Google Hacking

Search Engine Black-Ops

Joshua Brashars



Obligatory C.Y.A. Disclaimer:

I am in NO way, shape, or form affiliated with the almighty Google. Google is a registered trademark, owned by people that are almost completely, but not at all like me. Void where prohibited, actual colors may vary, see your dealer for details, batteries not included. So please, Google, don't sue me or pull the plug on me. I can't imagine a life without Google, and trying to makes me cry, just like at the end of Old Yeller. What a great movie.



Now that that's out of the way...



Who the heck is this guy?

- Based out of San Diego
- A moderator of http://johnny.ihackstuff.com/
- IT Support and Network Security
- A heck of a dancer
- Not as funny as he thinks he is...



Google Hacking?!

- What it is *not:*
 - NOT hacking into Google itself!
 - NOT something that requires "leet skillz"
 - NOT limited to security!
 - NOT related to the O'Reilly Book about SEO



Ok, so what is it then?

Simply put, mining data the Google search engine has already indexed.

- YES! It is easy...
- YES! Anyone can do it...
- YES! It can be very dangerous...
- YES! It is a great book written by Johnny Long...
- YES! That was a shameless plug...



Advanced Operators

Before we can walk, we must learn to run. In Google's terms, this means understanding advanced operators.



Advanced Operators

- Google advanced operators help refine searches.
- They are included as part of the standard Google Query.
- Advanced operators use 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

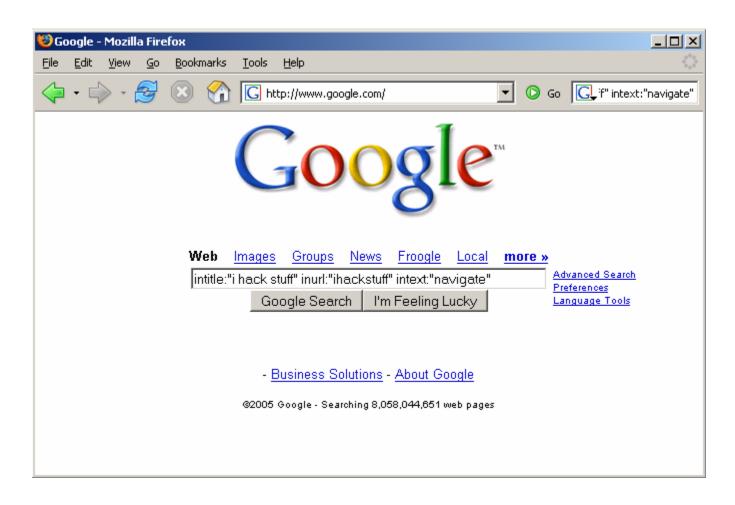
- intitle: Search page title
- inurl: Search URL
- site: limit results to a specific site
- link: other sites that link to our subject
- inanchor: search within hyperlinks
- filetype: Starting to see a patern yet?



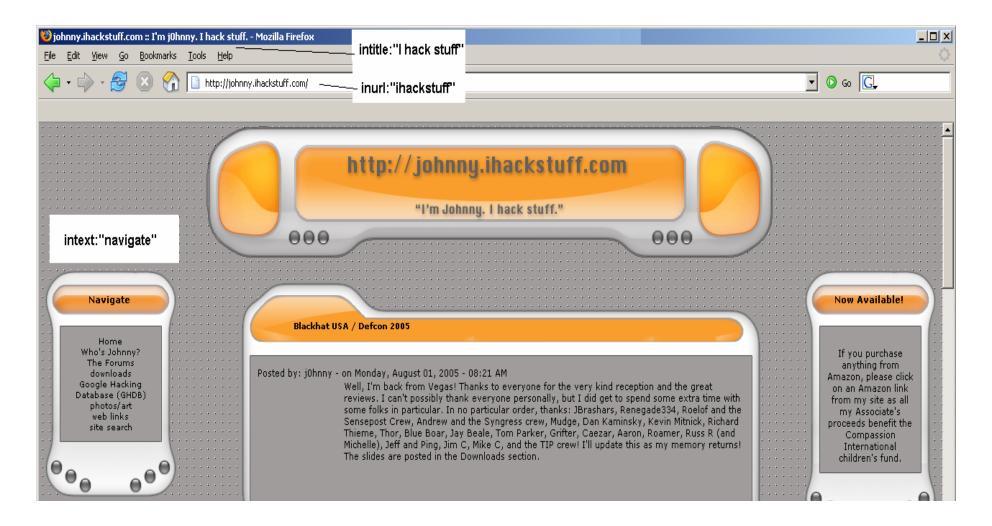
A note on numrange...

- Received a lot of press in the past
- Used for credit card and social security number searches.
- Sorry, that type of stuff is beyond the scope of this talk.

A crash course in Advanced Googling



Advanced Google Searching





Google Hacking Basics

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



Google Hacking Basics

...can have devastating results!

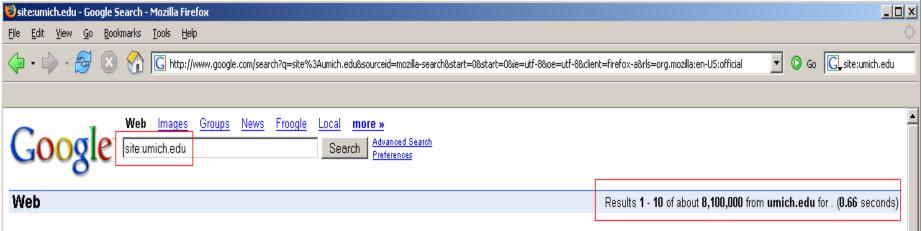




Basic Domain Crawling

- The site: operator narrows a search to a particular site, domain, or sub domain.
- Consider, site:umich.edu...





How to Apply

UM Chemical Biology. How to Apply. IT IS RECOMMENDED THAT APPLICANTS APPLY ONLINE. This will ensure the fastest response to your application. ... www.chembio.umich.edu/apply/ - 9k - Cached - Similar pages

UM SOE: Administration

Quick Links. Dean's Office · Student Travel Reimbursement · Educational and Soc. Justice Committee · Prospective Students · Who we are ... www.soe.umich.edu/administration/ - 14k - Cached - Similar pages

UM Office of the Provost: Arthur F. Thurnau Professorship

Arthur F. Thurnau Professorship. General Information. The Thurnau Professorships are named after Arthur F. Thurnau, a student at the University of Michigan ... www.provost.umich.edu/programs/thurnau/ - 14k - Cached - Similar pages

The University Record

The University of Michigan · News Services · The University Record Online, search, front · accolades · briefs · view events · submit events · UM employment ... www.umich.edu/~urecord/events_submission.shtml - 18k - Cached - Similar pages

Welcome UM Comprehensive Cancer Center

Institutional information about this Ann Arbor, Michigan facility, including access to the Patient Education Resource Center.

www.cancer.med.umich.edu/ - 12k - Jul 31, 2005 - Cached - Similar pages



Basic Domain Crawling

- Most obvious stuff floats to the top
- As a security tester (or an attacker) we need to get to the less obvious stuff
- www.umich.edu is way too obvious.



Basic Domain Filter

- To get rid of the most obvious junk, do a negative search!
 - site:umich.edu -site:www.umich.edu





Web Images Groups News Froogle Local more »

site:umich.edu -site:www.umich.edu

arch Advanced Search

Get the Google Toolbar

Web

Results 1 - 10 of about 5,340,000 from umich.edu for -site:www.umich.edu. (0.31 seconds)

UM-SSW: Information Request

Admissions and Financial Aid Links About the Area. About the Area. Admissions. MSW Program · Doctoral Program · Non-degree Enrollment. Fees & Expenses ... www.ssw.umich.edu/admissions-doctoral/infoform.html - 20k - <u>Cached</u> - <u>Similar pages</u>

UM School of Music - Prospective Students

Welcome to the University of Michigan School of Music! We look forward to working with you during your time at the UM. You probably have many questions that ... www.music.umich.edu/prospective_students/admitted.htm - 54k - Jul 31, 2005 - Cached - Similar pages

JOBS at the University of Michigan

University of Michigan Archived Posting (For Information Only. Do NOT Apply.)

Printable Version (opens new window). Posting No: T-045873-DW ...

websvcs.itcs.umich.edu/jobnet/job_posting.php?postingnumber=045873 - 8k - Cached - Similar pages

UM | Museum of Art (UMMA)

search · e-news · become a member · UMMA Logo · Exhibitions · Collections Galleries · Coming Soon · Past Exhibitions · For Students ... www.umma.umich.edu/view/past.html - 10k - Cached - Similar pages

How to Apply

UM Chemical Biology. How to Apply. IT IS RECOMMENDED THAT APPLICANTS APPLY ONLINE.

