Google Hacking for Penetration Testers

Using Google as a Security Testing Tool
Johnny Long
johnny@ihackstuff.com
What we’re doing

- I hate pimpin’, but we’re covering many techniques covered in the “Google Hacking” book.
- For much more detail, I encourage you to check out “Google Hacking for Penetration Testers” by Syngress Publishing.
Advanced Operators

Before we can walk, we must run. In Google’s terms this means understanding advanced operators.
Advanced Operators

- Google advanced operators help refine searches.
- They are included as part of a standard Google query.
- Advanced operators use a syntax such as the following:

  `operator:search_term`

- There’s no space between the operator, the colon, and the search term!
### Advanced Operators at a Glance

<table>
<thead>
<tr>
<th>Operator</th>
<th>Purpose</th>
<th>Mixes with other operators?</th>
<th>Can be used alone?</th>
<th>Does search work in</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Web</td>
</tr>
<tr>
<td>intitle</td>
<td>Search page title</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>allintitle</td>
<td>Search page title</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>inurl</td>
<td>Search URL</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>allinurl</td>
<td>Search URL</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>filetype</td>
<td>Search specific files</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>allintext</td>
<td>Search text of page only</td>
<td>not really</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>site</td>
<td>Search specific site</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>link</td>
<td>Search for links to pages</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>inanchor</td>
<td>Search link anchor text</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>numrange</td>
<td>Locate number</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>daterange</td>
<td>Search in date range</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>author</td>
<td>Group author search</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>group</td>
<td>Group name search</td>
<td>not really</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>insubject</td>
<td>Group subject search</td>
<td>yes</td>
<td>yes</td>
<td>like intitle</td>
</tr>
<tr>
<td>msgid</td>
<td>Group msgid search</td>
<td>no</td>
<td>yes</td>
<td>not really</td>
</tr>
</tbody>
</table>

Some operators can only be used to search specific areas of Google, as these columns show.

Advanced operators can be combined in some cases.

In other cases, mixing should be avoided.
Crash course in advanced operators

Some operators search overlapping areas. Consider site, inurl and filetype.

- **SITE:**
  - Site can not search port.

- **INURL:**
  - Inurl can search the whole URL, including port and filetype.

- **FILETYPE:**
  - Filetype can only search file extension, which may be hard to distinguish in long URLs.
There are many ways to find the same page. These individual queries could all help find the same page.
Advanced Google Searching

Put those individual queries together into one monster query and you only get that one specific result.

Adding advanced operators reduces the number of results adding focus to the search.
Google Hacking Basics

Putting operators together in intelligent ways can cause a seemingly innocuous query…

INURL:admin  INURL:orders  FILETYPE:php

osCommerce

om/catalog/admin/orders.php+filety
Google Hacking Basics

…can return devastating results!

<table>
<thead>
<tr>
<th>Customer Names</th>
<th>Order Amounts</th>
<th>Payment Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>$56.30</td>
<td>Delivered</td>
</tr>
<tr>
<td>Bob</td>
<td>$81.90</td>
<td>Delivered</td>
</tr>
<tr>
<td>Charlie</td>
<td>$69.50</td>
<td>Pending</td>
</tr>
<tr>
<td>Dana</td>
<td>$45.25</td>
<td>Delivered</td>
</tr>
<tr>
<td>Elena</td>
<td>$159.15</td>
<td>Delivered</td>
</tr>
</tbody>
</table>
Google Hacking Basics

Let’s take a look at some basic techniques:

Anonymous Googling
Special Characters
Anonymous Googling

The cache link is a great way to grab content after it’s deleted from the site. The question is, where exactly does that content come from?
Anonymous Googling

• Some folks use the cache link as an anonymizer, thinking the content comes from Google. Let’s take a closer look.

This line from the cached page’s header gives a clue as to what’s going on...
Anonymous Googling

This tcpdump output shows our network traffic while loading that cached page.

This is Google.

This is Phrack.

We touched Phrack’s web server. We’re not anonymous.
Anonymous Googling

- Obviously we touched the site, but why?
- Here’s more detailed tcpdump output:

```
0x0040  0d6c 4745 5420 2f67 7266 782f 3831 736d        .lGET./grfx/81sm
0x0050   626c 7565 2e6a 7067 2048 5454 502f 312e        blue.jpg.HTTP/1.
0x0060   310d 0a48 6f73 743a 2077 7777 2e70 6872        1..Host:.www.phr
0x0070   6163 6b2e 6f72 670d 0a43 6f6e 6e65 6374 ack.org..Connect
0x0080   696e3a 206b 6565 702d 616c 6976 650d ion:.keep-alive.
0x0090   0a52 6566 6572 6572 3a20 6874 7470 3a2f .Referer:.http:/
0x00a0   2f36 342e 3233 332e 3136 312e 3130 342f /64.233.161.104/
0x00b0   7365 6172 6368 3f71 3d63 6163 6865 3a4c search?q=cache:L
0x00c0   4251 5a49 7253 6b4d 6755 4a3a 7777 772e BQZIrSkMgUJ:www.
0x00d0   7068 7261 636b 2e6f 7267 2f2b 2b73 6974 phrack.org/++sit
0x00e0   3d6a 7067 2e6f 7267 2f2b 2b73 6974 653a e:www.phrack.org
0x00f0   2b70 6872 6163 6b62 666c 3d65 6e0d 0a55 +phrack&hl=en..U
```

An image loaded!
Anonymous Googling

This line spells it out. Let’s click this link and sniff the connection again....
Anonymous Googling

This time, the entire conversation was between us (192.168.2.32) and Google (64.233.167.104)
Anonymous Googling

• What made the difference? Let’s compare the two URLs:

• Original:
  

• Cached Text Only:
  

Adding &strip=1 to the end of the cached URL only shows Google’s text, not the target’s.
Anonymous Googling

- Anonymous Googling can be helpful, especially if combined with a proxy. Here’s a summary.

