Search Engine Security Auditing

By Russ McRee

Prerequisites

SiteDigger 2.0[1]
Google SOAP API[2]
MSN AppID[3]
MSNPawn 1.1[4]
Gooscan[5]
Windows .NET Framework[6]

Introduction

Of the plethora of vectors we endeavor to protect, one often overlooked is the level of exposure our clients and/or places of business may present to search engines via their web presence. Auditing for these exposures should be as common a practice as updating IDS/IPS signatures or keeping your systems patched. Unless you have been in a cave without Internet connectivity for the last few years, you've likely heard of Google hacking and the Google Hacking Database (GHDB) brought to you by Johnny Long.[7] Given that many far brighter than I have covered the specifics of Google hacking extensively, I shall not presume to bore you or reinvent the wheel. But I do believe I may be able to offer some insight on a toolkit of specific tools and certain queries used in a variety of search mechanisms that will produce results useful in ensuring that your assets are not unnecessarily exposed. The tools themselves are useful, and well worth exploring and utilizing; but remember, with a set of well-tuned queries, you can audit quite successfully. I'll list some useful queries as we explore a few tools as well as a summary list at the end of the column.

If you doubt the currency of this topic consider this. As this column was being written, the week of Cinco de Mayo (cheers), Secunia announced an AXIS Camera Control "SaveBMP()" Method Buffer Overflow.[8] Try querying *inurl: indexFrame.shtml Axis* and you will catch my drift.

One other thought: if you really take a liking to this topic and want to conduct your own research, or strengthen the defensive posture of those you protect, consider this. GHH, the "Google Hack Honeypot,"[9] will provide with an excellent platform for said research.

SiteDigger 2.0

If you have read up on popular information gathering tools that leverage Google, you are likely familiar with Foundstone's SiteDigger 2.0. The problem with SiteDigger, as well as other similar tools, is that they utilize a Google web services API license key. As of 12/5/2006 Google no longer offers keys for the SOAP Search API (the AJAX API will not work) and SiteDigger, at the time of this writing, has not been updated. Shanit at Foundstone says they are aware of the issue and are planning to overhaul all the tools in the near future but cannot provide specific dates at this time. Most of the Software Application Security Services (SASS) tools on the Foundstone site are somewhat aged, averaging at least 2.5 years without an update, yet there appears to be light at the end of the tunnel. But, unless you have on old API key, SiteDigger will not work. Contact me below for some assistance here, if you wish.

While it will not help you with SiteDigger (it throws an exception without as an API key), the folks at Sensepost (Wikto, BiDiBlah) have released Aura,[10] a proxy of sorts for doGoogleSearch API function calls.

There is also some interesting work being developed by the folks at sitening.com in their EvilAPI Perl scripts.[11]

There are other tools that will use Google without an API key, but with some risk and there are also tools for other search engines, like MSN. We'll cover both elements later in this column.

SiteDigger 2.0, assuming you have an API key, is an excellent tool. The interface is simple enough and is intuitive enough to allow immediate use. You are presented with four tabs, including *Search, Options, Submit Signature,* and *Raw Search.* Submit Signature allow you to do just that: submit a unique query you have written to the Foundstone Signature Database. I am not sure if this feature will result in a response from Foundstone, but regardless your new query will be

[1] http://www.foundstone.com/resources/freetools.htm

[2] No longer available, you must already have one

[3] http://search.msn.com/developer/default.aspx

[4] http://net-square.com/msnpawn/index.shtml

[5] http://johnny.ihackstuff.com/downloads/task,cat_view/gid,16/

[6] http://www.microsoft.com/downloads/details.aspx?familyid=0856EACB-4362-4B0D-8EDD-AAB15C5E04F5&displaylang=en

