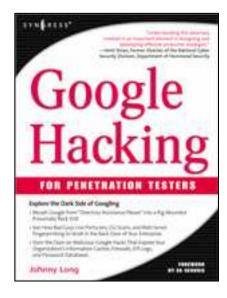
Google Hacking for Penetration Testers

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

search

Advanced operators can be combined in some cases.

Purpose Operator Mixes with Can be Does search work in other used operators? alone? Web **Images** Groups News intitle Search page yes yes yes yes yes yes title Search page allintitle no yes yes yes yes yes title Search URL inurl yes like intitle ves yes yes not really allinurl Search URL ves like intitle no yes yes yes Search filetype not really yes no yes yes no specific files not really allintext Search text of yes yes yes yes yes page only site Search not really ves yes yes yes no specific site Search for ∕link no yes yes no no not really links to pages inanchor Search link not yes yes yes yes yes anchor text really Locate no not really numrange ves yes no yes number daterange Search in yes no yes not not not really date range really really Group author author ves ves no no ves not really search Group name not really not really group yes no no yes search Group subject like like like intitle insubject yes yes yes intitle search intitle Group msgid msgid no yes not not yes not really

really

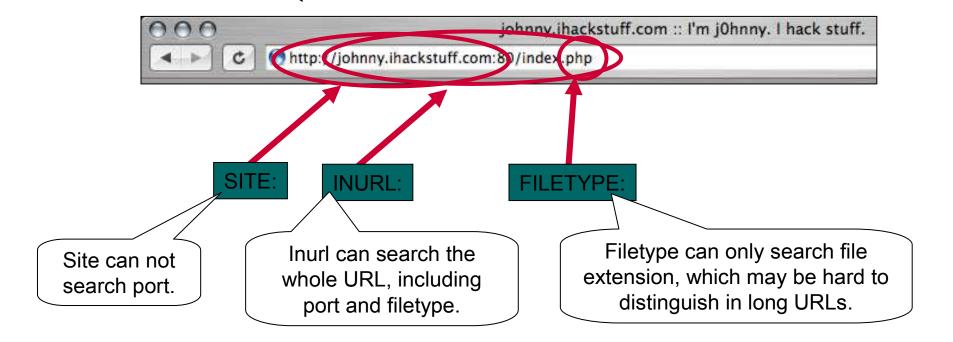
really

In other cases, mixing should be avoided.

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

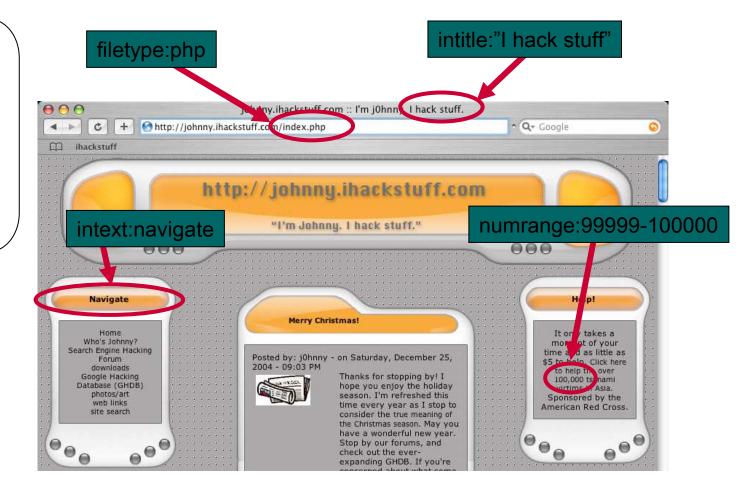
Crash course in advanced operators

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



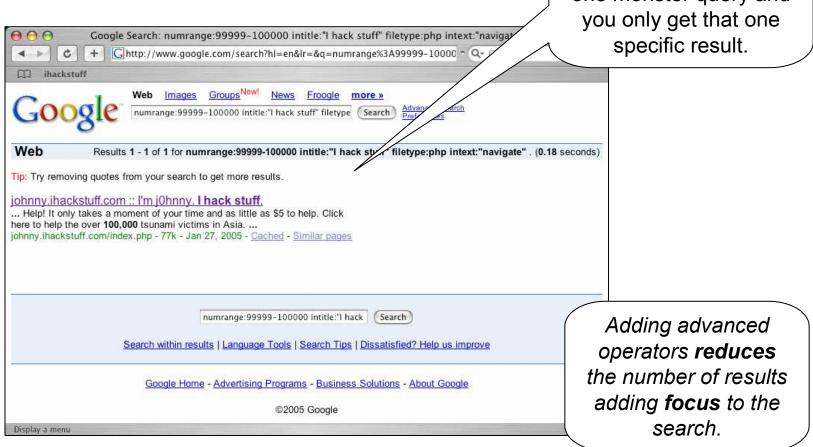
Advanced Google Searching

There are many ways to find the same page. These individual queries could all help find the same page.



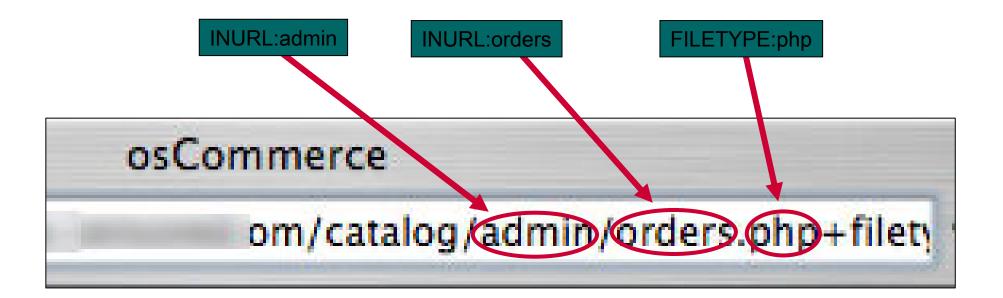
Advanced Google Searching

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



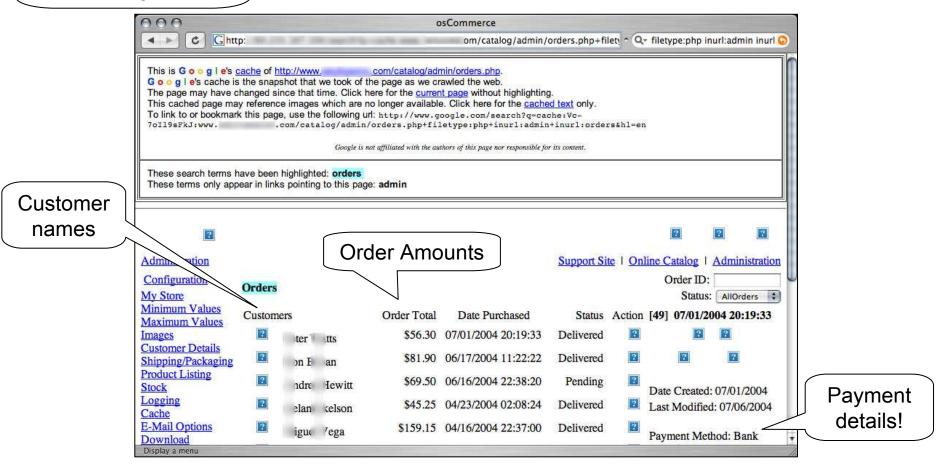
Google Hacking Basics

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



Google Hacking Basics

...can return devastating results!



Google Hacking Basics

Let's take a look at some basic techniques:

Anonymous Googling Special Characters



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



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

21:39:24.648422 IP 192.168.2.32.51670 > 64.233.167.104.80 21·39·24 719067 IP 64 233 167 104 80 > 192 168 2 32 51670 21:39:24.720351 IP 64.233.167.104.80 > 192.168.2.32.51670 21:39:24.731503 IP 192.168.2.32.51670 > 64.233.167.104.80 21:39:24.897987 IP 192.168.2.32.51672 > 82.165.25.125.80 21:39:24.902401 IP 192.168.2.32.51671 > 82.165.25.125.80 21:39:24.922716 IP 192.168.2.32.51673 > 82.165.25.125.80 21:39:24.927402 IP 192.168.2.32.51674 > 82.165.25.125.80 21:39:25.017288 IP 82.165.25.125.80 > 192.168.2.32.51672 21:39:25.019111 IP 82.165.25.125.80 > 192.168.2.32.51672 21:39:25.019228 IP 192.168.2.32.51672 > 82.165.25.125.80 21:39:25.023371 IP 82.165.25.125.80 > 192.168.2.32.51671 21:39:25.025388 IP 82.165.25.125.80 > 192.168.2.32.51671 21:39:25.025736 IP 192.168.2.32.51671 > 82.165.25.125.80 21·39·25 043418 IP 82 165 25 125 80 > 192 168 2 32 51673 21:39:25.045573 IP 82.165.25.125.80 > 192.168.2.32.51673 21:39:25.045707 IP 192.168.2.32.51673 > 82.165.25.125.80 21:39:25.052853 IP 82.165.25.125.80 > 192.168.2.32.51674

This is Google.

This is Phrack.

We touched Phrack's web server. We're not anonymous.

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

```
0x0040 0d6c 4745 5420 2f67 7266 782f 3831 736d 0x0050 626c 7565 2e6a 7067 2048 5454 502f 312e 0x0060 310d 0a48 6f73 743a 2077 7777 2e70 6872 0x0070 6163 6b2e 6f72 670d 0a43 6f6e 6e65 6374 0x0080 696f 6e3a 206b 6565 702d 616c 6976 650d 0x0090 0a52 6566 6572 6572 3a20 6874 7470 3a2f 0x00a0 2f36 342e 3233 332e 3136 312e 3130 342f 0x00b0 7365 6172 6368 3f71 3d63 6163 6865 3a4c 0x00c0 4251 5a49 7253 6b4d 6755 4a3a 7777 772e 0x00d0 7068 7261 636b 2e6f 7267 2f2b 2b73 6974 0x00e0 653a 7777 772e 7068 7261 636b 2e6f 7267 0x00f0 2b70 6872 6163 6b26 686c 3d65 6e0d 0a55
```

An image loaded!

ack.org..Connect ion:.keep-alive. .Referer:.http:/ /64.233.161.104/ search?q=cache:L BQZIrSkMgUJ:www. phrack.org/++sit e:www.phrack.org +phrack&hl=en..U



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

```
23:46:53.996067 IP 192.168.2.32.52912 > 64.233.167.104.80

23:46:54.025277 IP 64.233.167.104.80 > 192.168.2.32.52912

23:46:54.025345 IP 192.168.2.32.52912 > 64.233.167.104.80

23:46:54.025465 IP 192.168.2.32.52912 > 64.233.167.104.80

23:46:54.094007 IP 64.233.167.104.80 > 192.168.2.32.52912

23:46:54.124930 IP 64.233.167.104.80 > 192.168.2.32.52912

23:46:54.127202 IP 64.233.167.104.80 > 192.168.2.32.52912

23:46:54.128762 IP 64.233.167.104.80 > 192.168.2.32.52912

23:46:54.128836 IP 192.168.2.32.52912 > 64.233.167.104.80

23:47:54.130200 IP 192.168.2.32.52912 > 64.233.167.104.80

23:47:54.154500 IP 64.233.167.104.80 > 192.168.2.32.52912

23:47:54.154596 IP 192.168.2.32.52912 > 64.233.167.104.80
```

What made the difference? Let's compare the two URLS:

Original:

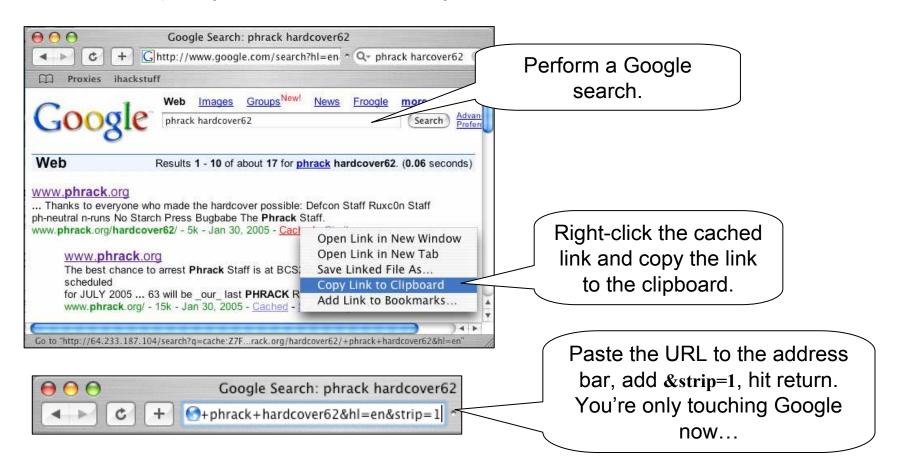
http://64.233.187.104/search?q=cache:Z7FntxDMrMIJ:www.phrack.org/hardcover62/+phrack+h ardcover62&hl=en

Cached Text Only:

http://64.233.187.104/search?q=cache:Z7FntxDMrMIJ:www.phrack.org/hardcover62/+phrack+h ardcover62&hl=en&lr=&strip=1

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

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

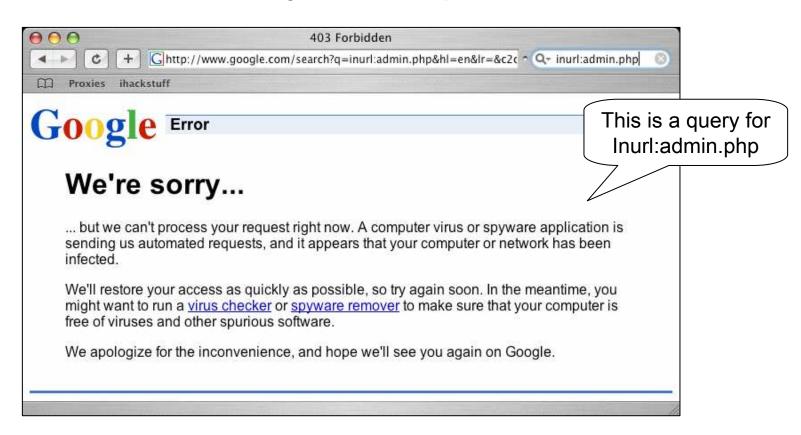


Special Search Characters

- 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.'



Google Hacker's workaround

Our original query looks like this:

http://www.google.com/search?q=inurl:admin.php&hl=en&lr=&c2coff=1&start=10&sa=N

Stripped down, the query looks like this:

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

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

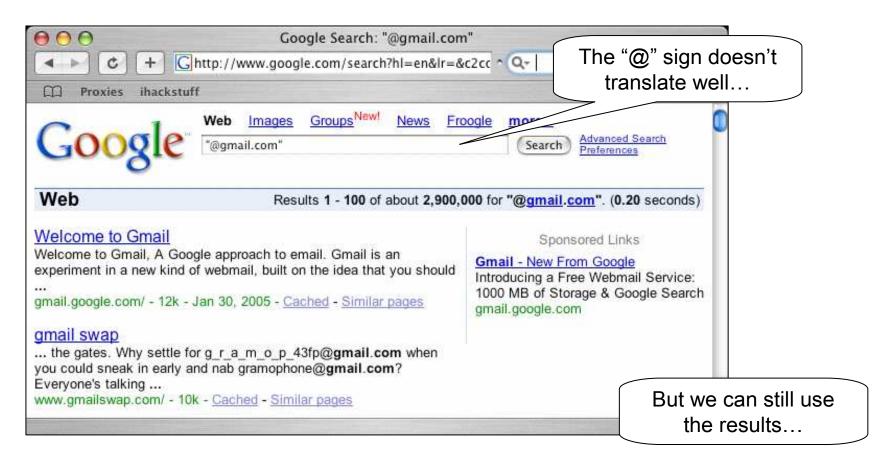
This works in the web interface as well.

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.



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

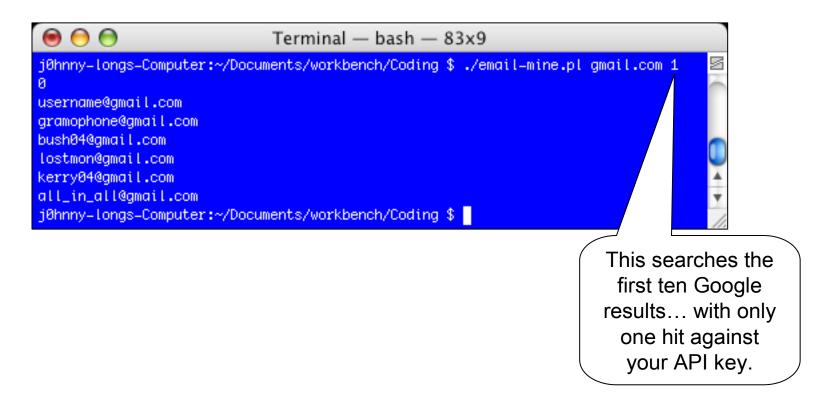
 We could then use regular expressions (like this puppy by Don Ranta) to troll through the results:

 $\begin{array}{l} [a-zA-Z0-9._-]+@(([a-zA-Z0-9_-]\{2,99\} \setminus)+[a-zA-Z]\{2,4\})|((25[0-5]|2[0-4][0-9]|1[0-9][0-9]|[1-9][0-9]|1-9] \\ 9][0-9]|[1-9]) \setminus (25[0-5]|2[0-4][0-9]|1[0-9]|1-9][0-9]|[1-9]) \setminus (25[0-5]|2[0-4][0-9]|1-9][0-9]|1-9] \\ 9]|[1-9]|0-9]|[1-9]) \setminus (25[0-5]|2[0-4][0-9]|1-9]|1-9][0-9]|1-9]|1-9])) \end{array}$

 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:



More Email Automation

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

movabletype@gmail.com fakubabe@gmail.com lostmon@gmail.com label@gmail.com charlescapps@gmail.com billgates@gmail.com ymtang@gmail.com tonyedgecombe@gmail.com ryawillifor@gmail.com jruderman@gmail.com itchy@gmail.com gramophone@gmail.com poojara@gmail.com london2012@gmail.com bush04@gmail.com fengfs@gmail.com username@gmail.com madrid2012@gmail.com somelabel@gmail.com bartjcannon@gmail.com fillmybox@gmail.com silverwolfwsc@gmail.com all in all@gmail.com mentzer@gmail.com kerry04@gmail.com presidentbush@gmail.com prabhav78@gmail.com

More email address locations

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

These queries locate email addresses in more "interesting" locations...

More email address locations

Query	Description
filetype:eml eml +intext: "Subject" +intext:"From"	Outlook express e-mail files contain e-mails with full headers
intitle:index.of inbox dbx	Outlook Express e-mail folder
filetype:wab wab	Outlook Mail address books contain sen- sitive e-mail information
filetype:pst inurl:"outlook.pst"	Outlook PST files can contain e-mails, cal- endaring, and address information
filetype:mbx mbx intext:Subject	Outlook versions 1–4 or Eudora mailbox files contain sensitive e-mail information
inurl:cgi-bin/printenv	Printenv script can reveal lots of informa- tion, including e-mail addresses and server information
inurl:forward filetype:forward -cvs	UNIX user e-mail forward files can list e- mail addresses
(filetype:mail filetype:eml filetype:mbox filetype:mbx) intext:password subject	Various generic e-mail files
"Most Submitted Forms and Scripts" "this section"	WebTrends statistics pages reveal directory information, client access statis- tics, e-mail addresses, and more
filetype:reg reg +intext: "internet account manager"	Windows registry files can reveal information such as usernames, POP3 passwords, e-mail addresses, and more
"This summary was generated by www.stat"	Wwwstat statistics information can reveal directory info, client access statistics, email addresses, and more

These queries locate email addresses in more "interesting" locations...

