Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

An Economic Analysis of Online Ad Fraud Deterrence

Min Chen, Subodha Kumar, Abhishek Ray

To cite this article:
Min Chen, Subodha Kumar, Abhishek Ray (2026) An Economic Analysis of Online Ad Fraud Deterrence. Management Science

Published online in Articles in Advance 06 Feb 2026

. https://doi.org/10.1287/mnsc.2022.02201

https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–28
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

An Economic Analysis of Online Ad Fraud Deterrence

Min Chen,[a,*] **Subodha Kumar,**[b] **Abhishek Ray**[a]

[a] Costello College of Business, George Mason University, Fairfax, Virginia 22030; [b] Fox School of Business, Temple University, Philadelphia, Pennsylvania 19122

*Corresponding author

Contact: mchen15@gmu.edu, https://orcid.org/0000-0002-3582-6991 (MC); subodha@temple.edu, https://orcid.org/0000-0002-4401-7950 (SK); aray8@gmu.edu, https://orcid.org/0000-0001-9963-1918 (AR)

Received: July 20, 2022
Revised: June 6, 2024; May 26, 2025
Accepted: September 28, 2025
Published Online in Articles in Advance: February 6, 2026

https://doi.org/10.1287/mnsc.2022.02201

Abstract. Ad fraud is increasingly becoming a major concern in online advertising, with publishers being one of the key sources of fraudulent ad traffic. Although ad fraud deterrence is a critical problem from both technical and economic perspectives, past research has not considered their interplay. Our paper fills this critical gap by building a game-theoretic model wherein an ad network (an intermediary between publishers and advertisers) strategically leverages its *technological tool* (the configuration of a given fraud detection technology) and *economic tool* (the payment to publishers) to deter ad fraud and maximize profits effectively. Our analysis generates several interesting findings. For example, as the fraud detection technology and fraud generation techniques improve, we show that although the ad network needs to respond by making the technology configuration stricter to dampen fraud motives and admit less ad traffic, it may sometimes need to increase the payment. Furthermore, although many stakeholders advocate instituting strict legislation and policies to reduce malicious publishers, we show that this may sometimes fail to reduce fraud traffic and even hurt an ad network's profit. In addition, our results provide other useful implications that present a new theoretical perspective on the incentive problems in ad fraud generation and detection. Our study also draws valuable insights for ad networks into implementing effective ad fraud deterrence policies and for advertisers to audit and monitor their ad campaign performance.

History: Accepted by Hemant Bhargava, information systems.
Funding: S. Kumar thanks the Temple Center for International Business Education and Research for partially supporting this research.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2022.02201.

Keywords: ad fraud • online advertising • ad networks • publishers • game-theoretical modeling

> Ad fraud is one of the biggest sources of income for organised crime—second only to the drugs trade. (Carlile 2023)

1. Introduction

Fueled by the rapid growth in internet usage, online advertising has experienced a substantial expansion in recent years. In the United States, digital ad spending reached a record high of $298.4 billion in 2024, marking a 10% increase over the previous year (SaleHoo 2025). This revenue surge has been primarily driven by the increasing number of *publishers* that provide advertisers with spaces to run digital ads (Kircher and Foerderer 2024). Intermediaries, such as *ad networks*, have aided this growth in publisher numbers by helping match advertisers seeking to reach their audiences with publishers aiming to monetize their ad-space inventory (D'Annunzio and Russo 2024, Jodzevica 2025). These ad networks receive payments from advertisers and share a portion of the revenue with participating publishers (Bhargava 2022). Since their inception, ad networks have become increasingly central to the digital advertising ecosystem. For example, Google's Display Network reaches over 90% of the global internet population (Gibbons 2025).

Despite their widespread adoption and central role in the digital advertising ecosystem, ad networks continue to grapple with persistent challenges of online *ad fraud*—the deliberate falsification of impressions, clicks, conversions, or other engagement metrics for illegal financial gain or other deceptive purposes (Rayobyte 2025). Although ad networks have implemented a range of technological and policy interventions to curb fraudulent activity (Google Ads 2025), ad fraud remains a significant and growing concern. Industry estimates place losses from online ad fraud at $84 billion in 2023, with projections that they will rise to $172 billion by 2028 (Dogtiev 2025).

These trends underscore the urgency of re-evaluating the strategic and operational mechanisms through which ad networks address online ad fraud and raise several critical questions. *How can an ad network effectively*

deploy its available tools to deter ad fraud? How should it adapt its policy decisions in response to a dynamically evolving market? Are some prevalent policies truly effective in reducing ad fraud or beneficial to the ad network? Despite the growing academic interest in ad fraud (Chen et al. 2015, Zhu et al. 2017b, Gordon et al. 2021), these questions have remained largely unexplored. This paper addresses these critical questions and reveals important managerial insights into effectively regulating the online advertising market.

1.1. Background and Motivation

We consider a typical online advertising setting where an ad network acts as an intermediary, aggregating the ad slots supply from publishers and selling them to advertisers (D'Annunzio and Russo 2020).[1] By providing ad spaces and generating ad traffic, publishers receive a share of the ad revenue collected by the ad network from advertisers (D'Annunzio and Russo 2024). Advertising on publishers' ad spaces generates ad traffic that is either *valid* (i.e., from genuine users) or *invalid/fraudulent* (i.e., from users who are not genuine). Within invalid ad traffic, a substantial portion can be fraudulently generated by publishers using automated tools, bots, or click/lead farms (Rayobyte 2025). These publishers are typically motivated by the potential economic benefits of generating fraudulent ad traffic to inflate engagement metrics and increase revenue.[2] Our study focuses on this type of invalid ad traffic generated by these publishers, one of the significant sources of ad fraud that most research has focused on (Saluja 2024).

Amid growing concerns over the prevalence of ad fraud, an ad network can make key strategic decisions to reduce the fraud incentives to maximize its profit. On the technology side, because an ad network typically uses a fraud detection system to identify and filter fraudulent ad traffic (Google 2025b), configuring this technology as either *stringent* to admit less traffic or *lenient* to admit more helps control the ad-traffic quality. Because of the imperfect nature of classification, perfect fraud detection without misclassifications (e.g., misdetecting fraud traffic as valid or the opposite) is technically impossible (Tanzako 2024). Thus, although a stringent configuration helps dampen the fraud motive, it risks disincentivizing publishers' participation in the ad network by gating and thus, not appropriately compensating a sizable portion of valid ad traffic. Conversely, although a lenient configuration helps increase publishers' participation, it can promote ad fraud because fraudulent ad traffic will likely be incorrectly rewarded. This is a *technological tool* that an ad network can employ to control the quality of the ad traffic in its network.

An important factor influencing the ad network's technology configuration decision is the underlying *quality profile* of the fraud detection technology. In a normative sense, a higher-quality technology yields more accurate classifications of the ad traffic than a lower-quality technology under the identical configuration (Mookerjee et al. 2011). Consequently, the ad network's technology configuration decision inherently depends on its technology's quality. Thus, we adopt a framework that captures the quality of a fraud detection technology and the associated technology configuration decisions to examine the dynamics between these two in the context of ad fraud.

On the economic side, the ad network's payment to publishers plays a crucial role in shaping publishers' fraud decisions (Zhu et al. 2017a). An ad network can strategically choose payment to reward publishers appropriately for the ad traffic that passes through its fraud detection technology (Dritsoula and Musacchio 2014). This is an *economic tool* that an ad network can leverage to influence publishers' participation and fraud incentives. A higher payment helps increase publishers' participation but risks attracting more fraudulent ad traffic. Similarly, a lower payment might dissuade fraud and reduce publisher participation and ad network revenue. Thus, the ad network's decision on publisher payments could have two opposing effects on fraud incentives. The interplay between these effects and those from the technology configuration decision complicates the analysis. A suboptimal policy might generate disincentives that lead to excessive ad fraud or the exclusion of a sizable number of publishers, thereby aggravating market inefficiencies.

Although ad fraud has been studied extensively (Wilbur and Zhu 2009; Chen et al. 2012a, 2015), to the best of our knowledge, prior research has not focused on an ad network's decision scenario that considers both *technological* and *economic* tools (discussed above) to curb online ad fraud. In this paper, we attempt to fill this gap by developing a game-theoretic model to reveal insights regarding how an ad network can strategically decide on its technology configuration (technological tool) and payment to publishers (economic tool) in a dynamically evolving online ad market and whether some of the ad network's prevalent policies are genuinely effective in deterring ad fraud. The following subsection elaborates on these research questions.

1.2. Research Questions and Contributions

It has long been argued that ad networks may lack strong incentives to proactively combat ad fraud because of revenues from undetected fraud traffic (Tsur 2024, Davies 2025). Essentially, this lack of ad fraud deterrence can be characterized as both a technology problem and an economic problem. Intuitively, publishers may perpetuate ad fraud if the economic gain is substantial. Similarly, an ad network may allow fraudulent ad traffic on its platform because of the

potential financial benefits of undetected fraud traffic (Cai et al. 2020). Therefore, contrary to prior studies that focus on either economic measures taken by ad networks (Wilbur and Zhu 2009) or technological advancements in fraud detection technology (Kitts et al. 2015) as the solution for curbing ad fraud, we consider the interplay between these two and characterize the ad network's strategic decisions to show how it can coordinate the economic tool with the technological tool to deter ad fraud and maximize self-interest effectively. This characterization enables us to analyze the ad network's responses to a dynamically evolving market, reassess the effectiveness of prevalent industry policies, and derive insights that defy some of the beliefs and intuitions of past research as elaborated below.

In recent years, ad networks have increasingly leveraged advancements in artificial intelligence (AI) and machine learning (ML) to enhance the performance of their fraud detection systems (Kim et al. 2023). Some argue that technological improvements enable more accurate detection of valid ad traffic, allowing ad networks to reward publishers more appropriately and potentially reduce the payment to publishers (Ranne 2024). However, improved detection capabilities also help detect ad fraud and deter fraud incentives, thereby increasing the advertisers' ad valuation and subsequent payment to publishers (Chen et al. 2015). These studies, however, focus only on economic incentives and overlook their interplay with the effects of technology configuration. To provide a better understanding, we first ask the following question. *How do improvements in fraud detection technology affect the ad network's payment to publishers?*

We find that as fraud detection technology improves, although an ad network should set the technology configuration more strictly to admit less traffic, it may increase or decrease the payment to publishers depending on market conditions. This result suggests that market leaders can leverage advanced detection capabilities by implementing and enforcing more stringent ad fraud detection policies. For example, Google has deployed a highly automated detection system (Google Ads 2025) and strengthened its enforcement through a series of new ad policy updates (Adegbola 2025). The result also underscores the importance of aligning economic levers (i.e., payment to publishers) with the ad network's technological capability in ad fraud detection. As the detection technology improves, increasing payments to publishers to sustain their participation and ensure a high-quality traffic ecosystem may be optimal. As explained in Proposition 1, this insight helps explain the increased payment to publishers because of the upward trend in the advertisers' valuations within the Google Ads platform (Octoboard 2024), reflecting the interplay between improved fraud mitigation and the need to maintain strong publisher incentives.

In addition to the ad network's fraud detection technology, another key determinant of the publishers' fraud incentives is the efficiency of generating fraudulent ad traffic (Saluja 2024). Over the past decade, the publishers' fraud generation efficiency has increased substantially driven by the evolution of ad fraud techniques from labor-intensive methods to highly automated, large-scale operations (Cai et al. 2020, Jackson 2025). This raises an important managerial question. *How does rising fraud generation efficiency affect the ad network's payment to publishers?*

Prior research suggests that ad networks can curb ad fraud by lowering payments to publishers because such economic measures help disincentivize fraud motives (Gordon et al. 2021). However, interestingly, this insight does not always hold because these studies have largely centered on economic countermeasures, often neglecting the complementary role of technological tools in shaping fraud incentives. By contrast, we show that when fraud generation efficiency rises, it may be optimal for an ad network to increase payments to publishers in conjunction with implementing more stringent technological configurations for fraud detection.

The interplay between economic and technological tools highlights the limitations of narrowly focused interventions relying primarily on economic disincentives to curb publishers' fraud motives. It also sheds light on strategic behaviors observed in industry practice. On one front, ad networks have adopted increasingly stringent enforcement measures to combat ad fraud, particularly in response to the rise of AI-generated threats. For example, Google recently announced the implementation of more than 30 new policy updates aimed at enhancing fraud detection and enforcement mechanisms (Adegbola 2025). On the other front, payments to publishers have also increased, in part because of higher ad valuations from advertisers (Veuno 2024). These developments collectively highlight the need for coordinated economic incentives alongside robust technological safeguards. In Proposition 2, we shed light on this important question.

In addition to analyzing the ad network's strategic responses to market changes, we also consider the problems arising from the interplay between improvements in fraud generation and detection technologies. Specifically, we ask the following question. *Is it beneficial for an ad network to continue investing in its fraud detection technology, particularly when faced with evolving fraud technology?* From a static perspective, we find that returns to technological investment may diminish with increasing technology quality. Interestingly, when viewed dynamically and accounting for the ever-changing nature of fraud generation techniques, we find that rising fraud

generation efficiency can significantly amplify the benefits of technological improvement.

This highlights the critical need for sustained innovation and investment in fraud detection systems, particularly given the adversarial nature of ad fraud, often described as a "cat-and-mouse" game in which fraudsters constantly develop new methods to evade detection. This helps provide a rationale for why Google maintains a large Ad Traffic Quality team dedicated to researching emerging threats and improving its fraud detection capabilities (Google Ads 2025) and its latest advancements in Generative Artificial Intelligence (GenAI) and investment in large language models (LLMs) for its fraud detection (Adegbola 2025). We discuss the results in Proposition 3.

One prevalent regulatory practice in the online advertising industry involves reducing the number of malicious publishers through legal deterrents and enforcement policies (Braun and Eklund 2019). For example, Google regularly scrutinizes its AdSense publishers and removes the suspicious ones from its network (Google Ads 2025). Given its widespread popularity and profound implications on ad fraud deterrence, it is important to gain an in-depth understanding of the effectiveness and implications of this policy. Therefore, we consider the following question. *Does reducing malicious publishers help decrease ad fraud and improve an ad network's profit?*

Although many believe that eliminating malicious publishers helps remove the "bad apples" from the ecosystem, thereby reducing ad fraud and benefiting the ad network (Jackson 2025), we show that this belief may not always be accurate. In Proposition 4 and Proposition 5, we find that under certain conditions, reducing the number of malicious publishers can fail to lower fraudulent ad traffic and may reduce the ad network's profits. This result underscores the limitations of the strategies aimed at fully eradicating ad fraud from the ecosystem. It helps explain why, despite extensive efforts to remove fraudulent actors, the overall rate of fraudulent ad traffic remained persistently high (Fraud Blocker 2025) and why the loss because of ad fraud continues to rise over the years (Dogtiev 2025).

We also extend the model in several directions to demonstrate that our core insights continue to hold when some settings are changed and obtain more intriguing results. For example, although the main model considers that advertisers can accurately estimate their ad valuation, in Section 5.1, we relax this assumption and consider a more general setting where advertisers' estimation may be imperfect. Our analysis reveals that the ad network and publishers can benefit from manipulating the advertisers' estimation of the true valuation. An important takeaway is that advertisers should not rely solely on the performance metrics reported by the ad network. Instead, they should implement independent mechanisms for evaluating campaign effectiveness and regularly monitor the performance of ad campaigns using unbiased, third-party, or in-house analytics.

The remainder of this paper is organized as follows. The next section briefly reviews the relevant literature and highlights our contributions. In Section 3, we introduce essential model elements and setup. In Section 4, we present the results of our analysis, outline the intuition behind the main results, and discuss their managerial implications. In Section 5, we examine three extensions to the main model to demonstrate the robustness of the results to our modeling choices. The paper concludes with managerial implications and possible avenues for future research in Section 6.

2. Literature Review

Our work is mainly related to two literature streams: (1) online ad fraud and (2) general fraud detection and economics. We briefly describe relevant work in the two streams and compare and contrast our work with the existing literature to highlight our contributions in the following two subsections.

2.1. Online Ad Fraud

Despite its tremendous growth in recent years (Mookerjee et al. 2012, Kumar et al. 2020, Liu et al. 2020), online advertising has been increasingly plagued by problems of ad fraud (Kshetri and Voas 2019). Specifically, with improved technology comes the problem of improved methods to impersonate genuine users' visits and increase fraudulent ad traffic online (Pu et al. 2022). Further, the more significant trend of advanced artificial intelligence-enabled and machine learning-enabled ad fraud poses immediate problems for ad networks and advertisers (Zhu et al. 2017b). For this reason, a sizable literature has been developed to study how to design or improve the techniques to detect fraudulent ad traffic (e.g., Nandini 2019).

Because underlying economic incentives also drive fraud motives, a significant body of literature examines the economics of ad fraud and its deterrence. Because fraud generation mechanisms differ across different models, they typically focus on ad fraud arising from a particular payment model (e.g., Asdemir et al. 2008, Mungamuru et al. 2008). Among the popular payment models (e.g., cost per mille (CPM), cost per click (CPC), and cost per action (CPA)), the CPC model is the most popular one, and the corresponding ad fraud, known as click fraud, has thus received the most attention in the literature. Wilbur and Zhu (2009) analyzes the effects of click fraud and shows that it is sometimes beneficial to allow for some fraud. Chen et al. (2015) studies click fraud in a setting of information asymmetry and finds that the CPC model could be unsustainable in certain conditions.

Although these studies focus on the economic aspect of ad fraud, they usually neglect that an ad network also possesses a technological tool to discipline the publishers' malicious actions. An ad network can configure its fraud detection technology stringently or leniently to determine ad-traffic quality. Ignoring the interplay with the effects of this technology configuration can lead to suboptimal policies by ad networks, which is perhaps why the ad fraud rate remains high despite the countermeasures. Our paper fills this gap by building a game-theoretic model that considers an ad network's strategic decisions on its technology configuration and economic incentives.

In summary, the key contributions of this paper to the online ad fraud literature are as follows. (i) Considering both decisions, our study reveals how they should be coordinated to deter fraudulent traffic. Notably, our analysis uncovers an intriguing interplay between the effects of technology and economic tools in disciplining fraud motives in a changing market. Because of this, some of the common beliefs about ad fraud deterrence are no longer true, and the pure economic countermeasures often adopted by ad networks can be suboptimal (see Propositions 1 and 2). (ii) Past research has advocated draconian policies, such as banning malicious publishers, to curb ad fraud (Nandini 2019). We investigate this banning policy and find that it not only fails to reduce fraudulent traffic but also hurts an ad network's profit under certain market conditions, suggesting that this prevalent regulatory practice may sometimes be ineffective in controlling ad fraud (see Proposition 4 and Proposition 5). (iii) Similar to previous studies (Wiatr et al. 2019), our result substantiates ongoing concerns over ad networks' negligence regarding fraud activities. Nevertheless, we provide an alternative explanation. Instead of focusing on an inherent lack of economic incentives (Cai et al. 2020), the ad network's leniency in combating ad fraud can also be attributed to technical deficiency.

