

Increasing Patients' Privacy While Using Search Engines and Social Media Websites

Ahmad Sleiman Ahmad¹, Obadah Hammoud², Mothanna Alkubaily³

¹ Al-Andalus University, Department of Medical Engineering, Alkadmous - Tartous, Syria

² Tartous University, Department of Information and Communications Technology, Tartous Syria

³ Tishreen University, Department of Mechanical and Electrical Engineering, Lattakia, Syria

Abstract: *Search engines and social media websites offer us great services, but they collect their users' information, which could be considered as a privacy violation. It requires some serious work to increase user's privacy level while using search engines services or surfing social media websites, and especially for patients, whom records are considered confident, and these websites are capable to create a medical profile for each of them. Our contribution in this paper is to propose a new method that enables us to increase users' privacy level while using existing search engines and social media websites without replacing them, in a way that disables those websites from accessing users' information.*

Keywords: Search Engines, Social Media, Google, Facebook, Privacy, Patients, Personal Data

1. Introduction

Search engines are considered as the main mean that enables users to access websites and different URLs across the internet. However, user's privacy might be in danger [1], especially if we are talking about medical information and patients records that are classified as secret data. It is possible to track patients and doctors search results about a specific disease or a specific medicine and its usage and side effects, and create a medical log for each user. The same thing can be said about social media websites, because they are able to collect information about each user in every single search process [2] he does, and every link he clicks on. This research aims to improve users' privacy level by filtering those websites dynamically. We will consider Google [www.google.com] as an example.

We've studied Search engines and social media ways of collecting user's sensitive information, it can create a complete log file for each user [3], like personal information (ex: places that user has accessed these websites from), his search history, and links he clicks on that implies on both search engines and social media websites [4]. Therefore, these search engines and social media website are able to create a medical log for all their users.

This research has been applied programmatically using PHP language, which is a free, efficient, and easy to learn language, used to develop web applications, and used by big companies like Google and Facebook [www.facebook.com]. We've used PHP v5, and [Byethost.com](http://www.byethost.com) [www.byethost.com] as a free webhost site.

The reminder of this paper is organized as follows: In section II, we reviews the related works. In section III, we present our proposal that aims to enables us to increase users' privacy level while using existing search engines and social media websites without replacing them. In section IV, we describes the system model and the studied performance parameters.

Finally, conclusion and future work are presented in section V.

2. Related Works

2.1 Search engines ways of collecting data about users

We can summaries search engines work mechanism [5] as the following:

- 1) User inserts his search query.
- 2) Search engine displays results.
- 3) User picks one of these results, and the website redirects him to the desired URL.

So, the search engine can collect data when the user interacts with it, that means, when he inserts the query, and clicks on the link.

2.2 Current available methods to beat the privacy violation problem

There are numerous ways of increasing user's privacy while searching, like [DuckDuckG](http://www.duckduckgo.com) [www.duckduckgo.com], search engine, whom developers have announced that it doesn't collect information about users or search results [6]. Maybe it's considered as a solution for this problem, but it has removed some existing benefits, for example, Google is considered one of the best search engines in retrieving related results to the search query, thanks to its algorithms, that means with this solution, user might not have the desired results. Another solution is to encrypt the data [7-9], but this method causes data overhead, and search engines are still able to collect data if the cipher key is weak in some cases.

3. Our Proposal

This research proposes a new method to improve user's privacy while keeping search engines and social media websites services, by: 1. Hiding user's personal information 2. denying these websites from collecting data about users

instead of encrypting it, which solves the overhead and privacy problems.

3.1 Hiding user's personal information

The idea of this solution is to use a proxy software. Since that some of proxy servers are known, we will propose a new method to use proxy technique, by creating a new website that translates any given URL, reads it, and returns the content to the end user, so we will name it 'Proxy website'. Whenever a user sends a request, the website will read the URL and display the result to the user, thus, the desired URL will deal with the website's server instead of the user, which keeps the user identity anonymous.

To achieve that in PHP, the user should insert the URL in a text box or by writing the URL after the website's URL in the location bar (For example: www.TheProxyWebsite.com/TheDesiredUrl). Page's HTML tags must be modified before displaying it, so that links leads to our proxy website to keep the user in the proxy mode (for example: `` to ``). 'allow_url_fopen' must be enabled in PHP Settings. So we can open the desired URL in PHP, and show the result as a web page. We have uploaded our website to a hosting server that exists in another country and checked both Google and ipchicken [www.ipchicken.com], and they are showing the server's location/IP instead of the original ones.