Network Mapping

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

 the site: operator narrows a search to a particular site, domain or subdomain.

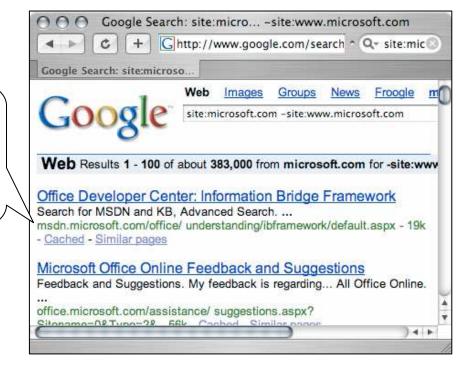


site: microsoft.com



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

Notice that the obvious "www" is missing, replaced by more interesting domains.



site: microsoft.com
-site:www.microsoft.com

 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

 $\hbox{-site:} support.microsoft.com$

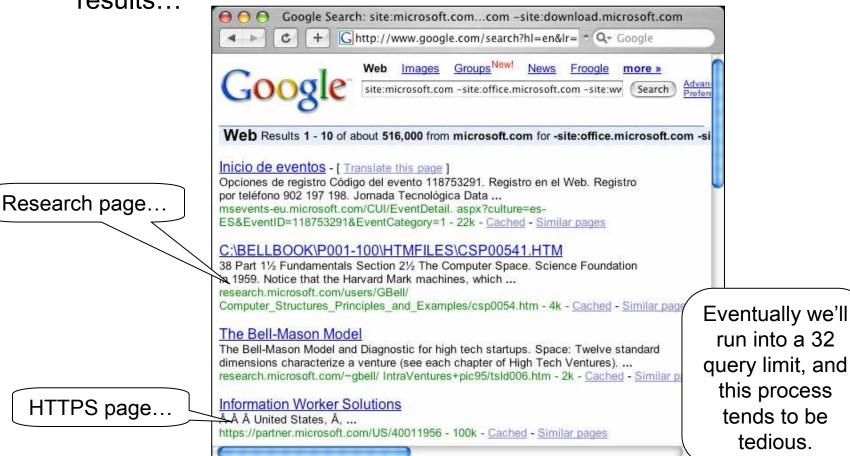
-site:download.microsoft.com

-site:office.microsoft.com

. . .

The results of such a big query reveal more interesting

results...



Intermediate Site Crawling

Using lynx to capture the Google results 900 Terminal - ssh - 80x24 page... z.v5b\$ lynx -dump "http://www.google.com/search?\ q=site:microsoft.com+-www.microsoft.com&num=100" > test.html -bash-2.05b\$ sed -n 's/\. [[:alpha:]]*:\/\/[[:alnum:]]*.microsoft.com\//& /p' te st.html | awk "{print \$2}" | sort -u http://download.microsoft.com/ http://go.microsoft.com/ http://msdn.microsoft.com/ ..returns the ..and sed and http://msevents.microsoft.com/ http://murl.microsoft.com/ same awk to process http://office.microsoft.com/ http://protect.microsoft.com/ results. the HTML... http://research.microsoft.com/ https://s.microsoft.com/ http://support.microsoft.com/ -bash-2.05b\$

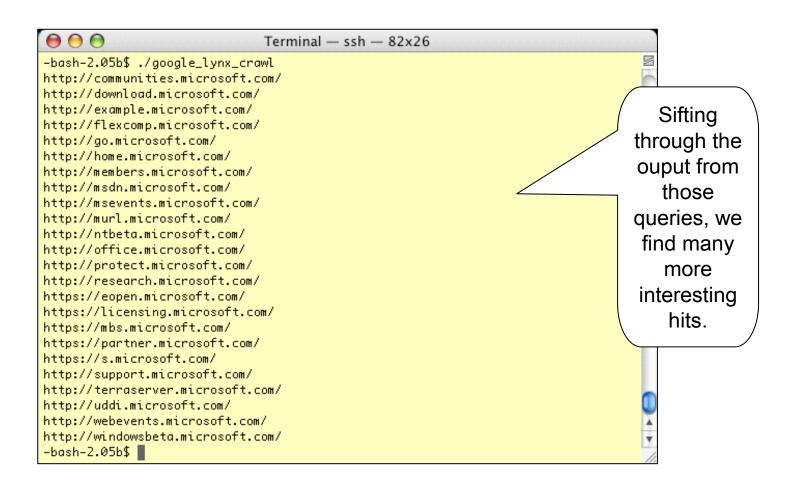
So what?

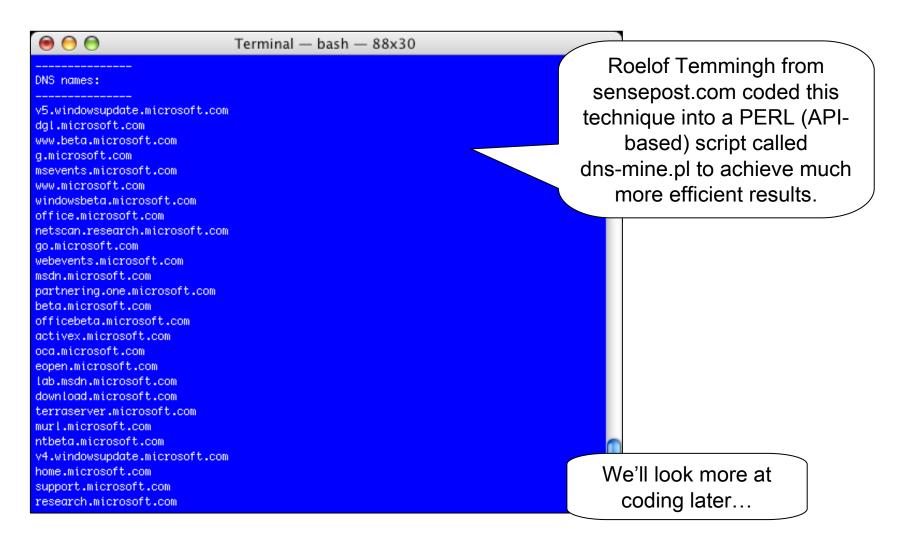
- Well, honestly, host and domain enumeration isn't new, but we're doing this without sending any packets to the target we're analyzing.
- This has several benefits:
 - Low profile. The target can't see your activity.
 - Results are "ranked" by Google. This means that the most public stuff floats to 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!

- 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.



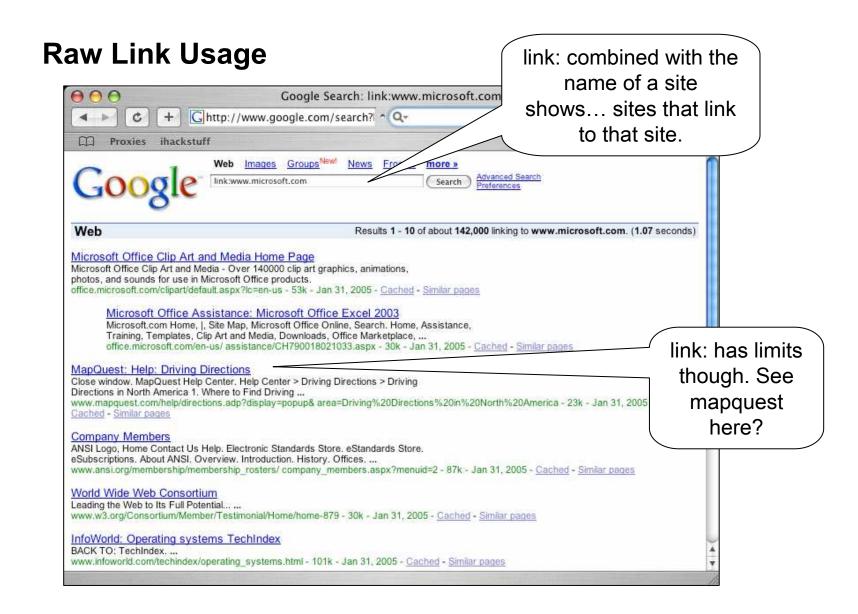
Searching for multiple common words like "web", "site", "email", and "about" along with site... appended to a file...



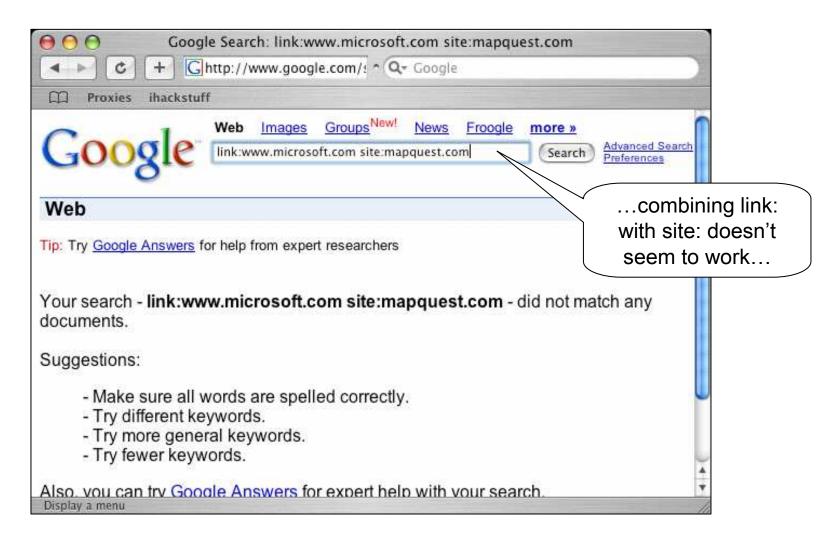


Too much noise, not enough signal...

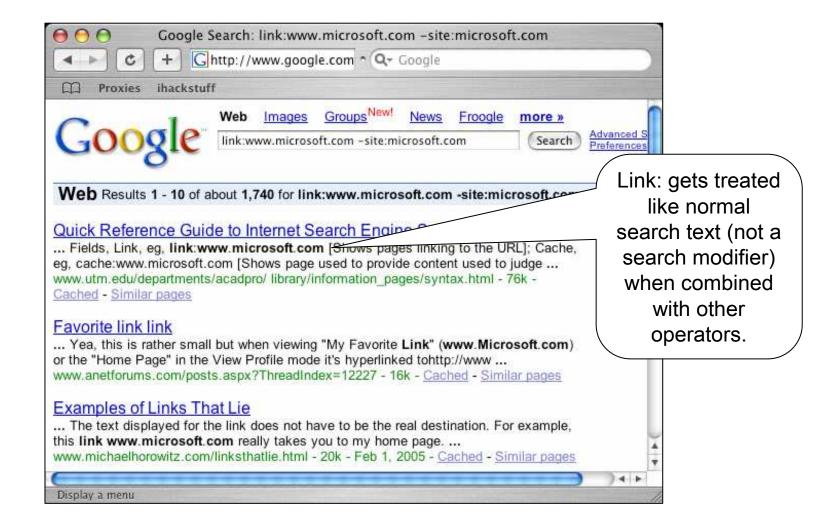
- Getting lists of hosts and (sub)domains is great. It gives you more targets, but there's another angle.
- Most systems are only as secure as their weakest link.
- If a poorly-secured company has a trust relationship with your target, that's your way in.
- Question: How can we determine site relationships with Google?
 - One Answer: the "link" operator.



Link has limits



Link has limits



Link has other limits

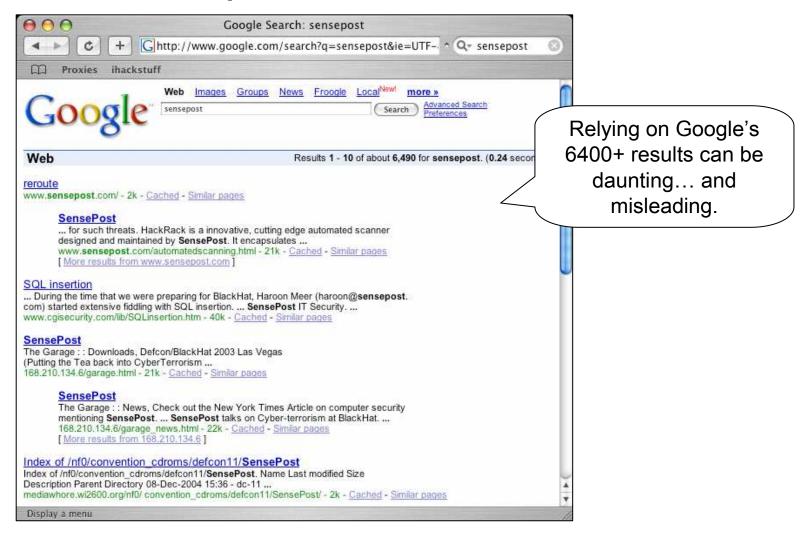
Google Search: link:www.microsoft.com Knowing that these Ghttp://www.google.com/search? a Qsites link to hackstuff www.microsoft.com News Froogle more » is great, but how link:www.microsoft.com relevant is this information? Results 1 - 10 of about 142,000 linking to www.microsoft.com. (1.07 seconds) Office Clip Art and Media Home Page ffice Clip Art and Media - Over 140000 clip art graphics, animations, d sounds for use in Microsoft Office products. rosoft.com/clipart/default.aspx?lc=en-us - 53k - Jan 31, 2005 - Cached - Similar pages Microsoft Office Assistance: Microsoft Office Excel 2003 Microsoft.com Home, J. Site Map, Microsoft Office Online, Search. Home, Assistance, Training, Templates, Clip Art and Media, Downloads, Office Marketplace, ... office.microsoft.com/en-us/assistance/CH790018021033.aspx - 30k - Jan 31, 2005 - Cached - Similar pages MapQuest: Help: Driving Directions Close window, MapQuest Help Center, Help Center > Driving Directions > Driving Directions in North America 1. Where to Find Driving ... www.mapquest.com/help/directions.adp?display=popup& area=Driving%20Directions%20in%20North%20America - 23k - Jan 31, 2005 -Cached - Similar pages Company Members Do we necessarily care about ANSI Logo, Home Contact Us Help. Electronic Standards Store. eStandards Store. eSubscriptions, About ANSI, Overview, Introduction, History, Offices, ... Google-ranked relationships? www.ansi.org/membership/membership rosters/ company members.aspx?menuid=2 - 87k -World Wide Web Consortium How do we get to REAL Leading the Web to Its Full Potential www.w3.org/Consortium/Member/Testimonial/Home/home-879 - 30k - Jan 31, 2005 - Cached relationships? InfoWorld: Operating systems TechIndex BACK TO: TechIndex. ... www.infoworld.com/techindex/operating_systems.html - 101k - Jan 31, 2005 - Cached - Similar pages

- Sensepost to the rescue again! =)
- BiLE (the Bi-directional Link Extractor), available from 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...

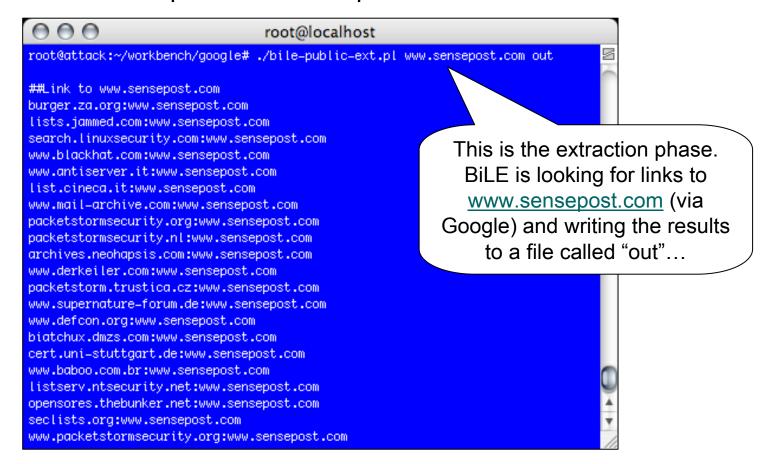
- A link from a site weighs more than a link to a site
 - Anyone can link to a site if they own web space (which is free to all)
- A link from a site with a lot of links weighs less that a link from a site with a small amount of links
 - This means specifically outbound links.
 - If a site has few outbound links, is is probably lighter.
 - There are obvious exceptions like link farms.

- A link to a site with a lot of links to the site weighs less that 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.

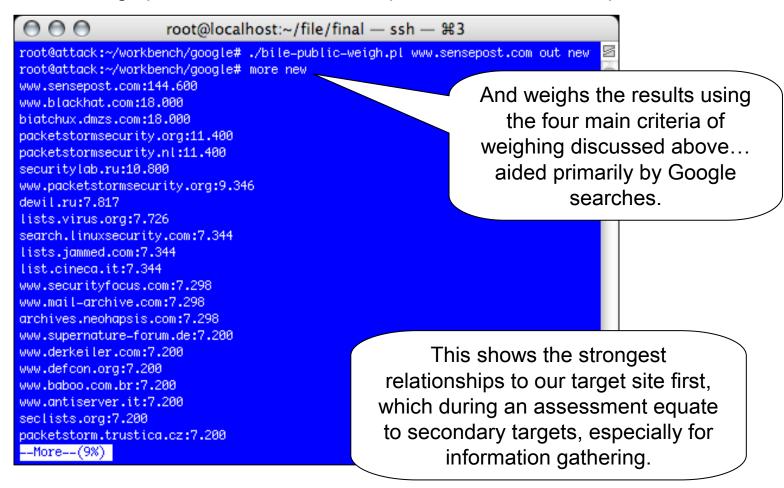
Who is Sensepost?