[7] http://johnny.ihackstuff.com

[8] http://secunia.com/advisories/25093/

[9] http://ghh.sourceforge.net/index.php

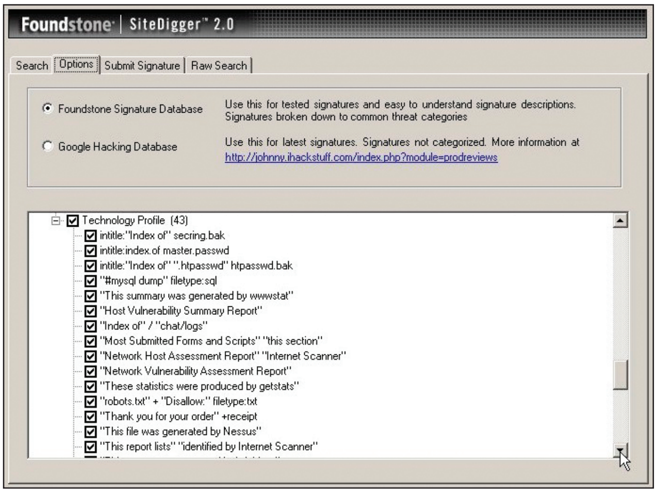
[10] http://www.sensepost.com/research/aura/

[11] http://sitening.com/evilapi/

stored locally in the Program Files\Foundstone Free Tools\Foundstone SiteDigger 2.0 directory. Here you will also note the FSDB and GHDB XML files where queries or "signatures" are tagged for use by SiteDigger. You can edit these if you wish but pay attention to your formatting.

The Raw Search view is little more than crafting a query directly in the Google interface so use it with a grain of salt.

Check out Options before you fire off your first Search. You will note radio buttons for either the Foundstone Signature Database or the GHDB. You can select queries based on relevance. If auditing a client you are absolutely certain should not be running an InterJak device (GHDB) or PHP-Nuke (FSDB), then disable the query. These scans do not take long as a whole, but tuning will definitely shave time off the bottom line. Figure 1 show the Options view.



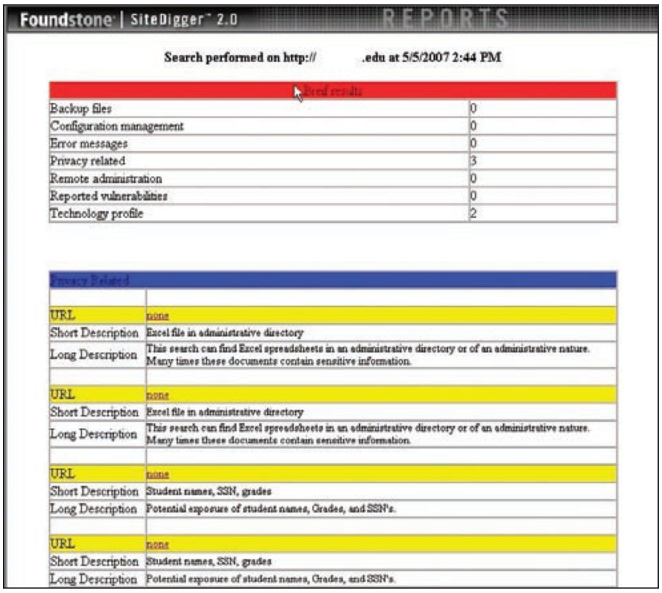
● Foundstone Signature Database — Use this for tested signatures and easy to understand signature descriptions. Signatures broken down to common threat categories

○ Google Hacking Database — Use this for latest signatures. Signatures not categorized. More information at http://johnny.ihackstuff.com/index.php?module=prodreviews

☑ Technology Profile (43)
- ☑ intitle:"Index of" secring.bak
- ☑ intitle:index.of master.passwd
- ☑ intitle:"Index of" ".htpasswd" htpasswd.bak
- ☑ "#mysql dump" filetype:sql
- ☑ "This summary was generated by wwwstat"
- ☑ "Host Vulnerability Summary Report"
- ☑ "Index of" / "chat/logs"
- ☑ "Most Submitted Forms and Scripts" "this section"
- ☑ "Network Host Assessment Report" "Internet Scanner"
- ☑ "Network Vulnerability Assessment Report"
- ☑ "These statistics were produced by getstats"
- ☑ "robots.txt" + "Disallow:" filetype:txt
- ☑ "Thank you for your order" +receipt
- ☑ "This file was generated by Nessus"
- ☑ "This report lists" "identified by Internet Scanner"

Figure 1 – SiteDigger Options view

As always, remember that it is recommended to point this tool and those like it only at sites you have permission to do so. It is less invasive than a true web application scanner, yes, but can still be perceived as hostile. I would only ask that if you decide to play at will, report anything you find that might be a true vulnerability to the site owner.