2.2. General Fraud Detection and Economics

In general, *fraud* is defined as an activity in which the perpetrator seeks to obtain financial gain through illicit methods (Abdallah et al. 2016). Research on fraud and its impacts has been broadly conducted along two dimensions: (1) fraud detection and prevention (Kumar et al. 2019) and (2) economic analysis of incentives for different actors to commit fraud. Fraud detection and prevention measures have grown drastically in financial institutions, with advanced AI and ML techniques being applied as countermeasures (Sailusha et al. 2020). Although technological advancements have heralded the introduction of methods related to nature-inspired computing and intelligent systems (Bouayad et al. 2019), fraud detection remains a significant challenge. Firms have discovered that fraud detection systems are prone to failure, may have a low accuracy rate, or generate many false alarms (Ji et al. 2016).

Research in this area has focused mostly on designing punitive monetary measures to elicit desired behavior rather than monetary rewards or incentives. For instance, Bennett et al. (1994) considers cases of regulatory measures, such as providing financial rewards to healthcare providers for adhering to quality standards while implementing severe punitive fines for violators. Other works consider implementing trust-based regulations, where lower penalties or in some cases, no penalties but relationship-based negotiations can be beneficial in eliciting desired honest behavior (Mendoza and Wielhouwer 2015). However, using financial incentives to elicit desired behavior in combination with nonfinancial punitive mechanisms has usually not been explored.

The economics of fraud in other contexts has also been an active area of research. From the earliest papers on the impact of fraud on markets (Macey and Miller 1990), where the major implication was how efficient markets promote the idea of trading for security against fraud activity, to more recent papers, where the analysis is of the impact of fraud on executive boards and decision making (Fich and Shivdasani 2007), research has evolved from a more market-oriented approach to analyzing impacts on major industries and related stakeholders. However, these models fail to consider how improvement in fraud generation mechanisms can influence the response of the impacted firm.

To summarize, our contributions to this stream of literature are as follows. (i) Our model presents a situation where an ad network can use both monetary and nonmonetary incentives to regulate the market as opposed to using only monetary or nonmonetary measures individually (see Propositions 1 and 2). (ii) The literature shows that improving technology typically has a diminishing marginal return (Ravichandran et al. 2017). In contrast, we show that in the ad network's case, improving technology can, interestingly, have increasing marginal returns when fraud generation efficiency improves (see Proposition 3). Insights from our analysis are critical for ad networks, especially given the steep rise in ad fraud across major platforms and publishers.

3. Model Description

An ad network, such as Google Ads, serves as an intermediary by effectively matching publishers' ad inventory with appropriate advertisements (Parti 2020). It plays a critical role in resolving coordination and efficiency challenges inherent in the online advertising marketplace (Choi et al. 2020). Consequently, consistent with anecdotes and the market structure in

the literature (D'Annunzio and Russo 2020), we consider three key sets of players in our model: (1) *advertisers* that pay for ad traffic to reach a targeted audience, (2) *publishers* that supply the inventory and infrastructure necessary to deliver ads, and (3) an *ad network* that connects advertisers to publishers by aggregating the ad supply from publishers and selling it to advertisers.[3] Next, we discuss the essential model elements. The key notations are summarized in Online Appendix EC.1.

3.1. The Publishers

As ad inventory owners, publishers often monetize their ad spaces by participating in an ad network platform. By agreeing to host the ads allotted by the ad network, publishers receive a share of ad revenues from advertisers for the appropriate ad traffic generated from users' visits (D'Annunzio and Russo 2024). In the example of Google's AdSense network, more than 58.5 million websites use it worldwide (Gibbons 2025). These publishers use Google Ad Manager to manage digital advertising and derive substantial revenue from the ad traffic (e.g., impressions, clicks, acquisitions, or actions) generated on their websites (Cai et al. 2020).[4] Although ad networks aim to attract ad traffic from genuine users' visits, the opportunity to earn through nongenuine visits also attracts fraudulent behavior from some publishers, which can use various tools, such as bots or other automated programs, to substantially inflate ad traffic and make lucrative profits (Jackson 2025). Because the tendency to generate fraudulent ad traffic can vary across publishers because of their idiosyncratic characteristics, we consider a publisher to be either *nonmalicious* (N) or *malicious* (M).

Nonmalicious publishers hardly generate fraud traffic because of a lack of malicious motives, technological limitations, or prohibitively high costs (e.g., moral or reputational costs) (Fulgoni 2016). For example, digital journalism by well-known publishers, such as *Forbes* and *The Wall Street Journal*, has negligible reports of ad fraud (Braun and Eklund 2019). Moreover, the potential reputational damage and associated revenue loss from being implicated in fraudulent activity serve as strong disincentives for most nonmalicious publishers (Zhu et al. 2017a, Nandini 2019). Thus, we consider that nonmalicious publishers generate only valid ad traffic. In contrast, malicious publishers can also generate fraudulent or invalid ad traffic to increase their ad revenue if doing so is profitable (Almeida and Gondim 2018, Fou 2020b).[5] We discuss these two types of publishers and their strategies in Sections 3.1.1 and 3.1.2, respectively.

3.1.1. Nonmalicious Publishers. Although publishers have different tendencies to generate fraud ad traffic, in general, the ad network does not know the *true type* (malicious or nonmalicious) of a publisher with certainty (Braun and Eklund 2019, Nagaraja and Shah 2019) other than some well-known and reputable nonmalicious publishers (e.g., major news agencies). Correspondingly, we consider that the ad network designates some reputable and well-known nonmalicious publishers to a "whitelist" and the remaining nonmalicious publishers, whose true type is uncertain to the ad network, to an "unknown" group.[6] For ease of exposition, we refer to the former specifically as the "whitelisted" publishers, and we refer to the latter as nonmalicious publishers unless otherwise noted.

Publishers receive a portion of the ad network's advertising revenue from charged ad traffic to incentivize participation (Zhu et al. 2017a). Let p_W be the ad network's *payment* to a whitelisted publisher for each ad-traffic unit, and let D_W be the amount of ad traffic generated by a whitelisted publisher.[7] The ad traffic of whitelisted publishers is not subject to the ad network's inspection or classification because of their known nonmalicious nature. Thus, the utility of a whitelisted publisher from participating in the ad network is

$$u_W = p_W D_W. \tag{1}$$

Unlike the whitelisted group, the "unknown" group consists of both nonmalicious (those that are not whitelisted) and malicious publishers.[8] Although the ad network does not know the true type of these publishers, it can classify publishers into malicious or nonmalicious based on the data collected from various sources.

The classification has errors because malicious publishers may be misclassified as nonmalicious and vice versa. To capture the imperfect nature of classifications, we use the index $i \in \{M, N\}$ to denote the publisher's *true* or *genuine* type in the "unknown" group as either malicious ($i = M$) or nonmalicious ($i = N$) and the index $j \in \{M, N\}$ to denote its *classified type* by the ad network. Furthermore, we refer to the publishers classified as malicious as *type* M ($j = M$) publishers and those classified as nonmalicious as *type* N ($j = N$) publishers. For simplicity, we consider that the classifier has symmetric performance on malicious and nonmalicious publishers, and we denote α as the classifier's sensitivity: that is, with probability α that a publisher's *classified type* is the same as its *true type*. We let $\alpha \in [\frac{1}{2}, 1)$ to indicate that this classifier is at least as good as a random classifier.

Because malicious publishers in the "unknown" group may produce ad fraud but their true type cannot be perfectly identified, the ad network employs a fraud detection technology to detect and filter out fraudulent *ad traffic* (Chen et al. 2012b). The detection and classification at the ad-traffic level are also imperfect;

that is, the ad network may misidentify some valid ad traffic as fraud and vice versa. Thus, nonmalicious publishers in the "unknown" group are hardly compensated for all genuine ad traffic that they generate (Zhu et al. 2017a). To formally characterize the ad network's fraud detection at the ad-traffic level, we define the ad traffic classified as fraudulent as the *positive* outcome and those classified as valid as the *negative* outcome. The former is filtered out and not charged to the advertisers, and the latter is the *charged ad traffic* to the advertiser.

We denote $s_j, j \in \{M,N\}$, as the *sensitivity* or *true-positive rate* (TPR) of the technology (i.e., the probability that fraud ad traffic is correctly identified (positive outcome)). Similarly, $(1-s_j)$ is the *false-negative rate* (i.e., the probability that fraud ad traffic is incorrectly identified as valid (negative outcome)). The subscript j indicates that the ad network may set different sensitivities to detect fraud ad traffic for the type M ($j=M$) and type N ($j=N$) publishers. We further define y_j as the *false-positive rate* (FPR) and $(1-y_j)$ as the *specificity* or *true-negative rate* of the technology. The TPR (sensitivity) and FPR are usually related under a given classification technology, and we provide a detailed discussion on this in Section 3.2.

We denote p_j as the ad network's *payment* for each *charged ad traffic* (i.e., ad traffic classified as valid by the fraud detection technology) from type j publishers, where $j \in \{M,N\}$ is the classified type. Let D_N be the amount of valid ad traffic generated by a nonmalicious publisher.[9] The expected revenue of type j nonmalicious publishers is $p_j(1-y_j)D_N$, where $(1-y_j)$ is the true-negative rate and captures the notion that some truly valid ad traffic may be incorrectly classified as fraud because of the imperfect fraud detection system. Such misclassifications can also lead to additional costs as the ad network may impose penalties on publishers (e.g., forfeiting ad revenue, account suspension) because of the detected ad fraud (D'Annunzio and Russo 2020).[10] This penalty is positively related to the classified fraud activity and the associated economic damages; thus, we define it as $wy_jp_jD_N$, where $y_jp_jD_N$ is the expected economic value of the classified ad fraud and $w>0$ is the cost parameter of the penalty.[11] Thus, the expected utility of a type j nonmalicious publisher from participating in the ad network is as follows:

$$u_{Nj} = p_j(1-y_j)D_N - wp_jy_jD_N. \tag{2}$$

In addition to participation in the ad network, publishers may also directly contract with advertisers and handle ad placements and scheduling using in-house or third-party services (Jackson 2025). Because the ability to monetize ad spaces without the ad network can vary significantly across nonmalicious publishers (D'Annunzio and Russo 2020), we consider that nonmalicious publishers (both the whitelisted group and the "unknown" group) have heterogeneous reservation utility that is uniformly distributed over [0, 1]. Thus, only those with a reservation utility less than or equal to u_W or u_{Nj} will participate in the ad network. In reality, although many publishers participate in ad networks, some still use in-house or third-party services to handle ad placements (D'Annunzio and Russo 2024, Jackson 2025). As such, our analysis focuses on the parameter space for the case where a subset of nonmalicious publishers is incentivized to participate in the ad network. For this purpose, we impose a technical condition in Online Appendix EC.4 to eliminate the cases where all nonmalicious publishers participate (i.e., full participation) and none of the nonmalicious publishers participate (i.e., zero participation) from the equilibrium outcome.

3.1.2. Malicious Publishers. Because malicious publishers may inflate their genuine ad traffic with fraudulent ad traffic to increase profits, we consider that their revenue arises from two sources. First, a malicious publisher can be endowed with some valid ad traffic denoted by $D_M = bD_N$, where $b>0$ and D_N is the valid ad traffic generated by a nonmalicious publisher.[12] For a type j malicious publisher, $j \in \{M,N\}$, its expected revenue from the valid ad traffic is $p_j(1-y_j)D_M$, where p_j is the ad network's payment for traffic deemed valid to publishers of type j and $(1-y_j)$ is the specificity (i.e., probability that valid traffic is correctly classified) of the ad fraud detection system for the corresponding type. Second, malicious publishers may also produce additional fraudulent ad traffic. Let x_j be the amount of the fraudulent ad traffic generated by a type j malicious publisher. The expected revenue from such traffic is $p_jx_j(1-s_j)$, where s_j is the *sensitivity* (TPR; i.e., $(1-s_j)$ is the probability that fraudulent ad traffic is misclassified as valid). In other words, the expected revenue from fraud comes from the fraudulent ad traffic that "bypasses" the ad network's detection system.

Although fraudulent ad traffic helps inflate ad revenue, it can also incur substantial costs to malicious publishers (Faou et al. 2016). First, there is a direct cost in producing fraudulent ad traffic that can bypass the detection system. Malicious publishers can incur significant expenses, such as developing, purchasing, maintaining, and deploying automated programs or recruiting a large group of "fraudsters" to generate ad fraud (Jackson 2025). Prior studies, such as Wilbur and Zhu (2009) and Zhu et al. (2017a), consider the cost of producing fraudulent traffic as an increasing and convex function. Thus, consistent with the literature, we consider cx_j^2 ($c>0$) as the cost of producing x_j fraud ad traffic for a type j malicious publisher.[13]

Second, once fraudulent activities are detected, malicious publishers may face severe penalties from the ad network, such as forfeiting of ad revenue, account suspension, or removal from the content network (D'Annunzio and Russo 2020). The penalty is positively related to the activities identified as fraud by the ad network and the associated economic damages. Because classifications are imperfect, the total classified fraud ad traffic for a type j malicious publisher is $y_j D_M + s_j x_j$, where the first term is valid traffic misclassified as fraud (i.e., false positives) and the second term is the detected fraud traffic (i.e., true positives). As such, we define the penalty as $wp_j(y_j D_M + s_j x_j)$, where $p_j(y_j D_M + s_j x_j)$ is the expected economic value of classified ad fraud and $w > 0$ is the cost parameter of the penalty. The expected utility of type j malicious publishers is expressed as

$$u_{Mj} = \underbrace{p_j(1-y_j)D_M}_{\text{Revenue from valid ad}} + \underbrace{p_j x_j(1-s_j) - cx_j^2}_{\text{Payoff from ad fraud}} - \underbrace{wp_j(y_j D_M + s_j x_j)}_{\text{Penalties on classified fraud}}, \tag{3}$$

where the first two terms on the right-hand side are the expected revenue and payoff from valid and fraudulent ad traffic, respectively, and the last term is the expected penalties.

With a variety of methods ranging from botnets (Kahn 2020) on one end to click fraud farms (Chen et al. 2015) on the other end, we term c the *fraud generation efficiency* to indicate how malicious publishers can manage the cost of generating fraud traffic efficiently. A lower c implies that a more cost-effective method generates fraud traffic, whereas a higher c implies the opposite. For simplicity, we set $D_N = D_W = 1$ and $D_M = b$ in the subsequent analysis. This shifts the boundary conditions, but it does not affect the main results qualitatively. As shown in Equations (2) and (3), the expected utilities of publishers in the unknown group depend on the values of s_j and y_j, whose values are related when the technology quality is fixed (Cavusoglu et al. 2009). In other words, one cannot change the value of one parameter (e.g., y_j) without affecting the other parameter (e.g., s_j) under a given technology quality. Next, we discuss the characterization of the quality profile of the ad network's ad fraud detection technology and the associated configuration choice.

3.2. Quality Profile and Configuration of Ad Fraud Detection Technology

Although ad networks employ various methods to detect fraudulent traffic, their imperfect nature inevitably yields classification errors essential to players' decisions. We use the framework of the receiver operating characteristics (ROC) curve as used in Mookerjee et al. (2011) to provide a formal study of the strategic interactions in the game. Using the framework, we capture both the *quality* of an ad fraud detection system (e.g., high versus low) and the associated *technology configuration* decisions (e.g., stringent versus lenient) to examine how they would influence the game dynamics.

An ROC curve is a widely used performance measurement graph for evaluating classification models (Zhu et al. 2017a). It is created by plotting the sensitivity (TPR or s_j) against the false-positive rate (FPR or y_j) at various thresholds. There are several parametric methods that model ROC curves analytically. For example, Cavusoglu et al. (2009) derives a nonlinear ROC function to characterize the technology profile of a system for classifying hackers and regular users. Clemencon and Vayatis (2009) proposes a method to estimate optimal ROC curves using linear approximation.

These studies strictly analyze the detection system's raw quality (i.e., r) on the ROC curve, but fraud generation technology may affect performance. In our study's context, an efficient fraud generation technology (i.e., c is lower) allows more fraud ad traffic to circumvent the detection system. As such, we adapt the work in Clemencon and Vayatis (2009) and approximate the ROC curve of the ad network's ad fraud detection technology using a linear function below:

$$y_j = (1 - \tilde{r})s_j, \tag{4}$$

where $\tilde{r} = \phi cr$. We denote by $\tilde{r}$ the *effective quality* of the ad network's detection system, which depends on both r, the *raw quality* of the detection technology, and c, the malicious publishers' fraud generation efficiency.[14] The effective quality $\tilde{r}$ increases as the raw technology quality improves (r increases) or decreases as the publishers' fraud generation efficiency improves (c decreases). A higher effective quality means that the detection system generates fewer false positives (i.e., misclassifying valid ad traffic as fraud) at a given sensitivity level. Note that ϕ is a scaling factor to ensure that the optimal configuration is neither negative (technically infeasible) nor leading to a trivial case where the amount of fraud traffic is zero as discussed in Online Appendix EC.4.

The quality profile of a classification technology results from a long-term, continuing effort and investment. Thus, it cannot be easily improved quickly or without a substantial cost (Richet 2022). Nevertheless, for a fixed classification technology, the ad network may choose different technology configuration points, which are pairs of (s_j, y_j) points on the ROC curve, to maximize its profits.[15] We use s_j (sensitivity) to represent the ad network's technology configuration for type j publishers. The corresponding y_j (FPR) is obtained given the functional relationship as in Equation (4).

Generally, a high-sensitivity configuration (i.e., a higher s_j) helps gate fraud traffic and is considered a

stringent configuration. In contrast, a low-sensitivity configuration (i.e., a lower s_j) has the opposite effect of admitting fraud traffic and is a *lenient* configuration. Although a stringent configuration helps discourage fraud motives, it can also result in a significant misdetection of valid traffic as the corresponding FPR (y_j) can also be high. Similarly, a lenient configuration may increase publishers' incentives to commit fraud but may also lead to better identification of valid traffic.

3.3. The Advertisers

The ad network receives payments from advertisers and shares a portion of the revenue with the publishers for using their ad spaces (Gordon et al. 2021). Let v be the advertisers' *ad valuation* for a *valid* or *genuine* ad-traffic unit (e.g., an impression, a click, or an action). Fraud detection systems are imperfect, so advertisers are charged for legal (true-negative) and fraudulent (false-negative) ad traffic. Publishers' participation, ad fraud generation decisions, and the ad network's technical configuration might result in advertisers being charged for more or less ad traffic than they receive. As such, advertisers often update their valuation of the charged ad traffic and revise their ad valuation during an ad campaign (Sun and Zhu 2013, Chen and Stallaert 2014).