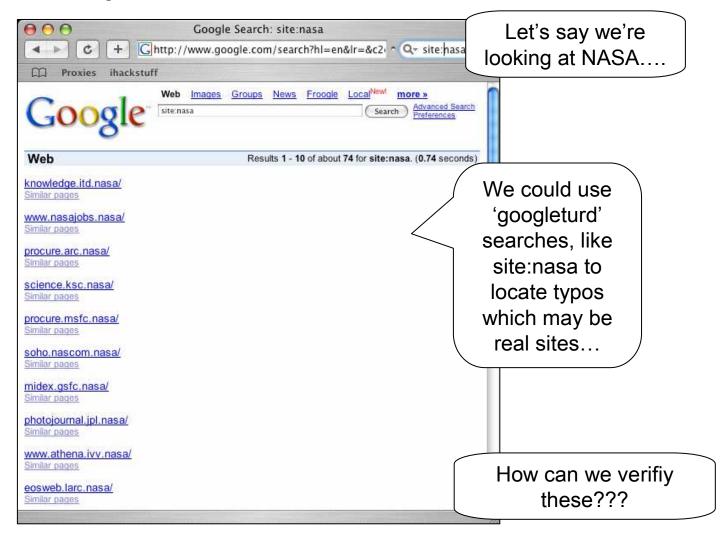
 It seems dizzying to pull all this together, but BiLE does wonders. Let's point it at sensepost.com:



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

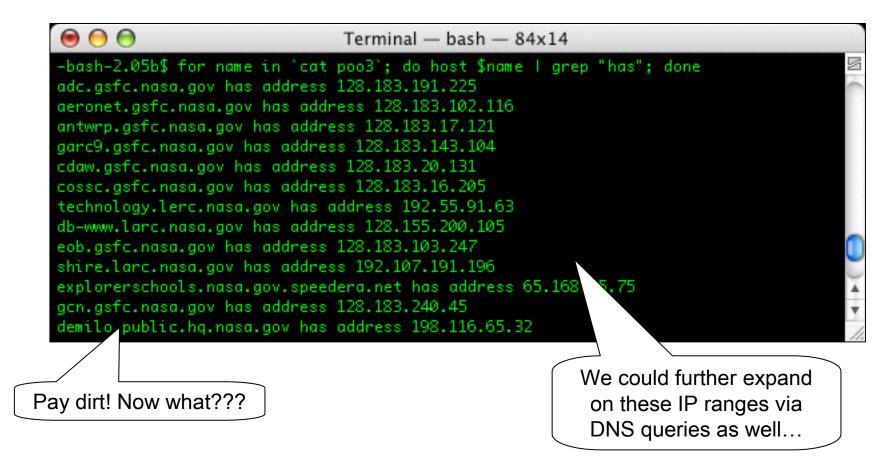


The next step...



Host verification...

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



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'...

[PDF] Verizon's 2003 Annual Report - Investor Information

File Format: PDF/Adobe Acrobat

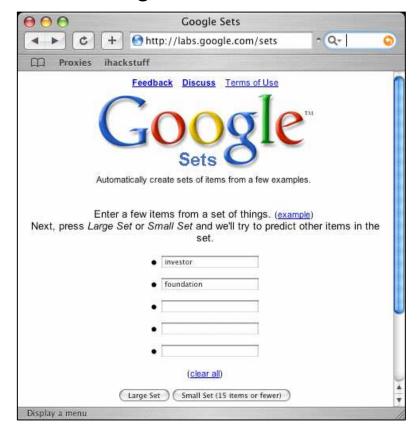
Page 1. Registered Shareowner Services Questions or requests for assistance regarding changes to or transfers of your registered ...

investor.verizon.com/2003annual/ download/vz_Investor_info.pdf - Similar pages

foundation.verizon.com/cybergrants/plsql/incomm.info?x_type_flag=DELETE
Similar pages

Google Sets

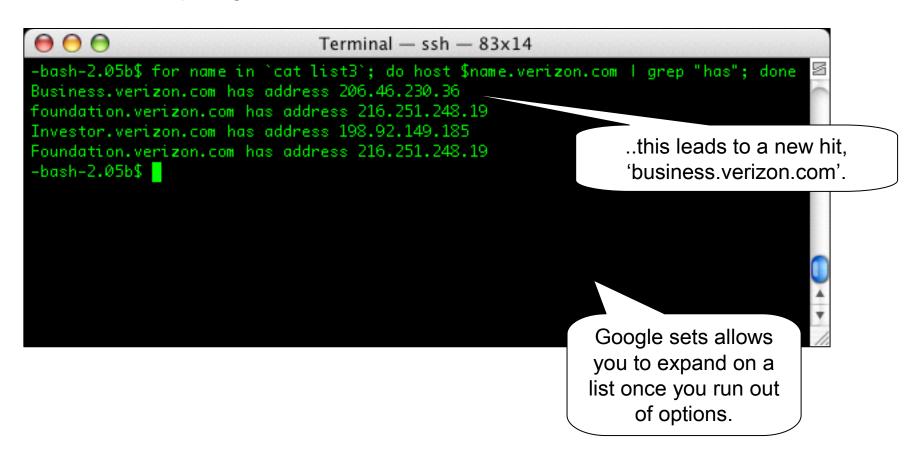
 Although this is a simple example, we can throw these two words into Google Sets....



Predicted Items
Investor
<u>foundation</u>
Second Foundation
Intermediate
Entrepreneur
Self Employed
Advanced
Roof
framing
Alliances
Careers
Community
<u>columns</u>
Completion
<u>Lender</u>
<u>Excavation</u>
BOARD MEMBERS
<u>walls</u>
<u>Application</u>
Earnings Reports
<u>IPOs</u>
<u>Financial</u>
<u>Certificate</u>
professional
Blusher
Research
<u>groundwork</u>
fundament
roof structure
Business Directory
floors
PF
Contact Us
Metal Roof
FOGA FOPA
Owner Occupied
STEM Methodology
<u>TEACHERS</u>
DISTRIBUTION

Expanding

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



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... =)

bhst03.verizon.com/

Similar pages

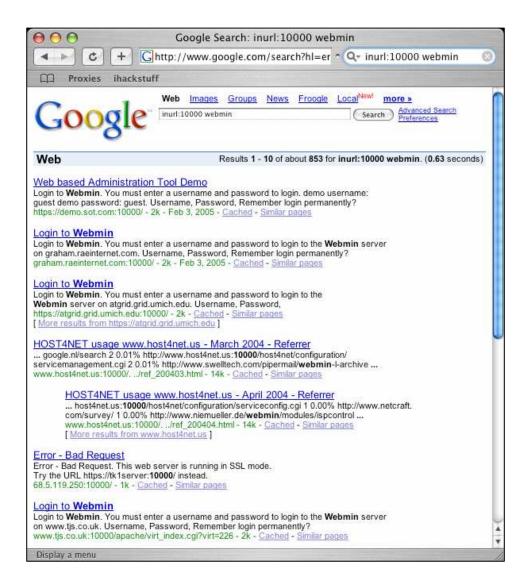
https://www33.verizon.com/wi-fi/login/locations/locations-remote.jsp Similar pages

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

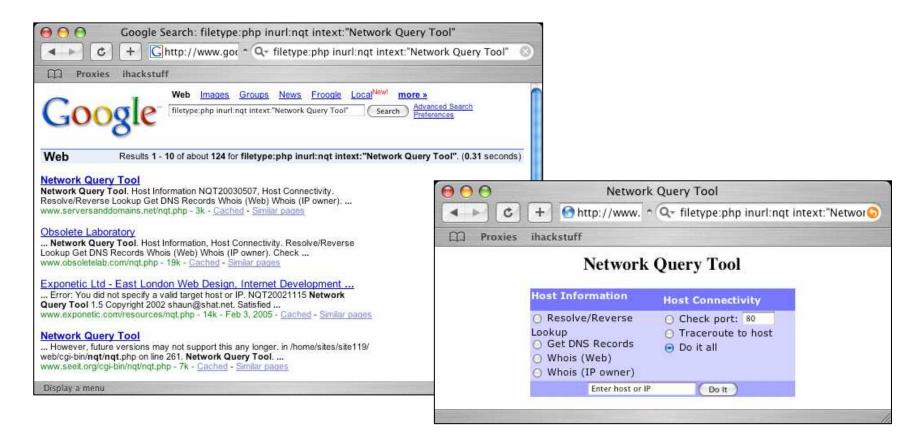
 Antoher way to go is to use a port number with inurl, combined with a negative intext search for that port number.

This search locates servers listening on port 8080.



Third party scanners

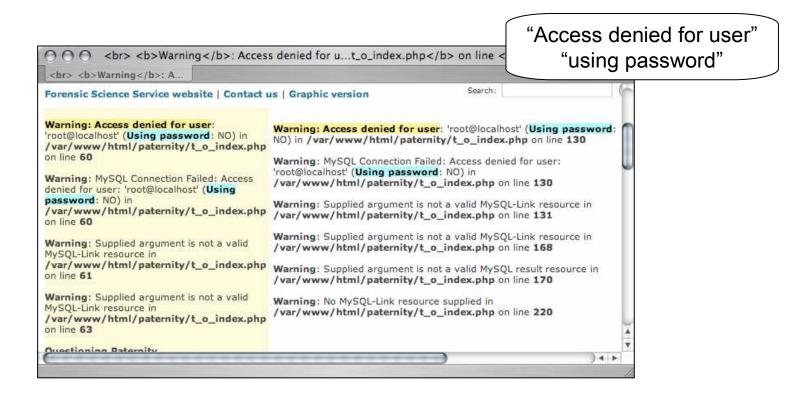
 When all else fails, Google for servers that can do your portscan for you!



Document Grinding and Database Digging

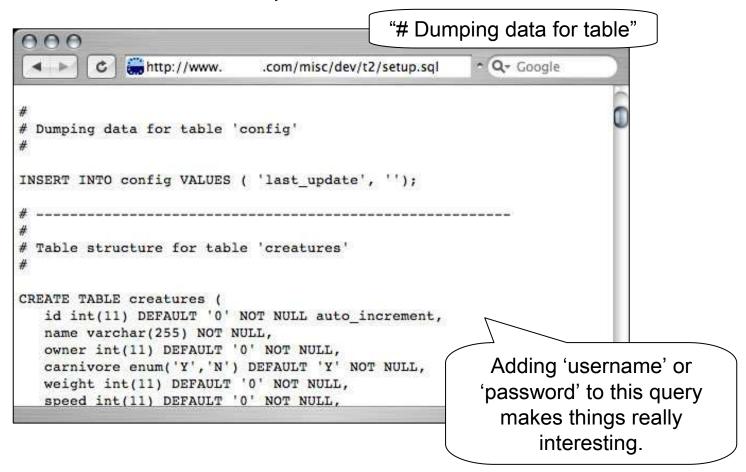
Documents and databases contain a wealth of information. Let's look at ways to foster abuse of SQL databases with Google.

SQL Usernames



SQL Schemas

Entire SQL Database dumps



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



ERROR: ORA-00933: SQL command not properly ended
:42000 -errorcodes:1

ERROR: ORA-00933: SQL command not properly ended
:42000 -errorcodes:1

ERROR: ORA-00933: SQL command not properly ended
:42000 -errorcodes:1

ERROR: ORA-00933: SQL command not properly ended
:42000 -errorcodes:1

ERROR: ORA-00933: SQL command not properly ended
:42000 -errorcodes:1

character string '>Mrs. Prisbrey"s Second Grade Class v'.

[Microsoft][ODBC SQL Server Driver][SQL Server]Unclosed quotation mark before the

"ORA-00933: SQL command not properly ended"

> "Unclosed quotation mark before the character string"

> > 4 1

SQL = "SELECT LinkBuilder_Enabled, FK_I INNER JOIN TB_Teacher ON TB_ClassPage WHERE PK_ClassPageID = 30365 AND Lin Class v"

Error Occurred While Proce...

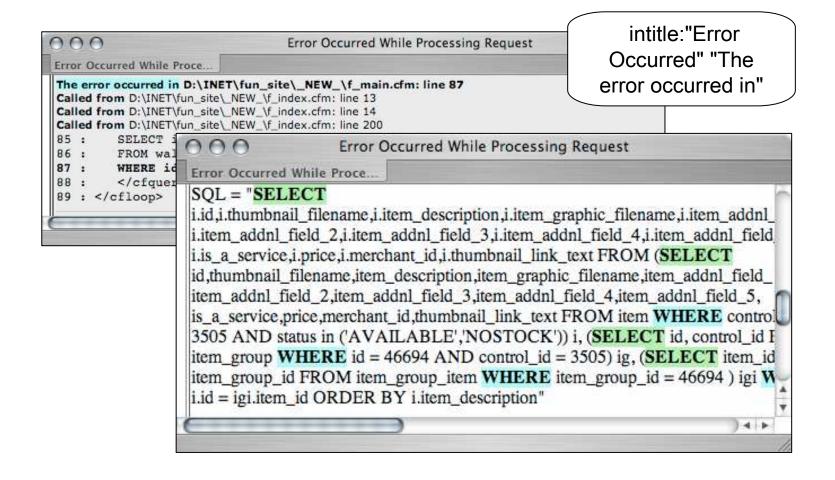
Error Detail: [MERANT][ODBC SQL Server Driver][SQL Server]Unclosed quotation mark before the character string 'Discover Platinum Card'.

- Our Apologie

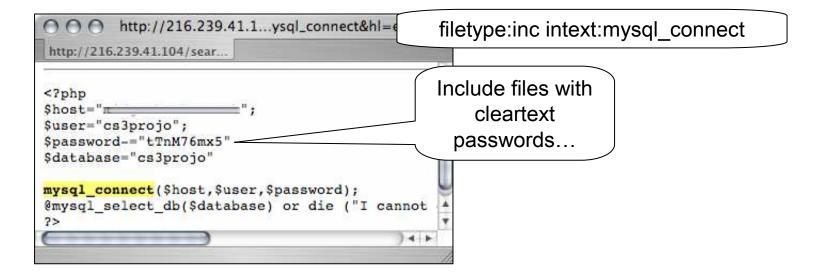
Our Apologies

SQL source

Getting lines of SQL source can aid an attacker.

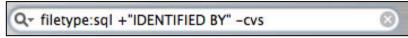


Going after SQL passwords



More SQL Passwords

- Question: What's the SQL syntax that can be used to set a passwords?
- (TWO WORDS)
- One Answer: "Identified by"





More SQL Passwords

The slightly more hardcore version...



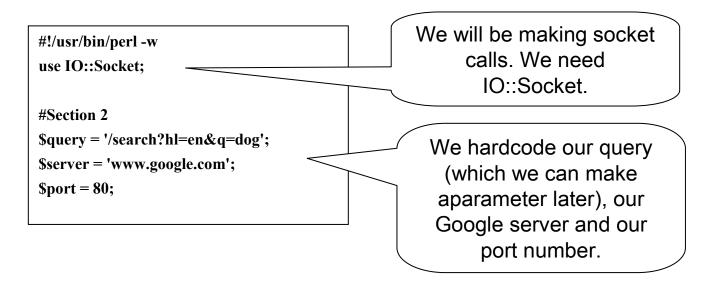
Various database detection queries

Query	Description			detection
	·			
inurl:nuke filetype:sql	php-nuke or postnuke CMS du	-		
filetype:sql password	SQL database dumps or batche commands	ed SQL		
filetype:sql "IDENTIFIED BY" –cvs	SQL database dumps or batche commands, focus on "IDENTIFI which can locate passwords			
"# Dumping data for table (username user users password)"	SQL database dumps or batche commands, focus on interestin			
"#mysql dump" filetype:sql	SQL database dumps			Database detection
"# Dumping data for table"	SQL database dumps			Dalabase delection
"# phpMyAdmin MySQL-Dump" filetype:txt	SQL database dumps created b phpMyAdmin	у		
"# phpMyAdmin MySQL-Dump" "INSERT INTO" -"the"	SQL database dumps created b phpMyAdmin (variation))+		
		Query		Description
		filetype:c password	fm "cfapplication nam I	e" ColdFusion source code
		filetype:n	ndb inurl:users.mdb	Microsoft Access user database
		inurl:ema	il filetype:mdb	Microsoft Access e-mail database
		inurl:backup filetype:mdb		Microsoft Access backup databases
		inurl:foru	ım filetype:mdb	Microsoft Access forum databases
		inurl:/db/	main.mdb	ASP-Nuke databases
		inurl:prof	files filetype:mdb	Microsoft Access user profile databases
		filetype:a MapPath(sp DBQ=" * Server. ("*.mdb")	Microsoft Access database connection string search
			dmin mdb	Microsoft Access administration databases

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."



Page Scraping with Perl

Page Scraping with Perl

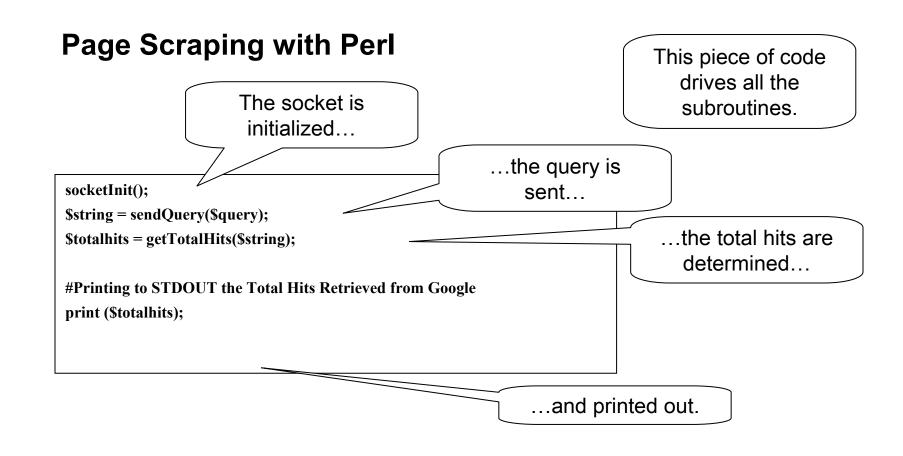
This subroutine sends the Google query (hardcoded above) and accepts one parameter, the Google query.

Google returned HTML is processed, and the line containing "of about" (our result line) is returned from this routine.

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

Page Scraping with Perl This subroutine takes one parameter (the results line from the sub getTotalHits(\$) Sendquery) my (Sourline) = (a);**\$hits=""**; "of about is located"... \$index = index(\$ourline, "of about"); \$str = substr(\$ourline, \$index, 30); ...the next 30 characters @buf=split(//,\$str); are grabbed... for (\$i = 0; \$i < 30; \$i++)if $(\$buf[\$i] = \sim /[0-9]/)$... all the digits are removed.... \$hits=\$hits.\$buf[\$i]; ...stored in \$hits... return \$hits; ...and returned.

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



CGI Scanning

/iisadmpwd/ /iisadmpwd/achg.htr /iisadmpwd/aexp.htr /iisadmpwd/aexp2.htr /iisadmpwd/aexp2b.htr Another automation example might involve chopping up a CGI scanner's vulnerability file...

inurl:/iisadmpwd/

inurl:/iisadmpwd/achg.htr inurl:/iisadmpwd/aexp.htr inurl:/iisadmpwd/aexp2.htr inurl:/iisadmpwd/aexp2b.htr ... converting the checks into Google queries, sending these queries to a Google scanner.

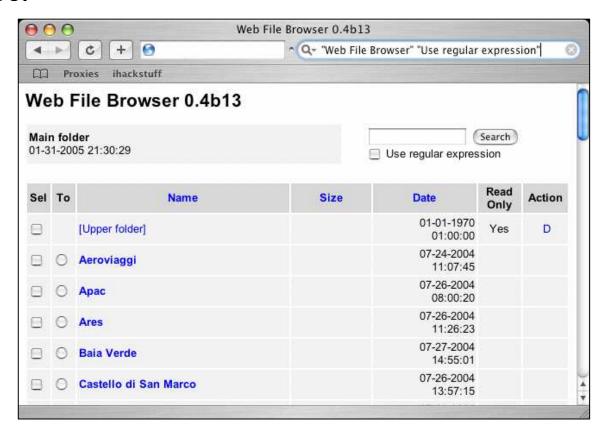
intitle:index.of /iisadmpwd/ intitle:index.of /iisadmpwd/achg.htr intitle:index.of /iisadmpwd/aexp.htr intitle:index.of /iisadmpwd/aexp2.htr intitle:index.of /iisadmpwd/aexp2b.htr

Web Servers, Login Portals, Network Hardware

Network devices can be soooo much fun to Google for...