1. Perform a Google search.
2. Right-click the cached link and copy the link to the clipboard.
3. Paste the URL to the address bar, add &strip=1, hit return. You’re only touching Google now…
Special Search Characters

• We’ll use some special characters in our examples. These characters have special meaning to Google.

• Always use these characters without surrounding spaces!
  • ( + ) force inclusion of something common
  • ( - ) exclude a search term
  • ( " ) use quotes around search phrases
  • ( . ) a single-character wildcard
  • ( * ) any word
  • ( | ) boolean ‘OR’
  • Parenthesis group queries (“master card” | mastercard)
Google’s PHP Blocker: “We’re Sorry..”

• Google has started blocking queries, most likely as a result of worms that slam Google with ‘evil queries.’

We're sorry...

... but we can’t process your request right now. A computer virus or spyware application is sending us automated requests, and it appears that your computer or network has been infected.

We'll restore your access as quickly as possible, so try again soon. In the meantime, you might want to run a virus checker or spyware remover to make sure that your computer is free of viruses and other spurious software.

We apologize for the inconvenience, and hope we'll see you again on Google.
Google Hacker’s workaround

- Our original query looks like this:


- Stripped down, the query looks like this:


- We can modify our query (inurl:something.php is bad) by changing the case of the file extension, like so:

  http://www.google.com/search?q=inurl:admin.PHP&start=10

This works in the web interface as well.
Pre-Assessment

There are many things to consider before testing a target, many of which Google can help with. One shining example is the collection of email addresses and usernames.
Trolling for Email Addresses

• A seemingly simple search uses the @ sign followed by the primary domain name.

The “@” sign doesn’t translate well…

But we can still use the results…
Automated Trolling for Email Addresses

• We could use a lynx to automate the download of the search results:

    lynx -dump http://www.google.com/search?q=@gmail.com > test.html

• We could then use regular expressions (like this puppy by Don Ranta) to troll through the results:


• Run through grep, this regexp would effectively find email addresses (including addresses containing IP numbers)
More Email Automation

• The ‘email miner’ PERL script by Roelof Temmingh at sensepost will effectively do the same thing, but via the Google API:

```
j0hny-longs-Computer:~/Documents/workbench/Coding $ ./email-mine.pl google.com 1
0
username@gmail.com
gramophone@gmail.com
bush04@gmail.com
lostmon@gmail.com
kerry04@gmail.com
all_in_all@gmail.com
j0hny-longs-Computer:~/Documents/workbench/Coding $
```

This searches the first ten Google results… with only one hit against your API key.
More Email Automation

Running the tool through 50 results (with a 5 parameter instead of 1) finds even more addresses.
More email address locations

<table>
<thead>
<tr>
<th>Query</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Internal Server Error&quot;</td>
<td>Apache server error could reveal admin e-mail address</td>
</tr>
<tr>
<td>int:=&quot;Execution of this script not permitted&quot;</td>
<td>Cgiwrap script can reveal lots of information, including e-mail addresses and even phone numbers</td>
</tr>
<tr>
<td>e-mail address filetype:csv</td>
<td>CSV files that could contain e-mail addresses</td>
</tr>
<tr>
<td>int:=&quot;index.of dead.letter&quot;</td>
<td>dead.letter UNIX file contains the contents of unfinished e-mails that can contain sensitive information</td>
</tr>
<tr>
<td>inurl:fcgi-bin/echo</td>
<td>fastcgi echo script can reveal lots of information, including e-mail addresses and server information</td>
</tr>
<tr>
<td>filetype:.pst .from .to .date</td>
<td>Finds Outlook PST files, which can contain e-mails, calendaring, and address information</td>
</tr>
<tr>
<td>int:=&quot;index.of inbox&quot;</td>
<td>Generic &quot;inbox&quot; search can locate e-mail caches</td>
</tr>
<tr>
<td>int:=&quot;Index Of&quot; -inurl:maillog maillog size</td>
<td>Maillog files can reveal usernames, e-mail addresses, user login/logout times, IP addresses, directories on the server, and more</td>
</tr>
<tr>
<td>inurl:email filetype: mdb</td>
<td>Microsoft Access databases that could contain e-mail information</td>
</tr>
<tr>
<td>filetype:xls inurl:&quot;email.xls&quot;</td>
<td>Microsoft Excel spreadsheets containing e-mail addresses</td>
</tr>
<tr>
<td>filetype:xls username password email</td>
<td>Microsoft Excel spreadsheets containing the words username, password, and email</td>
</tr>
<tr>
<td>int:=&quot;index.of inbox dbx&quot;</td>
<td>Outlook Express cleanup.log file can contain locations of e-mail information</td>
</tr>
</tbody>
</table>

These queries locate email addresses in more “interesting” locations…
More email address locations

