International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 18 (2018) pp. 13812-13816

A Survey on Online Click Fraud Execution and Analysis

Dr.R.Kayalvizhi[1], Kapil Khattar[2], Piya Mishra[3]

Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India

Abstract

With the drastic increment of usage of web, online promotional material plays a vital role in the department of online advertising. One method widely used for revenue generation in this domain involves charging for every click and supports the recognition of keywords and therefore the variety of competitors in advertising. Pay-Per-Click (PPC) model gives space or allows people or competitors to obtain revenue with duplicate or fake clicks (i.e., click fraud), that give rise to serious issues in the field of online advertising. To understand fraud, a vital issue is to detect these fake clicks over these window models, like decaying windows, jumping windows or slippy windows. Decaying window models are often terribly useful in shaping and determinant click fraud. However, though there are several algorithms detect click fraud in PPC model over decaying window models. PPC ad networks charge advertisers for each click on their ads. Click-fraud happens once a user or an automatic software package clicks on a commercial with a malicious intent and advertisers got to get those worthless clicks. Click-fraud has been tried to be a heavy drawback for the web promotional material trade. Though it's attracted a lot of attention from the community of safety, the immediate victims of click-fraud are found to be the advertisers and they still lack confidence within the click fraud detection techniques. Among several styles of click-fraud, botnets with the machine-controlled clickers are the foremost severe ones.

Keywords: Click fraud, Pay-Per-Click, Inorganic traffic, Fraud execution, fraud analysis.

INTRODUCTION

Click-fraud has been proved to be a serious issue for the online advertisement industry. Although it has caught a lot of attention from the immediate sufferers and victims of click-fraud, the advertisers, still lack confidence in the click-fraud detection techniques.

Among many forms of click-fraud, botnets with the automated clickers are the most severe ones Normal internet users are victimized by malicious attackers (e.g., bot-master of a botnet) and the attackers infect and use their machines to defraud advertisers. Since most modern operating systems already provide some kind of anti-malware service.

Advertisers and publishers use a wide range of revenue models. Among all the models, Pay-Per-Click (PPC) dominates and overpowers the others. In PPC, advertisers pay each time a user clicks on an ad. The biggest threat to the PPC advertisement is click fraud. Click-fraud is the practice of fraudulently clicking on online applications or websites or web ads with the motive of increasing revenues of third party websites or exhaustion of the advertiser's budget. Ad networks mostly use server-based techniques to detect advertising frauds as they do not have enough control over the client machines. They gather information from different sources about the user, the machine and the behaviour of the user. Then, they apply machine learning or pattern recognition techniques to identify suspicious clicks.

Basically mobile app fraud refers to the illegal activities performed by various ad campaigns for revenue generation which is illegitimate and is used to cause trouble for the advertisers directly leading to a loss in revenue and eliminating competition by false methods like:

- Device emulation using BlueStacks
- Malicious ad clicks for various web advertising campaigns.
- Fraudulent apps and background ad clicks using script.
- Illegal versions to hinder revenue generation for publisher.

In this paper we perform fraud analysis by converting the database, provided in CSV format as per the algorithms and then check and observe patterns so as to be able to execute the fraud. Patterns observed are Click Time, Convert Time, Publisher name, Sub Publisher name, Session time etc.

Google has developed the most stable program for the detection of click fraud which uses three approaches. Their detection system begins by automated filtering and then advanced algorithms are executed for detection and filtering out inorganic or fraudulent traffic in real time, which in turn saves cost for advertisers before they are charged money. But we cannot completely rely on these filters to catch all inorganic clicks, so to get better results, the Ad traffic quality team of Google performs manual and offline analysis for elimination of clicks that they consider as invalid before the advertisers are charged. Apart from all these, investigations are launched by Google based on advertiser's report of suspicious activities.

Whenever any fraudulent or malicious clicks are found, they are considered as invalid and credits are issued in the account.

A. UNDERSTANDING CLICK FRAUD

For understanding the click frauds, it is very important to know certain keywords. In this reference, the valuable thing is using the information, purchasing of a product, employment of service, and executing a transaction executed via a website's visitor which is the same throughout the lifetime of content provider.

A sponsored link usually will provide information like title, date, text, and other data in a sponsored result file. Whenever a query is exist in the result page of Search Engine (SERP) gives response as a URL Serviced by a search engine is called a sponsored link.

