Adversarial Information Retrieval Aspects of Sponsored Search

Bernard J. Jansen
College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA, 16801, USA
jjansen@acm.org

ABSTRACT

Search engines are commercial entities that require revenue to survive. The most prevalent revenue stream for search engines is sponsored search, where content providers have search engines service their links to users in response to queries or in a contextual manner on relevant Web sites. In exchange for providing this service, content providers pay search engines based on the number of clicks (i.e., a click being a visit by a user to the content providers Web page). This business model has proven to be very effective for the search engines, content providers, and searchers. However, click fraud, a unique form of adversarial information retrieval, threatens this business model and, therefore, the "free search" that has rapidly become indispensable to the daily lives of many people. In this paper, we outline how sponsored search is a unique form of information retrieval – not just a mode of advertising, what is click fraud, how click fraud happens, and what are some possible countermeasures.

1. INTRODUCTION

Web search engines provide information access to millions of users per day. For many people, Web search engines are now the primary method for finding information, news, and products, according to a recent report on Internet usage [11]. Given this importance, there is increasing attention being paid to search engine spam and other adversarial information retrieval (IR) techniques by content providers to secure undeserved highly ranked positions in the search engine results listings. However, most of the attention in adversarial IR is focused on the algorithmic listings.

Major search engines offer at least two types of results on a search engine results page (SERP), non-sponsored and sponsored results. Sponsored search is an increasingly important, popular, and uniquely contextual form of information interaction on the Web, and is subject to spamming (i.e., click fraud). However, sponsored search and adversarial techniques to subvert it have received little attention in the research community. This lack of consideration is surprising given that the negative effect of spam on the sponsored search process may have greater implications than on the algorithmic procedure.

Sponsored search is the process by which content providers pay Web search engines to display specific links in response to user queries alongside the algorithmic (a.k.a., organic or non-sponsored) links. The sponsored search mechanism plays a critical role in financing the "free" search provided by search engines that have rapidly become essential to many Web users. A distinctive type of interaction involving information push-and-pull, sponsored search also is increasingly important in locating information on the Web. Because of the uniquely dynamic contextual relationship among participants, sponsored search is a distinctive form of IR, and there are significant social and political repercussions if the process is significantly compromised.

In the paper, we provide an overview of sponsored search to demonstrate that it is a unique form of IR and much more than "just online advertising", which may be a common misperception [c.f., 13, 14]. We then discuss click fraud, highlighting how the sponsored search process is susceptible to spamming. We demonstrate how click fraud occurs. We conclude with a discussion of the implications of click fraud and possible mechanisms to combat it.

2. LITERATURE REVIEW

The primary business model for these search engines is sponsored search, where commercial corporations, small businesses, and other entities or individuals pay the search engines to service links that appear on SERP when searchers enter certain key phrases as queries. The content providers may also pay to have their listings presented on Web sites that the search engine's or the content provider deem relevant to the sponsored search links.

The economic impact of paid search is immense. Sponsored search was an $8 billion industry in 2004 and vital to the success of most major search engines. For example, Google received 99% of its $3.1 billion revenue from paid search in 2004; Yahoo! received 84% of its $3 billion, and AOL received 12% of its $1 billion, according to Tim McCarty of Time magazine [12]. In 2005, Web search engines displayed approximately 13 billion sponsored links in a given week, according to Nielsen/NetRatings (http://www.tekrati.com/firmnews/?id=5756). The investment firm Piper Jaffray estimates that online advertising will exceed $55 billion globally by 2010 (http://www.clickz.com/news/article.php/3569361). Without a doubt, sponsored search is now and the foreseeable future the primary business model for Web search engines [5].

For a review of how sponsored search works see [7]. Accounting is one reason that sponsored search is so popular for businesses and organizations. In most models of

AIRWeb'06, August 10, 2006, Seattle, Washington, USA.

advertising, there is little accountable with the cost being impressions (i.e., how many times and when a particular advertisement is shown). However, this is also the key area for an adversarial IR technique known as click fraud.

