Ethical Hacking and Countermeasures
Version 6

Module IV

Google Hacking

Module Objective

This module will familiarize you with:

- What is Google Hacking
- What a Hacker Can Do With Vulnerable Site
- Google Hacking Basics
- Google Advanced Operators
- Pre-Assessment
- Locating Exploits and Finding Targets
- Tracking Down Web Servers, Login Portals, and Network Hardware
- Google Hacking Tools

C|EH

TM

Certified Ethical Hacker

Google Hacking

→

Pre-Assessment

↓

What a Hacker Can Do
With Vulnerable Site

↓

Locating Exploits and Finding Targets

↓

Google Hacking Basics

↓

Tracking Down Web Servers,
Login Portals, and Network Hardware

↓

Google Advanced Operators

↓

Google Hacking Tools

CEH
Certified | Ethical | Hacker
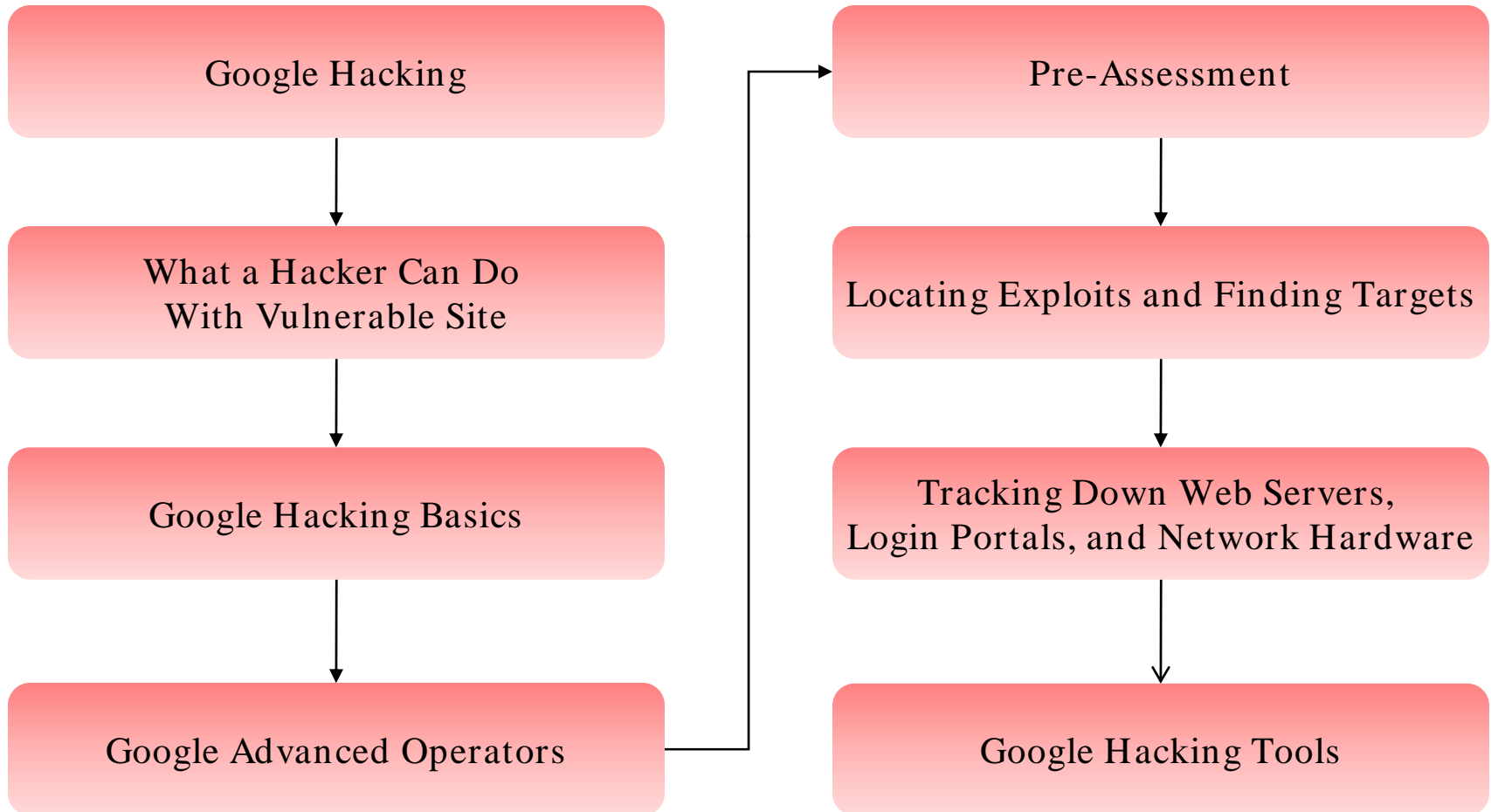
Google hacking is a term that refers to the art of creating complex search engine queries in order to filter through large amounts of search results for information related to computer security

In its malicious format, it can be used to detect websites that are vulnerable to numerous exploits and vulnerabilities as well as locate private, sensitive information about others, such as credit card numbers, social security numbers, and passwords

Google Hacking involves using Google operators to locate specific strings of text within search results

What a Hacker Can Do With Vulnerable Site

Information that the Google Hacking Database identifies:

- Advisories and server vulnerabilities

- Error messages that contain too much information

- Files containing passwords

- Sensitive directories

- Pages containing logon portals

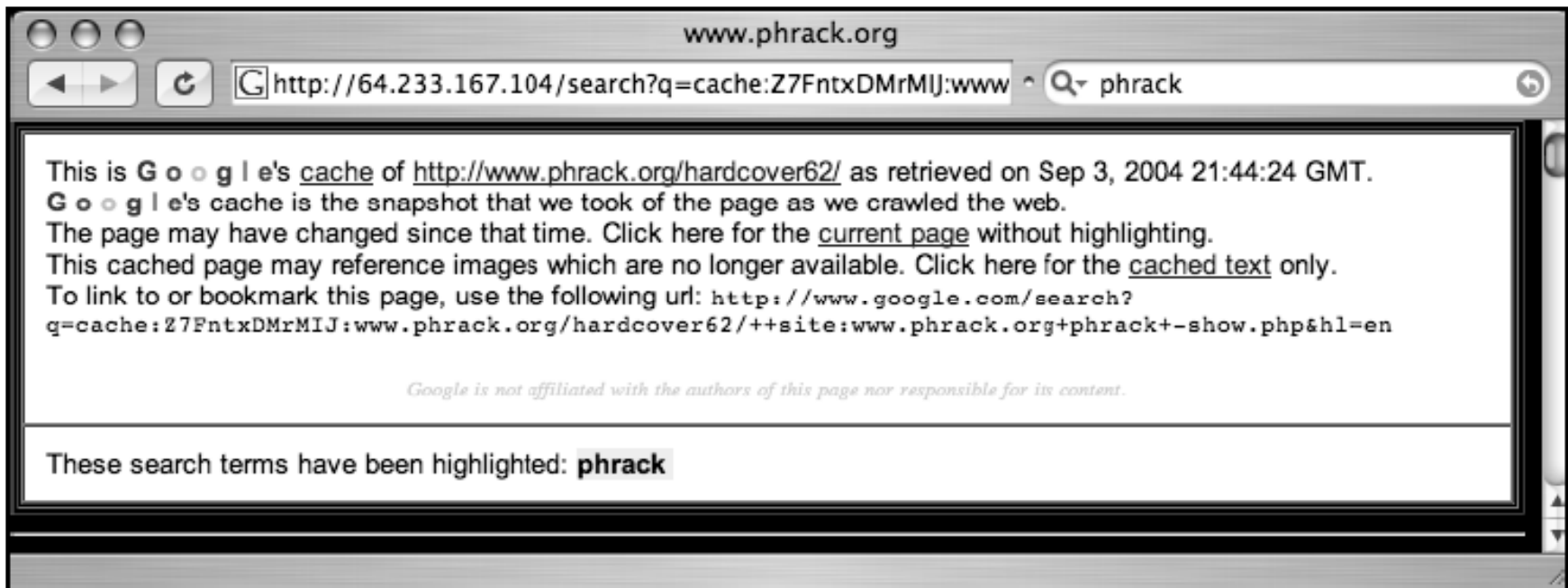- Pages containing network or vulnerability data such as firewall logs

Google Hacking Basics

Hackers can get a copy sensitive data even if plug on that pesky Web server is pulled off and they can crawl into entire website without even sending a single packet to server

If the web server does not get so much as a packet, it can not write any thing to log files

www.phrack.org

http://64.233.167.104/search?q=cache:Z7FntxDMrMIJ:www   phrack

This is Google's cache of http://www.phrack.org/hardcover62/ as retrieved on Sep 3, 2004 21:44:24 GMT.
Google's cache is the snapshot that we took of the page as we crawled the web.
The page may have changed since that time. Click here for the current page without highlighting.
This cached page may reference images which are no longer available. Click here for the cached text only.
To link to or bookmark this page, use the following url: http://www.google.com/search?
q=cache:Z7FntxDMrMIJ:www.phrack.org/hardcover62/++site:www.phrack.org+phrack+-show.php&hl=en

Google is not affiliated with the authors of this page nor responsible for its content.

These search terms have been highlighted: **phrack**

EC-Council

Google some times works as a proxy server which requires a Google translated URL and some minor URL modification

Translation URL is generated through Google's translation service, located at www.google.com/translate_t

If URL is entered in to "Translate a web page" field, by selecting a language pair and clicking on Translate button, Google will translate contents of Web page and generate a translation URL

EC-Council

Directory Listings

A directory listing is a type of Web page that lists files and directories that exist on a Web server

It is designed such that it is to be navigated by clicking directory links, directory listings typically have a title that describes the current directory, a list of files and directories that can be clicked

Just like an FTP server, directory listings offer a no-frills, easy-install solution for granting access to files that can be stored in categorized folders

Problems faced by directory listings are:

- They do not prevent users from downloading certain files or accessing certain directories hence they are not secure
- They can display information that helps an attacker learn specific technical details about Web server
- They do not discriminate between files that are meant to be public and those that are meant to remain behind the scenes
- They are often displayed accidentally, since many Web servers display a directory listing if a top-level index file is missing or invalid

CEH
Certified | Ethical | Hacker ™

Index of /security/dist

http://216.239.39.104/search?q=cache:B   Google

Index of /security/dist

Name	Last modified	Size	Description
Parent Directory		-	
KEYS	26-Mar-2003 03:34	4.3K	
c-library/	24-Mar-2004 02:46	-	
java-library/	17-Apr-2004 18:36	-	

Apache/2.0.52 (Unix) Server at xml.apache.org Port 80

C|EH — Certified Ethical Hacker

Since directory listings offer parent directory links and allow browsing through files and folders, attacker can find sensitive data simply by locating listings and browsing through them

Locating directory listings with Google is fairly straightforward as they begin with phrase "Index of," which shows in tittle

An obvious query to find this type of page might be *ntitle:index.of*, which can find pages with the term "index of" in the title of the document

intitle:index.of "parent directory" or *intitle:index.of "name size"* queries indeed provide directory listings by not only focusing on index.of in title but on keywords often found inside directory listings, such as parent directory, name, and size

EC-Council

Finding Specific Directories

This is easily accomplished by adding the name of the directory to the search query

To locate "admin" directories that are accessible from directory listings, queries such as *intitle:index.of.admin* or *intitle:index.of inurl:admin* will work well, as shown in the following figure

Google Search: intitle:index.of inurl:admin

http://www.google.com/search?hl intitle:index.of inurl:admin

Web Images Groups News Froogle **more »**

Google™

intitle:index.of inurl:admin   Search   Adva Prefe

Web   Results **1 - 10** of about **22,000** for **intitle:index.of inurl:admin**. (0.39 seconds)

Index of /admin/hurricane-plan
Index of /**admin**/hurricane-plan. Name Last modified Size Description Parent Directory - Disaster-Phone.pdf 01-Sep-2004 09:47 18K disaster ...
www.health.ufl.edu/admin/hurricane-plan/ - 3k - Cached - Similar pages

Index of /admin/dss/
Index of /**admin**/dss/. Name Last modified Size Description [DIR] Parent Directory
www.state.ak.us/admin/dss/ - 1k - Cached - Similar pages

Index of /admin/alumni/sharks/cim97
Index of /**admin**/alumni/sharks/cim97. Name Last modified Size Description Parent Directory - cim 25-Oct-1996 11:50 2.7K cimlinx.txt ...
www.brunel.ac.uk/admin/alumni/sharks/cim97/ - 3k - Cached - Similar pages

EC-Council

As the directory listing is in tree style, it is also possible to find specific files in a directory listing

To find WS_FTP log files, try a search such as $intitle:index.of\ ws\_ftp.log$, as shown in the Figure below:

Server Versioning

The information an attacker can use to determine the best method for attacking a Web server is the exact software version

An attacker can retrieve that information by connecting directly to the Web port of that server and issuing a request for the HTTP headers

Some typical directory listings provide the name of the server software as well as the version number at the bottom portion. These information are faked and attack can be done on web server

intitle:index.of "server at" query will locate all directory listings on the Web with *index of in the title and server at anywhere in the text of the page*

In addition to identifying the Web server version, it is also possible to determine the operating system of the server as well as modules and other software that is installed

Server versioning technique can be extended by including more details in the query

EC-Council

Server Versioning (cont'd)

Index of /

http://www.3do.com/   intitle:index.of "server at"

e3-euro2/	05-Dec-2003 19:52	–
e3-eurox/	05-Dec-2003 19:53	–
jacked/	05-Dec-2003 03:22	–
jumpgate/	04-Dec-2003 00:14	–
support/	05-Dec-2003 02:05	–

Apache/1.3.27 Server at www.3do.com Port 80
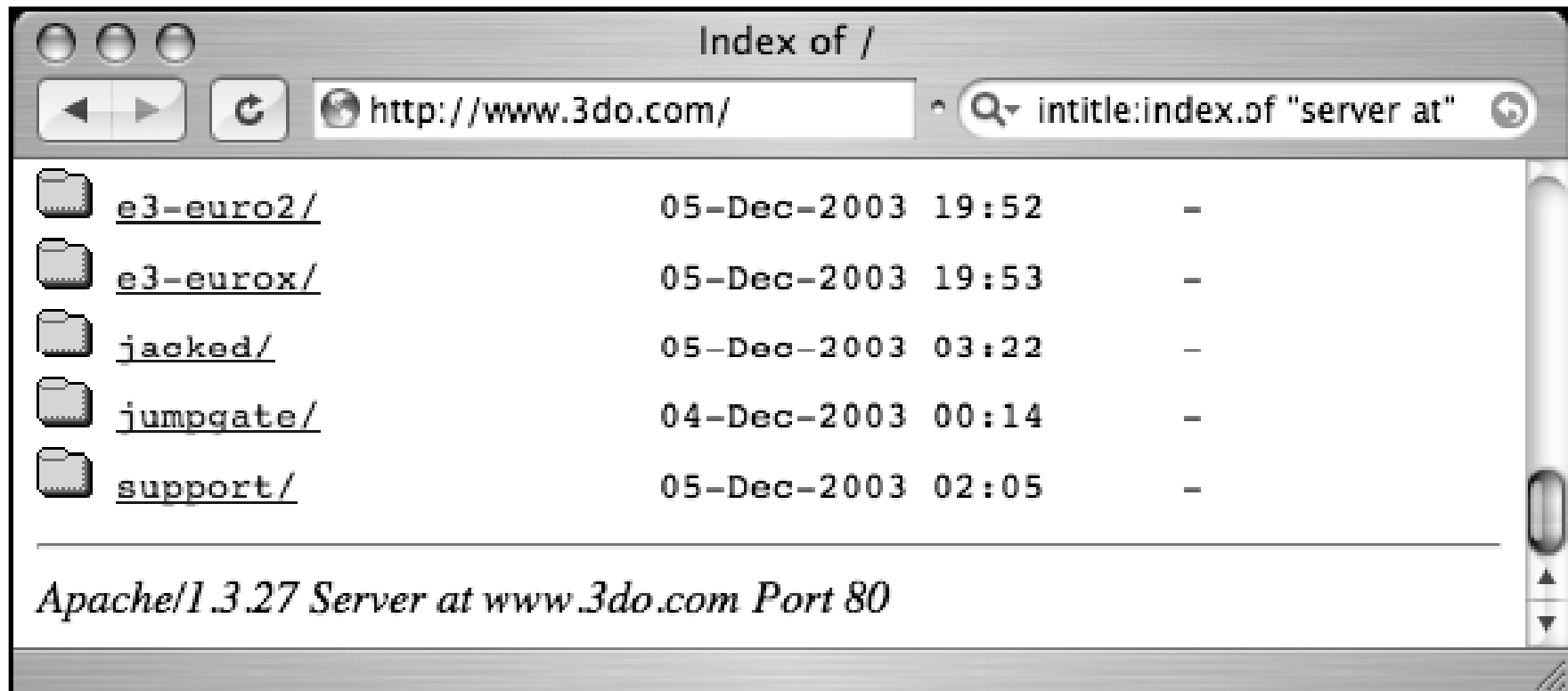
Going Out on a Limb: Traversal Techniques

