Click-Tracking Blocker: Privacy Preservation by Disabling Search Engines' Click-Tracking

Roberto Alberdeston, Erich Dondyk, Cliff C. Zou
Computer Science Division
Department of Electrical Engineering and Computer Science
University of Central Florida
Orlando, FL 32816 USA

***Abstract*—On a search engine result page, if a user clicks on any one of those contained URL links, whether it is a search result website (called organic link), or a search related advertisement link (called sponsored link), the user's click action will be tracked by returning back to the search engine first and then redirecting to the corresponding target website. This click-tracking is conducted by all three major search engines: Google, Bing, and Yahoo (although Bing does not track clicks on organic links). Many people are not aware that their clicks are tracked by search companies; and this click-tracking imposes a big privacy threat for privacy-concerned users.**

After discovering that all organic and sponsored links in a search engine result page actually encode the real URLs of target websites with simple encoding techniques, we have developed a browser plug-in program called *Click-Tracking Blocker* that modifies all URLs on search result pages once they are retrieved by users' computers, such that search engines will no longer be able to track users' clicking behaviors on both organic links and sponsored links. Click-Tracking Blocker can be readily and easily installed by end users to protect their privacy, works for all three major search engines, and does not affect users' search experience.

Keywords—***user privacy; search engine click tracking; sponsored links***

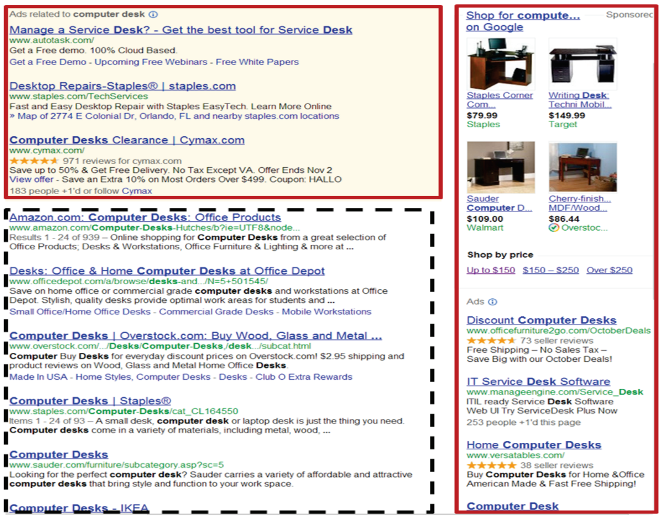
I. Introduction

Search engines are the backbone of today's Internet. They have redefined most aspects of human activities in modern society such as business, shopping, daily life and advertising. From a business standpoint, search engines are similar to the television or newspaper industry where most of their revenue comes from advertising [5]. Search engines collect significant data from their users for two main goals: improving ad relevancy and maximizing revenue.

At this time, U.S. search engine market is primarily concentrated in three companies: Google, Yahoo and Bing. As of September 12, 2012 the top-3 search engines in U.S. market share were Google (86%), Yahoo (7%) and Bing (4%) [1]. We primarily concentrate our research in the U.S. market and focus on these three search engines (after 2009, the Yahoo search engine results are managed by Bing [2]).

In virtually every search engine, if you search for a term, there are typically two types of URL link results contained in the search engine result pages (SERPs): *organic links* and *sponsored links*. "Organic links" are defined as hypertext links between websites with or without an explicit agreement to exchange links [3]. On SERPs, they are the search result website URL links satisfying what the user wants to search through a search engine. Organic links appear in the SERPs in the order according to the search engine's long term ranking algorithms. "Sponsored links" point to advertisement websites and they generate income for a search engine every time someone clicks them and search engine is paid from that click.

Organic and sponsored links of Google's SERPs are illustrated in Figure 1, which is the search result page for the search term "computer desk".