<table>
<thead>
<tr>
<th>Query</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>filetype:eml eml +intext: “Subject” +intext: “From”</code></td>
<td>Outlook express e-mail files contain e-mails with full headers</td>
</tr>
<tr>
<td><code>intitle:index.of.inbox dbx</code></td>
<td>Outlook Express e-mail folder</td>
</tr>
<tr>
<td><code>filetype:wab wab</code></td>
<td>Outlook Mail address books contain sensitive e-mail information</td>
</tr>
<tr>
<td><code>filetype:pst inurl:“outlook.pst”</code></td>
<td>Outlook PST files can contain e-mails, calendaring, and address information</td>
</tr>
<tr>
<td><code>filetype:mbx mbx intext:Subject</code></td>
<td>Outlook versions 1–4 or Eudora mailbox files contain sensitive e-mail information</td>
</tr>
<tr>
<td><code>inurl:.cgi-bin/printenv</code></td>
<td>Printenv script can reveal lots of information, including e-mail addresses and server information</td>
</tr>
<tr>
<td><code>inurl:forward filetype:forward -cvs</code></td>
<td>UNIX user e-mail forward files can list e-mail addresses</td>
</tr>
<tr>
<td>`( filetype:mail</td>
<td>filetype:eml</td>
</tr>
<tr>
<td>“Most Submitted Forms and Scripts” “this section”</td>
<td>WebTrends statistics pages reveal directory information, client access statistics, e-mail addresses, and more</td>
</tr>
<tr>
<td><code>filetype:reg reg +intext: “internet account manager”</code></td>
<td>Windows registry files can reveal information such as usernames, POP3 passwords, e-mail addresses, and more</td>
</tr>
<tr>
<td>“This summary was generated by wwwstat”</td>
<td>Wwwwstat statistics information can reveal directory info, client access statistics, e-mail addresses, and more</td>
</tr>
</tbody>
</table>

These queries locate email addresses in more “interesting” locations…
Network Mapping

Google is an indispensable tool for mapping out an Internet-connected network.
Basic Site Crawling

- the site: operator narrows a search to a particular site, domain or subdomain.
Basic Site Crawling

Most often, a site search makes the obvious stuff float to the top.

As a security tester, we need to get to the less obvious stuff.

www.microsoft.com is way too obvious...
Basic Site Crawling

• To get rid of the more obvious crap, do a negative search.

Notice that the obvious “www” is missing, replaced by more interesting domains.
Basic Site Crawling

- Repeating this process of site reduction, tracking what floats to the top leads to nasty big queries like:

  site:microsoft.com
  -site:www.microsoft.com
  -site:msdn.microsoft.com
  -site:support.microsoft.com
  -site:download.microsoft.com
  -site:office.microsoft.com

  ...
Basic Site Crawling

- The results of such a big query reveal more interesting results...

Eventually we’ll run into a 32 query limit, and this process tends to be tedious.
Intermediate Site Crawling

Using lynx to capture the Google results page...

..returns the same results.

..and sed and awk to process the HTML...

```bash
bash-2.05$ sed -n 's/\.\([[:alpha:]]*\)*/\1/' test.html | awk '{print $2}' | sort -u
http://download.microsoft.com/
http://go.microsoft.com/
http://msdn.microsoft.com/
http://msevents.microsoft.com/
http://murl.microsoft.com/
http://office.microsoft.com/
http://protect.microsoft.com/
http://research.microsoft.com/
https://s.microsoft.com/
http://support.microsoft.com/
bash-2.05$ 
```
So what?

• Well, honestly, host and domain enumeration isn’t new, but we’re doing this without sending any packets to the target we’re analyzing.

• This has several benefits:
  – Low profile. The target can’t see your activity.
  – Results are “ranked” by Google. This means that the most public stuff floats to the top. Some more “interesting stuff” trolls near the bottom.
  – “Hints” for follow-up recon. You aren’t just getting hosts and domain names, you get application information just by looking at the snippet returned from Google. One results page can be processed for many types of info. Email addresses, names, etc. More on this later on...
  – Since we’re getting data from several sources, we can focus on non obvious relationships. This is huge!

• Some down sides:
  – In some cases it may be faster and easier as a good guy to use traditional techniques and tools that connect to the target, but remember- the bad guys can still find and target you via Google!
Advanced Site Crawling

- Google frowns on automation, unless you use tools written with their API. Know what you’re running unless you don’t care about their terms of service.
- We could easily modify our lynx retrieval command to pull more results, but in many cases, more results won’t equal more unique hosts.
- So, we could also use another technique to locate hosts… plain old fashion common word queries.
Advanced Site Crawling

Searching for multiple common words like “web”, “site”, “email”, and “about” along with site… appended to a file…
Advanced Site Crawling

Sifting through the output from those queries, we find many more interesting hits.
Advanced Site Crawling

Roelof Temmingh from sensepost.com coded this technique into a PERL (API-based) script called dns-mine.pl to achieve much more efficient results.

We'll look more at coding later...
Too much noise, not enough signal…

• Getting lists of hosts and (sub)domains is great. It gives you more targets, but there’s another angle.
• Most systems are only as secure as their weakest link.
• If a poorly-secured company has a trust relationship with your target, that’s your way in.

• Question: How can we determine site relationships with Google?

• One Answer: the “link” operator.
Raw Link Usage

link: combined with the name of a site shows... sites that link to that site.

link: has limits though. See mapquest here?
Link has limits

...combining link: with site: doesn’t seem to work...
Link has limits

Link: gets treated like normal search text (not a search modifier) when combined with other operators.
Link has other limits

Knowing that these sites link to www.microsoft.com is great, but how relevant is this information?