Attackers use traversal techniques to expand a small foothold into a larger compromise

The query *intitle:index.of inurl:"/admin/ *"* is helped to traversal as shown in the figure:
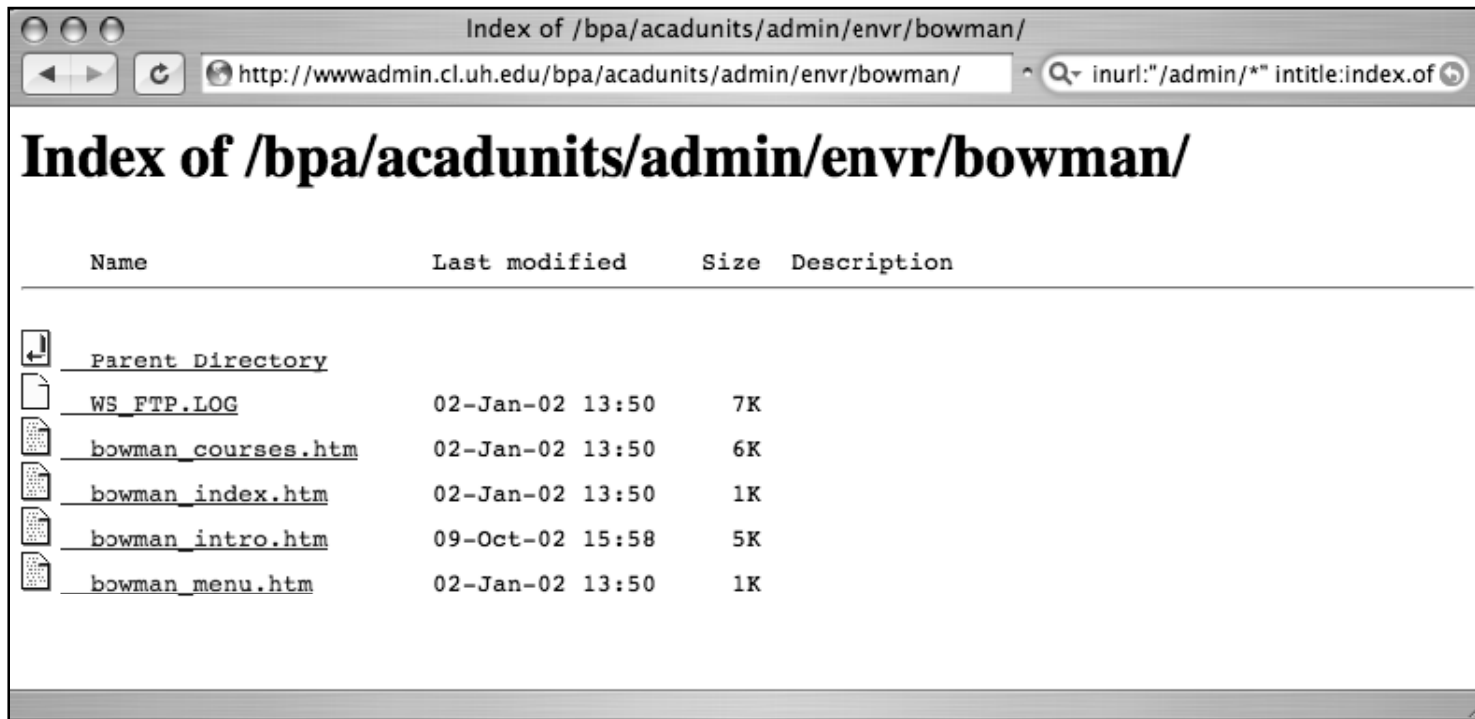
Directory Traversal

By clicking on the parent directory link the sub links under it will open. This is basic directory traversal

Regardless of walking through the directory tree , traversing outside the Google search wandering around on the target Web server is also be done

The word in the URL will be changed with other words

Poorly coded third-party software product installed in the server accepts directory names as arguments which allows users to view files above the web server directory

Automated tools can do a much better job of locating files and vulnerabilities

EC-Council

Incremental Substitution

This technique involves replacing numbers in a URL in an attempt to find directories or files that are hidden, or unlinked from other pages

By changing the numbers in the file names, the other files can be found

In some examples, substitution is used to modify the numbers in the URL to locate other files or directories that exist on the site

- /docs/bulletin/2.xls could be modified to /docs/bulletin/2.xls
- /DigLib_thumbnail/spmg/hel/0001/H/ could be changed to /DigLib_thumbnail/spmg/hel/0002/H/
- /gallery/wel008-1.jpg could be modified to /gallery/wel008-2.jpg

Extension Walking

File extensions and how filetype operator can be used to locate files with specific file extensions

HTM files can be easily searched with a query such as filetype:HTM HTM

Filetype searches require a search parameter and files ending in HTM always have HTM in the URL

After locating HTM files, substitution technique is used to find files with the same file name and different extension

Easiest way to determine names of backup files on a server is to locate a directory listing using *intitle:index.of* or to search for specific files with queries such as *intitle:index.of index.php.bak* or *inurl:index.php.bak*

If a system administrator or Web authoring program creates backup files with a .BAK extension in one directory, there is a good chance that BAK files will exist in other directories as well
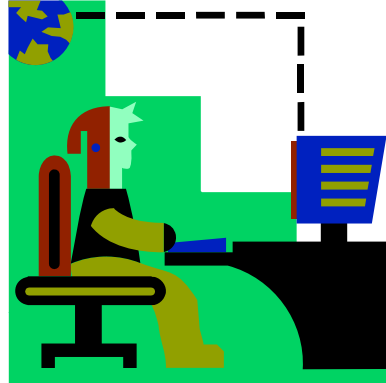
Google Advanced Operators

Site Operator

The site operator is absolutely invaluable during the information-gathering phase of an assessment

Site search can be used to gather information about the servers and hosts that a target hosts

Using simple reduction techniques, you can quickly get an idea about a target's online presence

Consider the simple example of site:washingtonpost.com – site:www.washingtonpost.com

This query effectively locates pages on the washingtonpost.com domain other than www.washingtonpost.com

Site Operator (cont'd)

EC-Council
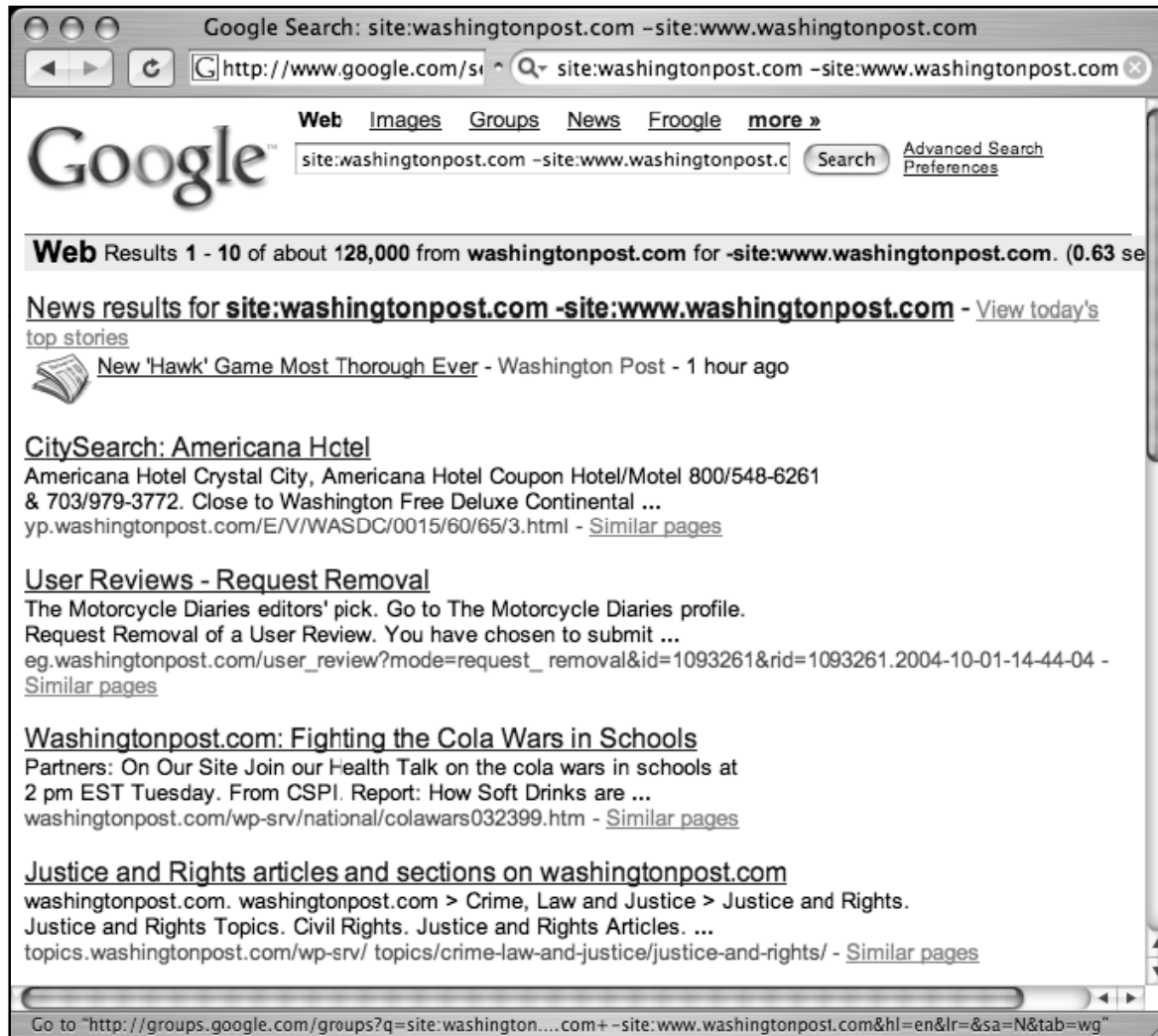
intitle:index.of

intitle:index.of is the universal search for directory listings

In most cases, this search applies only to Apache-based servers, but due to the overwhelming number of Apache-derived Web servers on the Internet, there is a good chance that the server you are profiling will be Apache-based

Error messages can reveal a great deal of information about a target

Often overlooked, error messages can provide insight into the application or operating system software a target is running, the architecture of the network the target is on, information about users on the system, and much more

Not only are error messages informative, they are prolific

A query of intitle: error results in over 55 million results

login | logon

Login portals can reveal the software and operating system of a target, and in many cases "self-help" documentation is linked from the main page of a login portal

These documents are designed to assist users who run into problems during the login process
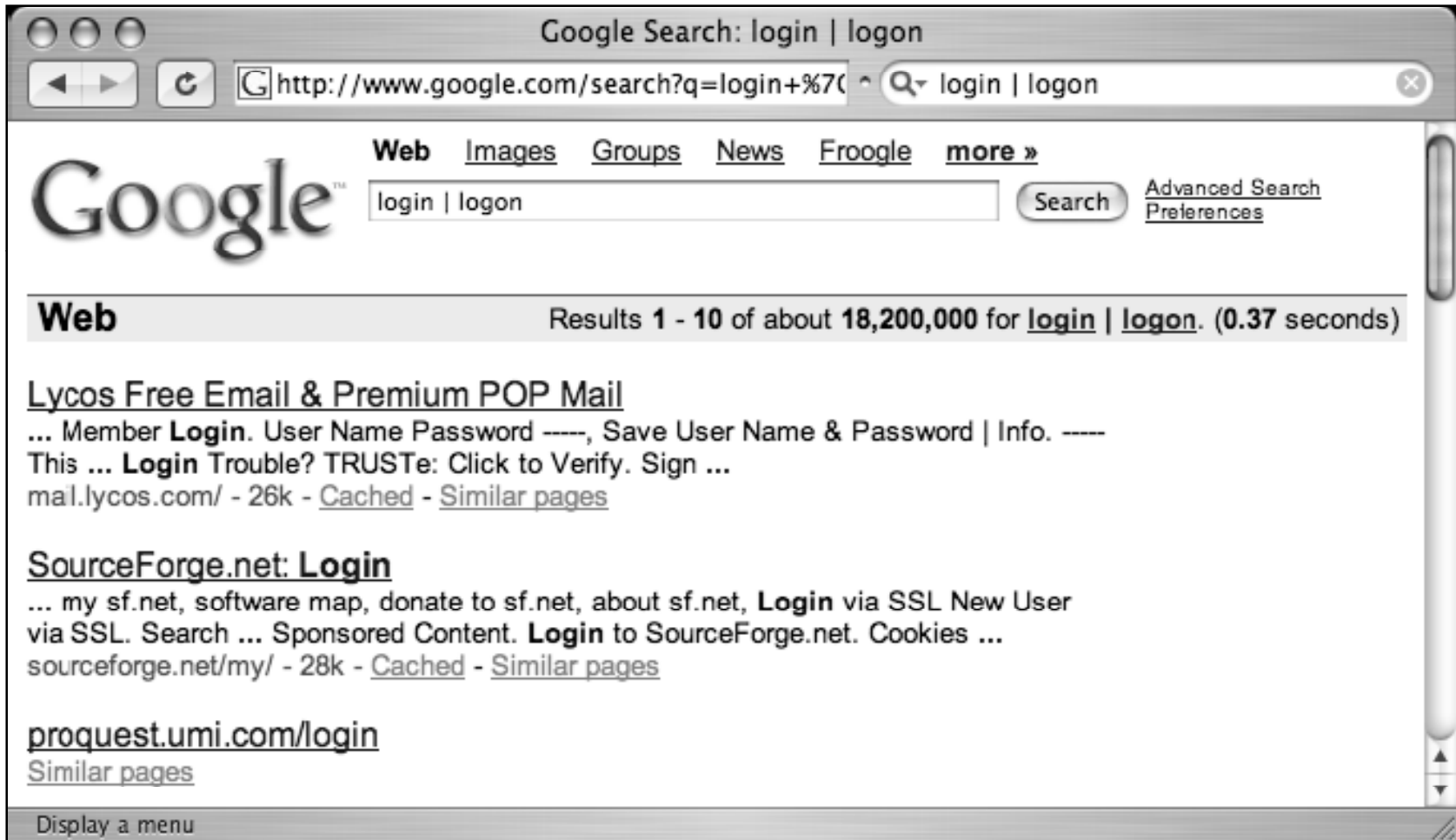
Whether the user has forgotten his or her password or even username, this document can provide clues that might help an attacker