3.2 Denying websites from collecting user's data

When the user searches for something and clicks on a link, the search engine saves that link as one of the links that this particular user has visited, as shown in Fig1.

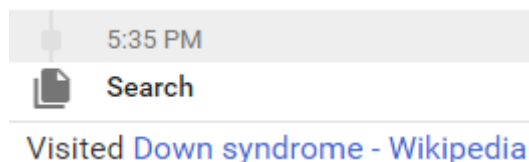


Figure 1: Visits log in Google

The reason why this link could be saved, is that the link doesn't direct the user to the requested URL directly, instead it redirects him to a new URL that belongs to the search engine company (where it gets saved), and then he gets redirected to the desired link, as show in Fig2.

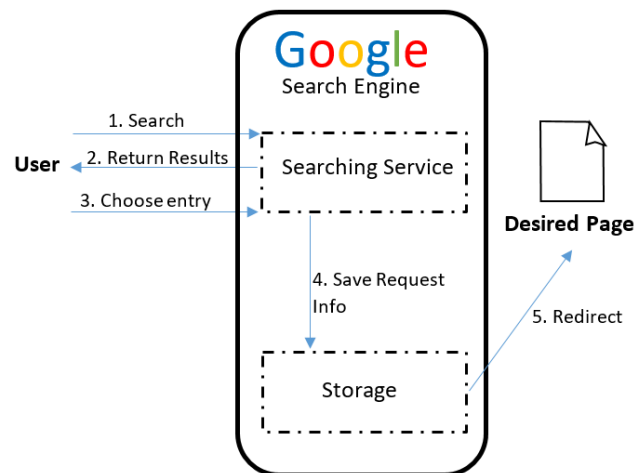


Figure 2: Redirections plan while searching in Google

We can check this theory by inspecting Google results page elements, where in HTML 'href' is responsible for the URL that the link links to, so if we want to redirect the user to www.drugs.com/diazepam.html, the HTML tag is written like this: ``

While in google, the desired link redirection won't be direct, for example, the URL is going to be like this (as shown in Fig3):

```
<a
href="/url?sa=t&rct=j&q=&esrc=s&source=web&cd=13&cad=rja&uact=8&ved=0ahUKEwiEzri5uefSAhUjJ5oKHYyCC30QFghXMAw&url=https%3A%2F%2Fwww.drugs.com%2Fdiazepam.html&usg=AFQjCNHw_NhL87acb113qOY0-7ffKGkRUw"
onmousedown="return
rwt(this,'','','13','AFQjCNHw_NhL87acb113qOY0-7ffKGkRUw','0ahUKEwiEzri5uefSAhUjJ5oKHYyCC30QFghXMAw','','event')"
data-
href="https://www.drugs.com/diazepam.html">
```

The same thing implies on social media websites with small differences, for example, let's consider Facebook, redirections which save the links don't replace the whole page, instead they use JavaScript, without moving actually to a new page, while the mobile mode (which doesn't use JS) uses the same Google strategy, by redirecting the user to a new page that saves information about him, and then redirecting him to the desired page, as shown in Fig4 and Fig5, (the desired link is www.cancer.gov/news-events/cancer-currents-blog/2017/platelet/immunotherapy).