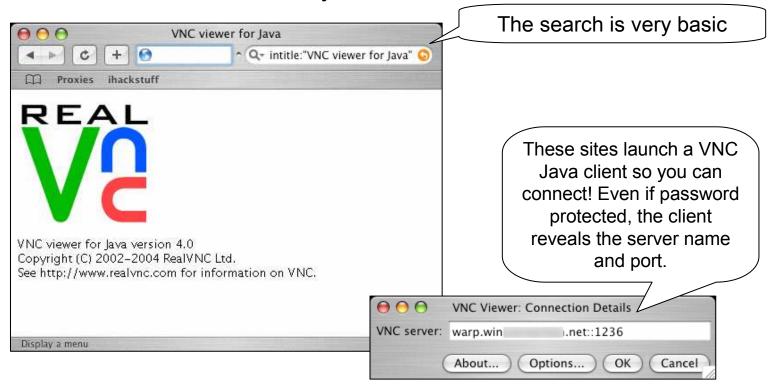
Web File Browser

 This program allows directory walking, file uploading, and more.



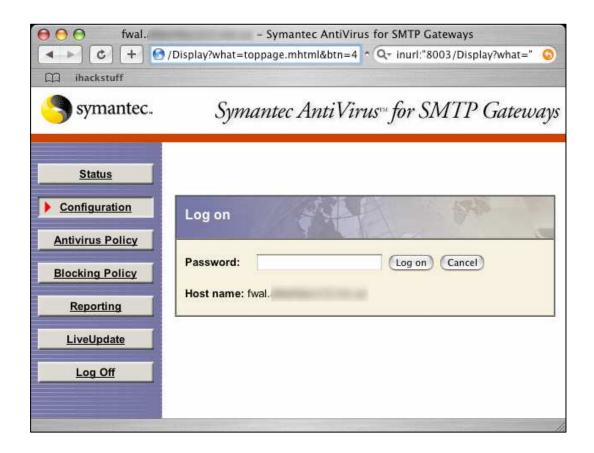
VNC Servers (with client)

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



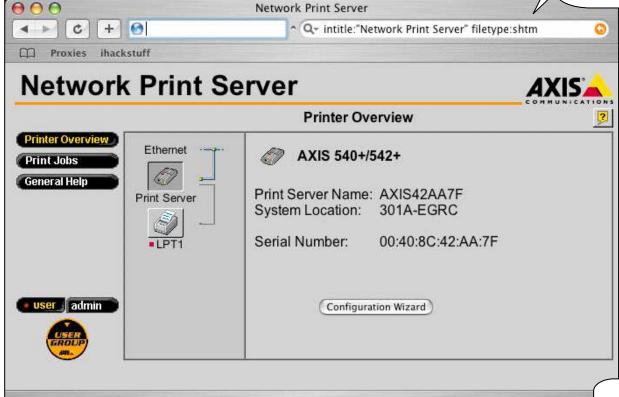
Thanks to lester for this one!

Symantec Anti-Virus SMTP Gateways



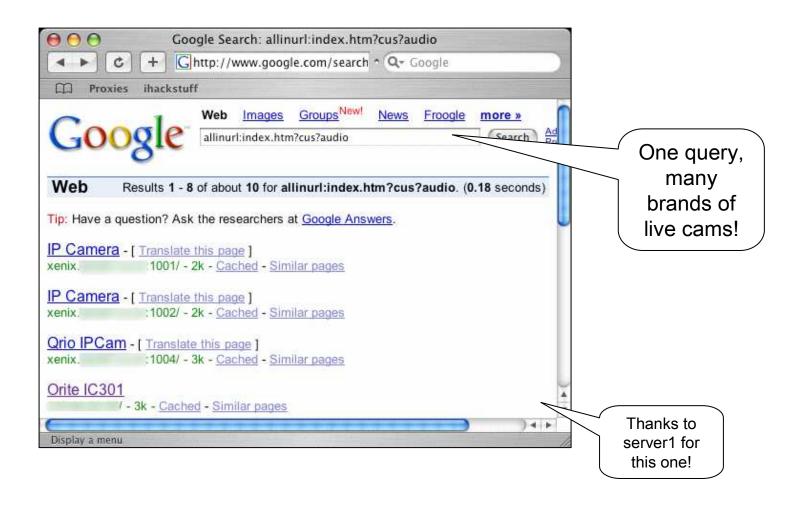
Axis Print Servers

Print server administration, Google-style!

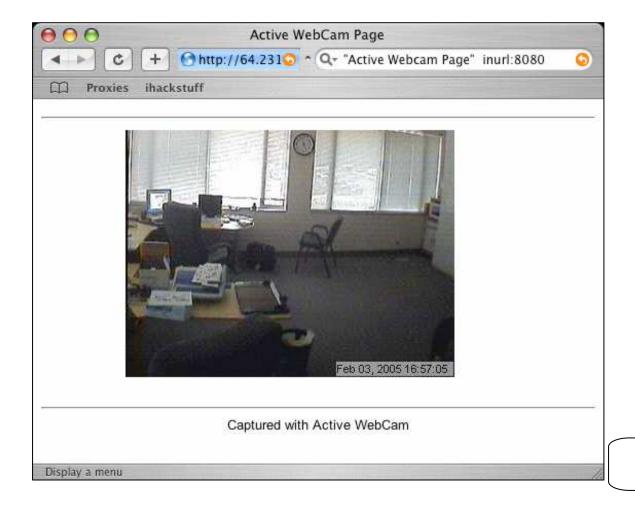


Thanks to murfie for this one!

Xenix, Sweex, Orite Web Cams



Active WebCam



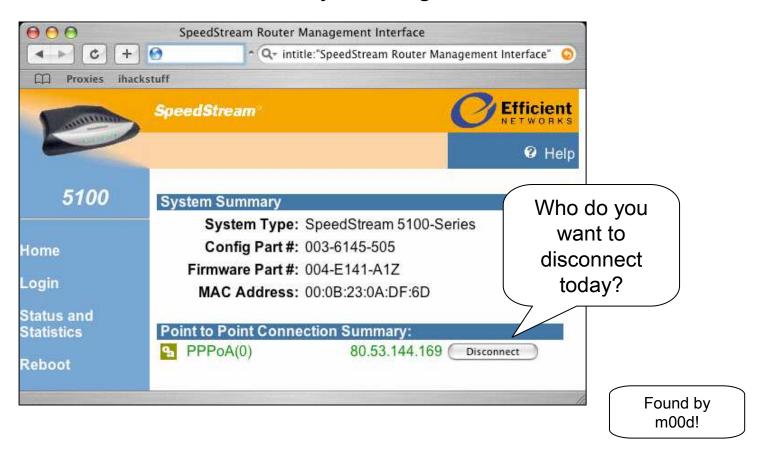
Thanks klouw!

Toshiba Network Cameras



Speedstream DSL Routers

Home broadband connectivity... Googled.



Belkin Routers

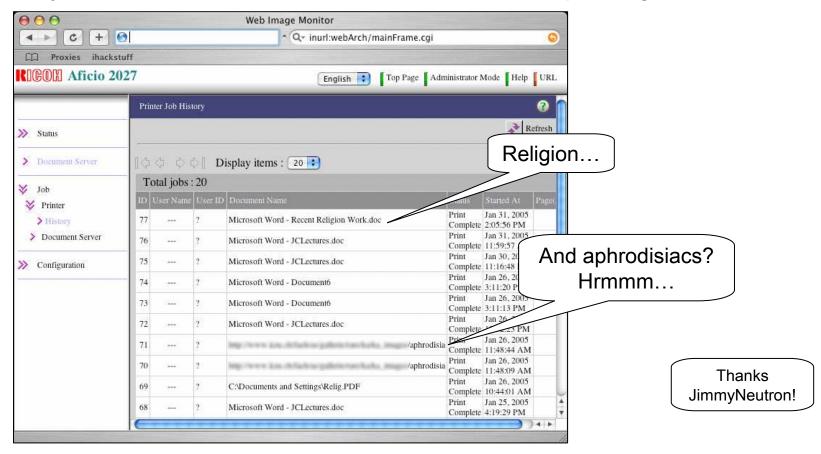
 Belkin routers have become a household name in connected households. The management interface shouldn't show up on Google... but it does.



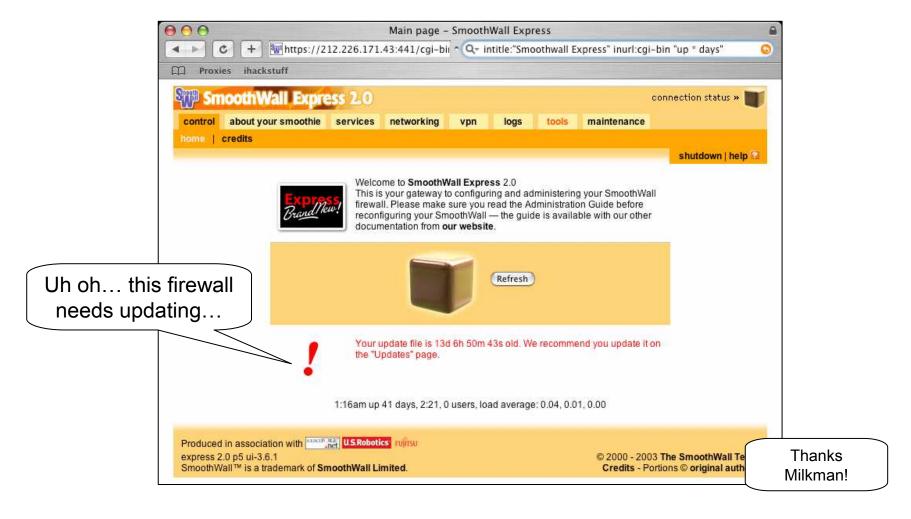
Thanks to darksun for this one!

Printers

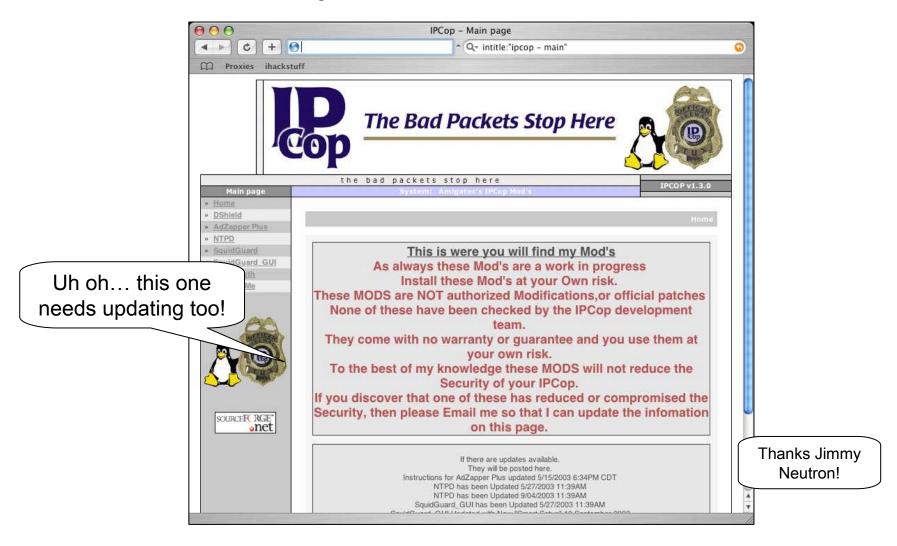
 Trolling printers through Google can be fun, especially when you can see and download what others are printing...



Firewalls - Smoothwall

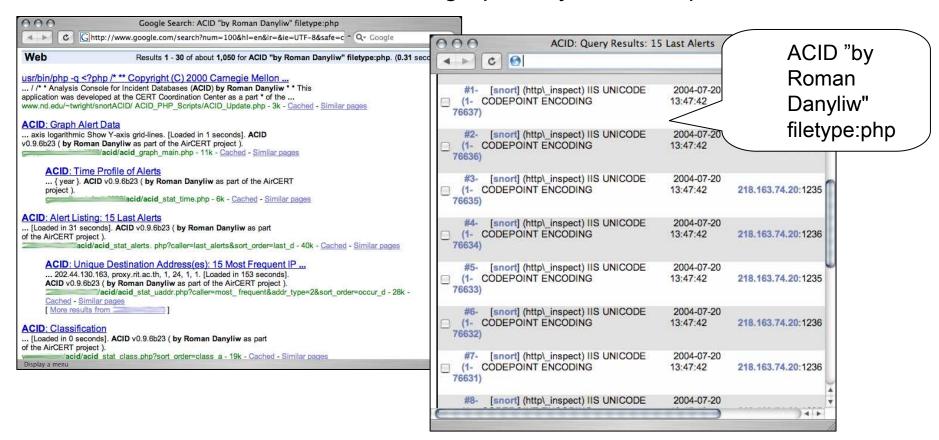


Firewalls - IPCop

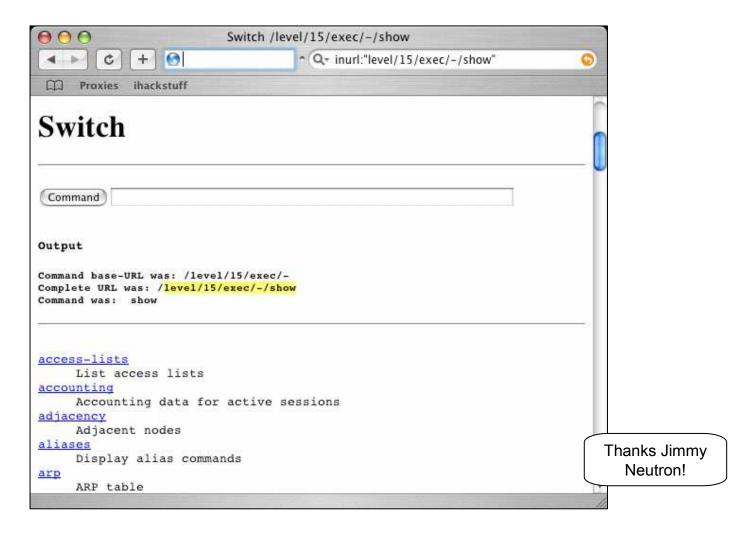


IDS Data: ACID

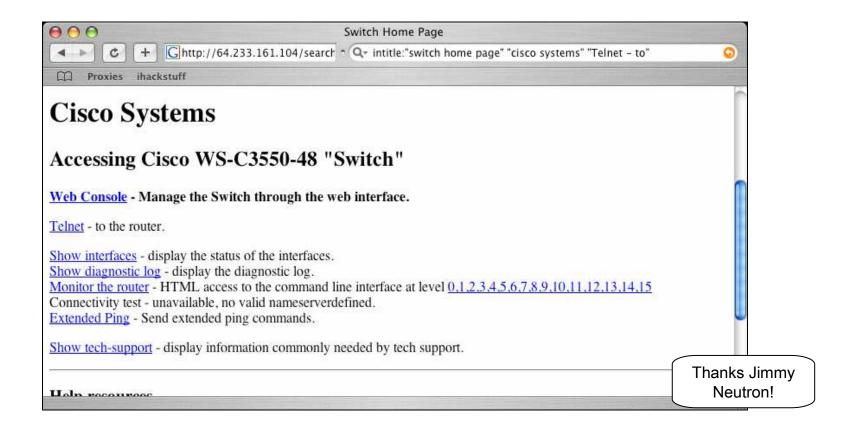
SNORT IDS data delivered graphically, served up fresh



Open Cisco Devices

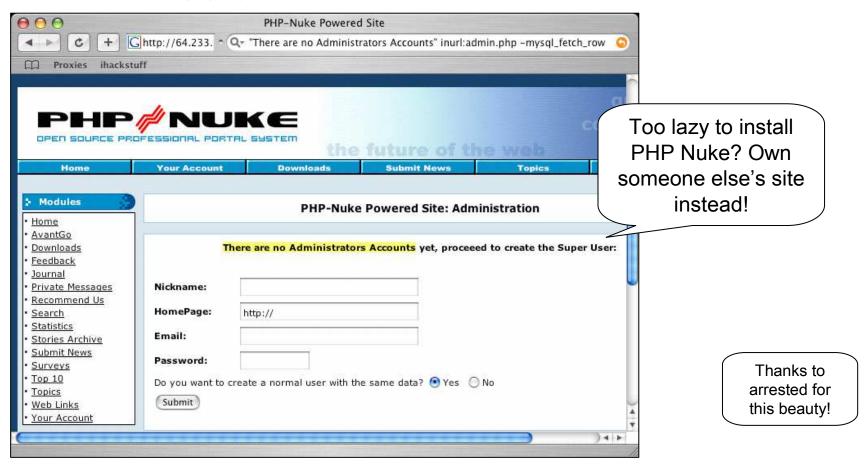


Cisco Switches

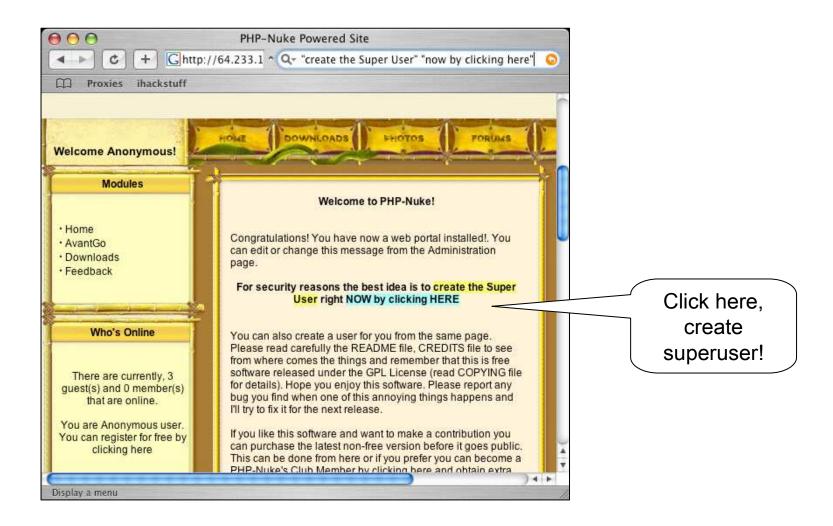


Wide Open PHP Nuke Sites

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



Open PHP Nuke... another way...



