網站安全攻防實務

Google Hacking

蔡一郎
大綱

- What is Google
- What is Google Hacking
- How to use the GHDB
- Google Hacking Tools
- Chose your keyword
- Demo and lab
What is Google

- Google was co-founded by Larry Page and Sergey Brin while they were students at Stanford University and the company was first incorporated as a privately held company on September 4, 1998.
- The powerful search engine.
- Using simple search interface.
- The shortest URL: g.cn
Making searching even easier

- iGoogleGuide
- http://www.googleguide.com/
Google's Boolean Modifiers

- AND is always implied
- OR: Escobar (Narcotics OR Cocaine)
- "-" = NOT: Escobar -Pablo
- "+" = MUST: Escobar +Roberto
- Use quotes for exact phrase matching
- Wildcards
  - Google supports word wildcards but NOT stemming
  - Find different?
    - 蔡一郎、蔡一*、蔡*郎、*一郎
    - airline、air*、*line
Basic Google Operator

- Exclude terms using the NOT operator (*minus sign*)
  - For example, searching **NCHC -NARL** will give you everything that has NCHC but not NARL

- Include common words using the AND operator (*plus sign*)
  - For example, searching **NCHC +NARL** will give you everything with the words NCHC and NARL

- Searching for exact phrases must be surrounded by double quotes
  - For example, "**NCHC and NARL**" will return all results that have NCHC and NARL as a phrase

- Wildcards are represented by an asterisk
  - Searching for **NCHC * "NARL"** will return all entries with NCHC any word NARL

- Google searching is not case sensitive so NCHC, nchc and NcHc are **all the same**
How Google Works

1. The web server sends the query to the index servers. The content inside the index servers is similar to the index in the back of a book—it tells which pages contain the words that match any particular query term.

2. The query travels to the doc servers, which actually retrieve the stored documents. Snippets are generated to describe each search result.

3. The search results are returned to the user in a fraction of a second.

Copyright © 2003 Google Inc. Used with permission.
How Do I Get Results

- Pick your keywords carefully & be specific
- Do NOT exceed 10 keywords
- Use Boolean modifiers
- Use advanced operators
- Google ignores some words
  - a, about, an, and, are, as, at, be, by, from, how, i, in, is, it, of, on, or, that, the, this, to, we, what, when, where, which, with

(From: Google 201, Advanced Googology - Patrick Crispen, CSU)
Some of the Advanced Google Search Techniques

- **site**
  - restricts a search to a particular site or domain
- **intitle**
  - finds strings in the title of a page
- **inurl**
  - finds strings in the url of a page
- **filetype**
  - finds specific types of files based on file extension
- **link**
  - searches for links to a site or url
- **inanchor**
  - finds text in the descriptive text of links
About filetype

- Everything listed at http://filext.com/
Advanced Google Search

- site: (.edu, .gov, honeynet.org.tw)
- filetype: (txt, xls, mdb, pdf, .log)
- Daterange: (julian date format)
- Intitle / allintitle
- Inurl / allinurl
Directory Listings

- Show server version information
  - Useful for an attacker
    - intitle:index.of server.at
    - intitle:index.of server.at site:nchc.org.tw
    - intitle:index.of "parent directory"
What is Google Hacking

- Johnny Long is the “grandfather” of Google hacking.
  - http://hackersforcharity.org/
- Google is much more than just a simple search interface and engine.
  - http://www.google.com/
- Google crawls public websites for information **every 6-8 weeks** using an automated search and record program called Googlebot.
- It is not hacking into Google.
How to use the GHDB

- GHDB=Google Hacking Database
  - http://hackersforcharity.org/ghdb/

- Cartage
  - Advisories and Vulnerabilities
  - Error Messages
  - Files containing juicy info
  - Files containing passwords
  - Files containing usernames
  - Footholds
  - Pages containing login portals
  - Pages containing network or vulnerability data
  - sensitive Directories
  - sensitive Online Shopping Info
  - Various Online Devices
  - Vulnerable Files
  - Vulnerable Servers
  - Web Server Detection

Find and Click.....
Google Hacking Tools

- **Gooscan** – Johnny Long’s free command line UNIX tool. It violates the Google TOS. Gooscan automates queries designed to find potential vulnerabilities on web pages against Google.
  - [http://johnny.ihackstuff.com/](http://johnny.ihackstuff.com/) (not found)

- **SiteDigger** – A Windows tool that searches Google’s cache to look for vulnerabilities, errors, configuration issues and proprietary information on websites. Must have **Google API license key**.

- **Wikto** – Wikto is a Windows based web server assessment tool that uses the Google hacking database (GHDB). This tool requires a Google developer license.
  - [http://www.sensepost.com/research/wikto](http://www.sensepost.com/research/wikto)

- **Advanced Dork** – AdvancedDork is a **Firefox extension** designed to quickly search for specific text inside Google’s Advanced Operators.
  - [https://addons.mozilla.org/firefox/2144](https://addons.mozilla.org/firefox/2144)
Google Hacking-search XSS

Keyword:
.gov.tw OR .com.tw
9i5t.cn/a.js
Google Hacking-Search

Keyword: index of /
Keyword: index of /passwd

- Google Hacking - Search
Keyword:
"index of /" 范逸臣 mp3 site:edu.tw
How to secure

