

Click Fraud

By Brendan Kitts, Benjamin LeBlanc, Ryan Meech, Parameshvyas Laxminarayan

Brendan Kitts, Benjamin LeBlanc, Ryan Meech, and Parameshvyas Laxminarayan of iProspect, 311 Arsenal Street, Watertown, MA. For questions, email: bkitts@excite.com.

The Case of the Missing Clicks

iProspect manages millions of dollars of advertising budgets on Pay-Per-Click (PPC) auctions for many of the largest companies in the world. We have developed our own bidding agent and tracking system (Kitts, et. al., 2004). We are Ambassadors for *Yahoo!*'s Paid Inclusion program. So you can imagine our surprise when we seemed to find an error in *Yahoo!*'s cost accounting.

Both the costs and clicks quoted to us by *Yahoo!* were lower than our independent tracking system was reporting. The under-charge was about 12.5%

Had *Yahoo!* made a mistake?

Little did we know, but *Yahoo!* had been efficiently removing huge numbers of clicks before reaching their customer reports. Think of this as like an Enron document shredding operation in reverse! They were removing fraud before it hit their advertisers. We had stumbled across a way of spying on *Yahoo!*'s Click Fraud Protection System in action.

Overture has created a truly revolutionary, market-driven, information retrieval system. The PPC model is fascinating because the relevance of paid search appears to be as good as classical information retrieval systems, at least on commercial queries (see for example, Jansen et. al., 2005). We applaud Overture's efforts in fighting the newly emerging problem of click fraud. However, the revelation that as many as 1 in 10 clicks are fraudulent leaves one with a queasy feeling in the pit of the stomach. The questions come thick and fast:

- Has all of the fraud been caught?
- What is it about paid search that makes it susceptible to fraud?
- What is the impact of the fraud?
- If fraud continues to grow, what is the future of the PPC model?

This article will introduce readers to the problem of "click fraud", examine the scope of the problem, and discuss methods for overcoming the problem.

What is Click Fraud?

Click fraud, the intentional clicking on PPC advertisements, where the perpetrator has no intention of buying the products or services advertised, is one of the fastest growing problems on the internet. Click fraud generally falls into two categories - clicking on competitors, and network fraud.

Clicking on competitors occurs when a company purposely clicks on a competitor so as to cost them money, use up their daily budgets, and force them off the auction.

John Carreras, President of Impact Displays, says that he knew he had a click fraud problem when he went to a major trade show. He returned to discover that his ad expenses had been 50% lower than normal. He surmised that his competitors were all at the trade show and weren't able to click on his ads! (Eroshenko, 2004)

Network fraud occurs when Website owners click on their own banner advertisements in order to generate revenue from the search engine who is serving the banner advertisement. Most people committing network fraud are small-time operators. However, there are also some professionals.

An End to Internet Advertising?

How could a few clicks do any harm? The doomsday scenario goes something like this. Since ad-clicking is easy and lucrative, an increasing number of fraudsters begin to take advantage of the program. PPC auctions are eventually flooded with fraudulent clicks. Awash with clicks that cost advertisers but generate no purchases, advertisers are crippled by massive advertising costs with almost no return. They stop or reduce their participation in PPC. Search engines lose their fees and can no longer support their operations. The industry turns upon itself as advertisers sue the search engines for delivering fraudulent clicks. Like a massive star collapsing into a black hole, a mass of fraudulent clicks could cause the implosion of the industry.

High flying Google executives are understandably concerned. George Reyes, the Chief Financial Officer of Google, who one would expect to play down this scenario says: "Click fraud is the biggest threat to the Internet economy" (Delaney, 2005). Stephen Messer, CEO of LinkShare expresses similar sentiments: "Click fraud is 'rampant' and 'staggering'.... it could wipe out ROI in search marketing in 2005" (Beal, 2004).

How Prevalent is Click Fraud?

It seems that almost every story on click fraud quotes some expert with an estimate of click fraud in the industry. What are the facts? We developed three methods for estimating the level of click fraud in the industry.

1. Statistical methods

Every website has a conversion rate (purchase rate) that can be calculated by dividing conversions by their clicks. The conversion rate of the site a , gives us the probability that a visitor will convert (purchase).