3. ADVERSARIAL IR ASPECT OF SPONSORED SEARCH

Sponsored search significantly reduces spam content that many times occurs with algorithmic listings. In fact, search engine spam was the primary motivation for the development of the sponsored search paradigm [1]. The reason that sponsored search helped reduce spam is that there is a cost motive for the provider and search engine to present relevant content, and the search engines have review processes consisting of both automated and manual aspects to help ensure this. These monetary factors significantly reduce spam content.

3.1 Click Fraud

However, there is the issue of click fraud with sponsored search. Click fraud is the intentional clicking on a sponsored link where the perpetrator does not intend to buy (or use) the products or services advertised. We use the term "buy", since most sponsored links are ecommerce related. However, more and more non-commercial entities are entering the sponsored search market. So, "buy" may soon be too restrictive. Regardless, click fraud is one of the fastest growing problems on the Web, according to iProspect (http://www.iprospect.com/media/newsletter_october_meech.htm?ipsrc=media&reftype=pi&sptype=osmx). Click fraud has not been widely perceived as search engine spamming [6], but its negative effect is severe.

Click fraud can take various forms, but the final result is usually the same. Content providers pay for unproductive traffic generated by perpetrators who repeatedly click on a content provider's sponsored link with no intention of buying anything. Click fraud produces revenue for the major search engines and the Web sites that display the links. This is because the clicks generate sales commissions based on the content provider's bid even if the click does not result in a sale. In sponsored search, content providers are contractually obligated to pay for all valid clicks. However, the search engine has discretion over what is valid. According to [10], based on an analysis of more than 1,000 content providers, Google and Yahoo!'s sponsored search programs suffered a click fraud rate of 12%, translating to more than $1.5 billion of Google's ad revenue in 2005. The 12% click fraud rate correlates well with that reported in [9]. However, some content providers complain that their individual click fraud rate is as high as 35% [10]. See [9] for an overview of the click fraud issue.

3.2 Click Fraud Implementation

Why and how does click fraud occur? As for the why, sometimes one content provider tries to deplete a competitor's sponsored search budget (most content providers have monetary limits for any period). In other instances, the owners of Web sites servicing sponsored links click on these links to generate commissions for themselves. Finally, some even more "ethically challenged" individuals set up fictitious Web sites targeted at high payoff sponsored terms. These Web sites exist solely to generate commissions for clicks on sponsored links. The owners of these Web sites typically use automated tools to both set up and generate clicks.

In the first case, depleting a competitor's budget, the motivation is usually to increase the cost of advertising for the other content provider, exhausting the rival's budget. Once the rival's link has dropped out of the search engine's listing, more traffic is diverted to the remaining sponsored links. This type of click fraud is fairly easy to implement. For example, see Figure 2.

Figure 2 contains a snippet from a SERP, namely the sponsored link section. From Figure 2, we see that the initial query "Jim Jansen" retrieved three sponsored results with the sponsored link **Jim Jansen** in the first position. With repeated submissions of the query "Jim Jansen" and subsequent clicks on the sponsored result, we see that the link **Jim Jansen** soon drops out of the sponsored listing once the daily budget has expired. The effects are (1) the content provider of the link **Jim Jansen** pays the search engine for each of these clicks, (2) once the budget for the link **Jim Jansen** is exhausted that link no longer appears, depriving the content provider of traffic, and (3) with the link J**im Jansen** gone, the other links move up in ranking. Studies show that about 30% of searches involve a click on a sponsored link [8] and that the higher a link is in the results listing, the more visits that Web site will receive [2, 3].

For the other two cases of click fraud, the motivation is money. This version of click fraud consists of Web site owners who service sponsored content on their sites and then click these links to generate commissions. See Figure 3 for an example.

