



Ethical Hacking and Countermeasures

Version 6

Module IV

Google Hacking

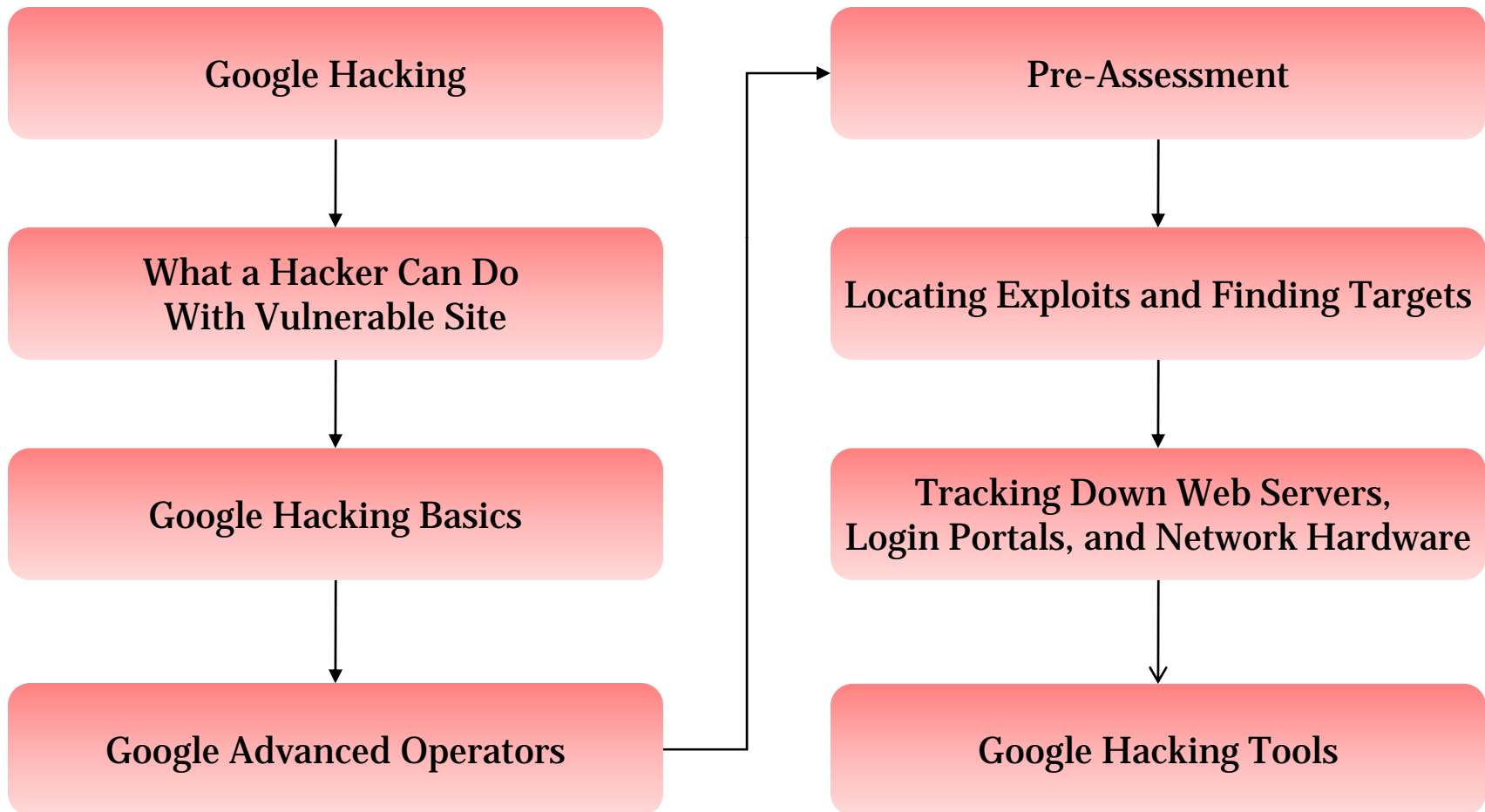
Module Objective

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



Module Flow



What is Google Hacking

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

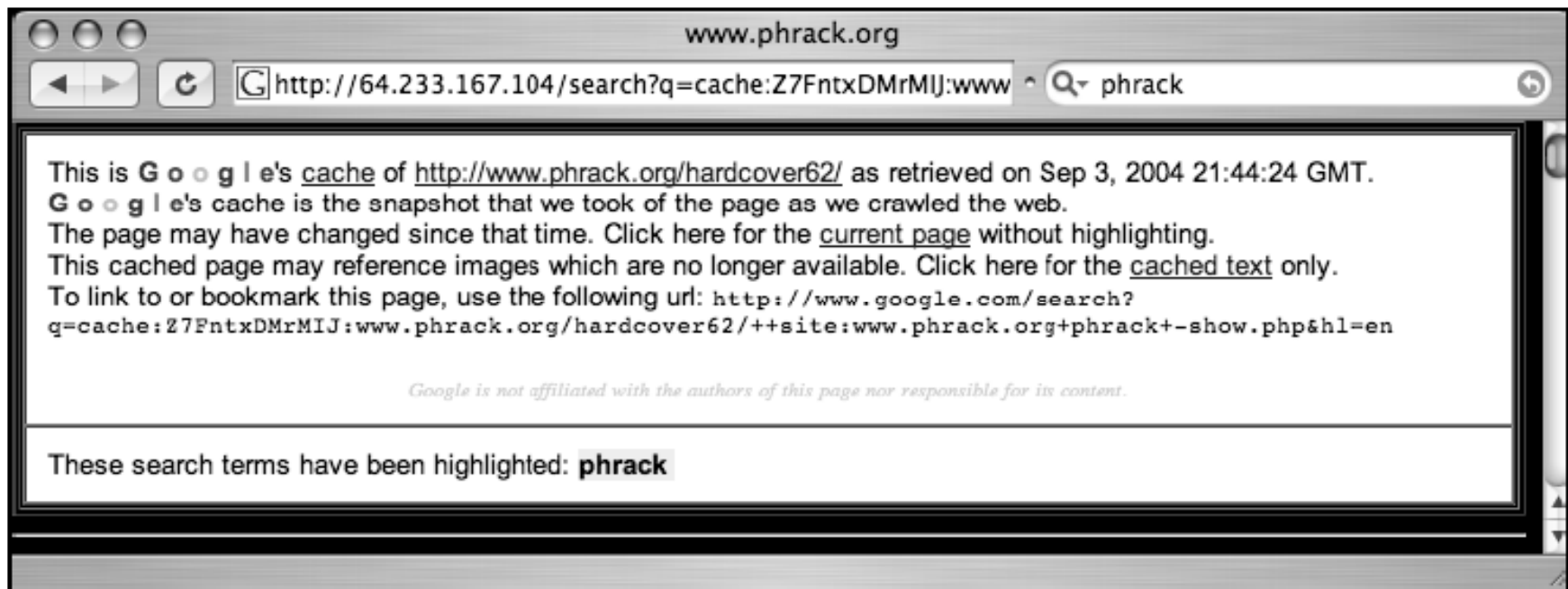


Google Hacking Basics

Anonymity with Caches

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

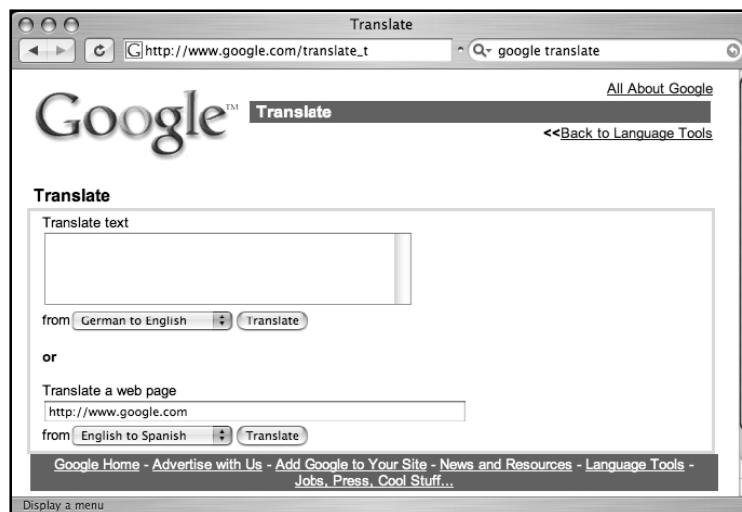


Using Google as a Proxy Server

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





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

Directory Listings(cont'd)



Locating Directory Listings

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 title

An obvious query to find this type of page might be *intitle: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



Locating Directory Listings (cont'd)

