just as susceptible to click fraud as CPC-based advertising.

From an advertiser's perspective, CPA-based advertising can be appealing because they only pay for predefined user actions, such as making a purchase or generating a sales lead. This means that advertisers only make payments to an ad network once they have derived tangible value from a click, ensuring a more direct return on their advertising investment. This model aligns the interests of advertisers and ad networks, as both parties benefit from genuine, high-quality user actions rather than mere clicks.

Ad fraud techniques

We can categorize types of ad fraud into the following 3 groups:

1. Click and Impression Fraud

Click and Impression Fraud is a category of ad fraud that involves artificially inflating the number of clicks or impressions on digital advertisements to generate fraudulent revenue or manipulate ad performance metrics. This type of fraud is carried out through various methods, such as using automated bots, click farms, or malicious software to simulate legitimate user interactions with ads. Click Fraud, also known as Bot Fraud, is a technique where fraudsters use automated scripts or bots to generate fake clicks on ads. These bots are designed to mimic human behavior, making it difficult for ad networks and publishers to detect the fraudulent activity. As a result, advertisers end up paying for clicks that have no genuine interest in their products or services, leading to wasted ad spend and skewed performance data.

Click Spamming (*9*), or Click Flooding, is another form of Click Fraud where a large number of clicks are generated in a short period, often from a single IP address or a group of IP addresses. This technique aims to exhaust an advertiser's daily budget quickly, preventing legitimate users from seeing and interacting with the ads. Click Injection is a more sophisticated form of Click Fraud that targets mobile apps. In this scheme, fraudsters use malware to intercept and falsify click data before it reaches the ad network. The malware injects fake clicks into the app, making it appear as though users have interacted with the ads, even when they haven't. Impression Fraud involves generating fake ad impressions without any genuine user interaction. Pixel Stuffing is one such technique, where a 1x1 pixel iframe containing an ad is hidden within a webpage, invisible to the user. Although the ad is not seen, it still counts as an impression, leading to fraudulent charges for the advertiser.

Ad Injection is another Impression Fraud method where unauthorized ads are inserted into websites or apps without the publisher's knowledge or consent (*10*, *11*). This can be done through browser extensions, plugins, or malware that overwrites the intended ad content with fraudulent ads. Ad Stacking is a technique where multiple ads are layered on top of each other within a single ad slot. Only the top ad is visible to the user, but impressions are counted for all the stacked ads, resulting in fraudulent charges for unseen advertisements. Install Farms are a form of Click and Impression Fraud specific to mobile app installations. In this scheme, fraudsters use large groups of low-paid workers or automated tools to download, install, and sometimes interact with mobile apps to inflate installation numbers and engagement metrics artificially. This deceives advertisers into believing that their app install campaigns are more successful than they actually are.

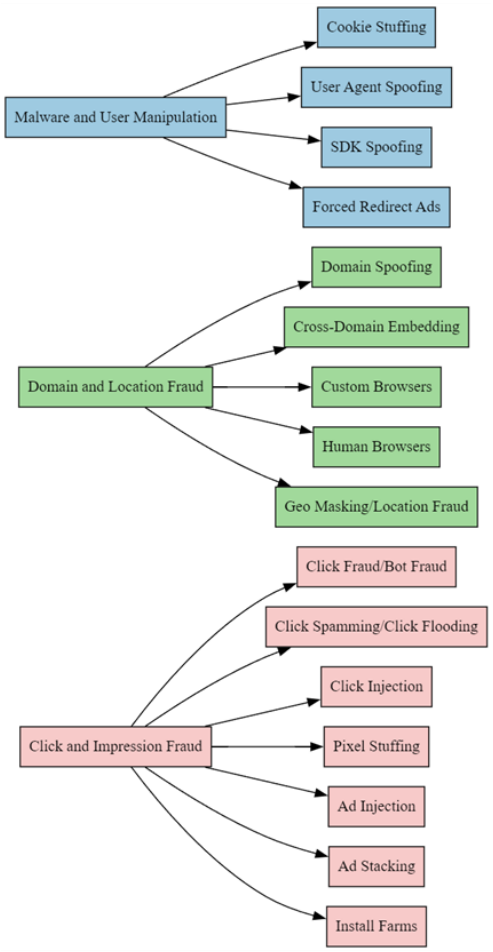

Figure 1

2. Domain and Location Fraud

Domain and Location Fraud is a category of ad fraud that involves misrepresenting the domain or geographic location where ads are being served. The goal of this type of fraud is to deceive advertisers into believing that their ads are appearing on high-quality, reputable websites or within their targeted geographic areas, when in reality, the ads are being displayed on low-quality or fraudulent sites, or in entirely different locations.

One common technique used in Domain and Location Fraud is Domain Spoofing. In this scheme, fraudsters create fake websites that mimic the appearance and content of legitimate, high-traffic websites. They then sell ad space on these spoofed domains to advertisers who believe they are purchasing ad inventory on the genuine, high-value sites. As a result, advertisers end up paying premium prices for ads that are actually being shown on low-quality or fraudulent websites (*12*). Cross-Domain Embedding is another method used in Domain Fraud. This technique involves embedding ad tags or content from a low-quality website into a high-quality, reputable site using iframes or other embedding techniques. When a user visits the reputable site, the embedded ad content from the low-quality site is loaded, generating fraudulent ad impressions and clicks (*13*).

Fraudsters may also use Custom Browsers or Human Browsers to perpetrate Domain and Location Fraud. Custom Browsers are modified web browsers that can be programmed to visit specific websites, click on ads, and simulate human behavior. Human Browsers, on the other hand, are actual human operators who are paid to browse websites and click on ads manually. These techniques are used to generate fraudulent traffic and ad interactions on specific domains, making it appear as though the sites are more popular and valuable than they actually are (*14*, *15*).