Two crucial performance metrics can affect the advertisers' ad valuation update. The first metric is *the amount of charged ad traffic*, which can be obtained using the tools provided by the ad network to track campaign performance. For example, Google offers ad-performance monitoring tools to help advertisers make investment decisions (Google 2025a). The second metric is *the amount of valid ad traffic*. Advertisers can often estimate this metric using campaign metrics and reports provided by the ad network and analytic tools from either their internal team or third-party services (Prokopets 2021).[16] For example, advertisers can use page-visit data to track ad-traffic conversions after visitors arrive at their websites and evaluate the campaign (Chen et al. 2015).

Let T_i, where $i \in \{M,N\}$, be the number of malicious (M) and nonmalicious (N) publishers in the unknown group, respectively, and $\sigma_W T_N$ ($\sigma_W \geq 0$) be the number of "whitelisted" publishers. The valid ad traffic from whitelisted publishers is $n_{gW} = u_W \sigma_W T_N$, where u_W is the proportion of participating whitelisted publishers. Because a type j publisher can be either malicious or nonmalicious because of imperfect classifications, we denote by $n_{gj} = \sum_{i=M,N} n_{gij}$ the valid ad traffic from type j publishers, where the subscript $i \in \{M,N\}$ indicates the publishers' true type. Thus, the valid ad traffic from type M publishers is $n_{gM} = n_{gNM} + n_{gMM}$, where $n_{gNM} = (1-\alpha)u_{NM}T_N$ and $n_{gMM} = \alpha b T_M$ are valid ad traffic from type M nonmalicious and malicious publishers, respectively, and α is the accuracy of classifying publishers' types as discussed earlier. The valid ad traffic from type N publishers is $n_{gN} = n_{gNN} + n_{gMN}$, where $n_{gNN} = \alpha u_{NN} T_N$ and $n_{gMN} = (1-\alpha) b T_M$ are the valid ad traffic from type N nonmalicious and malicious publishers, respectively. The total amount of valid ad traffic is $n_g = n_{gW} + \sum_{j=M,N} n_{gj}$. Similarly, the amount of fraudulent ad traffic from type j malicious publishers, denoted by n_{fj}, is $n_{fM} = \alpha x_M T_M$ and $n_{fN} = (1-\alpha) x_N T_M$, and the total amount of fraudulent ad traffic generated by malicious publishers is $n_f = n_{fM} + n_{fN}$.

Although all ad traffic from the whitelisted group is correctly charged, for publishers in the unknown group, not all of their valid traffic is correctly identified, and some fraudulent ones can go undetected because of the imperfect classification by fraud detection technology. Thus, the total amount of charged ad traffic is $n_c = n_{gW} + \sum_{j=M,N} (n_{gj}(1-y_j) + n_{fj}(1-s_j))$, where the first and second terms are the amount of charged ad traffic from the whitelisted group and the unknown group, respectively. The latter consists of the ad traffic that can be truly valid or invalid (but misclassified as valid) and is subject to the sensitivity (s_j) and false-positive rate (y_j) configured for the respective classified type j as explained in earlier subsections. Because advertisers can observe the amount of charged ad traffic (n_c) and assess the return from ad campaigns (n_g) using the supplied information and various analytics tools and services, we consider the advertisers' *updated ad valuation* after accounting for the fraud traffic and the classifications by the ad network's detection technology to be

$$\tilde{v} = v\left(\frac{n_g}{n_c}\right). \tag{5}$$

Intuitively, as n_g increases, the advertisers receive more valid ad traffic, and thus, they are willing to pay more for each charged ad traffic, in which case the updated ad valuation increases. Similarly, as n_c increases, the advertisers are charged for more ad traffic, and thus, they lower their valuation for each charged ad traffic. If advertisers are charged for more traffic than the valid ones that they truly receive (i.e., $n_g < n_c$), then a *downward revision* of their ad valuation occurs (i.e., $\tilde{v} < v$). Nevertheless, if advertisers receive more valid traffic than what the ad network charges (i.e., $n_g > n_c$), then there is a *upward revision* to their ad valuation (i.e., $\tilde{v} > v$). Lastly, where $n_g = n_c$, there is no revision to the advertisers' ad valuation (i.e., $\tilde{v} = v$). We further consider the advertisers' overestimation and underestimation of their ad campaign performance (i.e., n_g) in an extension in Section 5.1 and show that they do not change the main takeaways in the paper qualitatively.

3.4. The Ad Network

As the intermediary, the ad network receives a payment $\tilde{v}$ from advertisers for each charged ad traffic and shares it with publishers (D'Annunzio and Russo 2020). We denote p_W and p_j as the payment for each charged ad traffic to the whitelisted publishers and to the type j publishers of the unknown group, respectively.[17] Thus, the ad network's expected net ad revenue is the sum of net revenue from the whitelisted publishers (i.e., $(\tilde{v}-p_W)n_{gW}$) and type j publishers (i.e., $\sum_{j=M,N}(\tilde{v}-p_j)(n_{gj}(1-y_j)+n_{fj}(1-s_j))$), where n_{gW} is the ad traffic of whitelisted publishers and n_{gj} and n_{fj} are the valid and invalid ad traffic, respectively, from type j publishers in the unknown group as defined earlier in Section 3.3.

As discussed earlier, publishers may be subject to penalties, such as forfeiture of ad revenue, account suspension, termination, or even legal action, if they engage in fraudulent activities and violate the ad network's policies. In such cases, a portion of these penalties (e.g., forfeited revenue and compensatory or punitive damages from lawsuits) may be retained by the ad network and become its revenue. Let $\theta \in (0,1)$ be the fraction of penalty received by the ad network. Its expected revenue from penalties is $\theta w p_j(s_j n_{fj} + y_j n_{gMj})$ for type j malicious publishers and $\theta w p_j y_j n_{gNj}$ for the type j nonmalicious publisher, where $s_j n_{fj} + y_j n_{gMj}$ and $y_j n_{gNj}$ are the classified fraud ad traffic for each group and wp_j is the per-unit penalty for type j publishers.[18]

In addition to affecting the amount of charged ad traffic (n_c) and the advertisers' ad valuation ($\tilde{v}$), misclassifications of valid ad traffic can incur significant costs to the ad network. We term such cost the *mislabeling cost*, which can arise from three primary sources. First, publishers often contest the mislabeled valid ad traffic. These contests are typically handled by the ad network's internal team using a more sophisticated suite of tools or even manual investigations (Karpenkova 2020), incurring substantial costs to the ad network. Second, ad networks rely on the traffic-level classification results to identify fraudulent publishers (Trajcheva 2023), which are subject to strict sanctions, such as accounts deactivation (Google 2025b). More mislabeled valid ad traffic increases the chance of nonmalicious publishers being incorrectly labeled as fraudsters. Although these publishers can usually restore their accounts after appealing to the ad network, they may demand financial compensation from the ad network for the wrongful sanctions (Hu et al. 2016). For example, it cost Google $11 million to settle a lawsuit with only some of its AdSense publishers (Khoury 2019). Third, mislabeling substantial amounts or value of valid ad traffic as fraud can damage the ad network's reputation and undermine publishers' trust (Jacob 2023).

Thus, the ad network's mislabeling cost is related to the financial loss because of the misclassified valid ad traffic as mentioned above. Specifically, the expected amount of mislabeled valid ad traffic by type j publishers is $y_j n_{gj}$, where n_{gj} is the amount of valid ad traffic from type j publishers and y_j is the respective false-positive rate. The mislabeling cost is $k\sum_{j=M,N} y_j n_{gj} p_j$, where p_j is the ad network's payment for type j publishers and $k \in (0,1)$ is the cost parameter. As a result, the ad network's profit function can be expressed as follows:

$$\pi = \underbrace{(\tilde{v}-p_W)n_{gW} + \sum_{j=M,N}(\tilde{v}-p_j)(n_{gj}(1-y_j)+n_{fj}(1-s_j))}_{\text{Ad revenue from whitelist and unknown groups}} + \underbrace{\theta w \sum_{j=M,N} p_j\left(s_j n_{fj} + y_j \sum_{i=M,N} n_{gij}\right)}_{\text{Revenue from penalties}} - \underbrace{k\sum_{j=M,N} y_j n_{gj} p_j}_{\text{Mislabeling cost}}. \tag{6}$$

The ad network makes strategic decisions on the configuration of its ad fraud detection technology (s_j) and payment to its publishers of type j (p_j) and its publishers in the whitelist group (p_W) to maximize its profits.

3.5. Game Sequence

As discussed above, an ad network can influence the publishers' participation and fraud incentives using its economic tool (payment to publishers) and technological tool (configuration of its fraud detection technology). Generally, an ad network can calculate and anticipate the responses from the publishers and advertisers using various information from both internal and external sources and historical data (D'Annunzio and Russo 2020). Therefore, stage 1 of the game begins with the ad network offering a contract to publishers and selecting a technology configuration with full knowledge of the publishers' responses in their participation and fraud generation activities and the advertisers' responses to their valuation of ad traffic.

In our setting, in stage 1, the ad network decides on the payment that it offers to whitelisted publishers (p_W) and the type j publishers (p_j) in the unknown group. At this stage, it also strategically decides on its technology configuration for type j publishers (that is, the sensitivity (s_j) and the respective false-positive rate (y_j) in the ROC curve) given the quality profile of its ad fraud detection technology ($\tilde{r}$). The ad network's optimization problem in stage 1 is given in Equation (6).

In stage 2, conditional on their classified type $j \in \{M,N\}$, publishers make strategic decisions in response to the ad network's technology configuration and payment decisions. Specifically, malicious publishers determine the level of fraudulent traffic to generate

denoted by x_j, and nonmalicious publishers (both in the whitelisted and unknown groups) decide whether to participate in the network, weighing the expected utility against their outside options. As noted, the ad network often provides publishers with access to tools for monitoring and managing digital advertising performance (e.g., Google Ad Manager for AdSense publishers), enabling them to effectively learn or infer the network's configuration and payment policies (Cai et al. 2020). The malicious publishers' optimization problem in stage 2 is given in Equation (3), whereas the participation decision for the nonmalicious publishers is given in Equation (2).

Given the publishers' fraud generation activities and participation decisions in stage 2 and the ad network's stage 1 decisions, the amount of valid ad traffic (n_g) in the ad network and the amount of charged traffic (n_c) are realized. As discussed earlier, advertisers use this information to update their ad valuation ($\tilde{v}$), which is their payment to the ad network for each charged ad traffic (Sun and Zhu 2013, Chen and Stallaert 2014). The advertisers' ad valuation update in stage 3 is given in Equation (5). Finally, the payoffs are realized. The game sequence is depicted in Figure 1.

4. Results and Insights

In this section, we first examine the ad network's optimal decisions concerning its technology configuration and payment, and we solve the equilibrium for the game. We then investigate how the perturbation of essential market parameters impacts these equilibrium decisions and derive insights into the ad network's responses to the changing market. Finally, we examine the ad network's regulatory practice from the perspective of the ad network to shed light on debates on its effectiveness against ad fraud and offer important managerial implications and actionable recommendations.

We use the approach of backward induction to solve the game. We first begin with characterizing the advertisers' responses in stage 3 given the publishers' participation and fraud decisions in stage 2. We then analyze the malicious publishers' ad fraud generation decisions and nonmalicious publishers' participation decisions in stage 2. Finally, we examine the ad network's technology configuration and payment decisions in stage 1.

4.1. The Ad Network's Technology Configuration and Payment Decisions

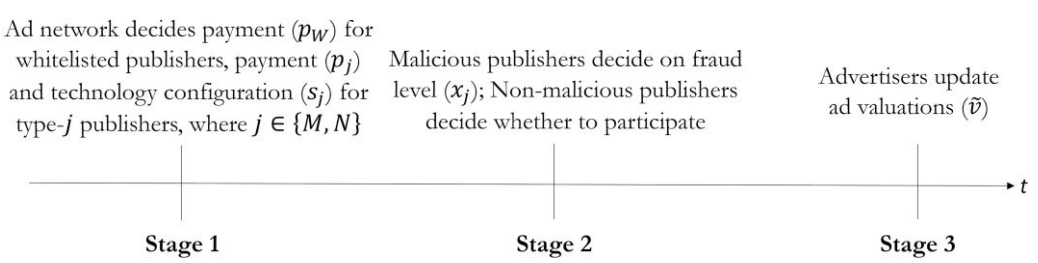
It is widely believed that ad networks may lack strong incentives to filter out fraudulent ad traffic as such traffic can still contribute to revenue generation (Fraud Blocker 2022). For example, Uber alleges that one of its ad network partners knowingly allowed ad fraud from up to 100 publishers in procuring $70 million of ad inventory over a two-year period (Taylor 2024). Motivated by such examples, one may ask whether an ad network deliberately admits all traffic or willingly allows undetected fraudulent ad traffic to maximize self-interest. To address this question, we examine the ad network's equilibrium decisions to unravel how technological and economic tools are coordinated to curb ad fraud and maximize self-interest.

Based on the game sequence outlined above, there exist several possibilities with respect to the decision variables s_j and y_j. Of these possibilities, $0 < \{s_j, y_j\} < 1$ represents an interior case where the ad network leverages its technological configuration strategically for both type N and type M publishers. In addition, we consider two other simple configuration policies. (1) The ad network admits all traffic from type N publishers, whereas it does fraud detection on the type M group (i.e., $s_N = y_N = 0$ and $0 < \{s_M, y_M\} < 1$). (2) The ad network admits all traffic from type N group but bans all traffic from type M group (i.e., $s_N = y_N = 0$ and $s_M = y_M = 1$). We characterize the equilibrium outcomes in Lemma 1. All proofs for lemmas and propositions are presented in Online Appendix EC.4.

Lemma 1. *When the ad network can leverage both technology configuration (s_j^*) and payment (p_W^*, p_j^*), where $j \in \{M, N\}$, we have the following.*

1. *The following constitutes an equilibrium outcome where the ad network uses the fraud detection system on ad traffic from both type M and type N publishers*

Figure 1. Game Sequence



(*i.e.*, $0 < \{s_M, s_N, y_M, y_N\} < 1$):

$$p_W^* = \frac{v}{2},$$

$$p_M^* = \frac{A_1(\alpha b T_M(1+w\theta-k)+(1-\alpha)(w+1)T_N v)}{A_2(4(1-\alpha)^2c^2T_N^2(w(1-\theta)+k)^2)},$$

$$p_N^* = \frac{A_3((1-\alpha)wT_M(1-\theta-(\theta+1)\tilde{r})+2\alpha cw(1-\theta)T_N(1-\tilde{r})^2)}{A_4(\alpha k^2T_N(1-\tilde{r})^2-2(1-\alpha)krT_M\phi(1-\tilde{r})-2(1-\alpha)cT_M\tilde{r}^2)},$$

$$s_M^* = \frac{A_5(w(1-\theta)((1-\alpha)T_N v+\alpha bT_M)+2\alpha bkT_M)}{A_6(\alpha bT_M(k+1-\theta(1-k))+(1-\alpha)(1-\theta)T_N v)}, y_M^* = (1-\tilde{r})s_M^*$$

$$s_N^* = \frac{\begin{array}{l}A_7(2\alpha cw(1-\theta)T_N(1-\tilde{r})+2\alpha ckT_N(1-\tilde{r})-2(1-\alpha)\tilde{r}T_M\\+(1-\alpha)wT_M(1-2\tilde{r}-\theta))\end{array}}{A_8(1+k+\tilde{r}(\theta(3-k)+1-k)-\theta(1-k))},$$

$$y_N^* = (1-\tilde{r})s_N^*.$$

This equilibrium is feasible in the parameter space χ.

2. *Outside the parameter space* χ, *there exists a possibility where the ad network treats all ad traffic from type N publishers as valid and uses the fraud detection system on ad traffic from type M publishers only* (*i.e.*, $s_N = y_N = 0$ *and* $0 < \{s_M, y_M\} < 1$). *The equilibrium outcome in this case is characterized below*:

$$p_{W,\chi'}^* = p_W^*,$$

$$p_{M,\chi'}^* = p_M^*,$$

$$p_{N,\chi'}^* = \frac{c(\alpha T_N v-(1-\alpha)bT_M)}{2\alpha cT_N+T_M(1-\alpha)},$$

$$s_{N,\chi'}^* = y_{N,\chi'}^* = 0,$$

$$s_{M,\chi'}^* = s_M^*, y_{M,\chi'}^* = (1-\tilde{r})s_{M,\chi'}^*.$$

This equilibrium is feasible in the parameter space χ'.

3. *Outside the parameter spaces* χ *and* χ', *there could exist a possibility where the ad network treats all ad traffic from type N publishers as valid and all ad traffic from type M publishers as fraudulent* (*i.e.*, $s_N = y_N = 0$ *and* $s_M = y_M = 1$). *The equilibrium outcome in this case is characterized below*:

$$p_{W,\chi''}^* = p_W^*, p_{N,\chi''}^* = p_{N,\chi'}^*,$$

$$s_{N,\chi''}^* = y_{N,\chi''}^* = 0, s_{M,\chi''}^* = y_{M,\chi''}^* = 1.$$

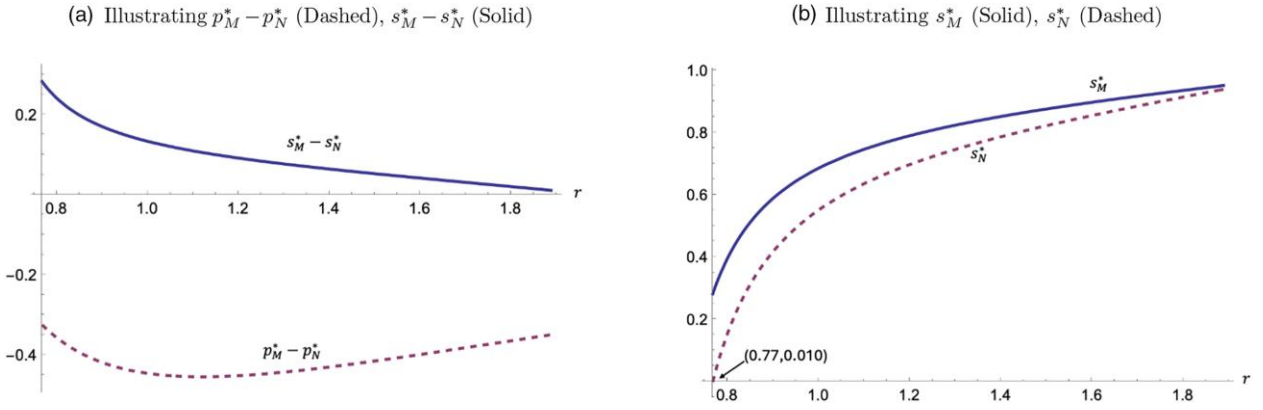
This equilibrium is feasible in the parameter space χ''.