Do we necessarily care about Google-ranked relationships? How do we get to REAL relationships?
Non-obvious site relationships

- Sensepost to the rescue again! =)
- BiLE (the Bi-directional Link Extractor), available from [http://www.sensepost.com/garage_portal.html](http://www.sensepost.com/garage_portal.html) helps us gather together links from Google and piece together these relationships.
- There’s much more detail on this process in their whitepaper, but let’s cover the basics…
Non-obvious site relationships

• A link from a site weighs more than a link to a site
  – Anyone can link to a site if they own web space (which is free to all)

• A link from a site with a lot of links weighs less than a link from a site with a small amount of links
  – This means specifically outbound links.
  – If a site has few outbound links, it is probably lighter.
  – There are obvious exceptions like link farms.
Non-obvious site relationships

• A link to a site with a lot of links to the site weighs less than a link to a site with a small amount of links to the site.
  – If external sources link to a site, it must be important (or more specifically popular)
  – This is basically how Google weighs a site.
• The site that was given as input parameter need not end up with the highest weight – a good indication that the provided site is not the central site of the organization.”
  – If after much research, the site you are investigating doesn’t weight the most, you’ve probably missed the target’s main site.
Who is Sensepost?

Relying on Google’s 6400+ results can be daunting… and misleading.
Non-obvious site relationships

• It seems dizzying to pull all this together, but BiLE does wonders. Let’s point it at sensepost.com:

This is the extraction phase. BiLE is looking for links to www.sensepost.com (via Google) and writing the results to a file called “out”...
Non-obvious site relationships

- This is the weigh phase. BiLE takes the output from the extraction phase...

And weighs the results using the four main criteria of weighing discussed above... aided primarily by Google searches.

This shows the strongest relationships to our target site first, which during an assessment equate to secondary targets, especially for information gathering.
The next step…

Let’s say we’re looking at NASA.…

We could use ‘googleturd’ searches, like site:nasa to locate typos which may be real sites…

How can we verify these???
Host verification...

- Cleaning the names and running DNS lookups is one way...

Pay dirt! Now what???

We could further expand on these IP ranges via DNS queries as well...
Expanding out…

• Once armed with a list of sites and domains, we could expand out the list in several ways. DNS queries are helpful, but what else can we do to get more names to try?
• From whatever source, let’s say we get two names from verizon, ‘foundation’ and investor’…
Google Sets

- Although this is a simple example, we can throw these two words into Google Sets....
Expanding

• Then, we can take all these words and perform DNS host lookups against each of these combinations:

```
bash-2.05b$ for name in `cat list3`; do host $name.verizon.com | grep "has"; done
Business.verizon.com has address 206.46.230.36
foundation.verizon.com has address 216.251.248.19
Investor.verizon.com has address 198.92.149.185
Foundation.verizon.com has address 216.251.248.19
```
..this leads to a new hit, ‘business.verizon.com’.

Google sets allows you to expand on a list once you run out of options.
Fuzzing

• Given hosts with numbers and “predictable” names, we could fuzz the numbers, performing DNS lookups on those names…

• I’ll let Roelof at sensepost discuss this topic, however… =)
Limitless mapping possibilities…

• Once you get rolling with Google mapping, especially automated recursive mapping, you’ll be AMAZED at how deep you can dig into the layout of a target.
Port scanning

- Although crude, there are ways to do basic “portscanning” with Google.
- First, combine inurl searches for a port with the name of a service that commonly listens on that port… (optionally combined with the site operator)
Inurl -intext scanning

- Another way to go is to use a port number with inurl, combined with a negative intext search for that port number.

This search locates servers listening on port 8080.
Third party scanners

- When all else fails, Google for servers that can do your portscan for you!
Document Grinding and Database Digging

Documents and databases contain a wealth of information. Let’s look at ways to foster abuse of SQL databases with Google.
SQL Usernames

“Access denied for user”
“using password”
SQL Schemas

• Entire SQL Database dumps

```
# Dumping data for table 'config'
#
INSERT INTO config VALUES ( 'last_update', '');
#
# Table structure for table 'creatures'
#
CREATE TABLE creatures (  
id int(11) DEFAULT '0' NOT NULL auto_increment,  
name varchar(255) NOT NULL,  
owner int(11) DEFAULT '0' NOT NULL,  
carnivore enum('Y','N') DEFAULT 'Y' NOT NULL,  
weight int(11) DEFAULT '0' NOT NULL,  
speed int(11) DEFAULT '0' NOT NULL,
```

Adding ‘username’ or ‘password’ to this query makes things really interesting.
SQL injection hints

"ORA-00933: SQL command not properly ended"

Improper command termination can be abused quite easily by an attacker.

"Unclosed quotation mark before the character string"
SQL source

• Getting lines of SQL source can aid an attacker.

intitle:"Error Occurred" "The error occurred in"
Going after SQL passwords

```php
<?php
$host = "127.0.0.1";
$user = "cs3projo";
$password = "LTnM76mx5";
$database = "cs3projo"

mysql_connect($host, $user, $password);
@mysql_select_db($database) or die ("I cannot
?>
```
More SQL Passwords

• Question: What’s the SQL syntax that can be used to set a password?

