Mobvista

TABLE OF CONTENTS

Foreword

The global mobile advertising and monetization market is developing at a rapid pace. According to data from App Annie, the number of mobile app downloads reached 204 billion in 2019 and is expected to hit 258 billion by 2022. Worldwide mobile ad spending is also growing every year: data from market research company eMarketer shows that mobile ad spending exceeded $220 billion in 2019 and is predicted to soar to $352.1 billion by 2022.

Where there is profit, there is potential for crime. The increasingly higher levels of mobile ad fraud have a deep impact on the entire advertising industry: according to WFA's conservative estimations, advertisers have lost over $20 billion in 2019 due to mobile ad fraud and that figure is predicted to skyrocket to $50 billion by 2025.

Mobile ad fraud has caused massive economic losses to advertisers, has affected the ad platforms' delivery quality, the reputation of advertising platforms, and it has even damaged the long-term development of the entire mobile advertising market.

As the global leading technology platform - Mobvista, which has established an ad network covering over 200 countries and more than 10 billion daily impressions, is at the forefront of the fight against mobile ad fraud.

To clean up the online advertising market and promote the healthy development of the mobile ad industry, Mobvista released the Mobile Ad Anti-Fraud White Paper 2.0. This white paper aims to clarify the current state of mobile ad fraud, fraud types, and anti-fraud strategies in the mobile marketing ecosystem, as well as to improve transparency and foster an environment dedicated to the sustainable growth for the entire mobile ad industry.

What Is Mobile Ad Fraud?

Mobile ad fraud is the attempt to defraud advertisers, publishers, or supply partners by exploiting mobile advertising technology. The main objective of fraudsters is to steal advertising budgets.

There are several types of mobile ad fraud, such as Click Spamming, Click Injection, SDK Spoofing, and Bots. In this chapter, we will take an in-depth look at these common fraud types.

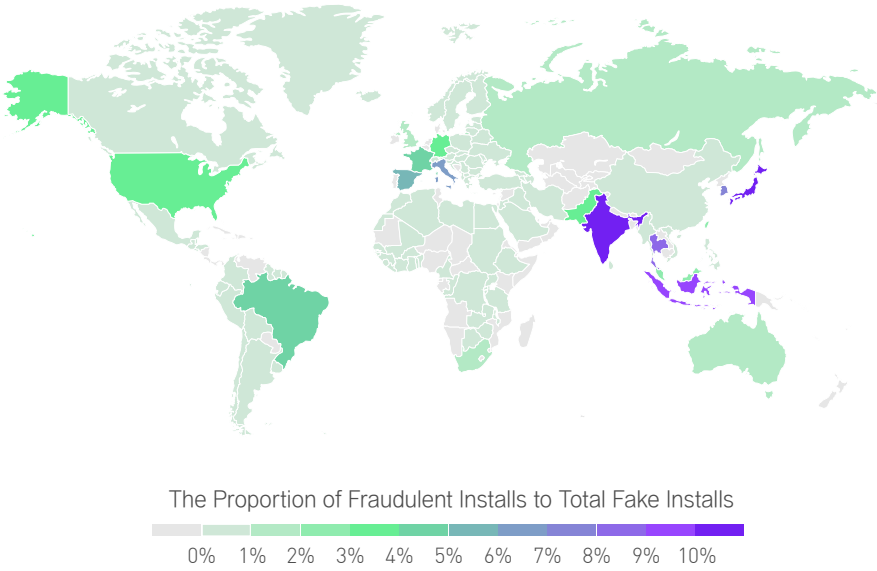
The Status of Mobile Ad Fraud

Based on the 10 billion daily impressions data and other proprietary data generated from a traffic monitoring system that consists of Mintegral (Mobvista's programmatic advertising platform) and Nativex (Mobvista's mobile marketing platform), covering over 200 countries and regions, Mobvista's anti-fraud technology team carried out an in-depth analysis of the total number of fraudulent installs globally between October 2019 and March 2020. The result reveals that the fraudulent traffic monitored and preemptively intercepted by Mintegral and Nativex accounts for 2.01% and 10.69% respectively of the total traffic recorded during that period. The distribution of fraudulent installs across regions, operating systems, and app categories shows the following characteristics:

