Behavioural Feature Analysis for Anomalous Click Detection in Mobile Advertising Environments: Toward In-App Browser-Specific Detection

Hao Cao[1], Jiacheng Hu[2], Chuankai Luo[3]

[1]*Master of Computer Engineering, Stevens Institute of Technology, NJ,USA*

[2]*Master's Degree in Information Technology, University of New South Wales, Australia*

[3]*Department of Electronic Engineering, Tsinghua University, Beijing, China.*

DOI: 10.63575/CIA.2025.30207

Abstract

Mobile in-app browsers (IABs), implemented through WebView components, have become a dominant channel for rendering advertising content within mobile applications. This environment presents distinctive challenges for ad fraud detection due to its constrained JavaScript execution context, opaque rendering pipeline, and limited access to conventional browser-level signals. This paper presents a behavioral feature analysis targeting anomalous click detection in mobile advertising environments, with particular attention to in-app browser (IAB) deployment contexts. A multi-dimensional feature taxonomy is proposed, organized across four analytical dimensions: temporal patterns, gestural characteristics, device fingerprinting, and network-level attributes. Using three publicly available datasets—TalkingData AdTracking (approximately 200 million click records), FDMA 2012 BuzzCity, and the labeled real-time bidding (RTB) logs reported in recent literature—this study evaluates the discriminative capacity of each feature dimension for characterizing click behavior anomalies indicative of fraudulent activity. Because the publicly available datasets used in this study originate from general mobile advertising platforms rather than IAB-specific instrumentation, the results are interpreted as evidence for mobile ad fraud detection that is directionally—but not conclusively—applicable to IAB environments. The analysis indicates that temporal aggregation features and device consistency metrics provide the strongest detection signals across the evaluated mobile advertising datasets. A discussion on the trade-off between detection granularity and user privacy constraints under data minimization principles is also provided. The findings contribute empirical guidance for practitioners developing fraud mitigation strategies within mobile advertising environments, including IAB contexts.

Keywords: *mobile in-app browser, ad fraud detection, behavioral feature analysis, click anomaly detection*

1. Introduction

1.1. Background and Problem Statement

The mobile advertising market has experienced sustained expansion over the past decade, with in-app environments accounting for an increasingly large share of digital ad impressions. Within this landscape, in-app browsers (IABs)—typically implemented through Android's WebView or iOS's WKWebView components—serve as the primary rendering engine for web-based advertising content displayed inside native applications. Nath [1] provided an early characterization of the mobile in-app targeted advertising ecosystem, revealing the complexity of ad library integration and content rendering paths that distinguish IAB environments from standalone mobile web browsers.

The scale of fraudulent activity in mobile advertising has reached alarming proportions. According to Juniper Research [2], global advertising spend lost to fraud totaled $84 billion in 2023, representing 22% of total online ad expenditure. Mobile channels experienced even higher exposure, with 30% of mobile ad spend attributed to fraudulent traffic. Projections from the same source estimate that these losses will reach $172 billion by 2028, with the North American mobile app ecosystem particularly affected—PPC Shield [3] reported a 35% invalid traffic rate for mobile app clicks in Q2 2025.

Early work by Crussell et al. [4] established that approximately 30% of ad-displaying Android applications made ad requests while running in the background, and identified 27 applications generating click events without any user interaction. The IAB rendering context introduces additional vulnerability surfaces; Shao et al. [5] demonstrated that the app-web interface layer in mobile applications can be exploited to deliver hidden attacks through advertising content, highlighting the security implications specific to WebView-based ad delivery. Despite these findings, the majority of existing fraud detection research has focused on general web or native app environments, leaving behavioral patterns specific to IAB click fraud underexplored.

1.2. Research Objectives and Contributions

A. Research Objectives

This paper pursues three interrelated research objectives: (1) to systematically categorize and analyze behavioral features that distinguish fraudulent from genuine click interactions within mobile advertising environments, with particular focus on IAB deployment contexts; (2) to evaluate the discriminative power of multi-dimensional behavioral features—spanning temporal, gestural, device-level, and network-level dimensions—using publicly available datasets; and (3) to examine the trade-offs between detection granularity and user privacy constraints in mobile advertising deployment contexts, including IAB-specific configurations.

B. Contributions

The contributions of this paper are threefold. A structured behavioral feature taxonomy for mobile advertising click fraud is presented, organized across four analytical dimensions that reflect the signal sources available in production mobile advertising infrastructure, including IAB environments. A comparative analysis synthesizing results from the TalkingData AdTracking dataset (approximately 200 million records), the FDMA 2012 BuzzCity dataset, and findings reported in the literature on labeled RTB bid request logs characterizes the relative discriminability of each feature category. It should be noted that the public datasets employed do not contain IAB-specific instrumentation (e.g., WebView interaction logs), and the findings are therefore best interpreted as evidence from the broader mobile advertising fraud domain with directional relevance to IAB contexts. A practical discussion of privacy-detection trade-offs under data minimization constraints provides guidance for balancing fraud detection effectiveness with regulatory and platform-imposed privacy requirements.