Geo Masking, or Location Fraud, is used to misrepresent the geographic location of ad impressions and clicks. Fraudsters use proxy servers, VPNs, or other tools to mask the true IP addresses of their traffic sources , making it appear as though the traffic is originating from a different location. This allows them to target ads to specific geographic regions fraudulently, or to generate fake traffic from high-value locations, such as the United States or Europe, even when the actual traffic is coming from lower-value regions. The impact of Domain and Location Fraud on the digital advertising ecosystem is significant. Advertisers end up wasting their ad budgets on fraudulent, low-quality traffic, while legitimate publishers and websites lose out on potential revenue. This type of fraud undermines the trust and transparency that are essential for the healthy functioning of the online advertising market.

3. Malware and User Manipulation

Malware and User Manipulation is a category of ad fraud that involves manipulating user data or redirecting users to malicious websites without their knowledge or consent. These techniques are designed to exploit vulnerabilities in the digital advertising ecosystem, allowing fraudsters to steal sensitive information, distribute malware, or generate fraudulent ad revenue.

Cookie Stuffing is used to manipulate user data by secretly placing multiple affiliate cookies on a user's device without their awareness. When the user makes a purchase or completes a desired action on the affiliate website, the fraudster receives a commission, even though they did not legitimately refer the user to the site. This technique allows fraudsters to steal commission revenue from legitimate affiliates and can also be used

to skew analytics data, making it difficult for advertisers to accurately attribute conversions and optimize their campaigns.

Another technique used in Malware and User Manipulation is User Agent Spoofing. A user agent is a string of text that identifies a user's browser, operating system, and other device information to websites. Fraudsters can manipulate the user agent string to make it appear as though the traffic is coming from a different device or browser than it actually is. This can be used to bypass ad targeting criteria, generate fake impressions or clicks, or exploit browser-specific vulnerabilities to deliver malware.

SDK Spoofing technique targets mobile apps and their associated Software Development Kits (SDKs). SDKs are tools that app developers use to integrate various features and functionalities into their apps, including ad serving and analytics. Fraudsters can create fake or modified versions of legitimate SDKs, which can then be used to generate fraudulent ad impressions, clicks, or installs. This type of fraud is particularly difficult to detect, as it operates within the app environment and can be hard to distinguish from legitimate SDK activity.

In Forced Redirect Ads scheme, users are forcibly redirected from a legitimate website to a malicious or fraudulent site without their consent. The malicious site may contain deceptive ads, phishing attempts, or malware downloads. Forced redirects can be triggered by hidden scripts, pop-ups, or other deceptive techniques, making it difficult for users to navigate back to the original site or avoid the malicious content.

These techniques not only defraud advertisers and steal revenue from legitimate publishers but also erode user trust and compromise the overall integrity of the online ecosystem. Malware infections and data breaches resulting from these fraudulent activities can lead to significant financial losses, reputational damage, and legal liabilities for all parties involved.

AI-based strategies tailored to tackle the various types of ad fraud

1. Anomaly Detection and Behavioral Analysis

At the core of this approach lies the continuous monitoring of ad traffic patterns and user behavior metrics. Advertisers can gain valuable insights into the genuineness of the interactions with their ads by collecting and analyzing data points such as click-through rates (CTR), conversion rates, session duration, and user navigation paths. Machine learning models play a crucial role in this process, as they are trained to identify patterns and detect anomalies that may indicate fraudulent activity.

The implementation of this strategy begins with the training of machine learning models. These models are fed vast amounts of historical data, allowing them to learn and understand the typical behavior patterns associated with legitimate ad interactions. Once trained, these models are deployed to continuously monitor the incoming ad traffic and user metrics in real-time.

As data flows into the system, the machine learning models analyze each data point, comparing it against the established baseline of normal behavior. They look for any sudden spikes in clicks or conversions originating from suspicious sources, such as bot networks or click farms. Additionally, the models scrutinize user engagement patterns, seeking out deviations from the norm, such as abnormally short session durations or erratic navigation paths.