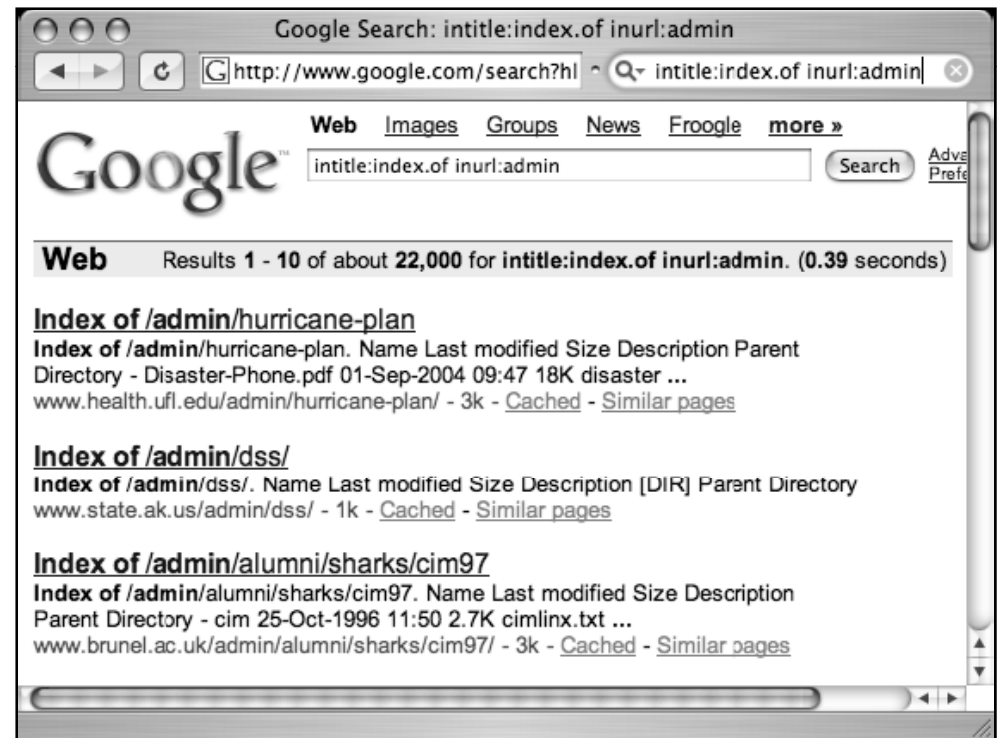


Finding Specific Directories

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



Finding Specific Files

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

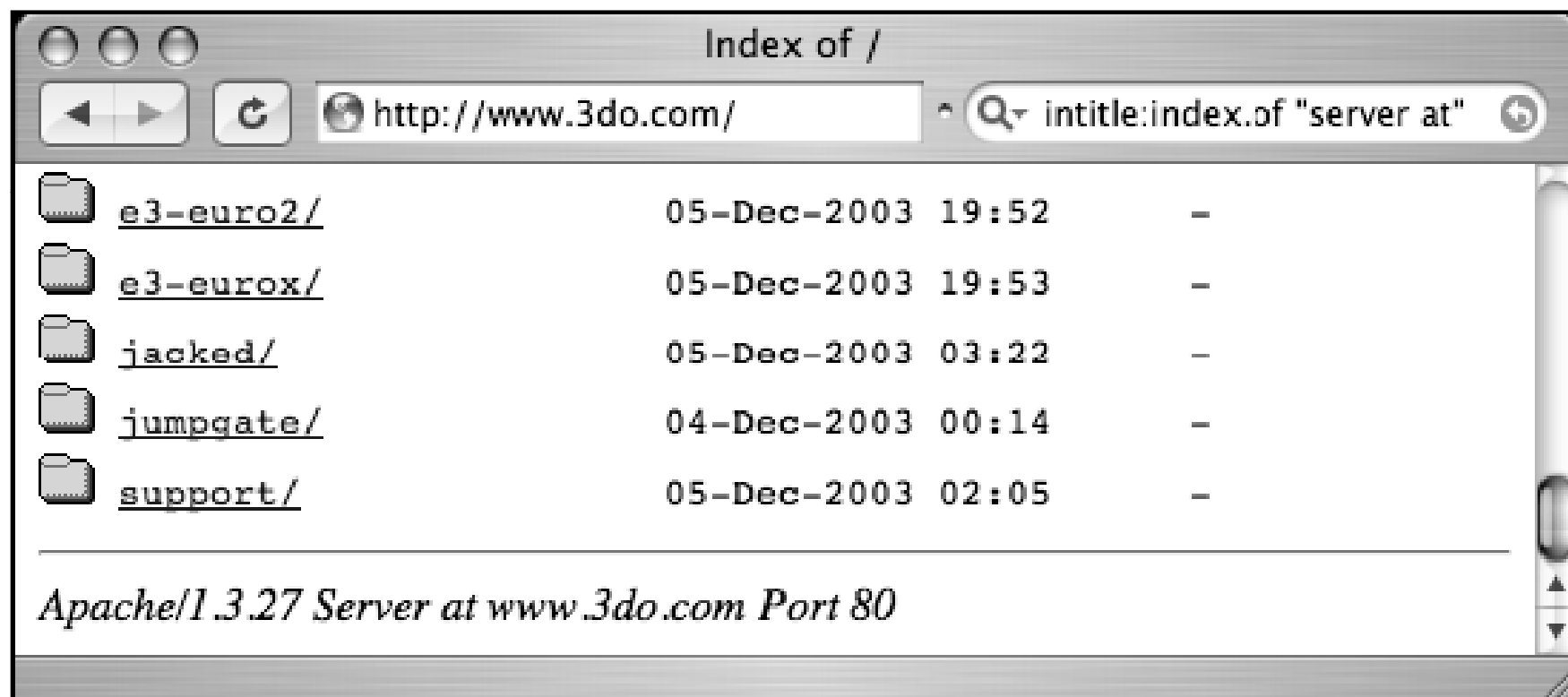