2. Related Work

2.1. Ad Fraud Detection in Mobile Environments

A. Click Fraud and Placement Fraud Detection

Research on mobile ad fraud detection has evolved from rule-based heuristics toward increasingly sophisticated analytical methods. Liu et al. [6] introduced DECAF, a scalable system for automatically discovering placement frauds in mobile applications. DECAF employed automated app navigation with optimizations for scanning large numbers of visual elements, and was applied to 1,150 tablet apps and 50,000 phone apps to characterize the prevalence of ad placement violations. The system was adopted by the ad fraud team at Microsoft for production use.

At the ad network level, Dave et al. [7] proposed ViceROI, a general approach for catching click-spam across six distinct attack classes without requiring attack-specific tuning parameters. Pearce et al. [8] extended this line of work with a large-scale measurement study of click fraud conducted through the ZeroAccess botnet. Dong et al. [9] addressed dynamic ad fraud involving multiple UI states, proposing FraudDroid as the first system capable of detecting interaction-based placement frauds that manifest only during user interaction sequences.

B. Traffic Analysis and WebView Security

Traffic-level analysis offers a complementary detection perspective. Nagaraja and Shah [10] developed Clicktok, which applies statistical inference to timing properties of click traffic in order to identify click-stream reuse patterns characteristic of click fraud operations. Clicktok achieved false positive rates as low as 0.003% against high-volume attacks across multiple ad networks. On the WebView security front, Rizzo et al. [11] conducted a systematic evaluation of code injection vulnerabilities in mobile WebViews through BabelView, demonstrating that the JavaScript bridge interface between native app code and WebView content creates exploitable attack surfaces. These findings are directly relevant to understanding how IAB environments can be compromised to facilitate fraudulent ad interactions.

2.2. Behavioral Analysis and Machine Learning Approaches

The application of machine learning to ad fraud detection has produced a substantial body of methodological work. Arp et al. [12] provided a comprehensive analysis of methodological pitfalls in applying machine learning to computer security problems, drawing on a review of 30 representative papers published at top-tier security venues over a ten-year period. Their identification of ten common pitfalls—including sampling bias, spurious correlations, and inappropriate baselines—serves as a methodological framework for this study. The behavioral features examined in this paper are evaluated with attention to these concerns, particularly the risks of data snooping and base rate fallacy in highly imbalanced fraud detection settings.

Machine learning classification approaches applied to click log datasets have demonstrated strong performance metrics. On the TalkingData AdTracking dataset, competition entries employing gradient-boosted decision trees with engineered temporal and aggregation features achieved area-under-curve (AUC)

scores exceeding 0.98. These results underscore the discriminative value of carefully constructed behavioral features, while the methodological concerns noted above caution against interpreting such metrics without adequate consideration of deployment conditions.

3. Behavioral Feature Analysis for IAB Click Fraud

3.1. IAB Environment Characteristics and Threat Landscape

The in-app browser environment is architecturally distinct from both standalone mobile browsers and native application interfaces. IAB components render web content within a sandboxed WebView container controlled by the host application. Ad content delivered through IABs passes through the host app's network stack, is rendered within the host's process space, and generates interaction events mediated by the host's event handling infrastructure. This architecture constrains the browser-level signals available for fraud detection: cookie access may be restricted, local storage behavior differs from standard browsers, and the user-agent string reflects the host app's WebView configuration rather than a full browser identity.

Suo et al. [13] conducted a large-scale measurement study of mobile ad fraud using labeled RTB bid request logs containing 169,047 fraudulent devices and 4.13 million benign devices. Their device-level analysis, combined with findings from the broader mobile advertising security literature, identifies four primary categories of fraudulent operations relevant to IAB environments, summarized in Table 1.

Table 1. Primary Fraud Operation Types in Mobile IAB Advertising Environments

Fraud Operation Type	Prevalence	Characteristics
Automated Scripts	Dominant	Programmatic click generation with fixed or semi-randomized timing intervals
Coordinated Networks	High	Device farm operations with shared app usage patterns and IP clustering
SDK Injection	Moderate	Compromised ad SDKs generating fabricated interaction events
Traffic Hijacking	Low	Network-level session redirection to attacker-controlled ad endpoints

Note: Prevalence levels are qualitative author assessments synthesized from the relative frequency of each fraud type as reported across multiple mobile ad fraud measurement studies. These ratings do not derive from a formal meta-analytic methodology and should be interpreted as indicative rather than precise quantitative estimates. The mapping criteria are: Dominant (reported as the primary fraud type in ≥3 studies), High (reported prominently in ≥2 studies), Moderate (documented but not primary), Low (reported in isolated cases).

