

Trent Garewal
5/10/2007
“Google Hacking”

Over the years, the term “hacker” has been used to categorize malicious intent in the form of compromising a target computer for monetary gain. While that definition may not be incorrect, it is also missing many parts that don't get talked about publicly through the media. Hacking in itself is just manipulating something to do what it was not intended to do. For example, Google hacking is the process of making complex search queries to Google.com's search feature to force Google to display things most people would not know how to find. People can also use these techniques to bring up pages that may have potential security holes and that are vulnerable to exploitation.

When Google was first created in 1998 by Larry Page and Sergey Brin, their main and shared goal was to be able to retrieve relevant information from a massive set of data (Google 1). What they did not know in 1998 was the amount of popularity Google would grow to. Today, Google is the largest open source search engine out there, processing over 112 million searches per day (1cog 1).

With Google having the largest index of the Internet there is bound to be pages accidentally thrown in there. Google's web crawler goes out and makes an index entry for just about every web page out there. Providing a complex search query can limit the search down to only hundreds of pages.

Google hacking was first really created around 2004. Most of the search queries involved having the code “inurl” which told Google to look for the following search only inside the web pages URL. With the inurl: command you could dig up parts of web pages that were not meant to be accessed standalone. For example, one of the most common Google hacking search queries thanks to Wikipedia is the “inurl:”ViewerFrame?Mode=” “ search. By searching this string, Google will find web pages containing “ViewerFrame?Mode=” inside the url of a web page. This happens to be the string for a popular web cam software. Google will turn up thousands of web pages, almost all of them just a link to a web cam being hosted over the internet.

While Google is now aware of such things, they have taken measures to counteract typical and obvious Google hacking. For example, after a few pages of searching with the “inurl:”ViewerFrame? Mode=” query, we are confronted with an error message from Google stating: “We're sorry... but your query looks similar to automated requests from a computer virus or spyware application. To protect our users, we can't process your request right now. “ A somewhat newer (2006) query shows SHOUTcast servers where you can search for music being sent over the Internet like a radio station. While the idea for a SHOUTcast server is to share music with people of similar interest sometimes they just weren't meant for anyone and everyone. The following: “(intitle:”SHOUTcast Administrator”)|(intext:”U SHOUTcast D.N.A.S. Status”)” does not produce the “We're sorry” error but only has about 185 results due to SHOUTcast fixing the problem from their end. When you search for that query you only get old version of the software that have not been updated.

The Google Hacking Database(ihackstuff.com 1) is a large resource database of search queries that yield exploiting results. They vary from intext:”This site is using phpGraphy” to show picture collections, to very blatant more complex attempts to find web servers with default phpmyadmin settings: “intitle:”AppServ Open Project *” "AppServ is a merging open source software installer package" -phpbb” Google is obviously working very well on stopping Google hacking. When you try to find a server using these methods to gain access to a phpmyadmin site Google won't even display the 1st 10 pages. It gives you the “We're sorry” error at the main page. While fewer Google hacking breakthroughs are being reported some old tricks still work. The newest addition to the Google Hacking Database is from 2006. Mainly, php was the most susceptible to Google hacking because it worked with lots of things in its URL You could easily find php errors, install directories, configuration files, and even administrator passwords. A few years ago, Google hacking worked very closely with SQL injections. People found their targets by the use of Google hacking.

Even simple searches without using complex arguments and commands can yield results that would compromise a company. Searching for “Host Vulnerability Summary Report” yields thousands

of PDF files with very detailed information on normally the company that is hosting the reports' vulnerabilities. Continuing to look through the Google Hacking Database shows that from 2005-2006 is when most of the entries were added to the database. Since then, the arguments to archive some of the results have been flagged by Google and some even temporarily ban your IP for up to 5 minutes. It looks to be some kind of flood ban since my IP has been doing Google hacking searches frequently for the past few days. I have found myself locked out of Google for a few minutes a couple times.

One of the most promising search query I have done recently that still works great is the “login: *” password: *” filetype:xls” search. This pulls hundreds of thousands of excel spread sheets that contain user name / password fields in them; normally for personal records. In just looking through them I have come across hundreds of passwords and user names for websites. Tons of housing companies and broker back end pages. While it looks to be after any complex search query I do, Google tends to stop me after page 10 or so with the infamous “We're sorry” error.