Documentation linked from login portals lists e-mail addresses, phone numbers, or URLs of human assistants who can help a troubled user regain lost access

These assistants, or help desk operators, are perfect targets for a social engineering attack
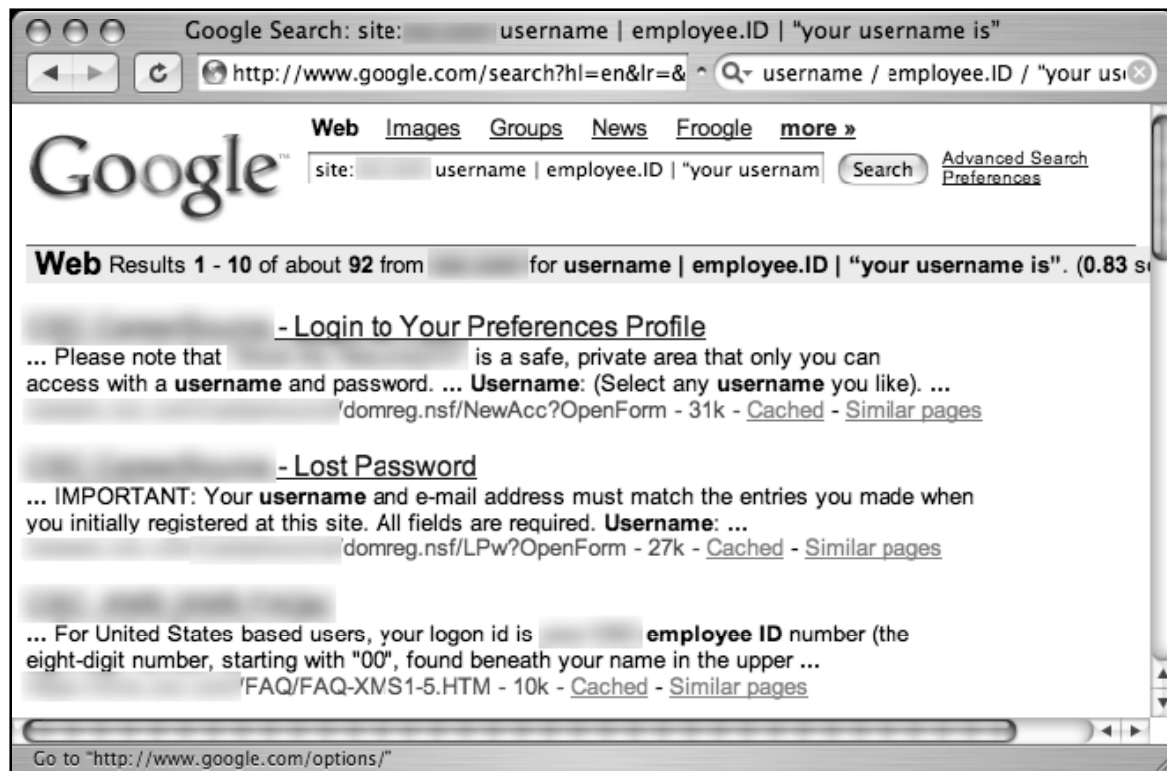
EC-Council

C|EH
Certified Ethical Hacker ™

Google Search: login | logon

http://www.google.com/search?q=login+%7(— login | logon

Google™

Web   Images   Groups   News   Froogle   **more »**

login | logon   [Search]   Advanced Search
Preferences

Web   Results **1 - 10** of about **18,200,000** for <u>login</u> | <u>logon</u>. (0.37 seconds)

Lycos Free Email & Premium POP Mail
... Member **Login**. User Name Password -----, Save User Name & Password | Info. -----
This ... **Login** Trouble? TRUSTe: Click to Verify. Sign ...
mail.lycos.com/ - 26k - Cached - Similar pages

SourceForge.net: Login
... my sf.net, software map, donate to sf.net, about sf.net, **Login** via SSL New User
via SSL. Search ... Sponsored Content. **Login** to SourceForge.net. Cookies ...
sourceforge.net/my/ - 28k - Cached - Similar pages

proquest.umi.com/login
Similar pages

Display a menu

There are many different ways to obtain a username from a target system

Even though a username is the less important half of most authentication mechanisms, it should at least be marginally protected from outsiders

The word password is so common on the Internet, there are over 73 million results for this one-word query

During an assessment, it is very likely that results for this query combined with a site operator will include pages that provide help to users who have forgotten their passwords

In some cases, this query will locate pages that provide policy information about the creation of a password

This type of information can be used in an intelligent-guessing or even a brute-force campaign against a password field

EC-Council

admin | administrator
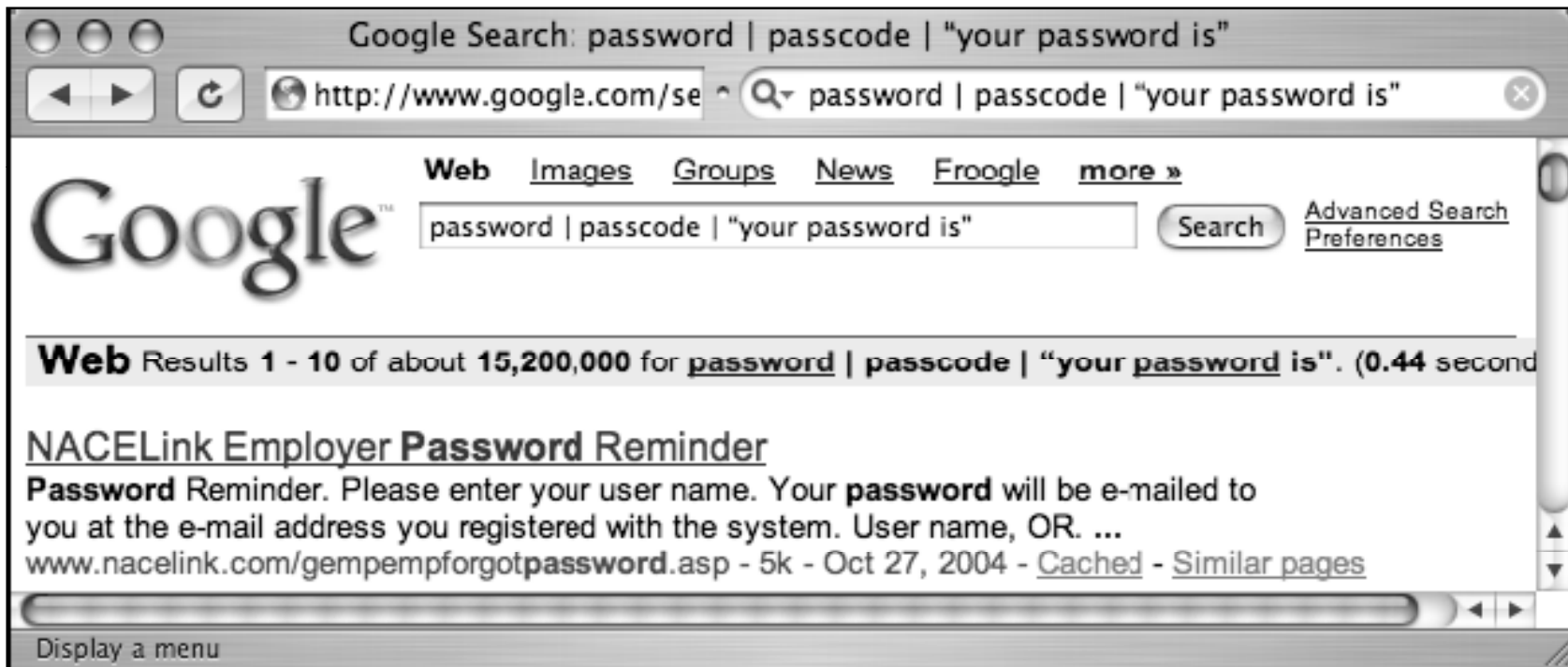
The word administrator is often used to describe the person in control of a network or system

The word administrator can also be used to locate administrative login pages, or login portals

The phrase Contact your system administrator is a fairly common phrase on the Web, as are several basic derivations

A query such as "please contact your * administrator" will return results that reference local, company, site, department, server, system, network, database, e-mail, and even tennis administrators

If a Web user is said to contact an administrator, chances are that the data has at least moderate importance to a security tester

EC-Council

CEH
Certified Ethical Hacker
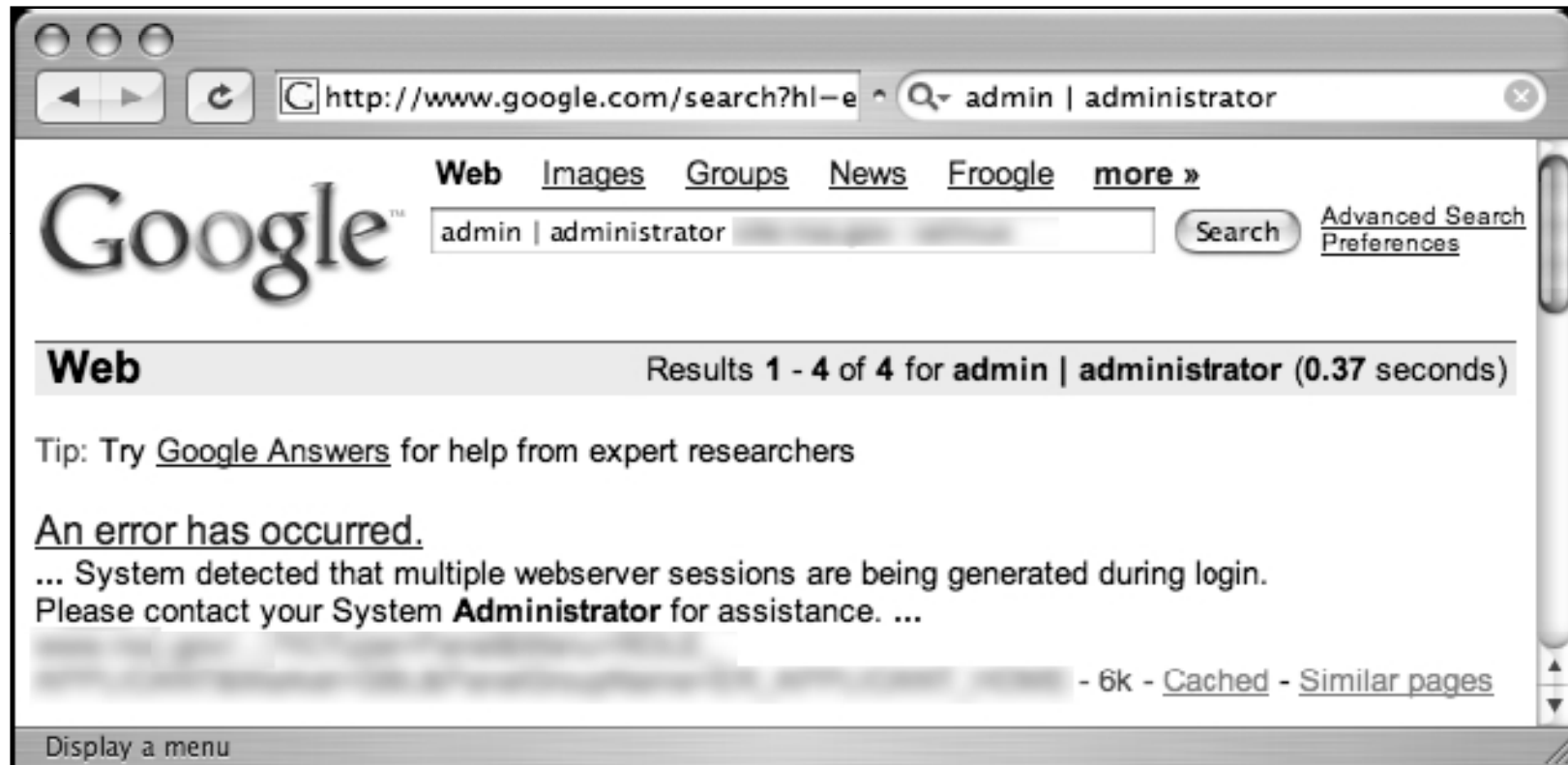
admin login Reveals Administrative Login Pages

The $-ext:html - ext:htm - ext:shtml - ext:asp - ext:php$ query uses ext, a synonym for the filetype operator, and is a negative query

It returns no results when used alone and should be combined with a site operator to work properly

The idea behind this query is to exclude some of the most common Internet file types in an attempt to find files that might be more interesting

Google   -ext:html -ext:htm -ext:shtml -ext:asp -ext:php   Search

inurl:temp | inurl:tmp | inurl:backup | inurl:bak

The *inurl:temp | inurl:tmp | inurl:backup | inurl:bak query, combined with the site* operator, searches for temporary or backup files or directories on a server

Although there are many possible naming conventions for temporary or backup files, this search focuses on the most common terms

Since this search uses the *inurl operator, it will also locate files that contain these terms as file extensions,* such as index.html.bak

Google Suggest LABS

`inurl:temp | inurl:tmp | inurl:backup | inurl:bak`   Search

EC-Council
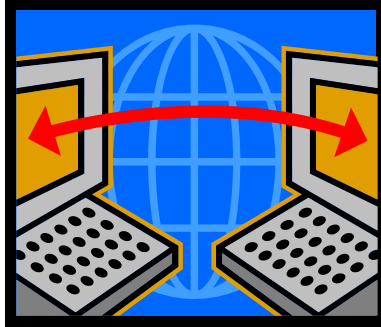
Pre-Assessment

The term intranet, despite more specific technical meanings, has become a generic term that describes a network confined to a small group

In most cases, the term intranet describes a closed or private network unavailable to the general public

Many sites have configured portals that allow access to an intranet from the Internet, bringing this typically closed network one step closer to the potential attackers

Unavailable to public

EC-Council
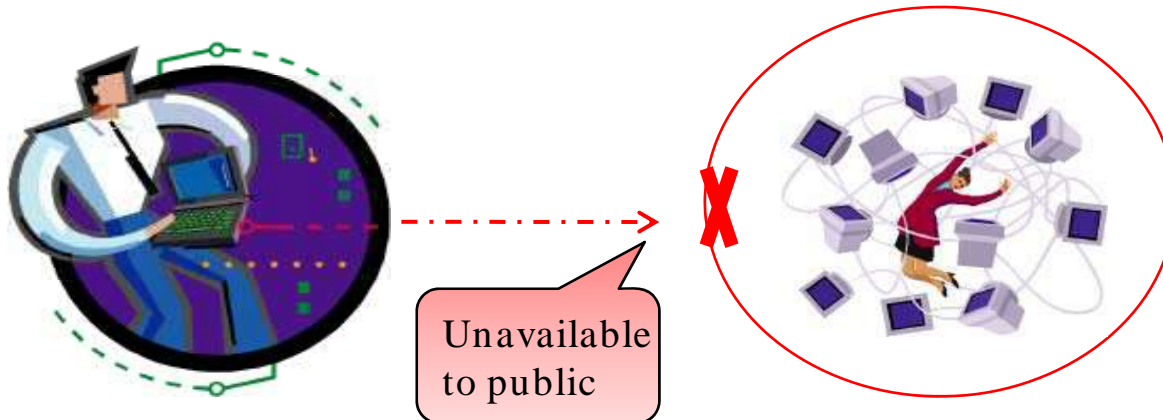
Locating Exploits and Finding Targets

One way to locate exploit code is to focus on the file extension of the source code and then search for specific content within that code

Since source code is the text-based representation of the difficult-to-read machine code, Google is well suited for this task

For example, a large number of exploits are written in C, which generally use source code ending in a .c extension

A query for *filetype:c exploit returns around 5,000 results, most of which* are exactly the types of programs you are looking for

These are the most popular sites hosting C source code containing the word *exploit, the* returned list is a good start for a list of bookmarks