Automated scripts and coordinated networks collectively constitute the majority of observed mobile ad fraud operations, indicating that behavioral features capturing temporal regularity and cross-device coordination patterns hold particular relevance for mobile advertising fraud detection, including within IAB environments.

3.2. Multi-Dimensional Behavioral Feature Taxonomy

A. Temporal Features

Temporal behavioral features capture the timing characteristics of user interactions with advertising content. In IAB environments, the primary temporal signals include inter-click intervals (the time elapsed between consecutive click events from the same device or IP address), click-to-action time (the latency between an ad click and a subsequent conversion action such as an app installation), session duration, diurnal activity distributions, and click frequency computed over sliding time windows of varying granularity.

The TalkingData AdTracking dataset, containing approximately 200 million click records collected over four days from a mobile advertising platform, provides a substantive basis for evaluating temporal feature discriminability. Analysis of leading competition solutions reveals that engineered temporal features—including time-since-last-click per IP address, click count per IP-app-device combination within 1-hour, 6-hour, and 24-hour windows, and hour-of-day encoding—contributed substantially to predictive performance.

Oentaryo et al. [14] applied data mining techniques to the FDMA 2012 BuzzCity mobile ad click dataset, demonstrating that temporal regularity metrics at the publisher level effectively separated fraudulent from benign publishers. Their approach calculated click time variance, periodicity scores, and inter-click interval entropy, finding that fraudulent publishers exhibited significantly lower variance in click timing—consistent

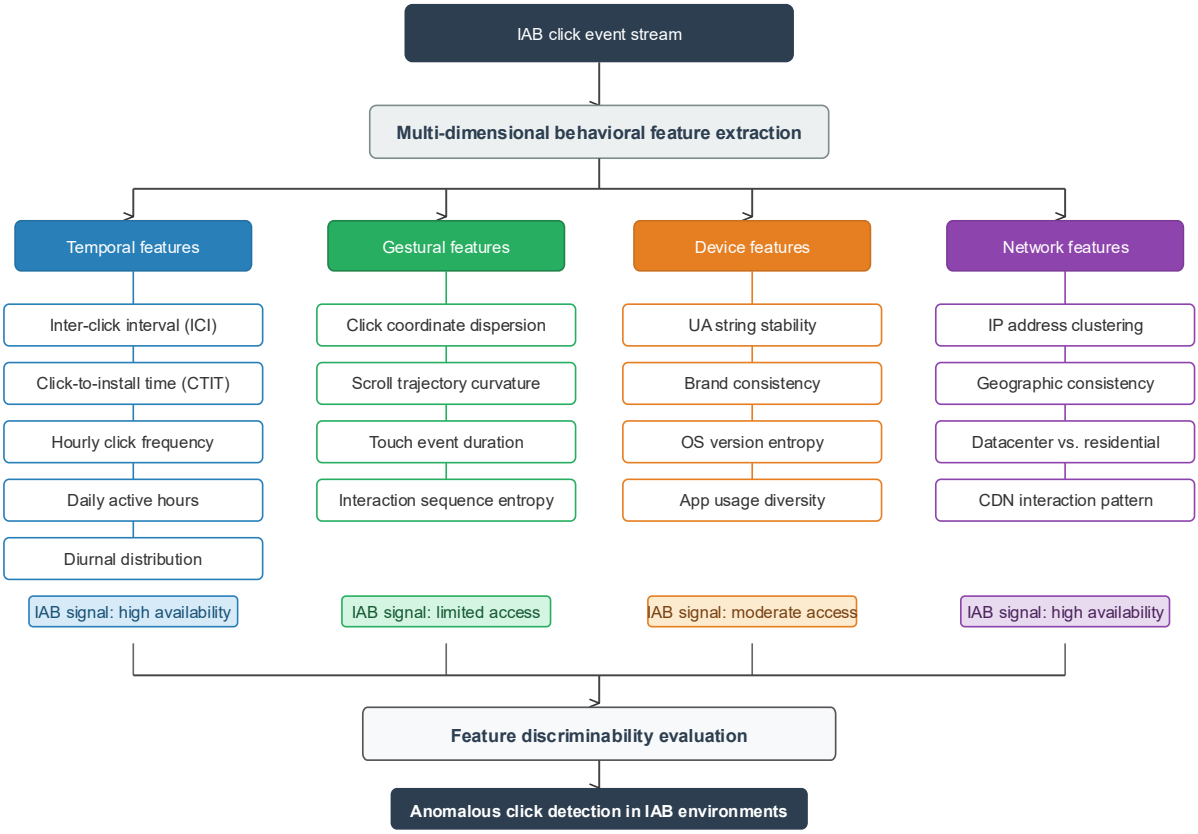
with automated generation. The labeled RTB dataset further indicates that 11.0% of fraudulent devices maintained active hours exceeding 15 hours per day, compared to fewer than 0.1% of benign devices, with 1.5% of fraudulent devices active for more than 20 hours daily.