- Anti-robot
  - configure "robot.txt" in the web root directory
  - Robots Exclusion Protocol or robots.txt protocol
  - It is a convention to prevent cooperating web spiders and other web robots from accessing all or part of a website.
allows all robots to visit all files because the wildcard "*" specifies all robots

User-agent: *
Disallow:

keeps all robots out

User-agent: *
Disallow: /
- all crawlers not to enter four directories of a website

User-agent: *
Disallow: /cgi-bin/
Disallow: /images/
Disallow: /tmp/
Disallow: /private/

- a specific crawler not to enter one specific directory

User-agent: BadBot # replace the 'BadBot' with the actual user-agent of the bot
Disallow: /private/
- all crawlers not to enter one specific file

  User-agent: *
  Disallow: /directory/file.html

- Sitemap


  Sitemap: http://www.google.com/hostednews/sitemap_index.xml
About Googlebot

- **Googlebot**
  - crawl pages from our web index and our news index

- **Googlebot-Mobile**
  - crawls pages for our mobile index

- **Googlebot-Image**
  - crawls pages for our image index

- **Mediapartners-Google**
  - crawls pages to determine AdSense content. We only use this bot to crawl your site if AdSense ads are displayed on your site.

- **Adsbot-Google**
  - crawls pages to measure AdWords landing page quality. We only use this bot if you use Google AdWords to advertise your site.
How do I block Googlebot

- Blocking Googlebot
  - Google uses several user-agents. You can block access to any of them by including the bot name on the User-agent line of an entry. Blocking Googlebot blocks all bots that begin with "Googlebot".

User-agent: Googlebot
Disallow: /
How do I allow Googlebot

- Allowing Googlebot

<table>
<thead>
<tr>
<th>User-agent: *</th>
<th>Disallow: /</th>
<th>Allow all robot</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-agent: Googlebot</td>
<td>Disallow:</td>
<td>Allow single robot</td>
</tr>
</tbody>
</table>
| User-agent: Googlebot | Disallow: / | }
The other chose

- robot.txt tool

# robots.txt tool from www.howrank.com
User-agent: *
Disallow: /
User-agent: Slurp
Disallow: /
User-agent: MSNBot
Disallow: /
User-agent: Teoma
Disallow: /
Disallow: /cgi-bin/
Example robots.txt

- User-agent: *
- Disallow: /images/
- Disallow: /stats/
- Disallow: /logs/
- Disallow: /admin/
- Disallow: /comment/
- User-agent: Googlebot
- Allow:
- User-agent: BecomeBot
- Disallow:
- Disallow: /
- Disallow: *
- User-agent: MSNBot
- Disallow:
- Disallow: /
- Disallow: *
Automatically create sets of items from a few examples.

When you’re tired of relating keywords yourself, let Google do it for you

- http://labs.google.com/sets
Lab-search

- filetype:htpasswd htpasswd
- intitle:"Index of" " .htpasswd" -intitle:"dist" -apache - htpasswd.c
- index.of.private (algo privado)
- intitle:index.of master.passwd
- inurl:passlist.txt (para encontrar listas de passwords)
- intitle:"Index of..etc” passwd
- intitle:admin intitle:login
- “Incorrect syntax near” (SQL script error)
- intitle:"the page cannot be found” inetmgr (debilidad en IIS4)
- intitle:index.of ws_ftp.ini
“Supplied arguments is not a valid PostgreSQL result” (possible debilidad SQL)

_vti_pvt password intitle:index.of (Frontpage)
inurl:backup intitle:index.of inurl:admin

“Index of /backup”
index.of.password
index.of.winnt
inurl:"auth_user_file.txt"

“Index of /admin”
“Index of /password”
“Index of /mail”
“Index of /” +passwd
Index of /” +.htaccess
Index of ftp +.mdb allinurl:/cgi-bin/ +mailto
Lab-search (cont.)

- allintitle: “index of/admin”
- allintitle: “index of/root”
- allintitle: sensitive filetype:doc
- allintitle: restricted filetype:mail
- allintitle: restricted filetype:doc site:gov
- administrator.pwd.index
- authors.pwd.index
- service.pwd.index
- filetype:config web
- gobal.asax index
- inurl:passwd filetype:txt
- inurl:admin filetype:db
- inurl:iisadmin
- inurl:"auth_user_file.txt"
Lab-search (cont.)

- inurl:"wwwroot/*.”
- allinurl: winnt/system32/ (get cmd.exe)
- allinurl:/bash_history
- intitle:"Index of" .sh_history
- intitle:"Index of" .bash_history
- intitle:"Index of" passwd
- intitle:"Index of" people.1st
- intitle:"Index of" pwd.db
- intitle:"Index of" etc/shadow
- intitle:"Index of" spwd
- intitle:"Index of" master.passwd
- intitle:"Index of" htpasswd
- intitle:"Index of" members OR accounts
- intitle:"Index of" user_carts OR user _cart
try... then tell me what happen

- "# -FrontPage-" inurl:service.pwd
- site:tw inurl:login.asp
- intitle:"Index of" passwords modified
- allinurl:auth_user_file.txt
- "Index of /backup"
- "parent directory" MP3 -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
try… then tell me what happen

- "parent directory" **DVDRip** -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
- "parent directory" **album** -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
- "Windows Vista" 94FBR
- "index of" intext:fckeditor inurl:fckeditor
- allinurl: winnt/system32/
- "VNC Desktop" inurl:5800
Q & A

蔡一郎 Steven Tsai
yilang@nchc.narl.org.tw
06-5050940-749