This will ensure the fastest response to your application. $\boldsymbol{\ldots}$

www.chembio.umich.edu/apply/ - 9k - Cached - Similar pages



Basic Domain Filter

- This has several benefits:
 - □ Low profile. The target can't see the activity.
 - Results are "ranked" by Google. This means that the most public stuff floats to the top. Some more interesting stuff trolls to the bottom.
 - □ Leads for follow up recon. You aren't just getting hosts and domain names, you get application data just by looking at the results snippet. One page of results can contain *tons* of info, such as e-mail addresses, names, etc...
 - □ We can explore non-obvious relationships. This is HUGE!



You're ranting, Josh...

There are downsides, though.

□ In many cases it would 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 through Google.

Google Translation as a proxy

- Use Google to do your work
- English to English translation
 - Still get the content, still readable, not your IP!
 - □ http://www.google.com/translate?u=http%3A%2F%2Fwww.umic h.edu&langpair=en%7Cen&hl=en&ie=UTF8





Google translation as a proxy

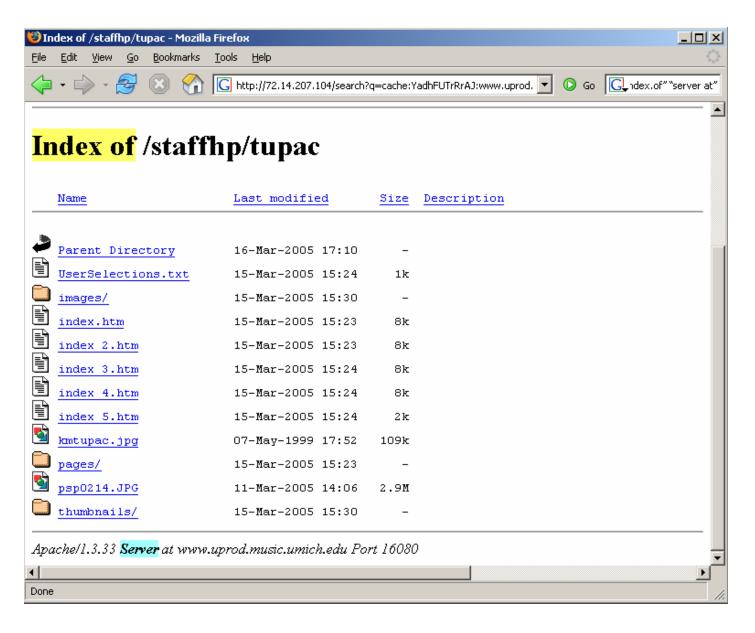
- The Caveat Images
 - Not truly anonymous
 - Images requested from the site will still be processed with our IP address
 - □ Still, it's a creative use of Google
 - □ Always test your proxies!
 - www.whatismyip.com



Server Identification

- Intitle:"index.of" "server at"
- There are two ways this is useful
 - If an attacker knows what version a server is, he may be able to locate an exploit for it
 - If an attacker has an exploit for a certain type of server, Google can ferret out some vulnerable hosts

Server Identification





More server identification queries

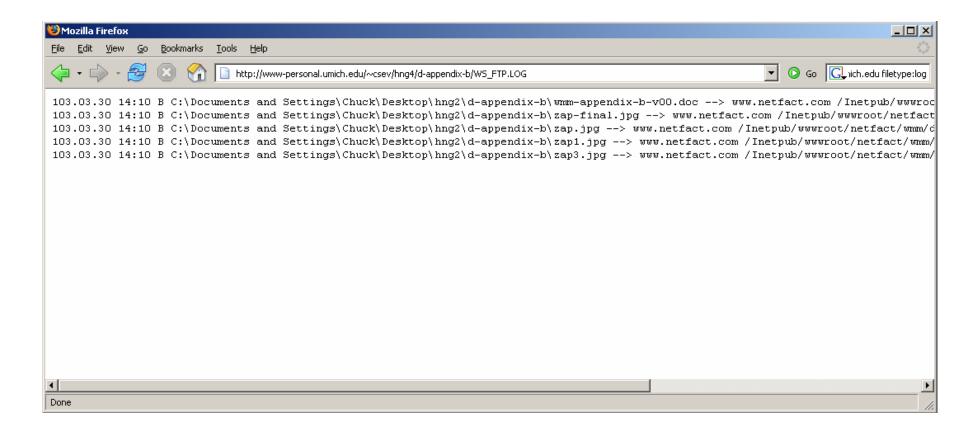
- "Apache/" "server at" intitle: "index.of"
- "Microsoft-IIS/* server at" intitle: "index.of"
- "Oracle HTTP Server Powered by Apache" intitle:"index.of"
- "Red Hat Secure/3.0 server at" intitle:"index.of"
- "Apache Tomcat/" intitle: "index.of"
- "AnWeb/1.42h" intitle:"index.of"



Finding specific files

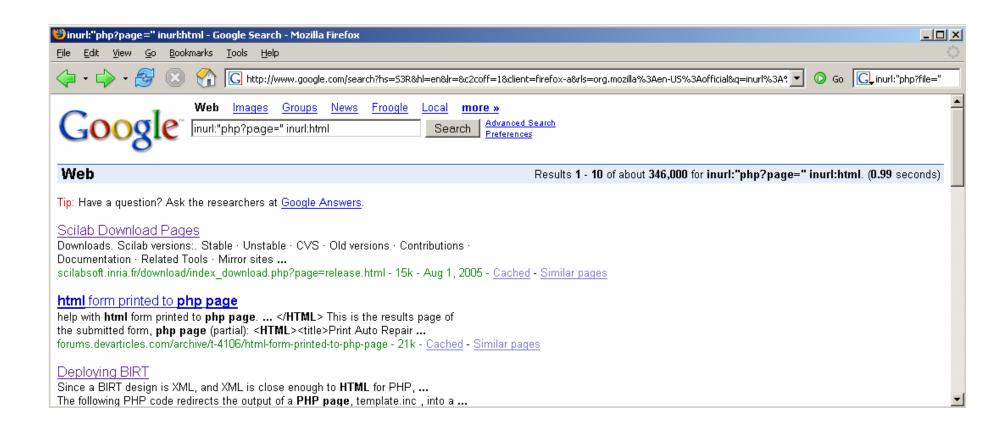
- The filetype: operator allows us to find specific types of files.
- Consider log files, such as ws_ftp.log
 - Log files often contain juicy info such as IP addresses, directory structures, and more...
 - □ Site:umich.edu filetype:log

site:umich.edu filetype:log



Directory Transversal

inurl:"php?page=" inurl:html



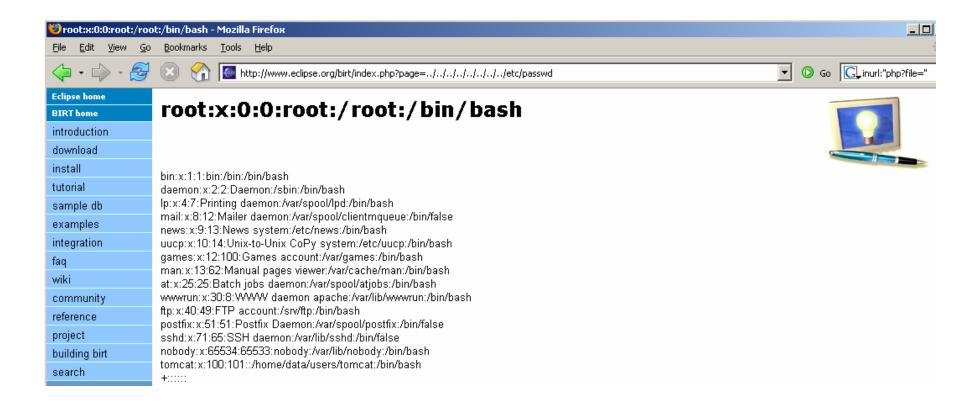
Directory Transversal

■ This...