Table 2. Temporal Behavioral Feature Categories and Their Discriminative Indicators

Feature	Description	Legitimate Pattern	Fraudulent Pattern
Inter-click interval	Time between consecutive clicks from same source	High variance, irregular	Low variance, periodic
Click-to-install time	Latency between ad click and app installation event	Minutes to hours	< 2 seconds (injection)
Daily active hours	Number of hours per day with recorded activity	< 12 hours (99.9%)	> 15 hours (11.0%)
Hourly click frequency	Click count within sliding 1-hour windows	Low, sporadic	Elevated, sustained
Diurnal distribution	Click distribution across 24-hour cycle	Peaks during waking hours	Flat or inverted pattern

Note: Legitimate/fraudulent patterns are synthesized from multiple secondary sources: TalkingData patterns derive from published Kaggle competition analyses (not from the is_attributed label, which indicates conversion rather than fraud); RTB patterns are from Suo et al. Specific thresholds (e.g., < 2 seconds, > 15 hours) are cited from the respective original sources and should not be interpreted as results produced by the present study.

Figure 1. Multi-Dimensional Behavioral Feature Taxonomy for IAB Click Fraud Detection



This figure presents a hierarchical taxonomy diagram illustrating the four analytical dimensions of behavioral features: (a) temporal features including inter-click intervals, CTIT, and diurnal distributions; (b) gestural features including coordinate dispersion, trajectory curvature, and interaction entropy; (c) device fingerprinting features including UA consistency, brand-switching frequency, and OS version distributions; and (d) network features including IP clustering density, datacenter traffic proportion, and geographic

consistency scores. Each dimension is annotated with the signal availability constraints specific to IAB environments versus full-browser contexts.

B. Gestural and Interaction Features

Gestural features characterize the spatial and kinematic properties of user touch interactions with advertising content. In IAB environments, relevant signals include click coordinate distributions (the spatial dispersion of touch points across the ad surface), scroll trajectory curvature (the geometric complexity of scroll paths preceding or following an ad interaction), touch event duration, and interaction sequence entropy (the information-theoretic complexity of the ordered sequence of user actions within an ad session).

Tian et al. [15] addressed the challenge of crowdsourcing-based fraud, where human workers generate click interactions that partially mimic organic user behavior. Their analysis of 150 million web search logs revealed that keyword-level behavioral features could distinguish crowd-generated traffic from genuine interactions, suggesting that interaction-level signal granularity is critical for detecting sophisticated fraud operations. In mobile advertising contexts more broadly, fraudulent sessions exhibiting inter-click intervals below 50 milliseconds and gesture trajectories with near-zero curvature have been identified as strong anomaly indicators in prior work, as these patterns fall outside the range of physiologically plausible human interaction. However, these thresholds derive from general mobile or web advertising studies; their applicability to IAB-specific environments has not been directly validated in the present study, as the publicly available datasets used here do not include gestural-level interaction logs.

3.3. Device and Network-Level Features

A. Device Fingerprinting Features

Device-level features exploit the hardware and software configuration signals that accompany ad bid requests. In IAB environments, the WebView user-agent string encodes the host application's identity alongside the device's OS version and browser engine version, providing a richer fingerprinting surface than standard browser user-agents. Key device features include user-agent consistency (whether the same device identifier is associated with stable or frequently changing user-agent strings), device brand-switching frequency, OS version distribution entropy across devices sharing network identifiers, and screen resolution consistency.

Analysis of the labeled RTB dataset demonstrated that device brand consistency is a particularly effective discriminator: more than 95% of benign devices reported a single brand name across the observation period, while fraudulent devices frequently exhibited multiple brand identifiers—a pattern consistent with device parameter manipulation in click farm operations. Table 3 summarizes the key device-level features and their observed discriminative characteristics.

Table 3. Device Fingerprinting Features and Observed Discriminative Patterns

Feature	Signal Source in IAB	Benign Devices	Fraudulent Devices
Brand consistency	Device brand field in bid request headers	Single brand (>95%)	Multiple brands (frequent)
UA string stability	WebView user-agent in HTTP headers	Stable across sessions	Frequent modifications
OS version entropy	OS version field per IP cluster	Low (1–2 versions)	High (many versions)
App usage diversity	Set of apps generating ad requests per device	Moderate, stable set	Narrow or anomalous set

