OWASP 香港 HONG KONG
The Open Web Application Security Project

Google HaCkinG

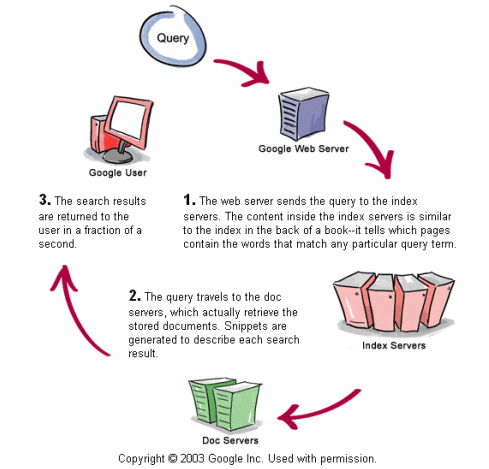
Anthony L@I, CISSP, CISM, CISA
M@rco Leung, SCJP, OCP

0xD8.0x30.0x3.0x12

Google HaCkinG What is that?

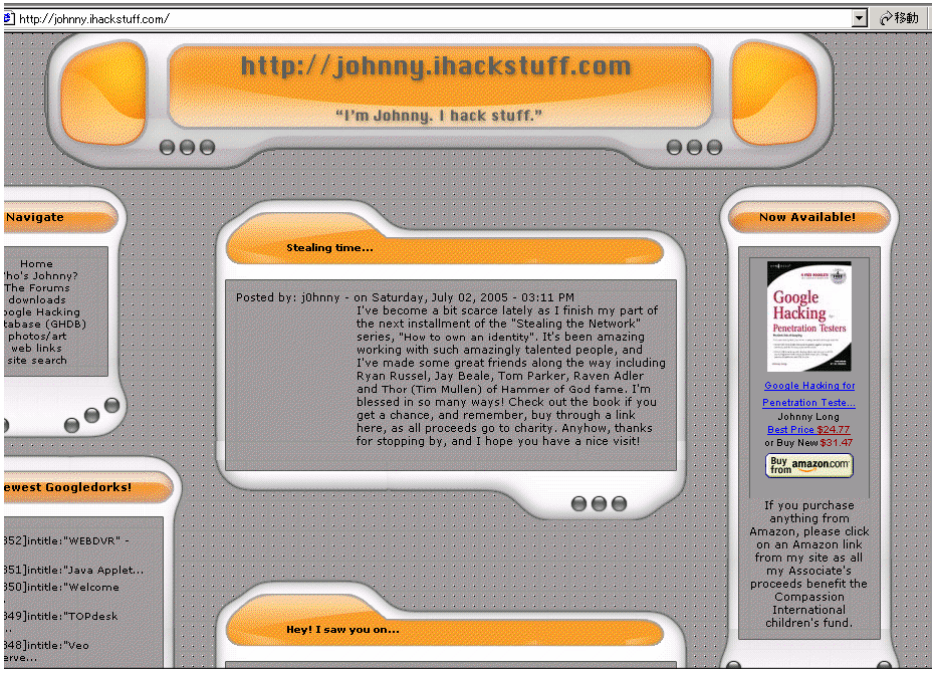
- Your developed web application is on the web. Meanwhile, your ignored vulnerabilities and setting are also there ☺
- Google is a great tools for searching like a well-trained anti-drugs dog on the web.
- Reminder: We are not liable for your intrusive and malicious action and intention.

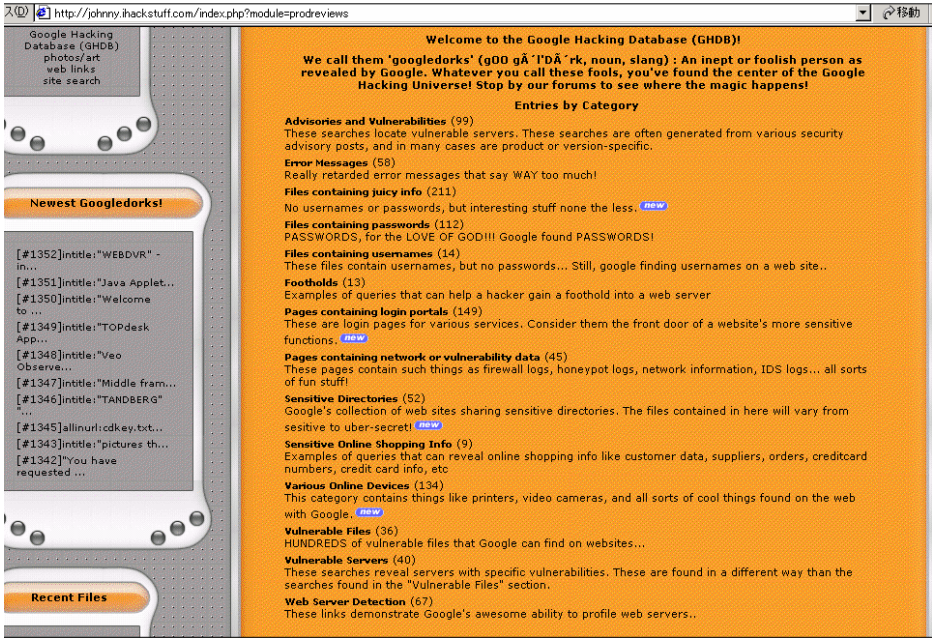
How Google works?