Fraud Install Rate by Country

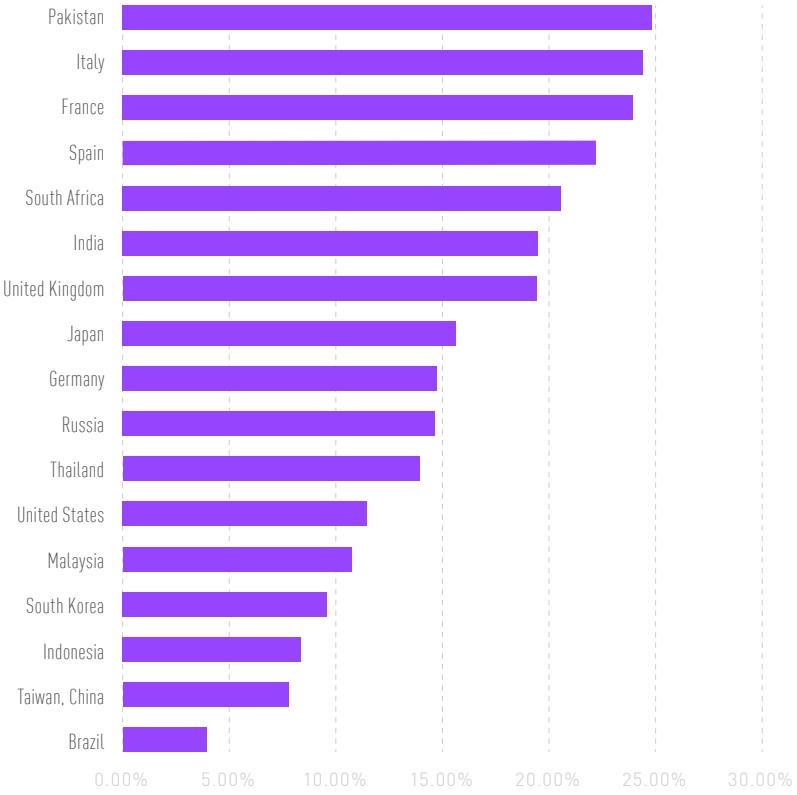
According to the traffic system created by Mintegral and Nativex, which covers more than 200 countries and regions, most of the fraud intercepted came from countries like Japan (12.81%), India (11.6%), and Indonesia (9.4%). (Note: Since the traffic of the above countries or regions makes up a relatively large percentage of Nativex's traffic system, the proportion of fraudulent installs in those or regions to total fraud installs is inevitably high. Therefore, the data provided here is of a limited reference value.)

Fraud Install Rate Countries and regions



To understand the severity of mobile ad fraud more clearly, we looked at the percentage of fraudulent installs in all countries or regions. Data shows that Pakistan (24.87%), Italy (24.47%), and France (23.97%) are at high risk.

The Percentage of Fraudulent Installs in All Countries or Regions

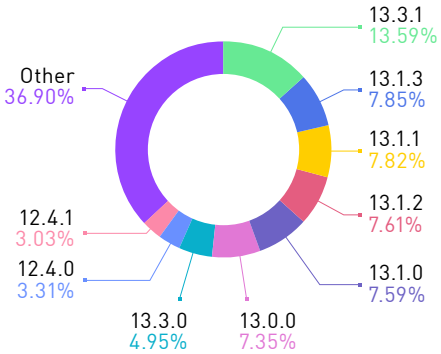


Fraud Install Rate by Operating System

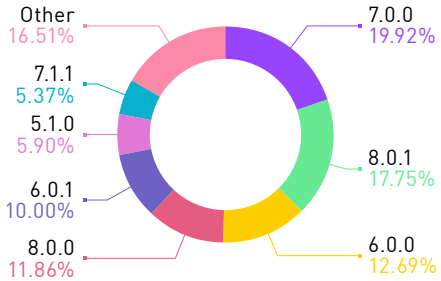
From an operating system perspective, the iOS platform accounts for 40.2% of fraudulent installs, while Android makes up for the remaining 59.8%.

With regards to the iOS platform, fraudulent installs are mainly happening on these OS versions: 13.3.1 (13.59%), 13.1.3 (7.85%), and 13.1.1 (7.82%). On the Android platform, fraudulent installs are mainly happening on the following OS versions: 7.0.0 (19.92%), 8.1.0 (17.75%), and 6.0.0 (12.69%).

The Proportion of Fraudulent Installs on Each iOS Version to the Total Fraudulent Installs on the iOS Platform

