Click Fraud

Bernard J. Jansen
The Pennsylvania State University

Unchecked, click fraud could undermine the sponsored search business model.

Today, Web search engines are the primary method for millions of users throughout the world to access information on a topic, navigate to Web sites, keep up with the news, and shop online.

Most major search engines generate revenue via *sponsored search*, a process whereby content providers pay for traffic from specific links the search engine's display in response to user queries. Search engines typically display these links alongside nonsponsored links, also known as organic or algorithmic links.

Sponsored search has become an integral part of the business model of most search engines and many online retailers, generating billions of dollars in revenue each year. As such, it plays a critical role in financing the "free" search provided by search engines that has become indispensable to many Web users.

Given the profound impact of sponsored search on Web content access, anything that compromises the process would have significant social, economic, and political repercussions. *Click fraud*, which involves the intentional clicking on sponsored links with the purpose of gaining undue monetary returns or harming a particular content provider, has the potential to do just that.

Click fraud is one of the fastest growing problems on the Web, according to search-engine marketing firm iProspect (www.iprospect.com). It can take various forms, but the result is usually the same: Content providers pay for unproductive traffic generated by perpetrators who repeatedly click on a sponsored link with no intention of giving value to that provider.

UNDERSTANDING CLICK FRAUD

To understand click fraud, it is necessary to define some key terms.

In this context, *value* is the use of information, employment of a service, purchase of a product, or execution of a transaction by a Web site visitor that is consistent with the content provider's goal.

A *sponsored result* is the title, text, and other material associated with a particular sponsored link. A *sponsored link* is a URL serviced by a search engine in response to a query in a search-engine results page (SERP) or in a contextual manner in a relevant Web site or e-mail.

A *click* is the act of initiating a visit to a Web site via a sponsored link and can be either

- *valid*—an intentional click that has a realistic probability of generating value once the visitor arrives at the Web site; or
- *invalid*—a click on a sponsored link that has no probability of generating value.

Invalid clicks can be either fraudulent—that is, malicious—or void. A *fraudulent click* is an intentional click on a sponsored link with no intention of generating value. *Identifiable* click fraud is a pattern of fraudulent clicks that can be distinguished from valid clicks, while *unidentifiable* click fraud is a pattern of fraudulent clicks that can't be distinguished from valid clicks.

A *void click* is an invalid click that isn't fraudulent or malicious—for example, a double click on a sponsored link or a click on a sponsored link when the Web site is down. A void click is identifiable if it can be distinguished from a set of valid clicks, and it's unidentifiable if it can't be distinguished.

IDENTIFYING CLICK FRAUD

Identifying click fraud is surprisingly difficult. Void clicks are relatively easy to identify based on aggregate user metrics such as time between clicks on the same sponsored link and time visiting a page. However, click fraud perpetrators attempt to make their clicks look like valid ones, thus analysis must consider individual as well as aggregate behavior.

Figure 1 represents the total "click space"of search-engine visits. Of a given body of visits, around 30 percent result in one or more clicks on sponsored links (www.internetnews.com/xSP/article. php/3502611).

Popular press reports indicate that search engines screen about 15 percent of all invalid clicks, or 5 percent of all clicks (www.businessweek.com/magazine/content/06_40/b4003001.htm). Search engines' accuracy in filtering invalid clicks isn't precisely known but can reasonably be estimated to be 80 percent or higher. This leads to the conclusion that invalid clicks make up about 6 percent of all search-engine visits.

Assuming that unidentifiable void clicks are nearly zero, slightly more than 1 percent of all search-engine visits result in an unidentifiable fraudulent click. Although a low percentage, this can translate into tens of millions

of dollars for billion-dollar-earning search engines.

In addition, these percentages are only for sponsored links off the SERP. No comparable estimates exist for fraudulent clicks on sponsored links from contextual Web sites, but complaints from content providers suggest that this rate is much higher (www.businessweek.com/technology/content/feb2006/tc20060227_930506.htm).

COMBATING CLICK FRAUD

There are few effective legal restrictions against click fraud, putting the onus for combating this threat on the search engines.

Aggressive monitoring

Many content providers have been critical of the major search engines for not aggressively pursuing click fraud. Doing so would raise the cost of click fraud and could reduce the number of perpetrators.