Using page-scraping techniques, you can isolate these sites by running a UNIX command against the dumped Google results page

```
grep Cached exp | awk -F" -" '{print $1}' | sort -u
```

Locating Exploits Via Common Code Strings

Another way to locate exploit code is to focus on common strings within the source code itself

One way to do this is to focus on common inclusions or header file references

For example, many C programs include the standard input/output library functions, which are referenced by an include statement such as #include <stdio.h> within the source code

A query like this would locate C source code that contained the word exploit, regardless of the file's extension:

- "#include <stdio.h>" exploit

C|E|H ™
Certified Ethical Hacker

EC-Council

Language	Extension (Optional)	Sample String
asp.net (C#)	Aspx	"<%@ Page Language="C#"" inherits
asp.net (VB)	Aspx	"<%@ Page Language="vb"" inherits
asp.net (VB)	Aspx	<%@ Page LANGUAGE="JScript"
C	C	"#include <stdio.h>"
C#	Cs	"using System;" class
c++	Cpp	"#include "stdafx.h""
Java	J, JAV	class public static
JavaScript	JS	"<script language="JavaScript">"

Perl	PERL, PL, PM	"#!/usr/bin/perl"
Python	Py	"#!/usr/bin/env"
VBScript	.vbs	"<%@ language="vbscript" %>"
Visual Basic	Vb	"Private Sub"

Locating Vulnerable Targets

Attackers are increasingly using Google to locate Web-based targets vulnerable to specific exploits

In fact, it's not uncommon for public vulnerability announcements to contain Google links to potentially vulnerable targets

Bugtraq: Coppermine Photo Gallery remote compromise

http://seclists.org/lists/bugtraq/2003/Apr/0117.html

Web Photo Gallery
Simple-to-use tools for superior photos. View the free demo today!
www.microsoft.com/imag

PHP Without Knowing PHP
Generate PHP On The Fly From Your Data. It Does All The Work !
PHPPageGenerator.com

---EXPLOIT---
Attached is a working exploit, upload this onto a vulnerable server and execute it like this:
/albums/userpics/Copperminer.jpg.php?[command]
Where command can be something like "id;uname%20-a" or "cat%20/etc/passwd"
Note 1: MSIE will display Copperminer.jpg.php as an image, but lynx will display the output of the command you gave it.
Note 2: http://www.google.com/search?q=allinurl%3A+/upload.php?album=

---TIMELINE---
mar 31, 2003 - Issue discovered, working exploit written.
mar 31, 2003 - Author contacted, problem aknowledged by author.
apr 05, 2003 - Patches released through Coppermine website.
apr 07, 2003 - Information disclosed.

---PATCH---

EC-Council

Develop a query string to locate vulnerable targets on the Web; the vendor's Web site is a good place to discover what exactly the product's Web pages look like

For example, some administrators might modify the format of a vendor-supplied Web page to fit the theme of the site

These types of modifications can impact the effectiveness of a Google search that targets a vendor-supplied page format

You can find that most sites look very similar and that nearly every site has a "powered by" message at the bottom of the main page

EC-Council
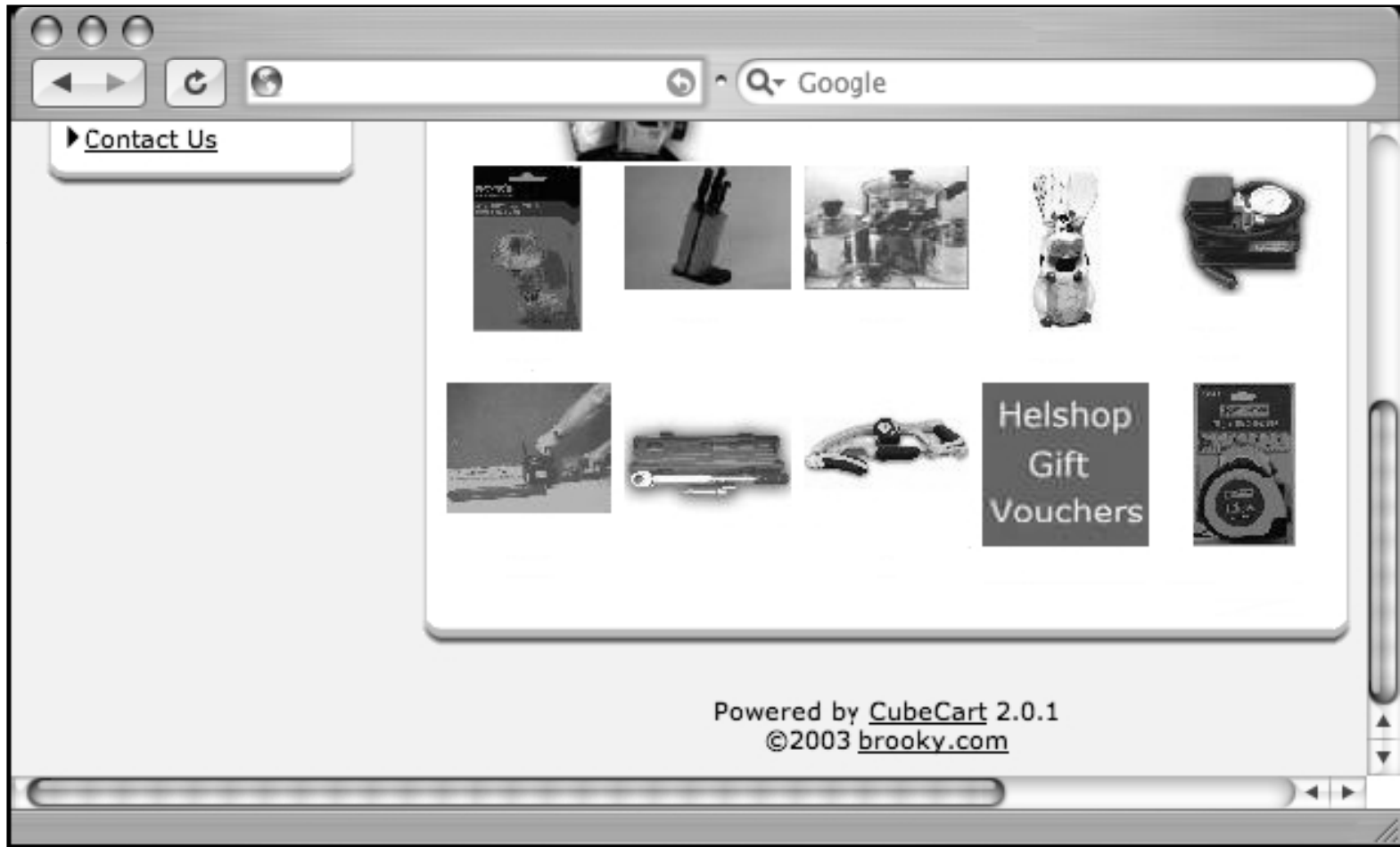
A hacker might use the source code of a program to discover ways to search for that software with Google

To find the best search string to locate potentially vulnerable targets, you can visit the Web page of the software vendor to find the source code of the offending software

In cases where source code is not available, an attacker might opt to simply download the offending software and run it on a machine he controls to get ideas for potential searches
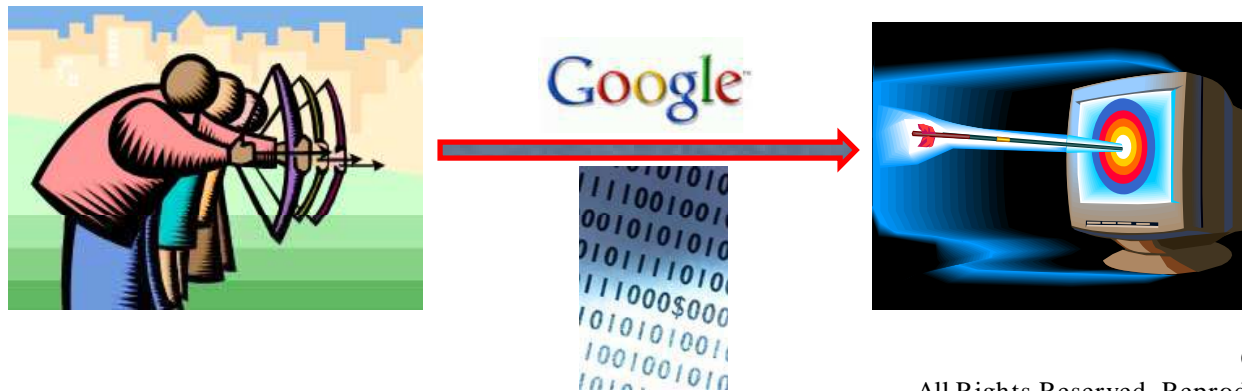
Query	Vulnerability
"Powered by A-CART"	A-CART 2.x vulnerable to cross-site scripting
inurl:"dispatch.php?atknodetype" | inurl:class.atkdateattribute.js.php	Achievo .8.x could allow remote code execution
intitle:guestbook "advanced guestbook 2.2 powered"	Advanced Guestbook v2.2 has an SQL injection problem that allows unauthorized access
"Powered by AJ-Fork v.167"	AJ-Fork, a fork based on the CuteNews 1.3.1 core, is susceptible to multiple vulnerabilities
"BlackBoard 1.5.1-f | ¬© 2003-4 by Yves Goergen"	BlackBoard 1.5.1 has a remote file inclusion vulnerability
"BosDates Calendar System " "powered by BosDates v3.2 by BosDev"	BosDates 3.2 is vulnerable to SQL injection
inurl:changepassword.cgi —cvs	changepassword.cgi allows for unlimited repeated failed login attempts
"Copyright ¬© 2002 Agustin Dondo Scripts"	CoolPHP 1.0 has multiple vulnerabilities
"Powered by CubeCart 2.0.1"	CubeCart 2.0.1 has an SQL injection vulnerability
"Powered *: newtelligence" ("dasBlog 1.6"| "dasBlog 1.5"| "dasBlog 1.4"|"dasBlog 1.3")	DasBlog versions 1.3–1.6 are susceptible to an HTML injection vulnerability in their request log
"Powered by DCP-Portal v5.5"	DCP-Portal version 5.5 is vulnerable to SQL injection
"2003 DUware All Rights Reserved"	DUForum 3.0 may allow a remote attacker to carry out SQL injection and HTML injection attacks
"inurl:/site/articles.asp?idcategory="	Dwc_Articles 1.6 has multiple input validation problems
inurl:custva.asp	EarlyImpact Productcart v1.5 contains multiple vulnerabilities
inurl:"/becommunity/community/ index.php?pageurl="	E-market prior to 1.4.0 contains various vulnerabilities
intitle:"EMUMAIL - Login" "Powered by EMU Webmail"	EMU Webmail 5.6 messaging product is susceptible to a cross-site scripting vulnerability
"Powered by FUDforum"	FUDforum 2.0.2 allows manipulation of arbitrary server files
"1999-2004 FuseTalk Inc" -site:fusetalk.com	FuseTalk forums (v4) are susceptible to cross-site scripting attacks
"Powered by My Blog" intext: "FuzzyMonkey.org"	FuzzyMonkey 2.11 has an SQL injection vulnerability
"Powered by Gallery v1.4.4"	Gallery 1.4.4 allows remote code execution
intitle:gallery inurl:setup "Gallery configuration"	Gallery default configuration files allow gallery modification
inurl:"messageboard/Forum.asp?"	GoSmart Message Board (specific versions) are susceptible to SQL injection attack and cross-site scripting attack
intitle:welcome.to.horde	Horde Mail prior to 2.2 has had several reported vulnerabilities
"Powered by IceWarp Software" inurl:mail	IceWarp Web Mail (versions prior to 5.2.8) is reported prone to multiple input validation vulnerabilities
"Ideal BB Version: 0.1" -idealbb.com	Ideal BB 0.1 is susceptible to multiple vulnerabilities
"Powered by Ikonboard 3.1.1"	IkonBoard 3.1.1 allows cross-site scripting
"Powered by Invision Power Board(U) v1.3 Final ¬© *"	Invision Power Board v1.3 is vulnerable to SQL injection
inurl:wiki/MediaWiki	MediaWiki 1.3.5 has a cross-site scripting vulnerability
"Powered by Megabook *" inurl:guestbook.cgi	MegaBook 2.0 is prone to multiple HTML injection vulnerabilities
"Powered by mnoGoSearch - free Web search engine software"	mnGoSearch 3.1.20 and 3.2.10 contain a buffer overflow vulnerability

intitle:"MRTG/RRD" 1.1* (inurl:mrtg.cgi | inurl:14all.cgi |traffic.cgi)	MRTG 1.1 allows viewing of arbitrary system files
filetype:cgi inurl:nbmember.cgi	nbmember.cgi 2.0 allows system and user information disclosure
"Powered by ocPortal" -demo -ocportal.com	ocPortal 1.0.3 allows remote file inclusion
intitle:"PHP Explorer" ext:php (inurl:phpexplorer.php | inurl:list.php | inurl:browse.php)	PHP Explorer scripts reveal server information and provides remote shell access
"create the Super User" "now by clicking here"	PHP-Nuke open configuration allows arbitrary creation of admin users
"Enter ip" inurl:"php-ping.php"	php-ping prior to version 1.2 may be prone to a remote command execution vulnerability
intitle:"phpremoteview" filetype: php "Name, Size,	phpRemoteView allows browsing of entire file system
inurl:"plog/register.php"	pLog installation scripts should be removed after install because they allow for program compromise
filetype:php inurl:index.php inurl: "module=subjects" inurl:"func= *" (listpages| viewpage | listcat)	Postnuke Modules Factory Subjects module has an SQL injection vulnerability
"Online Store - Powered by ProductCart"	ProductCart v1.5–1.6 and v2 are vulnerable to an SQL injection vulnerability
inurl:com_remository	ReMOSitory 4.5.1 1.09 module for Mambo is prone to an SQL injection vulnerability
inurl:"slxWeb.dll"	SalesLogix 2000.0 contains multiple remote vulnerabilities
"File Upload Manager v1.3" "rename to"	thepeak file upload manager allows arbitrary user to transfer files
filetype:cgi inurl:tseekdir.cgi	Turbo Seek 1.7.2 search engine reveals arbitrary file contents
inurl:ttt-webmaster.php	Turbo traffic trader Nitro v1.0 contains multiple vulnerabilities

ext:cgi inurl:ubb6_test.cgi	UBB trial version contains files that are not safe to keep online after going live
"Powered by: vBulletin * 3.0.1" inurl:newreply.php	vBulletin 3.0.1 allows arbitrary code execution
inurl:/cgi-bin/index.cgi inurl: topics inurl:viewcat= +intext: "WebAPP" -site:web-app.org	WebAPP 0.x has a serious reverse directory traversal vulnerability
intitle:"WebJeff - FileManager" intext:"login" intext:Pass|Passe	WebJeff-FileManager 1.x can reveal arbitrary system files
intitle:"Index of /" modified php.exe	Windows PHP parser allow an attacker to view arbitrary system files
"Powered by WowBB" -site: wowbb.com	WowBB 1.x affected by multiple input validation vulnerabilities
"Powered by YaPig V0.92b"	YaPiG 0.92b contains an HTML injection vulnerability

One of the oldest and most familiar techniques for locating vulnerable Web servers is through the use of a CGI scanner

These programs parse a list of known "bad" or vulnerable Web files and attempt to locate those files on a Web server

Based on various response codes, the scanner could detect the presence of these potentially vulnerable files

A CGI scanner can list vulnerable files and directories in a data file, such as:

```
/cgi-bin/userreg.cgi
/cgi-bin/cgiemail/uargg.txt
/random_banner/index.cgi
/random_banner/index.cgi
/cgi-bin/mailview.cgi
/cgi-bin/maillist.cgi
/iissamples/ISSamples/SQLQHit.asp
/iissamples/ISSamples/SQLQHit.asp
/SiteServer/admin/findvserver.asp
/scripts/cphost.dll
/cgi-bin/finger.cgi
```