Directory Transversal

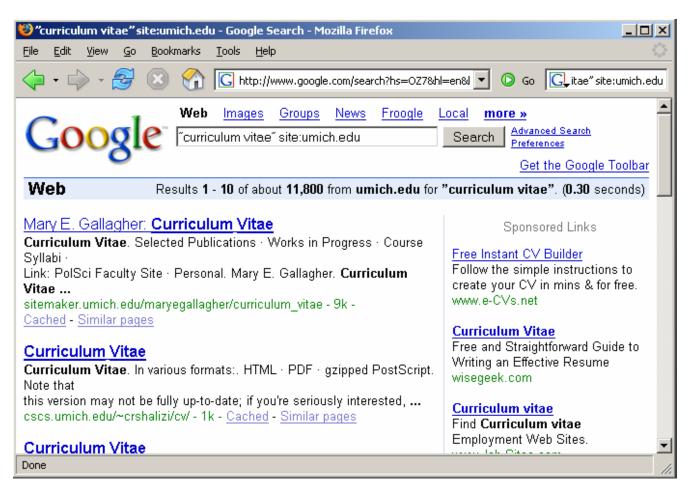
...becomes this!



Social Engineering

Resumes can be valuable!

"curriculum vitae" site:umich.edu





robots.txt

- Robots.txt can provide a roadmap for unknown, and potentially sensitive, directories and files.
- Robots.txt should not be spidered by the web server... but is that always the case?





- User-agent: *
- Disallow: /htbin/
- Disallow: /shtbin/
- Disallow: /stats/dynamic/
- Disallow: /stats/static/
- Disallow: /search/
- Disallow: /caen/EITC2004/
- Disallow: /ipe/studyabroad/funding/scholarships/
- Disallow: /caen/news/Volume_18/
- Disallow: /caen/news/Volume_19/
- Disallow: /caen/news/Volume_20/
- Disallow: /admin/dean/
- Disallow: /caen/systems/
- Disallow: /caen/staff/
- Disallow: /lost/
- Disallow: /class/eecs381/
- Disallow: /class/eecs493/



Zero-Packet Port Scanning

Why get your hands dirty when someone else will do it for you?



Whatchoo talkin' bout, Willis?

- Ok, before you throw things at me, allow me to clear up a few things about the phrase "zero packet" in this context:
 - □ Passive techniques are truly zero-packet. That's not what I'm talking about.
 - I'm talking about zero packets directly from source to target. Think proxy. It's about staying out of the targets logs.
 - Um... plus this is a talk about Google Hacking, sheesh!
 - □ Oh, come on, it's silly but it's still fun!



Zero-packet verification

- So, it takes a few packets from us to the target to verify and fingerprint hosts.
- Now, DNS resolution is no big deal, but port scanning is. This flags IDS systems.
- Is there an interesting way to do traditional recon without sending any packets directly from us to the target?

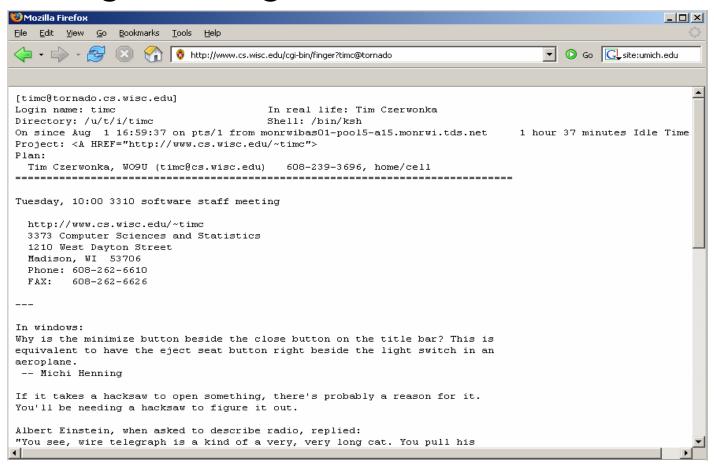


Everyone, say it with me...

(yes, even you in the front. Say it with me...)

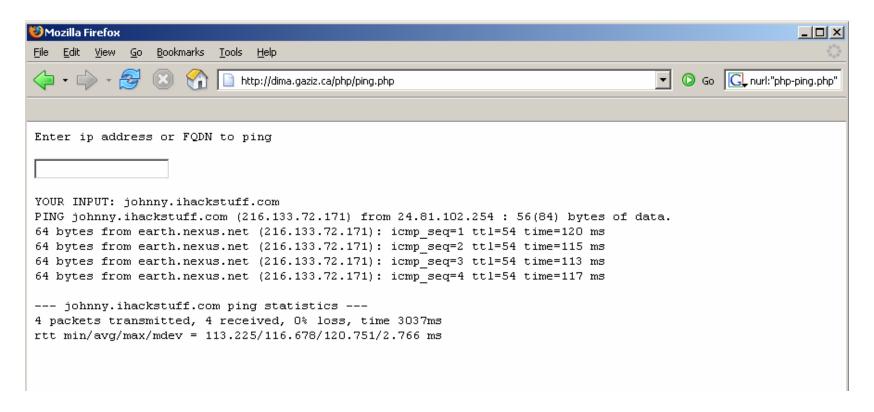
Old School! Finger...

inurl:/cgi-bin/finger?"in real life"



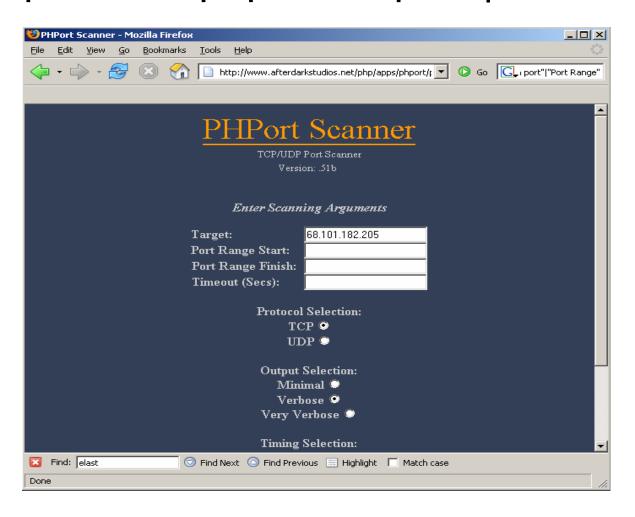
PHP Ping

"Enter ip" inurl:"php-ping.php"



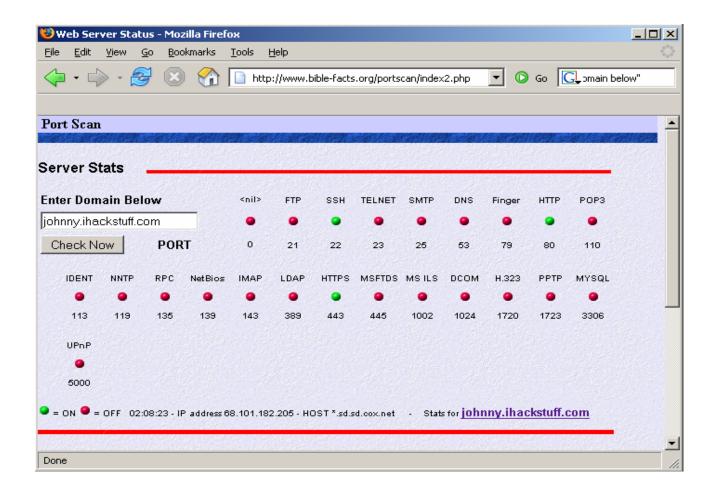
PHP Port Scanner

inurl:portscan.php "from port"|"Port Range"



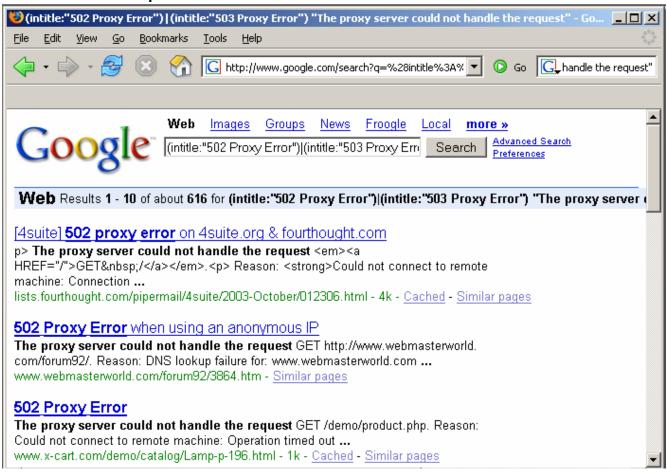
Yet another port scanner

"server status" "enter domain below"