Search performed on http://   .edu at 5/5/2007 2:44 PM

Brief results	
Backup files	0
Configuration management	0
Error messages	0
Privacy related	3
Remote administration	0
Reported vulnerabilities	0
Technology profile	2

Privacy Related	
URL	none
Short Description	Excel file in administrative directory
Long Description	This search can find Excel spreadsheets in an administrative directory or of an administrative nature. Many times these documents contain sensitive information.
URL	none
Short Description	Excel file in administrative directory
Long Description	This search can find Excel spreadsheets in an administrative directory or of an administrative nature. Many times these documents contain sensitive information.
URL	none
Short Description	Student names, SSN, grades
Long Description	Potential exposure of student names, Grades, and SSN's.
URL	none
Short Description	Student names, SSN, grades
Long Description	Potential exposure of student names, Grades, and SSN's.

Figure 2 – Too much information…

SiteDigger offers a nice findings export as well, which can be used to build a report for clients or management. Figure 2 shows the results of a scan I conducted against an .edu domain. Often these domains will result in PII disclosure (personally identifiable information) and this one was no exception.

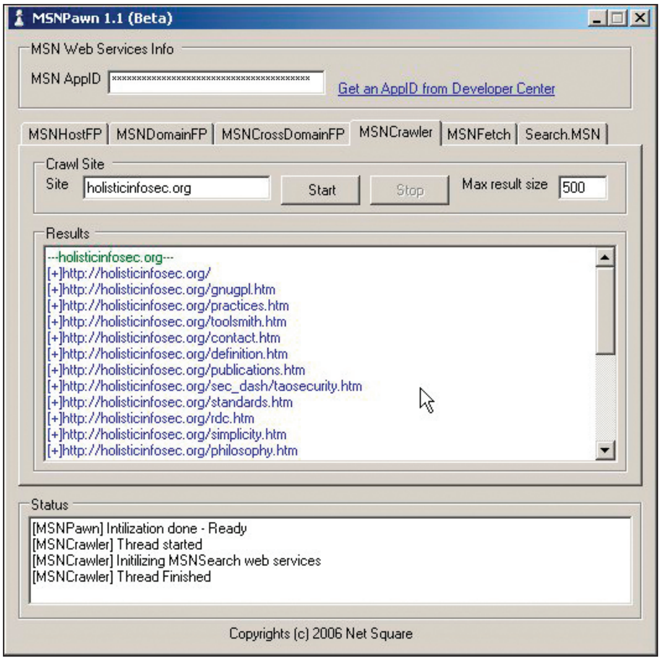
Even if you don't have a Google Web API key, download SiteDigger and explore the Options tab. If nothing else, consider it a valuable collection of queries you can use manually to audit the sites in your scope of interest. Again, I'll select a number of queries I believe useful to an auditor at the end of the column.

MSNPawn 1.1

MSNPawn is a free tool brought to you by net-square. Shreeraj Shah is one of net-square's founders, whose work on mod_security for Apache httpd is well documented and invaluable for use in hardening Apache web server installations. MSNPawn is still "beta" at the time of this writing, but nonetheless works reasonably well. Shreeraj indicates absolutely no further work on this tool, so what you see is what you get.

MSNPawn shares some similarities with SiteDigger but offers some additional features you might find useful or worthy of experimentation.