Some typical directory listings provide the name of the server software as well as the version number at the bottom portion. These information are faked and attack can be done on web server

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

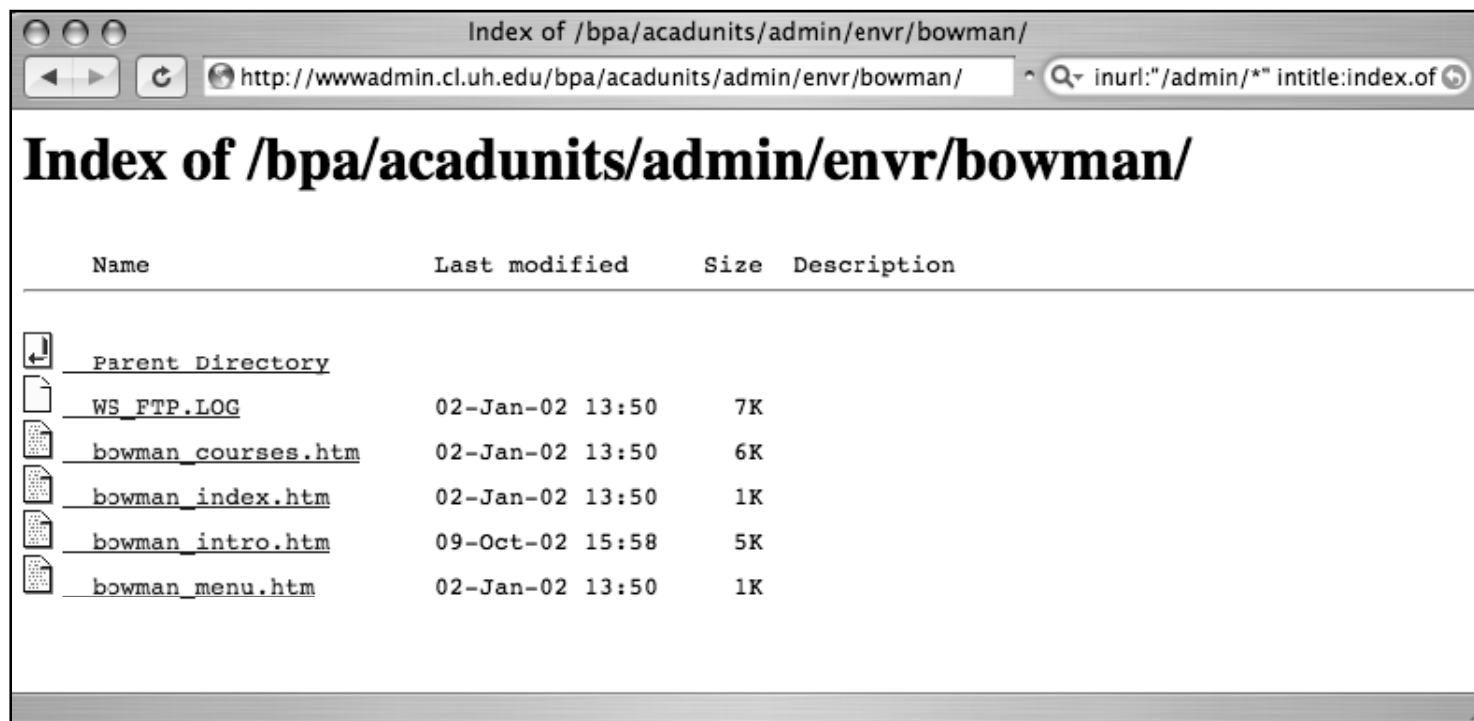
Server Versioning (cont'd)



