Abstract:

Search engines like Bing, Google, and Yahoo! are homes to trillions of pages of information. The problem that most users encounter when searching information on the web is not the quantity of information but it is the quality of the information. With a single search, it is possible that one can come across anything from good and informative sites to malicious and misinformed site. Studies show that users in the US alone land on malicious websites about 285 million times per month by clicking on search results from some of the major search engines. With those kinds of possibilities, it is imperative that we figure out how users can protect themselves when searching the web. In this research, I describe advantages and disadvantages of web portals. I also document any incidence of malicious/misinformed sites while comparing existing tools for web site screening. This research is important because it makes contributions towards identifying specific circumstances in which users encounter malicious/misinformed websites. It also identifies ways we can use existing tools to efficiently avoid malicious websites.

1. Introduction

The vast amount of information on the internet makes it easy for users to find malicious or misinformed web pages. Malicious and misinformed websites can be very harmful to users. Not only can they mislead users by spreading false and opinionated information, these sites could also be host to cyber threats like malware and phishing. It is reported that US users land on malicious websites on average about 9.5 million times per day [REF]. That is why it is important to have an adequate way for users to detect and protect themselves against these harmful web pages.

In this research, I considered three approaches to come up with an efficient way for users to protect themselves against malicious/misinformed sites. First I developed multiple case studies to test for incidents of malicious/misinformed websites. Using different queries on the same search engine, I examine how different user search queries can produce bad results. Secondly, to determine whether problem results are search engine specific, I conducted an investigation of the advantages and disadvantages of different search engines by using the same queries on three of the most popular search engines. Lastly, I use multiple web browsers to compare existing screening tools employed to protect users against harmful sites. The use of multiple web browsers to conduct the same searches is vital in determining whether browsers have an effect on misinformed search results. This research is unique because there are no comprehensive studies that use these approaches to find solutions to this problem.

2. Background

Since 1995, search engines have been the tool we used to gains to volumes of information from the world via internet. For example, Google went from a Stanford research project to one of the world’s major search engines [REF]. According to research group comScore, Google handled from 10,000 and 500,000 searches a day in 1998 and 99 to about 235 and 496 million searches a day in 2008 and 2009 respectively [REF]. As reported by McAfee Inc. in 2006, of all those searches over 181 million Google results per year lead to malicious websites. As stated by Juan Carlos Perez of the International Data Group, “search engines deliver links to dangerous Web sites”. It is important to know that because search engines accounts for 50% of all website
accesses. With the problem presented, several have attempted to address the issue with different approaches.

One approach cautions users to be on guard for misinformation on the web that detail issues of opinion rather than fact. Paul S. Piper, a Librarian from Western Washington University, points out that users can encounter malicious and misinformed websites by simply not paying close attention to a site’s URL [REF]. He maintains that any possible misspelling or change in the domain name could result in the access of counterfeit, fictitious, or malicious websites. Basically his approach is to educate users on the different signs to pay attention to while accessing different websites. Realistically, users are not always vigilant and sometimes it’s not easy to be vigilant of misinformation when using search engines. Most of the time, a user never recognizes or have not come across some of the websites that come up when conducting a search on a search engine.

Another approach is assessing the credibility of a website prior to accessing the page from the search engine. BJ Fogg and Hsiang Tseng don’t necessarily address malicious and misinformed websites, but their concentration on the concept of Computer Credibility is a foundation that researchers in the area of internet protection can build on [REF]. Many would consider being able to find the credibility of anything on the web is the first protection from misinformed sites. Although Fogg and Tseng don’t provide a checklist for how to check for credibility, they go to great lengths to explain the key concepts that relate credibility and they provide new ways to understand credibility [REF]. But, the downfall when trying to check for credibility of search results is that people make errors when trying to assess the “believability”. Users either perceive items that are not credible to be believable, or they just don’t trust the items that are credible [REF].

My approach is unique and overcomes the drawbacks because it examines how we can use the existing tools to do the work for us. I examine how we can efficiently use tools available to avoid the problems when assessing the credibility of different websites and to overcome the fact that most websites are not familiar to us, which means we cannot examine the URL.