EC-Council

C|EH ™
Certified | Ethical | Hacker

Example: search for *inurl:/cgi-bin/userreg.cgi*
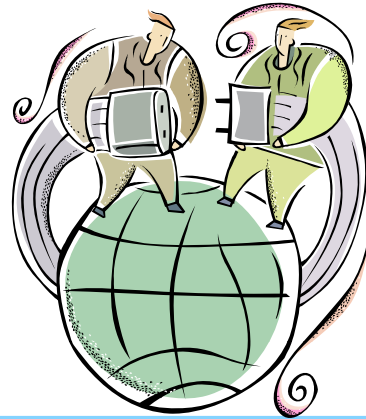
Tracking Down Web Servers, Login Portals, and Network Hardware

EC-Council
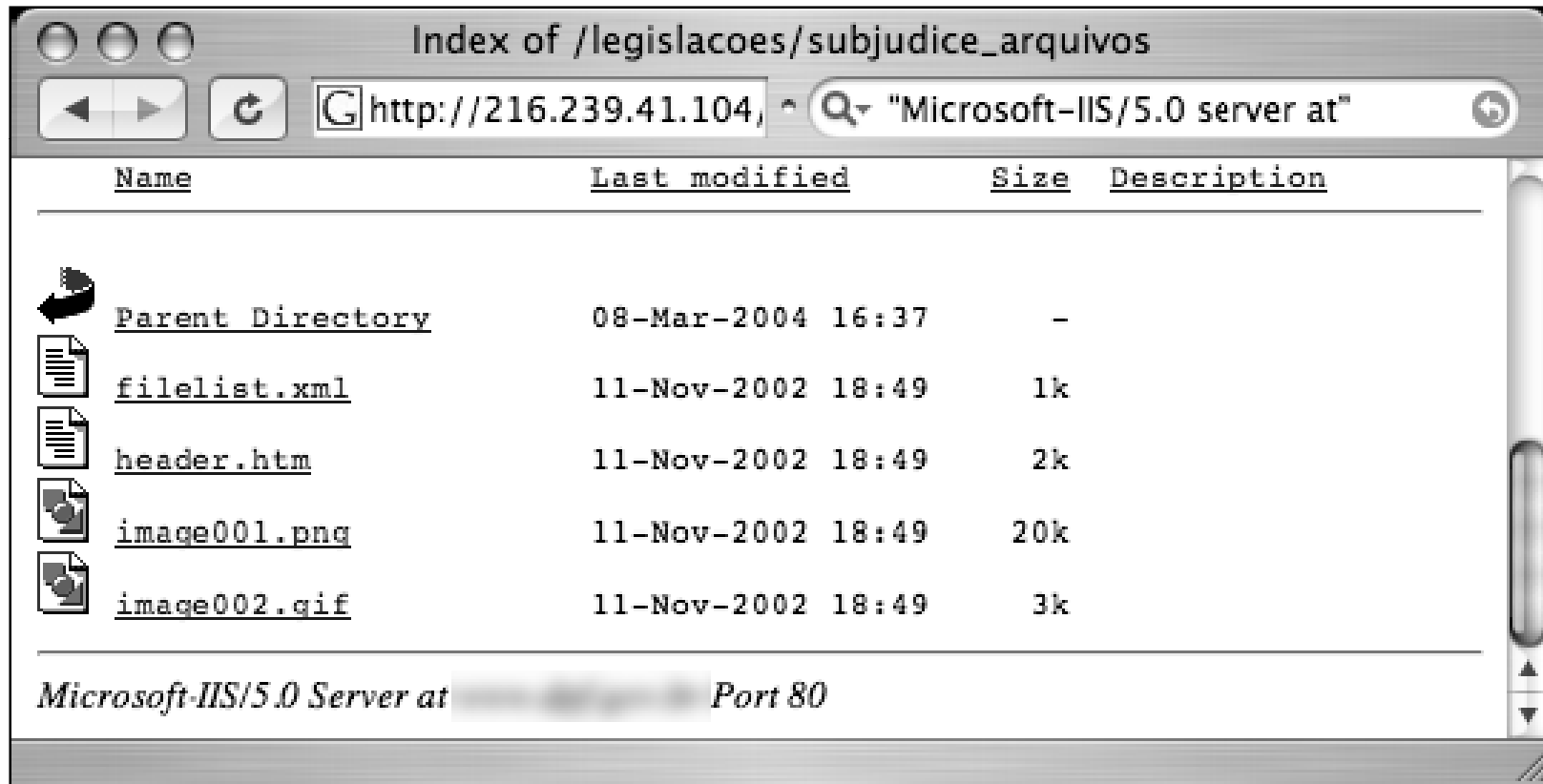
Web Server Software Error Messages

Error messages contain a lot of useful information, but in the context of locating specific servers, you can use portions of various error messages to locate servers running specific software versions

The best way to find error messages is to figure out what messages the server is capable of generating

You could gather these messages by examining the server source code or configuration files or by actually generating the errors on the server yourself

The best way to get this information from IIS is by examining the source code of the error pages themselves

IIS 5 and 6, by default, display static HTTP/1.1 error messages when the server encounters some sort of problem

These error pages are stored by default in the %SYSTEMROOT%\help\iisHelp\common directory

A query such as *intitle:"The page cannot be found" "please following" "Internet * Services"* can be used to search for IIS servers that present a 400 error
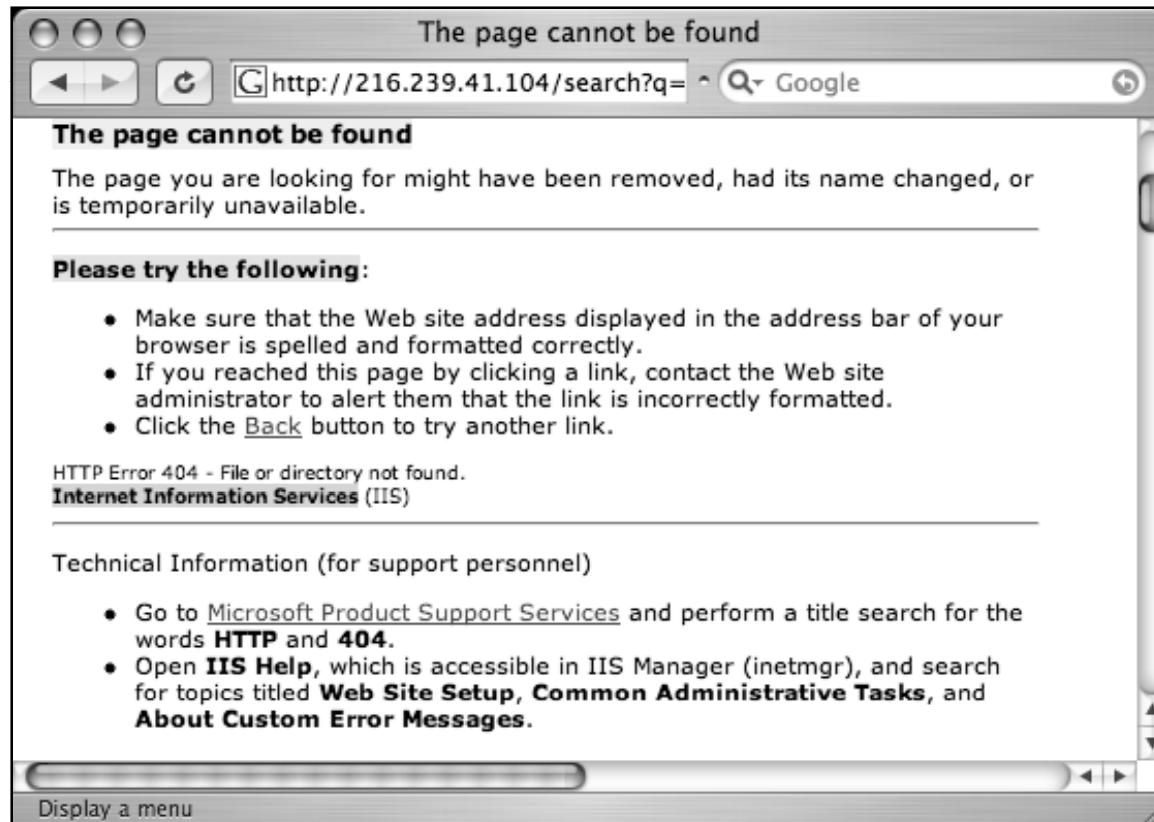
403.5	The page must be viewed with a high-security Web browser
403.6	You are not authorized to view this page
403.7	The page requires a client certificate
403.8	You are not authorized to view this page
403.9	The page cannot be displayed
403.10, 403.11	You are not authorized to view this page
403.12, 403.13	The page requires a valid client certificate
403.15	The page cannot be displayed
403.16, 403.17	The page requires a valid client certificate
404.1, 404b	The Web site cannot be found
405	The page cannot be displayed
406	The resource cannot be displayed
407	Proxy authentication required
410	The page does not exist
412	The page cannot be displayed
414	The page cannot be displayed
500, 500.11, 500.12, 500.13, 500.14, 500.15	The page cannot be displayed
502	The page cannot be displayed

400	The page cannot be found
401.1, 401.2, 401.3, 401.4, 401.5	You are not authorized to view this page
403.1, 403.2	The page cannot be displayed
403.3	The page cannot be saved
403.4	The page must be viewed over a secure channel

News–Journal Online – Beach news

http://216.23 | intext:"404 Object Not Found" Microsoft-IIS/5.0

HTTP/1.1 404 Object Not Found Server: Microsoft-IIS/5.0 Date: Wed, 08 Sep 2004 12:59:58 GMT
X-Powered-By: ASP.NET Content-Type: text/html

404 Object Not Found

HTTP/1.1 404 Object Not Found Server: Microsoft-IIS/5.0 Date: Wed, 08 Sep 2004 12:59:58 GMT
X-Powered-By: ASP.NET Content-Type: text/html

404 Object Not Found

HTTP/1.1 404 Object Not Found Server: Microsoft-IIS/5.0 Date: Wed, 08 Sep 2004 12:59:58 GMT

Display a menu

Apache Web servers can also be located by focusing on server-generated error messages

Some generic searches such as *"Apache/1.3.27 Server at"-intitle:index.of intitle:inf" or "Apache/1.3.27 Server at"-intitle:index.of intitle:error*
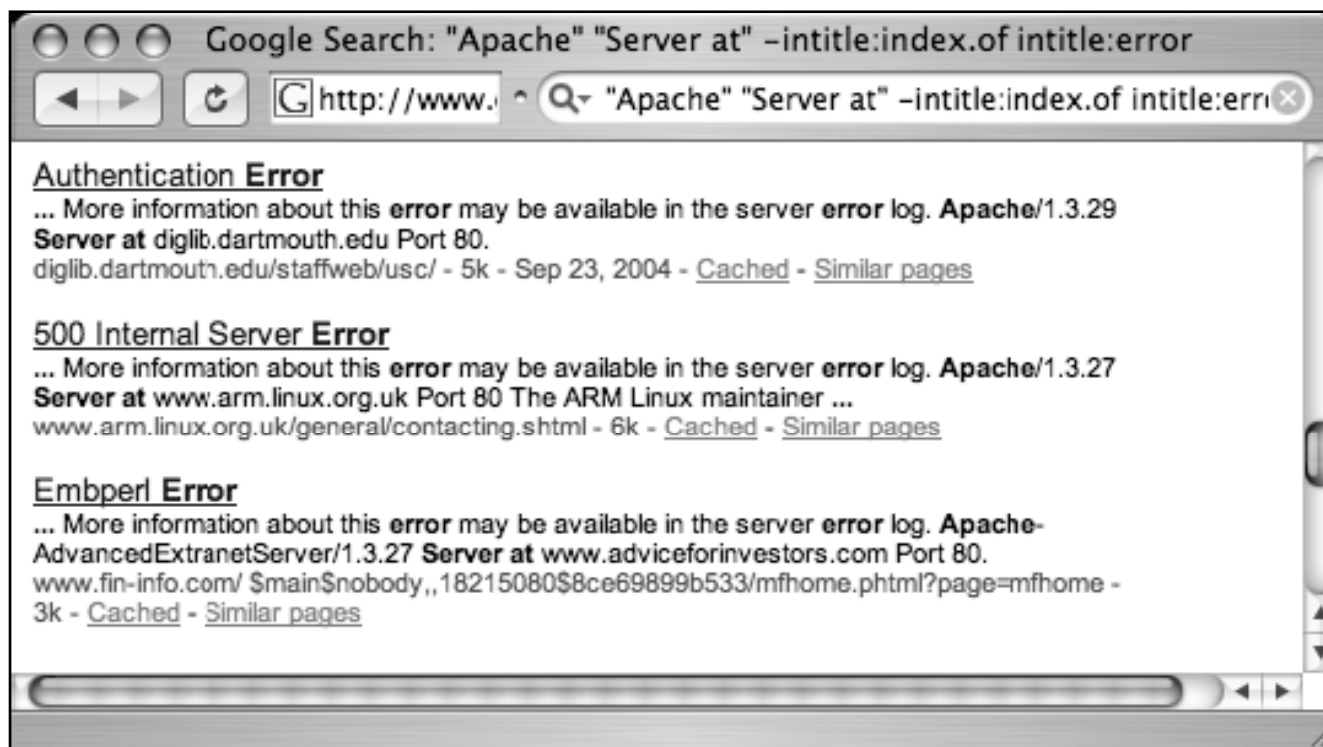
Bad gateway!	The proxy server received an invalid response from an upstream server.
Bad request!	Your browser (or proxy) sent a request that this server could not understand.
Access forbidden!	You don't have permission to access the requested directory. Either there is no index document or the directory is read-protected.
Resource is no longer available!	The requested URL is no longer available on this server and there is no forwarding address.
Server error!	The server encountered an internal error and was unable to complete your request.
Method not allowed!	A request with the method is not allowed for the requested URL.
No acceptable object found!	An appropriate representation of the requested resource could not be found on this server.
Object not found!	The requested URL was not found on this server.
Cannot process request!	The server does not support the action requested by the browser.
Precondition failed!	The precondition on the request for the URL failed positive evaluation.
Request entity too large!	The method does not allow the data transmitted, or the data volume exceeds the capacity limit.
Request time-out!	The server closed the network connection because the browser didn't finish the request within the specified time.
Submitted URI too large!	The length of the requested URL exceeds the capacity limit for this server. The request cannot be processed.
Service unavailable!	The server is temporarily unable to service your request due to maintenance downtime or capacity problems. Please try again later.
Authentication required!	This server could not verify that you are authorized to access the URL. You either supplied the wrong credentials (such as a bad password) or your browser doesn't understand how to supply the credentials required.