Now let's consider the activity from one particular user, which we identify by their Internet Protocol (IP) address. If the user clicks on an advertisement a large number of times and does not convert (purchase), this is like flipping a coin repeatedly, with a being probability of a head, and each time having that coin come up tails. The probability of this occurring at random can be calculated using the binomial distribution, where c_u are the number of clicks from user u , A_u the number of conversions from the user, and a is the conversion rate over all users. A user with a p-value less than a critical value of 0.01 will be regarded as probably fraudulent.

$$\text{fraud \%} = \frac{\sum_{u: p_u < 0.01} c_u}{\sum_{i=1}^N c_i} \quad \text{where } p_u = \text{bpdf}(A_u, c_u, a) = \binom{c_u}{A_u} a^{A_u} (1-a)^{(c_u - A_u)} \quad (21)$$

2. Search Engine removed fraud

Yahoo! does not bill for clicks that it considers to be fraudulent. We can therefore measure the difference between the clicks that are tracked from our own tracking systems c_i , against the clicks that *Yahoo!* charges c_j . This under-charge represents the amount of fraud that *Yahoo!* is detecting.

$$\text{fraud \%} = \frac{\sum_{j=1}^M c_j}{\sum_{i=1}^N c_i}$$

3. Consensus estimate from popular media

In recent years, economists have gained a deeper appreciation for the wisdom of crowds. A fascinating example of this is "ideas markets", where people buy shares to try to predict events in the world. Although we don't have an ideas market, there are still a large number of news reports on click fraud being produced by various media outlets, featuring interviews with various industry experts. We ran a web search in May 2005 and found every article we could on click fraud. Every time the story quoted an estimate of the rate of click fraud, we recorded it. We then took the median value. The results are shown in Figure 2.

Disposition of Fraud

The three methods estimated fraud across all industries at 17%, 12.5%, and 15% respectively¹. The Internet Protocol addresses (IPs) that we flagged as fraudulent through our statistical test comprised less than 1% of all IPs (Figure 1).

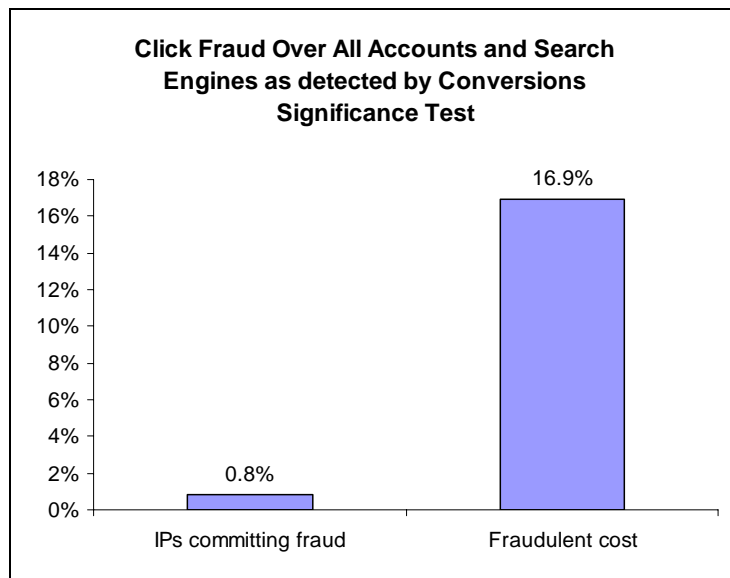


Figure 1: Fraud rates as estimated from a statistical detection regime with p-value of 0.01.

¹ The three statistics refer to percentage of charges where the IP's clicks to conversions were flagged as p<0.01, percentage of rebated clicks, and media percentage estimate.

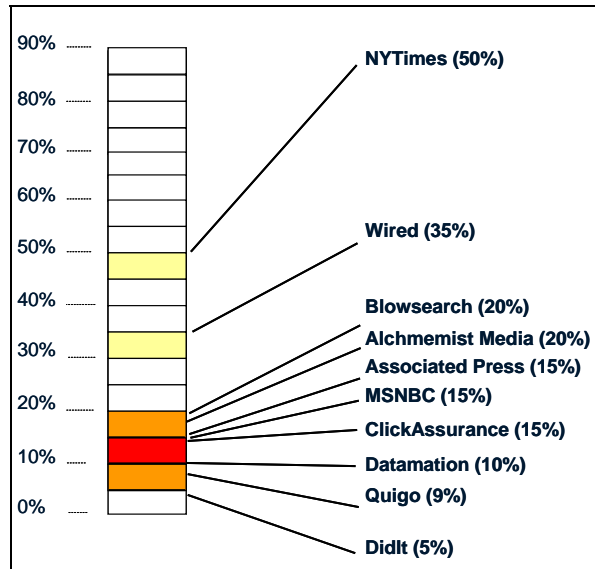


Figure 2: Fraud rates as estimated from a web search of media reports.