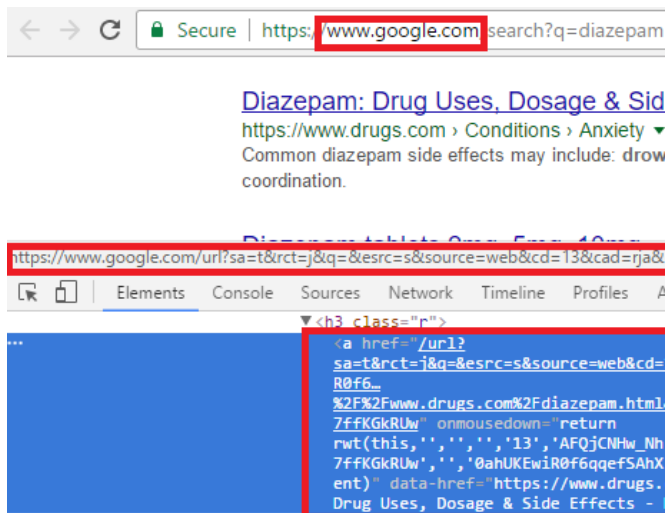


Figure 3: Links redirections in Google

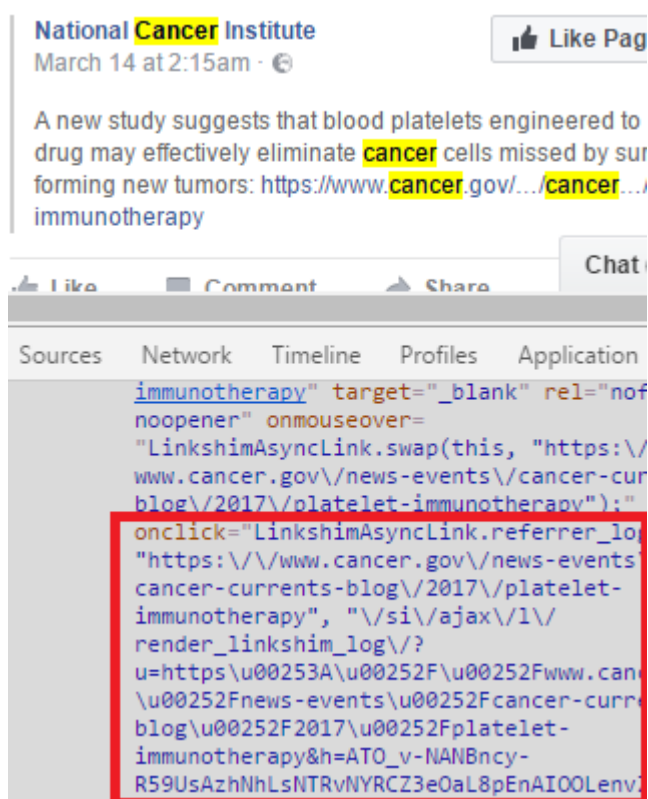


Figure 4: Following users' links with JavaScript in Facebook

This problem can be solved by creating a middleware that clones the results page, and filters it by deleting undesired links, and JS codes in Facebook case, and finally displaying the results to the end user, as shown in Fig 6.

To filter these links, we have to specify the redirection link pattern. After some experiments, we found that it takes the following pattern in Google case:

``

Where 'S' is a text string responsible of the redirection, and 'W' is the desired link. The middleware should convert the pattern to: ``

We can use multiple ways to achieve this, like Regex expressions.

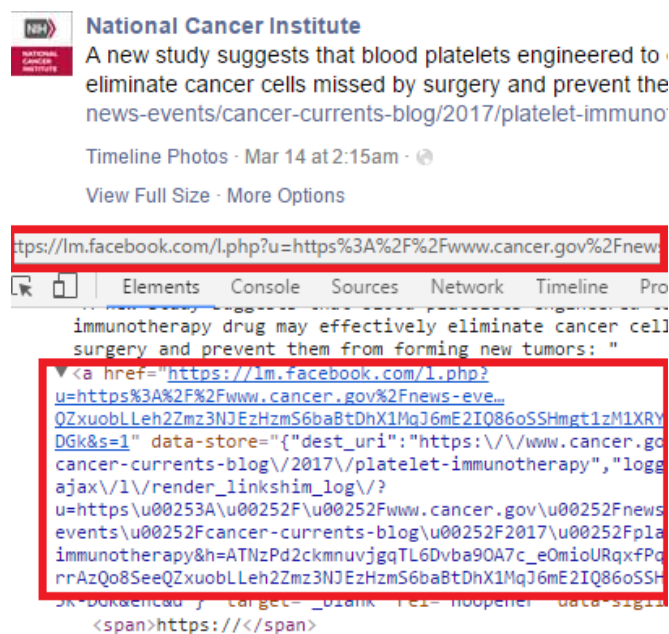


Figure 5: Following users' links with redirection in Facebook mobile mode

This idea has been achieved programmatically under PGoogle name, by making a new website that applies this idea, and uploading it to a free webhost server (byethost), and we used PHP as the programming language. Fig7 shows the work results. We can notice the difference by comparing with Fig3.

This application maintains multiple results pages, so choosing another page number or 'Next' shows the new filtered results page directly, and it has the same Google files distribution way, which enables it to be used as a default search engine inside web browsers. Another benefit of using this method is that it's executed as a 'Proxy website', and that protects other users' information which isn't related to search results, like IP, session info, etc., which improves confidentiality and users' anonymousness.