Ads related to **computer desk** ⓘ

Manage a Service **Desk**? - Get the best tool for Service **Desk**
www.autotask.com/
Get a Free demo. 100% Cloud Based.
Get a Free Demo - Upcoming Free Webinars - Free White Papers

Desktop Repairs-Staples® | staples.com
www.staples.com/TechServices
Fast and Easy Desktop Repair with Staples EasyTech. Learn More Online
» Map of 2774 E Colonial Dr, Orlando, FL and nearby staples.com locations

Computer Desks Clearance | Cymax.com
www.cymax.com/
★★★★★ 971 reviews for cymax.com
Save up to 50% & Get Free Delivery. No Tax Except VA. Offer Ends Nov 2
View offer - Save an Extra 10% on Most Orders Over $499. Coupon: HALLO
183 people +1'd or follow Cymax

Amazon.com: **Computer Desks**: Office Products
www.amazon.com/**Computer-Desks**-Hutches/b?ie=UTF8&node...
Results 1 - 24 of 939 – Online shopping for **Computer Desks** from a great selection of Office Products; Desks & Workstations, Office Furniture & Lighting & more at ...

Desks: Office & Home **Computer Desks** at Office Depot
www.officedepot.com/a/browse/**desks**-and.../N=5+501545/
Save on home office or commercial grade **computer desks** and workstations at Office Depot. Stylish, quality desks provide optimal work areas for students and ...
Small Office/Home Office Desks - Commercial Grade Desks - Mobile Workstations

Computer Desks | Overstock.com: Buy Wood, Glass and Metal ...
www.overstock.com/.../**Desks**/**Computer-Desks**,/**desk**.../subcat.html
Computer Buy **Desks** for everyday discount prices on Overstock.com! $2.95 shipping and product reviews on Wood, Glass and Metal Home Office **Desks**.
Made In USA - Home Styles, Computer Desks - Desks - Club O Extra Rewards

Computer Desks | Staples®
www.staples.com/**Computer-Desks**/cat_CL164550
Items 1 - 24 of 93 – A small desk, **computer desk** or laptop desk is just the thing you need. **Computer desks** come in a variety of materials, including metal, wood, ...

Computer Desks
www.sauder.com/furniture/subcategory.asp?sc=5
Looking for the perfect **computer desk**? Sauder carries a variety of affordable and attractive **computer desks** that bring style and function to your work space.

Computer Desks - IKEA

Shop for **compute...** on Google — Sponsored

Staples Corner Com...
$79.99
Staples

Writing **Desk**: Techni Mobil...
$149.99
Target

Sauder **Computer** D...
$109.00
Walmart

Cherry-finish... MDF/Wood...
$86.44
Overstoc...

Shop by price
Up to $150 $150 – $250 Over $250

Ads ⓘ

Discount **Computer Desks**
www.officefurniture2go.com/OctoberDeals
★★★★★ 73 seller reviews
Free Shipping – No Sales Tax – Save Big with our October Deals!

IT Service **Desk** Software
www.manageengine.com/Service_**Desk**
ITIL ready Service **Desk** Software
Web UI Try ServiceDesk Plus Now
253 people +1'd this page

Home **Computer Desks**
www.versatables.com/
★★★★★ 38 seller reviews
Buy **Computer Desks** for Home &Office
American Made & Fast Free Shipping!

Computer Desk

Figure 1: Organic vs. Sponsored Links in Google's SERPs. Links in the two solid rectangles are sponsored links; links in the dashed rectangle area are organic links showing the top few search result links.

Sponsored links appear in SERPs based on a number of factors which are different from the factors used to rank organic links. One major factor is the amount of money companies are willing to pay for each click they get to their sponsored ads, which is called pay-per-click (PPC). Another factor is the percentage of times the sponsored links are clicked, which is called click-through rate (CTR). Both factors are driven by financial motivation [4], and require search engines to have an accurate monitoring on user clicks on all sponsored links.

Some people have noticed the privacy issue of click-tracking in SERPs, and several programs have been developed to circumvent it, such as Google Anonymizer [13] and

Straight Google [14]. However, they only circumvent the click-tracking of organic links by Google; nothing is done to deal with the click-tracking on sponsored links. In addition, they do not consider other major search engines like Bing and Yahoo.

In this paper, we show that besides organic links, it is also easy to circumvent sponsored link click-tracking as well for all three major search engines with their current SERP construction format. Both organic and sponsored links for all three search engines actually contain the real URLs of target websites by using somewhat similar encoding techniques. We have developed a program called *Click-Tracking Blocker*, in the form of browser plug-in, to modify search result pages once they are retrieved by users' computers, such that search engines will no longer be able to track users' clicking behaviors on all organic links and sponsored links contained in SERPs. Click-Tracking Blocker works for all three major search engines, is transparent to end users, and does not affect users' search experience.