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



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

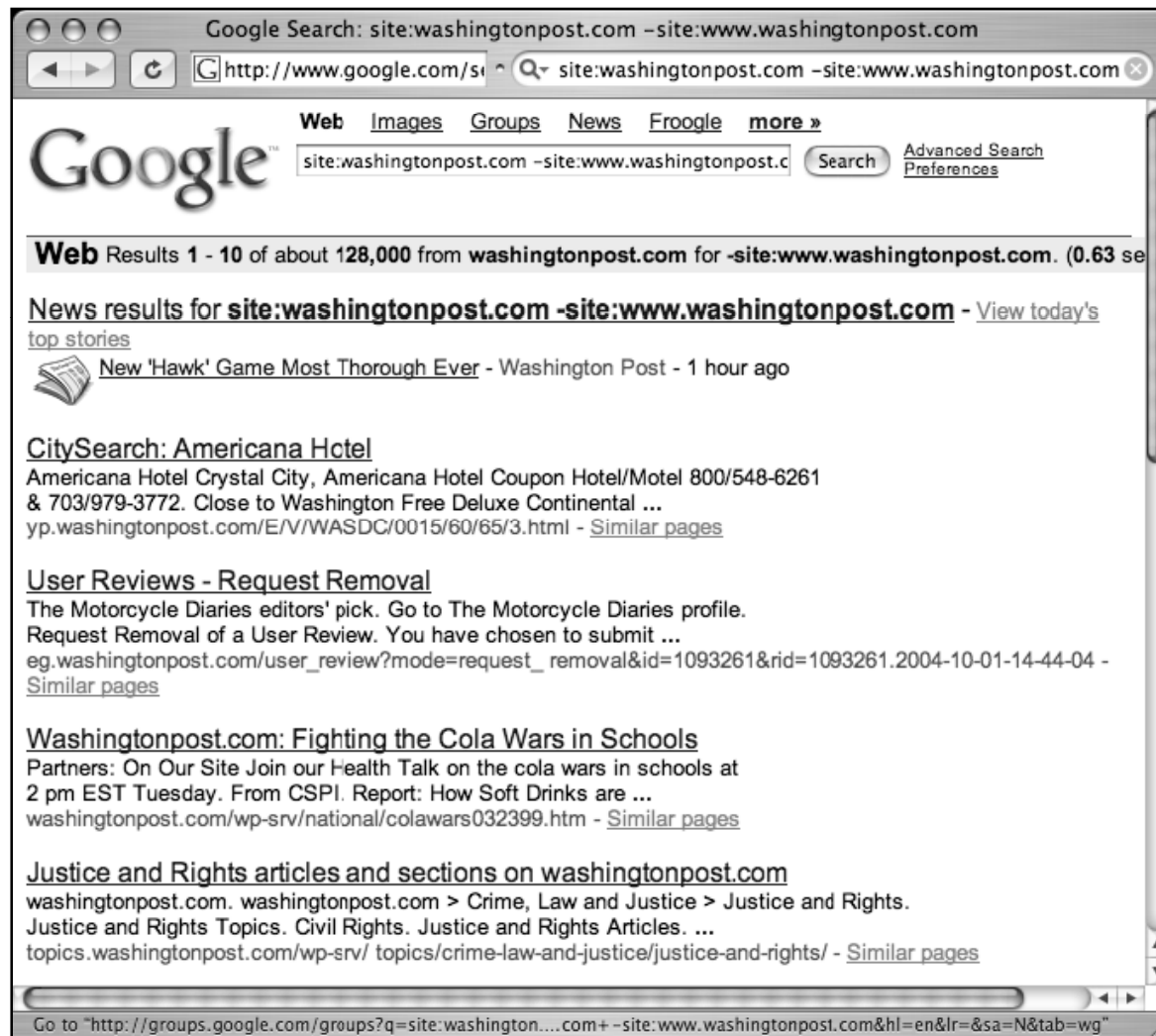
Using simple reduction techniques, you can quickly get an idea about a target's online presence

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)



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

Screenshot

[Web](#) [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more ▼](#)



[Advanced Search](#)
[Preferences](#)

Web

Results 1

[\[PDF\] Google Hacking 101](#)

File Format: PDF/Adobe Acrobat - [View as HTML](#)

Adobe Portable Document Format (**pdf**). • Adobe PostScript (ps). • Lotus 1-2-3 (wk1, wk2, **intitle:index**.of.etc. This search gets you access to the etc ...

[www.nebraskacert.org/CSF/CSF-Jun2005.pdf](#) - [Similar pages](#)

[Intitle Index Of Bluetooth Hack Pdf](#)

Intitle Index Of Bluetooth Hack Pdf. 0 keygen intitleindex. Mp3 mp4 download free symbian movie download free tuneup utilities 2005 crack for bwgen chets to ...

[intitleindexofbluetooth91107.blogspot.com/](#) - 26k - [Cached](#) - [Similar pages](#)

[Videoclips zoeken \"intitle:index of: hack pdf](#)

Je zocht op \"**intitle:index of: hack pdf** maar er werd niets gevonden! Terug - i feel myself rapidshare | **intitle:\"index.of\"** (mp3) yael naim | **Intitle:Index ...**

[www.top-clips.nl/home/woord.php?woord=%22intitle:index%20of:%20hack%20pdf](#) - 19k - [Cached](#) - [Similar pages](#)

[Intitle:google Hacking | Free eBooks Download](#)

Search Terms: free ebooks illegal, **intitle index of pdf** architecture html htm php asp cf jsp, intext google **hacking**, google book download **hack**, google **hack ...**

[freeebookdownload.net/pdf/intitle-google-hacking.shtml](#) - 11k - [Cached](#) - [Similar pages](#)

[Boediger.net intitle index of hack msn warez direct download links](#)

Best **intitle index of hack** msn the results from Google. MSN, Yahoo, Altavista fast and more **intitle index of hack** msn results - added 2008-01-26 ...