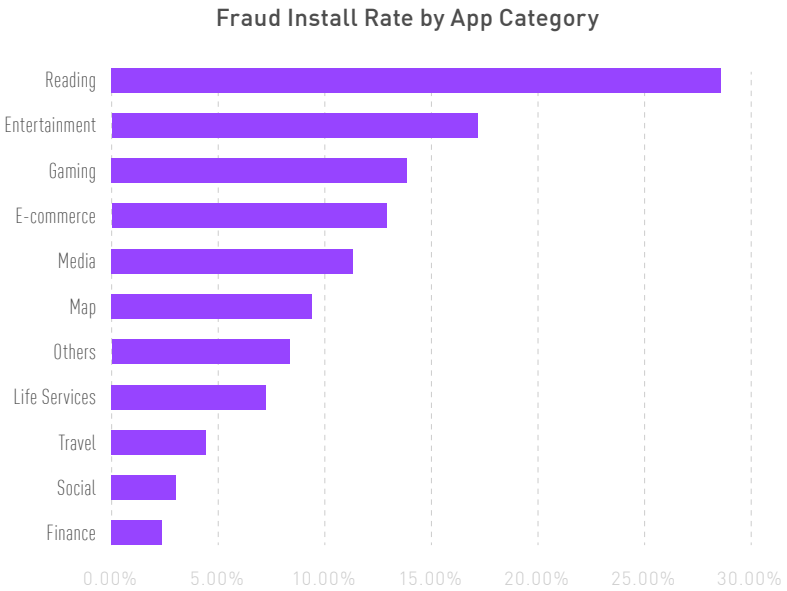


The Proportion of Fraudulent Installs on Each Android Version to the Total Fraudulent Installs on the Android Platform



Fraud Install Rate by App Category

Looking at stats on various app categories such as games, e-commerce, and social, we have noticed that the fraudulent install percentages are not the same. The top 3 highest categories in terms of fraudulent installs are reading apps (28.60%), entertainment apps (17.17%), and games (13.86%).

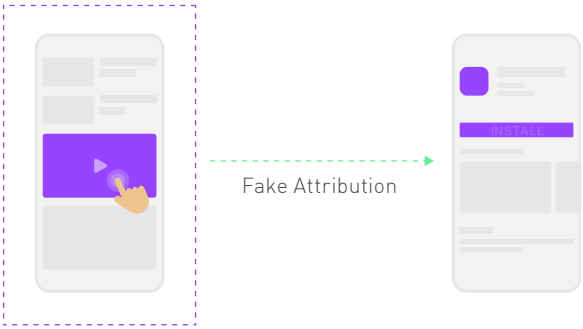


Common Fraud Types

In terms of traffic authenticity, popular fraud methods can be divided into three categories: attribution fraud, fake traffic, and illegal traffic.

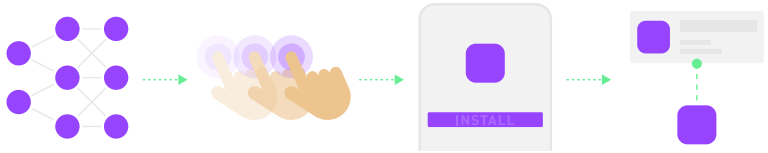
Attribution Fraud

Attribution fraud is where fraudsters attempt to steal the credit for organic app installations, not generated by them, by reporting fake impressions or fake clicks in an attempt to deliver the last engagement before the app is first launched by the user.



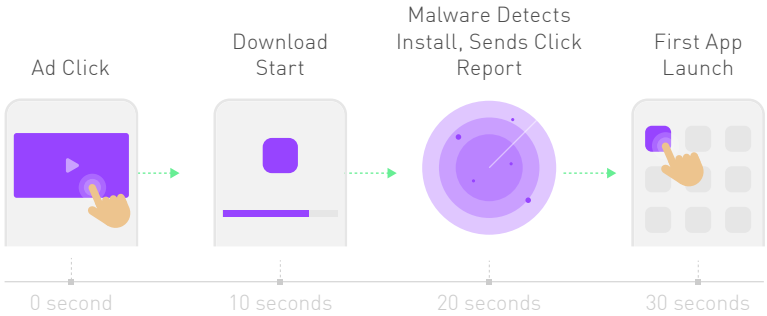
A. Click Spamming/Click Stuffing/Click Flood

Click Spamming, also called Click Stuffing, or Click Flood, is a case in which fraudsters send a huge number of fraudulent clicks in hopes of taking credit for organic app installations. This type of fraudulent traffic is characterized by low CVR (Conversion Rate) and abnormal distribution of CTIT (Click-to-Install Time).



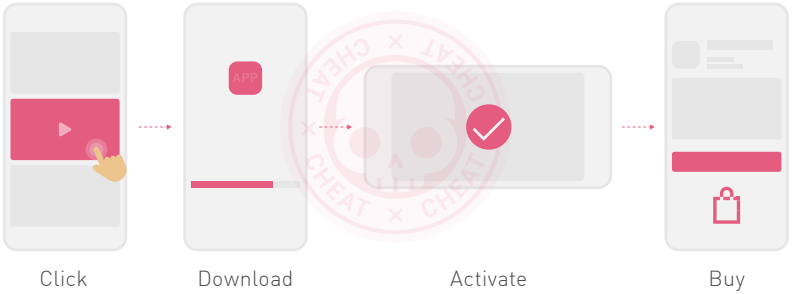
B. Click Injection/Install Hijacking

Click Injection, also known as Install Hijacking, is a sophisticated form of click-spamming. Click injection will use an app located on the user's device which "listens" app installation broadcasts. The fraudsters will be informed when new apps are installed on the device, triggering a click before the installation is complete, taking credit for the install. One key characteristic of this type of fraud is that the CTIT (Click-to-Install Time) is notably short, and the download time recorded by app stores is earlier than the click time.