The expressions for $A_1, \ldots, A_8$ *and the parameter spaces* χ, χ', *and* χ'' *are provided in Online Appendix* EC.4. *We use* χ' *and* χ'' *in the subscripts for cases* 2 *and* 3 *to indicate the equilibrium outcomes of the respective cases*.

Lemma 1 characterizes the ad network's equilibrium payment and technology configuration when facing both malicious and nonmalicious publishers under imperfect classification of publisher types and ad traffic. A few points are worth noting. First, the ad network's payment to the whitelisted publishers (p_W^*) is independent of the fraud detection technology and market parameters because as discussed earlier, these publishers can be unerringly identified, and their ad traffic is exempted from the ad network's inspection. Second, when the ad network is completely uninformed about the publishers' true type (i.e., $\alpha = \frac{1}{2}$), the ad network's equilibrium payment and technology configuration for these two classified types are identical (i.e., $p_M^* = p_N^*$ and $s_M^* = s_N^*$). Third, when the ad network can learn the publishers' type (i.e., $\alpha > \frac{1}{2}$), it usually adopts a stricter configuration (sets a higher sensitivity) for type M publishers and provides a higher payment to type N publishers as shown in Figure 2(a). Last, improvements in fraud detection technology prompt the ad network to respond with a stricter configuration to both types: that is, s_j^* increases in r as illustrated in Figure 2(b).

The last two cases characterized in Lemma 1 are degenerate cases that render the analysis of strategic technology configuration moot, particularly when examining the ad network's strategic responses to

Figure 2. (Color online) Equilibrium Payment and Technology Configuration with Parameter Values $w = 0.05, c = 0.5, k = 0.2, \phi = 1, T_N = 2, T_M = 1, \alpha = 0.65, b = 1, v = 2, \theta = 0.2, \sigma_W = 0.3$



Notes. Values of parameters are $\{v = 2, k = 0.2, c = 0.5, \alpha = 0.65, w = 0.05, T_N = 2, T_M = 1, \phi = b = 1, \theta = 0.2, \sigma_W = 0.3\}$. (a) $p_M^* - p_N^*$ (dashed line) and $s_M^* - s_N^*$ (solid line). (b) s_M^* (solid line) and s_N^* (dashed line).

market dynamics later. Therefore, we focus our analysis on the equilibrium where the ad network strategically sets its detection technology for both publisher types within the interior region (i.e., $0 < \{s_j, y_j\} < 1$). In addition, the ad network's policy toward the whitelisted group is unaffected by the fraud detection technology or prevailing market conditions as these publishers are perfectly identified, and their ad traffic is not subject to classification. As a result, their inclusion does not interact with the technological or economic tools under consideration. Therefore, in the subsequent analysis, we focus on the ad network's strategic deployment of its technological tool (i.e., the configuration of fraud detection technology) and economic tool (i.e., payment structure) to publishers in the unknown group, and we examine the implications of these decisions.

4.2. Impact of Market Changes on the Ad Network's Equilibrium Decisions

Given the highly dynamic nature of the online advertising market, a critical question is how the equilibrium decisions are influenced by parametric perturbations, particularly given the interplay between technological and economic tools. In this section, we present an analysis of this question to derive insights into how the ad network's equilibrium decisions are impacted by the changes in two critical market parameters: (1) the *raw quality* of the ad fraud detection technology (r), which is a crucial instrument in the hands of the ad network, and (2) the ad fraud generation efficiency (c), which is the driving force of fraud incentives by the malicious publishers.

4.2.1. How Does Improved Fraud Detection Technology Affect Payment to Publishers?

The ad network's payment is central to publishers' incentive to participate and produce ad fraud. It can be influenced by key market conditions, such as the quality of the ad fraud detection technology (Jacob 2023). This is particularly crucial because ad fraud detection technology has improved significantly driven by advancements in AI and ML (Sisodia and Sisodia 2023). Considering the potential impacts of improved ad fraud detection technology on malicious and nonmalicious publishers, the ad network must determine how to leverage the technology improvement to curb ad fraud (Rayobyte 2025). Because improved fraud detection technology enhances the accuracy of identifying valid traffic, it allows the ad network to compensate publishers more precisely. This has led some to argue that a better detection technology reduces the need for high monetary incentives, thereby lowering payments to publishers (Schiff 2016). However, improved fraud detection technology can also help detect more fraudulent ad traffic, helping to deter malicious publishers' fraud incentives and increase advertisers' valuation of ad traffic. As such, it could help increase the ad network's payment to the publishers (Chen et al. 2015).

Motivated by a lack of clear direction and the profound implications on the ad network's decision, in Proposition 1, we address this critical research question. *How does improved fraud detection technology affect the ad network's payment to publishers?* This question is more intriguing in our setting because the ad network can offer different payments to publishers depending on their classifications as type M (malicious) or type N (nonmalicious). On the one hand, increasing the payment has the risk of overincentivizing fraudulent ad traffic, particularly for type M publishers. On the other hand, decreasing the payment can discourage nonmalicious publishers' participation, particularly for type N publishers. So, should the ad network leverage its economic tool to dampen the incentives for type M publishers and increase the incentives for type N publishers?

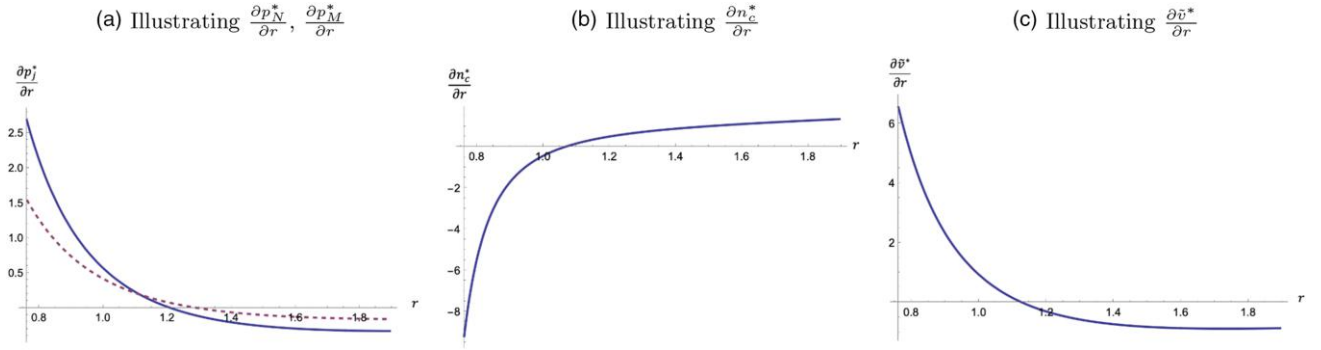
Proposition 1. *When the quality of the fraud detection technology is better (i.e., r is higher), the ad network may sometimes increase its payment to type M publishers (i.e., $\frac{\partial p_M^*}{\partial r} > 0$) and decrease its payment to type N publishers (i.e., $\frac{\partial p_N^*}{\partial r} < 0$). Specifically, $\frac{\partial p_N^*}{\partial r} < 0$ happens when $\hat{G}_1(r, \phi, w, T_N, T_M, \alpha, c, k, \theta) < 0$ and $\frac{\partial p_M^*}{\partial r} > 0$ happens when $\hat{G}_2(r, \phi, w, T_N, T_M, \alpha, c, k, \theta) > 0$, where expressions for $\hat{G}_1(\cdot)$ and $\hat{G}_2(\cdot)$ are given in Online Appendix* EC.4.

Contrary to the findings from the prior studies (Mungamuru 2010) and intuitions, Proposition 1 shows that in response to the improved fraud detection technology (r), the ad network may increase its payment to publishers, even to the type M publishers ($\frac{\partial p_M^*}{\partial r} > 0$). Specifically, this happens when $\hat{G}_1(\cdot) > 0$ and $\hat{G}_2(\cdot) > 0$, which are likely to hold when the technology quality (r) is low as illustrated in Figure 3(a). The mechanism can be explained as follows.

First, we note that improvements in technology quality (i.e., a higher r) always enable the ad network to respond with a more stringent configuration, thereby filtering more fraud traffic for both type M and type N publishers (i.e., $\frac{\partial s_j^*}{\partial r} > 0$) as shown in Figure 2(b).[19] When $\hat{G}_1(\cdot) > 0$ and $\hat{G}_2(\cdot) > 0$, which usually occurs when the technology quality (r) is low, this tightened policy can also filter more valid traffic (i.e., $\frac{\partial y_j^*}{\partial r} > 0$) (see Equation (4)). This leads to an *ad-traffic deflation effect* (that is, less traffic (n_c) is charged as valid (i.e., $\frac{\partial n_c^*}{\partial r} < 0$) (see Figure 3(b))), which could potentially hurt the ad network's revenue.

To counter this effect, the ad network also increases payments to publishers (i.e., $\frac{\partial p_N^*}{\partial r} > 0$ and $\frac{\partial p_M^*}{\partial r} > 0$) to induce the nonmalicious publishers' participation in both groups, thereby helping to increase the valid ad

Figure 3. (Color online) Effect of Improved Technology Quality (r)



Notes. Values of parameters are $\{v = 2, k = 0.2, c = 0.5, \alpha = 0.65, w = 0.05, T_N = 2, T_M = 1, \phi = b = 1, \theta = 0.25, \sigma_W = 0.2\}$. (a) $\frac{\partial p_N^*}{\partial r}$ (solid line) and $\frac{\partial p_M^*}{\partial r}$ (dashed line). (b) $\frac{\partial n_c^*}{\partial r}$ (solid line). (c) $\frac{\partial \tilde{v}^*}{\partial r}$ (solid line).

traffic in the network. This coupled with less fraud ad traffic being mistakenly charged to the advertisers leads to an *ad-value appreciation effect*, where advertisers are willing to pay more for each charged ad traffic (i.e., $\frac{\partial \tilde{v}^*}{\partial r} > 0$) (see Figure 3(c)). Overall, the ad-value appreciation effect outweighs the ad-traffic deflation effect, leading the ad network to raise payments as a strategic response (i.e., $\frac{\partial p_j^*}{\partial r} > 0$). In essence, in this case, the ad network prioritizes improvements in ad-traffic *quality* and *valuation* over simply increasing the *quantity* of ad traffic monetized.

When $\hat{G}_1(\cdot) < 0$ and $\hat{G}_2(\cdot) < 0$, which often occurs when the technology quality (r) is sufficiently high, the ad network would interestingly reverse its strategy to lower the payment (p_j) in response to technology improvement. In this case, a stricter configuration against fraud may lead to improved detection of valid ad traffic (i.e., $\frac{\partial y_j^*}{\partial r} < 0$) for both types of publishers. This has two effects. First, it leads to an *ad-traffic inflation effect*, wherein more ad traffic is charged to advertisers (i.e., $\frac{\partial n_c^*}{\partial r} > 0$) (see Figure 3(b)). This occurs even if the payment is lowered appropriately. This is because when nonmalicious publishers can be more appropriately rewarded for their valid ad traffic, the additional economic incentives (i.e., p_j) for inducing their participation can be lowered. Second, this also leads to an *ad-value depreciation effect*, wherein advertisers lower their valuation for each charged ad traffic as the ad network charges more traffic (i.e., $\frac{\partial \tilde{v}^*}{\partial r} < 0$) (see Figure 3(c)). Overall, the ad-traffic inflation effect and the lower payment to publishers dominate the ad-value depreciation effect, leading the ad network to benefit from reducing the payment to dampen the fraud incentives (i.e., $\frac{\partial p_j^*}{\partial r} < 0$).

Overall, Proposition 1 highlights how the ad network strategically coordinates its economic and technological tools to discipline the players' actions and maximize self-interest. As detection quality improves, the ad network always tightens its configuration to filter more fraud traffic and reduce fraud incentives. However, payments may increase to attract nonmalicious publishers and may fall to deter malicious publishers' fraud incentives depending on market conditions. Moreover, our result that the ad network may adopt a more lenient configuration under a lower detection quality (because $\frac{\partial s_j^*}{\partial r} > 0$) aligns with criticisms from both academia and industry regarding ad networks' insufficient efforts to curb fraud (Wiatr et al. 2019). Importantly, our result reveals that such leniency may reflect an underlying technical deficiency rather than a lack of economic incentives (Cai et al. 2020).

From a broader perspective, Proposition 1 links the ad network's regulatory decision problem to a larger body of research on government regulation in markets characterized by graft and corruption. Bennett et al. (1994) considers mechanisms that reward private healthcare providers for adhering to quality standards while implementing punitive mechanisms for violators. Similar insights have been echoed in later works (Bértoa et al. 2014). However, extant research has criticized financial incentives and punitive market regulation mechanisms as discussed in Section 2.2. Mendoza and Wielhouwer (2015) shows that using financial incentives to elicit desired behavior does not effectively regulate corruption or fraud in financial markets. In some cases, this may incentivize more fraud. Still, none of these works have focused on monetary incentives (similar to the ad network's payment arrangements) to manage the online advertising industry. Thus, Proposition 1 adds to this literature by proposing a practical way in which not only accurate detection of ad fraud (s_j, y_j in this case) but financial incentives for market participants play a role in deterring fraudulent traffic generation.

Besides the quality of the fraud detection technology, the efficiency of producing fraudulent ad traffic can also influence malicious publishers' fraud incentives.

This is particularly important as the recent technological advancements have also catalyzed fraud generation techniques. Thus, we examine how improvements in fraud generation efficiency would affect the ad network's optimal decisions.

4.2.2. How Does Rising Fraud Generation Efficiency Affect Payment to Publishers? Recent advancements in AI and machine learning have transformed ad fraud from labor-intensive schemes to highly automated, software-driven operations (Kshetri and Voas 2019, Cai et al. 2020). Emerging evidence finds that bot-driven fraud techniques have become increasingly efficient and cost effective (Almeida and Gondim 2018). In response, some ad networks have implemented economic countermeasures to reduce fraud incentives. A notable example is Google's smart pricing, which accounts for the impact of ad fraud by not charging the total cost for some clicks, thereby lowering publishers' earnings per unit of traffic (Calvert 2012). Although such economic tools play a critical role in deterring fraud, the question of how they should be coordinated with technological tools remains underexplored. Proposition 2 addresses this gap by examining the following question. *How does the rising fraud generation efficiency affect the ad network's payment to publishers?*

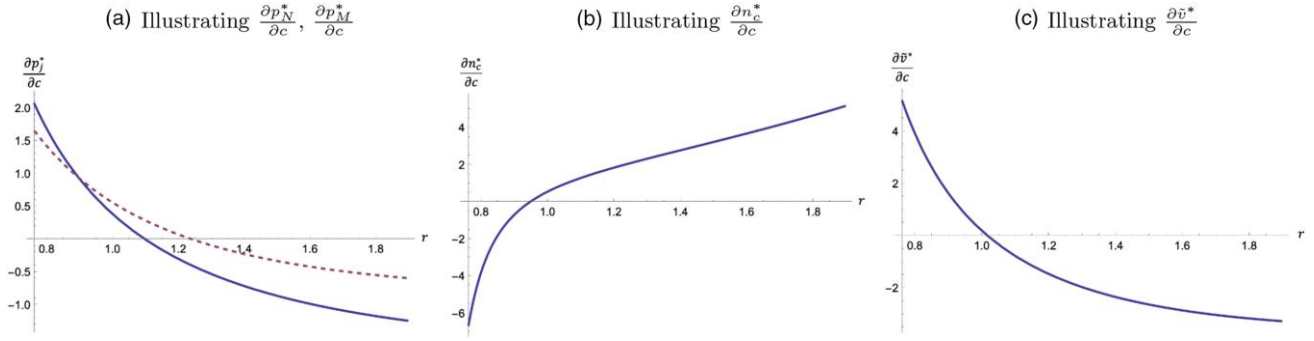
Proposition 2. *When the publishers' fraud generation efficiency is higher (i.e., c is smaller), the ad network may sometimes increase its payment to both type M and type N publishers (i.e., $\frac{\partial p_M^*}{\partial c} < 0$ and $\frac{\partial p_N^*}{\partial c} < 0$). Specifically, $\frac{\partial p_M^*}{\partial c} < 0$ happens when $\hat{G}_3(r,\phi,w,T_N,T_M,\alpha,c,k,\theta) < 0$ and $\frac{\partial p_N^*}{\partial c} < 0$ when $\hat{G}_4(r,\phi,w,T_N,T_M,\alpha,c,k,\theta) < 0$, where expressions for $\hat{G}_3(\cdot)$ and $\hat{G}_4(\cdot)$ are given in Online Appendix* EC.4.3.

Because improved fraud generation efficiency (i.e., a lower value of c) can increase fraud activity, one may expect that the ad network should counter this by reducing publisher payments to dampen malicious incentives (Stone-Gross et al. 2011). Interestingly, Proposition 2 shows that this intuition may not always hold. Specifically, when $\hat{G}_3(\cdot) < 0$ and $\hat{G}_4(\cdot) < 0$, which are likely to hold when technology quality (r) is sufficiently high, the ad network may respond to increased fraud generation efficiency (a lower value of c) with a higher payment to publishers, even for the type M publishers (i.e., $\frac{\partial p_N^*}{\partial c} < 0$ and $\frac{\partial p_M^*}{\partial c} < 0$), as illustrated in Figure 4(a). The mechanism can be explained as follows.

We note that in addition to a reduction in fraud generation cost, improved fraud generation efficiency also lowers the *effective quality* ($\tilde{r}$) of the ad network's fraud detection technology (i.e., $\tilde{r} = \phi c r$ is smaller for a lower c), rendering the ad fraud detection less effective. This is similar to the effect of a decrease in the *raw quality* of the ad fraud detection technology (r is lower). Thus, this leads to a similar mechanism as discussed in Proposition 1. Specifically, improved fraud generation efficiency (i.e., a lower c) always enables the ad network to respond with a more stringent configuration to effectively filter more fraud traffic for both type M and type N publishers (i.e., $\frac{\partial s_j^*}{\partial c} < 0$).

When $\hat{G}_3(\cdot) < 0$ and $\hat{G}_4(\cdot) < 0$, which often occurs when the technology quality (r) is high, this tightened policy leads to an *ad-traffic deflation effect*, whereby less traffic (n_c) is charged (i.e., $\frac{\partial n_c^*}{\partial c} > 0$) (see Figure 4(b)), potentially reducing the ad network's revenue. To counter this, the ad network increases payments to both type N and type M publishers (i.e., $\frac{\partial p_N^*}{\partial c} < 0$ and $\frac{\partial p_M^*}{\partial c} < 0$) to induce nonmalicious publishers' participation in both groups. This, in turn, leads to an *ad-value appreciation effect*, where advertisers are willing to pay more for each charged ad traffic (i.e., $\frac{\partial \tilde{v}^*}{\partial c} < 0$) (see Figure 4(c)). Overall, the ad-value appreciation effect outweighs the ad-traffic deflation effect, resulting in higher payments to publishers as the fraud generation efficiency increases (i.e., as c decreases).