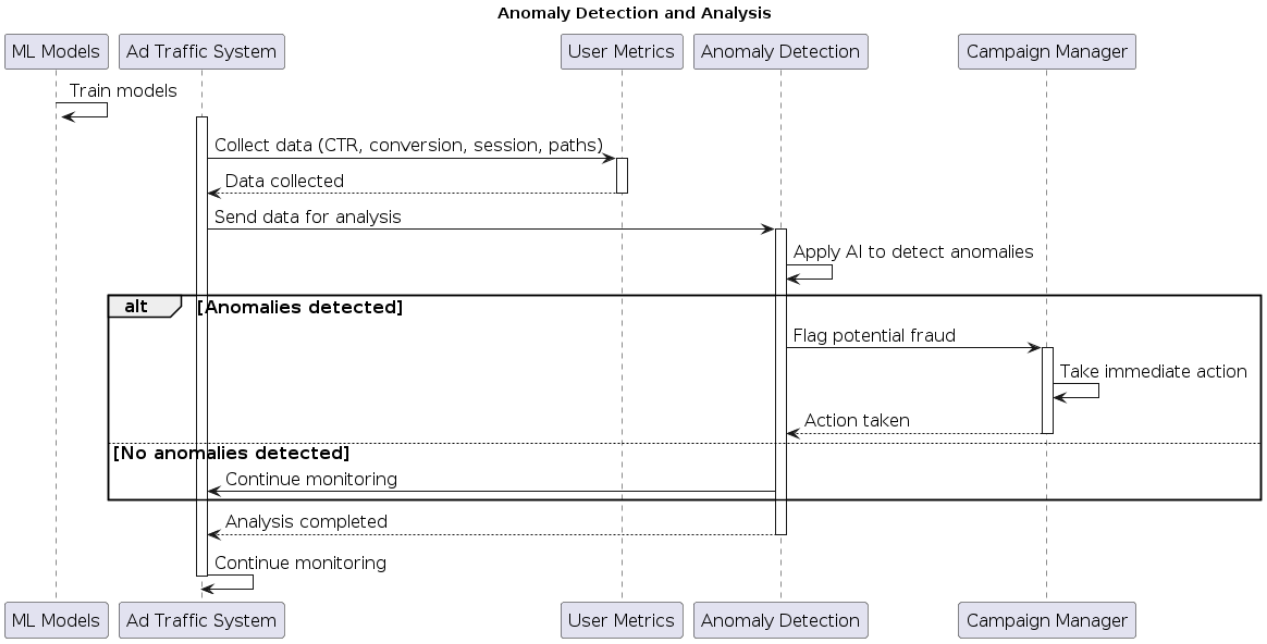

Figure 2

When an anomaly is detected, the system promptly flags it as potential ad fraud and alerts the campaign manager. This proactive approach allows advertisers to take immediate action, such as blocking the suspicious traffic sources or adjusting the targeting parameters to mitigate the impact on their campaigns. Advertisers can protect their ad spend and ensure that their campaigns reach genuine, interested users by swiftly identifying and addressing fraudulent activity.

The benefits of implementing AI-driven anomaly detection and behavioral analysis are manifold. Firstly, it enables advertisers to maintain the integrity of their ad campaigns by weeding out fraudulent interactions. This ensures that the performance metrics accurately reflect the true impact of the ads on the target audience. Secondly, it helps optimize ad spend by preventing fraudulent clicks and conversions from draining the

advertising budget. Advertisers can maximize the return on their investment by allocating resources towards genuine user interactions.

Behavioral analysis provides analysis of user engagement patterns, allowing advertisers to refine their targeting strategies and ad creatives. Advertisers can tailor their campaigns to better resonate with their audience, leading to improved ad relevance and higher conversion rates, by understanding how legitimate users interact with their ads.

Anomaly detection and behavioral analysis, powered by AI, offer a robust solution to the pervasive problem of ad fraud. Advertisers can safeguard the integrity of their campaigns by continuously monitoring ad traffic and user behavior, detecting anomalies in real-time, and enabling swift action against suspicious activity. This proactive approach not only protects ad spend but also ensures that campaigns reach genuine users, ultimately driving better results and ROI.

2. Domain Verification and Network Analysis

Domain spoofing occurs when fraudulent actors misrepresent low-quality or fraudulent websites as legitimate, high-value domains, tricking advertisers into placing ads on these sites. To combat this issue, an strategy involves AI-driven domain verification and network analysis techniques.

The core of this approach lies in the development of AI algorithms capable of mapping out the intricate relationships between various entities within the ad ecosystem. These entities include publishers, ad exchanges, and advertisers, each playing a crucial role in the delivery of ads to the intended audience. AI algorithms can uncover patterns and anomalies that may indicate instances of domain spoofing by analyzing the complex web of connections.

The implementation of this strategy begins with the training of machine learning models. These models are fed vast amounts of historical data on traffic flows and domain relationships, allowing them to learn and understand the expected patterns of legitimate ad delivery. Once trained, these models are deployed to continuously monitor the ad ecosystem, analyzing the real-time flow of traffic between different entities.

As data on traffic flows is fed into the system, the AI algorithms diligently compare the observed patterns against the expected norms. They look for discrepancies and anomalies that may indicate the presence of spoofed domains. For example, if a low-quality website suddenly starts receiving a high volume of traffic from reputable ad exchanges, it may raise a red flag for potential domain spoofing.

When discrepancies are detected, the system promptly alerts the advertiser, providing them with valuable insights into the potential instances of domain spoofing. This proactive approach empowers advertisers to take immediate action, such as blocking the fraudulent traffic sources or adjusting their ad placement strategies to avoid low-quality or fraudulent websites.