Security Cameras

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

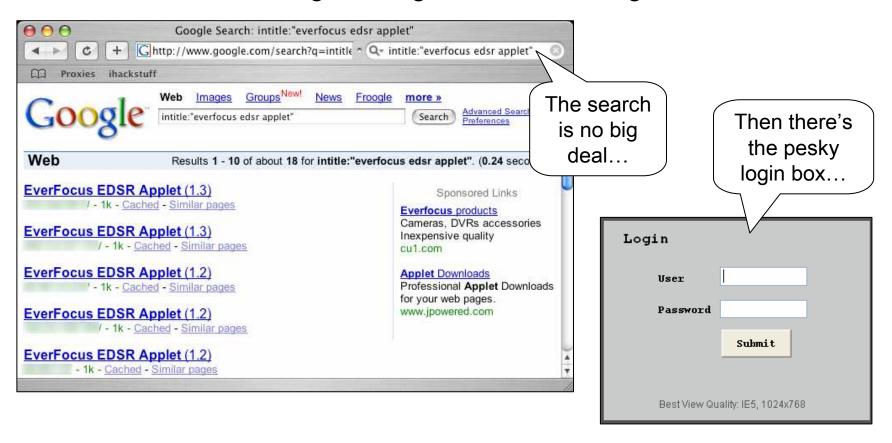


Security Cameras

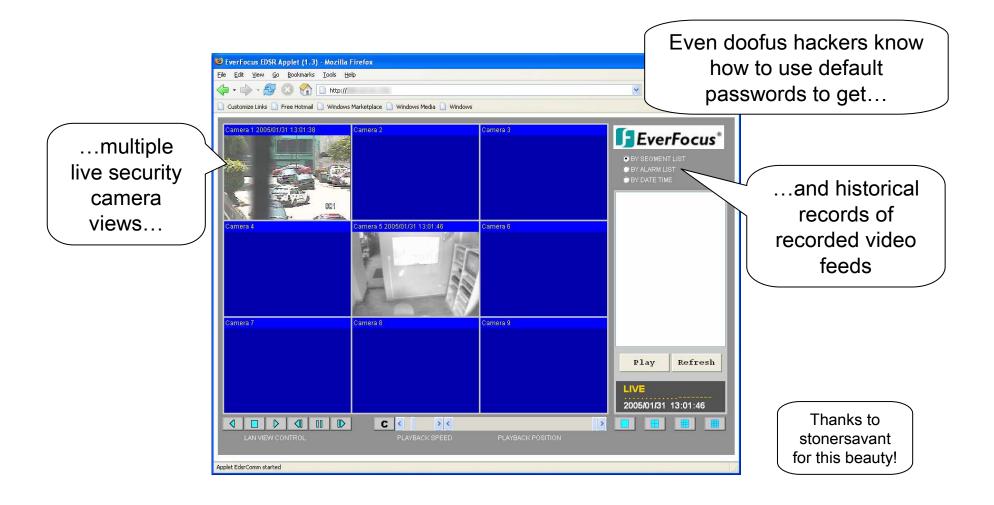


Time-lapse video recorders

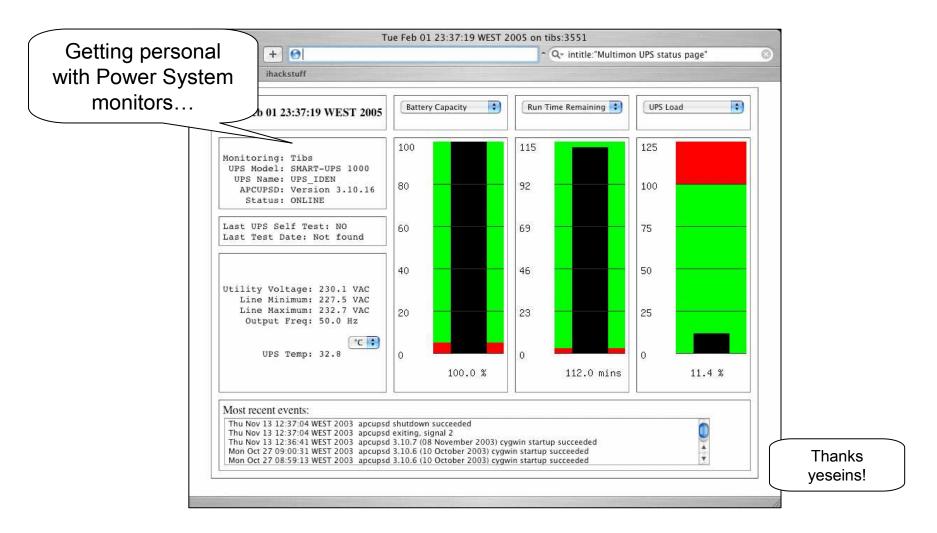
 A staple of any decent security system, these camera control units have gotten high-tech.. And Googlable...



Time lapse video recorders



UPS Monitors



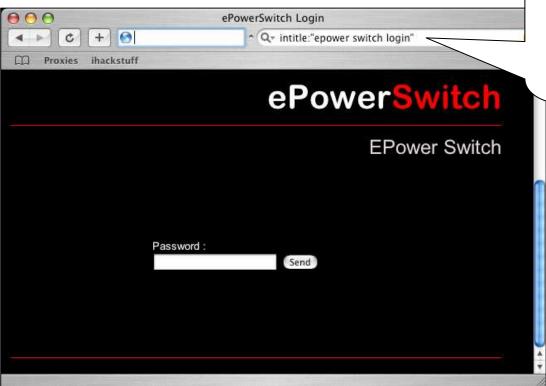
UPS Monitors

Oh wait.. Wrong kind of UPS...this is package tracking hacking... =P



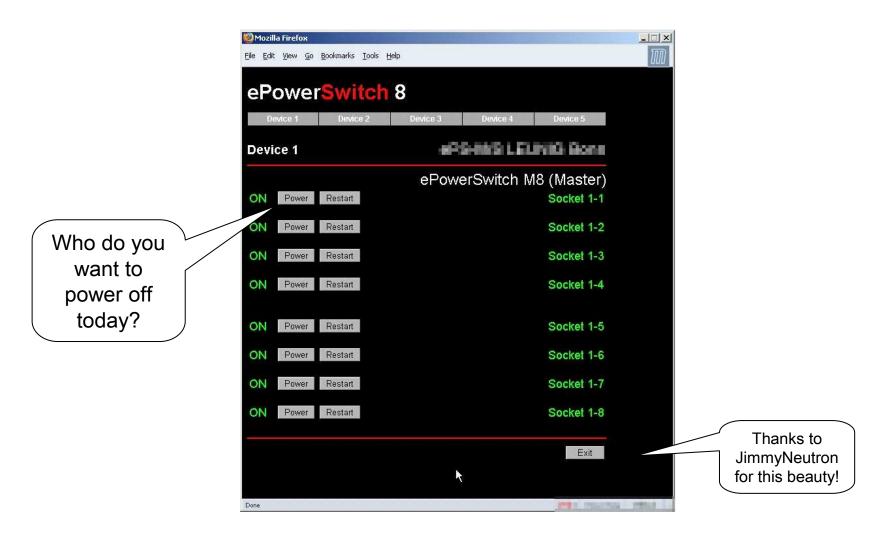
Hacking POWER Systems!

 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!



Google Phreaking

Question... Which is easier to hack with a web browser?

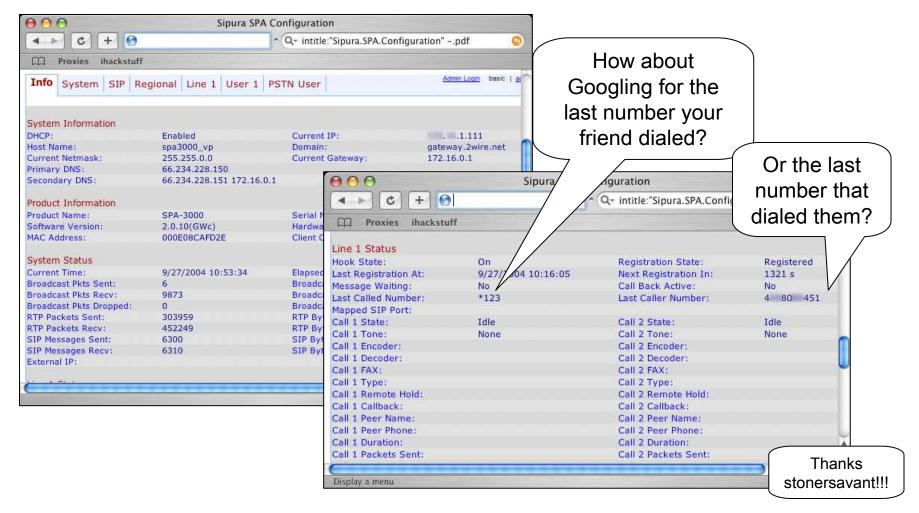
A: Sipura SPA 2000 IP Telephone

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

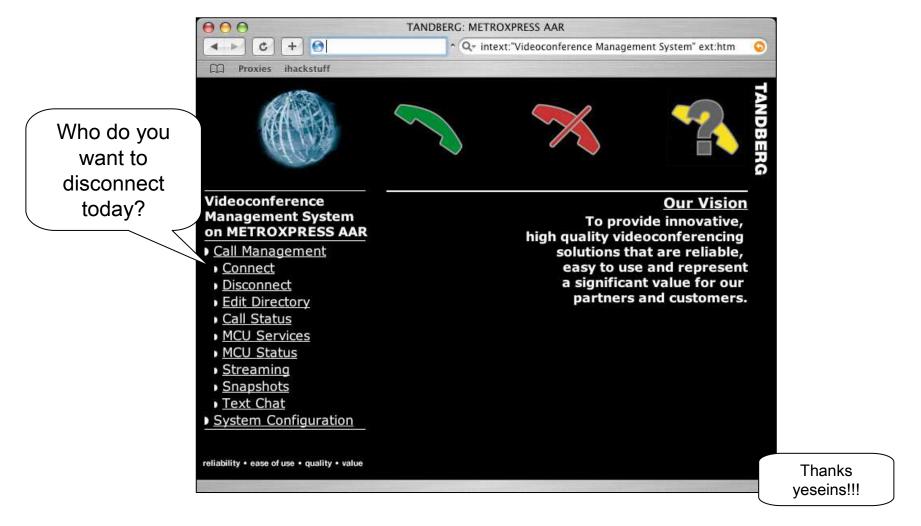
B: Vintage 1970's Rotary Phone

> QuickTime™ and a TIFF (Uncompressed) decompressor are needed to see this picture.

Sipura SPA IP Telephone

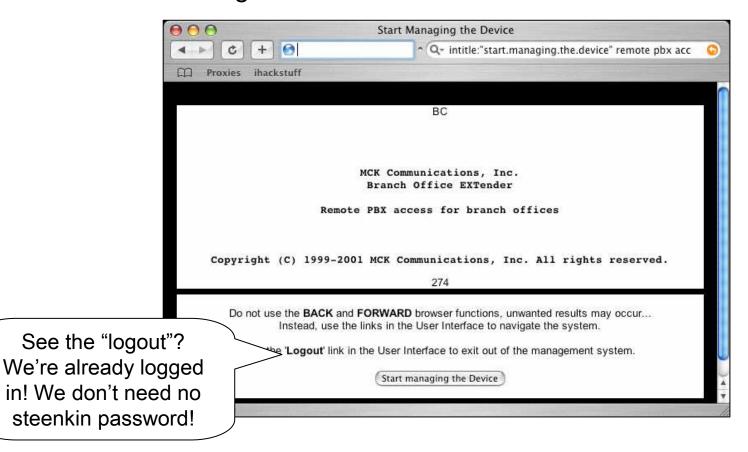


Videoconferencing

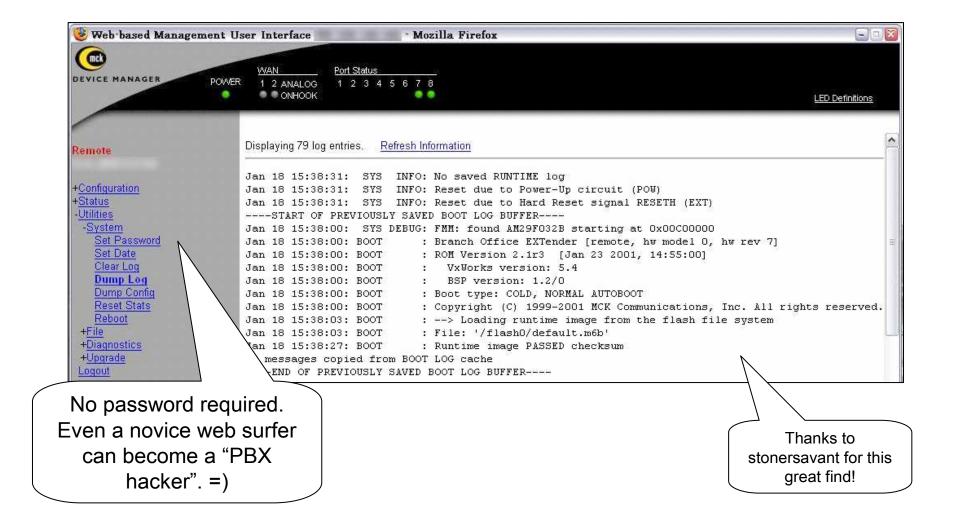


PBX Systems

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



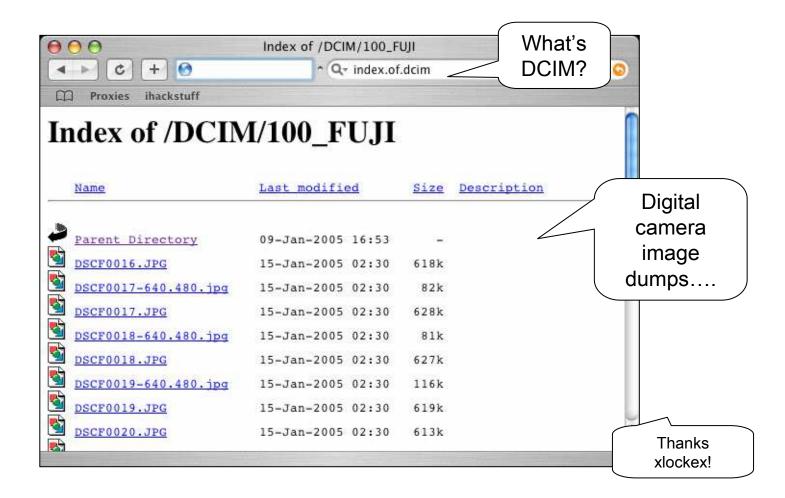
PBX Systems



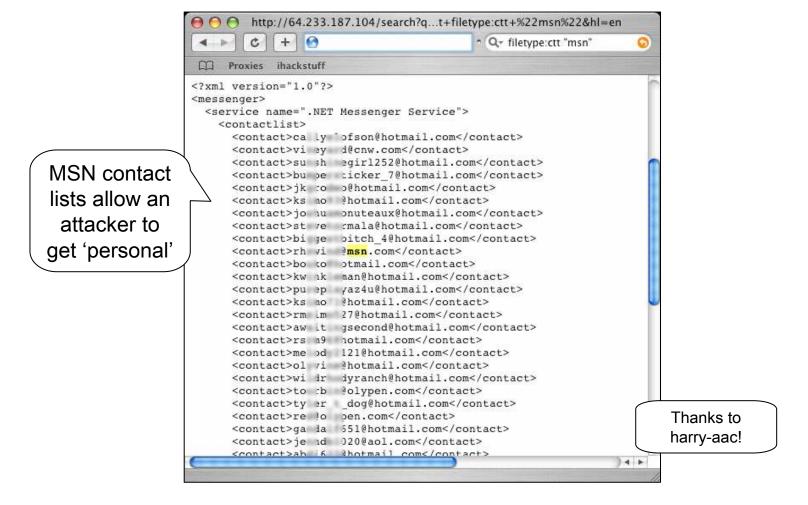
Usernames, Passwords and Secret Stuff, oh my!

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...

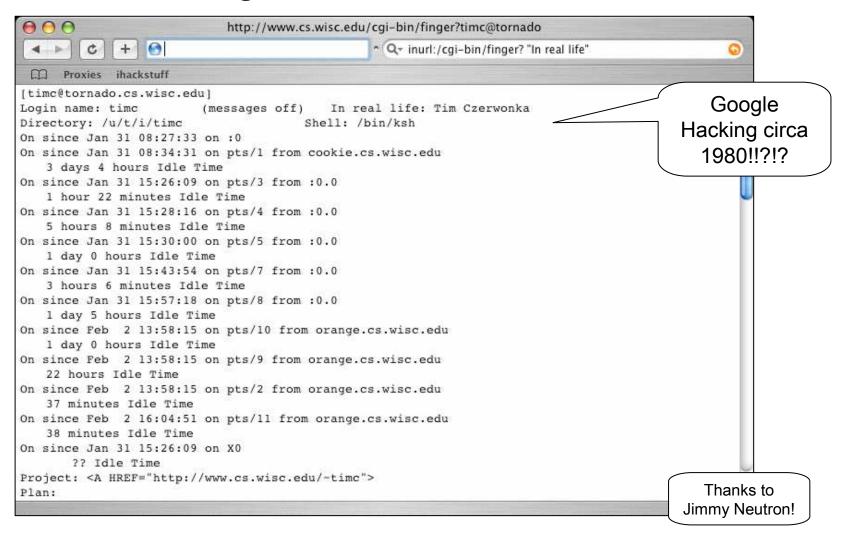
DCIM



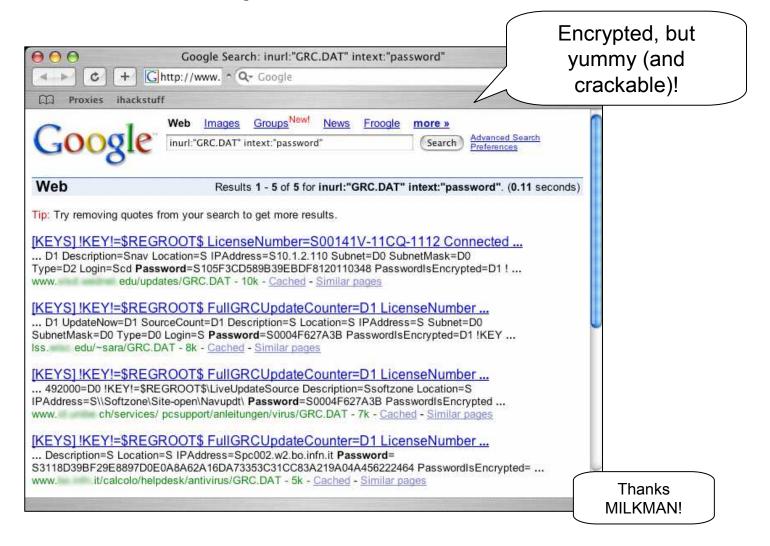
MSN Contact Lists



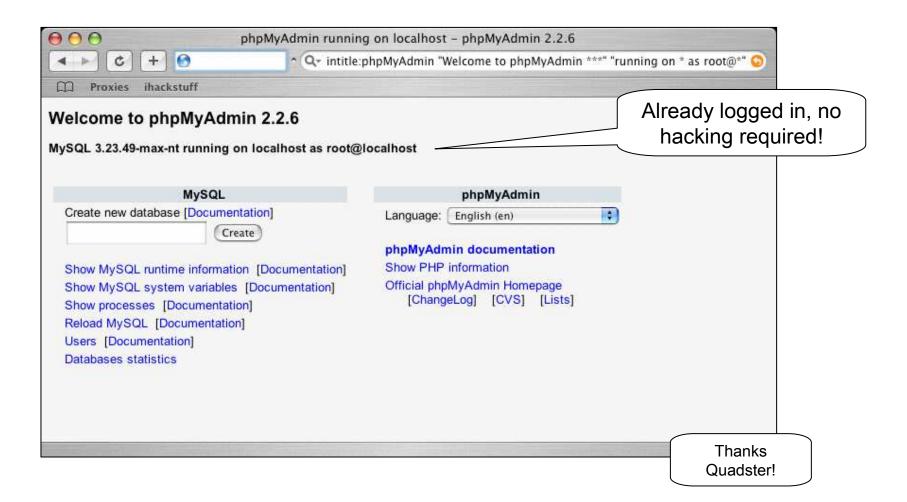
Old School! Finger...