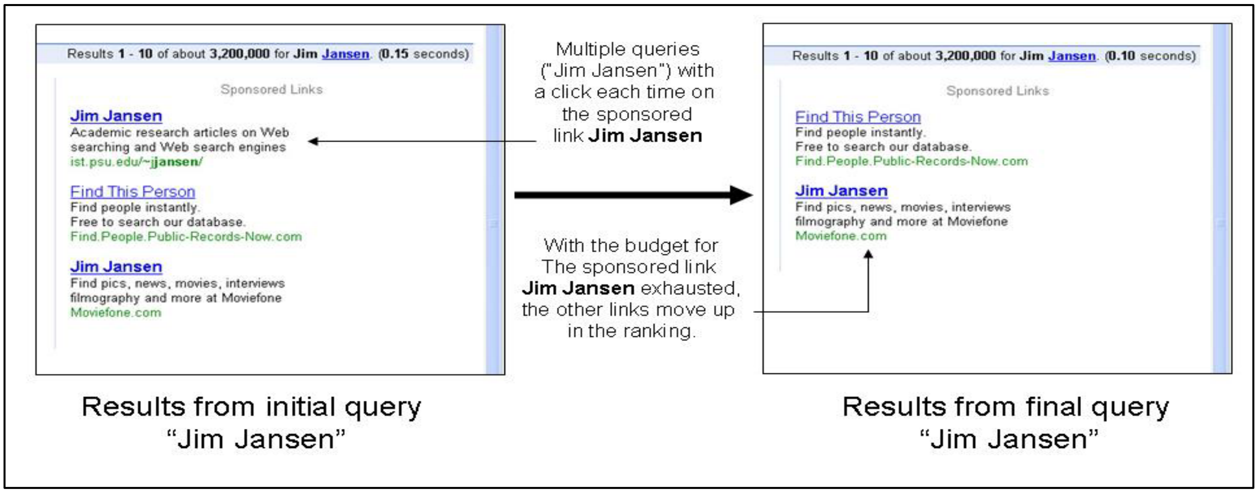
We see in Figure 3 that this particular Web site serves contextual sponsored links from Google. Many times these links appear during a normal visit to the Web site or one can trigger their appearance via searching conducted on the Web site. In Figure 3, the query "Jim Jansen" to the Web site's local content prompted the display of the sponsored link **Jim Jansen**. By clicking on this sponsored link, the content provider will pay the search engine, who will then split the payment with the Web site owner.

Numerous software packages are available that will assist in setting-up or nearly automatically setting-up a Web site targeted at high pay-off key words, automate the clicks, and disguise the Internet Protocol (IP) address. The process is marketed as something easy to do, as shown in Figure 4.

3.3 Click Fraud Prevention

What are some potential click fraud countermeasures?

- **Automated and Human Filters**: Search engines currently employ both automated and human filters in an attempt to identify current and prevent future click fraud. Search engine also appear to be making reasonable attempts to reimburse or not charge clients for click identified as click fraud. However, given the string of class action lawsuits, it is apparent significantly more needs to be done.



Results **1** - **10** of about **3,200,000** for **Jim Jansen**. (**0.15** seconds)

Sponsored Links

Jim Jansen
Academic research articles on Web searching and Web search engines
ist.psu.edu/~**jjansen/**

Find This Person
Find people instantly.
Free to search our database.
Find.People.Public-Records-Now.com

Jim Jansen
Find pics, news, movies, interviews filmography and more at Moviefone
Moviefone.com

Multiple queries ("Jim Jansen") with a click each time on the sponsored link **Jim Jansen**

With the budget for The sponsored link **Jim Jansen** exhausted, the other links move up in the ranking.

Results **1** - **10** of about **3,200,000** for **Jim Jansen**. (**0.10** seconds)

