



Google: A Hacker's Best Friend

In the last few years a number of news articles appeared that warned of the fact that hackers (or crackers if you will) make use of the google search engine to gain access to files they shouldn't be allowed to see or have access to. This knowledge is nothing new to some people but personally I have always wondered how exactly a thing like this works. VNUnet's James Middleton wrote an article in 2001 talking about hackers using a special search string on google to find sensitive banking data:

"One such posting on a security newsgroup claimed that searching using the string 'Index of / +banques +filetype:xls' eventually turned up sensitive Excel spreadsheets from French banks. The same technique could also be used to find password files"[1]

Another article that appeared on wired.com told us how Adrian Lamo, a hacker who made the news often the last couple of years, explained that google could be used to gain access to websites of big corporations.

"For example, typing the phrase "Select a database to view" -- a common phrase in the FileMaker Pro database interface -- into Google recently yielded about 200 links, almost all of which led to FileMaker databases accessible online."[2]

These articles kept on coming up in the online news. U.S. Military and Government websites were vulnerable because admin scripts could be found using google, medical files, personal records, everything suddenly seemed just one google search away. But these articles seemed to show up once every half year and always talked about it as if it was something new. Another thing was, the articles never explained how one would actually go about doing this. Almost never an example of a search string was given. The last time I read one of these articles I decided it was time to find out for myself, whether google actually could do all they say it can. The following is a report of my findings and a description of some techniques and search strings one could use.

Theory

The theory behind this is actually quite simple. Either you think of certain data you would like to acquire and try and imagine in what files this kind of data could be stored and you search for these files directly. (Search for *.xls files for example) Or you take the more interesting approach and you try to think of a certain software that allows you to perform certain tasks or to access certain things and you search for critical files of this software. An example could be a content management system. You read up on this particular content management system, check out of what files it exists and search for those. A great example is that of the databases mentioned above, where you know the string "view database" is used on pages that shouldn't be accessible to you and you then search for pages containing that string, or you check the software and notice that the option to view a database is linked on a webpage within this software called "viewdbase.htm" and you search for "viewdbase.htm"

The most important thing is to have a clear goal, to know what it is you want to find. Then search for these specific files or trademarks that these files have.

Google Search Options

Specific file types: *.xls, *.doc, *.pdf *.ps *.ppt *.rtf

Google allows you to search for specific file types, so instead of getting html-files as a result (websites) you get Microsoft excel files for example. The search string you would use would be this:

Filetype:xls (for excel files) or *filetype:doc* for word files.

But maybe more interesting would be searching for *.db files and *.mdb files. Google by the way doesn't tell you you can search for *.db and *.mdb files. I wonder what other file types one can search for. Things that come to mind are *.cfg files or *.pwd files, *.dat files, stuff like that. Try and think of something that might get you some interesting results.

Inurl

Another useful search option is the *inurl*: option which allows one to search for a certain word one would want to be in the url. This gives you the opportunity to search for specific directories/folders, especially in combination with the "index of" option, about which I will talk later on.

An example would be *inurl:admin* which would give you results of website urls that have the word "admin" in the url.

Index of

The index of option is another option that isn't especially thought of by the creators of google, but comes in very handy. If you use the "index of" string you will find directory listings of specific folders on servers. An example could be:

'index of' admin or *index.of.admin*

which would get you many directory listings of admin folders. (don't forget to use the quotes in this case since you are looking for the entire "index of" string, not just for "index" and "of")

Site

The site option allows you to come up with results that only belong to a certain domain name extension or to a specific site. For example one could search for .com sites or .box.sk sites or .nl sites, but also for results from just one site, but more interesting might be to search for specific military or government websites. An example of a search string would be:

Site:mil or *site:gov*

Site:neworder.box.sk "board"

Intitle

Intitle is another nice option. It allows you to search for html files that have a certain word or words in the title. The format would be *intitle:wordhere*. You could check out what words appear in the title of some online control panel or content management system and then search google for this word with the intitle option, to find these control panel pages.

Link

The Link option allows you to check which sites link to a specific site. As described in Hacking Exposed Third Edition, this could be useful:

These search engines provide a handy facility that allows you to search for all sites that have links back to the target organization's domain. This may not seem significant at first but let's explore the implications. Suppose someone in an organization decides to put up a rogue website at home or on the target network's site."[4]

Combining search options

The above mentioned search options might or might not be known to you, but even though they can amount to some interesting results, it's a fact that when you start combining them, that's when google's magic starts to show. For example, one could try this search string:

inurl:nasa.gov filetype:xls "restricted" or this one: site:mil filetype:xls "password" or maybe

site:mil "index of" admin

(I'm just producing these from the top of my head, I don't know whether they'd result in anything interesting, that's where you come in. You got to find a search string that gets the results you want.)