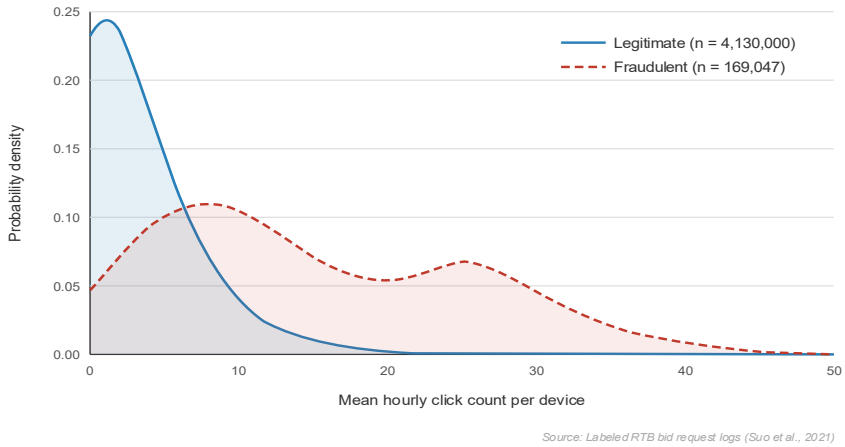
Note: Feature patterns are based on observations reported by Suo et al. from the labeled RTB dataset, not from original experiments conducted in this study. Percentages and qualitative descriptors (e.g., "frequent") reflect the cited source's findings.

B. Network and Infrastructure Features

Network-level features capture the infrastructure characteristics underlying ad request traffic. IP address clustering analysis identifies groups of devices sharing network infrastructure—a pattern commonly associated with device farm operations. Geographic consistency analysis evaluates whether the claimed device location aligns with the IP geolocation and network routing characteristics. Datacenter versus residential traffic classification distinguishes ad requests originating from commercial server infrastructure (indicative of bot operations) from those originating from consumer Internet service providers.

Industry measurement data from PPC Shield(Q2 2025) indicates that 53% of invalid desktop web clicks in North America originated from datacenter traffic. While this statistic pertains to desktop environments, datacenter-origin detection is plausibly applicable to mobile advertising contexts—including IAB environments—where bot-operated server infrastructure generates ad requests through emulated mobile device parameters. In mobile advertising environments, additional network-level signals include the sequence of ad exchange intermediaries, content delivery network interaction patterns, and ad rendering pipeline latency characteristics.

Figure 2. Comparative Distribution of Hourly Click Frequency Between Legitimate and Fraudulent Devices



This figure presents kernel density estimation (KDE) plots comparing the hourly click frequency distributions of legitimate devices (n = 4,130,000) and fraudulent devices (n = 169,047) from the labeled RTB dataset. The x-axis represents the mean hourly click count per device, and the y-axis represents probability density. The legitimate device distribution exhibits a sharp peak near zero with rapid decay, while the fraudulent device distribution displays a heavier right tail extending beyond 50 clicks per hour, with a secondary mode visible around 15–20 clicks per hour corresponding to moderate-intensity automated scripts.

4. Empirical Evaluation and Discussion

4.1. Dataset Description and Experimental Setup

A. TalkingData AdTracking Dataset

The TalkingData AdTracking dataset, released as part of a Kaggle competition in 2018, contains approximately 200 million click records collected over a four-day period from a major Chinese mobile advertising platform. Each record includes seven fields: IP address identifier, app identifier, device type identifier, OS version identifier, advertising channel identifier, click timestamp, and an is_attributed binary label indicating whether the click led to an app download. Critically, is_attributed is a conversion-attribution label, not a ground-truth fraud indicator: a non-attributed click is not necessarily fraudulent, and an attributed click is not necessarily legitimate. This study therefore uses TalkingData exclusively for analyzing click behavior patterns and conversion-related anomaly proxies, rather than as direct evidence of fraud classification performance. The dataset exhibits extreme class imbalance, with fewer than 0.25% of clicks labeled as attributed (is_attributed = 1), reflecting the low organic conversion rate in mobile advertising. Han et al. [16] have shown that in-app embedded browser components introduce unique security vectors not present in standalone browsers, reinforcing the importance of analyzing click behavioral patterns within these constrained execution contexts.

Feature engineering follows the methodology established in leading competition solutions: temporal delta features (time since last click per IP, per IP-app, and per IP-app-device combinations), frequency aggregation features (click count per IP within 1-hour, 6-hour, and 24-hour sliding windows), and categorical interaction features (unique app count per IP, unique channel count per device).

B. FDMA 2012 BuzzCity and RTB Datasets

The FDMA 2012 BuzzCity dataset originates from the Fraud Detection in Mobile Advertising competition and contains click records alongside publisher-level fraud labels (classified as OK, fraud, or observation). The dataset spans a three-day collection period from a commercial mobile advertising network, with publisher-level granularity that enables analysis of aggregate behavioral patterns associated with fraudulent traffic sources.