Sponsored Links

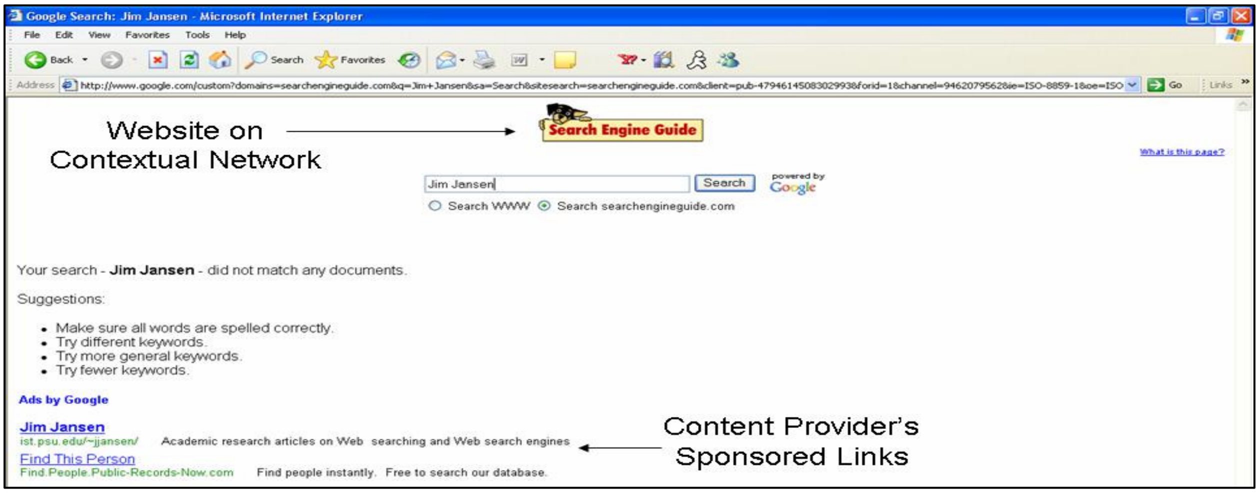
Find This Person
Find people instantly.
Free to search our database.
Find.People.Public-Records-Now.com

Jim Jansen
Find pics, news, movies, interviews filmography and more at Moviefone
Moviefone.com

Results from initial query "Jim Jansen"

Results from final query "Jim Jansen"

Figure 2. An Example of Click Fraud on the Sponsored Listings of a Web Search Engine.



Google Search: Jim Jansen - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.google.com/custom?domains=searchengineguide.com&q=Jim+Jansen&sa=Search&sitesearch=searchengineguide.com&client=pub-4794614508302993&forid=1&channel=9462079562&ie=ISO-8859-1&oe=ISO
Go Links
Search Engine Guide
What is this page?
Search
powered by Google
○ Search WWW ● Search searchengineguide.com

Website on Contextual Network

Jim Jansen

Your search - **Jim Jansen** - did not match any documents.

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.
- Try fewer keywords.

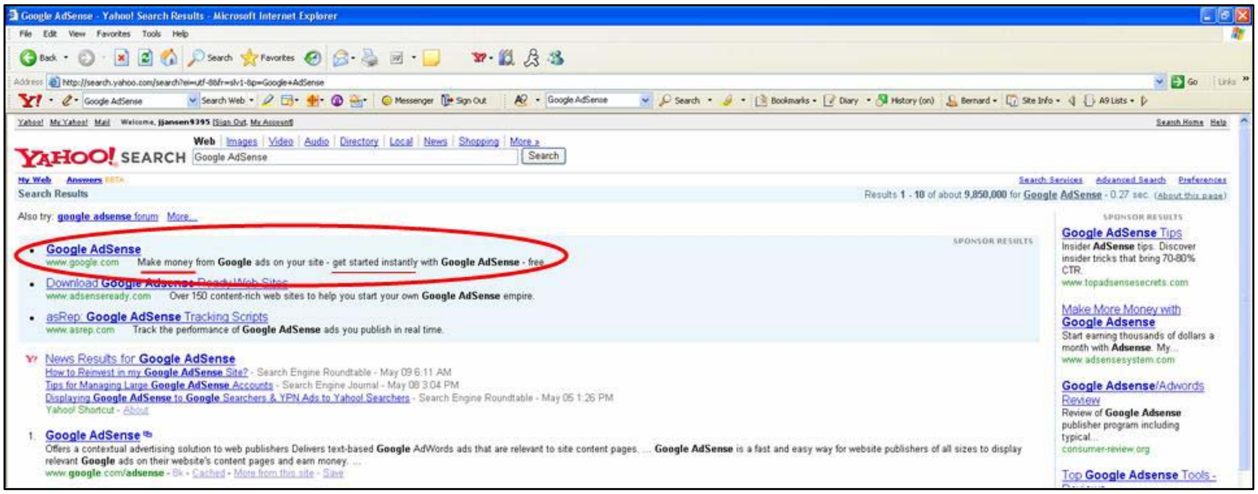
Ads by Google

Jim Jansen
ist.psu.edu/~jjansen/ Academic research articles on Web searching and Web search engines

Find This Person
Find.People.Public-Records-Now.com Find people instantly. Free to search our database.