Oddly enough, some of the older obvious ones still turn up pages. For example, “intitle:"Retina Report" "CONFIDENTIAL INFORMATION” pulls up reports created by Retnia Network Security Scanner, which contain basically a network map with detailed information on vulnerabilities. The only downside basically is once the report is published chances are its already all fixed.

Some companies have noticed they are being targeted by Google hacking and found ways to circumvent it. In 2006 UPS packages could be tracked using the search query “site:ups.com intitle:"Ups Package tracking" intext:"1Z ### ## ## #### ## #” UPS has since then changed the way pages are created to make sure man-in-the-middle attacks such as these are not used against their customers. While some devices and companies either don't know about it yet, or don't care. Searching for “intitle:"Live View / - AXIS" | inurl:view/view.shtml OR inurl:view/indexFrame.shtml | intitle:"MJPEG Live Demo" | "intext:Select preset position” still pulls up thousands of web cams around the world for public view.

Sensitive information is still out there, but Google hacking has lost its ways. However,

performing complex search queries is still considered Google hacking. Informat has an article with simple tips to make your searching easier(Informat 1). For example, Google doesn't include common words in its searches like 'and' and 'the.' You can get around this by using the + or – keys in your search to add or take away things from your searches. If you are looking for something specific, double quotes always works and is often the best way of finding exactly what you want. As with most operating systems, the asterisk key (*) is used as a wildcard. If you forgot something or if you were looking for any of blank, that would be the best bet to find it.

Typical operators are also used to narrow your searches “Site:” tells Google to search within a certain site for something. This can be useful if you are looking for something in a huge website such as microsoft.com. “filetype:” tells google to only show things if your file type. This can be useful if you are looking for PDF's of your search or even only .gif images.

Sometimes if something is outdated or you cant find what you are looking for anymore, Google's web crawlers take snapshots of every website it visits. For this, you can have Google pull up its cached version of the website with “cache:” This can be useful if the website is down for some reason and you are looking for content on it, although pictures are rarely displayed. “Intitle:” and” inurl:” are typical searches that are normally used in the malicious kind of Google hacking. Using these operators sometimes leads to the “We're sorry” error.

There have been tools created to automate Google hacking. GNUcitizen created the Google Hacking Database Tool for just the thing. It is a web application that allows simple clicks to add and shape complex queries without fully knowing what they do. In essence, it has been described as the “scriptkiddies” Google hacking tool(Gnu 1). Oddly enough this tool still works. I would think Google would have blocked search queries to be made from this site unless it doesn't forward the information saying it was sent by gnuCitizen.org.

In protection against such programs, acunetix.com has a vulnerability scanner that scans your web page against the Google Hacking Database to make sure your site is not publishing anything it

shouldn't. While this program is not free, there is a trial version(Acunetix 1).

While Google hacking was a big deal a few years ago, I think it has lost its flare. Google is very aware of whats going on and they have done a lot to protect people from it. Google understands that its search engine is if not the most powerful engine out there. I think that there will still be new searches that will pop up, but companies are now in the know that even search engines are watching the way they run their network. I believe that the entire Google hacking idea will die down in a few years. In doing my research I have come to far too many road blocks and countermeasures to make it viable. Foot printing in itself is considered Google hacking, but anything other than that will eventually fall off as the newest information I can find on the topic is dated in 2007.

As security becomes more and more powerful and important people will realize these things and Google hacking will eventually be nothing more than someone who understands the complexity that Google allows you to search for.

Sources

“Search Engine Statistics” {Cog 1} Online Document. Retrieved on May 1st, 2008 from: <http://www.1cog.com/search-engine-statistics.html>

“Google Hacking Mini-Guide” {Informat 1} Online Document. Retrieved April 27th 2008 from: <http://www.informat.com/articles/article.aspx?p=170880>

“Google Hacking Prevention” {Acunetix 1} Online Document. Retrieved April 27th 2008 from: <http://www.acunetix.com/vulnerability-scanner/google-hacking.htm>

“Google Hacking Database Tool” {Gnu 1} Online Document. Retrieved April 28th 2008 from: <http://www.gnucitizen.org/ghdb/application.htm>