Improving automated filters

Search engines currently use both automated and human filters to identify and prevent click fraud. They also appear to be making reasonable attempts to reimburse or not charge clients for fraudulent clicks. However, search engines need to incorporate more effective automated filters using sophisticated data mining technology and do a better job communicating these efforts to both customers and the general public.

Pay-per-action

One partial solution to the click fraud problem is for search engines to shift from a pay-per-click to a pay-per-action paradigm. With pay-per-action, the advertiser only pays if the visitor actually executes an action, such as purchasing a product. Snap (www.snap.com) is an example of a search engine that offers a sponsored search program to content providers based on pay-per-action.

However, a limitation of this approach is that a user might visit a sponsored link multiple times before making a purchase (www.atlassolutions.com/pdf/RankReportPart2.pdf). In addition, some of the traffic generated should be based on the content provider's ability to construct enticing sponsored links. Finally, having a sponsored link appear in a SERP has branding value.

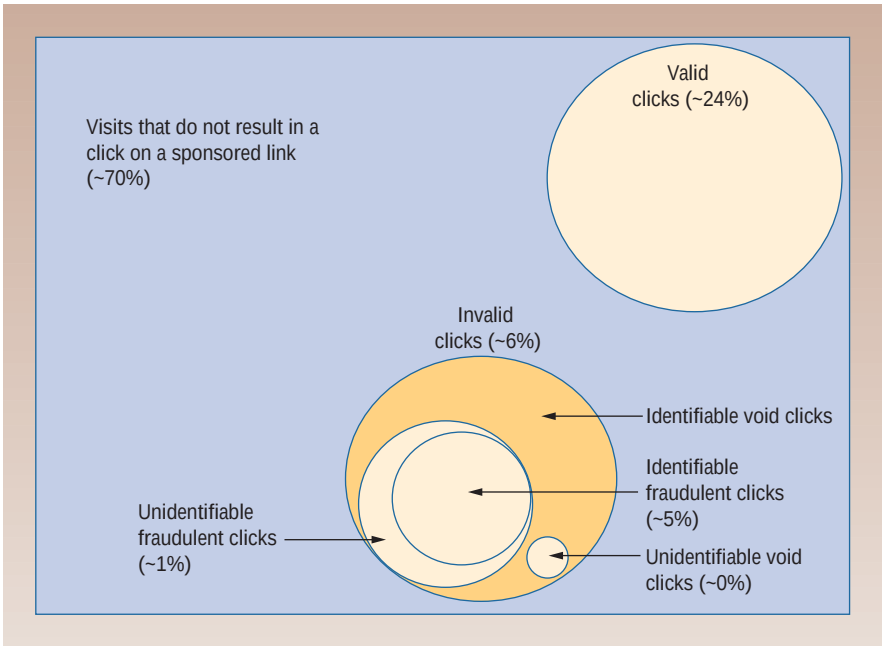

Figure 1. Click space. About 30 percent of all search-engine visits result in one or more clicks on sponsored links, most of which are valid. Slightly more than 1 percent of all search-engine visits result in an unidentifiable fraudulent click.

Cultivating trust

With sponsored search, content providers sign contracts with search engines to pay for all valid clicks, with the search engine determining which clicks are valid. Only if content providers and users trust the process will it become a successful long-term business model.

Although the major search engines do make efforts to identify click fraud, sponsored search currently isn't subject to independent auditing. However, emerging companies such as Click Forensics are beginning to document the frequency of click fraud (http://cbs3.com/national/topstories_story_127230350.html).

Google, Yahoo!, and other companies vying for a piece of the lucrative search-engine market continue to transform the sponsored search model, linking results to other information media such as telephones and television. However, click fraud undermines this process by reducing the value of Web site traffic to content providers and thereby decreasing revenue for the search engines.

Threats to the search engines' underlying business model are also threats to the free search services that these companies provide to millions of users, making click fraud a concern for all.

Bernard J. Jansen is an assistant professor in the College of Information Sciences and Technology at the Pennsylvania State University. Contact him at jjansen@acm.org.

Editor: Simon S.Y. Shim, Department of Computer Engineering, San Jose State University; sishim@email.sjsu.edu