Content Provider's Sponsored Links

Figure 3. An Example of Contextual Link Where Click Fraud Can Occur.



Google AdSense - Yahoo! Search Results - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://search.yahoo.com/search?ei=utf-8&fr=slv1-&p=Google+AdSense
Go Links
Google AdSense
Search Web
Messenger
Sign Out
Google AdSense
Search
Bookmarks
Diary
History (on)
Bernard
Site Info
A9 Lists
Yahoo! My Yahoo! Mail Welcome, jjansen9395 [Sign Out, My Account]
Search Home Help
Web Images Video Audio Directory Local News Shopping More »
YAHOO! SEARCH
Google AdSense
Search
My Web Answers BETA
Search Services Advanced Search Preferences

Search Results

Results **1** - **10** of about **9,850,000** for **Google AdSense** - 0.27 sec. (About this page)

Also try: **google adsense** forum More...

SPONSOR RESULTS

- **Google AdSense**
  www.google.com Make money from **Google** ads on your site - get started instantly with **Google AdSense** - free
- Download **Google Adsense** Ready Web Sites
  www.adsenseready.com Over 150 content-rich web sites to help you start your own **Google AdSense** empire.
- asRep: **Google AdSense** Tracking Scripts
  www.asrep.com Track the performance of **Google AdSense** ads you publish in real time.

News Results for **Google AdSense**
How to Reinvest in my **Google AdSense** Site? - Search Engine Roundtable - May 09 6:11 AM
Tips for Managing Large **Google AdSense** Accounts - Search Engine Journal - May 08 3:04 PM
Displaying **Google AdSense** to **Google** Searchers & YPN Ads to Yahoo! Searchers - Search Engine Roundtable - May 05 1:26 PM
Yahoo! Shortcut - About

1. **Google AdSense**
   Offers a contextual advertising solution to web publishers Delivers text-based **Google** AdWords ads that are relevant to site content pages. ... **Google AdSense** is a fast and easy way for website publishers of all sizes to display relevant **Google** ads on their website's content pages and earn money. ...
   www.**google**.com/**adsense** - 8k - Cached - More from this site - Save

SPONSOR RESULTS

Google AdSense Tips
Insider **AdSense** tips. Discover insider tricks that bring 70-80% CTR.
www.topadsensesecrets.com

Make More Money with **Google Adsense**
Start earning thousands of dollars a month with **Adsense**. My...
www.adsensesystem.com

Google Adsense/Adwords Review
Review of **Google Adsense** publisher program including typical...
consumer-review.org

Top **Google Adsense** Tools -

Figure 4. A Sponsored Link Concerning Google's AdSense Program.

Certainly, more sophisticated automated filters and data mining techniques need to be employed, more human effort needs to be engaged, and much better communication about these efforts to the customers and public.

- **Pay-per-action Paradigm**: One partial solution is a shift in paradigm from pay-per-click to pay-per-action. With pay-per-action, the advertiser only pays if the visitor actually executes an action, such as purchasing a product. However, pay-per-action is not the total answer though, as research reports that many searchers visit a sponsored link multiple times before a purchase [4]. Additionally, some of the traffic generated should be based on the ability of the content provider to construct enticing sponsored links.

- **Block Blacklisted IP addresses**: There are various databases of blacklisted IPs (c.f., http://www.declude.com/Articles.asp?ID=97 and http://www.moensted.dk/spam/), which maintain lists of IPs that are know email spammer sites. These spam lists are also known as Realtime Blackhole List (RBL). The company Mail Abuse Prevention System (MAPS) LLC actively maintains records of RBLs. These are IP addresses whose owners refuse to stop others from using their servers for spam. Email servers routinely block messages from these IPs. However, click fraud perpetrators also use these IP. It would seem reasonable that the search engines could take measures to block clicks from these IPs or reimburse content providers for clicks from these IP addresses.

- **Aggressive monitoring of click fraud perpetrators:** Click fraud is similar to what occurred in the online music industry. The Recording Industry Association of America's (RIAA) campaign against illegal file sharing via peer-to-peer networks is a good example of the effect that an aggressive operation can achieve (c.f., http://www.techweb.com/wire/story/TWB20031105S0006). The RIAA's effort significantly reduced illegal copying of copyrighted audio files. Many content providers have been critical of the major search engines for their lack of aggressive pursuit of click fraud abusers. Aggressive action against click fraud would raise the cost of click fraud, and could reduce the number of folks doing it.