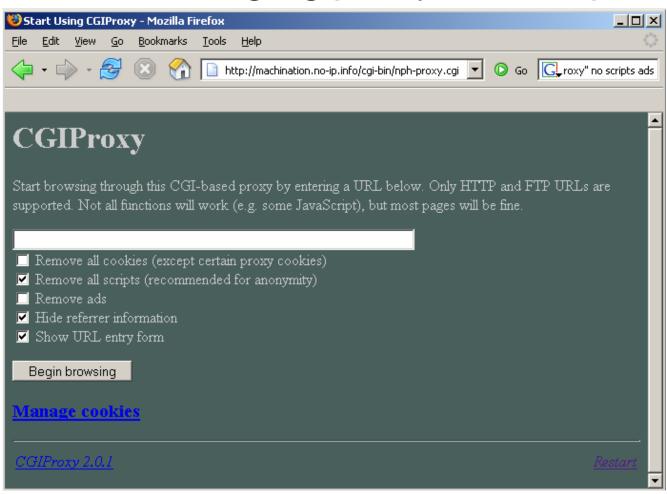
Locating proxy servers

(intitle:"502 Proxy Error")|(intitle:"503 Proxy Error") "The proxy server could not handle the request"



CGIProxy

intitle:"start using cgiproxy" no scripts ads



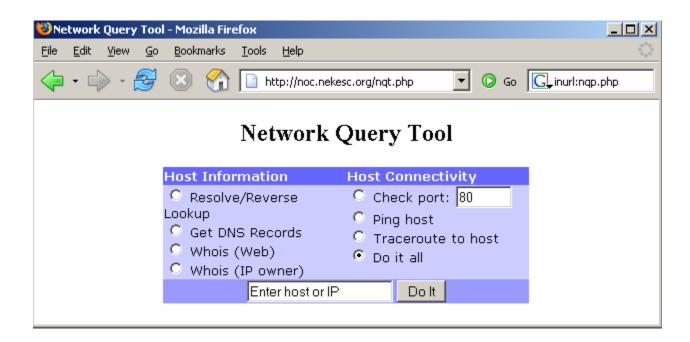
WebUtil

inurl:webutil.pl

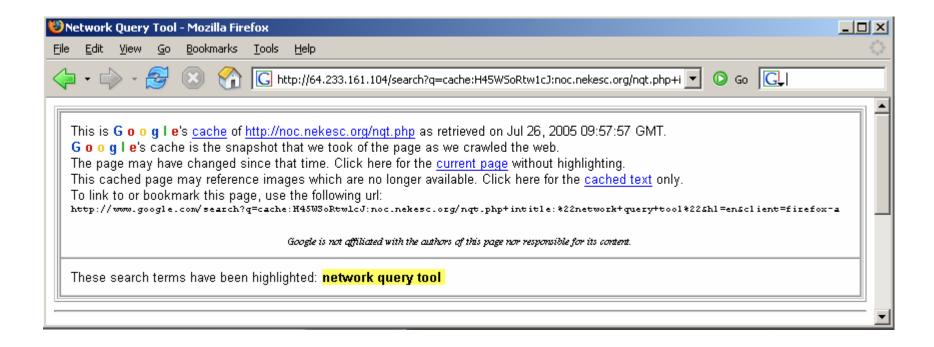


Network Query Tool

intitle:"network query tool"



Cache is your friend!





Zero-packet Recon

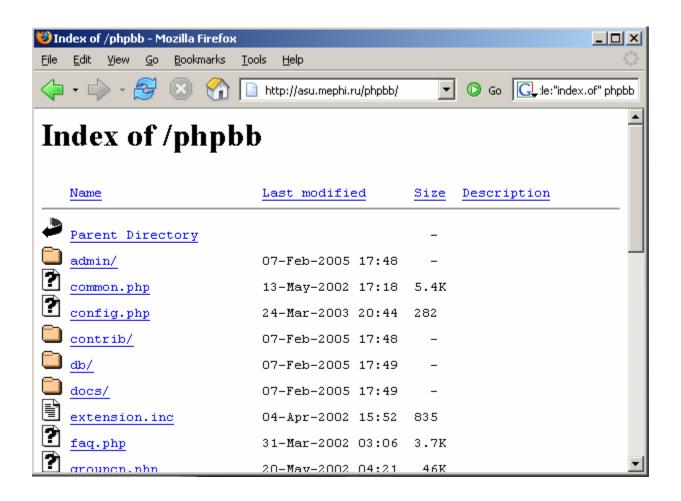
- The point is, Google can be used as an interesting, low-profile alternative to traditional recon techniques. We've used Google queries for low profile alternatives to
 - DNS resolution
 - Unix service queries
 - Network Recon
 - Web-based proxy services
 - Web crawling via cache



Directory Listings, a Google hackers best friend!

- intitle:"index of" "last modified"
 - □ Virtual file server, can reveal sensitive files web surfers shouldn't see
 - Index listings provide an x-ray into the system. Just because our target doesn't necessarily have directory listings, other sites with the same web apps might. This is handy!





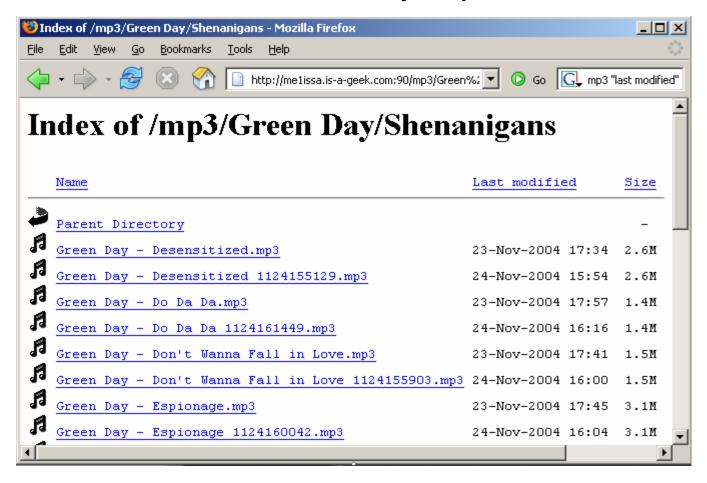


Who needs Kazaa?

- Peer to peer applications use nonstandard ports.
- Not always possible to install with given access.
- P2P Ports can be blocked at the firewall level.

Google to the rescue!

intitle:"index.of" Green Day mp3 last modified



Google Hacking Showcase, 2005!

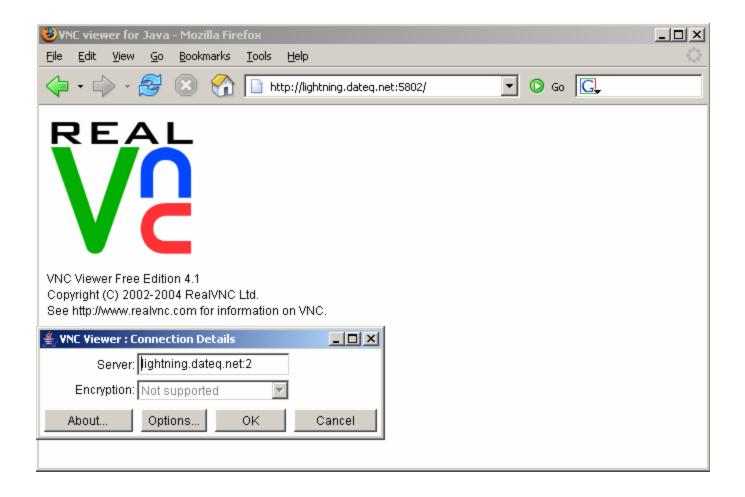
Let the games begin!

Each of these screenshots were found using nothing but Google.

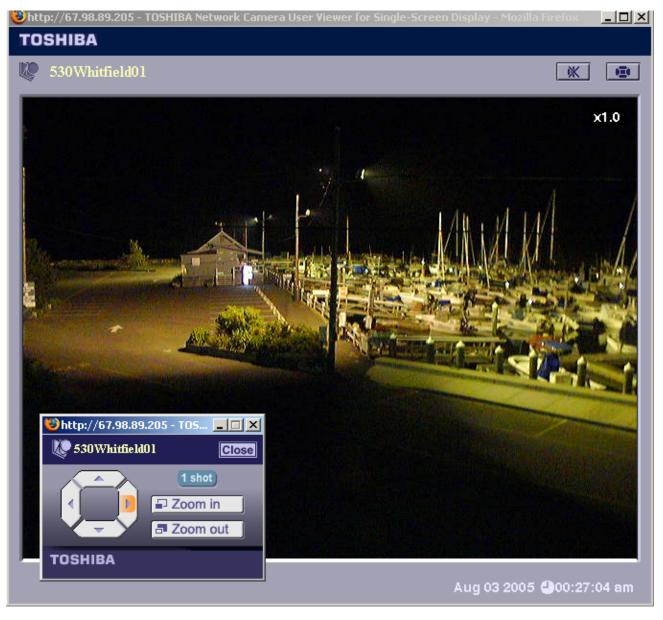
Here's some of the best of the worst:

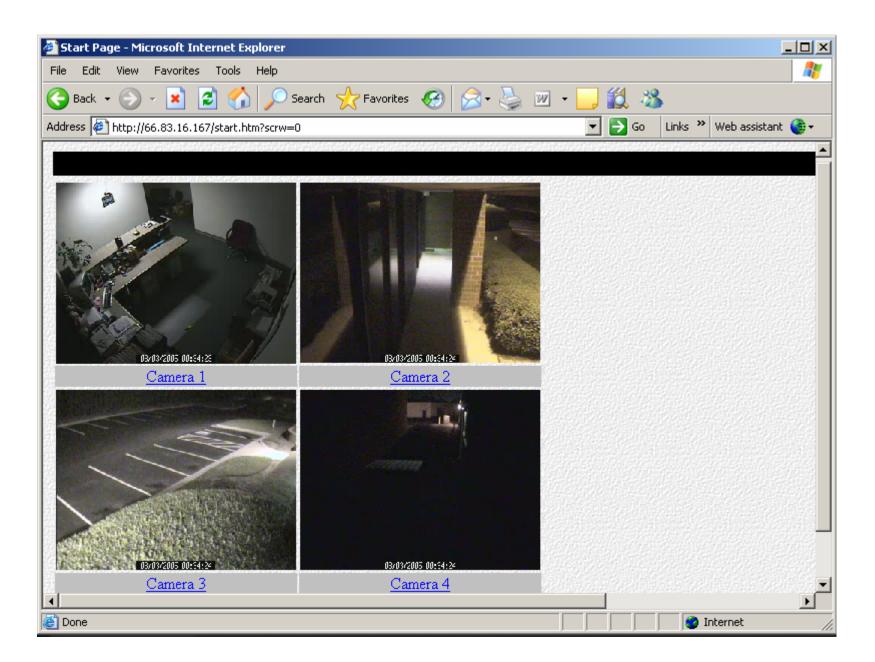


intitle:"VNC Viewer for Java"

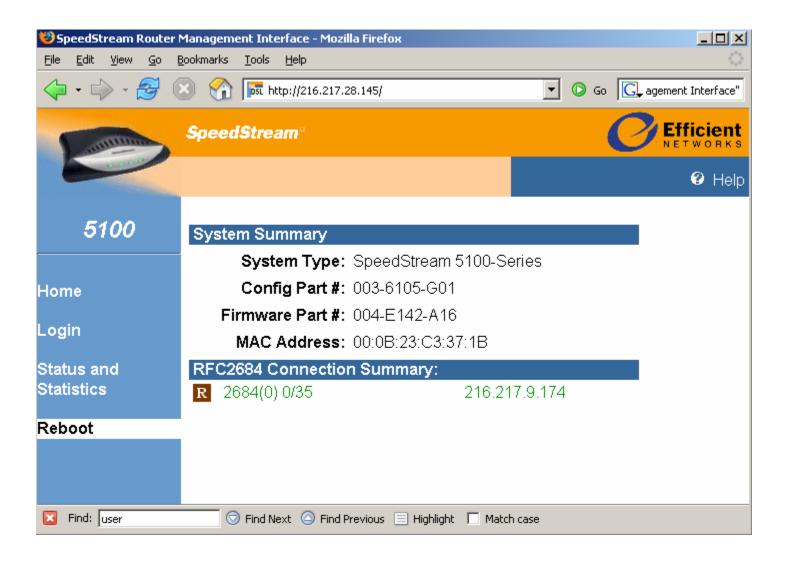


intitle:"toshiba network camera - user login"

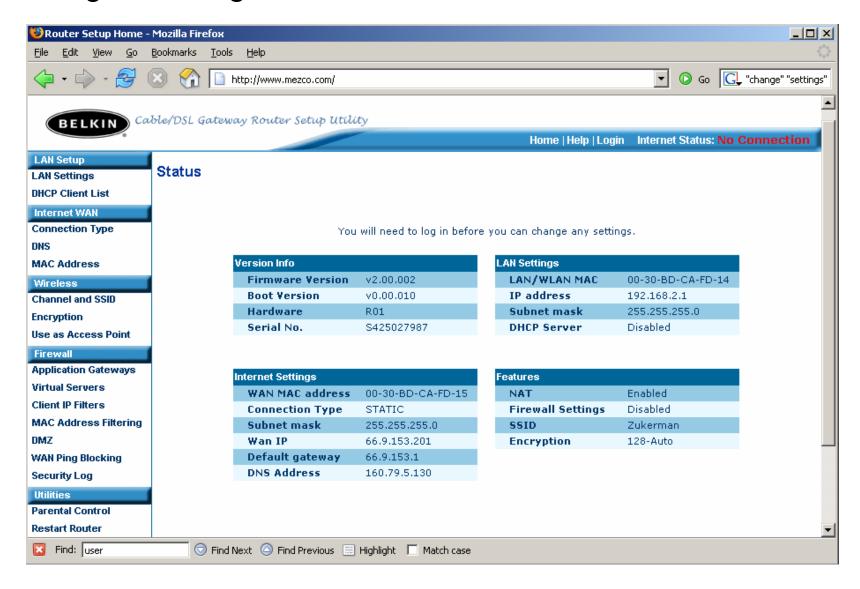




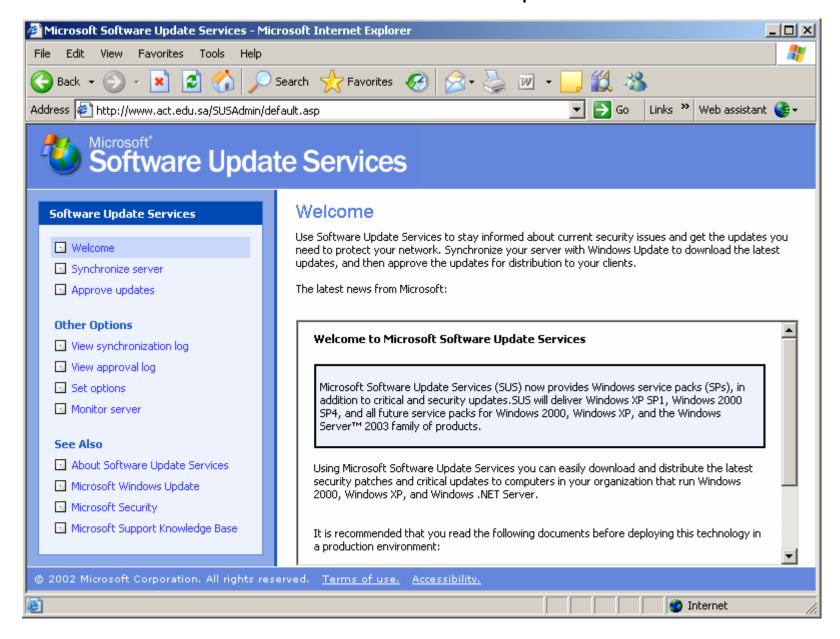
intitle: "Speedstream Router Management Interface"



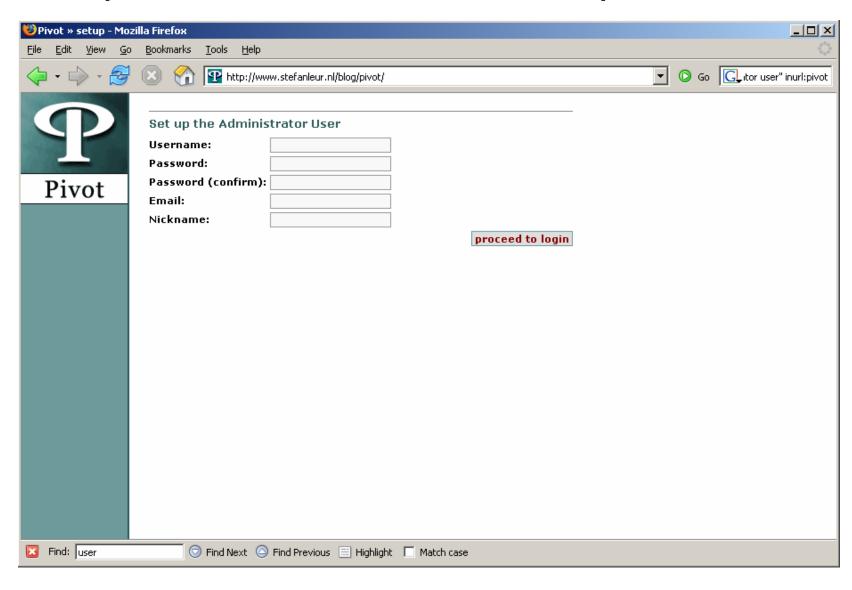
intitle: "Setup Home" "You will need to log in before" "change" "settings"