More privacy can be granted by disabling JS, or using the same way to filter JS codes.

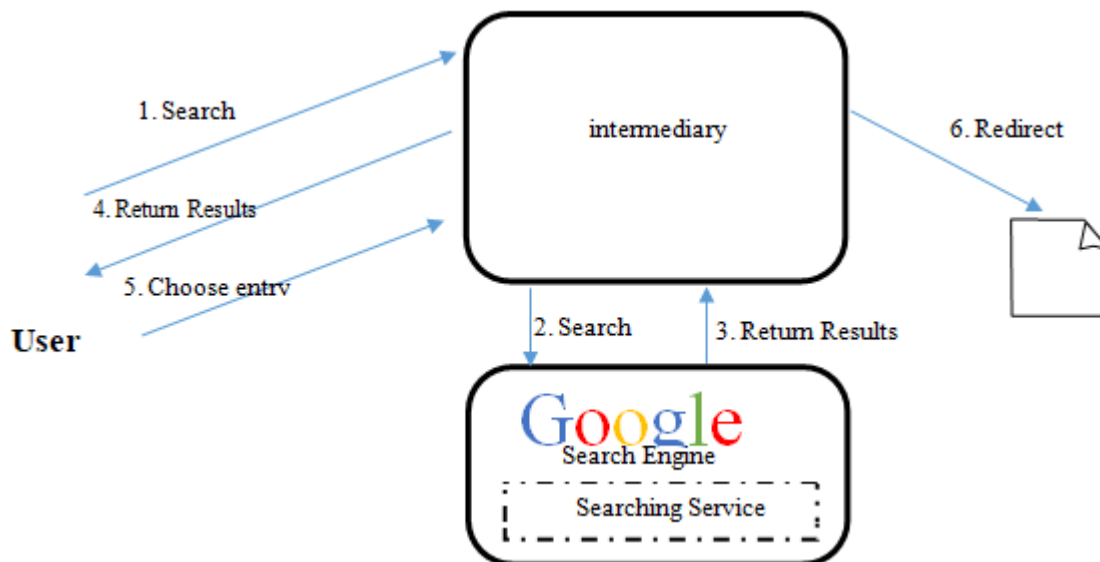


Figure 6: Filtering redirects plan

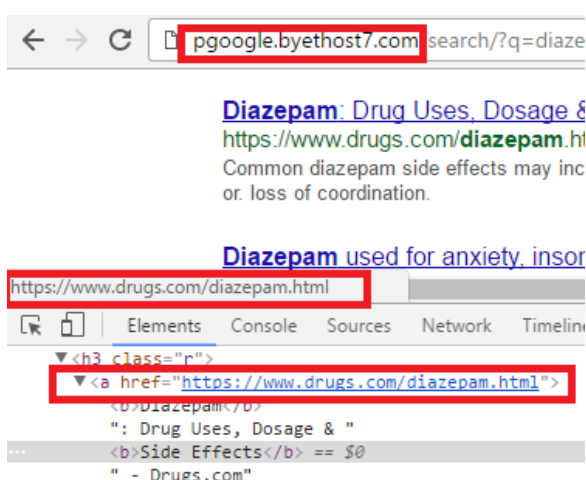


Figure 7: Links are direct in PGoogle

4. Results and Analysis

4.1 PGoogle Parameters study compared to Google

Results might change according to the used server. for faster results, localhost can be used, but it's better to use another server so it can play Proxy Server rule to increase confidentiality as mentioned before.

In this experiment, Byethost is used as a server (which is located in UK), and JS has been filtered.

We can expect the following:

- When using PGoogle, additional delay will show up because of filtering the results, while it saves loading JS time.
- When clicking on a link, PGoogle will show less delay, because it doesn't redirect the user to save request information, instead it redirects him to the requested link directly.