• (TWO WORDS)

• One Answer: “Identified by”
More SQL Passwords

- The slightly more hardcore version...
## Various database detection queries

<table>
<thead>
<tr>
<th>Query</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>inurl:nuke filetype:sql</td>
<td>php-nuke or postnuke CMS dumps</td>
</tr>
<tr>
<td>filetype:sql password</td>
<td>SQL database dumps or batched SQL commands</td>
</tr>
<tr>
<td>filetype:sql &quot;IDENTIFIED BY&quot; -cvs</td>
<td>SQL database dumps or batched SQL commands, focus on &quot;IDENTIFIED BY&quot;, which can locate passwords</td>
</tr>
<tr>
<td>&quot;# Dumping data for table (username</td>
<td>user</td>
</tr>
<tr>
<td>&quot;#mysql dump&quot; filetype:sql</td>
<td>SQL database dumps</td>
</tr>
<tr>
<td>&quot;# Dumping data for table&quot;</td>
<td>SQL database dumps</td>
</tr>
<tr>
<td>&quot;# phpMyAdmin MySQL-Dump&quot; filetype:txt</td>
<td>SQL database dumps created by phpMyAdmin</td>
</tr>
<tr>
<td>&quot;# phpMyAdmin MySQL-Dump&quot; &quot;INSERT INTO&quot; -&quot;the&quot;</td>
<td>SQL database dumps created by phpMyAdmin (variation)</td>
</tr>
</tbody>
</table>

**SQL dump detection**

**Database detection**

<table>
<thead>
<tr>
<th>Query</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>filetype:cfm &quot;cfapplication name&quot; password</td>
<td>ColdFusion source code</td>
</tr>
<tr>
<td>filetype:mdb inurl:users.mdb</td>
<td>Microsoft Access user database</td>
</tr>
<tr>
<td>inurl:email filetype:mdb</td>
<td>Microsoft Access e-mail database</td>
</tr>
<tr>
<td>inurl:backup filetype:mdb</td>
<td>Microsoft Access backup databases</td>
</tr>
<tr>
<td>inurl:forum filetype:mdb</td>
<td>Microsoft Access forum databases</td>
</tr>
<tr>
<td>inurl:/db/main.mdb</td>
<td>ASP-Nuke databases</td>
</tr>
<tr>
<td>inurl:profiles filetype:mdb</td>
<td>Microsoft Access user profile databases</td>
</tr>
<tr>
<td>filetype:asp DBQ=&quot;* Server. MapPath(&quot;&quot;*.mdb&quot;)&quot;</td>
<td>Microsoft Access database connection string search</td>
</tr>
<tr>
<td>allinurl: admin mdb</td>
<td>Microsoft Access administration databases</td>
</tr>
</tbody>
</table>
Automation

Page Scraping in Perl
API querying in Perl
Page Scraping with Perl

- Thie Perl code, by James Foster, provides a good framework for “page scraping” Google results.
- This method relies on manually querying Google, and searching the resultant HTML for the “interesting stuff.”

```perl
#!/usr/bin/perl -w
use IO::Socket;

#Section 2
$query = '/search?hl=en&q=dog';
$server = 'www.google.com';
$port = 80;
```

We will be making socket calls. We need `IO::Socket`.

We hardcode our query (which we can make a parameter later), our Google server and our port number.
Page Scraping with Perl

```perl
sub socketInit() {
    $socket = IO::Socket::INET->new(
        Proto => 'tcp',
        PeerAddr => $server,
        PeerPort => $port,
        Timeout => 10,
    );

    unless($socket) {
        die("Could not connect to $server:$port");
    }

    $socket->autoflush(1);
}

Next we have a very generic socket initialization subroutine.
```
This subroutine sends the Google query (hardcoded above) and accepts one parameter, the Google query.

Google returned HTML is processed, and the line containing “of about” (our result line) is returned from this routine.
sub getTotalHits($) {
  my ($ourline) = @_;  
  $hits="";
  $index = index($ourline, "of about");
  $str = substr($ourline, $index, 30);
  @buf=split(/,\$str);
  for ($i = 0; $i < 30; $i++) {
    if ($buf[$i] =~ /[0-9]/) {
      $hits=$hits.$buf[$i];
    }
  }
  return $hits;
}

This subroutine takes one parameter (the results line from the Sendquery)

"of about is located"...

...the next 30 characters are grabbed...

...all the digits are removed....

...stored in $hits...

...and returned.

Results 1 - 10 of about 46,600 for "james foster". (0.49 seconds)
Page Scraping with Perl

The socket is initialized…

socketInit();
$string = sendQuery($query);
$totalhits = getTotalHits($string);

…the query is sent…

#Printing to STDOUT the Total Hits Retrieved from Google
print ($totalhits);

…the total hits are determined…

This piece of code drives all the subroutines.

…and printed out.
Another automation example might involve chopping up a CGI scanner's vulnerability file…

… converting the checks into Google queries, sending these queries to a Google scanner.