The act of initiating or beginning to visit to a website via a paid link is referred to as a click and can be of two types:

- **Invalid clicks:** a click generated by a bot or a script on a sponsored or paid link with zero or no chances of value generation.
- **Valid clicks:** this is an click made on purpose or an intentional click that has good probability to generate value whenever a visitor browses through a website or webpage.

Invalid clicks are malicious or fraudulent. These clicks are intentionally generated on clicking a paid link with zero intention of value generation. Malicious clicks that are distinguishable from valid clicks by a pattern are called Identifiable click fraud whereas unidentifiable click frauds are a pattern of all the fraudulent clicks which can't be differentiated from valid clicks.

A malicious click is an illegitimate click which is neither fraudulent nor malicious that is, a multiple click on a sponsored link whenever a website's server is down and it is only identifiable if it can be differentiated from a set of valid clicks.

B. ORGANIC TRAFFIC

Organic traffic refers to the unpaid traffic. It belongs and includes the traffic obtained from:

- Google, Bing and Yahoo and other such search engines are major contributors of Organic Traffic.
- Twitter and Facebook are some of the social media sites showcasing organic traffic.

Organic traffic is also observed when visitors arrive at a particular site via referrals from other sites. This is done by clicking on a referral link on another website.

1. Directly typing in or clicking on a URL in a browser is another form of Organic Traffic.
2. This type of traffic is basically obtained organically, i.e., through a genuine source or advertising or via some source of paid promotions.

C. PAID TRAFFIC OR ORGANIC TRAFFIC

Paid traffic refers to when a party or host is spending money so as to attract visitors to their website. This includes:

1. Paid listings and advertising on various search engines.
2. Payments are made to enhance and promote the content or the link available on social media websites to promote among a larger audience.
3. Displaying web advertisements of various websites across the World Wide Web using Google Ad words.

D. CLICK FRAUD

Click fraud is a fraudulent method that is quite prevalent. It occurs on mostly on online advertisements with intention on Pay-per Click revenue generation. Such advertising is when the publisher or owner of the website posts advertisements or organises ad campaigns and are paid certain sum of money on the basis of the number of visitors visiting the site click on the organisations advertisement. In such situations fraud takes place when a system, a script or a person performs fraud or illegitimate clicks just to achieve certain targets which in turn result in generation of revenue. This is done by intentionally clicking on the target advertisement with little or no interest in the contents of the link.

E. PAY PER CLICK (PPC)

Advertising on Search engine is organic and one of the most prevalent forms of Pay per Click revenue generation. It permits the advertisers to place a bid for the placement of their advertisement on the sponsored links of the search engine whenever someone searches a certain keyword. **PPC** is the abbreviation for *pay-per-click*, a concept of the web market where each advertiser pays a fee. Each click on their advertisement thereafter generates revenue. It is basically a methodology to generate revenue or a way of buying the number of visits on their ad or site.

F. GOOGLE ADWORDS

Google AdWords is the most widely and popularly used example of PPC advertising system. It is a platform which enables several businesses to create web advertisements that appear on search engines like Google. AdWords operate as per the phenomenon of the pay-per-click model, wherein the users or advertisers bid on large number of keywords related to their advertisement and then for each visitor they get paid for each click on their ads.

For each search made in Google, Google jumps into a pool of already existing AdWords for advertisers and then comprehends a set of winners based on several factors. Based on the relevance of their keywords, the amount of their bid and popularity of the keyword the winners ad appears on a space allotted in the results page.

RELATED WORKS

The most renowned and prevalent model for advertising online is based on the concept of charging for every click [1] on the basis of the number of competing advertisers and how popular the keyword is. This mechanism of charging on the PPC basis gives room to individuals or organizations to generate and give rise to false clicks also known as click fraud. These clicks are a serious threat to the world of healthy and good web advertising. The most vital issue to detect such frauds is by identifying the duplicate clicks made over the decaying window models, like sliding windows or jumping windows. Although, there exists a large number of algorithms useful to detect the duplicate clicks but there is still an absence of effective solutions to detect those duplicate clicks over the windows. This paper [1], discusses the issue of detecting pay-per-click and duplicate clicks over decaying window models has been addressed and proposes two algorithms namely Group Blooms Filter (GBF) algorithm and Timing Blooms Filter (TBF) algorithm that require simpler operations and less storage space.