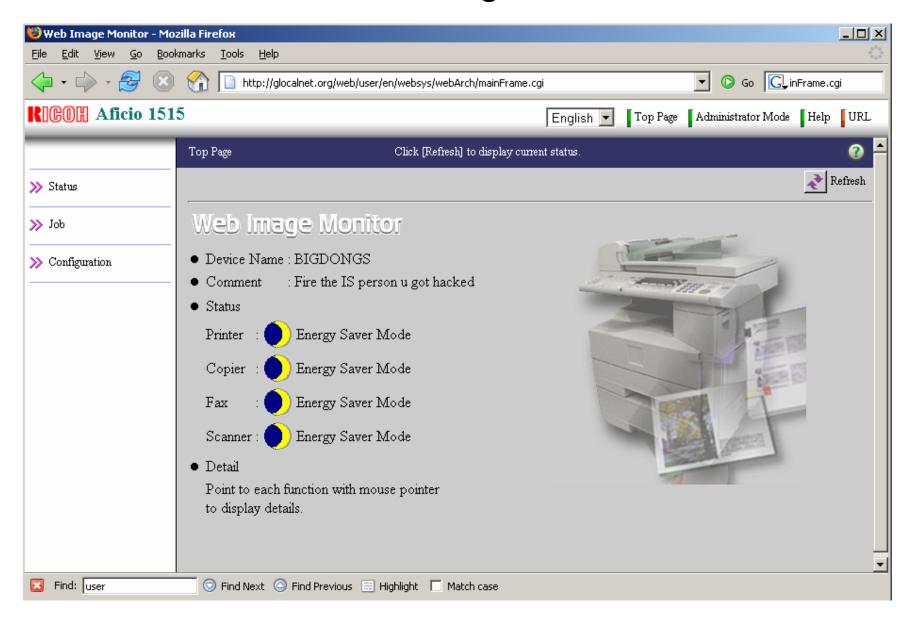
inurl:SUSAdmin intitle:"Microsoft Software Update Services"



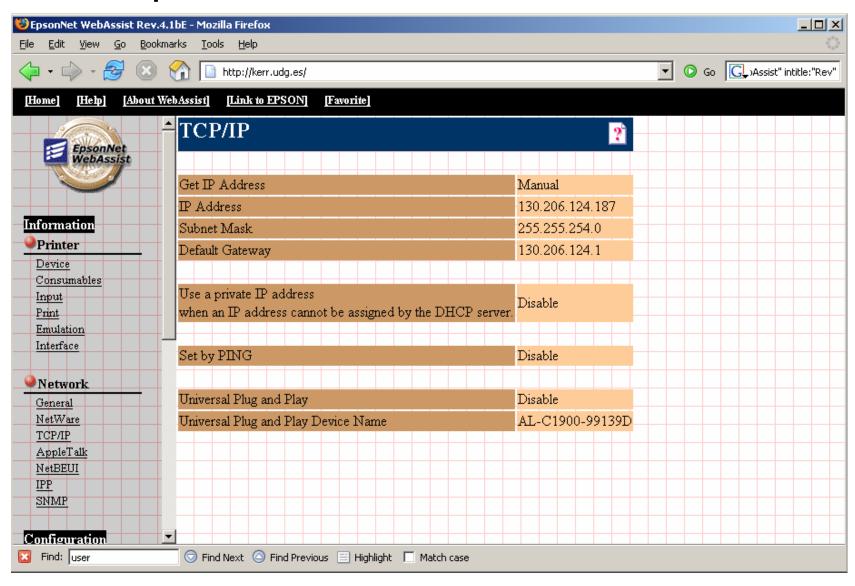
"set up administrator user" inurl:pivot



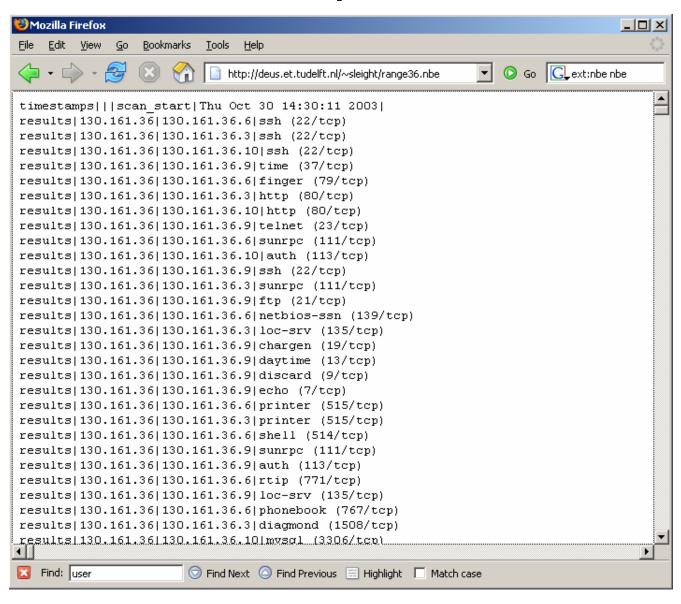
inurl:webArch/MainFrame.cgi



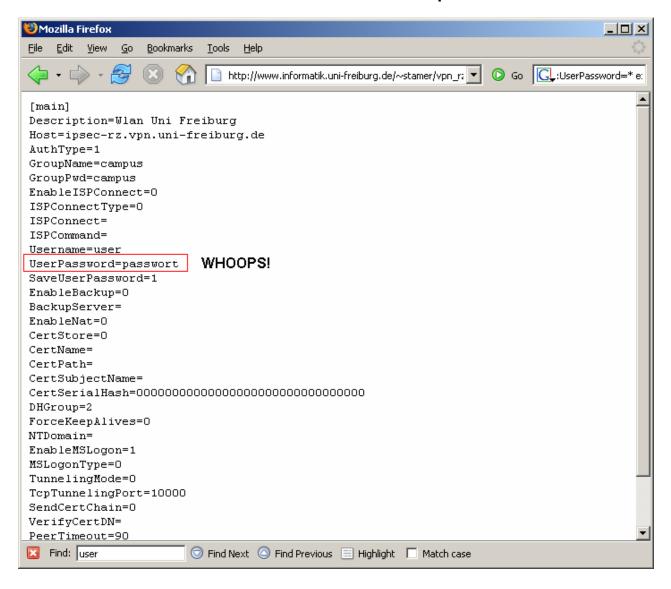
intitle: "EpsonNet WebAssist" intitle: "Rev"



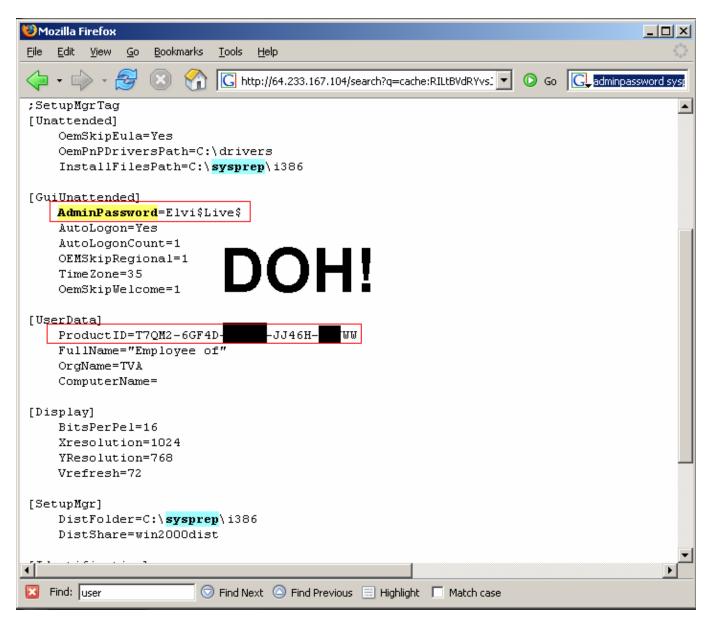
Nessus Scan output! ext:nbe nbe



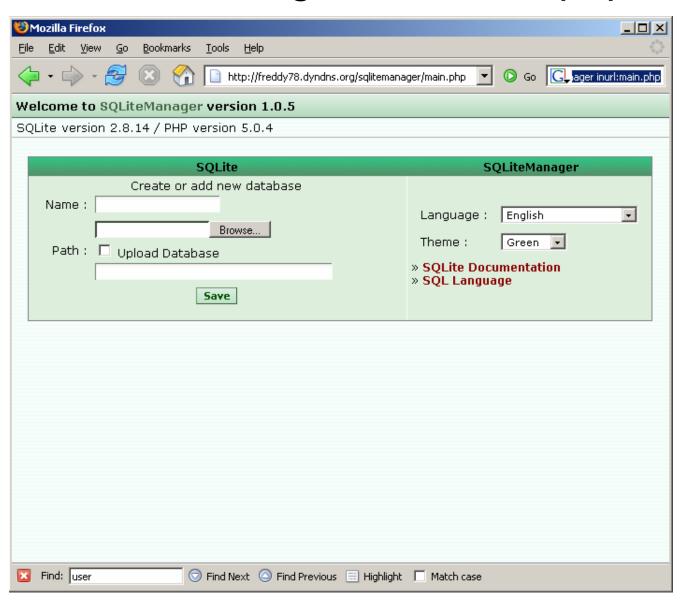
VPN User Profiles intext:Host=*.* intext:UserPassword=* ext:pcf