The rest of this paper is organized as follows. Section II provides discussion of related work. We describe the detailed analysis of click-tracking approaches used by the three major search engines in Section III. Then Section IV presents the design and implementation of our proposed Click-Tracking Blocker. We evaluate our prototype program and analyze the structural features of sponsored links in Section V. Finally Section VI concludes this paper.

II. Related Work

People have studied user privacy issue related to Internet search engines. Search engines collect rich data of their users, more than most people have known. For Internet users that have not logged into services belonging to search engines (such as Yahoo email or Gmail), search engines still collect rich information of users. [6][7] discussed what information could be recorded by a search engine using query logs, such as IP address, query term, and cookie based ID. Eckersley et al. [8] offered several tips for Internet users on how to protect their search privacy, but most of these tips are not practical or very hard for ordinary users to implement.

After AOL released a list of three months of search queries from 657,000 "anonymous" users in 2006 [9], people quickly realized how much private information could be derived solely based on query term profile, even to pinpoint an individual user [10]. Facing this privacy concern, Howe and Nissenbaum [11] developed *TrackMeNot*, a browser extension tool to achieve privacy in Web search by obfuscating users' queries within a stream of randomly generated noise queries. It can successfully prevent search engines to extract user private information based on search queries, but at the cost of generating excessive wasted network traffic to both the search engines and users' own Internet access. In addition, TrackMeNot does not deal with the privacy exposure by click-tracking conducted by search engines.

After the Firefox extension platform, Greasemonkey, appeared and allowed other developers to customize the way webpages look and function [12], several programs have been developed based on it to defend against click-tracking by Google search engine to certain extend. *Google Anonymizer* [13] disables organic link tracking on Google's SERPs and deletes known Google's user cookies from cookie folder. However, it does not prevent click-tracking on sponsored links in Google's SERPs. *Straight Google* [14] circumvents click tracking in most Google properties including YouTube, Groups, Docs and more. However, just like Google Anonymizer it cannot deal with click-tracking on Google's sponsored links as well. In addition, these programs focus on Google only; they do not deal with other major search engines like Yahoo and Bing.

The above privacy preservation tools, and our proposed Click-Tracking Blocker in this paper, will not be effective anymore if a search engine implements mouse cursor tracking on its SERPs [15]. However, mouse cursor tracking needs an SERP running a client-side JavaScript, which essentially acts like a keylogger and thus should not be allowed by default by browsers or search engines' privacy policy.

In addition, there are a few search engine service providers that claim to offer search engine services while keeping no search query logs or use of click tracking and cookies. Two examples of such search engine providers include DuckDuckGo.com and ixquick.com. However, their search performance and privacy is not well studied and few people use them at the current time.

III. Analysis of Click-Tracking used by Search Engines

As illustrated in Fig. 1, an SERP contains two categories of URL links, organic links and sponsored links. In this section, we explain the click-tracking algorithms used for both types of links by the three major search engines.

A. Click-Tracking Methods Used for Organic Links

Among the three major search engines, Bing does not track user's clicking on organic links. In other words, all organic links contain their true destination URLs and Bing cannot know which organic link on its SERP a user has clicked.

On the other hand, both Google and Yahoo conduct click-tracking on organic links by redirecting a user's click to a Google server or a Yahoo server after the user clicks any of the organic links. The search engine server, in turn, immediately redirects the user to the intended destination server. Therefore, many users do not notice this quick two-hop redirection behavior, and hence, never realize that their clicking actions have been monitored by these two search engines.

For better understanding, we describe the two different tracking approaches deployed by Google and Yahoo by using the following example link, i.e., assuming the destination URL in a search result page is:

```
http://buy.com/?id=2&ref=30
```

Figure 2: Example destination URL in a search engine result page

Google uses JavaScript in its SERP to substitute each destination organic link URL with the corresponding Google server's URL based on the "*onmousedown*" event. For the above destination URL example, its source code in Google's search result page HTML file would be:

```
<a href="http://buy.com/?id=2&ref=30" class=l
onmousedown="return rwt(this,",",",'1','******',",
'******',",",event)">
```

Figure 3: HTML source code for the organic link of the example URL in a Google search result page

The two "******" symbols above represent two ASCII strings that have unknown lengths. They are used by Google to track detailed information of the user who opens the SERP. The "onmousedown" event is activated whenever the link is left clicked or right clicked. Once it is activated, it generates a Google server's click-tracking URL like this:

```
http://www.google.com/url?^^^^^&url=http%3A%2F%2F
buy.com%2F%3Fid%3D2%26ref%3D30&^^^^^
```

Figure 4: Google's click-tracking URL generated by user's click on the example destination link. The real target URL is encoded and is right after the keyword '&url'.