Examples; The Good Stuff

Specific file types: *.xls, *.doc, *.pdf *.ps *.ppt *.rtf

To start out simple, you can try and search directly for files that you believe might hold interesting information. The obvious choices for me were things like:

Password, passwords, pwd, account, accounts, userid, uid, login, logins, secret, secrets, all followed by either *.doc or *.xls or *.db

This led me to quite some interesting results, especially with the *.db option but I actually also found some passwords.doc files, containing working passwords.

<http://www.doc.state.ok.us/Spreadsheets/private%20prison%20survey%20for%20web.xls>

<http://www.bmo.com/investorrelations/current/current/suppnew/private.xls>

http://www.nescaum.org/Greenhouse/Private/Participant_List.xls

http://www.dscr.dla.mil/aviationinvest/attendance_5Apr01.xls

http://web.nps.navy.mil/~drdolk/is3301/PART_IS3301.XLS

Admin.cfg

Admin.cfg is, most of the times, an admin configuration file of some sort. Many different software obviously use names like "config" or "admin" or "setup", etc. And most of the times these files contain sensitive information and thus, shouldn't be accessible for people browsing the web.

I tried a search for admin.cfg, using the following search string on google:

inurl:admin.cfg "index of"

This led me to many results of which many were useless. But some paid out. I found for example: <http://www.alternetwebdesign.com/cgi-bin/directimi/admin.cfg> Which contained a password. This was the admin password for a database located at <http://www.alternetwebdesign.com/cgi-bin/directimi/database.cgi?admin.cfg> This database contained sensitive client data of this particular company. I then proceeded to e-mail the company and tell them about the flaw. They replied to me in a very friendly manner and told me they appreciated my help and that they would take the necessary steps to solve the problem.

Webadmin

A short while back, while working on this article, I ran into this website:

<http://wacker-welt.de/webadmin/>

The website explains that "webadmin" is a small piece of software that allows one to remotely edit parts of a website, upload files, etc. The main page for the webadmin control centre is called "webeditor.php". So obviously, my next step was to visit google and use the inurl tag to find webeditor.php pages that I could reach. I used the following search string:

inurl:webeditor.php

and I found the following results:

<http://orbyonline.com/php/webeditor.php>
<http://www-user.tu-chemnitz.de/~hkri/Neuer%20Ordner/webeditor.php>
<http://artematrix.org/webeditor/webeditor.php>
<http://www.directinfo.hu/kapu/webeditor.php>

All these webeditor.php files were reachable by anyone, merely because the owners failed to (correctly) protect these pages by using .htaccess. This mistake allows whomever to change the webpages on the server and thus defacing the site, uploading files and thus possible gaining full access to the server.

In browsing through these sites I noticed that the file that allows one to upload files is called "file_upload.php", which I could then search for at google and find more examples.

http://www.hvcc.edu/~kantopet/ciss_225/examples/begphp/ch10/file_upload.php

A good example:

<http://www.pelicandecals.com/admin/webeditor.php>

The script allows you to change files, like in the above examples, including the index.php. In theory one could write or download whatever malicious script one wants, paste this code into an existing file or just upload it and well, the consequences are obvious.

there was also a link "[Return Administration](#)" and clicking on it took me to:

<http://www.pelicandecals.com/admin/administration.html>

Where there were customer addresses, where one could change pricing, etc.

Content Management Systems

Content Management Systems are software programs that allow a webmaster to edit, alter and control the content of his website. But the same goes for online control panels of websites.

The idea is to find out what files are for example the main files of these software programs. "cms.html" could be one or "panel.html" or "control.cfg" You find out what filenames a certain package uses, you then think of a good search string and hope you strike gold.