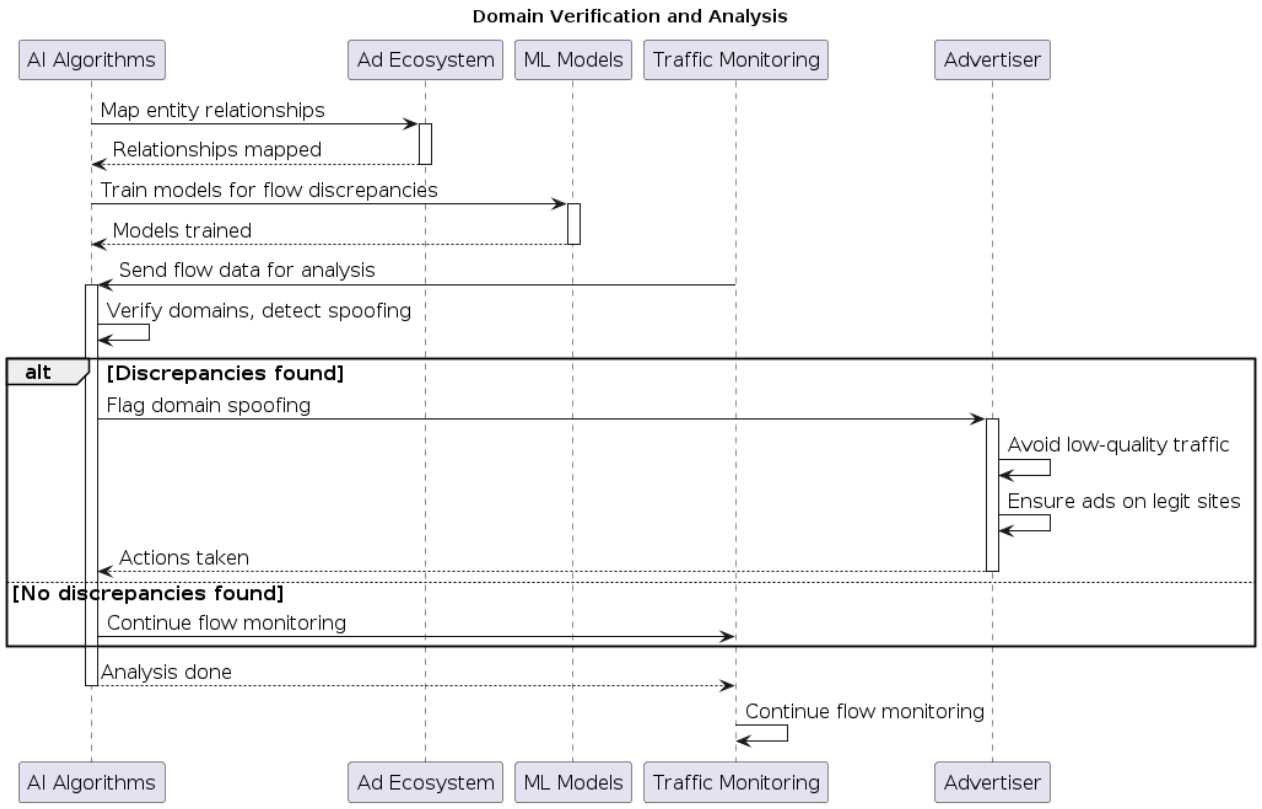

Figure 3

The expected outcomes of implementing AI-driven domain verification and network analysis are significant. Advertisers can ensure that their ads are served on legitimate, high-value websites by accurately verifying the authenticity of domains and detecting instances of spoofing. This helps to maximize the impact of their advertising spend, reaching genuine audiences and driving meaningful engagement.

Advertisers can protect their brand reputation and maintain the trust of their target audience by avoiding low-quality traffic and fraudulent websites. Serving ads on fraudulent or inappropriate websites can damage brand image and erode consumer confidence, leading to long-term negative consequences. Advertisers can contribute to the overall integrity and transparency of the digital advertising industry by mapping out the relationships between different entities and identifying patterns of fraudulent behavior.

