GOOGLE HACKING

Date: 10/31/2007

Krishna Prasad Yerrapragada
AGENDA

- Introduction
- What is Google Hacking/GHDB?
- GHDB – Johnny Long
- How it works?
- Possible Reasons
- Approaches to AVOID/RESOLVE
- Google’s Response (GHH)
- SPI Labs Solution
What is Hacking?

The act of gaining unauthorized access to computer systems for the purpose of stealing and corrupting data.


Types Of Hackers:
- Black Hats - Malicious hackers
- White Hats - Ethical hackers
- Grey Hats - Ambiguous

Search Engines
- Efficient (Google – most effective)
- Around 12 Billion Pages
- Starting point of many hacking activities. .. Can you believe it?
- Infact, One of the most interesting uses of Google
Google Hacking

**Definition**: Google hacking is a term that refers to the art of creating complex search engine queries in order to filter through large amounts of search results for information related to computer security.

- www.Wikipedia.com

**The whole Idea!!**

Web pages are:
- Crawled/Indexed (typically, once 2 weeks)
- Cached
- Hackers query this information (Reconnaissance)
- “inurl” and “allintitle”

- Once Indexed Its cached
  - a) Contact Google (http://www.google.com/remove.html)
  - b) Contact Other Search engines
- Google performs the dirty work (password embedded urls)
VULNERABILITIES

Almost 70% of Websites have vulnerabilities

- Known Vulnerabilities
  - Informally communicated
  - Chain emails

- Information Disclosure Vulnerabilities
  - Passwords
  - Administrative files
  - Sensitive customer information
  - Military information (Submarines, docking stations of Navy Ships)
  - System email id lists
  - Medical records
  - Bank account numbers

Crawlers - Just index/Cache what ever they find.
GHDB (Google Hack Database)

  (65K-70K members)

GHDB – A database containing Hacking queries

From the Google Hacking Database:
- Error messages that contain too much information.
- Password Files and Sensitive directories
- Pages containing logon portals.
- Pages containing network or vulnerability data such as firewall logs.
PRIMARY REASONS

- People Negligence – Called GoogleDorks
- Increase in number of Remote administrative tools
- Security holes in the Networks
- Poor site configuration
  
  e.g. Securing admin panel - .htaccess procedure
       (password protection on HTML documents)
Probable Solutions: Avoid/Resolve ??

- **Google’s Reaction**
  - Johnny’s Opinion (Google Security Alerts System)
  - GHH (Google Hack Honeypot, powered by the Google & GHDB)
  - Google Dork Detection (blocking evil queries)

**Google Hack Honeypot (GHH)** - reconnaissance against attackers

An information system resource whose value lies in unauthorized or illicit use of that resource

HONEYNET ARCHITECTURE

• An Architecture, not Product/Software
• High-interaction honeypot
• Populate it with live systems
• Every Packet entering or leaving is monitored, captured, analyzed

http://www.honeynet.org/papers/honeynet/
**Use of “robots.txt”**

- Instructs the Crawler not to crawl certain sections

**Syntax:**

```
User-Agent: [Spider or Bot name]
Disallow: [Directory or File Name]
```

- e.g.1: (This stops Google from viewing the directory)
  ```
  User-Agent: Googlebot
  Disallow: /private/privatefile.htm
  ```

- e.g.2: (All search engines are stopped.)
  ```
  User-agent: *
  Disallow: /cgi-bin/
  Disallow: /_borders/
  Disallow: /_derived/
  ```

- e.g.3: NO search engine can view anything on your site..
  ```
  User-agent: *
  Disallow: /
  ```

**Issues:**

- Not all Crawlers are Standards based
- Single Point of Risk
Automatic Scanners:
Web Vulnerability Scanners: Scan the website and point out potential security issues.
- Need to be configured properly.
- Not 100% efficient.
Examples: Nikto, Paros Proxy, WebScarab, WebInspect

SPIDYNAMICS (Web Inspect):
Approach – “Do your Own Crawl”

- Pick a scanning tool (possibly executing Java Script/Submit Forms)
- Appropriately configure the tool and kick it off
- Sort the results
  - Use a scanner to run queries
  - Scan the “SiteTree” (WebInspect displays the SiteTree in an explorer view)
  - Check for “/admin” folders
  - Check for “passwords” kind of files
  - Scan the content of the results
GENERAL RULES to AVOID HACKING

- Prevention - better than Cure

**Best Practices:**
- Security - development stage
- Access Controls
- Maintenance: Run Scanners
- Use “Robots.txt” carefully
- Change default error messages.
- Password Protection to critical data
- Password Encryption
References:

  url: http://searchsecurity.techtarget.com/sDefinition/0,,sid14_qci1151189,00.html

  url: http://www.castlecops.com/article-6466-nested-0-0.html


- SPI Labs, “Preventing Google Hacking, Steps to protect your Web Application”, Retrieved on 10/29/07

- Hendrick Speck and Philipp Thiele, European Graduate School, “Playing the Search Engine”, Retrieved on 10/29/07

  url: http://sectools.org/web-scanners.html - Vulnerability Scanners

- SourceForge.net, “What is GHH ?”, retrieved on 10/30/2007
  url: http://ghh.sourceforge.net/

  url: http://www.honeynet.org/papers/honeynet/

  url: http://www.outfront.net/tutorials_02/adv_tech/robots.htm

- www.Wikipedia.com
QUESTIONS ?
Thank You!