"^^^^^" represents an arbitrary ASCII strings including characters '&' and '='. Each of the special characters in the original destination URL, {/ : # & % [ ] + { } | ? =}, has been encoded into a three-character representation of its ASCII hexadecimal value, which is usually called "Percent-Encoding" [17]. For example, '/' has ASCII value of 0x2F and hence it is replaced by '%2F'. In the same way, ':' has ASCII value of 0x3A and is replaced by '%3A', '?' is '%3F', '=' is '%3D', and '&' is '%26'. Therefore, we can see that following the keyword '&url' is the encoded destination's URL, and further separated by '&' with another Google tracking ASCII string "^^^^^".

Similarly to Google, Yahoo uses JavaScript in its SERP to substitute each destination organic link URL with the corresponding Yahoo server URL. However, unlike Google, the tracking URL is not generated on the client side. The HTML anchor elements of each Yahoo organic link have an attribute called "dirtyhref", which contains the Yahoo tracking server URL. For the destination URL example shown in Fig. 2, the source code in a Yahoo result page's HTML file would be:

```
<a id="link-x" class="yschttl spt" data-bk="xxxx.x"
target="_blank" dirtyhref="/r/^^^^^/**http%3a//
buy.com/%3fid=2%26ref=30"
href="http://buy.com/?id=2&ref=30">
```

Figure 5: HTML source code for the organic link of the example destination URL in a Yahoo search result page

A left or right click on a Yahoo organic link triggers a JavaScript event which substitutes the destination URL stored in the "href" attribute with the Yahoo tracking server URL stored in the "dirtyhref" attribute. Fig. 6 shows the Yahoo tracking server URL generated for the above example. '^^^^^' represents an arbitrary ASCII strings. Compared with Google, Yahoo does not encode characters such as '/' and '=' in the original destination URL.

```
http://search.yahoo.com/r/^^^^^RU=http%3a//
buy.com/%3fid=2%26ref=30
```

Figure 6: Yahoo Tracking URL generated by user's click on the example organic link shown in figure 2

B. Click-Tracking Methods Used for Sponsored Links

All three search engines track user's clicking on their sponsored links. We find out that all of them use the similar encoding algorithm (with small variations) to encode destination URLs into their sponsored URL links. We introduce their encoding algorithms in this section.

For Google search engine, the URL of a typical sponsored link is illustrated in Fig. 7. A sponsored link usually contains three sections of ASCII strings. The first section has a long ASCII string "^^^^^" that is possibly used by Google to track additional information of a user. The second section starts with keyword string "&adurl=". This section is the encoded version of the target URL (final destination website). This section allows Google's tracking server to obtain the URL of the target website and redirect the user's HTTP request to the target website after click-tracking.

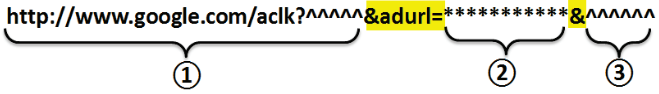


Figure 7: Three-section structure of a Google sponsored link. '^^^^' represents an ASCII string including '&' and '='; '****' represents an ASCII string excluding special character set {# & % [ ] + { } | ? =}

The third section starts from the first appearance of character '&' after "&adurl=". This section is optional; some Google sponsored links have it while others do not. The third section contains tags that are used by Google in a similar way as the query tags added by Google on its query URLs.

For the second section shown in Fig. 7 (the encoded target website URL) the encoding method is the same method as used in the Google tracking URL generated by 'onmousedown' JavaScript event in organic links (as shown in Fig. 4)—the only difference is that the encoding special character set is {# & % [ ] + { } | ? =} without ':' and '/'. Fig. 8 illustrates how Google encodes the example destination URL shown in Fig. 2 to generate the second section of its sponsored link.