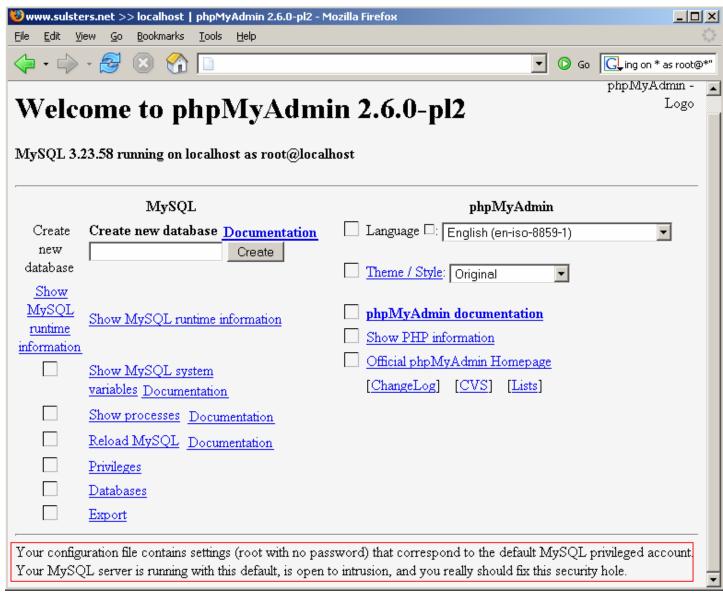
adminpassword sysprep filetype:inf



intext:SQLiteManager inurl:main.php



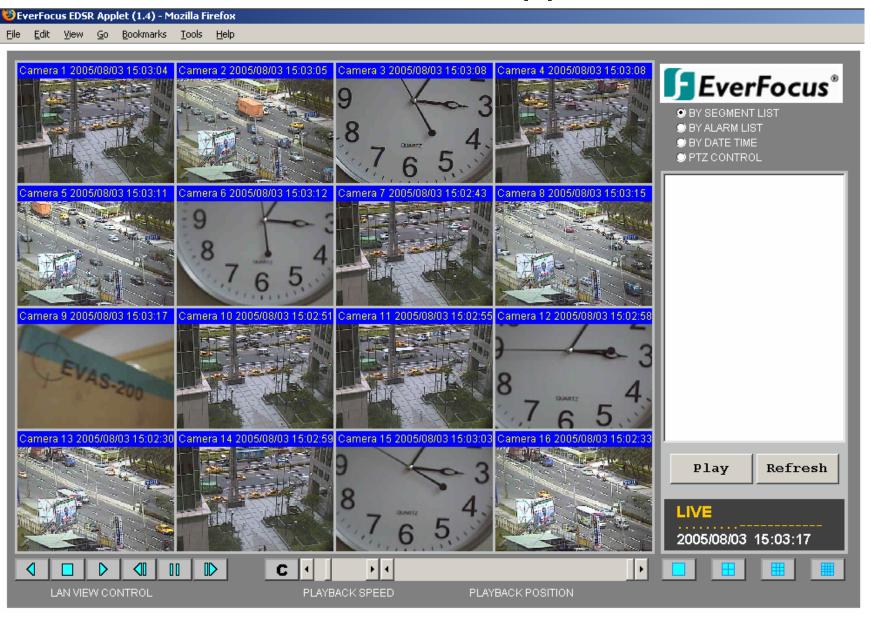
intitle:phpMyAdmin "Welcome to phpMyAdmin "*" "running on * as root@*"



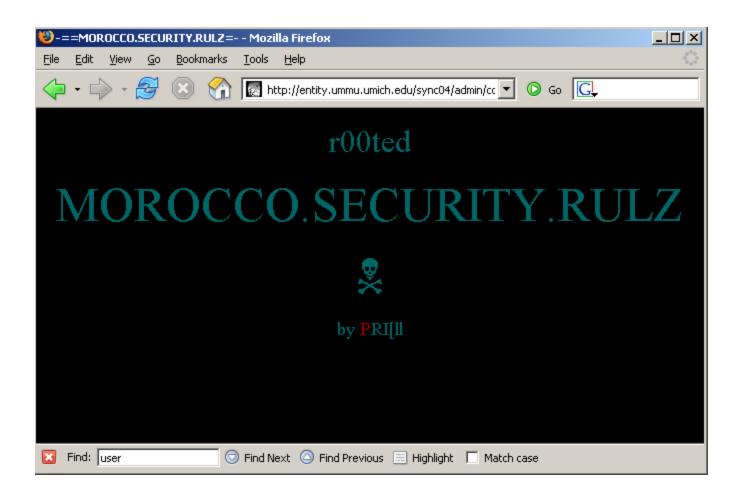
intitle: "Sipura.SPA.Configuration" -.pdf



intitle: "EverFocus" intitle: "Applet"



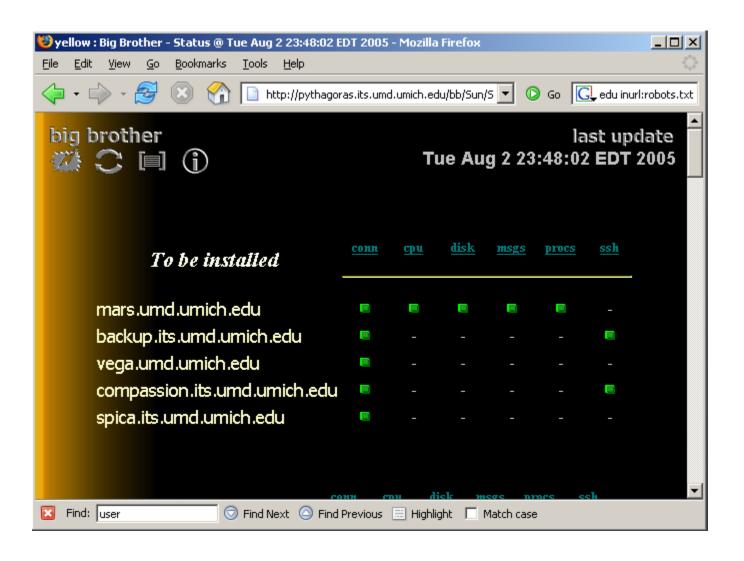
Even UMich is vulnerable...



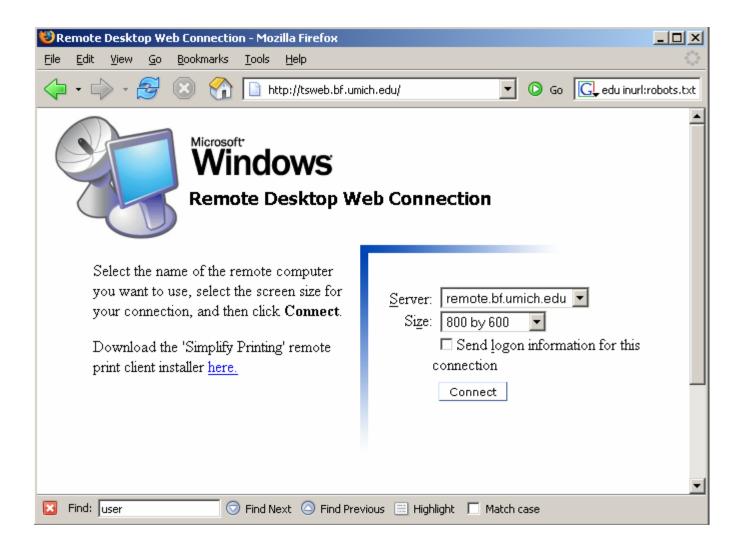
intitle:"TANDBERG" "This page requires a frame capable browser!" site:umich.edu



