GOOGLE CODE SEARCH

The Pitfalls of Copy/Paste
About the Presenter

- Tony Flick
- Optimist Developer->Tester->Pessimist
  ->Presenter on a Soap Box
What is Google Code Search

- http://www.google.com/codesearch
- “Google Code Search helps you find function definitions and sample code by giving you one place to search publicly accessible source code hosted on the Internet.”
Google Code Search vs. Normal Google Search Engine (Part I)

- More efficient
  - Regular expressions – allows complex string searches
    - supports POSIX (Portable Operating System Interface for Unix) extended regular expression syntax
Crash Course in Regular Expressions

- A regular expression is a string that utilizes certain syntax rules to describe or match a set of strings.
- Generally used to search and manipulate bodies of text based on certain patterns.
Regular Expressions II

- Regular Expressions Used in This Presentation
  - . Matches any single character.
  - [ ] Matches a single character that is contained within the brackets.
  - + Matches one or more of the preceding expression
  - ^ Matches only at the start of the line
  - \w Matches a word character [a-zA-Z_0-9]
  - \s Matches any white-space character [ \t\n\x0B\f\r]
  - \* Matches the * character [*]
  - "" Matches a double quote ["]
  - \( Matches the begin parenthesis [(]
  - \) Matches the end parenthesis [)]
  - | Logical OR
Google Code Search vs. Normal Google Search Engine (Part II)

- More efficient
  - Filter searches based on the following:
    - Programming language (e.g. lang:^(c|c#|c\+\+)$)
    - License type (e.g. license:bsd | gpl | mit)
    - Packages (e.g. package:"www.kernel.org")
    - Files (e.g. -file:\.cc$)
Nefarious Doings or Improving Security?

- Google Code Search can be used to find vulnerabilities and sensitive information.
  - What are you going to do with this information?
Vulnerabilities

- Buffer Overflows
  - A Buffer Overflow is a programming error which may result in a memory access exception and program termination, or in the event of the user being malicious, a breach of system security.

- SQL Injection
  - When an attacker can cause malicious SQL code to run by maliciously modifying data used to compose a SQL command.

- Cross-Site Scripting (XSS)
  - Cross-site scripting (XSS) is a type of computer security vulnerability typically found in web applications which allow malicious web users to inject HTML or client-side scripts into the web pages viewed by other users.

- Common Programming Errors
- Information Disclosure
Buffer Overflows

- Historically vulnerable functions
  - `strcpy - strcpy(lang:c)`
    - 2006 - 55,000 results
    - 2008 – 77,000 results
  - `sprintf - (sprintf(lang:c))`
    - 2006 - 4,000 results
    - 2008 – 8,000 results
  - `scanf - (scanf(lang:c))`
    - 2006 - 4,000 results
    - 2008 – 7,000 results
strcpy - strcpy\(\(\text{\textbackslash w+},\text{\textbackslash w+}\)\) lang:c

```c
106:   buf[0] = '\0';  /* start with empty string */
    strcpy(buf, str);
    sprintf(tmp, "%04x - ",$, i * dump_width);
```

www.openssl.org/source/openssl-0.9.6l.tar.gz - BSD - C
Historically vulnerable functions

- `(request.form | request.querystring)` lang:asp
  - 2006 - 10,000 results
  - 2008 – 36,000 results

- `(executequery) (request.getParameter)` lang:java
  - 2006 - 300 results
  - 2008 – 3,000 results
SQL Injection

- Example:
  - (executequery) (request.getParameter) lang:java

```java
76:   DbEngine db = DbEngine.borrowEngine(JEROMEDL_DOC);
    ResultSet resources = db.executeQuery(request.getParameter("xpath");
    ResourceIterator it = resources.getIterator();
```

[webapp/WEB-INF/src/org/jeromedl/servlets/ListResources.java]

[www.jeromedl.org/.../JeromeWebapp.tar.gz - Unknown License - Java]
Historically vulnerable functions

- (response.write) (request.form) lang:asp
  - 2006 - 4,000 results
  - 2008 – 11,900 results
- echo\s\$_GET lang:php
  - 2006 - 1,000 results
  - 2008 – 5,000 results
- echo\s\$_POST lang:php
  - 2006 - 1,000 results
  - 2008 – 4,000 results
XSS

- Example
  - echo\s\$_GET lang:php

```php
9: <?php
  echo \$_GET['id'].<br />'."\n";
  echo \$_GET['action'];
?
```
Common Programming Errors

- "<= 65553"
  - How many ports are there?
  - $2^{16} - 1 = 65535$ (Highest Port)
  - 2006 - 100 results
  - 2008 – 2 results

```java
210: int value = Integer.parseInt(_portText.getText());
    ok = ( (value > 0 ) && (value <= 65553) );
  } catch ( Exception e ) {
```

directory.fedora.redhat.com/.../fedora-directoryconsole-1.0.tar.gz - Unknown License - Java
Information Disclosure

- Notes inside comments
  - Too Much Information
Notes Inside Comments

- "// password =""
Private Code Released
  - confidential proprietary

103: *

* The information contained in this file is confidential and proprietary to Conexant Systems Inc.
* No part of this file may be reproduced or distributed, in any form
Notes Inside Comments

- “// FIXME security” – 29 results
- “// TODO security” – 43 results
- “// password = “ – 13 results
- backdoor password – 5,000 results
- “I think this code is broken” – 104 results
- “This is very insecure” – 6 results
- “This sucks” – 13,300 results
- “This really sucks” – 252 results
Client Specific Search

(7,000 results)

Client Specific Search

Copyright (c) 2002 Microsoft Corporation. All rights reserved.

download.microsoft.com/.../sscli_20021101.tgz - Unknown License - Perl
If your code is publicly accessible, Google will find it and people will try to compromise it.
Advice

- Use Secure Programming Practices
  - Avoid historically vulnerable functions and libraries
- Perform static source code analysis
Questions

- Questions Anyone?