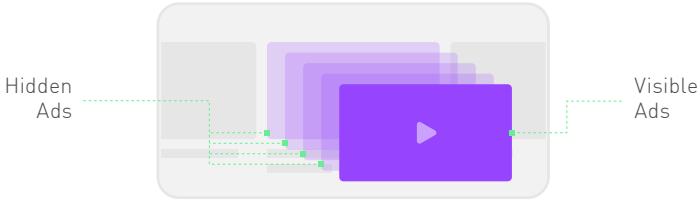
Fake Traffic

Fake traffic means that clicks, installs, and any in-app behavior exhibited during the advertising campaign are all invalid. As a result, advertisers will spend money out of their budgets but will fail to acquire real users.



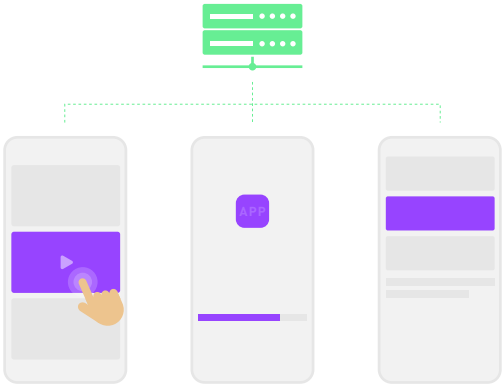
A. Ad Stacking

Ad stacking is a type of display and impression fraud where multiple ads are layered on top of each other in same ad placement. While only the top ad is visible, if a user clicks on the visible ad, fraudsters send multiple clicks of different ads in the background. A key characteristic of this type of fraud is that many ads are clicked on the same device within a very short amount of time.



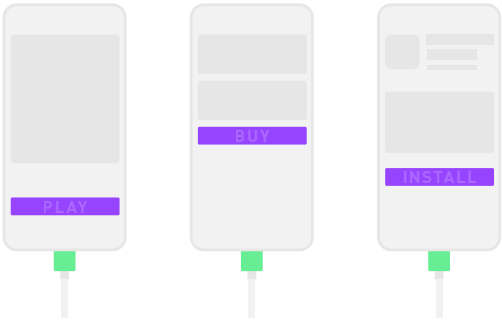
B. Bots

Bots are fake users created by fraudsters using automated scripts or software in order to simulate specific tasks, such as ad clicks, installs, and even in-app engagements, masquerading as legitimate users. This fraud type is characterized by low IP repetition rate, high new device rate, abnormal user behaviors, and unusual distribution of device models/operating systems.



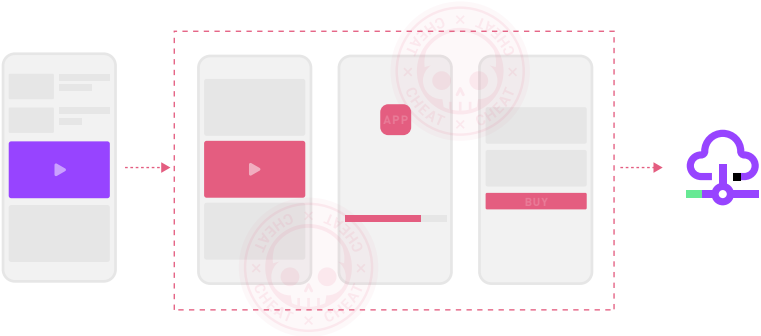
C. Device Farms

Device Farms refers to a fraudulent practice of purchasing a huge number of real devices for clicks, downloads, installs, or other various in-app behaviors, and concealing device information by modifying the devices' advertising identifiers. This type of fraud is characterized by dense IP dispersion, extremely high new device rate, abnormal user behaviors, and unusual distribution of models/ operating systems.



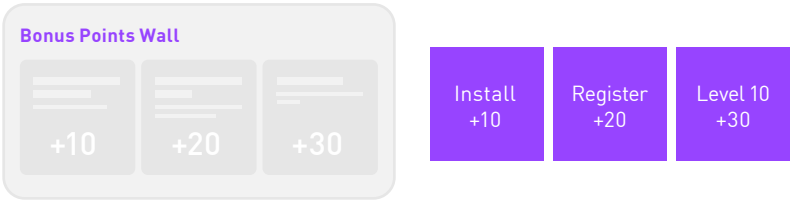
D. SDK Spoofing

SDK spoofing refers to generating legitimate-looking installs with data from real devices without any actual install happening. Fraudsters utilize a real device to create installs that look real to deplete an advertiser's budget. This type of fraud is characterized by the SDK version and app version of installs coming through not matching your latest version.