3. Real-time Monitoring and Ad Content Analysis

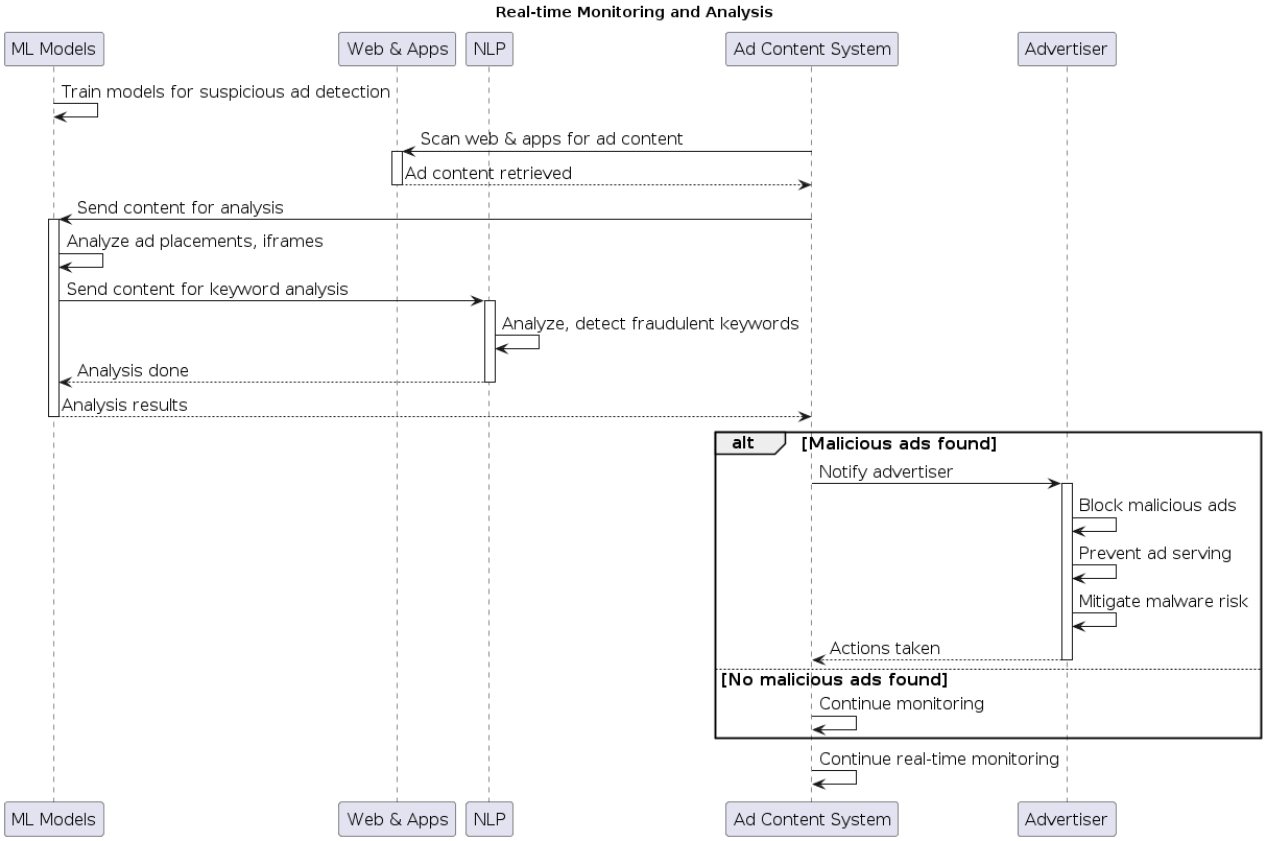

Figure 4

Advertisers strive to deliver relevant and engaging ads to their target audience, but the presence of malicious activities such as ad injection, pixel stuffing, and forced redirect ads poses significant risks. These fraudulent practices not only compromise the user experience but also expose users to potential malware infections and privacy breaches. To combat these threats, am strategy involves implementing AI-powered real-time monitoring and ad content analysis.

At the core of this approach lies the deployment of advanced machine learning models capable of continuously scanning web pages and mobile applications for suspicious ad placements and malicious content. These models are trained on vast amounts of historical data, learning to identify patterns and anomalies associated with fraudulent ad activities.

The implementation of this strategy begins with the integration of the AI-powered monitoring system into the ad serving infrastructure. As web pages and mobile apps are loaded, the system actively scans the content, analyzing the ad placements and their surrounding context. It looks for indicators of ad injection, such as the presence of hidden ads or unexpected ad placements that disrupt the user experience.

In addition to analyzing ad placements, the system leverages natural language processing (NLP) techniques to scrutinize the actual content of the ads. NLP algorithms are trained to identify keywords and phrases commonly associated with fraudulent activities, such as misleading claims, deceptive offers, or malicious URLs. The system can flag potential instances of fraud and take appropriate action by examining the language used in the ads. When suspicious ad content is detected, the system promptly notifies the advertiser, providing them with detailed information about the identified threat. This real-time alerting mechanism enables advertisers to swiftly block the malicious ads, preventing them from being served to users and mitigating the risk of harm.

The expected outcomes of implementing AI-powered real-time monitoring and ad content analysis are significant. Advertisers can prevent their users from being exposed to harmful information by continuously screening for malicious ads and detecting fraudulent activity in real time. This contributes to a safe and trustworthy advertising ecosystem, ensuring that users may interact with ads without worry of malware infections or privacy breaches.

Advertisers may protect their brand's reputation and retain the confidence of their target audience by stopping the delivering of malicious adverts. Serving fraudulent or deceptive ads can erode consumer confidence and damage the advertiser's credibility, leading to long-term negative consequences for their business. Real-time monitoring and ad content analysis also provide valuable insights into the evolving tactics of fraudulent actors. Through the examination of patterns and attributes of identified fraudulent advertisements, advertisers can acquire a more profound comprehension of the strategies employed by deceivers and modify their protection mechanisms correspondingly. This knowledge can be shared within the industry to contribute to the collective fight against ad fraud. To ensure the effectiveness of this strategy, continuous training and refinement of the AI models are essential. The machine learning algorithms must be updated with the latest data and patterns to maintain their accuracy and effectiveness in detecting new forms of ad fraud.

4. SDK Analysis and App Attribution Modeling

As mobile apps continue to grow, app install campaigns have become a vital tool for user acquisition and engagement. The insidious presence of SDK spoofing and click injection poses a significant threat, distorting metrics and draining ad budgets. To effectively combat these issues, harnessing the power of AI-driven SDK analysis and app attribution modeling is a good strategy. The heart of this approach lies in the creation of advanced machine learning models that possess the ability to meticulously analyze the intricate interactions between mobile apps and their integrated software development kits (SDKs). These SDKs that is enabling

various functionalities such as ad serving, analytics, and attribution tracking, can be scrutinized by AI algorithms to uncover anomalies and identify instances of fraudulent behavior.