EC-Council

Although this ASP message is fairly benign, some ASP error messages are much more revealing

Consider the query *"ASP.NET_SessionId"*"*data source="*, which locates unique strings found in ASP.NET application state dumps

These dumps reveal all sorts of information about the running application and the Web server that hosts that application
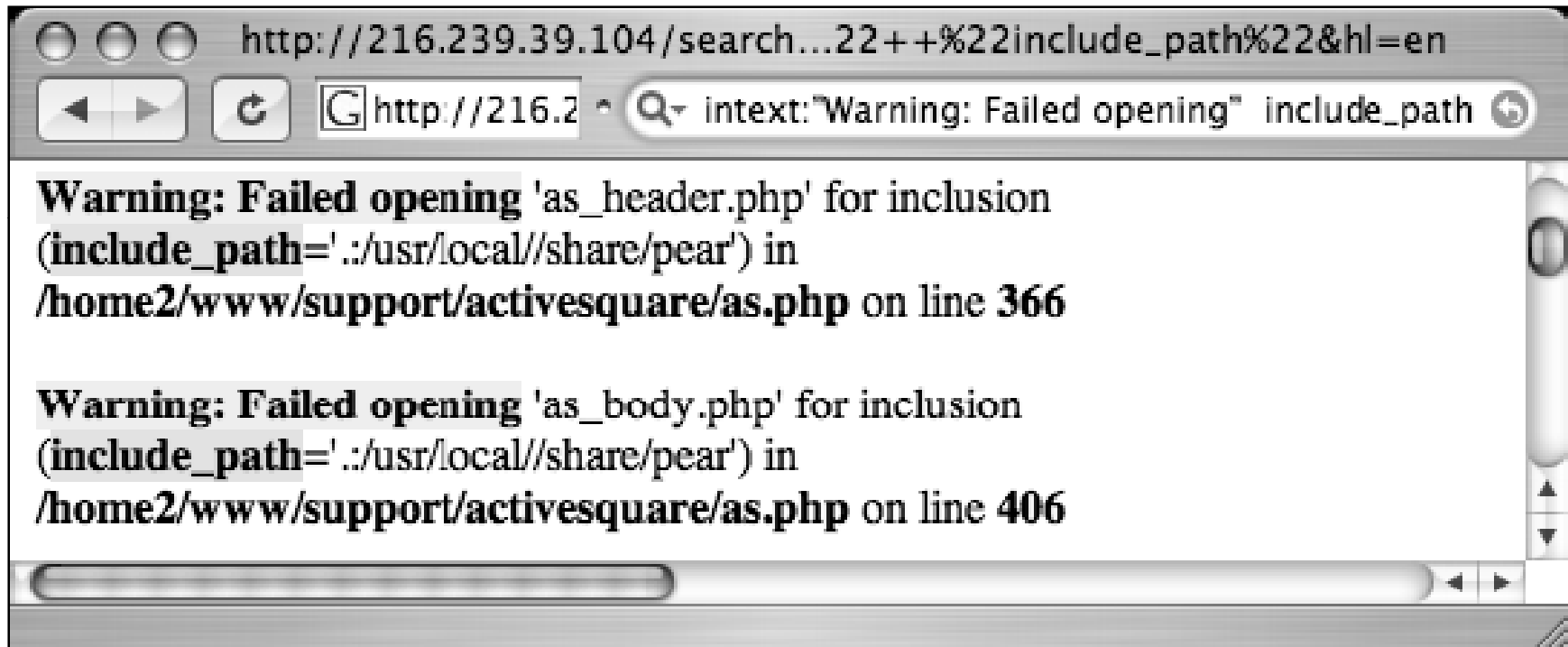
An advanced attacker can use encrypted password data and variable information in these stack traces to subvert the security of the application and perhaps the Web server itself

EC-Council

ASP Dumps Provide Dangerous Details

EC-Council

Many Errors Reveal Pathnames and Filenames

http://216.239.39.104/search...22++%22include_path%22&hl=en

http://216.2 ^ intext:"Warning: Failed opening" include_path

Warning: Failed opening 'as_header.php' for inclusion (**include_path**='.:/usr/local//share/pear') in **/home2/www/support/activesquare/as.php** on line **366**

Warning: Failed opening 'as_body.php' for inclusion (**include_path**='.:/usr/local//share/pear') in **/home2/www/support/activesquare/as.php** on line **406**

```
http://64.233.161.104/search?q=cac...bot.com+%22Server_Software".&hl=en

◀ ▶  ⟳  ⊕ http://64.233.1 ^  Q▾ "HTTP_FROM=googlebot" googlebot.com "Server_Softwa

HTTP_ACCEPT : text/html,text/plain
HTTP_USER_AGENT : Googlebot/2.1 (+http://www.google.com/bot.html)
GATEWAY_INTERFACE : CGI/1.1
HTTP_HOST : www.uib.no
SERVER_SOFTWARE : Apache/1.3.26 (Unix) PHP/4.2.1
SERVER_ADMIN : webmaster@uib.no
REMOTE_ADDR : 66.249.64.183
HTTP_IF_MODIFIED_SINCE : Tue, 31 Aug 2004 01:10:34 GMT
SCRIPT_NAME : /cgi-bin/env
SERVER_NAME : www.uib.no
DOCUMENT_ROOT : /www
REQUEST_URI : /cgi-bin/env
SCRIPT_FILENAME : /local/apache/cgi-bin/env
REQUEST_METHOD : GET
PATH : /usr/sbin:/usr/bin
HTTP_FROM : googlebot(at)google.com
SERVER_PORT : 80
```

Another way to locate specific types of servers or Web software is to search for default Web pages

Most Web software, including the Web server software itself, ships with one or more default or test pages

These pages can make it easy for a site administrator to test the installation of a Web server or application

Google crawls a Web server while it is in its earliest stages of installation, still displaying a set of default pages

In these cases there is generally a short window of time between the moment when Google crawls the site and when the intended content is actually placed on the server

Test Page for Apache Installation on Web Site

http://216.239.39.104/search?q=cache:Yi4alzX65zk

initle:"T

It Worked! The Apache Web Server is Installed on this Web Site!

If you can see this page, then the people who own this domain have just installed the Apache Web server software successfully. They now have to add content to this directory and replace this placeholder page, or else point the server at their real content.

If you are seeing this page instead of the site you expected, please **contact the administrator of the site involved.** (Try sending mail to `<Webmaster@domain>`.) Although this site is running the Apache software it almost certainly has no other connection to the Apache Group, so please do not send mail about this site or its contents to the Apache authors. If you do, your message will be **ignored**.

Display a menu

Many different types of Web server can be located by querying for default pages as well

Cisco Micro Webserver 200	"micro webserver home page"
Generic Appliance	"default web page" congratulations "hosting appliance"
HP appliance sa1	intitle:"default domain page" "congratulations" "hp web"
iPlanet/Many	intitle:"web server, enterprise edition"
Intel Netstructure	"congratulations on choosing" intel netstructure
JWS/1.0.3–2.0	allintitle:default home page java web server
J2EE/Many	intitle:"default j2ee home page"
Jigsaw/2.2.3	intitle:"jigsaw overview" "this is your"
Jigsaw/Many	intitle:"jigsaw overview"
KFSensor honeypot	"KF Web Server Home Page"
Kwiki	"Congratulations! You've created a new Kwiki website."
Matrix Appliance	"Welcome to your domain web page" matrix
NetWare 6	intitle:"welcome to netware 6"
Resin/Many	allintitle:Resin Default Home Page
Resin/Enterprise	allintitle:Resin-Enterprise Default Home Page
Sambar Server	intitle:"sambar server" "1997..2004 Sambar"
Sun AnswerBook Server	inurl:"Answerbook2options"
TivoConnect Server	inurl:/TiVoConnect

Outlook Web Access Default Portal

Query *allinurl:"exchange/logon.asp"*

EC-Council
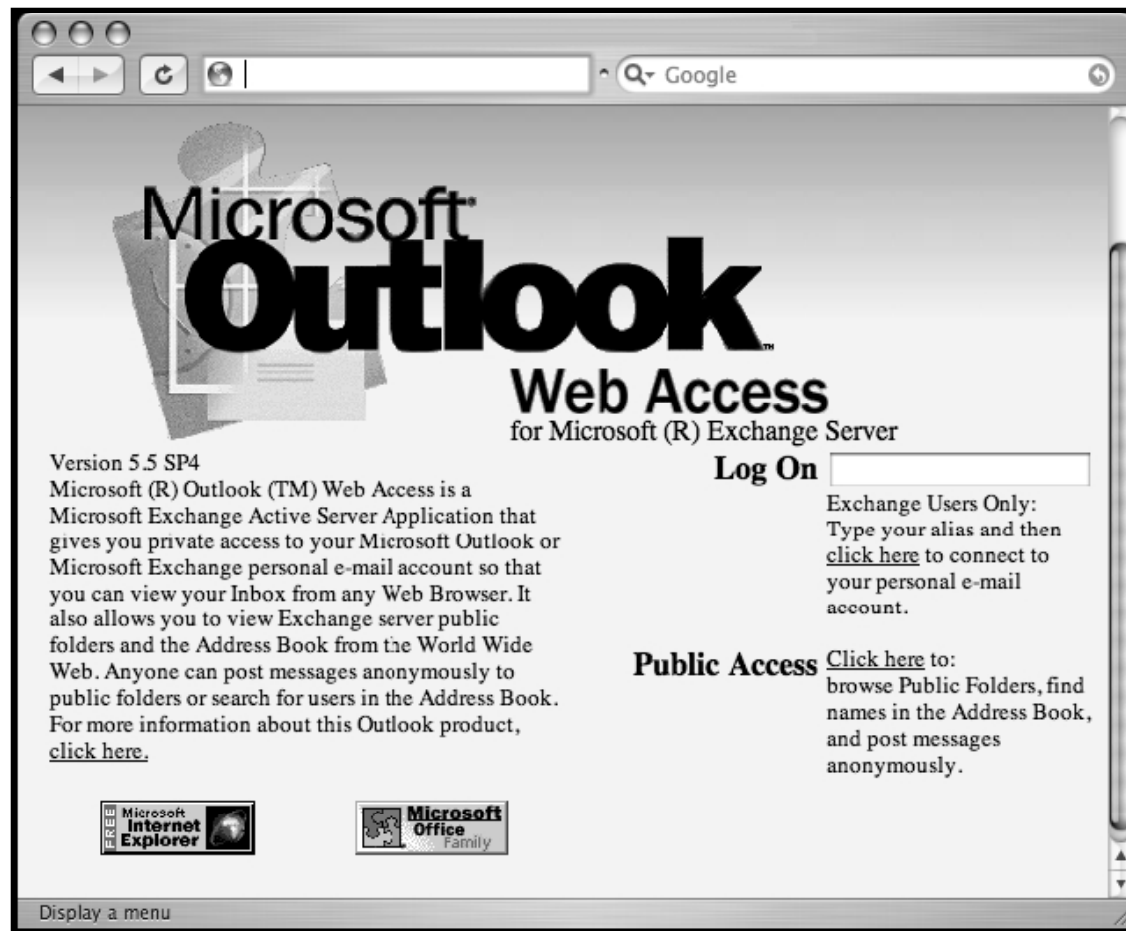
Searching for Passwords

Password data, one of the "Holy Grails" during a penetration test, should be protected
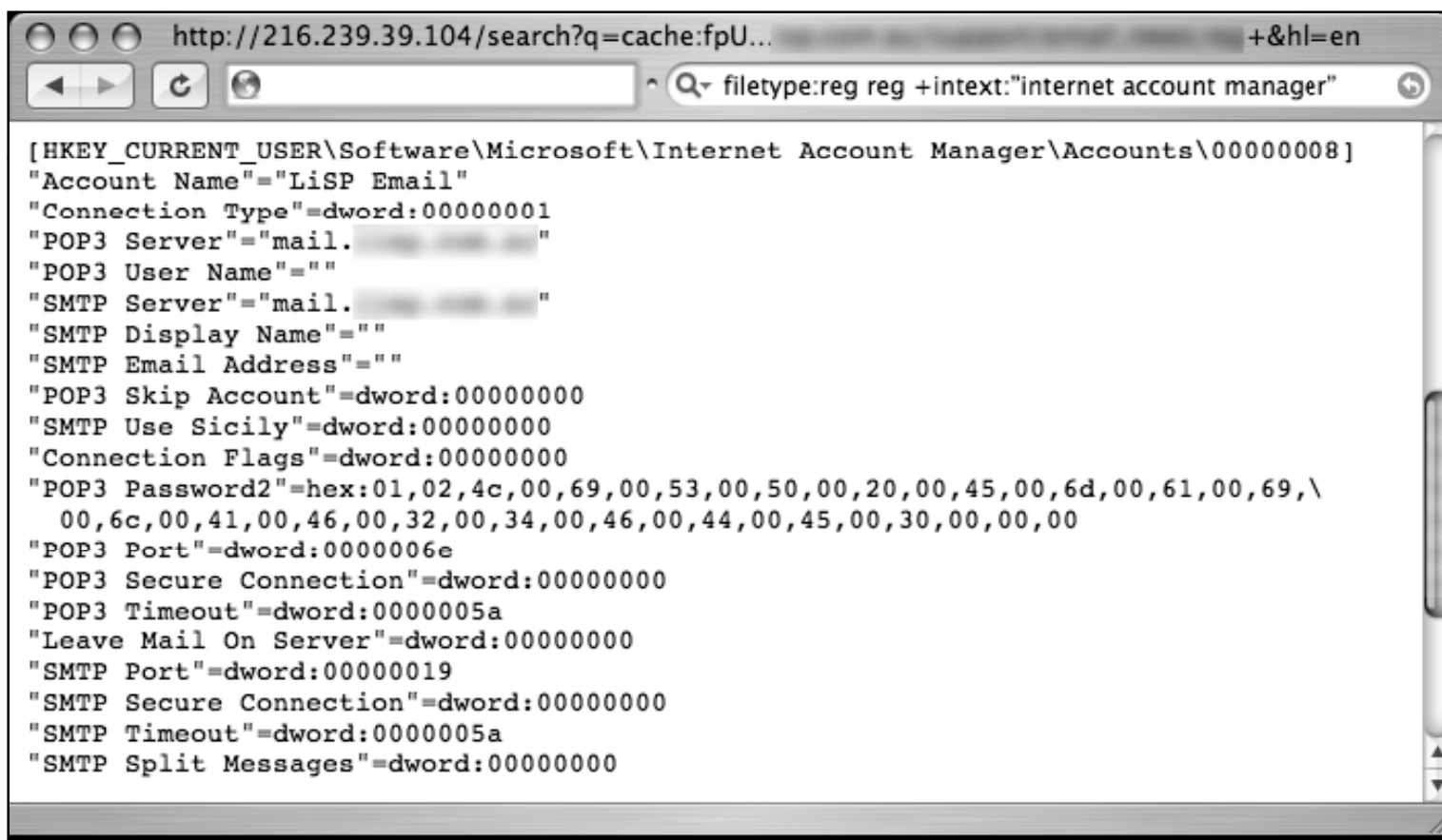
Unfortunately, many examples of Google queries can be used to locate passwords on the Web
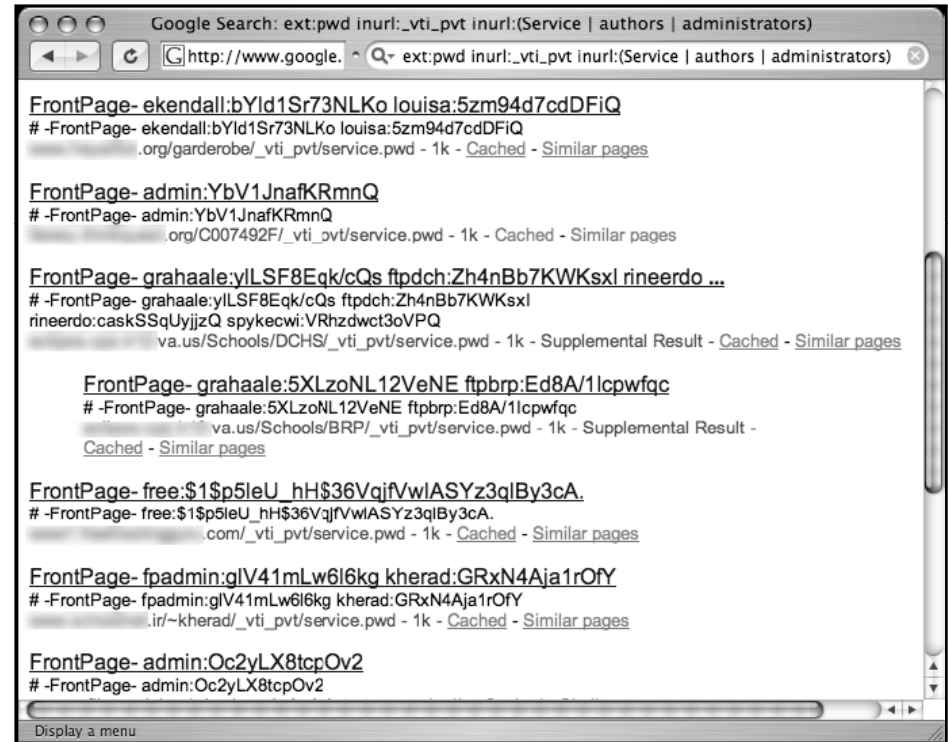
Query like *filetype:reg intext: "internet account manager"* could reveal interesting keys containing password data

```
http://216.239.39.104/search?q=cache:fpU...                    +&hl=en

◄ ► C 🌐                          ^ Q▾ filetype:reg reg +intext:"internet account manager"  ↺

[HKEY_CURRENT_USER\Software\Microsoft\Internet Account Manager\Accounts\00000008]
"Account Name"="LiSP Email"
"Connection Type"=dword:00000001
"POP3 Server"="mail.          "
"POP3 User Name"=""
"SMTP Server"="mail.          "
"SMTP Display Name"=""
"SMTP Email Address"=""
"POP3 Skip Account"=dword:00000000
"SMTP Use Sicily"=dword:00000000
"Connection Flags"=dword:00000000
"POP3 Password2"=hex:01,02,4c,00,69,00,53,00,50,00,20,00,45,00,6d,00,61,00,69,\
   00,6c,00,41,00,46,00,32,00,34,00,46,00,44,00,45,00,30,00,00,00
"POP3 Port"=dword:0000006e
"POP3 Secure Connection"=dword:00000000
"POP3 Timeout"=dword:0000005a
"Leave Mail On Server"=dword:00000000
"SMTP Port"=dword:00000019
"SMTP Secure Connection"=dword:00000000
"SMTP Timeout"=dword:0000005a
"SMTP Split Messages"=dword:00000000
```