Illegal Traffic

Illegal traffic happens when publishers use controversial methods to acquire users without the advertisers' consent, including downloading apps from unofficial app stores, incentivized traffic, using prohibited ad materials, pay-per-click/install/event scams or programs, deceptive clicks and downloads, background trojan installations, and more.



How to Fight Against Mobile Ad Fraud

An Introduction to the Anti Mobile Ad Fraud Industry

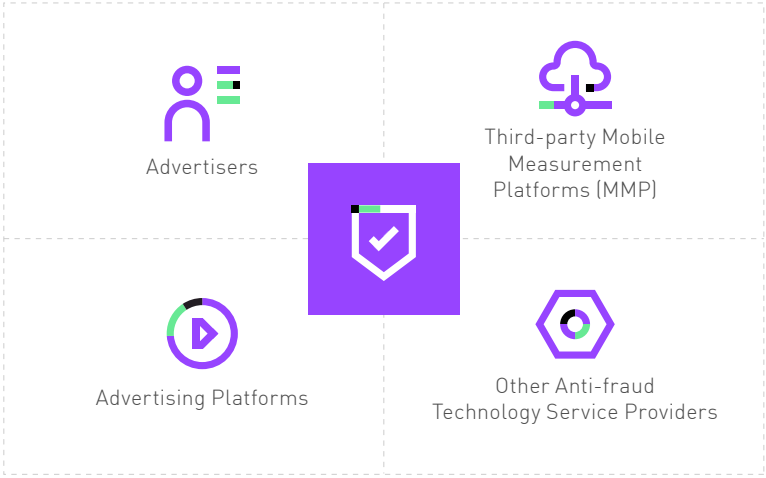
Key roles in the mobile anti-fraud industry:

a. Advertisers

b. Third-party detection platforms, including Adjust, AppsFlyer, TUNE, Kochava, and TalkingData, etc.

c. Advertising platforms

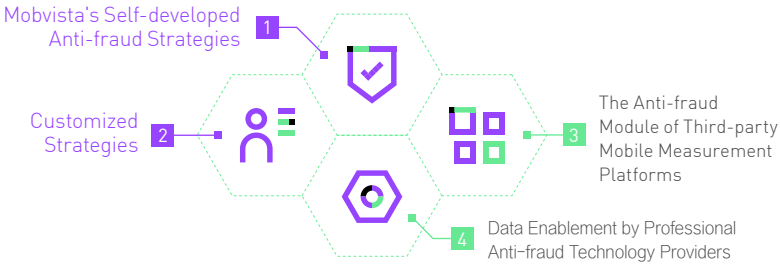
d. Other anti-fraud technology service providers, such as Machine, mFilterIt, as well as FraudScore and DataVisor.



Mobvista's Anti-Fraud Technologies

Introduction

At Mobvista, we have dedicated a considerable amount of research and manpower to our anti-fraud efforts. Our anti-fraud strategy combines the technologies used by mobile measurement platforms and other anti-fraud tech service providers, allowing us to deliver a multi-tiered anti-fraud solution.



Mobvista's anti-fraud suite is composed of our in-house anti-fraud solutions, customized strategies, using modules from third party anti-fraud suites and partnerships with other anti-fraud firms.

Through custom-built services, Mobvista developed anti-fraud strategies based on advertisers' specific needs. We can fight fraud using this technique as it is possible to look at key metrics, such as retention rate, based on the events data send back by advertisers.

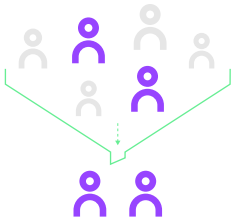
With the 3rd party anti-fraud modular approach, we are able to utilize features such as IP filtering from the IP blacklist provided by Adjust (and other modular technologies from other 3rd parties).

Additionally, Mobvista is currently working together with some anti-fraud companies, such as Fraudlogix from the U.S. and IP2Location from Malaysia,to fight aginst fraudulent ad activities.

Mobvista's in-house anti-fraud solutions are at the core of its anti-fraud system. To detect and prevent fraudulent traffic, the team has created a series of anti-fraud rules and models.

Rule-Based Anti-Fraud Solutions

So far, Mobvista's anti-fraud technology team has collected more than 60 key fraud features and focused on the industry's cutting-edge anti-fraud technologies. We have set up a comprehensive anti-fraud rule system and you can find out more details about some of these rules below.



