

ENCLOSURE 1

14 June 2013

I am the defense computer forensics expert. Please see Enclosure 2 for my CV. The purpose of this declaration is to provide the court with additional information regarding web caches. I will address Google Cache and the Way Back Machine separately, although most of the information applies to both.

Both the Google cache and the Way Back Machine contain only information about one point in time. The Google cache lets you know the page you are viewing is only a snapshot of how this particular page looked at a specific time. See Enclosure 3. Google even provides a warning to this effect "The current page could have changed in the meantime." The Way Back machine keeps multiple copies of the same web page captured on different days. Although when you view the pages the date and time captured do not appear on the page that displays the cached page. For any of the cache services you are essentially looking through a window at a picture of what occurred.

The web is alive. Web Pages by their very nature are made to be easily manipulated and changed. HTML (Hyper Text Markup Language), the foundation of almost all web pages, is about as simple as programing gets. Nearly anyone with a half an hour to spare can create one.

A web page can contain anything at any moment. Addressing the Most Wanted list, items could have easily been added or removed from the list at any time. And they likely were. There is no place to get a 100% definitive answer about what was contained on the page at every minute.¹ This list is meant to change as world events occur. Or when new "concealed documents or recordings most sought after by a country's journalists, activists, historians, lawyers, police, or human rights investigators" comes to light. There could be times when the entire list fit on one page.

The page could have been hacked and replaced with a nasty message. Very few websites are safe. See Enclosure 4-6. The entity that stores either cache could have been hacked. The hacker could have replaced the cached version with one of his own design. The integrity of the pages provided to the Way Back Machine could have been compromised before they even made it to the Way Back Machine's servers. The times and dated provided to the Way Back Machine could have be inaccurate.

All of the technologies that allow a web page appear on your screen are volatile. Accessibility to a site at a given time is unknown. Servers that host a web page could have been taken offline manually and been inaccessible for a period. There are many other possibilities, a server being inaccessible could be due to weather, the server crashing, an internet link going down, a DoS (Denial of service) attack, server overload, bandwidth saturation, etc.

The news changes every day all day. The CNN home page is live; it's meant to contain the most recent data at a given moment. What you see when you visit CNN.com is a single snapshot. What you see is the page at the time it is viewed. In most cases if a change is made on a web page after the user loads

¹ It is, however, possible that someone from Wikileaks.org could testify about what their page looked at a given point in time.

the page it does not change. The user that loaded the page 5 minutes ago would have no idea about the changes that occurred 1 minute ago. The modern web has mechanisms for updating pages after they load so if these pages have these mechanisms enabled they can retrieve new data to update information displayed on the page.

It is important to know when evaluating the accuracy of a copy to know certain things about how the copy was created. What algorithm, routine or program, was used to grab the page off of the server? How does the algorithm change the page in the capture process? There are a few ways to do this. Some may be more accurate than others. A web page is made up of many elements that can include things like images and videos. The most basic element of a web page is the HTML file. HTML files describe the structure of how a page should look by linking and formatting the elements into what you see in your browser. Images and Videos are not contained in the same file on the server. As such, the manner in which the files are copied makes a difference. For instance, are the images linked by the main HTML file captured along with the HTML? In other words does the algorithm grab images when it grabs the page? If it grabs the images does it change the underlying HTML code to reflect a new location of those images so the page displays correctly? If you moved a web page and changed nothing about it, images may or may not display. This is because there are multiple ways for a web page to "link" images. If the algorithm changes the links in the web page it is not the same file loaded from the server initially. This is only one example of content that may be changed when caching a page. Are the ads cached? Does the algorithm just ignore those sections of a page when it grabs it? Are videos captured along with images?

It is also important to evaluate the integrity of servers and storage locations. If someone has changed the page in the cache would you know? Has the page been compared to a backup or a hash to ensure integrity? Is the cache the only back up or does the server that contains the cache/archive have properly maintained backups?

Any live server it is susceptible to hack attempt. Not just small websites but sites like Twitter. See Enclosures 7 and 8.

The Wikileaks page is formatted like other wiki pages. This includes the Most Wanted Lists. Encyclopedia Britannica describes a wiki as "wiki, World Wide Web (WWW) site that can be modified or contributed to by users. Wikis can be dated to 1995, when American computer programmer Ward Cunningham created a new collaborative technology for organizing information on Web sites. Using a Hawaiian term meaning "quick," he called this new software WikiWikiWeb, attracted by its alliteration and also by its matching abbreviation (WWW)." So not only can site administrators add content to wiki pages, but so can the general public. Wiki's are meant to be a collaboration space.

When I did a search on Google for "wikileaks 2009 most wanted list" the top result was for the non-sort version. See Defense Exhibit F. The second result was for the sort version. See Prosecution Exhibit 109. Normally, when a user visits the Wikileaks Most Wanted Page they will be presented with the non-sort version of the page. The user could visit either page, and there is no link on the normal page to the sort version.

Below are the URLs for the referenced Enclosures should the Court want to visit the websites hosting the referenced news accounts.

URL for Enclosure 4

<http://rt.com/usa/anonymous-hacks-state-department-617/>

URL for Enclosure 5

http://www.pcworld.com/article/248644/us_government_online_security_website_hacked.html

URL for Enclosure 6

<http://www.techradar.com/us/news/internet/cia-website-and-fbi-hacked-by-lulzsec-966715>

URL for Enclosure 7

<http://techcrunch.com/2009/12/17/twitter-reportedly-hacked-by-iranian-cyber-army/>

URL for Enclosure 8

http://www.huffingtonpost.com/2010/09/21/twitter-hacked-twitter-vi_n_733338.html



Trent Struttman
Cyberagents, Inc.

14 June 13