3. Approach / Methodology

My approaches are aimed towards trying to figure out where the problem lies and what tools we need to employ to counter that problem. To examine how different user search queries can produce bad results, I conducted two academic searches that range from conservative to controversial. My intent was to determine whether a more contentious search query would produce more misinformed or malicious results. Throughout the experiment I use these 5 categories to classify my search results: relevant or not relevant, malicious/misinformed or not malicious/misinformed, and commercial /sponsored. An item was deemed relevant if it provides general, informative, and unbiased material pertaining to the subject. A search result was considered irrelevant but not malicious if provided unbiased and non-opinionated information that was not pertaining to the search topic. I characterize results as malicious/ misinformed if the content was biased, opinionated, or designed with a negative intent to get users to believe one aspect of a subject. Finally, I included a commercial/sponsored section just to keep up with how much advertised information is relevant or irrelevant.

My second approach was to consider the search engines. I used the same two queries on three different search engines. Studies have shown the people rarely look at results past the first two pages so I only observed the first 20 results including all of the sponsored ads. This helped me determine if a particular search engine did a better job than another with screening out
unnecessary information. While gathering separate calculations for each search engine, I calculated the percentage of information that fell into each category. I also reflected on which engine provided the most sponsored results and what percentage of it was irrelevant.

Because web browsers are our window to the vast amounts of information on the web, it is important to examine them when detecting and protecting against malicious or misinformed web pages. For my last approach I analyzed the role of the web browsers. To gather whether the browser was at fault for the access of the misinformed sites, I repeated the first two approaches on different 3 different web browsers. When it comes to security, most browsers focus on protecting against two general threats: malware and phishing attacks. But all web browsers use unique features to help users to protect themselves from harmful web pages. As a result, I summarized all of the tools that are available with each web browser. Not only did I summarize the available options, but I also categorized each feature to help evaluate which browser would be more efficient to protect users against the presented cyber threats.

4. The Experiment Search
   4.1 Resources Used:

   For the search engines, I choose the top three that comScore reported as the most used. Google, Yahoo, and Bing (one of the Microsoft sites) account for about 93 percent of the total searches in 2009 [REF]. Google was the most used engine by United States users with 65.0 percent of the searches conducted, followed by Yahoo! with 19.6 percent, and Microsoft Sites at 8.4 percent.

   I used the same tactics with the choices of web browsers. I decided to conduct the experiment with Google Chrome, Windows® Internet Explorer® 8, and Mozilla Firefox 3.6 because they are three of the most used web browsers. Below I include a summary of each browser and also a chart to characterize the features.

   **GOOGLE CHROME**

   Google maintains that "Chrome is designed to keep you safer and more secure on the web with built-in malware and phishing protection." It uses three core components to help protect from malicious and misinformed websites. The following list gives a summary of each component.

   1. Safe Browsing: An optional feature that shows a warning page if you encounter a website suspected of containing phishing or malware. Messages like those below appear on the screen with an option to proceed or turn away from the site:

      A. "Warning: Visiting this site may harm your computer!"

         - Google Chrome has detected that the site you're trying to visit may contain malware.

      B. "This is probably not the site you are looking for!"

         - This message comes up when the site you're trying to visit may be pretending to be another site. This is detected when Google notices that URL listed in the site's certificate doesn't match the site's actual URL.
C. >."The site's security certificate is not trusted!"

- This message means that the certificate wasn't issued by a recognized third-party organization. This is important because anyone can create a certificate.

2. Sandboxing: A security feature that is designed to automatically prevent malware from installing itself on your computer or using what happens in one browser tab to affect what happens in another.

3. Finally Chrome uses Auto Updating to ensure that your browsers is always current with the new security features and fixes.