IP-related

A. Large Number of Clicks/Installs from the Same IP Address

Under normal circumstances, due to the diverse distribution of users, the distribution of their devices' IP addresses should also be distinct. If a large number of clicks or installs originates from the same IP addresses or subnet, that may be fraud. (Please note: as the sizes of operators' IP address pools and their IP address allocation policies are different, and the possibility that multiple devices in the same LAN may be connected to the public network at the same time, as long as clicks or installs are not excessively concentrated on a few IP addresses, a certain degree of IP address duplication is normal.)

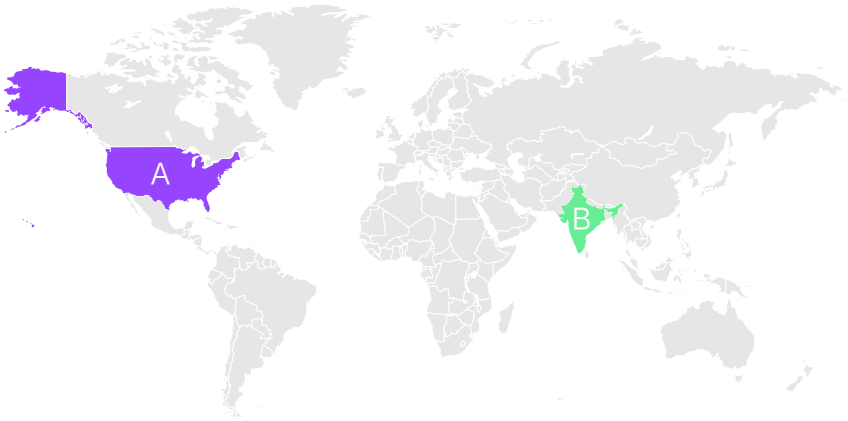
B. IP Blacklist

The IP blacklist mainly comes from third-party anti-fraud service providers and data accumulated by Mobvista over a long period of time. Generally speaking, the blacklist contains IP addresses that are impossible to be used by mobile devices (e.g. data centers), or highly risky IP addresses that have been used excessively for fraud.

C. IP mismatches between click and install

In most cases, the click and install IP addresses should be the same. If the inconsistent ratio between the click IPs and install IPs of a channel is too high, then that channel poses a significant risk of ad fraud.

Geography-related



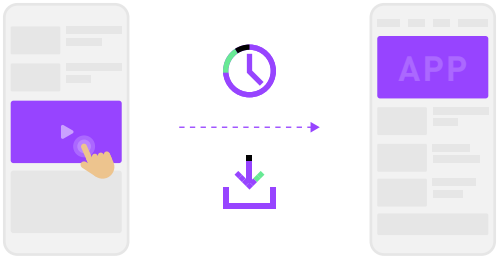
A. Country mismatches between clicks and installs

If, for example, ads are launched in the U.S., but the install IP addresses are from India, this is something that will rarely occur legitimately in reality, so if it happens on a large scale then it is very likely that these installs are fraudulent.

B. City mismatches between clicks and installs

Let's say, for example, that the IP addresses of devices used to click on ads are from Beijing, but the apps are installed on devices with Guangzhou IP addresses. This is very unlikely to happen legitimately, so if you notice this happening on a large scale, it is very likely that these installs are fake.

CTIT (Click to Install Time)-related



A. CTIT Too Short

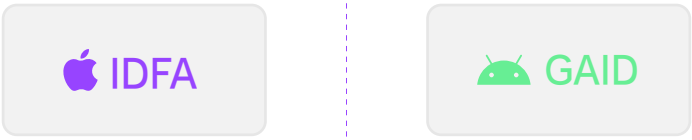
It usually takes a bit of time for users to click on an ad, jump to the app store's download page, download the app, and finally open the app for the first time. If the CTIT (Click to install time) is too short, then the install is likely a fraudulent install. How much time is considered "too short" depends on the local network speed, app package size, and mobile device performance.

B. Abnormal CTIT Distribution

When it comes to a single install, it may be possible that the time from clicking an ad to completing the installation process is extremely long, such as longer than 24 hours. However, if such a scenario occurs on a large scale, then it is highly possible that the traffic in question is fraudulent.

Device ID

Use to track the effectiveness of their advertising campaigns. iOS devices usually use IDFA, while Android devices tend to use GAID. As there are enough Device IDs available at any given time, the risk of clashes is extremely small, so it's usually safe to consider Device ID as a unique identifier.