This tool will allow you host, domain, and cross domain foot printing, which is all well and good, but I really like the MSNCrawler and MSNFetch options. MSNCrawler does precisely what is asked of it and will return the results (count of your choosing) right to the UI for cut and paste into a report or



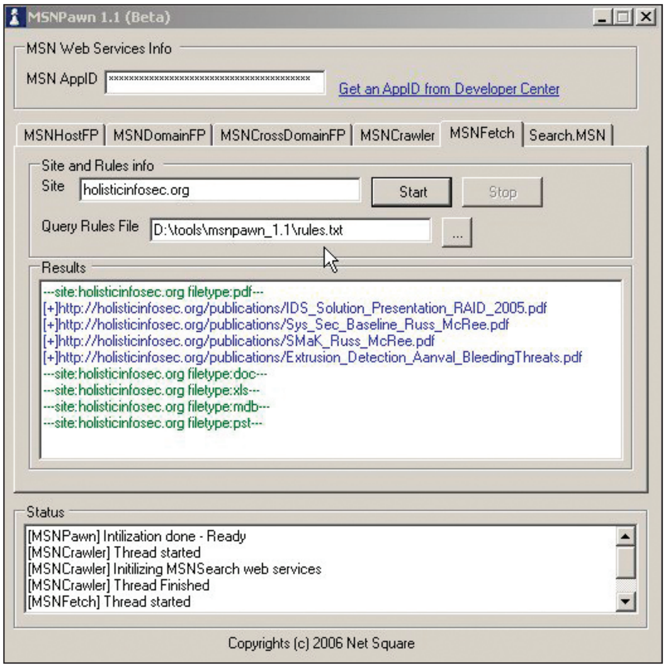
MSN AppID: **

Site: holisticinfosec.org   Max result size: 500

```
---holisticinfosec.org---
[+]http://holisticinfosec.org/
[+]http://holisticinfosec.org/gnugpl.htm
[+]http://holisticinfosec.org/practices.htm
[+]http://holisticinfosec.org/toolsmith.htm
[+]http://holisticinfosec.org/contact.htm
[+]http://holisticinfosec.org/definition.htm
[+]http://holisticinfosec.org/publications.htm
[+]http://holisticinfosec.org/sec_dash/taosecurity.htm
[+]http://holisticinfosec.org/standards.htm
[+]http://holisticinfosec.org/rdc.htm
[+]http://holisticinfosec.org/simplicity.htm
[+]http://holisticinfosec.org/philosophy.htm
```

```
[MSNPawn] Intilization done - Ready
[MSNCrawler] Thread started
[MSNCrawler] Initilizing MSNSearch web services
[MSNCrawler] Thread Finished
```

Figure 3 – MSNCrawler



MSN AppID: **

Site: holisticinfosec.org

Query Rules File: D:\tools\msnpawn_1.1\rules.txt

```
---site:holisticinfosec.org filetype:pdf---
[+]http://holisticinfosec.org/publications/IDS_Solution_Presentation_RAID_2005.pdf
[+]http://holisticinfosec.org/publications/Sys_Sec_Baseline_Russ_McRee.pdf
[+]http://holisticinfosec.org/publications/SMaK_Russ_McRee.pdf
[+]http://holisticinfosec.org/publications/Extrusion_Detection_Aanval_BleedingThreats.pdf
---site:holisticinfosec.org filetype:doc---
---site:holisticinfosec.org filetype:xls---
---site:holisticinfosec.org filetype:mdb---
---site:holisticinfosec.org filetype:pst---
```

```
[MSNPawn] Intilization done - Ready
[MSNCrawler] Thread started
[MSNCrawler] Initilizing MSNSearch web services
[MSNCrawler] Thread Finished
[MSNFetch] Thread started
```

Figure 4 – MSNFetch

findings log. Crawling is handy but crafting queries in the UI or calling your customized query file is really useful.

Navigate to the MSNFetch tab. Let's say you have a rules.txt file that contains:

filetype:pdf

filetype:doc

filetype:xls

filetype:mdb

filetype:pst

Tell MSNPawn where your rules.txt file is, enter the site to audit, and hit Start, it is that easy.

The API agreement for MSN limits you to 10,000 a day and allows a total of 50 results for each query. You will not "alienate" MSN, assuming you do not exceed these limits. You will find our next topic, however, is quite apt at drawing attention to yourself. You've been warned.

Gooscan

Gooscan is a little app written for Linux and can be downloaded at Johnny Long's site. It is described there as "a tool that automates queries against Google search appliances" but can be run against Google itself in direct violation of their Terms of Service. They have not enforced these terms to date, but do avoid being the first candidate.