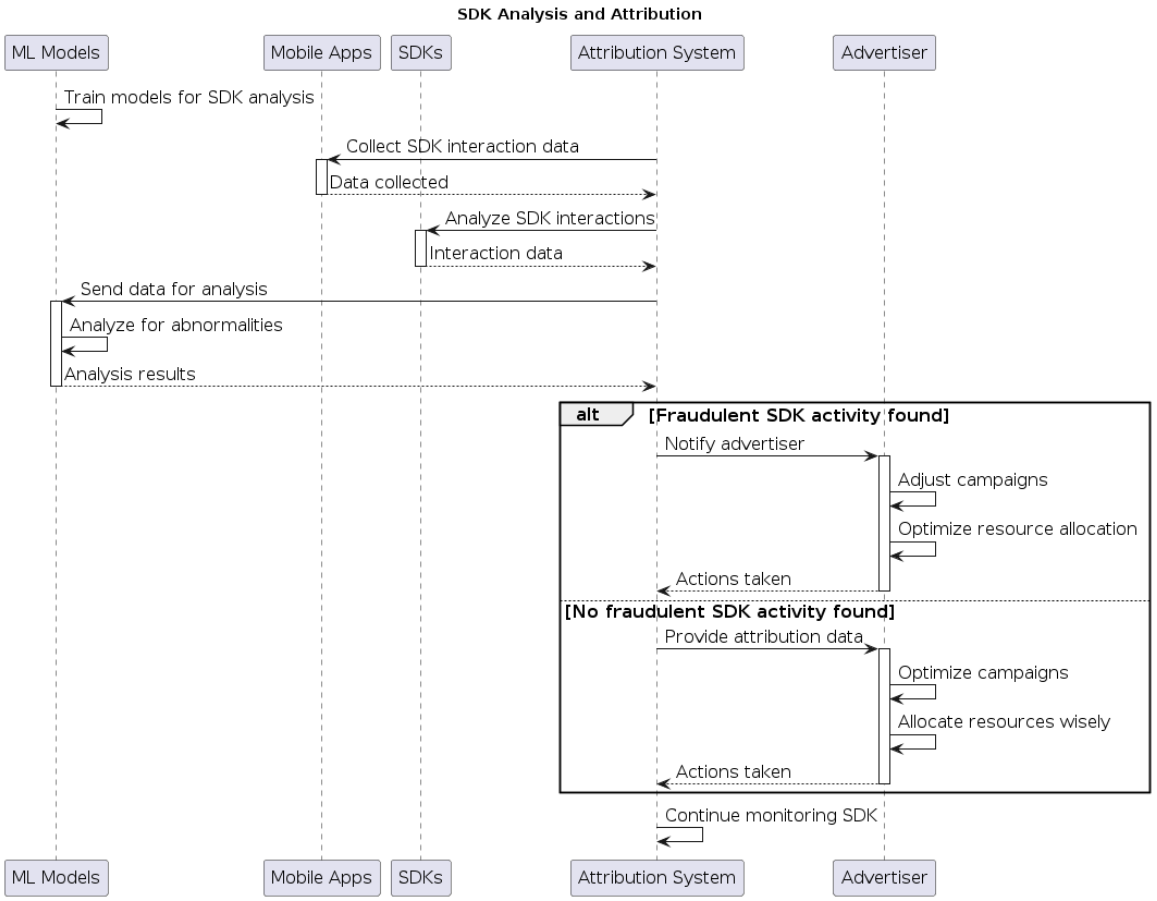

Figure 5

The first step involves amassing a wealth of data on SDK interactions spanning a diverse range of mobile apps. This rich dataset, encompassing click-through rates, install events, and user engagement metrics, forms the foundation upon which machine learning models are trained to discern the normal patterns of legitimate SDK behavior. With the models trained and ready, they are used to vigilantly monitor SDK interactions in real-time. Like digital detectives, they sift through the data, seeking out abnormalities and discrepancies that may signal the presence of SDK spoofing or click injection. Sudden spikes in click-through rates or glaring

mismatches between reported installs and actual user engagement serve as glaring red flags, triggering alarms for potentially fraudulent activity. When suspicious SDK behavior is uncovered, the system swiftly alerts the advertiser, arming them with comprehensive insights into the identified anomalies. This real-time alerting mechanism empowers advertisers to spring into action, launching investigations into the affected apps, fine-tuning campaign parameters, or swiftly blacklisting fraudulent sources.

Advertisers may make their app marketing efforts more efficient and spend resources more precisely by hiding fraudulent SDK activity and accurately attributing app installs to authentic sources. They can direct their efforts towards channels and partners that consistently deliver authentic, high-value users, while minimizing the wastage of ad spend on fraudulent installs.

Advertisers gain a crystal-clear understanding of their campaigns' true performance by ensuring the accuracy of attribution. With this knowledge, they can make informed, data-driven decisions to refine targeting strategies, adjust bidding models, and optimize creative assets based on genuine user engagement metrics. The result is an impressive boost in return on investment (ROI) and a more judicious allocation of marketing budgets. Beyond optimizing campaign performance, SDK analysis and app attribution modeling contribute to the greater good of the mobile advertising ecosystem. Advertisers play a vital role in fostering a fair and transparent environment for legitimate app developers and ad networks by identifying and thwarting fraudulent activities. This, in turn, nurtures trust and confidence among all stakeholders, paving the way for the long-term sustainability and growth of the industry. Fraudulent actors cunningly adapt their tactics, the machine learning algorithms must evolve in tandem, learning to detect novel patterns of SDK spoofing and click injection. Regular updates and retraining of the models by the latest data and trends, are essential to preserving their accuracy and effectiveness against ad fraud.

5. Collaborative Filtering and Industry Collaboration

Collaboration and information sharing through collaborative filtering techniques allow the industry to unlock the power of collective intelligence. This strengthens defenses and minimizes the financial impact of ad fraud. Industry-wide platforms or partnerships should be established to serve as central hubs for knowledge exchange. These platforms provide a space for advertisers, publishers, ad networks, and technology providers to share insights, best practices, and threat intelligence related to ad fraud. Advertisers, publishers, ad networks, and technology providers should willingly share their unique perspectives, experiences, and information on emerging fraud patterns, successful mitigation strategies, and innovative detection techniques. Open sharing of knowledge forms the foundation for collective intelligence.