A. Extremely High New Device Rate

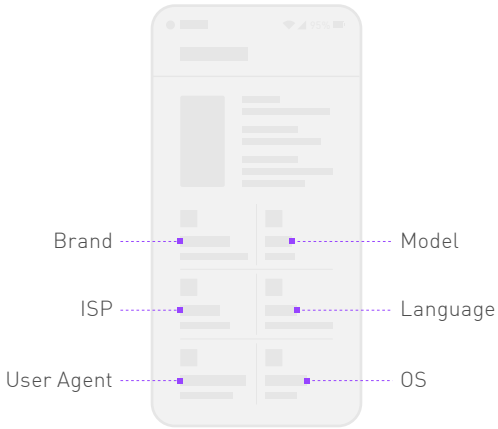
Device ID can be manually modified for privacy purposes, but most users will not actively modify it. This means that brand-new Device IDs appear either on devices that have not previously installed apps through ads or on fraud devices that attempt to install apps by frequently changing their Device ID. Therefore, if the percentage of new gadgets coming from a channel is extremely high, this is very likely to be fraud.

B. Device Blacklist

Based on the large amount of historical data, Mobvista will blacklist some Device IDs with a history of frequent fraudulent behavior. If an app is installed on a device whose Device ID is on the Device Blacklist, it's very likely that the installation is fraudulent.

Distribution of Device Parameters

The specific parameters of a mobile device such as brand, model, language, operator, user agent and operating system must be consistent with the actual local area. For example, if the brands and models of most devices used for app installs on a special channel are commonly found in the local area, then that is likely to be the result of ad fraud.



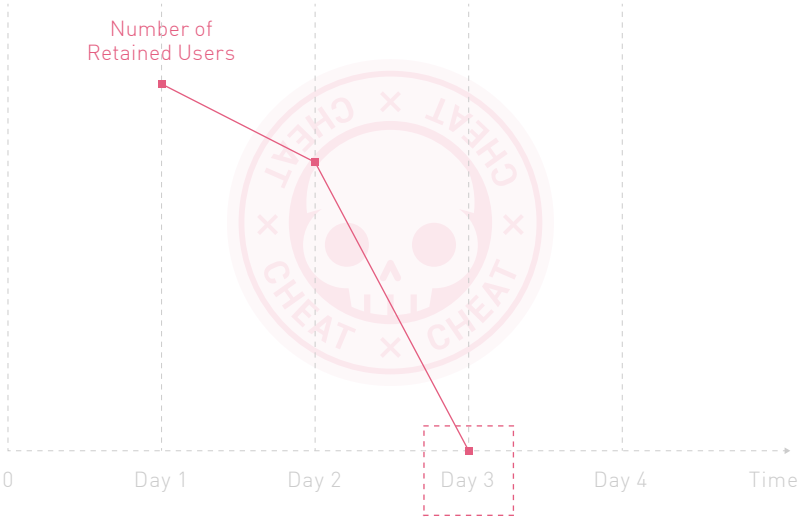
User Behavior

A. Extremely Low/High CVR

The CVR of normal channels is usually within a certain range, so pay special attention to the channel with particularly high or low CVR. For instance, if the CVR is too high, you need to analyze and determine whether that traffic is illegal or not.

B. Significant Decline in the Retention Curve

This refers to data on users brought by a certain channel that always falls sharply at a certain point in time in the game's/app's lifecycle. For example, user retention is normal the day after the app is installed, but it plummets to 0 on the third day.



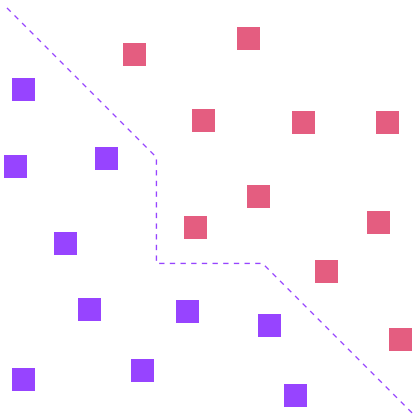
How to Handle Mixed Traffic

It's important to note that more and more fraudulent traffic is not the result of a single type of fraud, but rather a mix of different fraud types. For example, mixing the click spamming and bots traffic can not only improve a channel's CVR to a certain extent, but can also 'fix' the CTIT issue. This has made anti-fraud detection significantly more difficult.

To solve the problem caused by mixed traffic, it's crucial to increase statistical dimensions when we look at data distribution and channel indicators, such as CVR and CTIT. For example, we can calculate the CVR of each channel instead of calculating the overall CVR. And, we can calculate the CVR of each channel in each country, instead of only the CVR of each channel. This is increasing statistical dimension.

In addition to traditional dimensions such as channel and platform, other dimensions that can be added, including: device (e.g. brand, model, operating system and operator), geo (including city, IP subnet, province, state, and country), and time (e.g. click-to-install time).

If fraudsters mix click spamming traffic with bot traffic to make the original traffic indicators complement one another, this will make mobile ad fraud detection significantly more challenging. For example, click spamming traffic mainly occurs on Samsung mobile phones, while bots traffic mostly happens on Huawei mobile phones. By adding the mobile device brand statistical dimension to the equation, we can split up the original traffic indicators and see the real situation.