Figure 4. (Color online) Effect of Decreased Fraud Generation Efficiency (c)



Notes. Values of parameters are $\{v = 2, k = 0.2, c = 0.5, \alpha = 0.65, w = 0.05, T_N = 2, T_M = 1, \phi = b = 1, \theta = 0.2, \sigma_W = 0.3\}$. (a) $\frac{\partial p_N^*}{\partial c}$ (solid line) and $\frac{\partial p_M^*}{\partial c}$ (dashed line). (b) $\frac{\partial n_c^*}{\partial c}$ (solid line). (c) $\frac{\partial \tilde{v}^*}{\partial c}$ (solid line).

When $\hat{G}_3(\cdot) > 0$ and $\hat{G}_4(\cdot) > 0$, which usually occur when technology quality (r) is not high, the ad network would interestingly reverse its strategy by lowering payments (p_j) in response to improved fraud generation efficiency (a lower c). In this case, a stricter configuration against fraud leads to an *ad-traffic inflation effect*, whereby more ad traffic is charged to advertisers (i.e., $\frac{\partial n_c^*}{\partial c} < 0$) (see Figure 4(b)). As a result, advertisers lower their valuation for each charged ad traffic, resulting in an *ad-value depreciation effect* (i.e., $\frac{\partial \tilde{v}^*}{\partial c} > 0$) (see Figure 4(c)). This leads the ad network to lower the payments to publishers accordingly as the fraud generation efficiency increases.

These findings suggest that relying solely on either economic or technological interventions is unlikely to be sufficient to address the challenges posed by rising fraud generation efficiency. This could explain why ad fraud rates remain persistently high despite ad networks' adoption of selective economic countermeasures (Kshetri and Voas 2019). Instead, given ongoing investment in ad fraud detection technology that has built resilience against ad fraud (Sisodia and Sisodia 2023), ad networks, such as Google AdSense, should consider jointly deploying both tools to tackle the challenges from the rising fraud generation efficiency. A high-quality detection system enables the ad network to adopt stricter configurations to curb fraudulent activity, whereas increased payments can incentivize participation by nonmalicious publishers. Because improved detection allows for more precise filtering, higher payments are unlikely to overincentivize malicious publishers' fraud activities.

This perspective also helps explain observed strategic behavior by ad networks. On one front, networks have adopted stricter measures to combat ad fraud, particularly in response to AI-generated threats, such as impersonation ads. For example, Google's 2024 Ads Safety Report announced over 30 new policy updates to strengthen fraud enforcement (Adegbola 2025). On the other front, publisher payments have increased, partly because of higher ad valuations from advertisers (Veuno 2024). These developments highlight the need to coordinate economic incentives with technological safeguards to sustain a trustworthy advertising ecosystem.

The use of reward and punishment mechanisms is well established in the economics and policy literature. For example, Andreoni et al. (2003) show that although rewards alone often fail to curb selfish behavior, punishments can foster cooperation by reducing incentives for selfish actions. Proposition 2 connects these findings and builds on how rewards and punishments may not always have to be used as explicit monetary tools as discussed in Section 2.2. Specifically, in contrast to past literature where reward and punishment were measured in explicit monetary terms, Proposition 2 presents a situation where the evolving technology environment within which ad networks, publishers, and advertisers operate can promote reward and punishment.

Although the analysis thus far has focused on the ad network's responses to changing market conditions, an ad network can also strategically influence market outcomes. For example, an ad network may invest in improving fraud detection technology to influence publisher behavior or implement regulatory policies aimed at reducing fraud traffic and improving market outcomes. This raises two important questions. (1) Does the ad network benefit from such continued investment in fraud detection, particularly in a dynamic market? (2) Do such regulatory interventions effectively improve outcomes? In the next subsection, we proceed to analyze these questions in detail.

4.3. The Ad Network's Technology Investment Decision and Regulatory Policy Analysis

Various perspectives have been espoused by both the academia and industry practitioners on what an ad network could do to regulate the market and improve market outcomes through measures such as technological improvements (Zhu et al. 2017b) or implementation of effective penalty schemes (Zorz 2020). For example, some studies suggest that technology improvements could drive fraud numbers down and hence, benefit both the advertisers and the ad network (Klym and Clark 2019), whereas others suggest that reducing bad "actors" (publishers) from the networks can go a long way in improving returns from online advertising (Nandini 2019). Building upon earlier results, we derive additional managerial insights into these discussions by exploring how changes in key market conditions affect the ad network's profit in this section. We begin by analyzing the return of the ad network's investment in its fraud detection technology.

4.3.1. Should an Ad Network Continue Investing in Fraud Detection Technology? The high volume of ad fraud is often attributed to the lack of sufficient incentives for ad networks to fight ad fraud. Dritsoula and Musacchio (2014) and Cai et al. (2020) show how ad networks should fight ad fraud in theory but are seldom interested in investing effort to combat fraudulent traffic in practice. Similarly, Chen et al. (2015) show that an ad network may sometimes have insufficient incentives to improve its fraud detection technologies. However, technological improvements are often recommended as a strategic weapon for ad networks to counter ad fraud (Zhu et al. 2017b, Klym and Clark 2019). As such, an intriguing question is as follows. *Does investing in improving fraud detection technology always benefit an ad network, particularly under evolving*

fraud generation efficiency? To shed light on this debate, we analyze how the ad network's incentive to fight ad fraud is influenced by its technology improvement.

Our analysis aims to provide general insights into whether technology improvement is likely to benefit an ad network with a related cost instead of characterizing the optimal improvement level. Therefore, we consider the cost of technology improvement to be convex in the quality improvement (Chen et al. 2021) instead of specifying a functional form. We focus on the benefits side of technological investment, and we examine the impact of technology investment on the ad network's profit from two perspectives. First, we adopt a static view to assess how the ad network's profit is affected by technology improvements. Second, we take a dynamic perspective accounting for market dynamics (specifically, the evolving nature of the fraud generation technologies (Zhu et al. 2017b, Almeida and Gondim 2018)) to examine how the increased fraud generation efficiency moderates the ad network's incentive in technology improvement. The result is presented below.

Proposition 3. *When the quality of the ad network's fraud detection technology (i.e., r) is higher, interestingly, the return from the investment can be amplified by the improved fraud generation efficiency (i.e., a lower value of c; i.e., $\frac{\partial^2\pi^*}{\partial r\partial c}<0$), even though such return can sometimes be diminishing when considering only the technology improvement (i.e., $\frac{\partial^2\pi^*}{\partial r^2}<0$). Specifically, $\frac{\partial^2\pi^*}{\partial r^2}<0$ and $\frac{\partial^2\pi^*}{\partial r\partial c}<0$ hold when $\hat{G}_5(r,w,T_M,T_N,\alpha,c,k,\phi,\theta)<0$ and $\hat{G}_6(r,w,T_M,T_N,\alpha,c,k,\phi,\theta)<0$, where the expressions for $\hat{G}_5$ and $\hat{G}_6$ are given in Online Appendix* EC.4.4.

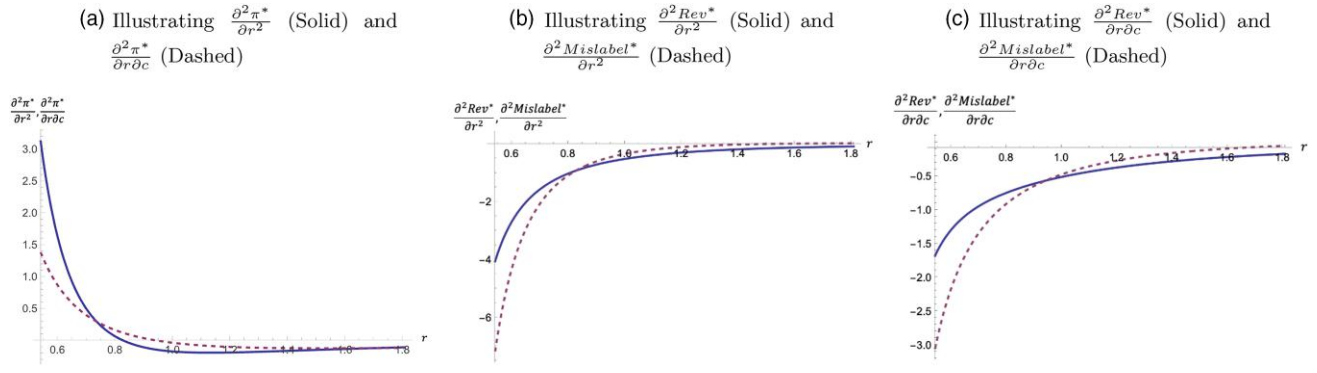
Investment in improving fraud detection technology has become an increasingly critical decision for ad networks. Although ad networks have reacted to the growing prevalence of ad fraud by employing advanced AI and ML techniques to improve fraud detection (Choi et al. 2020), industry evidence suggests that such improvements may not always be beneficial (Fou 2020a) and may even lead to losses for ad networks (Zhu et al. 2017b). Moreover, the improved fraud generation efficiency also raises concerns about the long-term sustainability of continued investment in what has often been characterized as a "cat-and-mouse" game with fraudsters.

Proposition 3 provides important insights into these strategic considerations. Specifically, we find that although ad networks always benefit from improvements in fraud detection technology, the marginal return to such investment in technology can diminish (i.e., $\frac{\partial^2\pi^*}{\partial r^2}<0$) when $\hat{G}_5(\cdot)<0$, a condition that is likely to occur when the technology quality (r) is sufficiently high as depicted in Figure 5(a). The mechanism can be explained as follows.

Technically, two critical forces influencing the benefit from technology improvement are its impact on the revenue from ad traffic (i.e., $Rev=(\tilde{v}-p_W)n_{gW}+\sum_{j=M,N}(\tilde{v}-p_j)(n_{gj}(1-y_j)+n_{fj}(1-s_j))+\theta w\sum_{j=M,N}p_j(s_jn_{fj}+y_j\sum_{i=M,N}n_{gij})$) and the associated mislabeling cost (i.e., $Mislabel=k(y_Mn_{gM}p_M+y_Nn_{gN}p_N)$) as defined in Equation (6). When $\hat{G}_5(\cdot)>0$, which is likely to occur when r is small, technology improvement leads to a diminishing return to the gross ad revenue (i.e., $\frac{\partial^2 Rev^*}{\partial r^2}<0$) (see Figure 5(b)). However, this is dominated by a larger marginal increase in the savings from mislabeling cost (i.e., $\frac{\partial^2 Mislabel^*}{\partial r^2}<\frac{\partial^2 Rev^*}{\partial r^2}<0$ as depicted in Figure 5(b)), leading to an overall increasing marginal return on profit (i.e., $\frac{\partial^2\pi^*}{\partial r^2}>0$) (see Figure 5(a)).

When $\hat{G}_5(\cdot)<0$, which usually occurs when the technology quality (r) is sufficiently high, the marginal benefit from reductions in mislabeling cost declines substantially and is outweighed by the diminishing return to gross ad revenue (i.e., $\frac{\partial^2 Rev^*}{\partial r^2}<\frac{\partial^2 Mislabel^*}{\partial r^2}<0$) (see Figure 5(b)), leading to a diminishing marginal return on profit (i.e., $\frac{\partial^2\pi^*}{\partial r^2}<0$) (see Figure 5(a)). As such, although the ad network can leverage the improved

Figure 5. (Color online) Illustration of Proposition 3



Notes. Values of parameters are $\{v=2,k=0.1,c=0.5,\alpha=0.65,w=0.05,T_N=2,T_M=1,\phi=b=1,\theta=0.2,\sigma_W=0.3\}$. (a) $\frac{\partial^2\pi^*}{\partial r^2}$ (solid line) and $\frac{\partial^2\pi^*}{\partial r\partial c}$ (dashed line). (b) $\frac{\partial^2 Rev^*}{\partial r^2}$ (solid line) and $\frac{\partial^2 Mislabel^*}{\partial r^2}$ (dashed line). (c) $\frac{\partial^2 Rev^*}{\partial r\partial c}$ (solid line) and $\frac{\partial^2 Mislabel^*}{\partial r\partial c}$ (dashed line).

technology to curb ad fraud and improve its ad revenue, the marginal benefits from such investments can diminish as the technology quality surpasses a certain threshold.

Although the marginal return to improved fraud detection can diminish, we find that, interestingly, it can be augmented by the evolving nature of the online ad market that is characterized by falling costs and increased ad fraud generation efficiency (Jackson 2025) (i.e., $\frac{\partial^2 \pi^*}{\partial r \partial c} < 0$ when $\hat{G}_6(\cdot) < 0$), which usually occurs when the technology quality (r) is high (see Figure 5(a)). This is because in these parameter regions, the moderating effect of improved fraud generation efficiency (i.e., a lower c) on the gross ad revenue is sufficiently large that it dominates the moderating effect from the mislabeling cost (i.e., $\frac{\partial^2 Rev^*}{\partial r \partial c} < \frac{\partial^2 Mislabel^*}{\partial r \partial c} < 0$) as shown in Figure 5(c), which amplifies the marginal benefit from technology improvement (i.e., $\frac{\partial^2 \pi^*}{\partial r \partial c} < 0$) (see Figure 5(a)).

The result in Proposition 3 suggests that when considering technology quality in isolation, there may exist a "tipping point" beyond which further investments in fraud detection will unlikely increase the ad network's profit. Nevertheless, when adopting a dynamic view that incorporates market evolution, particularly the rising fraud generation efficiency, we show that this insight may no longer hold in practice given the continual advancement of fraud techniques (Kahn 2024). Interestingly, Proposition 3 shows an intriguing moderating effect; as malicious publishers become more efficient in generating fraud, the benefits of improving detection technology are amplified. This indicates that this theory of the "tipping point" for technology investment may not always hold. Therefore, in the presence of rapidly advancing fraud generation capabilities, continued investment in detection technology may remain strategically beneficial for the ad network.

This result has important implications. In particular, it underscores the necessity of sustained innovation in fraud detection systems given the adversarial nature of ad fraud, a dynamic in which fraudsters continuously develop new tactics to evade detection. This perspective helps explain why firms, such as Google, have made substantial, ongoing investments in fraud detection technology. For example, Google maintains a dedicated Ad Traffic Quality team focused on identifying emerging threats and advancing detection technologies (Google Ads 2025). In 2024, the company leveraged recent advancements in generative AI by integrating large language models into its detection infrastructure, introducing more than 50 enhancements aimed at improving the speed and precision of fraud identification at scale (Adegbola 2025).

The above findings highlight the limitations of previous research on online advertising, which has predominantly adopted a static lens and overlooked the critical influence of market dynamics on strategic behavior (e.g., Dritsoula and Musacchio 2014, Chen et al. 2015) as discussed in Section 2.1. Thus, ad networks that remain reluctant to invest in detection technology should re-evaluate their strategies with greater attention to the evolving nature of the market. The above analysis reveals the importance of market dynamics for the ad network's strategic deliberations. In the next section, we turn to a widely used regulatory practice to assess its effectiveness in deterring ad fraud and its broader implications for the ad network's performance.

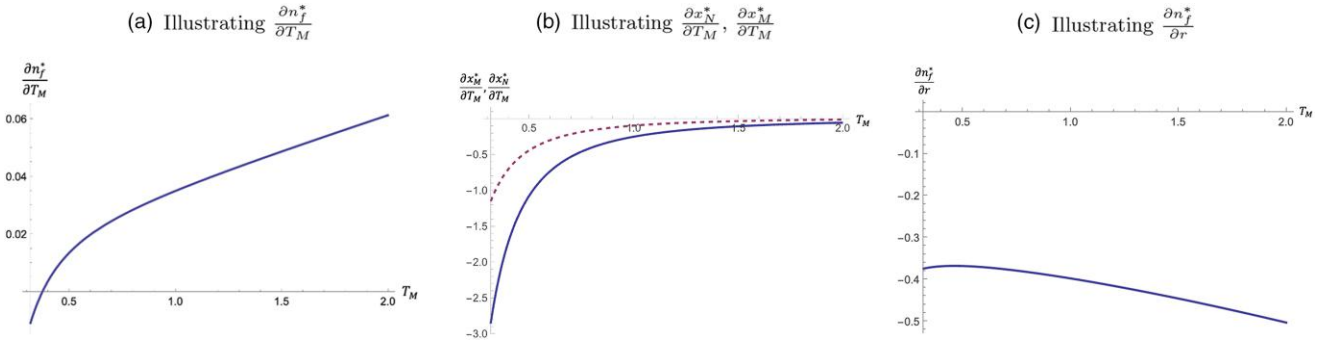
4.3.2. Is Reducing Malicious Publishers Effective Against Ad Fraud and Profitable?

Past research suggests that ad networks can improve returns and benefit from having fewer malicious publishers in the network (Nandini 2019). Several mechanisms can help reduce malicious publishers in the ad network. For example, legal and regulatory interventions may target the sources of ad fraud directly (Kahn 2023), whereas broader societal awareness campaigns can help diminish tolerance for fraudulent traffic. Similarly, ad networks can implement measures against fraudulent publishers and developers by excluding them from the network or pursuing costly legal action (Cai et al. 2020). For instance, Google often investigates AdSense publishers with unusual ad-traffic patterns (Biswas and Roy 2021) and may suspend or even terminate these publishers' contracts based on its internal analysis (D'Annunzio and Russo 2020). However, an important question remains. *Does a reduction in the number of malicious publishers help eliminate or reduce the amount of fraudulent ad traffic?* We analyze this question and report the results below.

Proposition 4. *When the size of malicious publishers* (T_M) *decreases, the amount of fraudulent ad traffic in the ad network* (n_f) *may sometimes actually be higher* (*i.e.*, $\frac{\partial n_f^*}{\partial T_M} < 0$) *when* $\hat{G}_7(\alpha, c, r, k, T_N, T_M, w, \phi, \theta) < 0$, *where the expression for* $\hat{G}_7(\cdot)$ *is given in Online Appendix* EC.4.5.

Past research has found that reducing the number of publishers with malicious intent from the ad network can effectively reduce fraudulent ad traffic. For example, Faou et al. (2016) proposes removing affiliate websites of some primary malicious publishers to suppress fraudulent activities. Similar ideas have been presented to counter other types of fraud. For example, Braun and Eklund (2019) consider *blacklisting* publishers responsible for misinformation campaigns to curb fake news. Interestingly, Proposition 4 shows

Figure 6. (Color online) Illustration of Proposition 4



Notes. Values of parameters are $\{v = 2, k = 0.15, c = 0.5, \alpha = 0.6, w = 0.05, T_N = 2, \phi = b = r = 1, \theta = 0.2, \sigma_W = 0.3\}$. (a) $\frac{\partial n_f^*}{\partial T_M}$ (solid line). (b) $\frac{\partial x_N^*}{\partial T_M}$ (solid line) and $\frac{\partial x_M^*}{\partial T_M}$ (dashed line). (c) $\frac{\partial n_f^*}{\partial r}$ (solid line).

that this approach may not always be effective at reducing fraudulent ad traffic in the network (see Figure 6(a)). The mechanism is as follows.