Frontpage Server Extensions HTML Administration Forms

"You can remotely administer the FrontPage Server Extensions from any computer connected to the Internet by using the FrontPage Server Extensions HTML Administration Forms, a set of Web pages that allow you to administer the FrontPage Server Extensions remotely.[3]

Well, that's what Microsoft's manual has to say about it. This means, users with access to these forms are able to perform a number of administrative functions, remotely. And that means, these forms should be well protected from non-authorized people. Now how would one go about finding non-protected forms over the internet? The first thing we do is try to find out what files these scripts consist of. A short visit to the Microsoft website or a peek into the frontpage manual tells us that the main page for these administration forms is a file called "fpadmin.htm". So that's what we need to search for. Now to find a correct search string that will get us the results we want. When a default install is performed, the files get installed in a directory called "admin". Putting to use what we have learned about google search options and the theory behind this technique, a good search string might be:

inurl:fpadmin.htm "index of" admin or maybe inurl:admin/fpadmin.htm

Well, these were the results I got:

http://www.lehigh.edu/~ineduc/degree_programs/tbte/admin/

<http://blackadder.eng.monash.edu.au/frontpage/admin/>

http://www.lehigh.edu/collegeofeducation/degree_programs/tbte/admin/

<http://www.vsl.gifu-u.ac.jp/freeman/frontpage4/admin/>

[http://www.tech-geeks.org/contrib/loveless/e-smith-fp-](http://www.tech-geeks.org/contrib/loveless/e-smith-fp-2002/frontpage/version5.0/admin/1033/fpadmin.htm)

[2002/frontpage/version5.0/admin/1033/fpadmin.htm](http://www.tech-geeks.org/contrib/loveless/e-smith-fp-2002/frontpage/version5.0/admin/1033/fpadmin.htm)

<http://fp.nsk.fio.ru/admin/1033/fpadmin.htm>

But the frontpage manual says more:

"Because of the security implications of making remote FrontPage administration possible from Web browsers, the HTML Administration Forms are not active when they are first installed." [3]

This means that some of these could be active and thus useful to us and some might not.

There is of course, only one way to find out and that is to perform one of the possible administrative functions and see if you get results. I for one decided not to go that far, because it would mean breaking the law. But I'm not here to teach ethics, or at least not today.

Freesco Router

The Freesco router software for Linux as a default, installs a small web browser which allows owners to control the router through the http protocol. In other words, a website automatically gets setup that allows you to control the router. The default password and login for this control panel is "admin" and "admin". Many people who use freesco don't know this. You could search for these Freesco router control websites by using a string such as:

intitle:"freesco control panel" or "check the connection" which are words that either are in the title of these pages or on the pages itself. That's what it's all about; you check out a certain software, find the part you'd want to be able to reach and figure out which search string would get you the good results.

Extra Tips

- Remember English is the most used language online, but it's not the only one. Try and search for words or strings that are specific to your language or French or German, etc. For example "beheer" is a Dutch word for "administration" or "privat" is German for "private".
- You can check vulnerability scanners' scan lists for interesting search strings you might want to use or combine with your own strings. Check <http://paris2k.at.box.sk/tools/listings/> for some examples.
- Search for files like "config.inc.php" or "mysql.cfg" that could contain mySQL password and username combinations. Try to think of good search strings using words like PHP, SQL, mySQL, etc.
- Try things like: *inurl:admin "index of" "database"* or *inurl:phpmyadmin "index of"* or *inurl:mysql "index of" site:neworder.box.sk intitle:index.of* or *intitle:index.of.private* (= intitle:"index of private")

Conclusion

The internet is a network to which hundreds of thousands, if not millions of web servers are connected and in theory, all data can be reached, unless properly protected. Both software designers and end users should pay more attention to default installation security configuration and security policy. In the end, there are always going to be people who make mistakes, use default installs, use poorly secured software or just don't care or still believe there's no danger in putting this kind of data online. And in the end there's also always going to be curious people who love to find that interesting information they have been hoping for. Google can help you considerably, in locating this kind of information and it's easy and fun.

Sidenote

I have used in this article, "live examples" because "foobar examples" in this case wouldn't have been very useful. I hope you choose to learn from these examples and not use them to commit malicious acts. Think of a great search string yourself and don't abuse the ones I have shown to explain the technique a little bit. (Guess I couldn't stay away from the ethics lesson after all)

Afterword

To not abandon tradition I would like to take the chance to greet some people. People like JLP, Rattlesnake, Drew, X, Tek, Sean, Marek, Resolution and others... you all know who you are, Thanks for helping me out numerous times with numerous different things.

Bibliography

1. Google not 'hackers' best friend', James Middleton, VNUnet.com, 2001
<http://www.vnunet.com/News/1127162>
 2. Google: Net Hacker Tool du Jour, Christopher Null, wired.com, 2003
<http://www.wired.com/news/infostructure/0,1377,57897,00.html>
 3. Microsoft FrontPage 2000 Server Extensions Resource Kit
http://sciris.shu.edu/Manuals/FrontPage/serk/adhtm_1.htm
 4. Hacking Exposed Third Edition, McClure, Scambray, Kurtz, ISBN: 0-07-219381-6
<http://www.osborne.com>
-