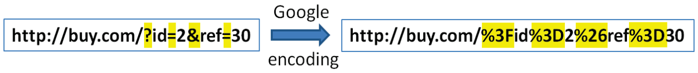

http://buy.com/?id=2&ref=30
Google
encoding
http://buy.com/%3Fid%3D2%26ref%3D30

Figure 8: Encoding of the example destination URL in the second section of a Google sponsored link

Bing search engine implements its sponsored links in a very similar way to Google. For the same example destination URL shown in Fig. 2, Bing's sponsored link URL is:

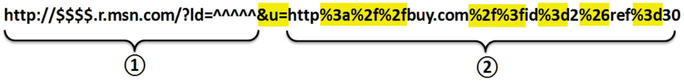

http://$$$$.r.msn.com/?ld=^^^^^&u=http%3a%2f%2fbuy.com%2f%3fid%3d2%26ref%3d30
①
②

Figure 9: Illustration of the example URL in a Bing's sponsored link

A Bing's sponsored link contains only 2 sections. In section 1 '$$$$' represents a numerical number. Section 2 is separated from Section 1 by the keyword "&u=" instead of "&adurl=" used in Google. In addition, the special character set encoded by Bing in a destination URL is identical to the set used in Google's organic link: {/ : # & % [ ] + { } | ? =}.

For Yahoo search engine, the sponsored link URL for the example destination website is:

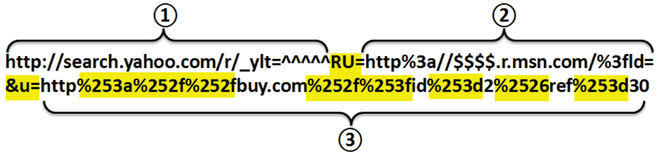

①
②
http://search.yahoo.com/r/_ylt=^^^^^RU=http%3a//$$$$.r.msn.com/%3fld=
&u=http%253a%252f%252fbuy.com%252f%253fid%253d2%2526ref%253d30
③

Figure 10: Illustrate of the example destination URL in a Yahoo's sponsored link. The keywords separating the three sections are '/**' and '&u='

Microsoft has taken over the search advertising operation of Yahoo and formed a *Yahoo! and Microsoft Search Alliance* [16]. Therefore, a Yahoo's sponsored link encapsulates Bing's sponsored link as shown in Fig. 10. One unique feature of the Yahoo sponsored link is that in its encoding of a destination URL, the character '%' is encoded again by its ASCII value of 0x25. That's why all encoded characters in the third section of a Yahoo sponsored link always start with '%25'.

IV. Proposed Click-Tracking Blocker

After analyzing the click-tracking methods used by the three major search engines, we have developed a browser plug-in program, called *Click-Tracking Blocker*, which can automatically replace organic links and sponsored links on a search engine result page with the true URLs of their corresponding destination websites. In this way, a user's clicking action will not be tracked by search engines and at the same time her user experience of searching will not be affected. In this section, we introduce the decoding algorithms used in our Click-Tracking Blocker.

A. Click-Tracking Blocking for Organic Links

Among the three major search engines, Bing does not track user's clicking on organic links. For Google and Yahoo, it is easy to remove their click-tracking on organic links since each of their organic links, as shown in Fig. 3 and Fig. 5, contains the original destination website URL in the HTML file. Thus for Google's search result pages, we simply remove the "onmousedown" JavaScript event from the anchor elements of every organic link. For Yahoo's search result pages, we simply remove the "dirtyhref" attribute from the anchor elements of each organic link.

B. Click-Tracking Blocking for Sponsored Links

As analyzed in Section III, sponsored links in all three search engines always embed the true destination URLs without any encryption. This allows us to extract the destination URLs to avoid tracking by search engines.