Compile it thus: *gcc -o gooscan gooscan.c*

Run it first like this: *./gooscan*, and you will note sample strings and parameters to pass.

Running it for the purpose of auditing a single site is easily done by passing *–s* to limit the scan to a specific site. You can also call a file containing a list of queries by passing *–i*. Your installation will include a data_files directory that is laden with search nuggets useful to the auditor.

You might try this Gooscan query where MySite is the site you are auditing and MyAudit is your output file: *./gooscan -t www.google.com -i data_files/gdork.gs -s MySite.net -o MyAudit.htm*

More simply, if you have a single query of concern for a specific site craft you can drop the data file and pass the string with the query switch *–q*.

./gooscan -t www.google.com -q "intitle:Remote.Desktop.Web.Connection inurl:tsweb" -s MySite.net -o MyAudit.htm.

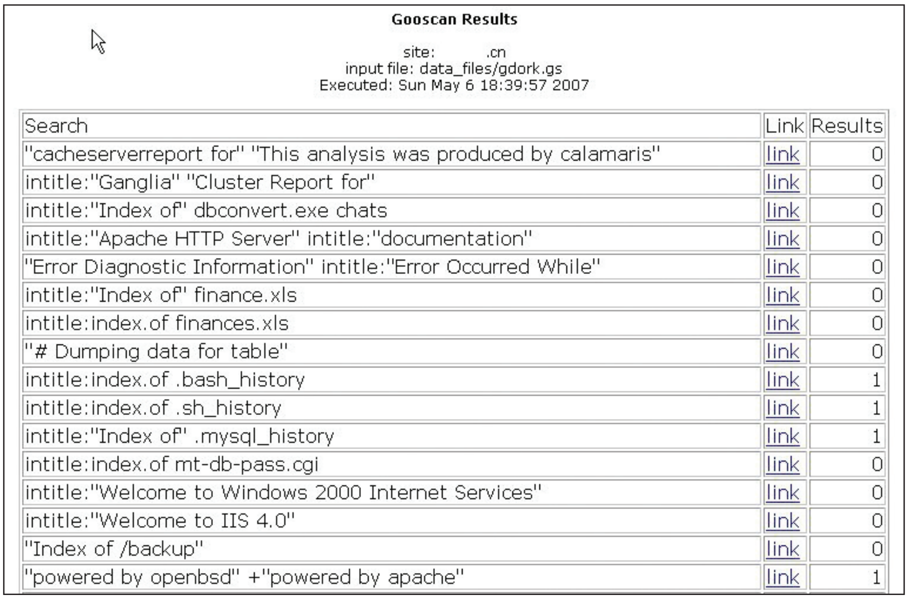
One of the coolest things about this tool is the output file. Pull it up in a browser and you will see everything you need to directly confirm the finding including a link. Good stuff, no doubt.

The results in Figure 5 could lead us to believe we have a horribly configured OpenBSD server running Apache, and exposing truly dangerous "hidden" files.

Of all the tools offered this month, I far and away prefer this one as it offers endless possibilities. It'll also get you in trouble with the big kahuna of search engines, so proceed with caution.

Searches

Following are searches found in the GHDB, the FSDB, or randomly online, but are those I consider potentially useful to the Search Engine Security Auditor. I claim absolutely no



Gooscan Results

site: .cn
input file: data_files/gdork.gs
Executed: Sun May 6 18:39:57 2007

Search	Link	Results
"cacheserverreport for" "This analysis was produced by calamaris"	link	0
intitle:"Ganglia" "Cluster Report for"	link	0
intitle:"Index of" dbconvert.exe chats	link	0
intitle:"Apache HTTP Server" intitle:"documentation"	link	0
"Error Diagnostic Information" intitle:"Error Occurred While"	link	0
intitle:"Index of" finance.xls	link	0
intitle:index.of finances.xls	link	0
"# Dumping data for table"	link	0
intitle:index.of .bash_history	link	1
intitle:index.of .sh_history	link	1
intitle:"Index of" .mysql_history	link	1
intitle:index.of mt-db-pass.cgi	link	0
intitle:"Welcome to Windows 2000 Internet Services"	link	0
intitle:"Welcome to IIS 4.0"	link	0
"Index of /backup"	link	0
"powered by openbsd" +"powered by apache"	link	1

Figure 5 – Gooscan gold

credit for any of these queries, nor any of the hard work that went into the research and development of the tools described in this column. I am merely reciting what I have found useful, hoping you find it useful as well, and applauding “those who went before.”