It's Good To Be Rational

One would expect that if 15% of clicks were fraudulent, and the search engines were not offering rebates (which they are), then the search engine would generate 15% more revenue.

However, a curious relationship called Ryan's Theorem (Kitts, et. al., 2005) suggests that there may be no change in search engine fees, and bidders may be completely unaffected by changes in *ambient* click fraud - if the bidders are rational.

If an advertiser is rational, their bid price for clicks should track the actual conversion value of the click. If there is a sudden influx of fraud (e.g., 1/G clicks are now valid), the rational response will be for advertisers to drop their bid prices by the same factor (1/G). The result is that there is no change in cost, acquisitions or Cost- Per- Acquisition.

The search engine doesn't even make any more money, because although there are now *G* times the clicks, their clicks are valued at 1/G of their previous price. The critical feature is that the bidders need to value their clicks based upon current conditions.

What should be done?

In 2005, we are beginning to see the crest from an oncoming wave of litigation against the search engines. Lane's Gifts and Collectibles LLC, filed suit in Miller County Arkansas Circuit Court against Google, *Yahoo!*, Ask Jeeves and others, alleging that they charged for fraudulent clicks (Delaney, 2005).

Click Defense Inc. filed suit against Google, alleging losses of over \$5 million from fraudulent clicks.

Google and *Yahoo!* are both working overtime to hand out refund credits.

Does litigation make sense? Will search engine rebates - honorable as they are - eliminate the problem?

We have seen that rational bidding - accurate pricing - can address the problem of ambient click fraud such as network fraud². Sadly, it cannot fix all sources of fraud. The most rational bidder in the world cannot survive if they've been targeted by competitor clicking. In order to eliminate all forms of fraud, two options seem promising.

The Pay Per Purchase (PPP) model could be adopted: Under PPP, the search engines would only be paid after the advertiser achieves a conversion. PPP is undesirable because advertisers would report conversions, and so it opens the door to advertiser fraud. It would also be less lucrative for search engines, since a large number of irrational bidders who are not valuing their clicks properly today would suddenly become perfectly rational.

The second option involves no major change to the PPC model. Search engines could give advertisers the ability to block certain IP addresses from viewing their advertisement. A blocked searcher would still have access to the search engine's natural search results, as well as paid listings from other advertisers.

Whether a customer is uninterested, fraudulent, or likes clicking on their advertisement because they're not familiar with the Internet, the solution for the advertiser is the same. They need to avoid showing their advertisement to that customer. Advertiser-initiated IP blocking would

- (a) avoid search engines paying commissions to sites which are fraudulent.
- (b) encourage web sites displaying the advertisements to improve their quality.
- (c) shift the fraud detection effort from one centralized authority, to hundreds of thousands of interested advertisers. The information processing problem also becomes simpler - while it is hard for a central authority to detect "fraud", it is easy for advertisers to list their "non-converting" IP addresses.

The PPC model works because thousands of advertisers are targeting their advertisements to find converting customers, and hide their listings from non-converting customers. Empower this network with the ability to block IPs, and sources of fraud should rapidly lose their traffic. At least, that's the theory.

References

- Beal, A. (2004), Is Click-Fraud an Issue For Search Marketing?, WebProNews, December 21, 2004, <http://www.webpronews.com/news/ebusinessnews/wpn-45-20041221IsClickFraudandIssueforSearchMarketing.html>
- Delaney, K. (2005), Click Fraud: Web Outfits have a Costly Problem, Marketers Worry about Bills Inflated by People Gaming the Search Ad-System, *Wall Street Journal*, April 6, 2005, Page A1.
- Eroshenko, D. (2004), Click Fraud: The State of the Industry, *Pay Per Click Analyst*, October 19, <http://www.payperclickanalyst.com/content/templates/default.aspx?a=68&z=1>
- Jansen, B.J. & Renick, M. (2005). *An Examination of Sponsored Results for E-commerce Web Searching*, Technical Report, School of Information Sciences and Technology. The Pennsylvania State University. University Park, PA. Copy available upon request from: jjansen@acm.org.

² Search-engine specific features such as minimum bids can break Ryan's theorem. For brevity we avoid introducing these complexities.

- Kitts, B. and LeBlanc, B. (2004), Optimal Bidding on Keyword Auctions, *Electronic Markets*, Vol. 14, No. 3.
- Kitts, B. Laxminarayan, P., LeBlanc, B. and Meech, R. (2005) A Formal Analysis of Search Auctions Including Predictions on Click Fraud and Bidding Tactics, *ACM Conference on E-Commerce - Workshop on Sponsored Search*, Vancouver, UK. June 2005.