The pseudo-code for extracting destination URLs from sponsored links in Google's SERPs is:

Algorithm 1: **Google Sponsored Link Extractor**
Function *SponsoredURL2TargetURL* (SponsoredURL)

```
1   startIndex = index of "&adurl=" from SponsoredURL
2   IF '&' exists after startIndex   //Section 3 exists in the URL
3       endIndex = index of '&' in SponsoredURL after startIndex
4   ELSE   // Sponsored link does not contain Section 3
5       endIndex = last index of SponsoredURL
6   encodeTargetURL = SponsoredURL[startIndex to endIndex]
7   stringArray[] = divide encodeTargetURL at regular
                        expression '%'
8   FOR all stringArray[] indexes
9       ASCIIValue = Hexadecimal value of first two characters
                        of stringArray[]
10      ASCIIChar = convert ASCIIValue to its one-byte character
11      tempString = replace first two characters of stringArray[]
                        with ASCIIChar
12      TargetURL = TargetURL + tempString
13  return TargetURL
```

For a sponsored link in Bing's search result pages, compared with Google's, the keyword before encoded destination URL is "&u=" instead of "&adurl=", and no section 3 exists. Thus "*Bing Sponsored Link Extractor*" is almost identical to Algorithm 1 above, except changing the keyword in Line 1 as "&u=" and removing the codes on Line 2, 3, and 4.

For a sponsored link in Yahoo's search result pages, Fig. 10 shows that after the keyword "RU=" there is in fact a complete Bing's sponsored link. Thus "*Yahoo Sponsored Link Extractor*" algorithm is:

1. URLstring = Sub-string of SponsoredURL after keyword "RU="
2. Replace all "%25" in URLstring to character "%"
3. Run "Bing Sponsored Link Extractor" using URLstring as the input

C. Prototype Implementation

We have developed a prototype of the proposed Click-Tracking Blocker as a Firefox browser plug-in program. Readers can download and try out the plug-in code from: http://www.cs.ucf.edu/~czou/clickTrackBlocker/

When a user uses any of the three search engines in a browser, this plug-in intercepts the source code (HTML file) of a search engine result page, modifies URLs and their anchor elements on the page according to algorithms described above, and then displays the page in the browser.

This HTML page modification process is transparent to end users. Since it does not change a search result page's outlook and layout, it does not affect a user's search experience.

To intercept the search result source code, on Mozilla's Firefox browser we utilize the augmented browsing extension "Greasemonkey" [18]. Greasemonkey is a platform that allows other program developers to customize the way webpages look and function on FireFox browser. It enables us to execute our JavaScript code on the search result document prior to displaying it to a user. Therefore, for our current prototype plug-in, a user needs to first install Greasemonkey extension, and then install our plug-in as a "User Script" contained within Greasemonkey. Fig. 11 shows the FireFox browser screenshots after installing the plug-in.

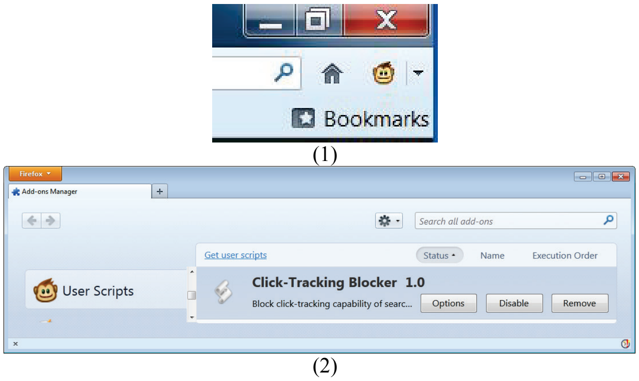


Figure 11: Prototyped Click-Tracking Blocker. (1) The monkey logo shows that Greasemonkey is enabled; (2) Click-Tracking Blocker runs as a "User Script" in Greasemonkey.

Greasemonkey is for FireFox browser only. Using the similar augmented browsing extension "Tampermonkey" [19] for Google's Chrome and "Trixie" [20] for Microsoft's Internet Explorer, we are able to implement the JavaScript code of our plug-in in these two major browsers as well.

V. Evaluation

We installed the prototype Click-Tracking Blocker on our computers and tested on many search result pages. Page layout, destination website links were not affected by the plug-in, and no click-tracking URL requests to search engines' servers have ever been generated from our computers. In order to further test the accuracy and effectiveness of the click-tracking blocker algorithms, we need to use a program to check tens of thousands of search result pages automatically.