Advertising is an essential element ensuring the success of majority of businesses today. The internet is so fast growing and developing that it is a piece of cake to come up with a complete business model via online advertising which is required for the understanding and promotion of such models and to ensure the continuity of the Internet [4]. However, quite often it is seen that large number of content developers and publishers opt for dishonest means to adapt to the automated machinery or methodology or equipments to obtain high traffic and profit by defrauding the advertisers. Similarly, it has been seen that most advertising agencies use such mechanism to generate click on their competitors so as to exhaust the budget and target of rival companies. This paper prevents the automated click frauds using detection techniques that are executed post the fraud execution and propose a mechanism to prevent such click frauds using a new concept of clickable CAPTCHAs.

Malware adaptation by means of false promotion activity is profoundly unmistakable [8]. It is incredibly lucrative and by and large straightforward. What's more, there is a more elevated amount of malware that adjusts by this instrument today than whenever in late memory beforehand. Despite the fact that this kind of assaults are not new, the current and current techniques that offer ascent to malware-produced promotion movement appear to be natural, while remaining basically undetected are for the most part novel. Web publicizing or the online universe of promoting is muddled and by and large misconstrued. It is ineffectively regulated yet very much supported in the meantime. There is next to zero straightforwardness at all and there emerges an irreconcilable circumstance. This paper gives a short overview of this belief system, plots a case of examination of current snap extortion malware. This paper, additionally exhibits another way to deal with handle and battle the kind of malware adaptation. The level of malware taking part in a few or the other type of snap extortion is exceedingly critical. Information and measurements demonstrate that this issue is a long way from contained, particularly in the versatile world. There is a prompt prerequisite to comprehend the security group of malware that adapts by means of snap misrepresentation.

The level of malware partaking in some sort of snap distortion is colossal [7]. Data demonstrate this issue is far from contained, especially in the versatile space. There is a prerequisite for perception in the security gathering of mal item that adjusts by methods for snap blackmail. There are opportunities to join unique educational lists today to manufacture neutralizing activity structures for malware-created click blackmail. This is basic to clients and furthermore the entire online exchange organic framework. Since it is a gigantic vector by which gangsters can profit, it is a key locale to address for a broad advanced bad behavior interference method.

The article [3] investigates the thickness, moderateness and concentricity of pursuit promoting crowd sourcing click misrepresentation. Base on this, it advances the snap misrepresentation location show in view of grouping investigation. The model incorporates three stages: preprocessing, amass distinguishing and post-handling. In the preprocessing step, the question that is more averse to be deceitfully clicked is evacuated. In the gathering location step, a crowd sourcing click extortion assemble is identical to a group. DP-Means bunching strategy is utilized to distinguish malignant gatherings. In the post-handling step, request clicks checked by botch are separated. The meeting, adaptability and exactness are checked by the recreation information and one week click information of a web crawler organization.

Among the fluctuated sorts of malware, botnets are rising as the most genuine risk against digital security as they give a dispersed stage to numerous stole exercises like propelling conveyed refusal of administration assaults against fundamental targets, malware spread, phishing, and tap on misrepresentation. The characterizing normal for botnets is that the utilization of charge and control channels through that they'll be refreshed and coordinated. As of late, botnet recognition has been a critical explore subject related with digital risk and digital wrongdoing counteractive action. This paper [9] could be a review of botnet and botnet identification. The review illuminates botnet improvement and talks about botnet recognition strategies. This data classifies the botnets identity system in main 4 categories:

1. Signature based
2. Inconsistency based
3. DNS-based
4. Mining based

This briefs the botnet detection techniques for each category and gives us knowledge of botnet identification methods.

Click fraud represents a heavy drain on advertising budgets and may seriously hurt the viability of the net advertising market. The paper [6] proposes a completely unique framework for prediction of click fraud in mobile advertising that consists of feature choice mistreatment algorithmic Feature Elimination (RFE) and classification through Hellinger Distance call Tree (HDDT). RFE is chosen

for the feature choice because it has given better results as compared to wrapper approach once evaluated using completely different classifiers. HDDT is additionally designated as classifier to deal with category imbalance issue gift within the information set. The efficiency of planned framework is investigated on the info set provided by Buzzcity and compared with J48, RepTreelog it boost, and random forest. Results show that accuracy achieved by planned framework is 64.07% which is best as compared to existing strategies.