**CGI Scanning**

/iisadmpwd/
iisadmpwd/achg.htr
/iisadmpwd/aexp.htr
/iisadmpwd/aexp2.htr
/iisadmpwd/aexp2b.htr

inurl:/iisadmpwd/
inurl:/iisadmpwd/achg.htr
inurl:/iisadmpwd/aexp.htr
inurl:/iisadmpwd/aexp2.htr
inurl:/iisadmpwd/aexp2b.htr

intitle:index.of /iisadmpwd/
intitle:index.of /iisadmpwd/achg.htr
intitle:index.of /iisadmpwd/aexp.htr
intitle:index.of /iisadmpwd/aexp2.htr
intitle:index.of /iisadmpwd/aexp2b.htr
Web Servers, Login Portals, Network Hardware

Network devices can be sooo much fun to Google for…
Web File Browser

• This program allows directory walking, file uploading, and more.
VNC Servers (with client)

- VNC (Virtual Network Computing) allows you to control a workstation remotely.

The search is very basic

These sites launch a VNC Java client so you can connect! Even if password protected, the client reveals the server name and port.

Thanks to lester for this one!
Symantec Anti-Virus SMTP Gateways
Axis Print Servers

Print server administration, Google-style!

Thanks to murfie for this one!
Xenix, Sweex, Orite Web Cams

One query, many brands of live cams!

Thanks to server1 for this one!
Active WebCam

Thanks klouv!
Toshiba Network Cameras

intitle:"toshiba network camera - User Login"

Found by WarriorClown!
Speedstream DSL Routers

- Home broadband connectivity… Googled.

Who do you want to disconnect today?

Found by m00d!
Belkin Routers

- Belkin routers have become a household name in connected households. The management interface shouldn’t show up on Google… but it does.
Printers

- Trolling printers through Google can be fun, especially when you can see and download what others are printing...

Religion…

And aphrodisiacs? Hrmmm…

Thanks JimmyNeutron!
Firewalls - Smoothwall

Uh oh… this firewall needs updating…

Thanks Milkman!
Firewalls - IPCop

Uh oh… this one needs updating too!

Thanks Jimmy Neutron!
IDS Data: ACID

- SNORT IDS data delivered graphically, served up fresh
Open Cisco Devices

Thanks Jimmy Neutron!
Cisco Switches

Cisco Systems

Accessing Cisco WS-C3550-48 "Switch"

- Web Console - Manage the Switch through the web interface.
- Telnet - to the router.
- Show interfaces - display the status of the interfaces.
- Show diagnostic log - display the diagnostic log.
- Monitor the router - HTML access to the command line interface at level 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
- Connectivity test - unavailable, no valid nameserverdefined.
- Extended Ping - Send extended ping commands.
- Show tech-support - display information commonly needed by tech support.

Help resources

Thanks Jimmy Neutron!
Wide Open PHP Nuke Sites

- PHP Nuke allows for the creation of a full-featured web site with little effort.

Too lazy to install PHP Nuke? Own someone else’s site instead!

Thanks to arrested for this beauty!
Open PHP Nuke... another way...

Click here, create superuser!
Security Cameras

- Although many cameras are multi-purpose, certain brands tend to be used more for security work.
Security Cameras

Not sure what “Woodie” is, but I'm not clicking it....

Thanks murfie!
Time-lapse video recorders

- A staple of any decent security system, these camera control units have gotten high-tech. And Googlable…

The search is no big deal…

Then there’s the pesky login box…
Time lapse video recorders

...multiple live security camera views...

Even doofus hackers know how to use default passwords to get...

...and historical records of recorded video feeds

Thanks to stonersavant for this beauty!
UPS Monitors

Getting personal with Power System monitors...

Thanks yeseins!
Oh wait.. Wrong kind of UPS…this is package tracking hacking… =P

Thanks Digital Spirit!
Hacking POWER Systems!

• Ain’t technology grand? This product allows web management of power outlets!

Google search locates login page. What does any decent hacker do to a login page?
Hacking Power Systems!

Who do you want to power off today?

Thanks to JimmyNeutron for this beauty!
Google Phreaking

- Question… Which is easier to hack with a web browser?

A: Sipura SPA 2000 IP Telephone

B: Vintage 1970’s Rotary Phone

QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.
Sipura SPA IP Telephone

How about Googling for the last number your friend dialed?

Or the last number that dialed them?

Thanks stonersavant!!!
Videoconferencing

Who do you want to disconnect today?

Thanks yeseins!!!
PBX Systems

- Web-based management interfaces open the door for a creative Google Hacker.

See the “logout”? We’re already logged in! We don’t need no steenkin password!
PBX Systems

No password required. Even a novice web surfer can become a “PBX hacker”. =)

Thanks to stonersavant for this great find!
Usernames, Passwords and Secret Stuff, oh my!

There’s all sorts of stuff out there that people probably didn’t mean to make public. Let’s take a look at some examples…
DCIM

What’s DCIM?

Digital camera image dumps....

Thanks xlockex!
MSN Contact Lists

MSN contact lists allow an attacker to get ‘personal’

Thanks to harry-aac!
Old School! Finger…

Google Hacking circa 1980!!?!?

Thanks to Jimmy Neutron!
Norton AntiVirus Corporate Passwords

Encrypted, but yummy (and crackable)!

Thanks MILKMAN!
Open SQL servers

Already logged in, no hacking required!

Thanks Quadster!
ServU FTP Passwords

ServU FTP Daemon passwords, super encrypto! =P

Thanks to vs1400 for this one!
Netscape History Files

Oops.. POP email passwords!

Thanks to digital.revolution for this one!
IPSec Final Encryption Keys

I only skimmed ‘Applied Cryptography’.. But this looks bad...