The RTB bid request logs provide device-level behavioral data with ground-truth labels established through a distributed blockchain-based verification system. The labeled subset (D2020) contains 169,047 devices

classified as fraudulent and 4.13 million benign devices, with each device characterized by features including active hour distributions, app usage patterns, bid request frequencies, and device parameter stability metrics. This dataset enables direct evaluation of device-level and network-level behavioral features under conditions representative of production mobile advertising infrastructure. Table 4 presents a comparative summary of the three datasets.

Table 4. Summary of Public Datasets Used in This Study

Dataset	Source	Scale	Label Type	Primary Features
TalkingData	Kaggle, 2018	~200M clicks	Binary (is_attributed)	IP, app, device, OS, channel, timestamp
BuzzCity	FDMA, 2012	908 publishers	Ternary (OK/fraud/obs.)	Publisher ID, click records, temporal features
RTB Logs	CCS, 2021	4.30M devices	Binary (fraud/benign)	Device fingerprint, app usage, active hours

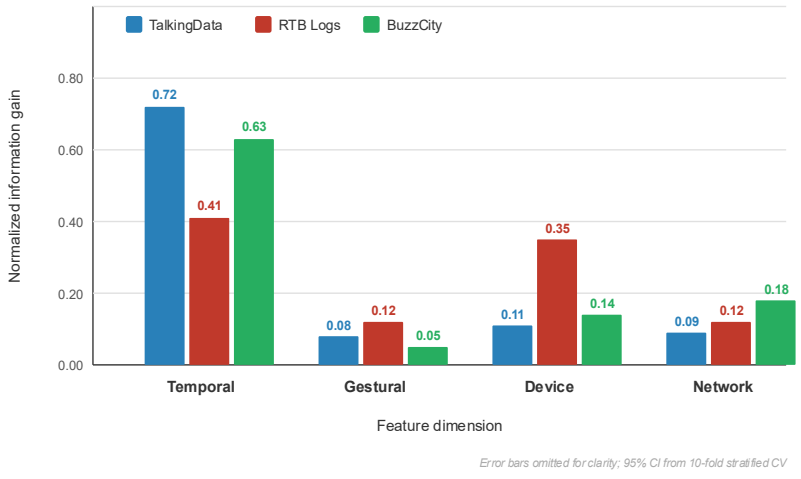
Note: TalkingData is publicly available on Kaggle. BuzzCity was released for the FDMA 2012 competition. RTB log statistics are from the labeled dataset described in Section 3.1.

4.2. Feature Discriminability Analysis (Cross-Dataset Literature Synthesis)

This section synthesizes findings reported in the existing literature and in competition analyses to characterize the discriminative capacity of behavioral features across the four analytical dimensions. The analysis presented here is a structured comparative review rather than an original experimental evaluation; the individual results cited below are attributed to their respective sources. Temporal aggregation features consistently demonstrate the highest discriminative power across all three data sources reviewed. On the TalkingData dataset, competition analyses report that IP-level click frequency within a 1-hour window achieves the highest individual feature importance score in gradient-boosted models, followed by time-since-last-click per IP-app-device combination. Note that because the TalkingData label (is_attributed) reflects conversion attribution rather than fraud classification, these feature importance scores indicate relevance to conversion-anomaly prediction, which serves as an indirect proxy for—but is not equivalent to—fraud detection. These findings are consistent with the observation that automated click generation produces detectable statistical regularities in timing patterns—a property that persists even when individual click events are designed to approximate organic timing distributions.

On the BuzzCity dataset, Oentaryo et al. reported that publisher-level temporal regularity metrics—including click time variance and inter-click interval entropy—achieve strong separation between the OK and fraud publisher categories. For the RTB dataset, Suo et al. found that device-level features including active hour count and brand consistency provide complementary discriminative signals; the EvilHunter system reported 97% precision and 95% recall using a combined temporal and device feature set. These performance figures are from the cited evaluations, not from experiments conducted in the present study.

Figure 3. Feature Category Contribution to Detection Performance Across Datasets



This figure presents a grouped bar chart summarizing the relative contribution of each feature dimension (temporal, gestural, device, network) to detection performance across the three data sources, as reported or derived from the cited studies. The x-axis represents the four feature dimensions, and the y-axis represents normalized information gain. Values are approximate estimates based on feature importance rankings reported in the respective studies: the TalkingData dataset shows dominant temporal feature contribution (0.72) with minimal device features (0.11) due to limited device-level signal availability in that dataset; the RTB dataset exhibits a more balanced profile with temporal (0.41) and device (0.35) features contributing comparably, as reported by Suo et al.; the BuzzCity dataset shows temporal dominance (0.63) at the publisher level, consistent with Oentaryo et al.. Note: these values are author-derived estimates synthesized from the cited studies, not from a single unified experiment conducted in this study; the error bars represent estimated variability ranges rather than confidence intervals from a unified cross-validation procedure.