Two effects can occur on the total amount of fraud ad traffic (n_f) when malicious publishers are reduced in an ad network (as T_M decreases). First, the size of publishers with fraud motives decreases. This is a *direct effect* that lowers fraud activities in the network. Second, as T_M decreases and fewer publishers may commit ad fraud, the ad network also adjusts its payment and technology configuration. This is an *indirect effect* that could unintentionally heighten the fraud incentives for the remaining malicious publishers (i.e., $\frac{\partial x_f^*}{\partial T_M} < 0$) as illustrated in Figure 6(b). When $\hat{G}_7(\cdot) < 0$, which is likely to occur when the size of malicious publishers (T_M) is not very large, the indirect effect of increased fraud activities by the remaining malicious publishers dominates the direct effect, leading to an overall increase in fraudulent ad traffic (i.e., $\frac{\partial n_f^*}{\partial T_M} < 0$) as depicted in Figure 6(a).

From a managerial perspective, Proposition 4 shows that reducing malicious publishers may not always be an effective strategy for improving ad network quality. Instead, a more robust approach involves enhancing the fraud detection technology (r). As Figure 6(c) shows, improvements in fraud detection quality can help reduce fraud traffic effectively, even in the absence of a reduction in the number of malicious publishers (i.e., $\frac{\partial n_f^*}{\partial r} < 0$). This is because improved technology quality results in better detection and revenue share for publishers. As a result, the incentives to engage in fraudulent activity are dampened, thereby inducing publishers to reduce fraud traffic generation. In this context, technological advancement serves as a more effective lever to reduce fraud traffic generation.

We show that reducing malicious publishers does not always help lower ad fraud. But, is this profitable to the ad network because many suggest that it can benefit from reducing malicious actors from the network (see, e.g., Nandini 2019)? We examine this question and summarize the result below.

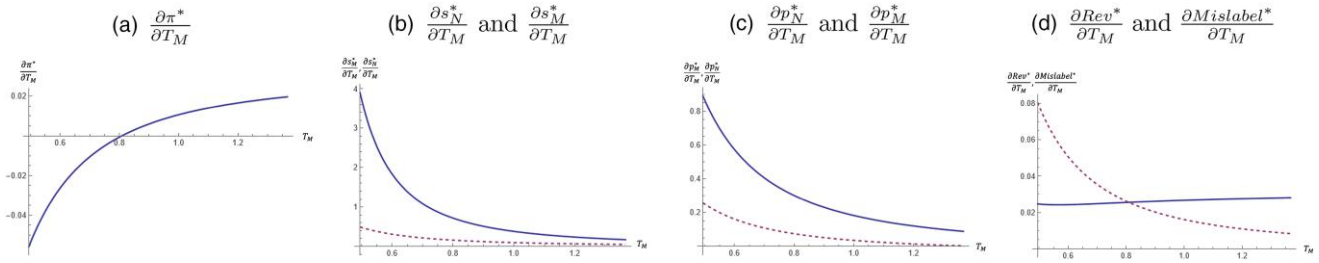
Proposition 5. *When the size of malicious publishers (T_M) is smaller, the ad network's profit may actually sometimes decrease. Specifically, $\frac{\partial \pi^*}{\partial T_M} > 0$ when $\hat{G}_8(\alpha, c, r, k, T_N, T_M, w, \phi, \theta) > 0$, where the expression for $\hat{G}_8(\cdot)$ is given in Online Appendix EC.4.6.*

Because reducing malicious publishers helps improve the quality of participants in an ad network, which in turn, could promote ad-traffic quality and encourage nonmalicious publishers' participation, one may expect this to benefit the ad network too (Klym and Clark 2019, Zorz 2020). Interestingly, Proposition 5 shows that this intuition may not always hold, and the ad network can sometimes be worse off by reducing the malicious publishers in the ad network ($\frac{\partial \pi^*}{\partial T_M} > 0$) under certain conditions as illustrated in Figure 7(a). The reasoning is as follows.

Reducing the size of malicious publishers has two critical implications. First, it results in a *direct* effect of decreased valid ad traffic because of the reduced malicious publishers. Second, it can influence the ad network's payment and technology configuration decisions, which in turn, *indirectly* influence the participation and fraud incentives from the remaining publishers, thereby affecting the ad network's profit. Specifically, a decrease in malicious publisher population (a lower T_M) can prompt the ad network to respond with both a loosening configuration (i.e., $\frac{\partial s_j^*}{\partial T_M} > 0$) (see Figure 7(b)) and a lower payment (i.e., $\frac{\partial p_j^*}{\partial T_M} > 0$) (see Figure 7(c)). These two effects together can result in lower ad revenue (i.e., $\frac{\partial Rev^*}{\partial T_M} > 0$) and reduced mislabeling cost (i.e., $\frac{\partial Mislabel^*}{\partial T_M} > 0$) as depicted in Figure 7(d), where *Rev* and *Mislabel* are defined earlier in Section 4.3.1.

When $\hat{G}_8(\cdot) > 0$, which usually occurs when the size of malicious publishers (T_M) is large, a reduction in

Figure 7. (Color online) Illustration of Proposition 5



Notes. Values of parameters are $\{v = 1, k = 0.15, r = 1, c = 0.5, \alpha = 0.62, w = 0.2, T_N = 2, \phi = 1, b = 0.05, \theta = 0.2, \sigma_W = 0.3\}$. (a) $\frac{\partial \pi^*}{\partial T_M}$ (solid line). (b) $\frac{\partial s_N^*}{\partial T_M}$ (solid line) and $\frac{\partial s_M^*}{\partial T_M}$ (dashed line). (c) $\frac{\partial p_N^*}{\partial T_M}$ (solid line) and $\frac{\partial p_M^*}{\partial T_M}$ (dashed line). (d) $\frac{\partial Rev^*}{\partial T_M}$ (solid line) and $\frac{\partial Mislabel^*}{\partial T_M}$ (dashed line).

malicious publishers (i.e., a lower T_M) can result in a greater loss in the gross ad revenue than the corresponding savings in the mislabeling cost (i.e., $\frac{\partial Rev^*}{\partial T_M} > \frac{\partial Mislabel^*}{\partial T_M}$) (see Figure 7(d)), effectively decreasing the overall profits for the ad network (i.e., $\frac{\partial \pi^*}{\partial T_M} > 0$) (see Figure 7(a)). In contrast, when $\hat{G}_8(\cdot) < 0$, a reduction in malicious publishers can generate substantial savings in mislabeling costs that outweigh the decline in gross ad revenue (i.e., $\frac{\partial Mislabel^*}{\partial T_M} > \frac{\partial Rev^*}{\partial T_M}$) (see Figure 7(d)), thereby effectively benefiting the ad network (i.e., $\frac{\partial \pi^*}{\partial T_M} < 0$) as illustrated in Figure 7(a).

Propositions 4 and Proposition 5 yield novel insights. First, Proposition 4 shows that reducing malicious publishers may inadvertently induce exceedingly more fraudulent traffic in the network. Although past research has noted the risk of erroneously removing nonmalicious publishers through such strategies (Asdemir et al. 2008), our analysis reveals a more fundamental concern; even when only malicious publishers are accurately excluded, the remaining malicious actors may respond by intensifying their fraudulent behavior, thereby exacerbating the very problem that the intervention aims to solve. Second, Proposition 5 shows that reducing malicious publishers can, under certain conditions, negatively affect the ad network's profitability. As such, it is sometimes in the interest of ad networks to allow some fraud, which echoes the broad literature on ad fraud (e.g., Wilbur and Zhu 2009, Chen et al. 2015) as discussed in Section 2.1.

The results from Proposition 4 and Proposition 5 have important policy implications. Specifically, they show that ad networks should move beyond purely punitive or exclusionary strategies and instead, adopt a more nuanced, efficiency-oriented approach. From a normative standpoint, this involves prioritizing the management and mitigation of ad fraud rather than its complete eradication. This helps explain why the overall rate of fraudulent ad traffic has remained persistently high (Fraud Blocker 2025) and why the economic losses attributable to ad fraud have continued to rise over time (Dogtiev 2025), despite widespread efforts to eliminate fraudulent actors. These patterns underscore the limitations of eradication-focused policies and highlight the need for more adaptive and strategically balanced interventions. This implication also aligns with well-established trade-offs in other adversarial domains, such as cybersecurity, where the complete elimination of risk is often economically inefficient and technically infeasible.

5. Extensions

In this section, we consider several scenarios extending the main model in Section 3 to demonstrate that the core insights are robust in these extensions. First, we consider participating advertisers' imperfect valuation updates. This extends advertisers' perfect estimation of valid ad traffic in the main model to represent reality. Second, the main model considers that malicious publishers also produce some valid ad traffic in addition to generating ad fraud. We show that our main takeaways remain qualitatively similar, even if malicious publishers only generate fraudulent ads. Last, we consider an alternative specification of the ROC curve, which also depends on the amount of fraudulent ad traffic. We find that our core insights remain qualitatively similar.

5.1. Imperfect Valuation Update

Our main model considers that advertisers can estimate the amount of valid traffic (n_g) and use it to update their ad-traffic valuation accordingly. Although this appropriately reflects the case where many advertisers can resort to third-party or in-house services to audit their ad performances and accurately estimate n_g (Cai et al. 2020, Prokopets 2021), there could also be situations wherein their estimates can deviate from the actual performance (Wiatr et al. 2019, Gordon et al. 2021). Thus, we extend this setting by considering this valuation update by advertisers to be imperfect.

For this purpose, we introduce a parameter $\gamma > 0$ and let $n_g' = \gamma n_g$ and $\tilde{v}' = v\left(\frac{n_g'}{n_c}\right) = \gamma \tilde{v}$ denote the advertisers' estimate of valid ad traffic and the corresponding ad

valuation update, respectively. In effect, γ represents to what extent the advertisers' estimate of true valid traffic n_g is accurate. Intuitively, when $\gamma = 1$, advertisers perfectly estimate n_g and update the valuation of the charged ad traffic accordingly, a case that is considered in our main model in Section 3. However, when $\gamma \neq 1$, misestimation leads to an imperfect valuation update (that is, $\tilde{v}' \neq \tilde{v}$), and there are two cases.

First, when $\gamma < 1$, the advertisers *underestimate* the amount of valid traffic (n_g) and the true valuation of each charged ad traffic (i.e., $\tilde{v}' < \tilde{v}$). Second, when $\gamma > 1$, advertisers *overestimate* the amount of valid traffic and the true valuation of charged ad traffic (i.e., $\tilde{v}' > \tilde{v}$). We analyze the effect of advertisers' misestimation of ad-traffic valuation on the equilibrium decisions and report the result below. As discussed earlier, we focus on the equilibrium where the ad network strategically sets its detection technology for both publisher types within the interior region (the same as case 1 in Lemma 1) for this and the subsequent extensions. Additionally, we impose the condition in Online Appendix EC.5.1.1 to eliminate the trivial case where none or all publishers participate in the ad network. Lemma 2 provides the characterization of the interior case equilibrium in this scenario, with full proof with expressions in Online Appendix EC.5.1.1.

Lemma 2. *When the advertisers' ad valuation update is imperfect, in equilibrium, the ad network's decisions on technology configuration* ($s^*_{j,I}$) *and payment* ($p^*_{j,I}$) *for type j publishers,* $j \in \{M, N\}$, *are a function of* γ *as shown below*:

$$p^*_{W,I} = \gamma p^*_W,$$
$$p^*_{M,I} = p^*_M + (\gamma - 1)B_1,$$
$$p^*_{N,I} = p^*_N + (\gamma - 1)B_2,$$
$$s^*_{M,I} = B_3(\gamma),$$
$$s^*_{N,I} = B_4(\gamma),$$

where expressions for B_1, B_2, B_3, B_4 *are in Online Appendix EC.5.1.1.*

In practice, the valuation of valid ad traffic is a combination of factors, such as ad location and geographic location of the audience (Choi et al. 2020). Advertisers typically estimate their ad campaign performance and revise their valuation of charged ad traffic using these factors together with the ad campaign report from the ad network (Prokopets 2021). When influenced by misinformation, advertisers are likely to incorrectly estimate the actual valuation of the charged traffic, which may hurt their utilities. Therefore, misreporting of ad campaign performance metrics, if going undetected, can provide additional malicious incentives. Indeed, Gordon et al. (2021) identifies overreporting or misrepresenting of audience metrics to increase ad revenues as one of the basic motivations for ad fraud. Yet, no study formally considers this aspect of misrepresentation in the ad fraud context. Thus, important questions are as follows. How does advertisers' misestimation affect the profits of the ad network and the publishers, and how does it affect whether they can benefit from manipulating advertisers' estimation? We examine these questions and summarize the results below.

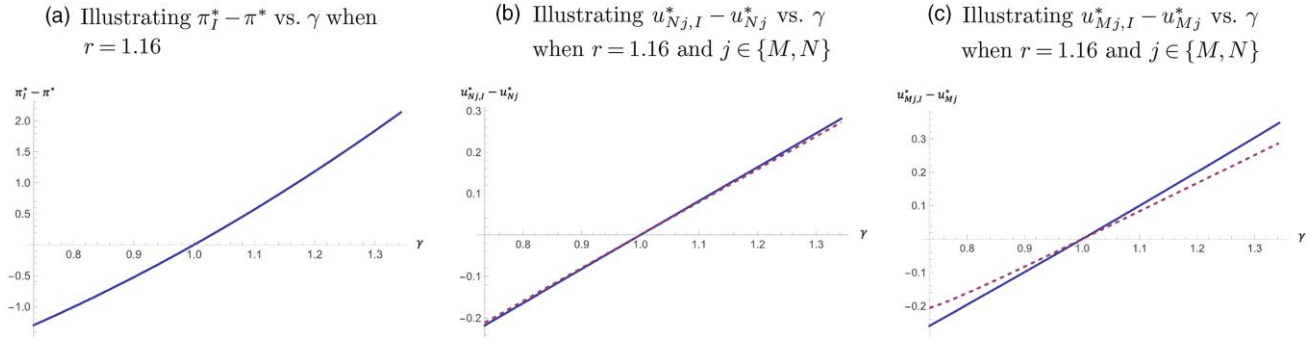
Proposition 6. *When advertisers imperfectly update the valuation of ad traffic, this may, interestingly, sometimes benefit both the ad network and the publishers. Specifically, we have the following.*

1. *The ad network's profit is higher* (*i.e.*, $\pi^*_I > \pi^*$) *when* $\hat{G}_9(\alpha, c, k, r, T_M, T_N, w, \phi, \theta, \gamma) > 0$.
2. *Type j nonmalicious publishers benefit* (*i.e.*, $u^*_{Nj,I} > u^*_{Nj}$) *when* $\hat{G}_{10}(\alpha, c, k, r, T_M, T_N, w, \phi, \theta, \gamma) > 0$ *for* $j = N$ *and* $\hat{G}_{11}(\alpha, c, k, r, T_M, T_N, w, \phi, \gamma) > 0$ *for* $j = M$.
3. *Type j malicious publishers benefit* (*i.e.*, $u^*_{Mj,I} > u^*_{Mj}$) *when* $\hat{G}_{12}(\alpha, c, k, r, T_M, T_N, w, \phi, \theta, \gamma) > 0$ *for* $j = M$ *and* $\hat{G}_{13}(\alpha, c, k, r, T_M, T_N, w, \phi, \theta, \gamma) > 0$ *for* $j = N$.

Expressions for $\hat{G}_9(\cdot), \hat{G}_{10}(\cdot), \ldots, \hat{G}_{13}(\cdot)$ *are given in Online Appendix* EC.5.1.2.

As illustrated in Figure 8(a), the ad network can be better off when advertisers imperfectly learn or calibrate their ad valuation (i.e., $\pi^*_I > \pi^*$), which usually

Figure 8. (Color online) Illustration of Proposition 6



Notes. Values of parameters are $\{v = 2, k = 0.2, r = 1.16, c = 0.4, \alpha = 0.65, w = 0.05, T_N = 2, T_M = 1, \phi = 1, b = 1, \theta = 0.2, \sigma_W = 0.3\}$. (a) $\pi^*_I - \pi^*$ (solid line). (b) $u^*_{NN,I} - u^*_{NN}$ (solid line) and $u^*_{NM,I} - u^*_{NM}$ (dashed line). (c) $u^*_{MN,I} - u^*_{MN}$ (solid line) and $u^*_{MM,I} - u^*_{MM}$ (dashed line).

occurs when advertisers overestimate the true ad valuation (i.e., $\gamma > 1$ and $\tilde{v}' > \tilde{v}$). Similarly, malicious and nonmalicious publishers may also benefit from such overestimation by the advertisers regardless of their classified type j (i.e., $u^*_{ij,I} > u^*_{ij}$) as shown in panels (b) and (c) of Figure 8. Proposition 6 highlights an important result; both the ad network and the publishers are incentivized to manipulate the advertisers' estimation of the true ad valuation. As a recent example, Facebook was accused of knowingly inflating and misleading its "potential reach" metric for years, which is meant to assist advertisers in estimating their ad budget and setting up ad campaigns (Graham 2021). When influenced by such misinformation, advertisers will likely misestimate the true valuation of the ad traffic.

The critical facet of this extension is how advertisers' imperfect valuation update impacts the ad network's payment and profits. The fact that $\tilde{v}'$ is impacted positively by γ indicates that both the ad network and the publishers have an incentive to inflate the true ad valuation. Given such a strong motive, this suggests that advertisers should consider the use of third-party services[20] or building in-house teams to conduct independent and unbiased ad-performance analysis and regularly monitor the performance of ad campaigns.

Overall, our insight from Proposition 6 aligns with past work in the finance literature (Han et al. 2021). Specifically, although analyst earning estimates are considered reliable performance indicators among academics and traders, evidence is emerging that these forecasts are becoming increasingly vulnerable to management *interference*. Sometimes, such interference is done to communicate inside information and reduce analyst overconfidence. In other cases, it is done to mislead analysts and dampen their expectations so that forecasts are beatable upon earnings release. Parallels from this scenario to the ad network scenario are easy to see. The higher the traffic valuation estimation from the actual value, the higher the likelihood that the ad network can extract more from advertisers for ads on publisher websites. This increases ad network profits while allowing publishers to earn more through either improved payment or higher earnings per unit traffic.