Promoting misrepresentation, strikingly click extortion, is a developing worry for the web publicizing business. The utilization of snap bots, malware that mechanically taps on advertisements to produce beguiling activity, has enduring upgraded in the course of the last a long time. Though the safety Business has focused on analyst work and expelling pernicious parallels identified with click bots, a superior comprehension of however fraudsters work inside the promotion conspire is should have been ready to disturb it speedily. This paper [2] gives a top to bottom analysis of the publicizing misrepresentation topic utilized by Boaxxe, a malware represent considerable authority in click extortion. By watching its exercises all through a 7-month longitudinal examination, we tend to could create of guide of the performing artists worried in the plan empowering this misleading action.

Table 1: Comparison on existing methods

Ref. Id	Year	Proposed Method	Advantages	Disadvantages
1	2008	Group Blooms Filter algorithm and Timing Blooms Filter algorithm	• Requires simpler operations and less storage space • Low false positive rate	Theoretical analysis was made
4	2012	automated click-fraud based on clickable CAPTCHAs	• Click fraud is identified based on valid users	Loading CAPTCHAs requires time and space
3	2015	DP-Means clustering method	• Is used to detect malicious groups. • The convergence, scalability and accuracy are verified by the simulation data.	One week click data is needed.
6	2015	Feature selection using Recursive Feature Elimination (RFE) and classification through Hellinger Distance Decision Tree (HDDT)	• Accuracy is 64.07%	Validated for mobile advertisement.
2	2016	Social Network Analysis (SNA) Analysis	• Identify the key actors • Uses minimum key actors to identify the click fraud.	Need 7 months of data

The table 1 shows the some of the click fraud identification methodologies used, and the advantages as well as disadvantages of those methods. Each click on their advertisement thereafter generates revenue. It is necessary to identify and stop the process of click fraud introducer in order to protect the advertisers in internet.

CONCLUSION

With the fast development of the Internet, online commercial plays a more imperative part in the publicizing market. One of the current and broadly utilized income models for web based promoting includes charging for each snap in view of the prevalence of catch phrases and the quantity of contending sponsors. This compensation per-click show leaves space for people or adversary organizations to produce false snaps (i.e., click misrepresentation), which posture difficult issues to the improvement of sound web based promoting market. To recognize click extortion, a critical issue is to identify copy clicks over rotting window models, for example, bouncing windows and sliding windows. Rotting window models can be exceptionally useful in characterizing and deciding snap misrepresentation. Be that as it may, in spite of the fact that there are accessible calculations to recognize copies, there is as yet an absence of pragmatic and viable answers for identify click misrepresentation in pay-per-click streams over rotting window models.

REFERENCES

[1] Zhang, Linfeng, and Yong Guan. "Detecting click fraud in pay-per-click streams of online advertising networks." *Distributed Computing Systems, 2008. ICDCS'08. The 28th International Conference on*. IEEE, 2008.

[2] Faou, Matthieu, et al. "Follow the traffic: Stopping click fraud by disrupting the value chain" *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*. IEEE, 2016.

[3] Jiarui, Xu, and Li Chen. "Detecting Crowd sourcing Click Fraud in Search Advertising Based on Clustering Analysis." Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2015 IEEE 12th Intl Conf on. IEEE, 2015.

[4] Costa, Rodrigo Alves, Ruy JGB de Queiroz, and ElmanoRamalhoCavalcanti. "A Proposal to Prevent Click-Fraud Using Clickable CAPTCHAs." *Software Security and Reliability Companion (SERE-C), 2012 IEEE Sixth International Conference on*. IEEE, 2012.

[5] Costa, Rodrigo Alves, Ruy JGB de Queiroz, and ElmanoRamalhoCavalcanti. "A Proposal to Prevent Click-Fraud Using Clickable CAPTCHAs." *Software Security and Reliability Companion (SERE-C), 2012 IEEE Sixth International Conference on*. IEEE, 2012.

[6] Taneja, Mayank, et al. "Prediction of click frauds in mobile advertising" *Contemporary Computing (IC3), 2015 Eighth International Conference on*. IEEE, 2015.

[7] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection for discrete sequences: A survey." *IEEE Transactions on Knowledge and Data Engineering* 24.5 (2012): 823-839.

[8] Blizard, Tommy, and Nikola Livic. "Click-fraud monetizing malware: A survey and case study." *Malicious and Unwanted Software (MALWARE), 2012 7th International Conference on*. IEEE, 2012.

[9] Feily, Maryam, AlirezaShahrestani, and SureswaranRamadass. "A survey of botnet and botnet detection" *Emerging Security Information, Systems and Technologies, 2009.SECURWARE'09. Third International Conference on*. IEEE, 2009.