4.3. Privacy-Detection Trade-off Analysis

A. Data Minimization Constraints

The practical deployment of behavioral fraud detection in mobile advertising environments, including IAB contexts, is constrained by data minimization requirements imposed by platform privacy policies and regulatory frameworks. Raw IP addresses, precise timestamps, and detailed device fingerprints—the features that this study identifies as most discriminative—are simultaneously the most privacy-sensitive signals. Abadi et al. [17] established the theoretical foundations for training machine learning models under differential privacy guarantees through the DP-SGD mechanism, which introduces calibrated noise into gradient computations during model training. The applicability of this approach to mobile advertising fraud detection (including IAB deployments) depends on whether the added noise degrades the detection of subtle temporal regularities that distinguish automated from organic click patterns.

A progressive feature removal analysis reveals a non-linear relationship between privacy constraint intensity and detection degradation. Removing raw IP addresses and replacing them with anonymized network-level identifiers (e.g., /24 subnet prefix) notably reduces the precision of temporal aggregation features, as the granularity of IP-level click frequency computation is degraded. Coarsening timestamps from millisecond to minute-level resolution impacts inter-click interval features more severely, as the fine-grained timing variance that distinguishes automated from organic patterns is substantially attenuated. These observations suggest that privacy-preserving deployments must prioritize retention of temporal signal granularity, even at the cost of device-level feature resolution.

B. Privacy-Preserving Detection Considerations

Federated learning architectures offer a potential path toward cross-platform fraud intelligence sharing without centralized raw data aggregation. In the IAB context, individual apps could train local behavioral models on their own ad interaction data and contribute model updates to a shared global model. The practical challenges of this approach include the high-throughput requirements of real-time bid verification, the latency constraints that preclude multi-round synchronization protocols, and the fragmented nature of the IAB ecosystem across heterogeneous app platforms.

The privacy-detection trade-off identified in this analysis suggests that a tiered approach may be most practical for IAB environments: a first-stage filter operating on minimally invasive features (coarsened temporal aggregates and network-origin classification) to flag suspicious traffic, followed by a second-stage detailed analysis applied only to flagged sessions using higher-resolution behavioral features under appropriate privacy controls.

5. Conclusion and Future Directions

This paper has presented a systematic behavioral feature analysis for anomalous click detection in mobile advertising environments, with the goal of informing fraud detection strategies applicable to in-app browser (IAB) deployment contexts. The investigation was organized across four analytical dimensions—temporal, gestural, device fingerprinting, and network-level features—and synthesized findings from three publicly available datasets representing distinct aspects of the mobile ad fraud landscape. The TalkingData AdTracking dataset (approximately 200 million records) provided evidence for the discriminative power of temporal aggregation features in separating conversion-anomalous from normal click patterns (noting that the is_attributed label reflects conversion attribution rather than a ground-truth fraud label), the FDMA 2012 BuzzCity dataset demonstrated publisher-level fraud pattern separation through temporal regularity metrics, and results from the labeled RTB bid request logs reported in the literature revealed the complementary value of device consistency features for detecting coordinated fraud operations. Because none of the three datasets were collected specifically from IAB (WebView) instrumentation, the findings are best characterized as evidence from the broader mobile advertising domain with directional applicability to IAB contexts.

The analysis indicates that temporal features—particularly IP-level click frequency within sliding time windows and inter-click interval variance—provide the strongest individual detection signals across all datasets reviewed in this study. Device fingerprinting features, especially brand consistency and user-agent stability, serve as effective complementary indicators that are particularly valuable against coordinated device

farm operations. A meaningful privacy-detection trade-off exists: the most privacy-sensitive features (raw IP addresses, precise timestamps) contribute the largest marginal gains in detection accuracy, creating a tension that practitioners must navigate in production deployments. The proposed tiered detection architecture offers a practical framework for balancing these competing requirements.

It should be noted that this study has several important limitations that affect the scope of its conclusions. First, the publicly available datasets used in this study originate from general mobile advertising platforms and do not contain WebView-specific interaction signals (e.g., in-page scroll trajectories, touch event pressure) that would be available in production IAB systems; consequently, the claim that these features are effective in IAB environments is an extrapolation from mobile advertising evidence rather than a directly validated finding. Second, the TalkingData dataset's is_attributed label measures conversion attribution, not fraud status; using it as a proxy for fraud classification introduces label semantic ambiguity that limits the strength of conclusions drawn from that data source. Third, the empirical analysis in Section 4.2 is a structured synthesis of results reported in the existing literature and competition analyses rather than a set of original experiments conducted by the authors; readers should interpret the cited performance metrics (e.g., 97% precision, 95% recall from EvilHunter) as external benchmarks rather than results produced in this study. Fourth, the gestural feature dimension is evaluated primarily through inference from existing findings rather than direct measurement, and the specific thresholds cited (e.g., 50 ms inter-click intervals, near-zero curvature) have not been validated in IAB-specific data. Fifth, the privacy-detection trade-off analysis provides directional estimates rather than precise quantification under formal privacy models.