5.2. Only Fraud Traffic from Malicious Publishers

The main model considers malicious publishers to generate both fraudulent and valid ad traffic. However, there may be situations where malicious publishers only generate fraudulent ad traffic. For example, some malicious publishers run fake website operations that use bots to generate fraudulent ad traffic and earn ad revenue (Richet 2022). Thus, we analyze the setting where the malicious publishers do not produce valid ad traffic and summarize the results below.

Lemma 3. *When malicious publishers only generate fraudulent ad traffic (i.e., $b = 0$) in equilibrium, the ad network's technology configuration ($s^*_{j,f}$) and payment ($p^*_{j,f}$) decisions for type j publishers, where $j \in \{M, N\}$, can be characterized as follows*:

$$p^*_{W,f} = p^*_W,$$

$$p^*_{M,f} = \frac{A_9(\alpha b T_M(1 + w\theta - k) + (1-\alpha)(w+1)T_N v)}{A_{10}(4(1-\alpha)^2 c^2 T_N^2 (w(1-\theta)+k)^2)},$$

$$p^*_{N,f} = \frac{A_{11}((1-\alpha)wT_M(1-\theta-(\theta+1)\tilde{r}) + 2\alpha cw(1-\theta)T_N(1-\tilde{r})^2)}{A_{12}(\alpha k^2 T_N(1-\tilde{r})^2 - 2(1-\alpha)krT_M\phi(1-\tilde{r}) - 2(1-\alpha)cT_M\tilde{r}^2)},$$

$$s^*_{M,f} = \frac{A_{13}(w(1-\theta)((1-\alpha)T_N v + \alpha b T_M) + 2\alpha b k T_M)}{A_{14}(\alpha b T_M(k+1-\theta(1-k)) + (1-\alpha)(1-\theta)T_N v)},$$

$$s^*_{N,f} = \frac{\begin{array}{l}A_{15}(2\alpha cw(1-\theta)T_N(1-\tilde{r}) + 2\alpha ckT_N(1-\tilde{r}) - 2(1-\alpha)\tilde{r}T_M \\ + (1-\alpha)wT_M(1-2\tilde{r}-\theta))\end{array}}{A_{16}(1+k+\tilde{r}(\theta(3-k)+1-k)-\theta(1-k))}.$$

where $A_9, A_{10}, \ldots, A_{16}$ are provided in Online Appendix EC.5.2.1.

As shown in Lemma 3, this extension presents a special scenario of our main model for $b = 0$; that is, malicious publishers generate no valid ad traffic. We have analyzed the extension and reported the results in Online Appendix EC.5.2 for brevity. We find that although most of our results remain qualitatively similar, the ad network may always benefit from reducing the malicious publisher size (i.e., $\frac{\partial \pi^*}{\partial T_M} < 0$). This happens because advertisers value only valid ad traffic (i.e., $\tilde{v}$ increases in n_g). When malicious publishers do not generate valid ad traffic, a reduction in these publishers no longer lowers the valid ad traffic. As a result, when this adverse effect is no longer present, the ad network can leverage this decreased fraud motive by adjusting its payment and technology configuration appropriately to improve its profit.

It is worth noting that this setting of absolutely no valid ad traffic from malicious publishers (i.e., $b = 0$) presents an extreme case because malicious publishers generally have valid content and can attract genuine ad traffic. However, they weigh the benefit of committing ad fraud against the risk/cost and may generate additional fraudulent ad traffic to inflate overall ad traffic and ad revenue if doing so is profitable (Fou 2020b). In addition, from the technical perspective, detecting publishers with pure fraudulent ad traffic can be relatively straightforward. As such, publishers need to conflate the fraudulent ad traffic with the valid traffic to avoid being easily detected. Nevertheless, it can be shown that the result that reduced malicious publisher size could hurt the ad network's profit (i.e., $\frac{\partial \pi^*}{\partial T_M} > 0$ continues to hold when malicious publishers generate a nonzero amount of valid traffic ($b > 0$)), albeit a very small value (e.g., $b = 0.01$).

5.3. Fraud Ad Traffic Impacts the ROC Curve

The main model considers that the ROC curve for the fraud detection technology is characterized by the effective quality of the detection technology ($\tilde{r}$), which depends on the raw quality (r) and the fraud generation efficiency (c). However, the ROC curve may also depend on the amount of fraud ad traffic (x_j) because it is likely more challenging to detect valid traffic when the fraudulent ad traffic increases. As such, in this extension, we consider the case where the ROC curve is also a function of $x_j, j \in \{N, M\}$. Specifically, we revise the ROC curve specification to the following equation:

$$y_j = (1 - \tilde{r})x_j s_j,$$

where x_j is the amount of fraudulent ad traffic generated by type j publishers. Intuitively, as the fraud traffic (x_j) increases, the false-positive rate (y_j) also increases for a given technology configuration at s_j. As such, an increase in x_j decreases the true-negative rate ($1 - \text{FRP}$), indicating that detecting valid ad traffic becomes more challenging. Because of the analytical intractability, we examine this new setting using numerical analysis. We show that an interior equilibrium exists and that the major insights from the main model remain qualitatively similar. The detailed results and discussions are reported in Online Appendix EC.5.3 for brevity.

6. Discussion and Conclusion

Ad fraud is a fundamental problem for the online advertising industry and a rising concern for all stakeholders, including advertisers, publishers, and ad networks. Fraudulent ad traffic, when undetected, can harm the overall quality of ad traffic and ultimately, undermine the trust and faith of advertisers to use the online channel. Past research on ad fraud focuses on either the technological aspect of fraud detection or economic policies of fraud deterrence (Wilbur and Zhu 2009, Chen et al. 2015). Nevertheless, it overlooks the interplay of the two and thus, can run the risk of designing suboptimal policies that lead to inefficient market outcomes.

Our work aims to fill this gap in the literature by developing a game-theoretic model to examine the equilibrium outcome when both economic and technological tools are used to discipline the market. In our framework, an ad network operates as an intermediary platform, collecting ad revenues from advertisers and sharing a portion of these revenues with publishers, which are either malicious or nonmalicious. Although the true types of these publishers are not directly observable, the ad network can imperfectly infer their types and classify them into malicious (type M publishers) or nonmalicious (type N publishers). This enables the ad network to implement differentiated technology configurations for fraud detection and payment policies for each group. These strategic decisions shape publishers' behavior and influence the advertisers' returns from online advertising, ultimately affecting their valuation for ad traffic. By characterizing the ad network's strategic choices, we provide insights into how it balances technological and economic levers to mitigate ad fraud and effectively regulate the market.

6.1. Managerial Insights

In this study, we seek to address several critical questions with substantial managerial implications, particularly on the intricate interplay between the ad network's technological and economic tools and the effectiveness of its policies in mitigating ad fraud. The first question that we examine is as follows. How do improvements to ad fraud detection technology impact the ad network's payment to publishers? In contrast to prior research suggesting that improved fraud detection can enable ad networks to reduce payments to publishers (Mungamuru 2010), our findings indicate that this is not always true. Specifically, we find that as the fraud detection technology improves, whereas the technology configuration should be set more strictly, it may be optimal for the ad network to increase its payments to publishers in some cases. This result has important managerial implications.

First, it reveals how ad networks should strategically leverage enhanced technological capabilities for their ad fraud detection. Technological capability can generally be characterized by the ad network's market share and its cumulative investment in ad fraud detection infrastructure. This result suggests that market leaders are well positioned to adopt and enforce more stringent ad fraud detection policies, effectively capitalizing on their advanced technological capability. For example, Google has deployed a highly advanced and automated detection system comprising "over 200 sophisticated filters" designed to identify and block invalid traffic in near real time (Google Ads 2025). As a result of this effort, Google has restricted ads on 1.3 billion publisher pages using this sophisticated suite of tools to enforce strict ad policies and curb fraudulent activities (Adegbola 2025).

Second, the result underscores the importance of strategically aligning economic levers (i.e., payment to publishers) with the ad network's technological capability in ad fraud detection. As the ad fraud detection technology improves, the ad network may find it optimal to increase payments to publishers in order to sustain participation and ensure a high-quality traffic ecosystem. This insight offers a partial explanation for the increased payment to publishers because of the observed upward trend in the average cost per click within the Google Ads platform in recent years

(Octoboard 2024), reflecting the interplay between improved fraud mitigation and the need to maintain strong publisher incentives.

In the second research question, we study the following. How does improved fraud generation efficiency affect the ad network's payment to publishers? Contrary to the prevailing notion that reducing economic incentives can help curtail publishers' growing fraud motives (Stone-Gross et al. 2011), our findings suggest that relying solely on economic disincentives may be insufficient to address the evolving nature of ad fraud and can, in some cases, result in suboptimal outcomes for the ad network. Specifically, we find that in response to improvements in fraud generation efficiency, the ad network may need to increase the payment to publishers in conjunction with implementing more stringent technological configurations for fraud detection.

This interplay between economic and technological tools has important implications for ad networks' policy formulation. It underscores the limitations of narrowly focused interventions that rely primarily on economic disincentives to suppress publishers' fraud motives. Instead, our result advocates for a more comprehensive strategy, in which ad networks strategically coordinate both technological and economic instruments to deter fraud more effectively and sustain a resilient advertising ecosystem. This also provides insight into observed strategic behavior by ad networks. On one front, ad networks have implemented increasingly stringent measures to combat ad fraud, particularly in response to the emergence of AI-generated fraudulent activities, such as impersonation ads featuring public figures. For instance, in its 2024 Ads Safety Report, Google announced that it had introduced more than 30 new policy updates aimed at strengthening enforcement against ad fraud (Adegbola 2025). On the other front, the payment to publishers has also increased driven in part by rising ad valuations from advertisers (Veuno 2024), further reinforcing the need for coordinated economic incentives alongside robust technological safeguards.

In general, many ad networks implement a range of payment models, including cost per mille, cost per click, and cost per action. The cost structure associated with fraud generation (i.e., fraud generation efficiency) varies across these models. For instance, the efficiency of generating fraudulent impressions in CPM models can often be higher than that in performance-based models, like CPC (which charges per click) and CPA (which charges per conversion or action). Another important implication of the above finding is that an ad network should consider enacting customized policies instead of a one-size-fits-all solution for different advertising products. This recommendation aligns with anecdotal evidence. For example, Google shares more with publishers in its AdSense for Content program (which primarily operates under a CPM model) than publishers in its AdSense for Search product (which uses both CPC and CPM models) (Sharma 2025).

Next, we study the following research question. Should an ad network continue investing in fraud detection technology, particularly in this era of evolving techniques in fraud generation? Although an ad network can have diminishing returns from technology investment, our findings show that from a dynamic perspective accounting for the evolving fraud generation techniques, the improved fraud generation efficiency can amplify the benefits of technology improvement. This result has important implications for an ad network's technology investment strategy. Specifically, it highlights the critical need for sustained innovation and investment in fraud detection systems, particularly given the adversarial nature of ad fraud, often described as a "cat-and-mouse" game in which fraudsters constantly develop new methods to evade detection. This helps provide a rationale for why Google maintains a large Ad Traffic Quality team dedicated to researching emerging threats and improving its fraud detection capabilities (Google Ads 2025). Recently, Google has leveraged the latest advancements in GenAI and invested in large language models for its fraud detection. In 2024, Google introduced more than 50 enhancements to its LLMs, enabling faster and more accurate detection of fraudulent activities at scale (Adegbola 2025).

A related question is whether reducing the number of malicious publishers constitutes an effective strategy for decreasing fraudulent ad traffic and improving the ad network's profit. Many believe that eliminating malicious publishers helps remove the "bad apples" from the ecosystem, thereby reducing ad fraud and benefiting the ad network (Jackson 2025). However, we show that this belief may not always be true and that reducing the number of malicious publishers can induce more fraud traffic and hurt ad network revenue under certain market conditions. For example, in 2024, Google reported taking broader site-level enforcement action on over 220,000 publisher sites and suspending 39.2 million advertiser accounts (Schwartz 2025). Despite these extensive efforts to eliminate fraudulent actors, the overall rate of fraudulent ad traffic remains persistently high (Fraud Blocker 2025), and the cost because of ad fraud has steadily increased over the years (Dogtiev 2025), highlighting the limitations of these strategies aimed at fully eradicating ad fraud from the ecosystem.

The implication of our findings is that ad networks should not focus solely on punitive or exclusionary measures but rather, should adopt a more nuanced, efficiency-driven approach to fraud governance. Normatively, this entails prioritizing the management and

mitigation of ad fraud over its complete eradication. This aligns with well-established trade-offs in other adversarial domains, such as cybersecurity, where the complete elimination of risk is often economically inefficient and technically infeasible. Moreover, our results highlight the potential of a collaborative detection ecosystem. Advertisers, many of which already invest in fraud detection through in-house systems or third-party solutions, represent an underleveraged asset in enhancing fraud identification and mitigation (Tarasewicz 2024). Ad networks can design incentives or mechanisms to encourage advertiser participation in fraud detection, thereby strengthening system-wide monitoring and responsiveness. This perspective also helps explain industry-wide calls for greater cooperation in addressing invalid traffic. For example, Google has emphasized that "industry collaboration is an essential element of the fight against invalid traffic," and it has actively participated in crossplatform working groups, including those led by the Interactive Advertising Bureau (IAB) Tech Laboratory and the Trustworthy Accountability Group (Google 2024).

Additionally, our findings support the use of neutral third-party auditors to enhance trust and accuracy in fraud detection efforts. Prior research has advocated for such third-party involvement to address information asymmetries in digital advertising markets (Wilbur and Zhu 2009; Chen et al. 2012b, 2015), and our result provides further theoretical support for these proposals. Integrating third-party verification can reduce information asymmetry and help align the incentives of advertisers, publishers, and the ad network. Taken together, these insights inform a broader rethinking of fraud governance strategies in digital platforms. Rather than viewing fraud solely as a threat to be eradicated, ad networks can instead develop resilience-based policies that tolerate manageable levels of fraud while preserving long-term ecosystem health.

Advertisers' estimation of ad valuation plays a central role in guiding campaign decisions but is inherently susceptible to misreporting and manipulation. This raises a critical question. How does advertisers' misestimation of campaign performance affect the profits of the ad network and the publishers? Our analysis reveals that both the ad network and the publishers can benefit from misreporting ad campaign metrics to manipulate advertisers toward overestimating the performance. This introduces a critical vulnerability in the advertising ecosystem, especially when advertisers rely exclusively on performance data provided by the ad network.

This result underscores a broader incentive misalignment between advertisers and intermediaries. When ad networks control both ad delivery and performance measurement, there is limited independent verification of campaign effectiveness, thereby increasing the risk of biased or inflated performance signals. A key takeaway, therefore, is that advertisers should treat network-supplied performance metrics with caution and consider implementing independent mechanisms for campaign tracking and evaluation. For example, CPA advertisers can use server-side page-visit data to validate conversion events and assess attribution accuracy (Chen et al. 2015). Advertisers can also use tools, such as vanity Uniform Resource Locators (URLs), branded hashtags, and coupons coded for specific ad campaigns, to track user engagement and responses. More systematically, advertisers can conduct postcampaign analysis and consumer surveys or leverage in-house teams or third-party services, such as brand-tracking platforms, to obtain a more reliable and holistic view of campaign effectiveness (Davies 2025).

6.2. Future Research Directions

Although our model captures the strategic decisions by the ad network and the publishers quite realistically, it has some limitations and therefore, can pave the way for new research questions. For example, our model considers an interior case where the ad network does not incentivize all malicious publishers to participate. Although this appropriately reflects the reality where an ad network captures a proportion of the market, one may wonder what would happen in the case of full participation. As such, future studies may extend our work to consider this corner case. Similarly, built upon a game-theoretical model to analyze online ad fraud, our study provides theoretical results for understanding the incentive problems in ad fraud generation and detection. Based on our model and results, future research can empirically examine how the publishers' incentives to generate online ad fraud and the ad networks' technology and economic decisions are driven by different marketing conditions. In addition, with the rising applications of blockchain technology in online advertising, another potential research direction could focus on analyzing how such technologies change how ad networks price their service offerings and their downstream effect on advertisers and publishers (Davies 2025).

Acknowledgments

The authors thank the department editor, the associate editor, and anonymous reviewers for thoughtful and constructive comments throughout the review process.

Endnotes

[1] Online Appendix EC.2.1 provides an overview of how ad networks facilitate the matching of publisher-supplied ad inventory with advertiser demand.

[2] We elaborate on the prevalent payment models in Online Appendix EC.2.2 and the related ad fraud in Online Appendix EC.2.3. Our model does not assume a specific type of ad traffic or fraud.

[3] We elaborate on the online advertising ecosystem and the crucial coordination role of the ad network in Online Appendix EC.2.1.

[4] We elaborate on the prevalent payment models in online advertising in Online Appendix EC.2.2. The ad traffic considered in our study is not tied to a specific payment model. Thus, the insights are not affected by a specific type of ad traffic and the associated ad fraud.

[5] We elaborate on the major types of online ad fraud in Online Appendix EC.2.3 and the challenges in deterring ad fraud in Online Appendix EC.2.4.

[6] This represents a case where the ad network is certain of the true type of some publishers in the network. We have also examined a case without a whitelist and found that the results remain qualitatively robust.

[7] We elaborate on the ad network's payment arrangement with publishers in Online Appendix EC.2.5.

[8] Similarly, there could be a "blacklist" for publishers known to be malicious and producing ad fraud. We do not model a blacklist explicitly for two reasons. First, ad networks enforce stringent policies, including preventive and punitive measures, to deter or eliminate known malicious publishers from their platforms. We provide further details in Online Appendix EC.2.6. As a result, these blacklisted publishers are deemed noncompliant with the ad network's policies and consequently, are not permitted to participate. Second, we have analyzed a case where these publishers are allowed to participate in the network even after being included in the "blacklist." Our analysis shows that the ad network would optimally adopt a strategy (through payment or technology configuration) that results in zero fraudulent ad traffic from this group. This contradicts the premise of their malicious nature and effectively removes this "blacklist" from consideration.

[9] Similar to Dellarocas et al. (2013), we consider that nonmalicious publishers generate the same amount of valid ad traffic. This also helps focus analysis on their participation decisions (i.e., whether to participate in the ad network).

[10] We elaborate on ad fraud detection and the ad network's penalty on publishers in Online Appendix EC.2.7.

[11] We consider a case where nonmalicious publishers can appeal to the ad network, which can then correctly waive/refund such penalties (i.e., $w = 0$ in u_{Nj}), and we find that our core results are qualitatively similar.

[12] In Section 5.2, we further examine a case where malicious publishers are not endowed with any valid ad traffic (i.e., $b = 0$) and demonstrate that our fundamental findings apply.

[13] We consider the cost of generating ad fraud to be the same for all malicious publishers to focus on how the ad network's strategic decisions affect their collective ad fraud generation activities. Further, we consider that malicious publishers will participate in the network when permitted as they generally have a lower (or no) ability to monetize than the nonmalicious publishers.