A. *Standalone Code for Evaluation Purpose*

Extraction of organic links is straightforward since all organic links contain the actual destinations' URLs. Therefore we only need to test the performance of Click-Tracking Blocker in recovering/decoding sponsored links. For this purpose, we have developed a standalone application in Java that can run by itself without support of web browsers. Readers can also download this code from our website. This application repeatedly performs the following test:

- Sends search queries (a predefined list of common search phrases) using Google, Bing, and Yahoo search engines.
- Extracts sponsored link URLs in the resulting page with the assistance of *Jsoup* Java HTML Parser library [21].
- Executes the proposed click-tracking blocker algorithms to retrieve destination URLs from sponsored links.
- For each sponsored link, retrieves the destination website page source code through the newly extracted destination URL and through the original sponsored link URL, respectively.
- Uses Java *DiffUtils* library [22], calculates the percentage difference between these two webpage source code files.

B. *Evaluation Results*

In our experiments, the standalone application tested our click-tracking blocker algorithms in sponsored link URLs obtained from each search engine until it has successfully compared the obtained webpage contents from 10,000 sponsored links.

During the testing some sponsored link URLs cannot be tested because our standalone application cannot obtain the destination webpage content. By further analyzing these problematic sponsored links, we find three main causes. First, some of these problematic websites were either offline or have wrong URL during our tests. Second, some did not support the HTTP protocol used by our JavaScript in the request message. Third, some URLs were third-party advertisement websites that used special and complicated JavaScript redirect. In this case, the standalone application was unable to reach the destination URL because, once it reached the third-party advertisement server, Java did not provide a JavaScript environment in which the redirection code could be executed. In our experiment, 7% of the Google tests, 19.4% of Bing tests, and 18.2% of Yahoo tests were inconclusive because these sponsored links on SERPs were unreachable.

Based on the percentage of webpage content differences calculated by Java *DiffUtils* [22] in our program, Table 1 lists the difference distribution for the 10,000 successfully-tested sponsored links for each search engine, respectively.

Table 1: Distribution of webpage content difference among 10,000 sponsored links

Difference	Google	Bing	Yahoo
0%	36.44%	40.93%	29.11%
< 1%	44.01%	40.67%	53.68%
1 ~ 2%	4.79%	4.67%	3.53%
2 ~ 3%	2.51%	2.48%	2.12%
3 ~ 4%	3.03%	2.54%	1.07%
4 ~ 5%	3.44%	1.15%	0.99%
5 ~ 6%	1.44%	1.37%	1.11%
6 ~ 7%	0.77%	0.57%	0.52%
7 ~ 8%	0.87%	0.80%	0.72%
8 ~ 9%	1.07%	1.58%	0.84%
9 ~ 10%	0.72%	0.86%	0.50%
10 ~ 100%	0.85%	2.33%	5.77%

From Table 1, we can see that for all three search engines, more than 80% of webpages are either completely identical or have less than 1% differences. Thus our Click-Tracking Blocker has satisfactory performance in not changing user's search experience.

In addition to testing the effectiveness of destination URL extraction algorithms, the standalone program also collects information about the characteristics of all tested 10,000 sponsored link URLs. Figures 7, 9 and 10 show that sponsored link URLs in all three search engines always start with the corresponding search engine's webserver together with some tracking information, i.e., ASCII strings '^^^^^' in Section 1 in these three figures. We suspect that a search engine could track more client information if this Section 1 is longer. Fig. 12 shows the average length and its standard deviation of Section 1 of Google, Bing and Yahoo sponsored link URLs.

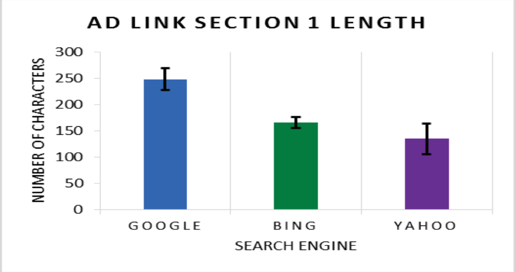


Figure 12: Average number of characters (and its standard deviation) in Section 1 in sponsored link URLs

The sponsored links in Bing and Yahoo always end with the encoded destination URLs as shown in Fig. 9 and 10. On the other hand, as shown in Fig. 7, a Google's sponsored link may or may not have an additional string (Section 3) after the encoded destination URL. In the following we analyze the structure of Section 3 when it exists.

Through analyzing the 10,000 sponsored links obtained from Google SERPs by our standalone program, we find out that overall 14.87% links do not have Section 3. But when Section 3 exists, it could possibly contain three different parameters as:

http://www.google.com/aclk?^^^^^^^^^^^^^^^^^^^^&rct=j&q=<search term>&ctype=27

Figure 13: Structure of Section 3 when it exists in Google sponsored links

The parameter "q=<search term>" shows the original search term in generating this search result page. The distribution of these parameters is summarized in Table 2.