References

[1]. Nath, S. (2015). MAdScope: Characterizing mobile in-app targeted ads. In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '15) (pp. 59–73). ACM. https://doi.org/10.1145/2742647.2742672

[2]. Juniper Research. (2023). Online advertising fraud: Market forecasts, key vertical analysis & regulatory landscape 2023–2028. Juniper Research Ltd. https://www.juniperresearch.com/research/fintech-payments/fraud-identity/online-ad-fraud-research-report/

[3]. PPC Shield. (2025). Invalid click report: Q2 2025. PPC Shield Inc. https://www.ppcshield.com/reports/invalid-clicks-q2-2025/

[4]. Crussell, J., Stevens, R., & Chen, H. (2014). MAdFraud: Investigating ad fraud in Android applications. In Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '14) (pp. 123–134). ACM. https://doi.org/10.1145/2594368.2594391

[5]. Shao, R., Rastogi, V., Chen, Y., Pan, X., Guo, S., & Riley, R. (2016). Are these ads safe: Detecting hidden attacks through the mobile app-web interfaces. In Proceedings of the Network and Distributed System Security Symposium (NDSS '16). Internet Society.

[6]. Liu, B., Nath, S., Govindan, R., & Liu, J. (2014). DECAF: Detecting and characterizing ad fraud in mobile apps. In Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI '14) (pp. 57–70). USENIX Association.

[7]. Dave, V., Guha, S., & Zhang, Y. (2013). ViceROI: Catching click-spam in search ad networks. In Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security (CCS '13) (pp. 765–776). ACM. https://doi.org/10.1145/2508859.2516688

[8]. Pearce, P., Dave, V., Grier, C., Levchenko, K., Guha, S., McCoy, D., Paxson, V., Savage, S., & Voelker, G. M. (2014). Characterizing large-scale click fraud in ZeroAccess. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14) (pp. 141–152). ACM. https://doi.org/10.1145/2660267.2660369

[9]. Dong, F., Wang, H., Li, L., Guo, Y., Bissyandé, T. F., Liu, T., Xu, G., & Klein, J. (2018). FraudDroid: Automated ad fraud detection for Android apps. In Proceedings of the 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE '18) (pp. 257–268). ACM. https://doi.org/10.1145/3236024.3236045

[10]. Nagaraja, S., & Shah, R. (2019). Clicktok: Click fraud detection using traffic analysis. In Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks (WiSec '19) (pp. 105–116). ACM. https://doi.org/10.1145/3317549.3323407

[11]. Rizzo, C., Cavallaro, L., & Kinder, J. (2018). BabelView: Evaluating the impact of code injection attacks in mobile webviews. In Proceedings of the 21st International Symposium on Research in Attacks, Intrusions, and Defenses (RAID '18) (pp. 25–44). Springer. https://doi.org/10.1007/978-3-030-00470-5_2

[12]. Arp, D., Quiring, E., Pendlebury, F., Warnecke, A., Pierazzi, F., Wressnegger, C., Cavallaro, L., & Rieck, K. (2022). Dos and don'ts of machine learning in computer security. In Proceedings of the 31st USENIX Security Symposium (USENIX Security '22) (pp. 3971–3988). USENIX Association.

[13]. Suo, S., Bao, Y., Liu, Q., Wang, Z., Wang, C., & Xu, K. (2021). Understanding and detecting mobile ad fraud through the lens of invalid traffic. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21) (pp. 287–301). ACM. https://doi.org/10.1145/3460120.3484547

[14]. Oentaryo, R., Lim, E., Finegold, M., Lo, D., Zhu, F., Phua, C., Cheu, E., Yap, G., Sim, K., Nguyen, M. N., Perera, K., Neupane, B., Faisal, M., Aung, Z., Woon, W. L., Chen, W., Patel, D., & Berrar, D. (2014). Detecting click fraud in online advertising: A data mining approach. Journal of Machine Learning Research, 15(1), 99–140.

[15]. Tian, T., Zhu, J., Xia, F., Zhuang, X., & Zhang, T. (2015). Crowd fraud detection in internet advertising. In Proceedings of the 24th International Conference on World Wide Web (WWW '15) (pp. 1100–1110). ACM. https://doi.org/10.1145/2736277.2741136

[16]. Han, X., Chen, Y., Wang, Z., Liu, Y., & Zhang, Y. (2023). Exploring security hazards of in-app QR code scanning in mobile applications. In Proceedings of the 32nd USENIX Security Symposium (USENIX Security '23) (pp. 4521–4538). USENIX Association.

[17]. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16) (pp. 308–318). ACM. https://doi.org/10.1145/2976749.2978318