error | warning



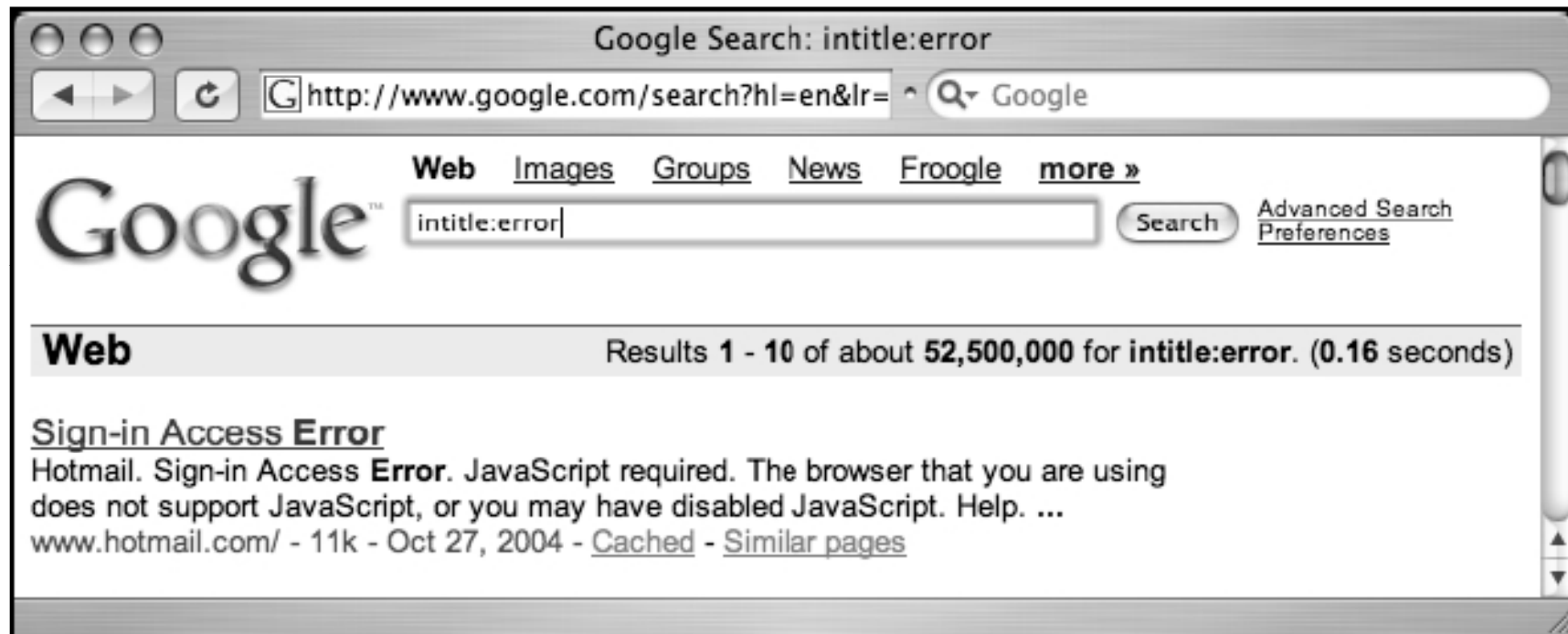
Error messages can reveal a great deal of information about a target

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

error | warning (cont'd)



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

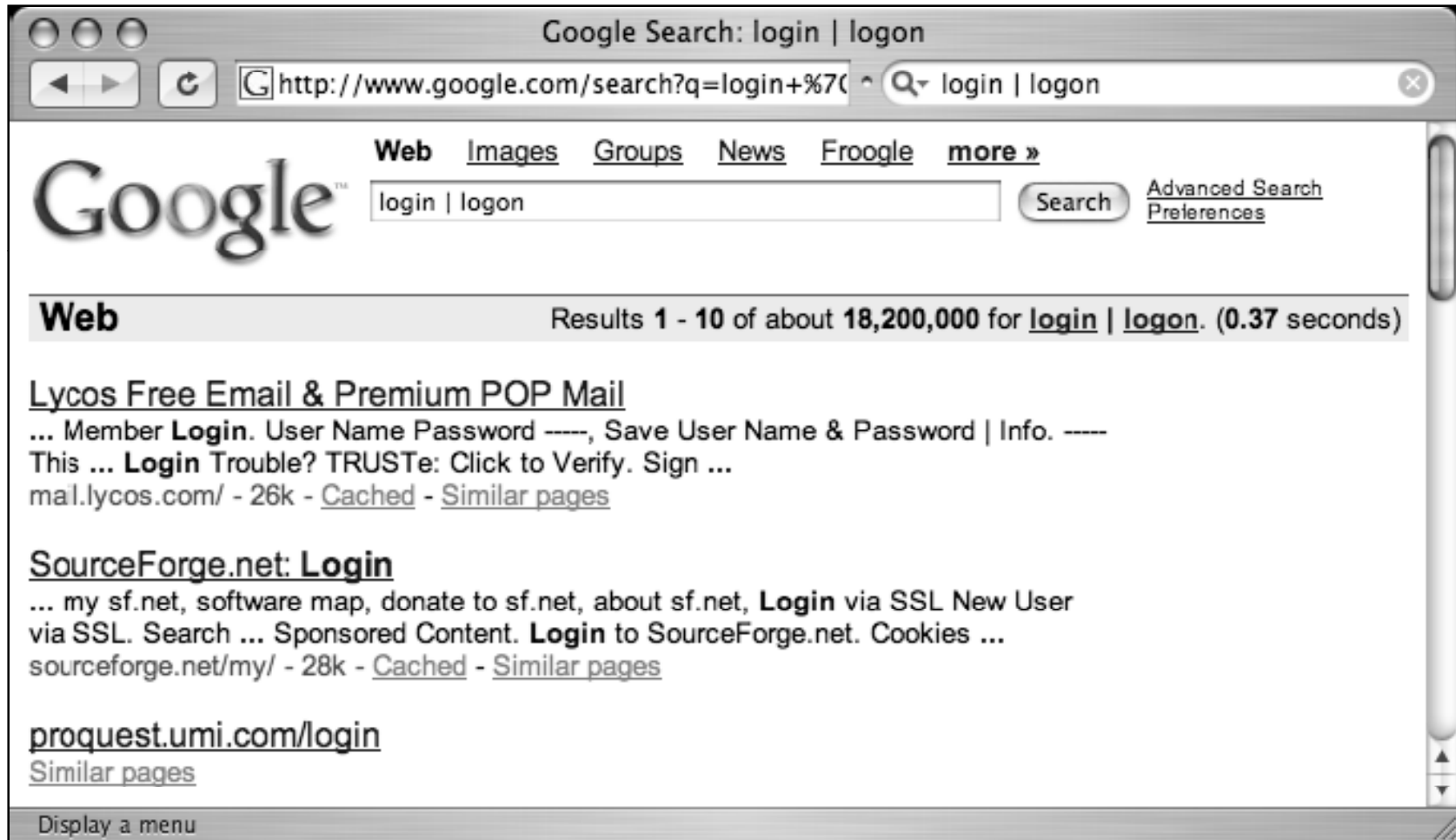
Whether the user has forgotten his or her password or even username, this document can provide clues that might help an attacker