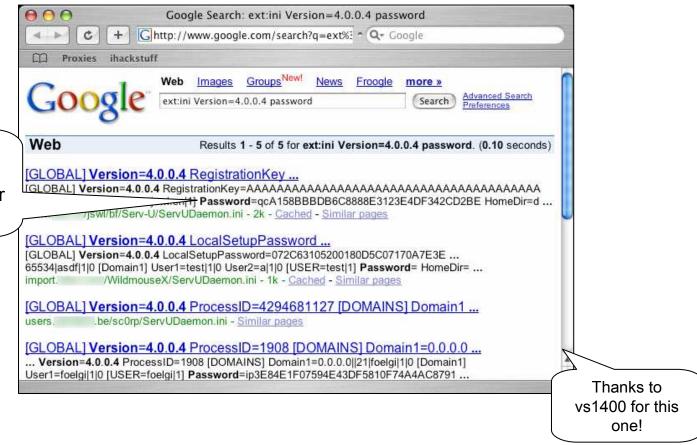
Norton AntiVirus Corporate Passwords



Open SQL servers

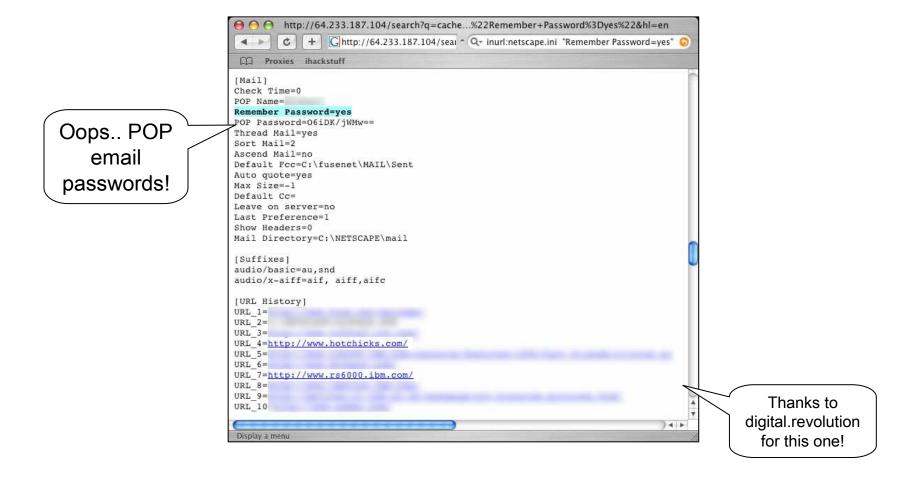


ServU FTP Passwords

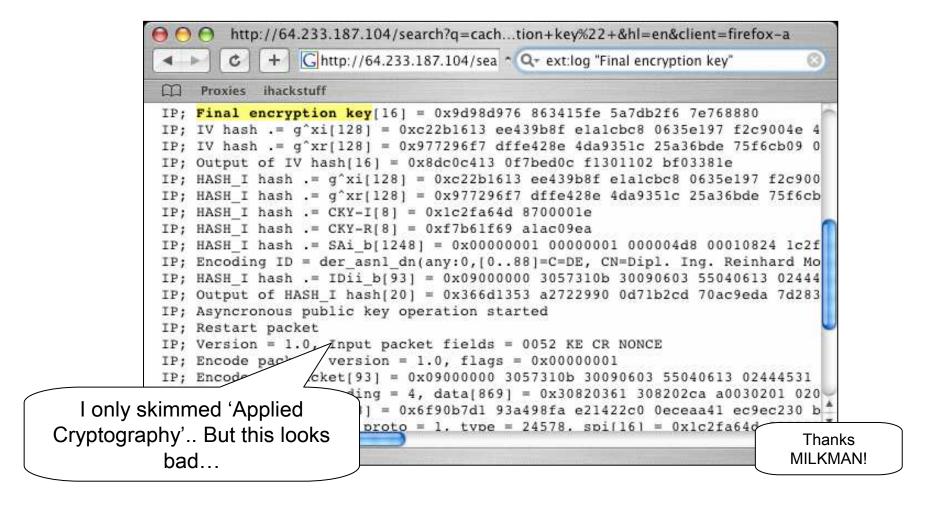


ServU FTP
Daemon
passwords, super
encrypto! =P

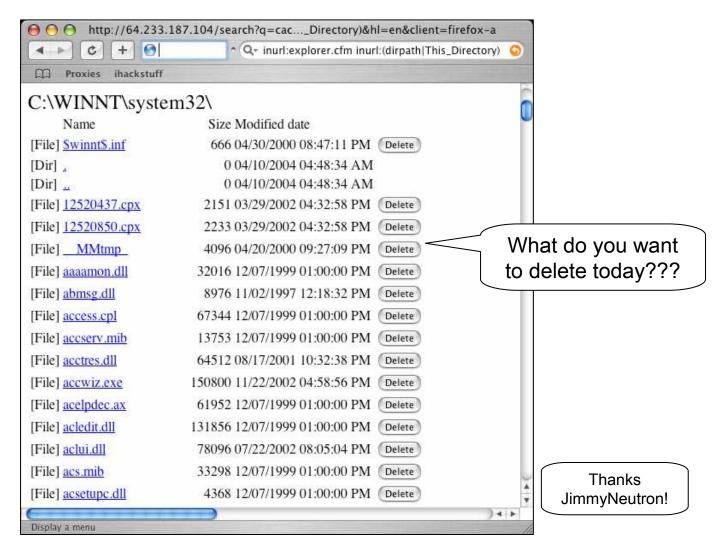
Netscape History Files



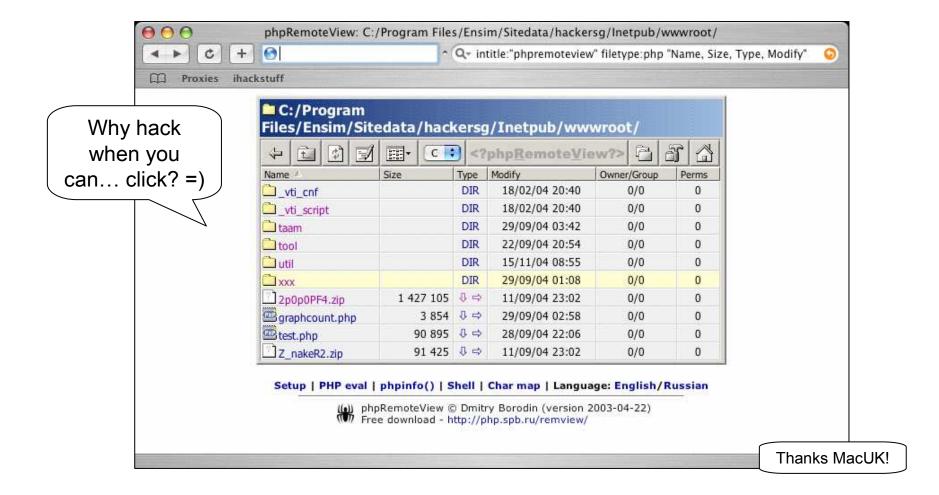
IPSec Final Encryption Keys



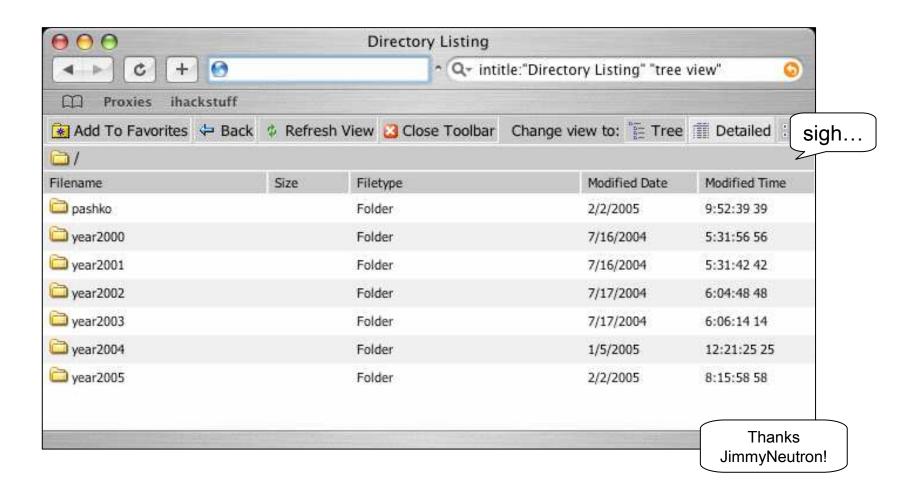
Explorer. EXPLORER!?!?!



More Explorers?!?!



More Explorers?!?!

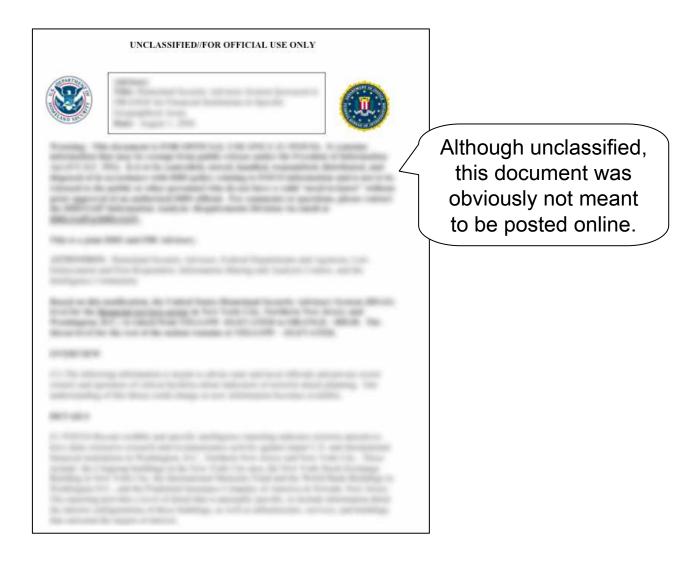


Sensitive Government Documents

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



FOUO Documents



FOUO Documents



FOUO "Prevention Guides", like this 19 page beauty, can give bad guys horrible ideas.

Locked out!

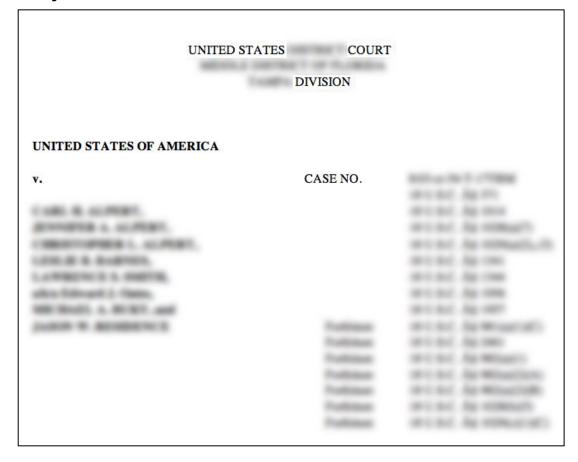
Some sites lock down sensitive data...



Credit card info on the web?

• How can this happen? Let's take a tour of some of the possibilities...

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



	Platinum
account number:	
account #-	
	account
#:	
	account #:
#:	account
	statement account #
Bank	#

How much detail is too much detail? =)

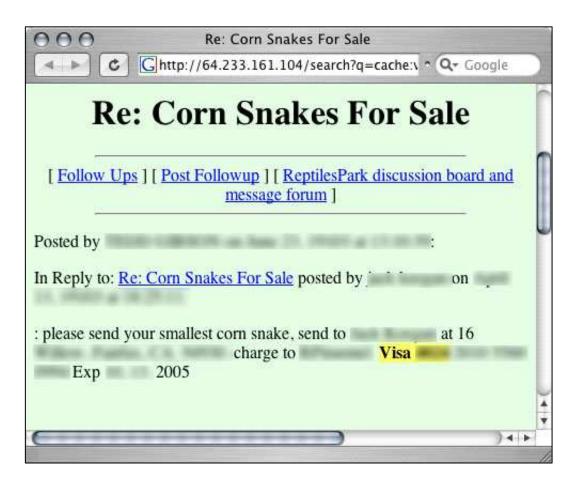
	CREDIT CARD					
			COUNT	ACCOUNT NAME	AMOU	NT
		1382	23.		\$.00
	Bank)59(13		S:	.00
	Bank (9040	97		\$.00
	t Bank)59(13		\$3	.00
	Bank	1004	96.		\$1	.00
	Bank	1004	96		S	.00
	Bank)59(13		\$	1.00
	r Bank	100	64.		\$	1.00
	Bank	7100	17		S	1.00
	Bank	100	164		s	0.00
	Bank	081	119		\$	1.00

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

	BANK		COUNT	ACCOUNT NAME	
					AMOUNT
	t Bank	054	13		\$1
		131	23		\$3
	Bank	90-	97		\$7
	Bank	711	02		\$2
	22 Bank	100	96		\$3
	Bank	282	27		\$11
	Bank	63	91		\$2
		90:	10		\$3
	of the Allegaria	and the	. boson	(98 apr)	

A tale of a corn snake

Is this for real? Either way it's pretty sad...



Getting shell.. the easy way

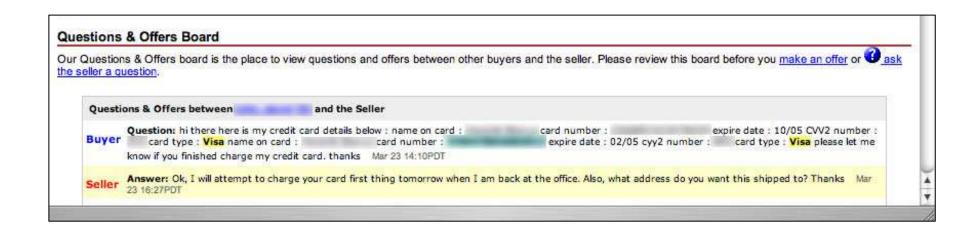
 Now I've heard the term 'using your credit card online' but this is ridiculous!

```
applying for a shell acct
  . To: haing talk @ haing org

    Subject: approving for a sheet acct

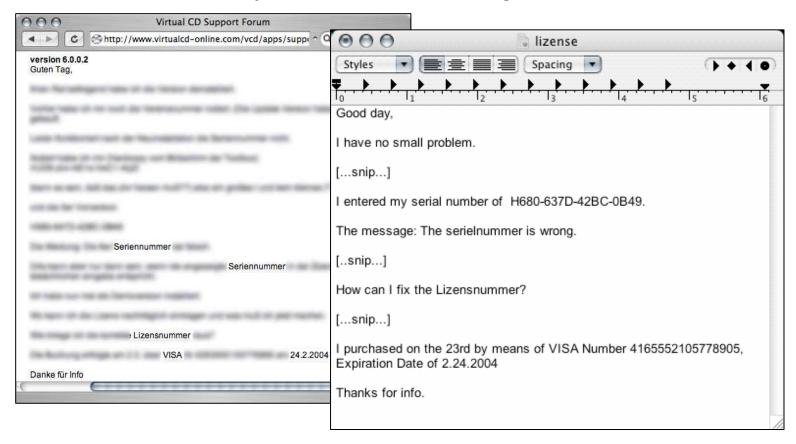
  • From: pwer possicolar competition bottom comp
  . Date: Sal (9) (3ct 1999) 14 19 17 (1MT
hi there i am interested to buy a shell acct for my son.
i m 2 busy because of some reasons. This is my only way to communicate
with my son... so i want 2 buy a shell acct using my creditcard
the login name and password will be:
login name:
password: 4 104 10
Down here will be the details about me.
name: Lorvain Alden
1865 Sountain View Stead
Fortuna, CA, 95540
tel#: "% "% "
Billing Information:
Name on card:
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....



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.

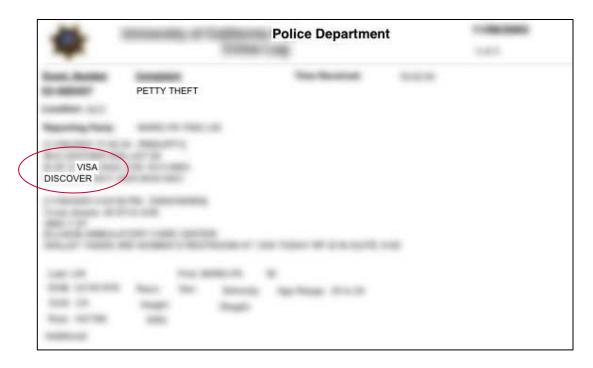


Police Crime reports

- 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.

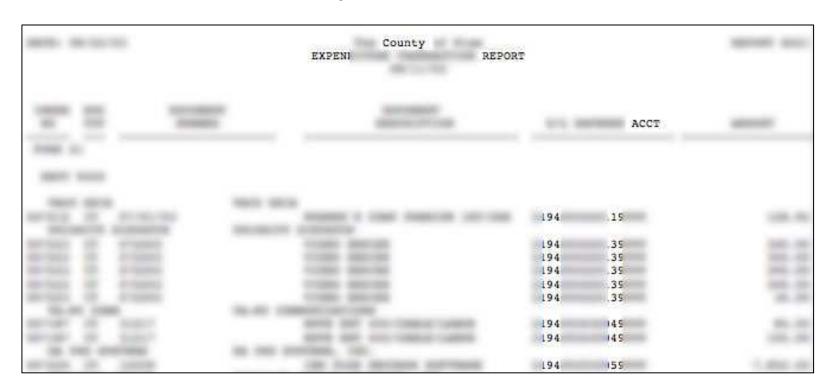
Crime shouldn't pay...

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



A few might fall through the cracks....

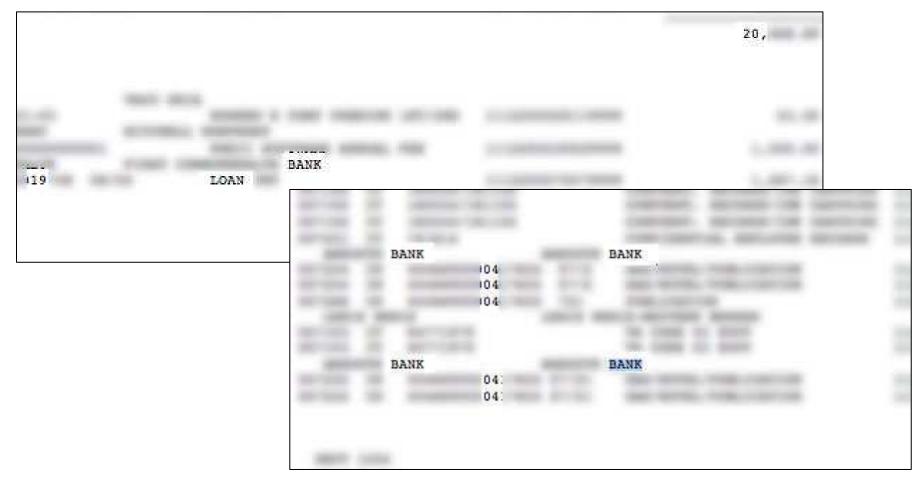
• It's not uncommon for expense reports to be generated. This one is for a county.



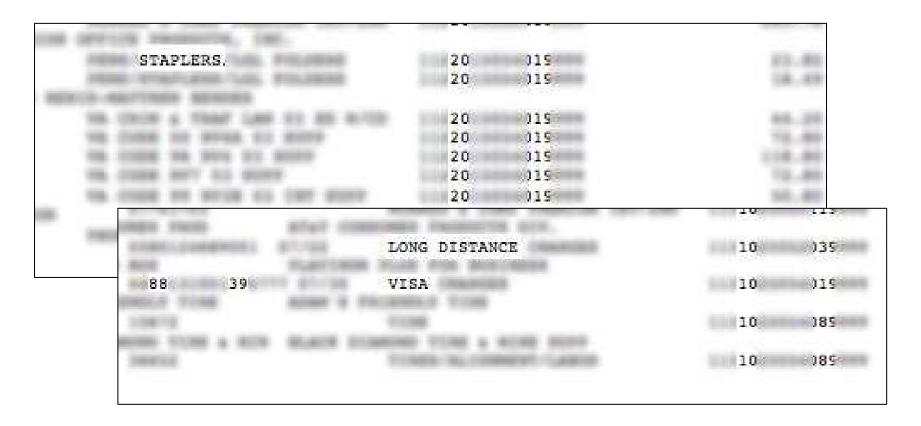
• Bank account numbers....