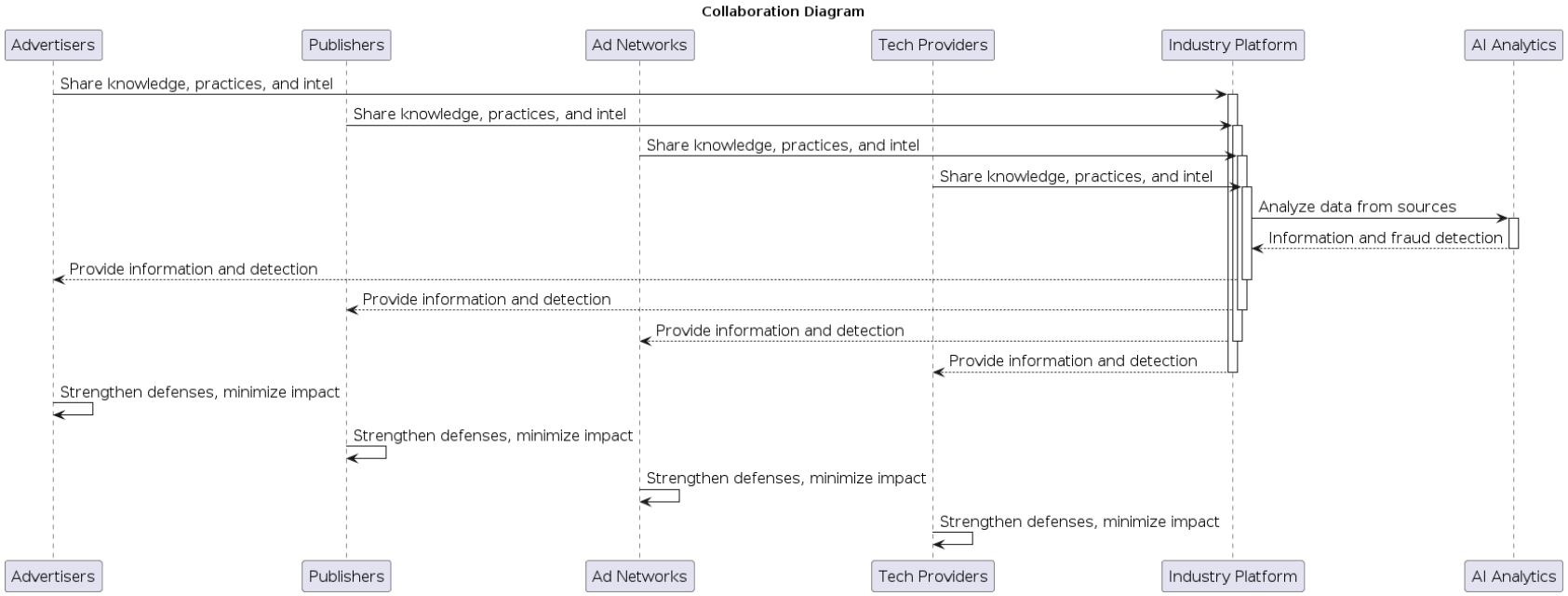

Figure 6

AI-driven analytics should be applied to the collected data. Machine learning algorithms aggregate and analyze the vast amounts of contributed data. These algorithms identify hidden patterns, detect anomalies, and uncover emerging trends in ad fraud activities. The industry can take proactive measures to strengthen their defenses against fraudulent activities using these valuable insights. Advertisers can adapt their strategies, implement more effective detection and prevention measures, and minimize the financial impact of fraudulent traffic on their advertising investments. Publishers, ad networks, and technology providers can also leverage the collective intelligence to enhance their own fraud detection capabilities and contribute to the overall integrity of the digital advertising ecosystem. Regular meetings, workshops, and forums should be organized to facilitate the continuous exchange of ideas, discuss emerging trends, and share success stories. Ongoing communication and collaboration ensure that the industry remains vigilant and adaptable in the face of ever-changing ad fraud tactics. Industry-wide standards and best practices for combating ad fraud should be developed throughout the collaborative filtering and industry collaboration process. Stakeholders should work together to establish common guidelines, protocols, and frameworks that ensure a consistent and effective approach to fraud detection and prevention. Standardization helps level the playing field and makes it more difficult for fraudsters to exploit vulnerabilities.

Collaborative filtering and industry collaboration is an iterative and ongoing process. When new threats emerge, the industry must continue to share information, leverage AI-driven analytics, and adapt their strategies accordingly. Maintaining a collaborative mindset and working together as an industry allows advertisers, publishers, ad networks, and technology providers to combat fraudulent actors and ensure a transparent and trustworthy advertising environment.

The sequence of collaborative filtering and industry collaboration involves establishing shared platforms, active participation and data contribution from stakeholders, applying AI-driven analytics, implementing strengthened defenses, ongoing collaboration and engagement, and developing industry-wide standards. This approach enables the industry to use the power of collective intelligence, adapt to evolving ad fraud tactics, and minimize the financial impact of fraudulent activities.

Conclusion

Ad fraud poses a significant challenge to the digital advertising industry, causing substantial financial losses and eroding trust between advertisers, publishers, and consumers. The complexity and ever-evolving nature of ad fraud require innovative and adaptive strategies to effectively combat this pervasive issue. Artificial intelligence (AI) emerges as a powerful tool in the fight against ad fraud, offering advanced capabilities in anomaly detection, behavioral analysis, and real-time monitoring.

AI-driven anomaly detection and behavioral analysis play a crucial role in identifying suspicious patterns and deviations from typical user engagement metrics. Machine learning models can be trained to continuously monitor ad traffic, click-through rates, conversion rates, and navigation paths, flagging any sudden spikes or abnormal geolocation patterns indicative of fraudulent activity. This proactive approach enables advertisers to take immediate action and mitigate the impact of ad fraud on their campaigns. Additionally, AI algorithms can be leveraged for domain verification and network analysis, accurately identifying instances of domain spoofing and ensuring that ads are served on legitimate, high-value websites.

Real-time monitoring and ad content analysis further enhance the effectiveness of AI-based strategies against ad fraud. Machine learning models can scan web pages and mobile apps for suspicious ad placements, hidden ads, or malicious iframes, while natural language processing techniques can analyze ad content to identify keywords associated with fraudulent activities. This continuous monitoring allows advertisers to quickly detect and block malicious ads, preventing them from reaching users and minimizing the risk of malware infections or privacy breaches. Moreover, AI algorithms can be employed to analyze software development kits (SDKs) integrated into mobile apps, accurately attributing app installs to legitimate sources and detecting fraudulent SDK activity, such as click injection or SDK spoofing.