Thanks MILKMAN!
Explorer. EXPLORER!?!?!

```
C:\WINNT\system32\n
<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Modified date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swinnts.inf</td>
<td>666777777777777</td>
<td>04/30/2000 08:47:11 PM</td>
</tr>
<tr>
<td>Dir</td>
<td></td>
<td>04/10/2004 04:48:34 AM</td>
</tr>
<tr>
<td>Dir</td>
<td></td>
<td>04/10/2004 04:48:34 AM</td>
</tr>
<tr>
<td>File: 12520437.cpx</td>
<td>215655555555555</td>
<td>03/29/2002 04:32:58 PM</td>
</tr>
<tr>
<td>File: 12520850.cpx</td>
<td>22332222222222222</td>
<td>03/29/2002 04:32:58 PM</td>
</tr>
<tr>
<td>File: Mtmp</td>
<td>4096409640964096</td>
<td>04/20/2000 09:27:09 PM</td>
</tr>
<tr>
<td>File: aamon.dll</td>
<td>32016320163201632016</td>
<td>12/07/1999 01:00:00 PM</td>
</tr>
<tr>
<td>File: abrsg.dll</td>
<td>8976897689768976</td>
<td>12/07/1999 12:18:32 PM</td>
</tr>
<tr>
<td>File: access.cpl</td>
<td>673446734467344</td>
<td>12/07/1999 01:00:00 PM</td>
</tr>
<tr>
<td>File: accesv.mib</td>
<td>13753137531375313753</td>
<td>12/07/1999 01:00:00 PM</td>
</tr>
<tr>
<td>File: acctres.dll</td>
<td>64512645126451264512</td>
<td>08/17/2001 10:32:38 PM</td>
</tr>
<tr>
<td>File: accviz.exe</td>
<td>150800150800150800150800</td>
<td>11/22/2002 04:58:56 PM</td>
</tr>
<tr>
<td>File: acelpdec.ax</td>
<td>61952619526195261952</td>
<td>12/07/1999 01:00:00 PM</td>
</tr>
<tr>
<td>File: acledit.dll</td>
<td>131856131856131856131856</td>
<td>12/07/1999 01:00:00 PM</td>
</tr>
<tr>
<td>File: aclui.dll</td>
<td>78096780967809678096</td>
<td>07/22/2002 08:05:04 PM</td>
</tr>
<tr>
<td>File: ace.mib</td>
<td>33298332983329833298</td>
<td>12/07/1999 01:00:00 PM</td>
</tr>
<tr>
<td>File: acsetupc.dll</td>
<td>43684368436843684368</td>
<td>12/07/1999 01:00:00 PM</td>
</tr>
</tbody>
</table>

What do you want to delete today???

Thanks JimmyNeutron!
More Explorers?!?!

Why hack when you can... click? =)
More Explorers?!?!

Thanks JimmyNeutron!

sigh…

Thanks JimmyNeutron!
Sensitive Government Documents

• Question: Are sensitive, non-public Government documents on the web?
• Answer: Yes.

Once these documents hit the Net, the media has a feeding frenzy, and people start copying and posting the docs...
Although unclassified, this document was obviously not meant to be posted online.
FOUO "Prevention Guides", like this 19 page beauty, can give bad guys horrible ideas.
Locked out!

- Some sites lock down sensitive data..

- However, the Google cache image still remains.
Credit card info on the web?

• How can this happen? Let’s take a tour of some of the possibilities…
Court Documents

• Court cases sometimes give TONS of detail about cases, especially fraud.
Court Documents

account number: Platinum

account #:

account #: account

account #: account

account #: account

account #: account

Bank #:

statement # account #
**Court Documents**

- How much detail is too much detail? =)
Court Documents

- Of course, fraud accounts are closed pretty quickly, no?

<table>
<thead>
<tr>
<th>BANK</th>
<th>ACCOUNT NUMBER</th>
<th>ACCOUNT NAME</th>
<th>AMOUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>t Bank</td>
<td>05 13</td>
<td></td>
<td>$1</td>
</tr>
<tr>
<td>Bank</td>
<td>13 23</td>
<td></td>
<td>$3</td>
</tr>
<tr>
<td>Bank</td>
<td>90 97</td>
<td></td>
<td>$7</td>
</tr>
<tr>
<td>Bank</td>
<td>71 02</td>
<td></td>
<td>$2</td>
</tr>
<tr>
<td>Bank</td>
<td>10 96</td>
<td></td>
<td>$3</td>
</tr>
<tr>
<td>Bank</td>
<td>28 27</td>
<td></td>
<td>$11</td>
</tr>
<tr>
<td>Bank</td>
<td>63 91</td>
<td></td>
<td>$2</td>
</tr>
<tr>
<td>Bank</td>
<td>90 10</td>
<td></td>
<td>$2</td>
</tr>
</tbody>
</table>
A tale of a corn snake

• Is this for real? Either way it’s pretty sad...
Getting shell.. the easy way

• Now I’ve heard the term ‘using your credit card online’ but this is ridiculous!

applying for a shell acct

- To: helpdesk@example.org
- Subject: applying for a shell acct
- From: johndoe@example.com
- Date: Sat, 09 Oct 1999 14:12:00 GMT

Hi there I am interested to buy a shell acct for my son.
I am busy because of some reasons. This is my only way to communicate with my son... so I want to buy a shell acct using my credit card.

The login name and password will be:
login name: johndoe
password: secret

Below are the details about me.
name: John Doe
age: 35
Address: 123 Main St, Anytown, CA 90210
Fortune, CA, 95540

Tel#: 555-1212
Billing Information:
Name on card: John Doe
Credit card #: 42888360198347
Expiration date: 01/00
Type of card: VISA

Thank you and I hope to hear from you soon.
Some people just don’t get it….

<table>
<thead>
<tr>
<th>Questions &amp; Offers between</th>
<th>Question: hi there here is my credit card details below</th>