Search for password information, *intext:(password | passcode | pass) intext:(username | userid | user), combines common words for passwords and* user IDs into one query

name: = "momo"; password: = "momo"; URL: = "password.htm" ...
name: = "momo"; password: = "momo"; URL: = "password.htm"; END_FILE
_____ net/password.log - 1k - Supplemental Result - Cached - Similar pages

name: = "jbhunt"; password: = "jbhunt"; URL: = "http://home.nc.rr. ...
name: = "jbhunt"; password: = "jbhunt"; URL: = "http:_____/clay123/ref23.
html"; Beth Haas name: = "BHaas"; password: = "Beth Haas"; URL: = "http ...
_____.com/clay123/password.log - 2k - Supplemental Result - Cached - Similar pages

name: = "dv21"; password: = "dv21_2004"; URL: = "intern.htm"; name ... - [
Translate this page]
name: = "dv21"; password: = "dv21_2004"; URL: = "intern.htm"; name: = "dv22"; password: =
"dv22_2004"; URL: = "intern.htm"; name: = "dv23"; password ...
_____.de/grossmann/password.log - 1k - Cached - Similar pages

EC-Council

Google Hacking Tools

Hacker group releases automated 'Google hacking' tool

Jaikumar Vijayan

February 22, 2008 (Computerworld) The Cult of the Dead Cow hacker group has released an open-source tool designed to enable IT workers to quickly scan their Web sites for security vulnerabilities and at-risk sensitive data, using a collection of specially crafted Google search terms.

The group, which refers to itself as the cDc, acknowledged that the Goolag Scanner tool could also be used by malicious attackers to look for vulnerable Web sites. "We're not stupid," a cDc member who goes by the name "Oxblood Ruffin" said. "We know some bored teenagers and criminals will try to exploit vulnerabilities [using the new tool]."

But such uses are "not something that we or anyone can control," Ruffin added. "What we're trying to do is two things: 1) to provide a very easy and legitimate tool for security professionals to test their own Web sites for vulnerabilities, and 2) to raise awareness about Web security in and of itself."

Goolag Scanner is a Windows-based auditing tool that was built around the concept of "Google hacking," a form of vulnerability research created by a hacker who uses the name "Johnny I Hack Stuff." Google hacking involves the use of certain types of search queries to look for Web site vulnerabilities. More than 1,500 such queries -- or Googledorks, as they and the people who leave their Web sites exposed to them are sometimes known -- have been compiled into a database by Johnny I Hack Stuff over the past few years.

Source: *http://www.computerworld.com/*

The Google Hacking Database (GHDB) contains queries that identify sensitive data such as portal logon pages, logs with network security information, and so on

Visit http://johnny.ihackstuff.com

Google Hacking Database (GHDB)
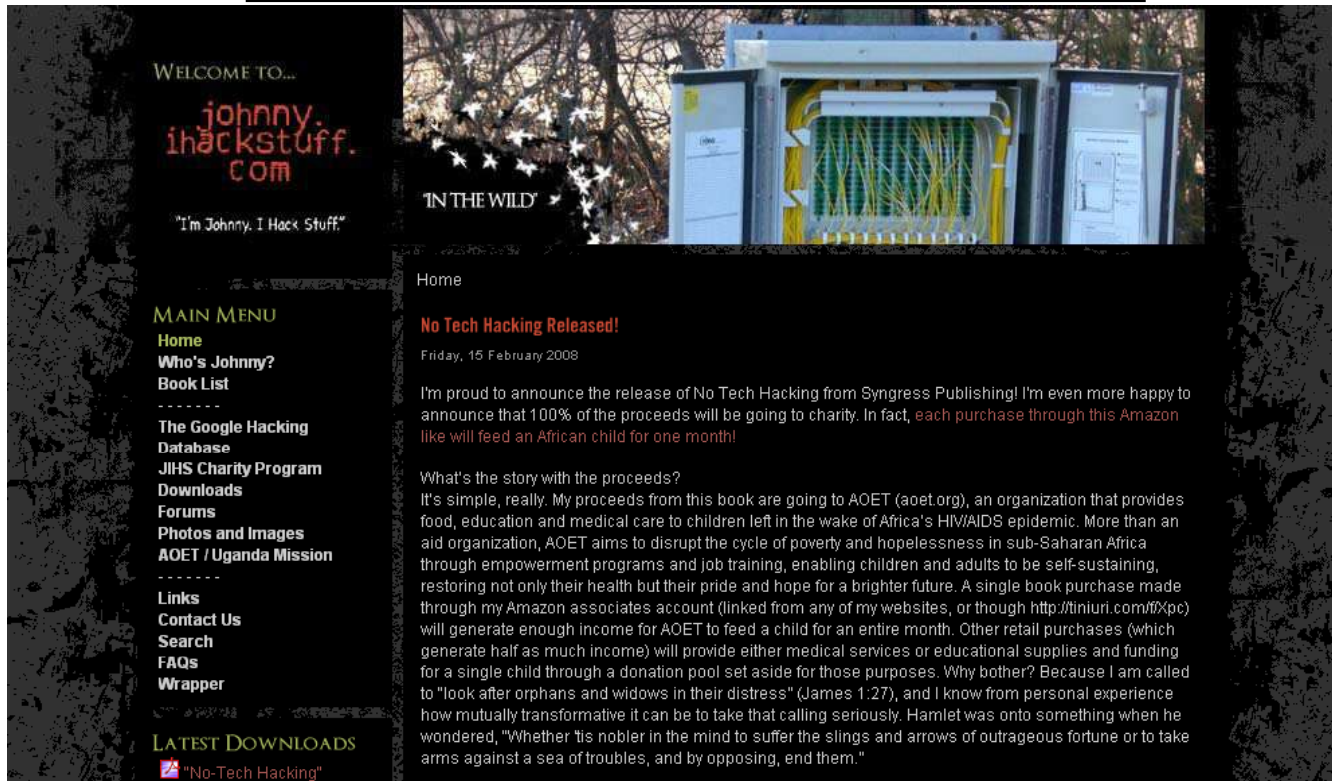
--==[Category View]==--

Welcome to the Google Hacking Database (GHDB)!

We call them 'googledorks': Inept or foolish people as revealed by Google. Whatever you call these fools, you've found the center of the Google Hacking Universe! Stop by our forums to see where the magic happens!

Advisories and Vulnerabilities (215 entries)
These searches locate vulnerable servers. These searches are often generated from various security advisory posts, and in many cases are product or version-specific.

Error Messages (68 entries)
Really retarded error messages that say WAY too much!

Files containing juicy info (230 entries)
No usernames or passwords, but interesting stuff none the less.

Files containing passwords (135 entries)
PASSWORDS, for the LOVE OF GOD!!! Google found PASSWORDS!

Files containing usernames (15 entries)
These files contain usernames, but no passwords... Still, google finding usernames on a web site..

Footholds (21 entries)
Examples of queries that can help a hacker gain a foothold into a web server

Pages containing login portals (232 entries)
These are login pages for various services. Consider them the front door of a website's more sensitive functions.

Pages containing network or vulnerability data (59 entries)
These pages contain such things as firewall logs, honeypot logs, network information, IDS logs... all sorts of fun stuff!

Sensitive Directories (61 entries)

SiteDigger Tool

SiteDigger searches Google's cache to look for vulnerabilities, errors, configuration issues, proprietary information, and interesting security nuggets on websites

Gooscan
johnny.ihackstuff.com
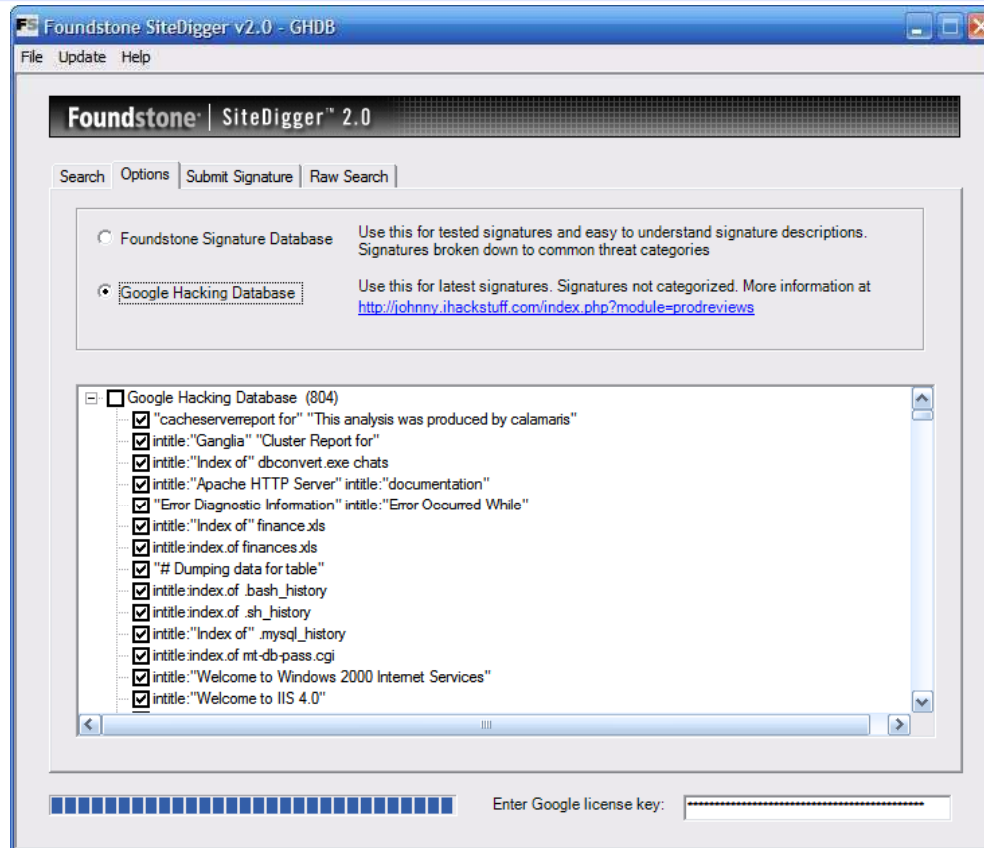
Gooscan is a tool that automates queries against Google search appliances

But it can be run against Google itself in direct violation of their Terms of Service

For the security professional, gooscan serves as a front end for an external server assessment and aids in the information-gathering phase of a vulnerability assessment

For the web server administrator, gooscan helps discover what the web community may already know about a site thanks to Google's search appliance

Goolink Scanner

It removes the cache information from your searches and only collects and displays the links

This is very handy for finding vulnerable sites wide open to google and googlebots