```
01
                                                          01
                                                          01
                       CONFDEN
                                                          01
                                                          01
BANK
                                                         11
                                                         11
    69
         41
                                                         11
                                                         118
                                                         11
                                                         06
     69
           76!
     69
           76!
                                                         06
```

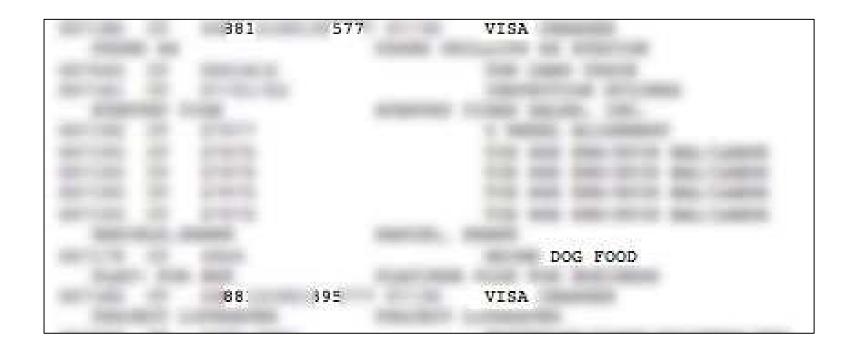
• Bank loan information... \$20,000 + transactions



• Oh boy...



Somebody has to pay for all this stuff....

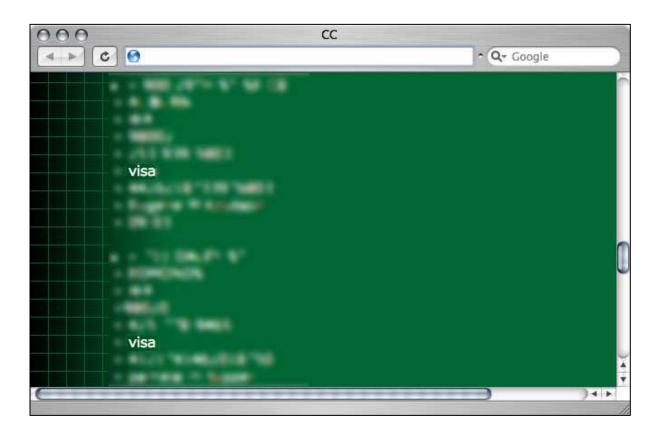


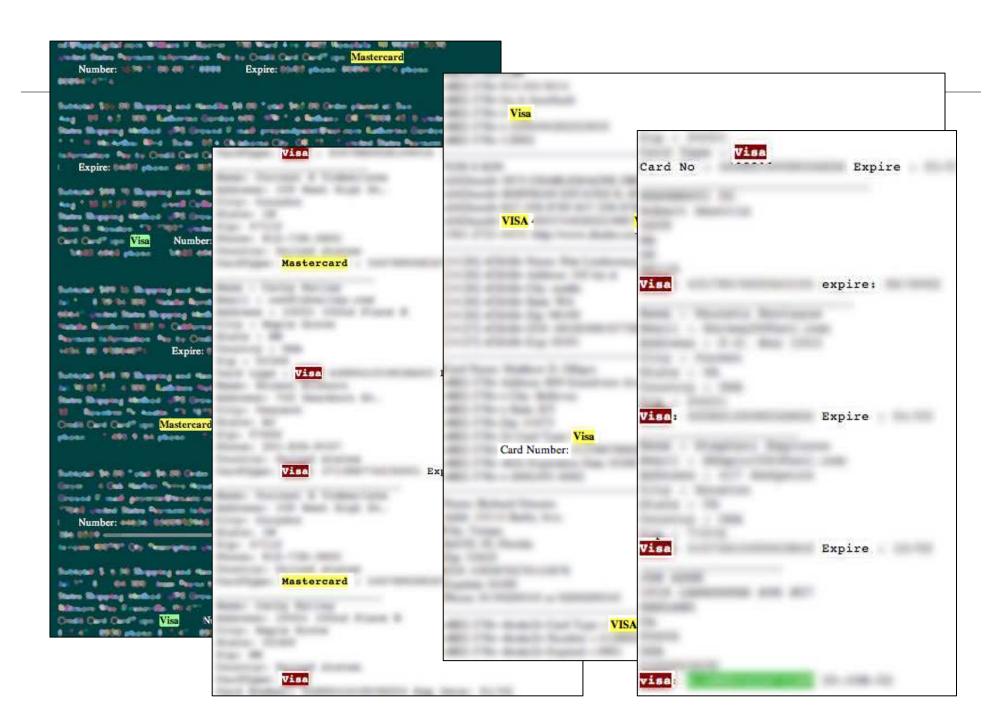
• That's one heck of a video series.... \$300+

VIDEO SERIES VIDEO SERIES	94005 39 94005 39	3 3 3
VIDEO SERIES	9400! 39	3
VIDEO SERIES	9400: 39	3
VIDEO SERIES	9400: 39	~~~
	49	
	49	

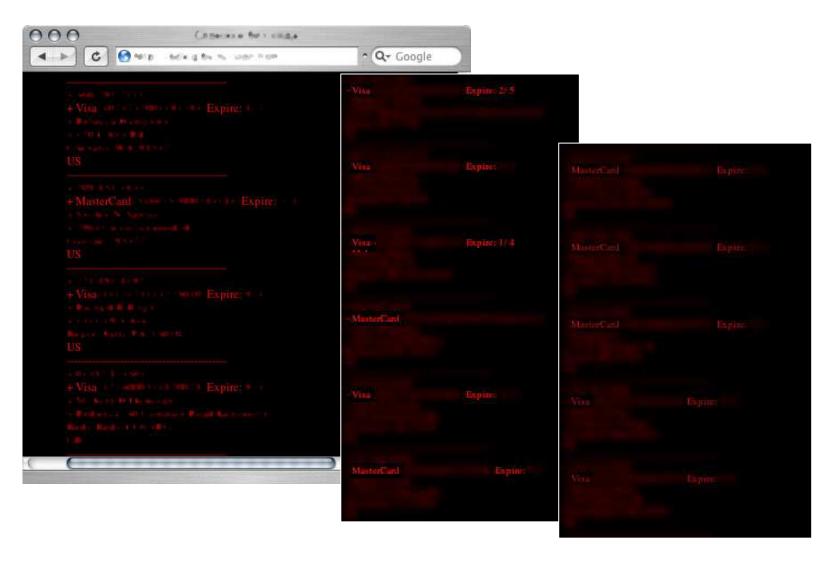
Credit cards... Google hacker's gold...

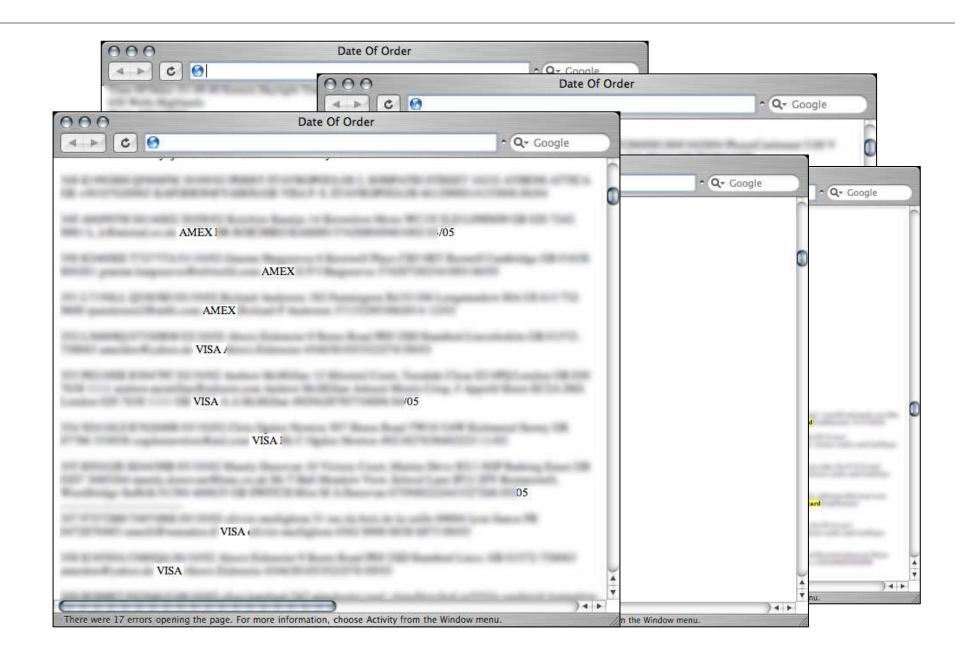
- The legend of finding credit cards online is true...
- I just get bored sifting through them all....

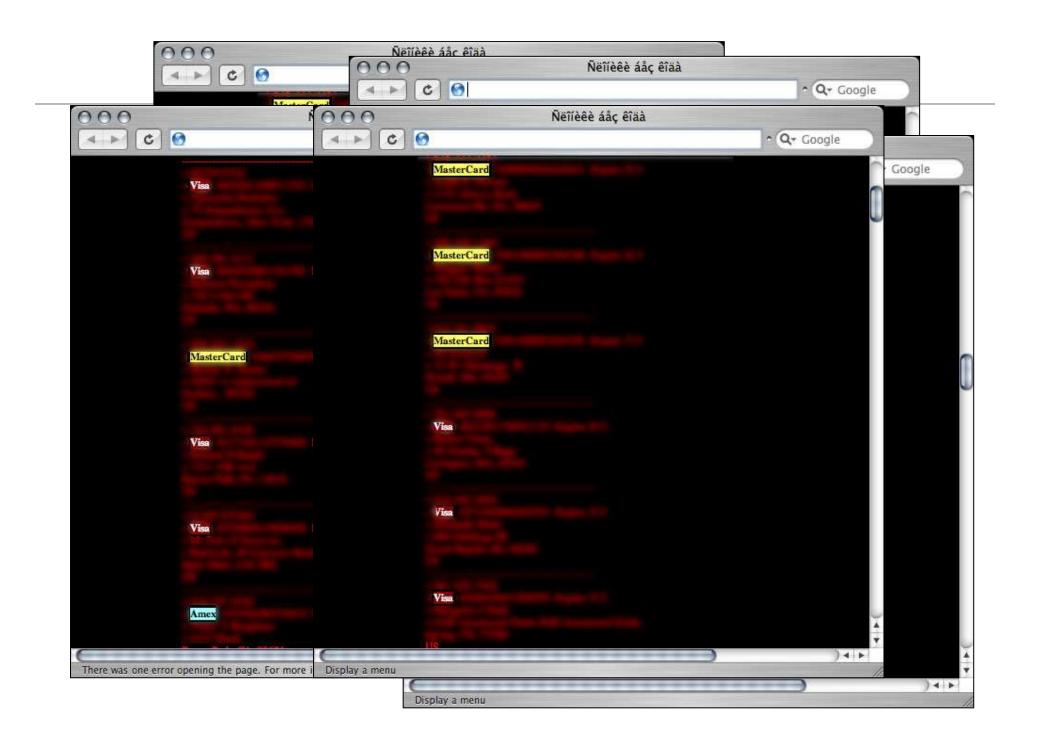


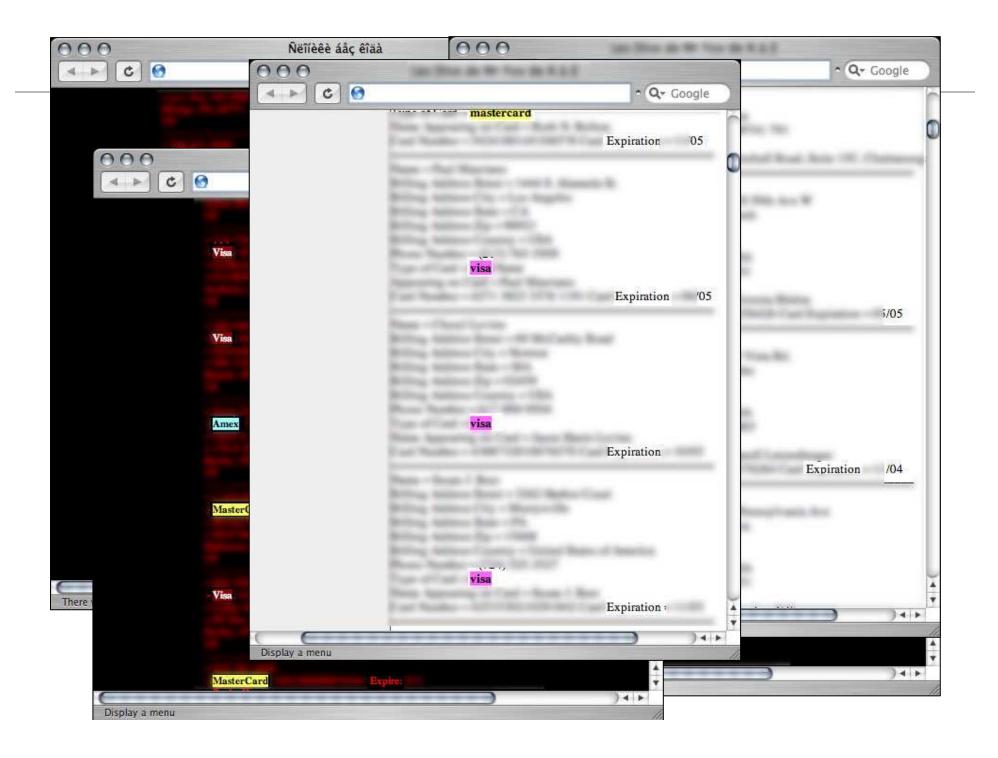


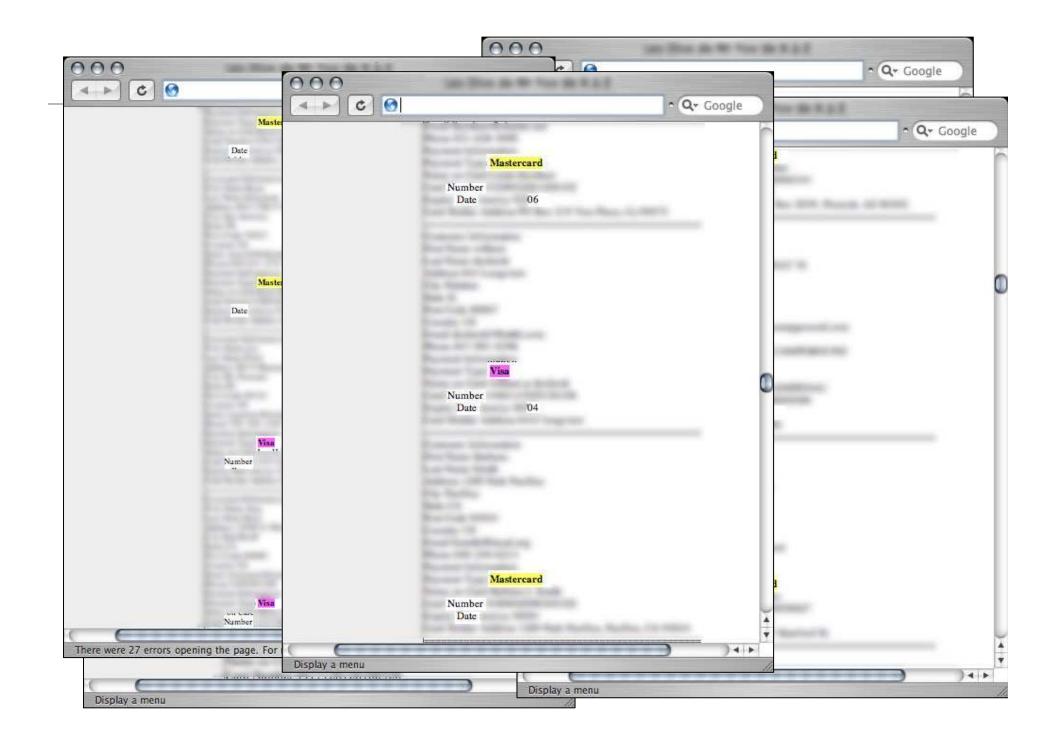
Credit Listings



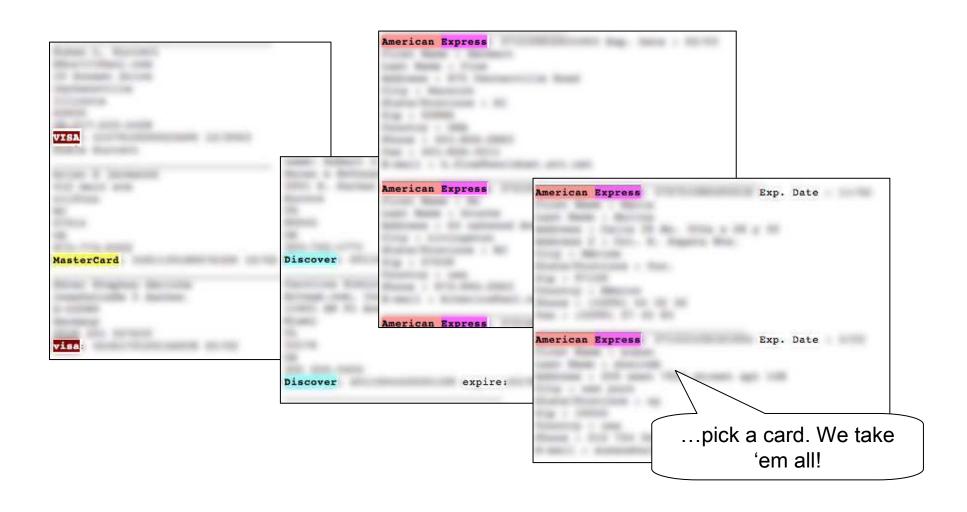








Pick a card any card...

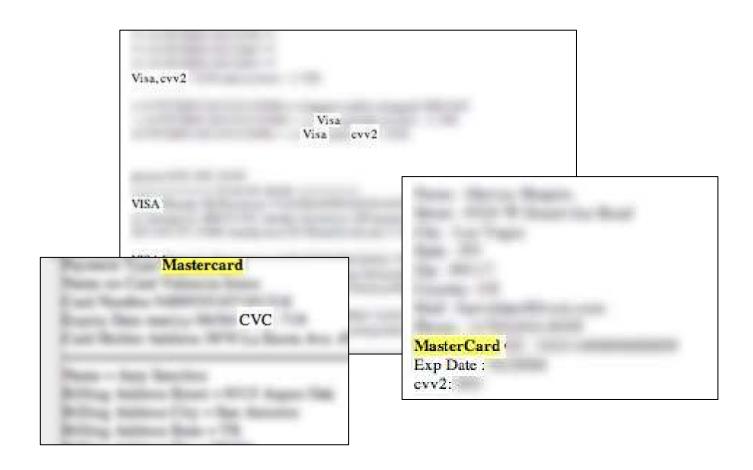


Credit Validation

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

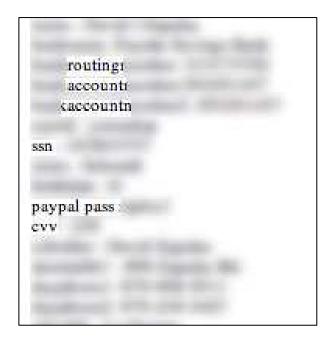
- Answer: That little code on the back of the card.
- Bonus question: What's that code called?
- Answer: A "CVV" code.

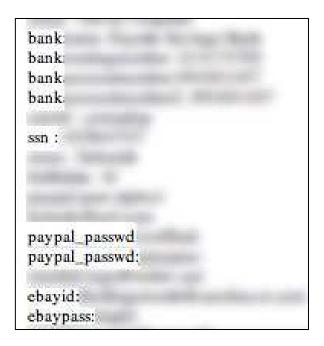
Credit Card Numbers, Expiration Date and CVV numbers, oh my!

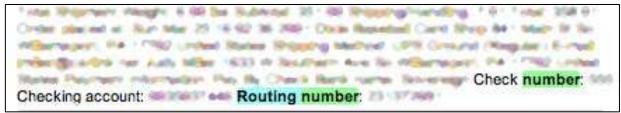


That's not all....

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.

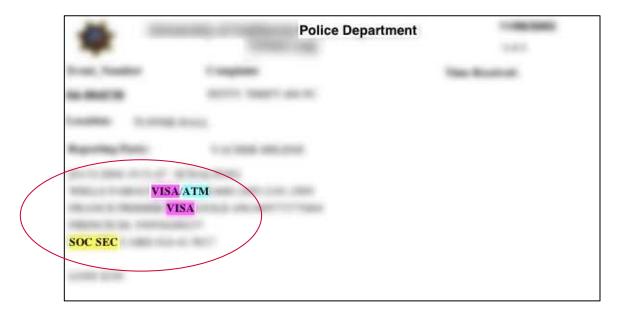
SSN's in source code

 Well, they could be hardcoded into a healthcare system... and uhmmm... put on the web...

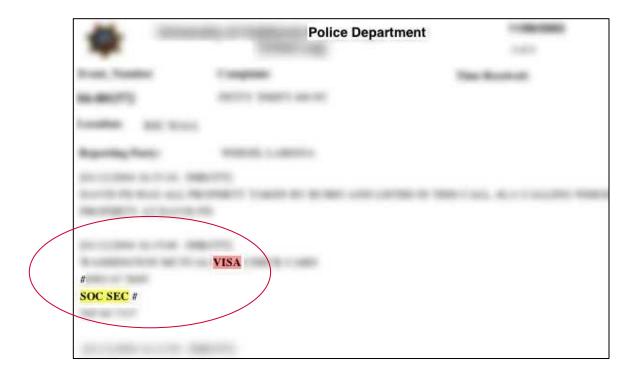
```
000
                         - Software Development Inventory
           Ghttp://64.233.161.104/search?q=cache:LYt9FTWEQXgJ:www.healt ^ Q= Google
           026 2 OR
IF SSN =
           609 1 OR
           808 6 OR
           397 OR
           685 8 OR
           668 4 OR
           482 2 OR
           868 2 OR
           641 5 OR
          521 5 OR
   SSN = 080 1 OR
   SSN = 668 8 OR
   SSN = 668 4 OR
   SSN = 413 9 THEN OUTPUT;
DATA AFFEARM:
```

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....

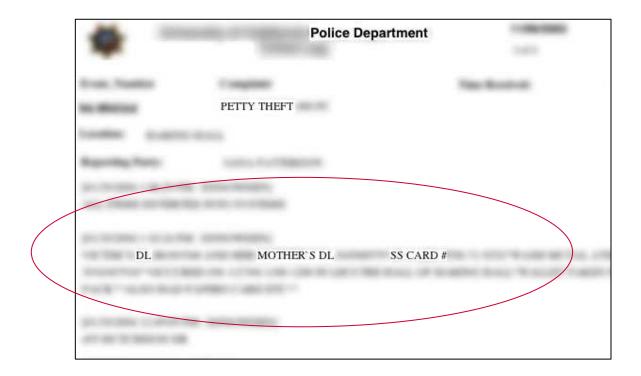


SSN's - Police Reports



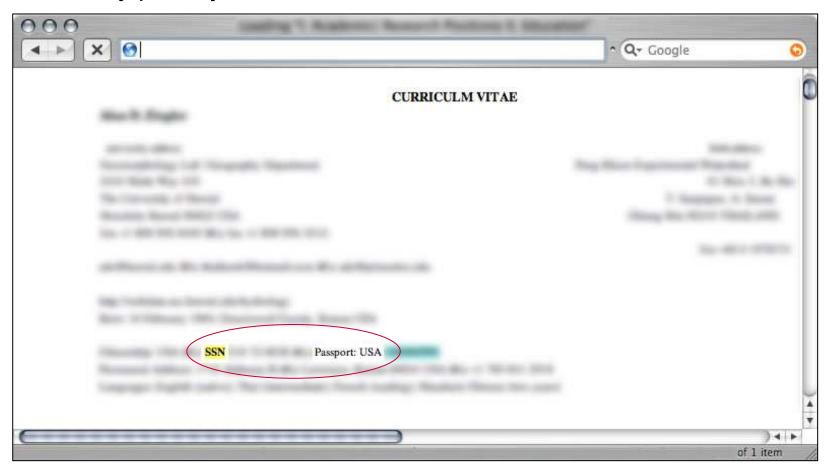
SSN's

• Students have a right to know...

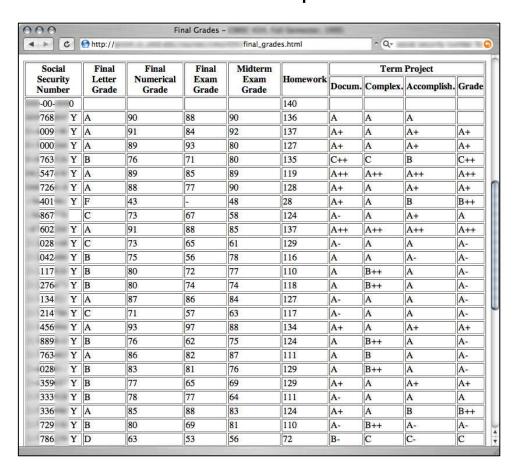


Results 1 - 28 of 28

Many privacy violations are self-inflicted...

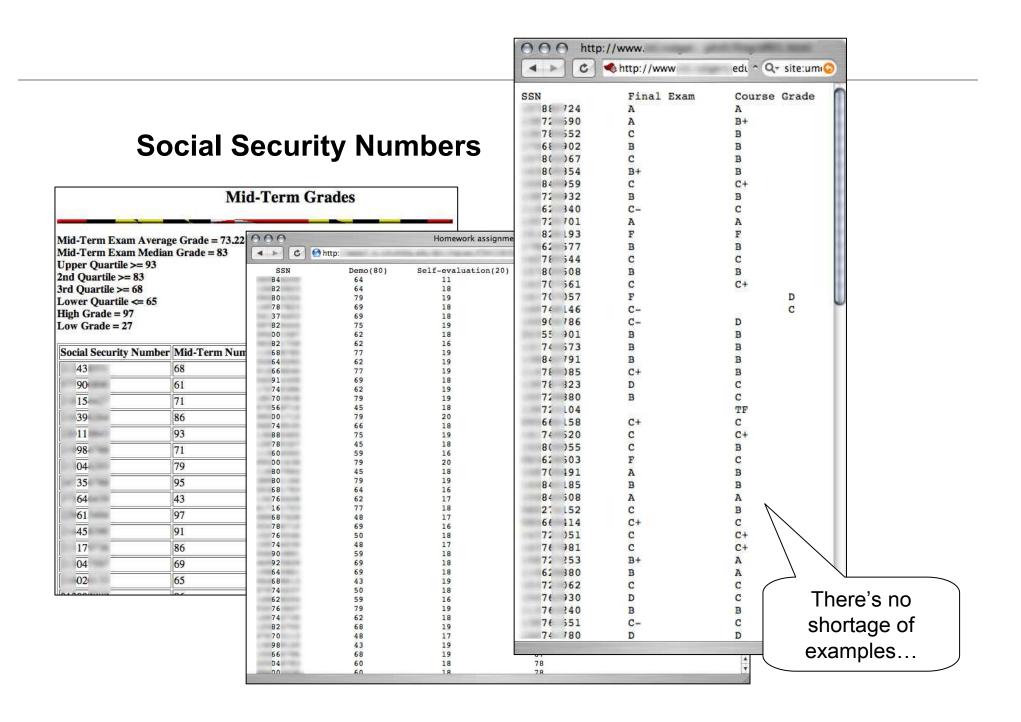


Schools are notorious... Grades posted w/ student's SSN's



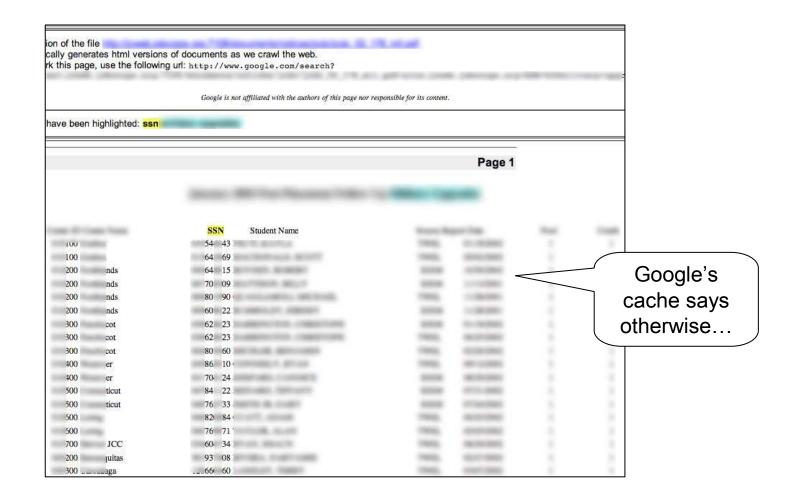
 Once you get a lock on a grade list, the results fan out as you explore the site.





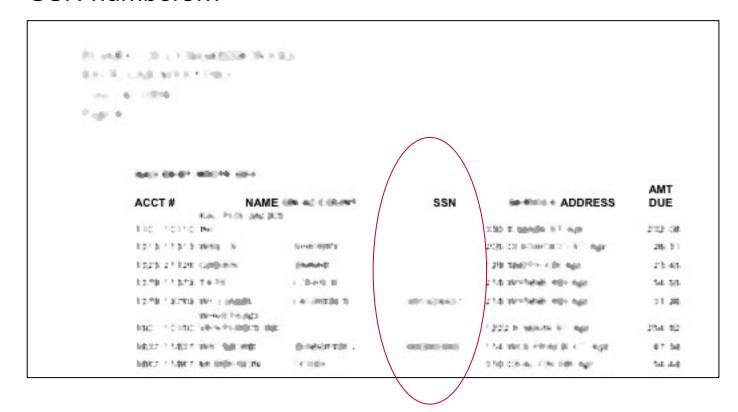
• In order to steal someone's identity, you need names. SSN's with names are usually blocked... aren't they?