<th>Seller Answer: Ok, I will attempt to charge your card first thing tomorrow when I am back at the office. Also, what address do you want this shipped to? Thanks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buyer</td>
<td>card type: Visa name on card:</td>
<td>Mar 23 14:10PDT</td>
</tr>
<tr>
<td></td>
<td>card number:</td>
<td>expire date: 02/05 cvv2 number:</td>
</tr>
<tr>
<td></td>
<td>expire date: 10/05 CVV2 number:</td>
<td></td>
</tr>
<tr>
<td>Seller</td>
<td>card type: Visa</td>
<td>Mar 23 16:27PDT</td>
</tr>
<tr>
<td></td>
<td>name on card:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>card number:</td>
<td></td>
</tr>
</tbody>
</table>
Getting serialz... wha-hay!! and MORE!

- This is a very generous person. He's willing to give his software serial numbers and his credit card info to the whole world. Generosity like this could change the world.
Police Crime reports

• Two questions:
  • Are police reports public record?
    • YES.
  • Are they on the web?
    • YES.
  • Many states have begun placing campus police crime reports on the web. Students have a right to know what crimes take place on campus.
Crime shouldn’t pay…

• I’m thinking there should be a process for filtering these reports.

• A few might fall through the cracks….
Expense Reports

• It’s not uncommon for expense reports to be generated. This one is for a county.
Expense Reports

• Bank account numbers….
Expense Reports

- Bank loan information… $20,000 + transactions
Expense Reports

• Oh boy…
Expense Reports...

- Somebody has to pay for all this stuff....
Expense Reports

- That’s one heck of a video series…. $300+
Credit cards… Google hacker’s gold…

- The legend of finding credit cards online is true…
- I just get bored sifting through them all…
Credit Listings

- Visa
  - Expire: 2/5
  - US

- MasterCard
  - Expire: 1/6
  - US

- Visa
  - Expire: 2/6
  - US

- MasterCard
  - Expire: 1/7

- Visa
  - Expire: 2/7
  - US

- MasterCard
  - Expire: 1/8

- Visa
  - Expire: 2/8
  - US
Pick a card any card...

...pick a card. We take 'em all!
Credit Validation

Question: What keeps someone from using a pilfered credit card number and expiration date to make an online purchase?

• Answer: That little code on the back of the card.

• Bonus question: What’s that code called?

• Answer: A “CVV” code.
Credit Card Numbers, Expiration Date and CVV numbers, oh my!
That’s not all….

• Credit cards are sooo 1990’s =)
Getting more personal

- Question: What’s the one 9 digit number you shouldn’t give to ANYONE?
- Answer: SSN
- Bonus question: What can you do with someone’s SSN?
- Answer: Steal their identity.

- How do SSN’s get on the web? Let’s take a look at some possibilities.
SSN’s in source code

- Well, they could be hardcoded into a healthcare system... and uhmmm... put on the web...
Crime shouldn’t pay…

• Remember the police reports? Since the credit card accounts in them are no good, maybe we should troll them some more….
SSN’s - Police Reports
SSN’s

• Students have a right to know…
Social Security Numbers

- Many privacy violations are self-inflicted…
Social Security Numbers

- Schools are notorious… Grades posted w/ student’s SSN’s
Social Security Numbers

- Once you get a lock on a grade list, the results fan out as you explore the site.
There's no shortage of examples...
Social Security Numbers

• In order to steal someone’s identity, you need names. SSN’s with names are usually blocked… aren’t they?
Social Security Numbers

Google's cache says otherwise…
A tale of one city

• A city document outlining residents who are in debt to the city… A little report of names, addresses, amount owed and SSN numbers…

<table>
<thead>
<tr>
<th>ACCT #</th>
<th>NAME (First Name Last Name)</th>
<th>SSN</th>
<th>ADDRESS</th>
<th>AMT DUE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Additional entries and details follow)
A tale of one city

• Or perhaps more than a little report…
A tale of one city

- Hundreds of city residents’ personal information posted to the web… 90% including SSN and address.
What we’ve done...

• We’ve skimmed “Google Hacking for Penetration Testers” by Syngress Publishing, which doesn’t seem to suck.
• We’ve looked at some great tools by Roelof Temmingh. Check out Sensepost.com.
• We’ve invaded the privacy of millions.
• We’re all still awake. Right?
Thanks!

• Thanks to God for the gift of life.
• Thanks to my family for the gift of love.
• Thanks to my friends for filling in the blanks.
• Thanks to the moderators of ihackstuff.com: Murfie, Jimmy Neutron, ThePsyko, Wasabi, l00m, Stonersavant
• Thanks to Roelof T for the great code, and to the current Google Masters: murfie, jimmyneutron, klouw, l00m, stonersavant, MILKMAN, ThePsyko, cybercide, yeseins, wolveso, Deadlink, crash_monkey, zoro25, digital.revolution, Renegade334, wasabi, urban, sfd, mlynch, Peefy, Vipsta, noAcces, brasileiro, john, Z!nCh