EC-Council
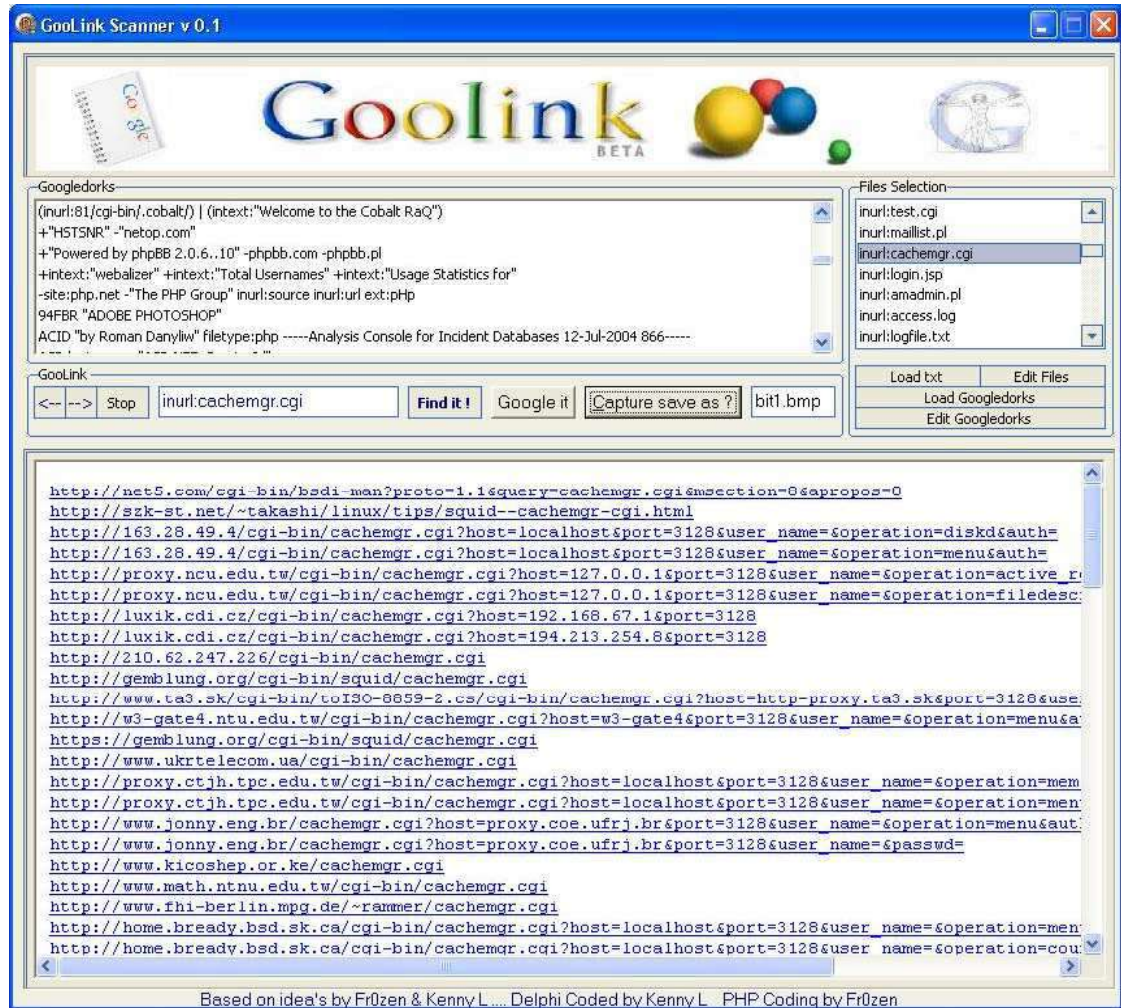
Goolag Scanner

Goolag Scanner enables everyone to audit his/her own web site via Google

It uses one xml-based configuration file for its settings
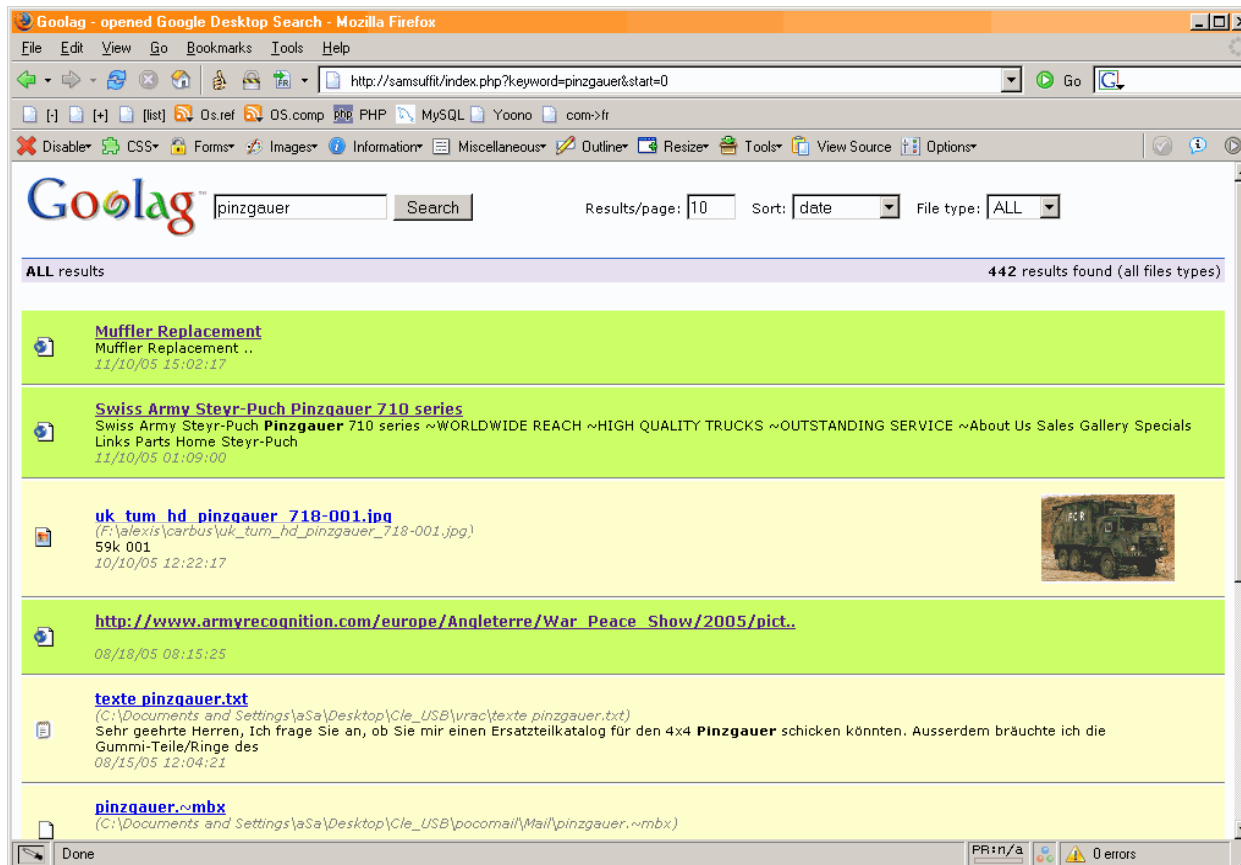
Tool: Google Hacks
code.google.com/p/googlehacks/

Google Hacks is a compilation of carefully crafted Google searches that expose novel functionality from Google's search and map services

You can use it to view a timeline of your search results, view a map, search for music, search for books, and perform many other specific kinds of searches

You can also use this program to use google as a proxy

EC-Council

Google Hacks

File   Edit   Help

Search String :   hack   Search

Type

- ◯ Music   ⦿ Book   ◯ Video   ◯ Tools
- ◯ Hacks   ◯ Proxy   ◯ Lyrics   ◯ Fonts
- ◯ Application ◯ Torrent   ◯ Cache   ◯ Web hosting

File Types

- ☑ .pdf   ☑ .rtf
- ☐ .txt   ☐ .chm
- ☐ .lit   ☐ .pps
- ☐ .rar   ☐ .zip
- ☑ .doc   ☐ .odt

This program is for educational purposes only.

Google Hacks Input

EC-Council

Google Hacks: Screenshot

Google Hacks Output

Google Hack Honeypot is the reaction to a new type of malicious web traffic: search engine hackers

It is designed to provide reconnaissance against attackers that use search engines as a hacking tool against resources
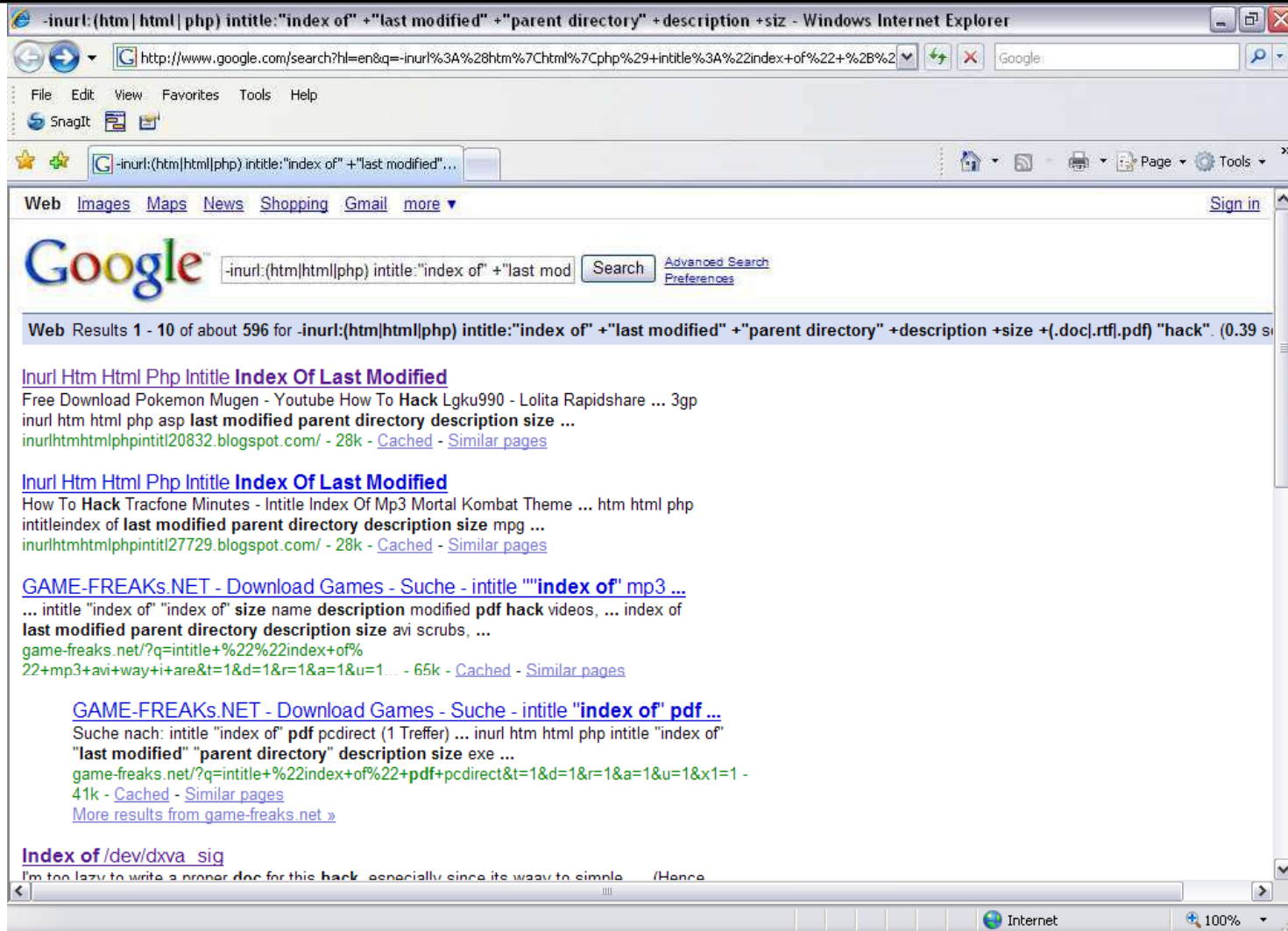
ghh
The "Google Hack" Honeypot
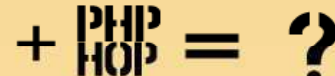
Google Protocol is a little app that when installed, registers two extra protocols similar to the http: and the ftp: protocols under windows, namely google: and lucky:

Urls starting with the 'google:' refer to the corresponding google search

Urls starting with the 'lucky:' refer to the top Google result

Google Cartography

Google Cartography uses the Google API to find web pages referring to street names

Initial street and region criteria are combined to form a search query, which is then executed by the Google API

Each URL from the Google results is fetched and the content of the pages converted into text

The text is then processed using regular expressions designed to capture information relating to the relationship between streets

EC-Council

Summary
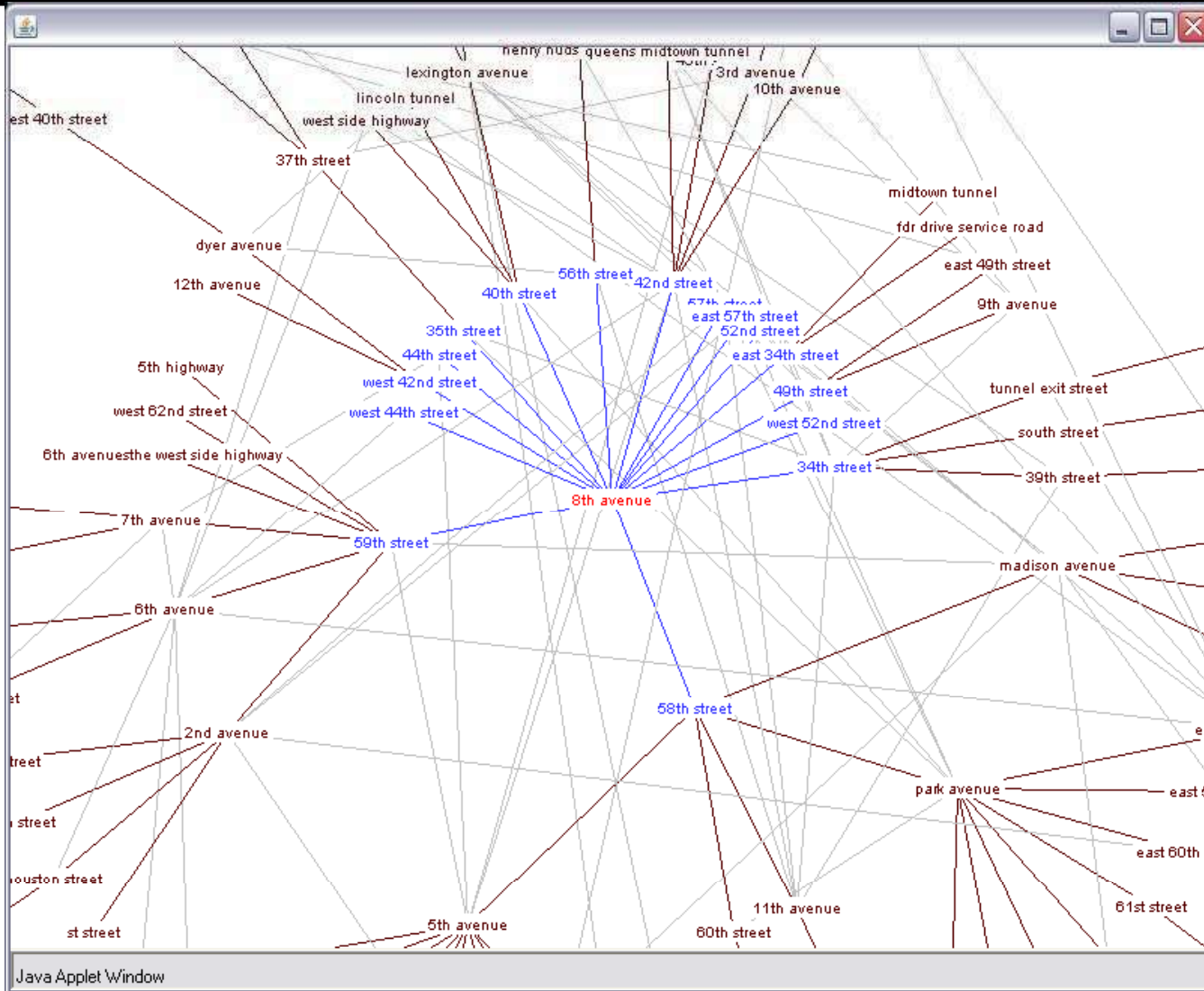
In this module, Google hacking techniques have been reviewed
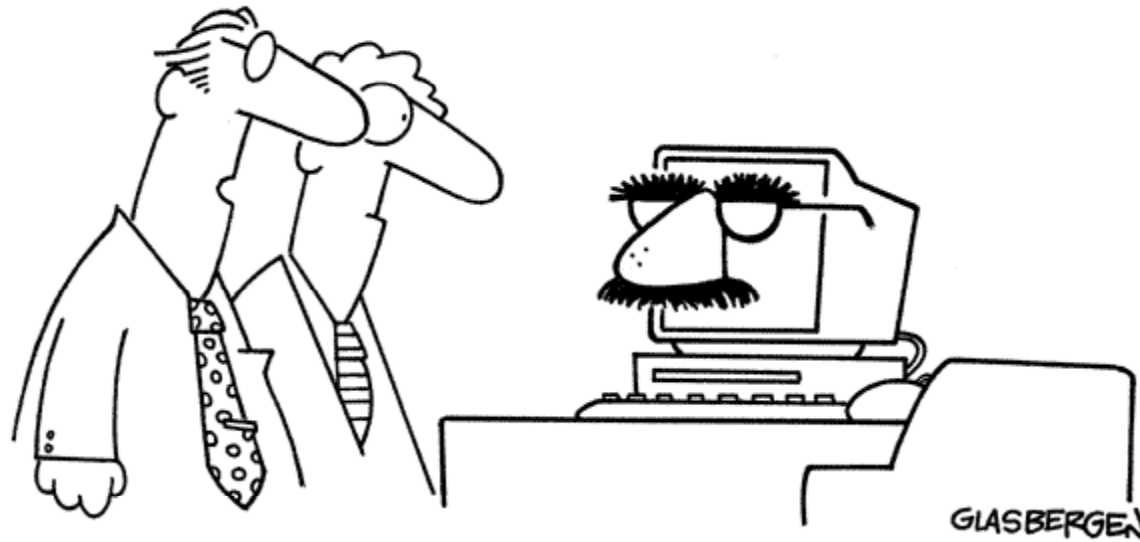
The following Google hacking techniques have been discussed:

- Software Error Messages
- Defaultpages
- Explanation of techniques to reveal password
- Locating targets
- Searching for passwords

"I'm sure there are better ways to disguise sensitive information, but we don't have a big budget."