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



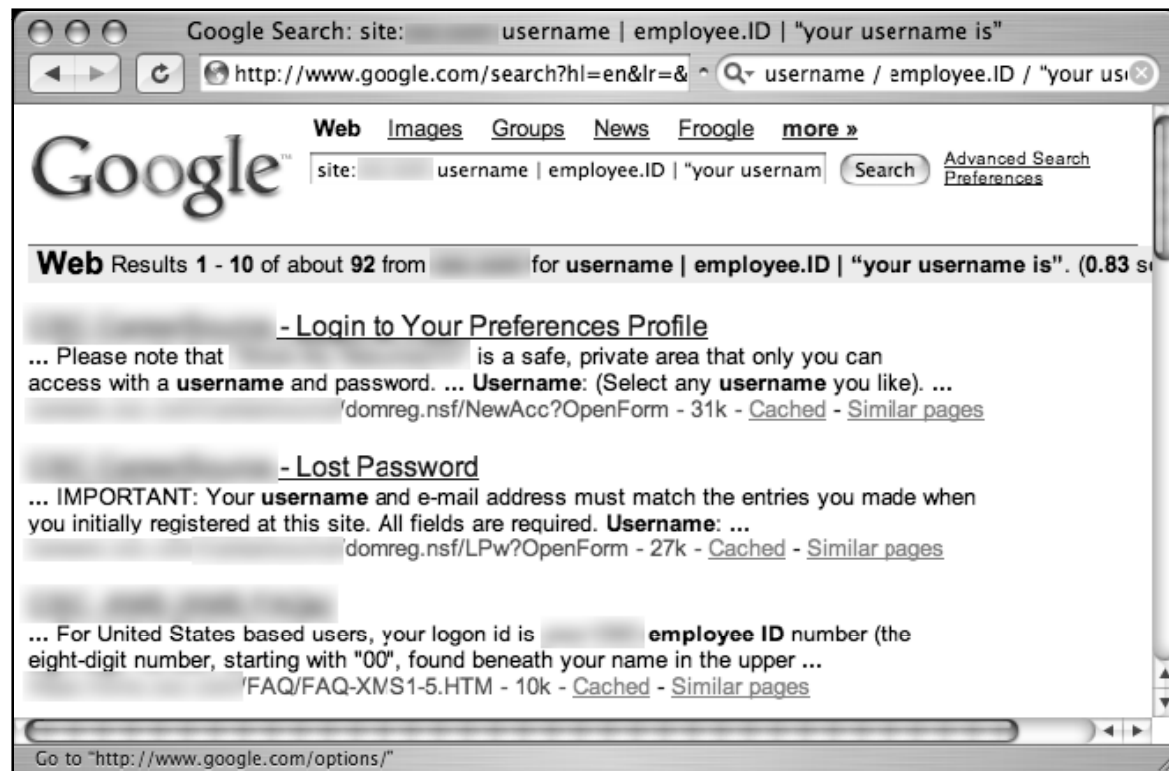
login | logon (cont'd)



username | userid | employee.ID | “your username is”

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



password | passcode | “your password is”

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

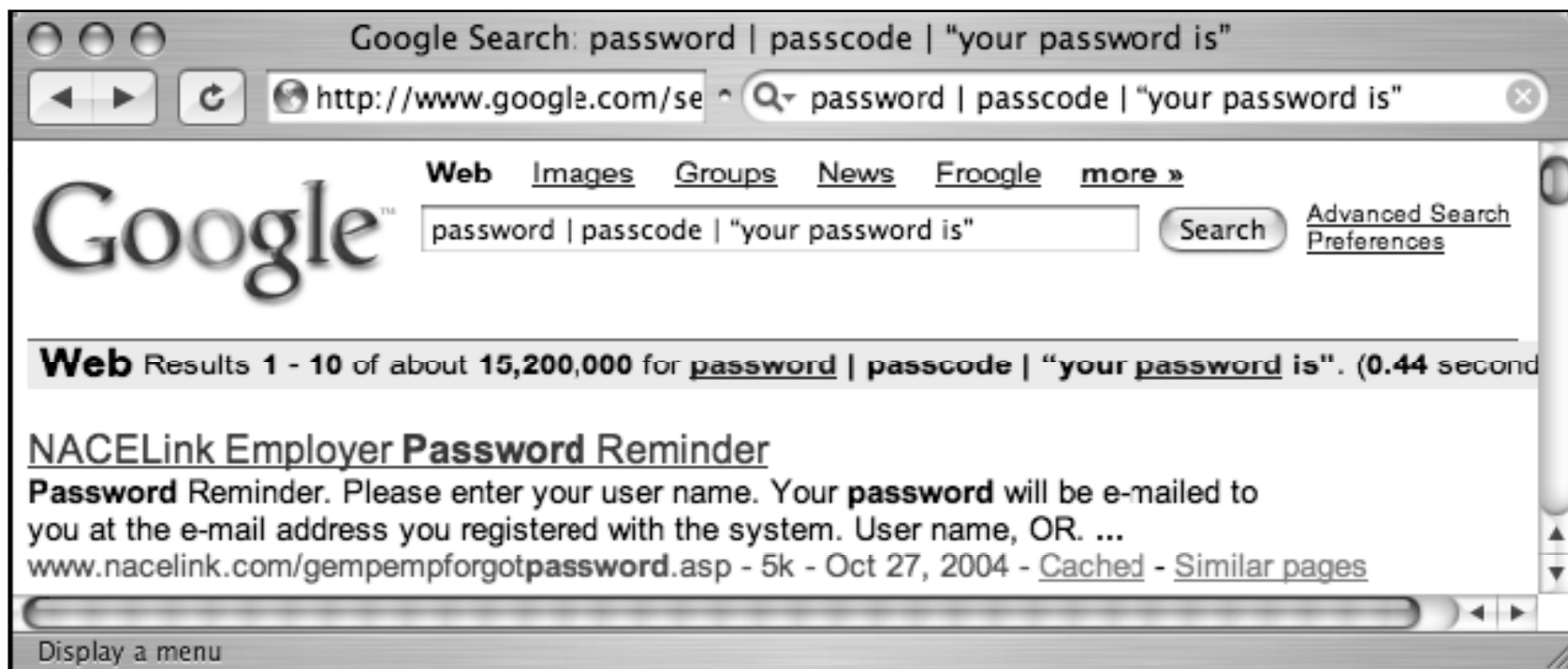
During an assessment, it is very likely that results for this query combined with a site operator will include pages that provide help to users who have forgotten their passwords