- **Search engines must make efforts to ensure trust**: In sponsored search, content providers sign contracts to pay for all valid clicks, with the search engine determining what is a valid click. Trust is a, if not the, critical element in the sponsored search paradigm. Although the major search engines do make efforts to identify click fraud, sponsored search is not subject to independent auditing. San Antonio-based Click Forensics Inc. recently set up a free service that intends to issue quarterly reports on the frequency of click fraud [10]. Whether through independent auditing or internal efforts, content providers and searchers must have trust in the process if it is to be a long-term business model.

4. CONCLUSION

It appears that the sponsored search model will have increasing impact as new players enter the field. Within this extremely dynamic paradigm, click fraud threatens the entire process. Although media reports rates as high as 50%, studies in the area indicate click fraud rates of between 12% and 16% [9]. This translates into billions of dollars per year, and it jeopardizes the entire model as it decreases trust in the system, which is the basis of any IR process. In this regard, the onus is on the search engines and related researchers to develop methods to combat this threat. These methods run the gamut from technological, to business processes, to regulatory measures.

5. REFERENCES

[1] Battelle, J., *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed our Culture*. New York: Penguin Group, 2005.

[2] Brooks, N., *The Atlas Rank Report I: How Search Engine Rank Impacts Traffic*, Accessed on 1 August 2004 on the World Wide Web at http://www.atlasdmt.com/media/pdfs/insights/RankReport.pdf.

[3] Brooks, N., *The Atlas Rank Report II: How Search Engine Rank Impacts Conversions*, Accessed on 15 January 2005 on the World Wide Web at http://www.atlasonepoint.com/pdf/AtlasRankReportPart2.pdf.

[4] Brooks, N., Repeat Search Behavior: Implications for Advertisers, *Bulletin of the American Society for Information Science and Technology*, vol. 32, pp. 16-17, 2006.

[5] Chiang, K.-P., Clicking Instead of Walking: Consumers Searching for Information in the Electronic Marketplace, *Bulletin of the American Society for Information Science and Technology*, vol. 32, pp. 9-10, 2006.

[6] Gyongyi, Z. and Garcia-Molina, H., Web Spam Taxonomy, in *Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb '05), the 14th International World Wide Web Conference (WWW2005)*, 2005. Chiba, Japan. 10-14 May. pp. 39-47.

[7] Jansen, B. J., Paid Search, *IEEE Computer*, Forthcoming.

[8] Jansen, B. J. and Resnick, M., An examination of searchers' perceptions of non-sponsored and sponsored links during ecommerce Web searching, *Journal of the American Society for Information Science and Technology*, forthcoming.

[9] Kitts, B., LeBlanc, B., Meech, R., and Laxminarayan, P., Click Fraud, *Bulletin of the American Society for Information Science and Technology*, vol. 32, pp. 14-16, 2005.

[10] Liedtke, M., *Click Fraud Concerns Hound Google*, in *ABC News Money*, 2006 http://abcnews.go.com/Technology/wireStory?id=1934655&CMP=OTC-RSSFeeds0312.

[11] Madden, M., *Internet Penetration and Impact*, Accessed on 9 May 2006 on the World Wide Web at http://www.pewinternet.org/PPF/r/182/report_display.asp.

[12] McCarthy, T., Yahoo! Goes to Hollywood, *Time*, vol. 165, pp. 50-53, 2005.

[13] Metaxas, P. T. and DeStefano, J., Web Spam, Propaganda and Trust, in *Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb '05), the 14th International World Wide Web Conference (WWW2005)*, 2005. Chiba, Japan. 10-14 May. pp. 70-78.

[14] Nicholson, S., Sierra, T., Eseryel, U. Y., Park, J.-H., Barkow, P., Pozo, E. J., and Ward, J., How much of it is real? Analysis of paid placement in Web search engine results, *Journal of the American Society of Information Science and Technology*, vol. 57, pp. 448-461, 2006.