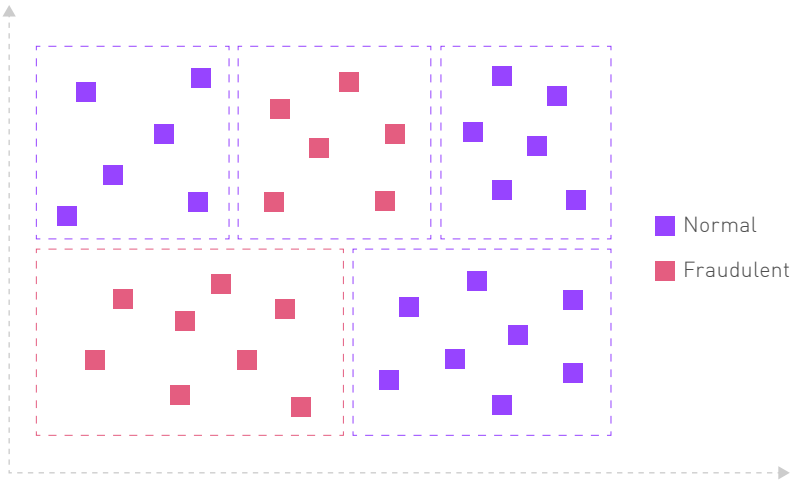


ML-Based Anti-Fraud Solutions

It has always been standard practice in the online advertising industry to prevent specific fraudulent traffic by following a series of anti-fraud rules. Rule-based anti-fraud solutions are quite effective but they cannot adapt to new fraud models. As fraud has become much more complex and sophisticated, one of the best approaches is to use intelligent technologies such as machine learning. Machine Learning (ML) models can use big data to automatically discover and learn potential or new fraud patterns.

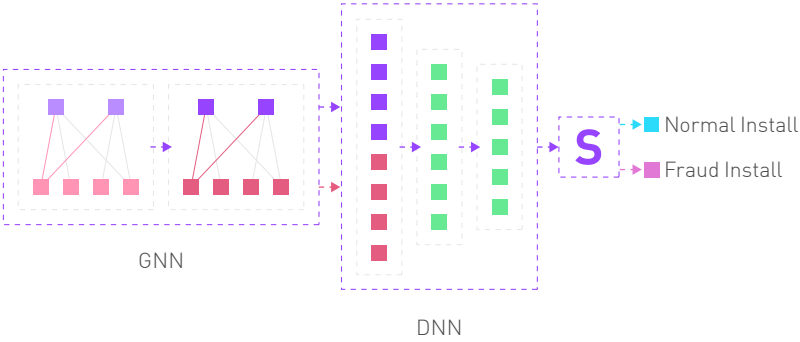
Unsupervised Machine Learning Anti-Fraud Models

Unsupervised machine learning models do not need to label data in advance, and those models can automatically detect potential or new fraud models by exploring and grouping the data available. Commonly used unsupervised machine learning includes clustering and anomaly detection. For example, we use DPC (Density Peak Clustering) for clustering, the human experts will perform further analysis based on the clustering results, to determine whether the corresponding traffic is fraudulent or not.



Supervised Machine Learning Anti-Fraud Models

With the labeled data, we also can train some supervised machine learning anti-fraud models. Commonly used supervised machine learning anti-fraud models are Random Forest, GBDT and Graph Neural Networks (GNN). For instance, we train a GNN-based model to detect bots install fraud with an accuracy up to 92%.



Conclusion

In this white paper, we have introduced several common mobile ad fraud types and we have talked about Mobvista's anti-fraud solutions. The Group Vice President Ramon Zhu claimed that, the battle between fraud and anti-fraud is an ongoing battle that will last for a long time. With fraud models and technologies constantly improving, anti-fraud technologies also have to keep up. It's essential for all parties involved, including advertisers, media, ad agencies, third parties, and other key industry players, to come together and fight against fraud traffic in order to keep the industry growing, become more transparent and maintain a standard that all parties can comply with .

About Mobvista

Mobvista is a leading technology platform dedicated to driving global business growth in the digital age.

With global technology and rich industry experience, Mobvista helps customers utilize advanced technologies such as big data, artificial intelligence, and elastic cloud computing cluster management to connect China and the rest of the world, helping customers build forward-looking business models and guaranteeing effective market access for all.

Mobvista was founded in Guangzhou, China, in 2013, and listed on the Main Board of the Stock Exchange of Hong Kong (01860.HK) since December 2018, hitherto has over 700 employees with offices in 16 cities across the world.

Mobvista

www.mobvista.com