TABLE 2: DISTRIBUTION OF THE THREE PARAMETERS IN SECTION 3 IN GOOGLE SPONSORED LINKS

Different Cases	**Percentage Among 10,000 Links**
Has 2 parameters: &rct=j&q=<search term>	99.49%
Has all 3 parameters: &rct=j&q=<search term>&ctype=27	0.5%

VI. CONCLUSION

Many people do not know that all three major search engines, Google, Bing and Yahoo, track which links on their search engine result pages (SERPs) users have clicked, even for organic links that have nothing to do with the ads business. Click-tracking is essential for search engines' advertisement venue for their sponsored links, and is important to help improving personalized search result pages for different users in terms of those organic links. However, such click-tracking could also be a big privacy concern for many Internet users. We discovered that all three search engines have used a similar approach to encode destination website URLs into their organic and sponsored links. Based on this observation, we have developed a browser plug-in called Click-Tracking Blocker, which could effectively prevent search engines from tracking users' clicking behaviors on SERPs on both organic links and sponsored links. The browser plug-in is transparent to users and does not affect users Internet search experience.

REFERENCES

[1] Search Engine Market Share Data. [online] http://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4

[2] Ryan Singel. "Yahoo Gives Up, Turns Search Over to Bing". July 29, 2009. [online]http://www.wired.com/business/2009/07/yahoo-gives-up/

[3] Organic linking. [online] http://en.wikipedia.org/wiki/Organic_linking

[4] Y. Zhu, G. Wang, J. Yang, D. Wang, J. Yan, J. Hu, Z. Chen, "Optimizing search engine revenue in sponsored search", *ACM SIGIR conference on Research and development in information retrieval* (SIGIR), pp. 588-595, 2009.

[5] Meghan Kelly, "96% of Google's Revenue is advertising, who buys it?", [online] http://venturebeat.com/2012/01/29/google-advertising/

[6] Mary Brandel, "What search engines store about you", July 20, 2007. [online] http://features.techworld.com/networking/3550/what-search-engines-store-about-you/

[7] Omer Tene, "What Google Knows: Privacy and Internet Search Engines", Utah Law Review, Vol 2008, No 4, p. 1433-1492.

[8] P. Eckersley, S. Schoen, K. Bankston, D. Slater, "Six Tips to Protect your Search Privacy", September 14, 2006. [online] https://www.eff.org/wp/six-tips-protect-your-search-privacy

[9] Saul Hansell, "Marketers Trace Paths Users Leave on Internet", *New York Times*, Sept. 15, 2006. [online] http://www.nytimes.com/2006/08/15/technology/15search.html

[10] M. Barbaro, T. Zeller, Jr., "A Face Is Exposed for AOL Searcher No. 4417749," *New York Times*, Aug. 9, 2006. [online] http://www.nytimes.com/2006/08/09/technology/09aol.html

[11] D.C. Howe, H. Nissenbaum, "TrackMeNot: Resisting Surveillance in Web Search", in I. Kerr, V. Steeves, C. Lucock (Eds): *In Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society* (2009), pp. 417-436.

[12] Greasemonkey. [online] http://www.greasespot.net/

[13] Google Anonymizer. [online] http://userscripts.org/scripts/show/10448

[14] Straight Google. [online] http://userscripts.org/scripts/show/121261

[15] J. Huang, R.W. White, S. Dumais. "No clicks, no problem: using cursor movements to understand and improve search". In *SIGCHI Conference on Human Factors in Computing Systems* (CHI), pp. 1225-1234, 2011.

[16] Yahoo! and Microsoft Search Alliance. [online] http://www.searchalliance.com

[17] RFC3986: Uniform Resource Identifier (URI): Generic Syntax [online] http://tools.ietf.org/html/rfc3986

[18] Greasemonkey. [online] http://www.greasespot.net/

[19] Tampermonkey. [online] https://code.google.com/p/tampermonkey/

[20] Trixie. [online] http://trixie.software.informer.com/

[21] Jsoup Java HTML Parser. [online] http://jsoup.org/

[22] DiffUtils. [online] http://code.google.com/p/java-diff-utils/