4.2 Simulation Setup and Results

We have tested 10 search processes for the same word (Google), on both Google and PGoogle, using the same PC in the same conditions (web browser, OS, same hour), and

used Google chrome inspector to show delays, the results where the following (in seconds):

Table 1: Time delays in both Google & PGoogle

PGoogle		Google	
Redirect	Request	Redirect	Request
3.74	4.99	5.77	5.33
2.76	3.27	3.01	5.09
3.45	4.1	5.68	5.33
3.03	4.08	3.21	5.58
3.08	4.32	3.64	5.06
4.28	6.46	4.1	5.13
3.06	5.58	4.36	4.91
2.96	4.12	6.28	5.05
2.93	6.2	3.95	5.68
2.73	4.18	5.48	5.33
Average	Average	Average	Average
3.202	4.73	5.548	5.249
Final Average		Final Average	
3.966		5.3985	

It's noticeable that there's no big difference, because that the extra work in Google (redirecting and saving requests), and PGoogle (filtering) doesn't lead to a big delay, and these changes differ from a request to another. We can say that PGoogle is faster than Google, but it might not be noticeable to the end user.

While using PGoogle, user is located in the hosting servers' location instead of user's location, that means the server is playing proxy server role that leads to additional confidentiality.

5. Conclusion and Future Work

This research proposes a new method to improve user's security level and privacy while using search engines or social media websites to prevent these websites from collecting users' data especially patients' medical information, and diseases they might have, and we approved programmatically the following:

- 1) Improving privacy and confidentiality while searching because the application is on another server which leads to a Proxy Server.
- 2) Improving privacy and confidentiality while clicking on links, because the redirects that save request's parameters are filtered.
- 3) Delay improvement, because of loading less files (by filtering external JS files calls) and filtering undesirable redirections.

As a result, we can use existing search engines and social media websites, with a higher level of privacy and confidentiality, and a smaller time delay.

References

- [1] Shu-Sheng Liawa and Hsiu-Mei Huang. "An investigation of user attitudes toward search engines as an information retrieval tool". Computers in Human Behavior, Volume 19, Issue 6, Pages 751–765. November 2003.
- [2] Hassan Aljifri and Diego Sánchez Navarro "Search engines and privacy". Computers & Security, Volume 23, Issue 5, Pages 379–388, July 2004.
- [3] Omer Tene. "What Google Knows: Privacy and Internet Search Engines", College of Management Academic Studies, October 2007.
- [4] Patricia Sánchez Abril, Avner Levin and Alissa Del Riego. "Blurred Boundaries: Social Media Privacy and the Twenty-First-Century Employee", American Business Law Journal, Volume 49, Issue 1, Pages 63–124. March 2012.
- [5] W. Bruce Croft, Donald Metzler and Trevor Strohman. "Search Engines: Information Retrieval in Practice". W.B. Croft, D. Metzler, T. Strohman. 2015.
- [6] Africa Hands. "Duckduckgo". Technical Services Quarterly, Volume 29, Issue 4, pages 345-347. Sep 2012.
- [7] Michael A. Halcrow, Dustin C. Kirkland, David B. Kumhyr and Kylene J. Smith. "Method and apparatus for conducting a confidential search". Patent US 7552327 B2, International Business Machines Corporation, USA. Jun 2009.
- [8] Jordi Castellà-Rocaa, Alexandre Viejoa and Jordi Herrera-Joancomartí. "Preserving user's privacy in web search engines". Computer Communications, Volume 32, Issues 13–14, 17, Pages 1541–1551 August 2009.
- [9] Albin Petit, Sonia Ben Mokhtar, Lionel Brunie and Harald Kosch. "Towards Efficient and Accurate Privacy Preserving Web Search". Proceedings of the 9th Workshop on Middleware for Next Generation Internet Computing, ACM, December 2014.

Author Profile



Obadah Hammoud is currently a Teaching Assistant at the Information Technology Engineering Department in Tartous University- Faculty of Information and Telecommunications Technology Engineering- Tartous, SYRIA.



Ahmad Sleiman Ahmad is currently Assistant Professor in the Biomedical Engineering Department at Al-Andalus University, Alkadmous, SYRIA.. He received the D.E.A. degree in Industrial Control System and Human from the Valenciennes University-France and the PhD degree in Electrical Engineering from Ecole Centrale de Lyon- France in 1992. His research interest includes Reliability, Modelling and Simulation in electrical engineering



Mothanna Alkubaily is currently Assistant Professor in the Communication and Electronic section at Mechanical and Electrical Engineering Department, Tishreen University, Lattakia, SYRIA. He received the M.S. degree and PhD degree in Application-Level Multicast Protocols from the University of Technology of Compiegne (UTC, France) in 2006 and 2009 respectively. His research interest includes Wireless Sensor Networks, MANET, VANET and Application-Level Multicast.