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



password | passcode | “your password is” (cont’d)



admin | administrator

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

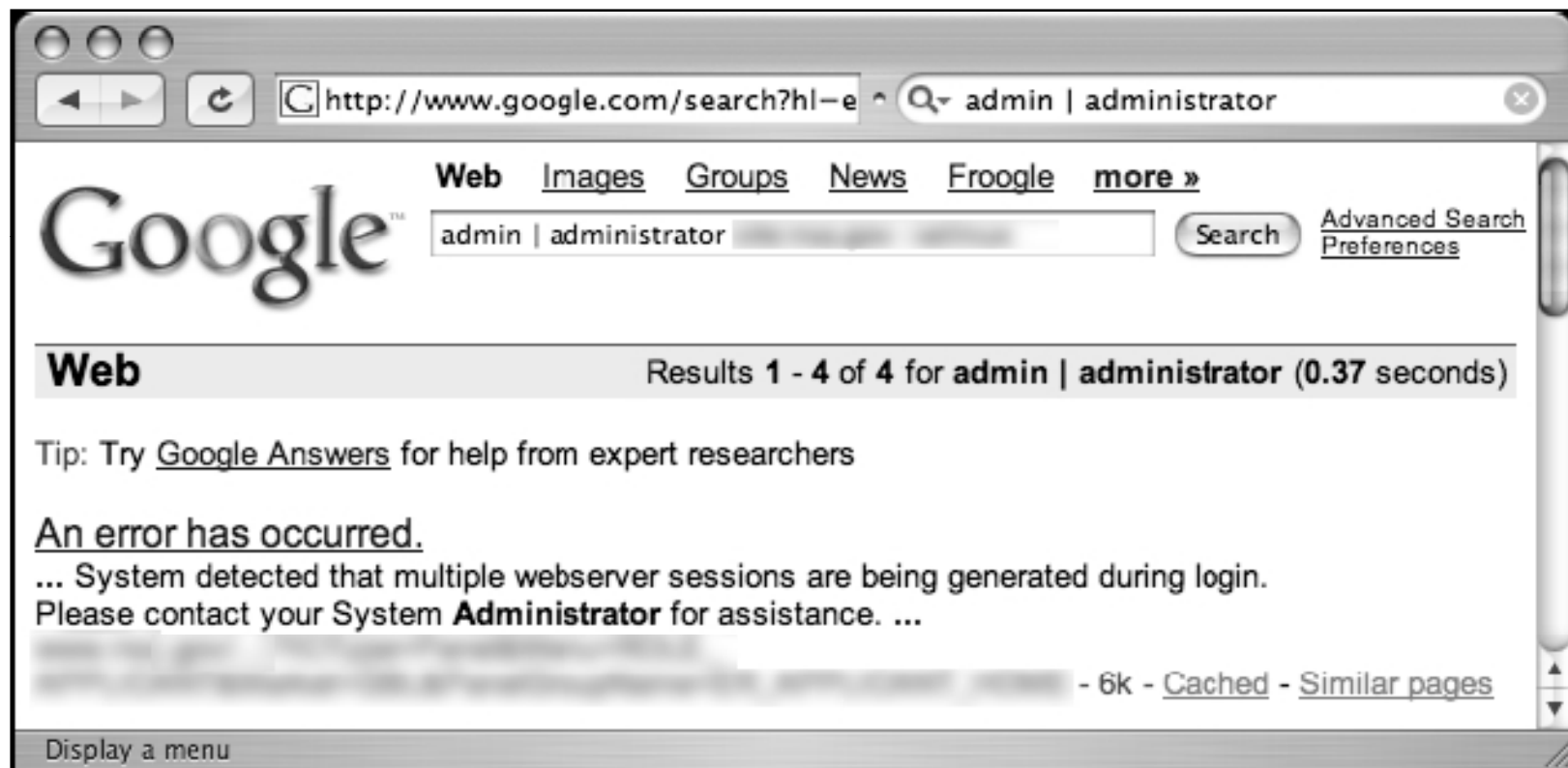
The word administrator can also be used to locate administrative login pages, or login portals