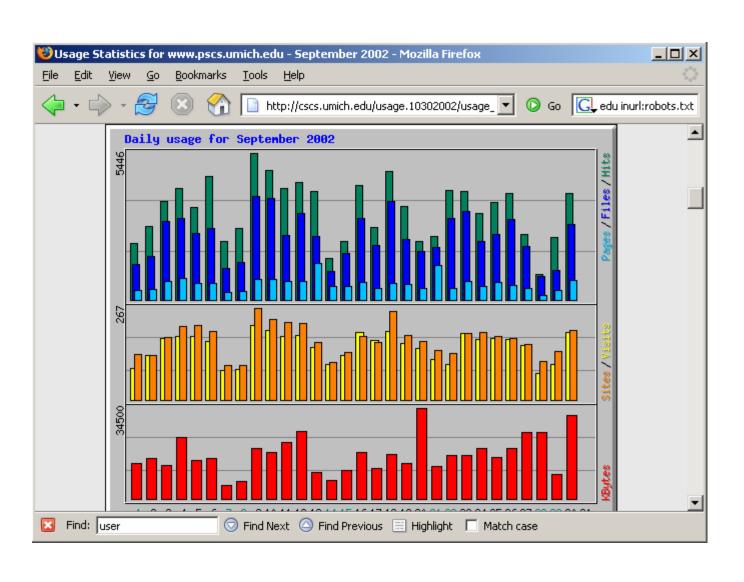
intitle: "Big Brother - Status" inurl:bb



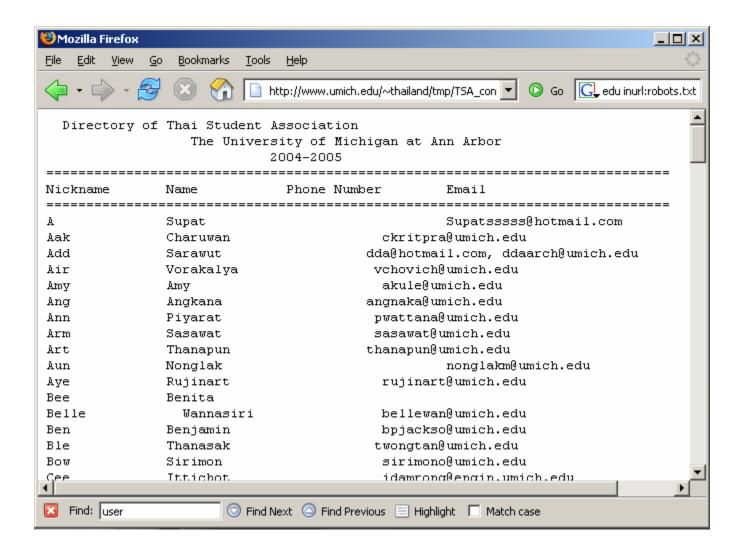
intitle:Remote.Desktop.Web.Connection inurl:tsweb



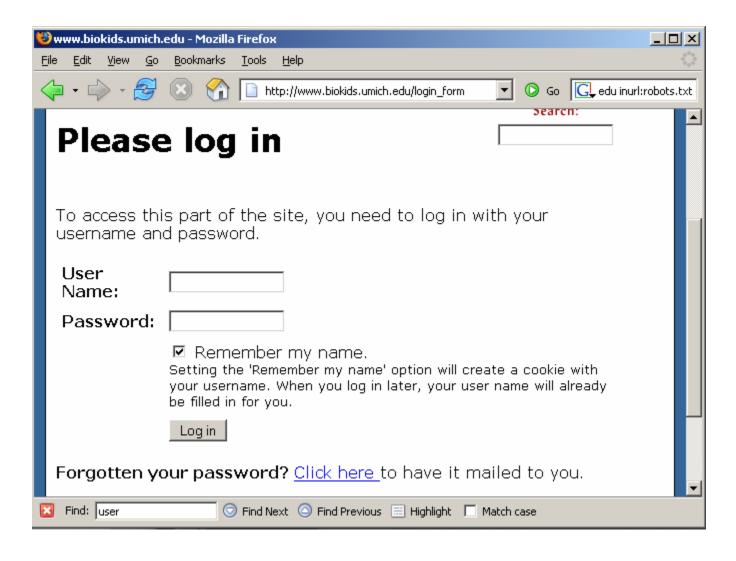
+intext:"webalizer" +intext:"Total Usernames" +intext:"Usage Statistics for"



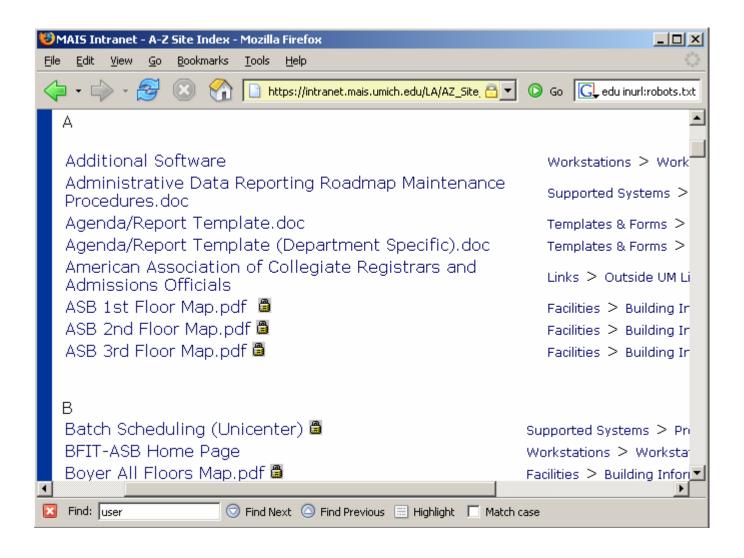
inurl:/tmp



"please log in"



intitle:intranet inurl:intranet +intext:"human resources"



inurl: "exchange/logon.asp" OR intitle: "Microsoft Outlook Web Access - Logon"





My personal favorite...

site:umich.edu filetype:mbox





So, what can be done?

- Preventative maintenance
 - Disable directory listings if you do not need them.
 - Password protect sensitive directories
 - □ Robots.txt
 - But don't let Google crawl it ;)
 - □ Don't use default passwords!
 - Do I really need to say this?
 - □ Google's removal page
 - http://www.google.com/remove.html



Go hack yourself, pal!

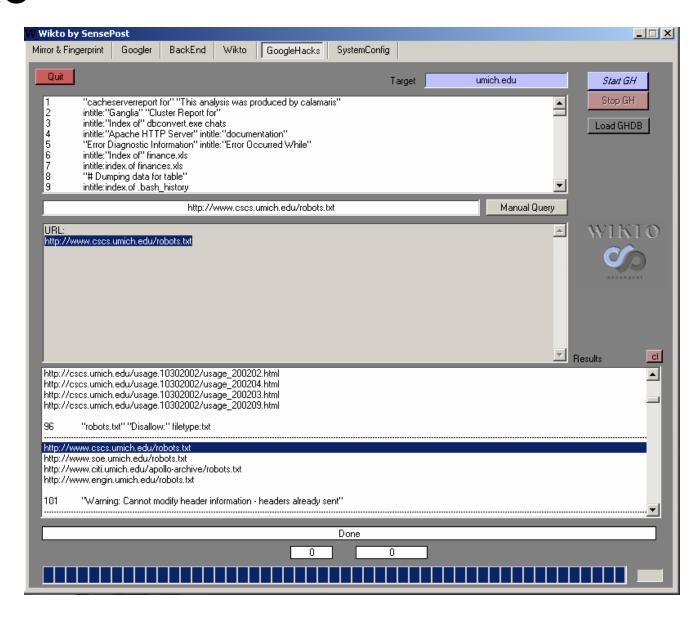
- Wikto from Sensepost.
- Athena
- Gooscan
 - Note: Gooscan violates Google's TOS
 - You really do not want Google pissed at you. Remember Old Yeller? Sadder than that.



WIKTO, by Sensepost

- Automates Google Hack Scanning
- Available for free from <u>www.sensepost.com</u>
- Requires a valid Google API Key
- Designed to allow site owners to test themselves for vulnerabilities

Wikto





Thanks!

- UMich for having me out
- Johnny Long for being a mentor and a friend
- The whole team at http://johnny.ihackstuff.com
- The endless (misguided?) loving support of my family and friends, and co-workers
- The 7-11 by my house, for always being there for me when I need them.

Without the help of all of these people and more, none of this would be possible and I might still be jockeying tapes at the video store.