[14] In Section 5.3, we consider an alternative ROC curve whose effective quality also depends on the amount of fraud traffic generated by malicious publishers. We find that our core insights remain qualitatively similar.

[15] We elaborate on how the ad network can set the pairs of sensitivity and FPR on the ROC curve (i.e., configure its ad fraud detection system) in Online Appendix EC.2.8.

[16] We elaborate on the tracking of performance for ad campaigns and publishers in Online Appendix EC.2.9.

[17] In Online Appendix EC.2.1, we elaborate on how an ad network coordinates the ad placement and matching and how it decides on the payment to publishers in practice.

[18] We have analyzed a case where the penalties are not included in the ad network's revenue and found that our major results remain qualitatively robust.

[19] For Propositions 1–5, we present additional comparative statics results in Online Appendix EC.3.

[20] See https://camphouse.io/blog/ad-performance-tracking.

References

Abdallah A, Maarof MA, Zainal A (2016) Fraud detection system: A survey. *J. Network Comput. Appl.* 68:90–113.

Adegbola A (2025) Google: 39 million advertisers suspended; 5.5 billion ads removed in 2024. Accessed May 25, 2025, https://searchengineland.com/google-ads-2024-suspensions-removals-454338.

Almeida PS, Gondim JJC (2018) Click fraud detection and prevention system for ad networks. *Enigma J. Inform. Security Cryptography* 5(1):27–39.

Andreoni J, Harbaugh W, Vesterlund L (2003) The carrot or the stick: Rewards, punishments, and cooperation. *Amer. Econom. Rev.* 93(3):893–902.

Asdemir K, Yurtseven Ö, Yahya MA (2008) An economic model of click fraud in publisher networks. *Internat. J. Electronic Commerce* 13(2):61–90.

Bennett S, Dakpallah G, Garner P, Gilson L, Nittayaramphong S, Zurita B, Zwi A (1994) Carrot and stick: State mechanisms to influence private provider behaviour. *Health Policy Planning* 9(1):1–13.

Bértoa FC, Rodríguez-Teruel J, Barberà O, Barrio A (2014) The carrot and the stick: Party regulation and politics in democratic Spain. *South Eur. Soc. Politics* 19(1):89–112.

Bhargava HK (2022) The creator economy: Managing ecosystem supply, revenue sharing, and platform design. *Management Sci.* 68(7):5233–5251.

Biswas R, Roy S (2021) Botnet traffic identification using neural networks. *Multimedia Tools Appl.* 80(16):24147–24171.

Bouayad L, Padmanabhan B, Chari K (2019) Audit policies under the sentinel effect: Deterrence-driven algorithms. *Inform. Systems Res.* 30(2):466–485.

Braun JA, Eklund JL (2019) Fake news, real money: Ad tech platforms, profit-driven hoaxes, and the business of journalism. *Digital Journalism* 7(1):1–21.

Cai Y, Yee GO, Gu YX, Lung CH (2020) Threats to online advertising and countermeasures: A technical survey. *Digital Threats Res. Practice* 1(2):1–27.

Calvert G (2012) Smart pricing grows the pie. Accessed May 25, 2025, https://research.google/pubs/pub38097/.

Carlile A (2023) Lord Carlile: "Unravelling the sinister link between advertising fraud and terrorism." Accessed May 25, 2025, https://www.politics.co.uk/comment/2023/12/27/lord-carlile-unravelling-the-sinister-link-between-advertising-fraud-and-terrorism/.

Cavusoglu H, Raghunathan S, Cavusoglu H (2009) Configuration of and interaction between information security technologies: The case of firewalls and intrusion detection systems. *Inform. Systems Res.* 20(2):198–217.

Chen J, Stallaert J (2014) An economic analysis of online advertising using behavioral targeting. *MIS Quart.* 38(2):429–450.

Chen M, Pang MS, Kumar S (2021) Do you have a room for us in your IT? An economic analysis of shared IT services and implications for IT industries. *MIS Quart.* 45(1):225–268.

Chen M, Jacob VS, Radhakrishnan S, Ryu YU (2012a) The effect of fraud investigation cost on pay-per-click advertising. *Workshop Econom. Inform. Security (Berlin, Germany).*

Chen M, Jacob VS, Radhakrishnan S, Ryu YU (2012b) The effect of third party investigation on pay-per-click advertising. *Internat. Conf. Inform. Systems* (AIS, Atlanta).

Chen M, Jacob VS, Radhakrishnan S, Ryu YU (2015) Can payment-per-click induce improvements in click fraud identification technologies? *Inform. Systems Res.* 26(4):754–772.

Choi H, Mela CF, Balseiro SR, Leary A (2020) Online display advertising markets: A literature review and future directions. *Inform. Systems Res.* 31(2):556–575.

Clemencon S, Vayatis N (2009) Tree-based ranking methods. *IEEE Trans. Inform. Theory* 55(9):4316–4336.

D'Annunzio A, Russo A (2020) Ad-networks and consumer tracking. *Management Sci.* 66(11):5040–5058.

D'Annunzio A, Russo A (2024) Intermediaries in the online advertising market. *Marketing Sci.* 43(1):33–53.

Davies R (2025) The state of digital advertising fraud in 2024. Accessed May 25, 2025, https://adsdax.com/the-state-of-digital-advertising-fraud-in-2024/.

Dellarocas C, Katona Z, Rand W (2013) Media, aggregators, and the link economy: Strategic hyperlink formation in content networks. *Management Sci.* 59(10):2360–2379.

Dogtiev A (2025) Ad fraud statistics (2025). *Bus. Apps* (August 12), https://www.businessofapps.com/ads/ad-fraud/research/ad-fraud-statistics/.

Dritsoula L, Musacchio J (2014) A game of clicks: Economic incentives to fight click fraud in ad networks. *ACM SIGMETRICS Performance Evaluation Rev.* 41(4):12–15.

Faou M, Lemay A, Décary-Hétu D, Calvet J, Labrèche F, Jean M, Dupont B, Fernande JM (2016) Follow the traffic: Stopping click fraud by disrupting the value chain. *2016 14th Annual Conf. Privacy Security Trust (PST)* (IEEE, Piscataway, NJ), 464–476.

Fich EM, Shivdasani A (2007) Financial fraud, director reputation, and shareholder wealth. *J. Financial Econom.* 86(2):306–336.

Fou A (2020a) Examples of incorrect measurements by "black box" fraud detection. *Forbes* (July 24), https://www.forbes.com/sites/augustinefou/2020/07/24/how-to-select-a-fraud-verification-vendor/.

Fou A (2020b) Insult to injury—How ad fraud harmed good publishers for years. *Forbes* (June 12), https://www.forbes.com/sites/augustinefou/2020/06/12/insult-to-injuryhow-ad-fraud-harmed-good-publishers-for-years.

Fraud Blocker (2022) Ad fraud is a problem sellers don't want to solve. (April 5), https://fraudblocker.com/articles/ad-fraud-is-a-problem-that-sellers-dont-want-to-solve.

Fraud Blocker (2025) Invalid click rate: Averages and benchmarks for Google Ads. (November 2), https://fraudblocker.com/articles/invalid-click-rate-benchmarks-for-google-ads.

Fulgoni GM (2016) Fraud in digital advertising: A multibillion-dollar black hole: How marketers can minimize losses caused by bogus web traffic. *J. Advertising Res.* 56(2):122–125.

Gibbons M (2025) 50 Google Ads statistics to know in 2025 and beyond. Accessed May 25, 2025, https://www.webfx.com/blog/marketing/google-ads-statistics/.

Google (2024) What is invalid traffic? Accessed May 25, 2025, https://www.google.com/ads/adtrafficquality/invalid-activity/.

Google (2025a) [UA] Google Ads performance report [legacy]. Accessed May 25, 2025, https://support.google.com/analytics/answer/9944574.

Google (2025b) Prevention of invalid clicks and impressions. Accessed May 25, 2025, https://support.google.com/admanager/answer/1298900.

Google Ads (2025) How does Google prevent invalid activity? Accessed May 25, 2025, https://www.google.com/ads/adtrafficquality/how-we-prevent-it/.

Gordon BR, Jerath K, Katona Z, Narayanan S, Shin J, Wilbur KC (2021) Inefficiencies in digital advertising markets. *J. Marketing* 85(1):7–25.

Graham M (2021) Facebook knew ad metrics were inflated, but ignored the problem to make more money, lawsuit claims. Accessed May 25, 2025, https://www.cnbc.com/2021/02/18/facebook-knew-ad-metrics-were-inflated-but-ignored-the-problem-lawsuit-claims.html.

Han H, Tang JJ, Tang Q (2021) Goodwill impairment, securities analysts, and information transparency. *Eur. Accounting Rev.* 30(4):767–799.

Hu Y, Shin J, Tang Z (2016) Incentive problems in performance-based online advertising pricing: Cost per click vs. cost per action. *Management Sci.* 62(7):2022–2038.

Jackson J (2025) Click fraud uncovered: Ultimate guide to protecting ad spend. Accessed May 25, 2025, https://hitprobe.com/blog/click-fraud-complete-guide#types-of-click-fraud.

Jacob A (2023) A publishers guide to ad fraud. Accessed May 25, 2025, https://www.monetizemore.com/blog/a-publishers-guide-to-ad-fraud/.

Ji Y, Kumar S, Mookerjee V (2016) When being hot is not cool: Monitoring hot lists for information security. *Inform. Systems Res.* 27(4):897–918.

Jodzevica A (2025) 15 best ad networks for publishers in 2025. Accessed May 25, 2025, https://setupad.com/blog/best-ad-networks-for-publishers/.

Kahn R (2020) What is a bot? How do bots impact your digital ad campaigns? Accessed May 25, 2025, https://www.anura.io/blog/what-is-a-bot-and-how-do-bots-impact-digital-ad-campaigns.

Kahn R (2023) Why aren't there laws to stop ad fraud? Accessed May 25, 2025, https://www.anura.io/blog/why-arent-there-laws-to-stop-ad-fraud.

Kahn R (2024) Ad fraud cost advertisers $125 billion in 2023. Accessed May 25, 2025, https://www.anura.io/blog/ad-fraud-cost-advertisers-125-billion-in-2023.

Karpenkova A (2020) Facebook ad account is disabled? Here's what you can do. Accessed May 25, 2025, https://joinative.com/facebook-ad-accound-ban.

Khoury GE (2019) Google settles AdSense publisher lawsuit for $11M. Accessed May 25, 2025, https://www.findlaw.com/legalblogs/technologist/google-settles-adsense-publisher-lawsuit-for-11m/.

Kim J, Lee KH, Kim J (2023) Linking blockchain technology and digital advertising: How blockchain technology can enhance digital advertising to be more effective, efficient, and trustworthy. *J. Bus. Res.* 160:113819.

Kircher T, Foerderer J (2024) Ban targeted advertising? An empirical investigation of the consequences for app development. *Management Sci.* 70(2):1070–1092.

Kitts B, Zhang JY, Wu G, Brandi W, Beasley J, Morrill K, Ettedgui J, et al. (2015) Click fraud detection: Adversarial pattern recognition over 5 years at Microsoft. Abou-Nasr M, Lessmann S, Stahlbock R, Weiss GM, eds. *Real World Data Mining Applications* (Springer International Publishing, Cham, Switzerland), 181–201.

Klym N, Clark D (2019) The future of the ad-supported internet ecosystem. MIT Internet Policy Research Initiative technical report, Massachusetts Institute of Technology, Cambridge.

Kshetri N, Voas J (2019) Online advertising fraud. *Computer* 52(1):58–61.

Kumar S, Tan Y, Wei L (2020) When to play your advertisement? Optimal insertion policy of behavioral advertisement. *Inform. Systems Res.* 31(2):589–606.

Kumar N, Venugopal D, Qiu L, Kumar S (2019) Detecting anomalous online reviewers: An unsupervised approach using mixture models. *J. Management Inform. Systems* 36(4):1313–1346.

Liu D, Kumar S, Mookerjee VS (2020) Flexible and committed advertising contracts in electronic retailing. *Inform. Systems Res.* 31(2):323–339.

Macey JR, Miller GP (1990) Good finance, bad economics: An analysis of the fraud-on-the-market theory. *Stanford Law Rev.* 42(4):1059–1092.

Mendoza JP, Wielhouwer JL (2015) Only the carrot, not the stick: Incorporating trust into the enforcement of regulation. *PLoS One* 10(2):117–122.

Mookerjee R, Kumar S, Mookerjee V (2012) To show or not show: Using user profiling to manage internet advertisement campaigns. *Interfaces* 42(5):449–464.

Mookerjee V, Mookerjee R, Bensoussan A, Yue WT (2011) When hackers talk: Managing information security under variable attack rates and knowledge dissemination. *Inform. Systems Res.* 22(3):606–623.

Mungamuru B (2010) Managing the quality of cost-per-click traffic. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Mungamuru B, Weis S, Garcia-Molina H (2008) Should ad networks bother fighting click fraud? (Yes, they should.) Technical report, Stanford University, Stanford, CA.

Nagaraja S, Shah R (2019) Clicktok: Click fraud detection using traffic analysis. *Proc. 12th Conf. Security Privacy Wireless Mobile Networks* (ACM, New York), 105–116.

Nandini C (2019) Detecting and preventing click fraud: The economic and legal aspects. *IUP Law Rev.* 9(2):12–50.

Octoboard (2024) Google Ads PPC trends 2024: CPC analysis. Accessed May 25, 2025, https://www.octoboard.com/blog/ppc-analytics/google-ads-paid-advertising-cpc-trends-2024.

Parti A (2020) What is an ad network and how does it work? Accessed May 25, 2025, https://rocketium.com/academy/what-is-an-ad-network-and-how-does-it-work/.

Prokopets E (2021) How to measure advertising campaign effectiveness offline and online. Accessed May 25, 2025, https://resources.latana.com/post/measure-campaign-effectiveness/.

Pu J, Nian T, Qiu L, Cheng HK (2022) Platform policies and sellers' competition in agency selling in the presence of online quality misrepresentation. *J. Management Inform. Systems* 39(1):159–186.

Ranne K (2024) Effective ad fraud solutions for businesses. Accessed May 25, 2025, https://www.nexd.com/blog/ad-fraud-prevention/.

Ravichandran T, Han S, Mithas S (2017) Mitigating diminishing returns to R&D: The role of information technology in innovation. *Inform. Systems Res.* 28(4):812–827.

Rayobyte (2025) What to know about ad fraud (and the ultimate guide to prevent it). Accessed May 25, 2025, https://rayobyte.com/blog/what-is-ad-fraud/.

Richet JL (2022) How cybercriminal communities grow and change: An investigation of ad-fraud communities. *Tech. Forecasting Soc. Change* 174:121282.

Sailusha R, Gnaneswar V, Ramesh R, Rao GR (2020) Credit card fraud detection using machine learning. *4th Internat. Conf. Intelligent Comput. Control Systems (Madurai, India)*, 1264–1270.

SaleHoo (2025) US digital ad spend growth (2017–2028): Key drivers, challenges & outlook. Accessed May 25, 2025, https://www.salehoo.com/learn/united-states-digital-ads-spending.

Saluja S (2024) Are publishers the real victims of ad fraud? Accessed May 25, 2025, https://www.mile.tech/blog/are-publishers-the-real-victims-of-ad-fraud.

Schiff A (2016) Ad networks are starting to get their anti-fraud ducks in a row. Accessed May 25, 2025, https://www.adexchanger.com/mobile/ad-networks-starting-get-anti-fraud-ducks-row/.

Schwartz B (2025) Google Ads suspended 200% more advertisers (39.2M) and removed 5.5 billion ads. Accessed May 25, 2025, https://www.seroundtable.com/google-ads-safety-report-39242.html.

Sharma R (2025) Google AdSense updates: Revenue-share structure and moving to CPM. Accessed May 25, 2025, https://www.mile.tech/blog/google-adsense-updates-revenue-share-structure-moving-to-cpm.

Sisodia D, Sisodia DS (2023) A hybrid data-level sampling approach in learning from skewed user-click data for click fraud detection in online advertising. *Expert Systems* 40(2):e13147.

Stone-Gross B, Stevens R, Zarras A, Kemmerer R, Kruegel C, Vigna G (2011) Understanding fraudulent activities in online ad exchanges. *Proc. 2011 ACM SIGCOMM Conf. Internet Measurement* (ACM, New York), 279–294.

Sun M, Zhu F (2013) Ad revenue and content commercialization: Evidence from blogs. *Management Sci.* 59(10):2314–2331.

Tanzako A (2024) Latest in ad fraud detection. Accessed May 25, 2025, https://clickpatrol.com/latest-in-ad-fraud-detection/.

Tarasewicz A (2024) Safeguarding the digital ad ecosystem: Strategies for ad fraud prevention in 2024. Accessed May 25, 2025, https://www.admonsters.com/safeguarding-the-digital-ad-ecosystem-strategies-for-ad-fraud-prevention-in-2024/.

Taylor L (2024) It's time to stop blaming ad networks for ad fraud. Accessed May 25, 2025, https://www.trafficguard.ai/news/its-time-to-stop-blaming-ad-networks-for-ad-fraud.

Trajcheva S (2023) What does Google do to prevent click fraud? Accessed May 25, 2025, https://cheq.ai/blog/what-does-google-do-to-prevent-click-fraud/.

Tsur E (2024) How ad fraud detection is saving money for customers and online retailers. Accessed May 25, 2025, https://www.memcyco.com/how-ad-fraud-detection-is-saving-money-for-customers-and-online-retailers/.

Veuno (2024) Ad platform cost rises and predictions. Accessed May 25, 2025, https://www.veuno.com/ad-platform-cost-rises-and-predictions/.

Wiatr R, Lyutenko V, Demczuk M, Słota R, Kitowski J (2019) Click-fraud detection for online advertising. *Internat. Conf. Parallel Processing Appl. Math.* (Springer), 261–271.

Wilbur KC, Zhu Y (2009) Click fraud. *Marketing Sci.* 28(2):293–308.

Zhu X, Tao H, Wu Z, Cao J, Kalish K, Kayne J (2017a) Ad fraud categorization and detection methods. *Fraud Prevention in Online Digital Advertising*, Springer Briefs in Computer Science (Springer International Publishing, Cham, Switzerland), 25–38.

Zhu X, Tao H, Wu Z, Cao J, Kalish K, Kayne J (2017b) Ad fraud detection tools and systems. *Fraud Prevention in Online Digital Advertising*, Springer Briefs in Computer Science (Springer International Publishing, Cham, Switzerland), 45–49.

Zorz Z (2020) What is ad fraud and how can advertisers fight against it? Accessed May 25, 2025, https://www.helpnetsecurity.com/2020/11/06/what-is-ad-fraud-and-how-can-advertisers-fight-against-it/.