**WINDOWS® INTERNET EXPLORER® 8**

Microsoft calls its latest web browser, Windows® Internet Explorer® 8, “a safer online experience”. They believe that it’s important for users to be aware of the cybercriminals and the methods they used to infiltrate your computer. They go to great lengths to present about three different ways you can be attacked while using the internet and they also explain how Explorer protects you against each attack. Even though Explorer focus on fighting against malware and phishing attacks like most web browsers, it maintains that its new and enhanced security features provide a level of protection like no other. Explorer also boast that it is one of the only browsers to include it own complete Anti-virus software ‘right out of the box.”

Internet Explorer uses the SmartScreen® filter to build on their phishing protection with added malware protection. Its function is to present a warning or blocking screen when you visit a malicious site or attempt to download a malicious program. This new software is designed to automatically protect against fake security warnings and websites that offer free games, movies or TV shows to trick you into installing malicious software. Although these features are activated by default, Microsoft included an option to choose your preferred security level. Furthermore, Internet Explorer introduces new, hard to spot threats like “cross-site scripting” and “click-jacking.” So when it’s all said and done Explorer use a combination of innovating technology and education to protect users from a host of cyber-threats.

**MOZILLA FIREFOX 3.6**

“Keeping your personal info personal and your online interests away from the bad guys” is the security promise of Mozilla Firefox. Firefox uses a series of methods like anti malware and phishing to protect against malicious/misinformed websites. These customizable methods are designed to warn you away from the bad sites and also explained to you why it isn’t safe to use. Unlike Internet Explorer, Firefox doesn’t come incorporated with its own anti-virus software, but it does “integrate elegantly with your Windows antivirus software”.

Firefox also offers Instant Web Site ID, which allows you to instantly check the identity overview of any website. Finally Firefox believes it’s the safest way to surf the internet because its open-source security strategy that allows it to find and fix security issues in record time.
Environmental Conditions

I used two search topics, which remained the same throughout the experiment. The first search topic is Stephen Crane, an American author. This search is considering my conservative search because Crane is not a contentious topic. Because this is an academic search, I consider search results with information about his works, biography, study topics, and academic papers relevant material. Material is considered not malicious if the information does not appear to be misleading or harmful to a user with academic intent. The second search topic is “abortion”. This is the chosen topic because it is a topic that is likely to cause disagreement and disputes between people with differing views. Material on abortion is considered relevant if it contains unbiased information like: facts, statics, information educating users about safe abortion methods and
Clinics, and educational materials about alternative methods to abortion. Advertisement for specific abortion clinics or methods would be deemed irrelevant but not malicious. Information would be considered malicious and misinformed to an academic study if it contains biased or opinionated material.

The second and third conditions are the search engines and web browsers. As stated before, I conducted the same search on the three search engines and repeated the search 3 more times on the different web browsers.

4.3 Experiment Results

<table>
<thead>
<tr>
<th>Results by Search Query</th>
<th>Stephen Crane</th>
<th>Abortion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Relevant</td>
<td>15%</td>
<td>49%</td>
</tr>
<tr>
<td>Malicious/Misinformed</td>
<td>0%</td>
<td>39%</td>
</tr>
<tr>
<td>Commercial/Sponsored</td>
<td>17.2%</td>
<td>32%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Engine Results</th>
<th>Yahoo!</th>
<th>Google</th>
<th>Bing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Relevant</td>
<td>38.8%</td>
<td>32.5%</td>
<td>28.3%</td>
</tr>
<tr>
<td>Malicious/Misinformed</td>
<td>13.8%</td>
<td>17.5%</td>
<td>19.1%</td>
</tr>
<tr>
<td>Commercial/Sponsored</td>
<td>30.8%</td>
<td>27.5%</td>
<td>18.3%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Web Browser Results</th>
<th>Internet Explorer</th>
<th>Google Chrome</th>
<th>Mozilla Firefox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Relevant</td>
<td>34.1%</td>
<td>35%</td>
<td>33.3%</td>
</tr>
<tr>
<td>Malicious/Misinformed</td>
<td>15%</td>
<td>15.8%</td>
<td>16.6%</td>
</tr>
<tr>
<td>Commercial/Sponsored</td>
<td>24.1%</td>
<td>24.1%</td>
<td>25.8%</td>
</tr>
</tbody>
</table>

4. Conclusions