Google Search Engine

- Advanced Search Operators
  - site:
    - restrict search to a specific web site
  - filetype:
    - search within the content of particular file type
  - cache:
    - displays the version of a web page
  - link:
    - search word within hyperlinks
  - intext:, allintext:
    - searches words within a specific website, but ignores the URLs and page titles
  - intitle:, allintitle:
    - search words within the title of a document
  - inurl:, allinurl:
    - search words within the URL of a document





Different Categories of Queries

- Programmer/Developer Bad Practice(s)
- Operating System Vulnerability and Information Exposure (Configuration file, default admin access path)
- Application server security
- Database-related information exposure (Setup queries, database maintenance etc)
- Others

Example (1)

- "A syntax error has occurred" filetype:ihtml intext:LOGIN -"[removed]"

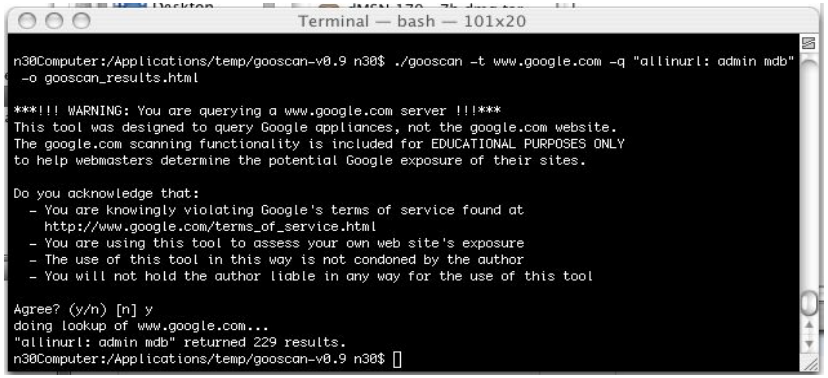
Example (2)

- filetype:ora tnsnames

Tools

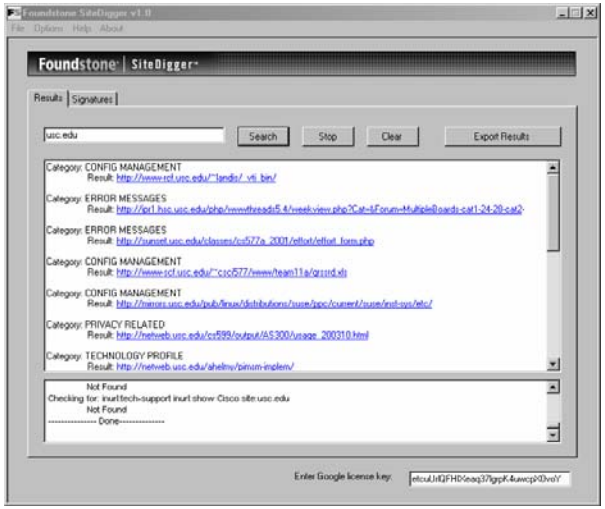
- Using Web interface
  - Athena
  - GooScan
- Using Web Service API
  - SiteDigger

Automated Tools - GooScan



Tools - Athena

Tools - SiteDigger



Tools - SiteDigger

- Version 2 features (tentative release 15th July)
  - Proxy support / Google appliance support
  - XML signatures in OASIS WAS format
  - Adding signatures for OWASP top 10
  - Signature contribution option
  - Raw search tab
  - Configurable # of results

Protection from Google Hackers (1)

- Keep your sensitive data off the web
  - Don't put any secure information on a web site
- Consider removing your site from Google's index
  - http://www.google.com/remove.html
- Use Meta tags
  - Prevent all robots from indexing a page
    - <META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">
  - Prevent only Google robots from indexing a page
    - <META NAME="GOOGLEBOT" CONTENT="NOINDEX, NOFOLLOW">
  - Allow robots to index the page, but don't follow the outgoing links
    - <META NAME="ROBOTS" CONTENT="NOFOLLOW">

Protection from Google Hackers (2)

- Use a robots.txt file
  - http://www.yoursite.com/robots.txt
  - Document to indicate which parts of the server should not be accessed
  - Remove entire website
    - place the following robots.txt file:

```
User-agent: *
Disallow: /
```

  - Remove part of web site
    - Remove all pages under a particular directory

```
User-agent: Googlebot
Disallow: /lemurs
```

    - Remove all files of specific file type

```
User-agent: Googlebot
Disallow: /*.gif$
```

Countermeasures

- Keep sensitive data off the web!!
- Perform periodic Google Assessments
  - Update robots.txt
  - Use meta-tags: NOARCHIVE
  - http://www.google.com/remove.html.

SUMMARY

- How Google works
- Information disclosure with Google
- Tools
- Countermeasures