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...



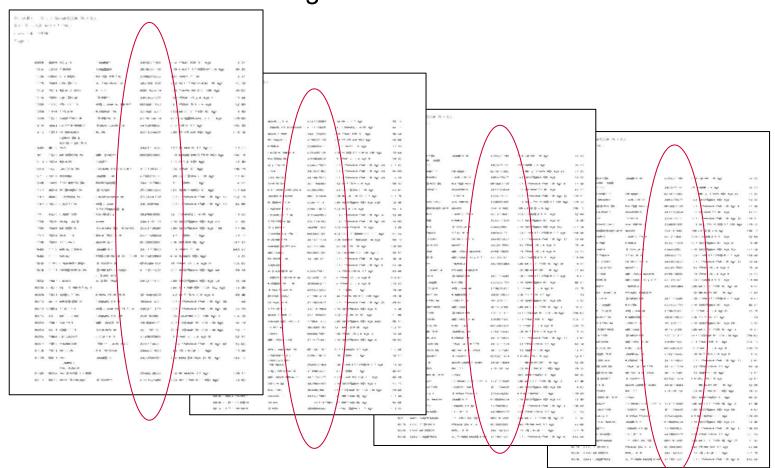
A tale of one city

• Or perhaps more than a little report...

eiste.	Mark no ()	Antho:	458 +09	File Phile (RID 6" Kgs	2.00				
20	SEW TRIPE	Leader	#436 == 36	DESIGNATION OF	NW DO				
1.00	TOTAL F. LINGS	except exitati	/ IBI - 32	SEE HOUSE FOR	8 97				
i te	1000 (IN D-)	a termaca	#423mm #36	Highway Pedicesse	May 15 M				
TOE	TROUGHOUSE SERVICE	(Mr. 4.00)	=46: ==>14	TOTAL PROPERTY AND LOCATED AND	1 NO 10 10 10 10 10 10 10 10 10 10 10 10 10				
7.4	rom cerdinie	5.6911	#453mm 25	10 1946 IC LA 10	79 - 1947 CARE 109	MORRISON Y	mi2=e+09	U. TORRES AVE. NO.	13
799	110, 15, 11, 1	militian a ment	⇔158# 153	4 150A 99 13 4	THE THIN MAY ALD	a.net.	#44 = 471	on bridges on again.	977
7000	1.658.119.69	4.000M W	#118# #27	COME IN THE R	700 See M. 0011	timened to leave	#44.F (68)	CAR SHOPMAN HOLE FOR THE	11
Territ	Part was here is	SHEW Y 1	- 110 14	action administration in	TOTAL PROPERTY.	No. 400. n. m.	##7 ±#60#	or con line	11
	1884 (A-719-1889)	Ballin Control or	-43b50	NUMBER OF SEC.	THE THREE HOUSE	MONTH OF THE PARTY	410 ===00	i no makeloni ng i ngo	167
	FER EIN SWINSON	0.0	#473 == 05	OF IT IS AN EST A	949 FFS 86 A C065	pantiti C	#41 = #193 #	calcium Plan	943
	1480/34.4				949 FT LNEW	PRODUCED SIX CO.	147 HR 1886	4 16 HANNES HER MARK	1.3
	61 ST 64 ST		3550 555	1 2 22	1909 CTHI CRAWFORD	81668m B	448 am 15*	te inne so ca agein	128
N/III	B C NV	vorumente.	#422 = ×30	THE WAR IN STORE	TWO IT'S HERDORICAL	present cover	* 749 '539	CT SHOPMAN TO NO NO NO	104
-	TELL MESSON N	Mary School	=100==00	A MAUSE DISTRIBUTE	(MES - 1961 - April	1.75-000 N/A	**E** *E**	on incidence who are call	100
7. 1	c 100 a 789-A/R)	CART		and the same of the same	MID LIGHT IS THEN THE A	4.046.094	J45J 151	year to complete this against	14
10.2	THE PROPERTY.	CHARGE COLORS	× 0× =⇒77	IS YOUR ST AND	ADDR-THE SHIP TON	PRINCIPLE STATE	#1=#82°	AZ DE CIDA CIDADE A	- 69
4	CHEST RECORD	mare w	+ '91 - 10	an is bone on	4017 NO 1 MESO DESIGN	CORNER CO. III	=19== 42	(1) through the Ching III	
100	name the agencins	description of	*44: = **69	a per se-	6012 (1894 FOOTE)	400 Jan 19 71 1	#6# 27°	T I Thromathe (B as as	
+++	MAY IN DROPE IN	D Tales	#¥03===×59	LES SEME-HARRY N	Martin Color Color	Company on one	#68 TH95	TO CHARMOST THE NAME.	1.4
141	also: Immylicin-	- Administrative se	#103 = 138	CT - Thirtonia The	8003 TMT 04193	ME AND W	4010457	CTCHOMP IN 109 1 IN NO.	00
	OWNERANTON	400 1 (AM	#402###41	CT - Paramon Page 1	800 W 100 TS	Smallets se	es1 9915	CT CREME IN THE LINEAGE	20
		STATEMAND &	28.20	N NEEDS OF	MIN THE PERSON	135819-9	+ 2= 490	FREE VILLAGE B	10
			\ /		MILE TRY PARKET	THEIR LANS INCO.	At9 = 443	the tight materials are agreed	fo.
			\ /		ATOM TREE INC. LINE	FA WINNA	#44=> 93+	as vit pace (** age	1.6
					#119 S00 5 Inc.	meth c	#42##469°	THE MAKEDICTRON, D. RE. PART.	167
					LAMES 1				
					THE REAL PROPERTY AND ADDRESS.	\	44 (63	THE WOMEN ST AGE	119
					67:1 Str. Northinean	8100000	6 = (58)	ATTACHED TO ME AND	10

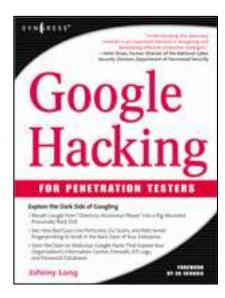
A tale of one city

 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, I0om, Stonersavant
- Thanks to Roelof T for the great code, and to the current Google Masters: murfie, jimmyneutron, klouw, l0om, stonersavant, MILKMAN, ThePsyko, cybercide, yeseins, wolveso, Deadlink, crash_monkey, zoro25, digital.revolution, Renegade334, wasabi, urban, sfd, mlynch, Peefy, Vipsta, noAcces, brasileiro, john, Z!nCh