This list not complete, nor comprehensive. Visit the GHDB or search for others. You can also tweak any and all for your use. Remember, there is no limit to modification.

There is a great presentation[12] from a couple of members of ISSA España at the FIST Conference in 2005 in Madrid. It is in Spanish but you will certainly benefit from giving it a read. Thanks to Pedro Pablo Pérez García and Gonzalo Álvarez Marañón. Many of the queries listed below are referred to in their presentation.

To best use these on your site, remember to preface these queries with site:MySite.com where MySite is your site.

Things you may not want exposed on sites in your scope of interest:

```
“Welcome to phpMyAdmin” “ Create new database”
“VNC Desktop” inurl:5800
intitle:Remote.Desktop.Web.Connection inurl:tsweb
allintitle:Outlook Web Access Logon (there are variations on this one)
inurl:”auth_user_file.txt”
“Index of /admin”
“Index of /password”
“Index of /mail”
“Index of /” +passwd
“Index of /” +password.txt
“Index of /” +.htaccess
index of ftp +.mdb allinurl:/cgi-bin/ +mailto
alliurl:phpinfo.php
administrators.pwd.index
authors.pwd.index
service.pwd.index
filetype:config web
allintitle: “index of/admin”
allintitle: “index of/root”
allintitle: sensitive filetype:doc
allintitle: restricted filetype :mail
allintitle: restricted filetype:doc site:gov
inurl:passwd filetype:txt
inurl:admin filetype:db
inurl:iisadmin
inurl:”auth_user_file.txt”
inurl:”wwwroot/*.”
“WS_FTP.LOG“
allinurl:winnt/system32/
allinurl:/bash_history
intitle:”Index of” .sh_history
intitle:”Index of” .bash_history
intitle:”index of” passwd
intitle:”index of” people.lst
intitle:”index of” pwd.db
intitle:”index of” etc/shadow
intitle:”index of” spwd
intitle:”index of” master.passwd
intitle:”index of” htpasswd
intitle:”index of” members OR accounts
intitle:”index of” user_carts OR user_cart
```

Benefits and Drawbacks

The benefits are endless and obvious. Who wants to find out they have been unnecessarily exposing corporate secrets, PII, or critical administrative files? Knowledge is power, for the hacker, yes, but also for you. Trust me, they are looking, you should too.

The drawbacks? Perhaps the hours you will spend down the rabbit hole chasing results of interest. But more likely, API license limitations are a real drawback and so too is the prospect of ending up Google banned. Can you so say “dynamic IP address” children?

The costs to use these methods to protect your enterprise are nil, just time. Use it!

In Conclusion

Keep in mind these points. First, you will find your share of honeypots using these tools at random, depending on the queries you use. There’s a good one here: http://worm.ccert.edu.cn/Goopot/indexof.html. Second, again, always be cautious about violating the terms of service of the search engine you query. Third, help out those Johnny Long refers to as Googledorks. If you actually stumble across something dangerous to the domain owner, lend a hand and let them know.

There is a great deal to be learned using these methods. If you wish to solidify these skills you may wish to entertain the SANS Stay Sharp (soon to be STAR) class Google Hacking and Defense. Regardless, take advantage of these tools and techniques. Remember, protect your own. Cheers…until next month.

Acknowledgements

— Johnny Long for paving the way for the Search Engine Security Auditor.

— Ed Skoudis for the inspiration to think like the bad guy and ensure that you know their tools as well as they do.

About the Author

Russ McRee, GCIH, CISSP, is a security analyst working for Expedia. He is a member of numerous security organizations and maintains holisticinfosec.org. Contact him at russ@holisticinfosec.org.

[12] http://www.iec.csic.es/gonzalo/descargas/ProteccioncontraHackingconGoogle.pdf