Collaborative filtering and industry collaboration are vital components of a comprehensive approach to combating ad fraud. Establishing industry-wide platforms or partnerships facilitates the sharing of insights, best practices, and threat intelligence among advertisers, publishers, ad networks, and technology providers. AI-driven analytics can aggregate and analyze data from diverse sources, identifying emerging trends and

enhancing fraud detection capabilities. Through collaboration and collective intelligence, advertisers can stay ahead of evolving ad fraud tactics, strengthen their defenses, and minimize the financial impact of fraudulent traffic on their advertising investments. As the digital advertising landscape continues to evolve, the integration of AI-based strategies will be essential in maintaining the integrity and effectiveness of online advertising, fostering a more transparent and trustworthy ecosystem for all stakeholders involved.

he success of AI-driven fraud detection heavily relies on the quality and availability of training data. Obtaining comprehensive and representative datasets that capture diverse ad fraud scenarios can be challenging, especially when dealing with emerging or sophisticated fraud techniques.

Inconsistencies or biases in the data used to train AI models may lead to suboptimal performance or false positives in fraud detection.

Fraudsters renew their tactics to evade detection, exploiting weaknesses in existing AI models (*16*, *17*). If AI algorithms become more sophisticated in identifying fraudulent activities, fraudsters may develop countermeasures or adapt their techniques to circumvent detection. The race between fraudsters and AI-based defense mechanisms requires constant monitoring, updates, and retraining of AI models to keep pace with the threat. Implementing AI-based strategies for ad fraud detection often requires significant computational resources, especially when dealing with large-scale ad networks and real-time monitoring. Scaling AI algorithms to handle the massive volumes of ad traffic and user interactions can be computationally intensive and may require substantial investment in infrastructure and hardware. False positives occur when legitimate ad traffic or user behavior is incorrectly flagged as fraudulent, leading to revenue losses for publishers and advertisers. False negatives, on the other hand, occur when fraudulent activities go undetected, allowing fraudsters to continue their malicious practices unchecked.

References

1. R. Briggs, N. Hollis, Advertising on the Web: is there response before click-through?, *Journal of Advertising research*. **37** (1997)p. 33+.

2. X. Drèze, F.-X. Hussherr, Internet advertising: Is anybody watching? *Journal of Interactive Marketing* **17**, 8–23 (2003).

3. T. Graepel, J. Q. Candela, T. Borchert, R. Herbrich, Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. (2010).

4. K. Fjell, Online advertising: Pay-per-view versus pay-per-click—A comment. *Journal of Revenue and Pricing Management* (2009).

5. K. K. Kapoor, Y. K. Dwivedi, N. C. Piercy, Pay-per-click advertising: A literature review. *The Marketing Review* **16**, 183–202 (2016).

6. X. He, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, J. Q. Candela, "Practical Lessons from Predicting Clicks on Ads at Facebook" in *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising* (Association for Computing Machinery, New York, NY, USA, 2014)*ADKDD'14*, pp. 1–9.

7. M. Richardson, E. Dominowska, R. Ragno, "Predicting clicks: estimating the click-through rate for new ads" in *Proceedings of the 16th International Conference on World Wide Web* (Association for Computing Machinery, New York, NY, USA, 2007)*WWW '07*, pp. 521–530.

8. L. Shi, B. Li, Predict the click-through rate and average cost per click for keywords using machine learning methodologies. *Proceedings of the International Conference on* (2016).

9. S. Ş. Kaya, B. Çavdaroğlu, K. S. Şensoy, "Detection of click spamming in mobile advertising" in *Advances in Operational Research in the Balkans* (Springer International Publishing, Cham, 2020), pp. 251–263.

10. M. Ikram, N. Vallina-Rodriguez, S. Seneviratne, M. A. Kaafar, V. Paxson, "An Analysis of the Privacy and Security Risks of Android VPN Permission-enabled Apps" in *Proceedings of the 2016 Internet Measurement Conference* (Association for Computing Machinery, New York, NY, USA, 2016)*IMC '16*, pp. 349–364.

11. N. Provos, D. McNamee, P. Mavrommatis, K. Wang, N. Modadugu, The ghost in the browser: Analysis of web-based malware. *HotBots*, 4–4 (2007).

12. K. Springborn, P. Barford, Impression fraud in on-line advertising via pay-per-view networks. *USENIX Secur Symp*, 211–226 (2013).

13. C. M. R. Haider, A. Iqbal, A. H. Rahman, An ensemble learning based approach for impression fraud detection in mobile advertising. *Journal of Network and* (2018).

14. A. A. Metwally, D. Agrawal, A. E. Abbadi, Using association rules for fraud detection in web advertising networks. *VLDB J.*, 169–180 (2005).

15. R. J. Oentaryo, E.-P. Lim, M. Finegold, D. Lo, F. Zhu, C. Phua, E. Cheu, G.-E. Yap, K. Sim, M. N. Nguyen, K. Perera, B. Neupane, M. Faisal, Z. Aung, W. Woon, W. Chen, D. Patel, D. Berrar, Detecting click fraud in online advertising: a data mining approach. *J. Mach. Learn. Res.* **15**, 99–140 (2014).

16. Z. Li, K. Zhang, Y. Xie, F. Yu, X. Wang, "Knowing your enemy: understanding and detecting malicious web advertising" in *Proceedings of the 2012 ACM Conference on Computer and Communications Security* (Association for Computing Machinery, New York, NY, USA, 2012)*CCS '12*, pp. 674–686.

17. R. S. Owen, "Online Advertising Fraud" in *Electronic Commerce: Concepts, Methodologies, Tools, and Applications* (IGI Global, 2008), pp. 1598–1605.