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

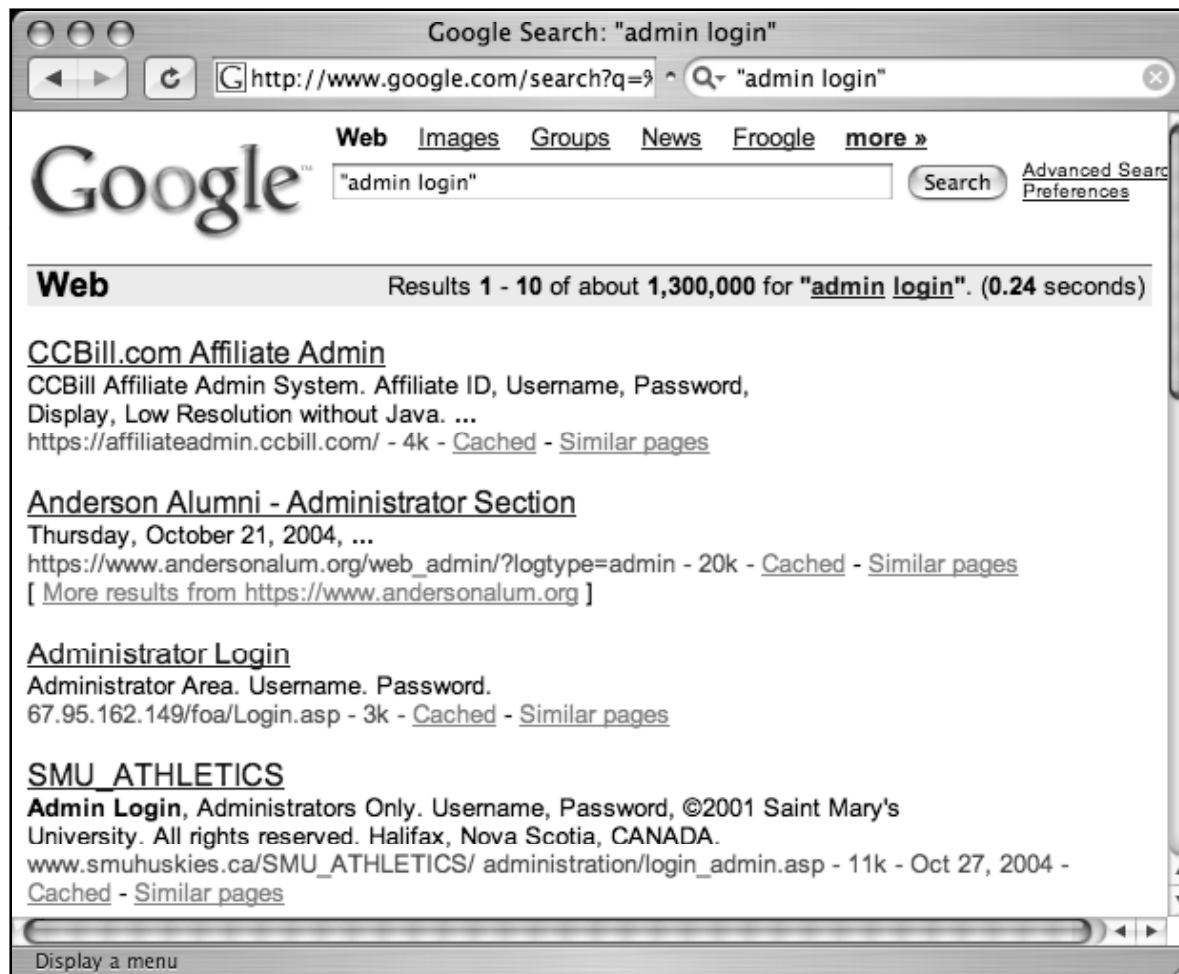
A query such as “please contact your * administrator” will return results that reference local, company, site, department, server, system, network, database, e-mail, and even tennis administrators

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





admin login Reveals Administrative Login Pages

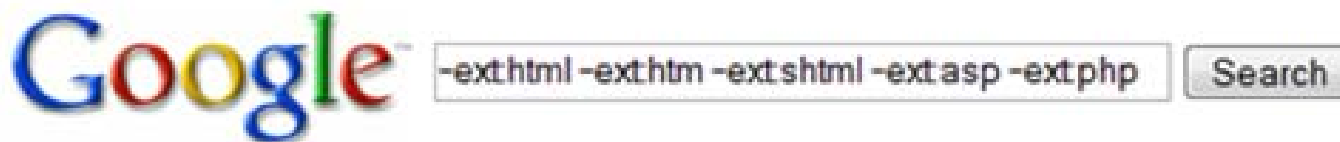


`–ext:html –ext:htm
–ext:shtml –ext:asp –ext:php`

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



–ext:html –ext:htm –ext:shtml –
ext:asp –ext:php (cont'd)



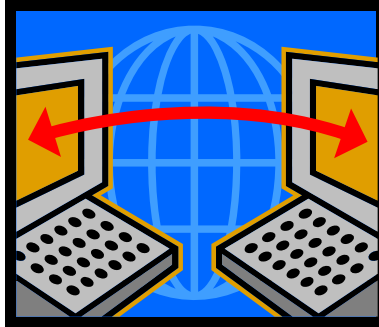
`inurl:temp` | `inurl:tmp` | `inurl:backup` | `inurl:bak`

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`



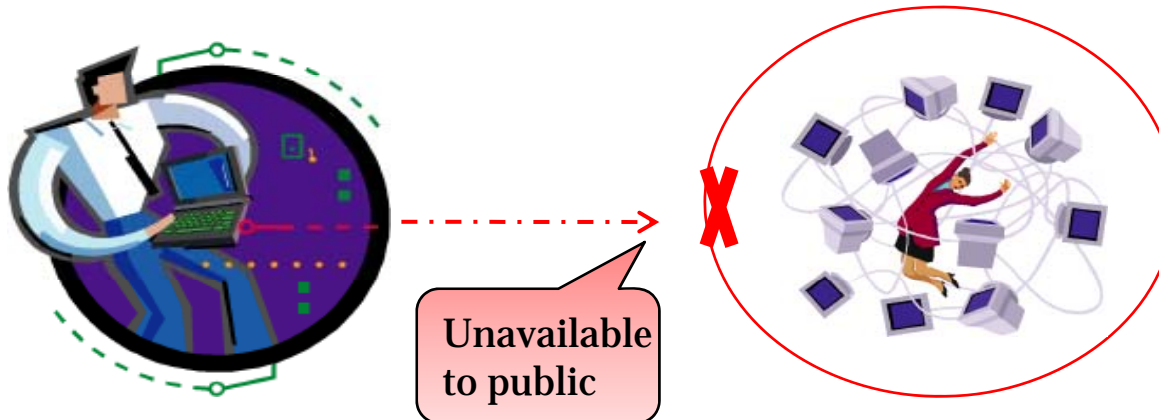


Pre-Assessment

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





Locating Exploits and Finding Targets

Locating Public Exploit Sites

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`

Searching for Exploit Code with Nonstandard Extensions



Locating Source Code with Common Strings

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

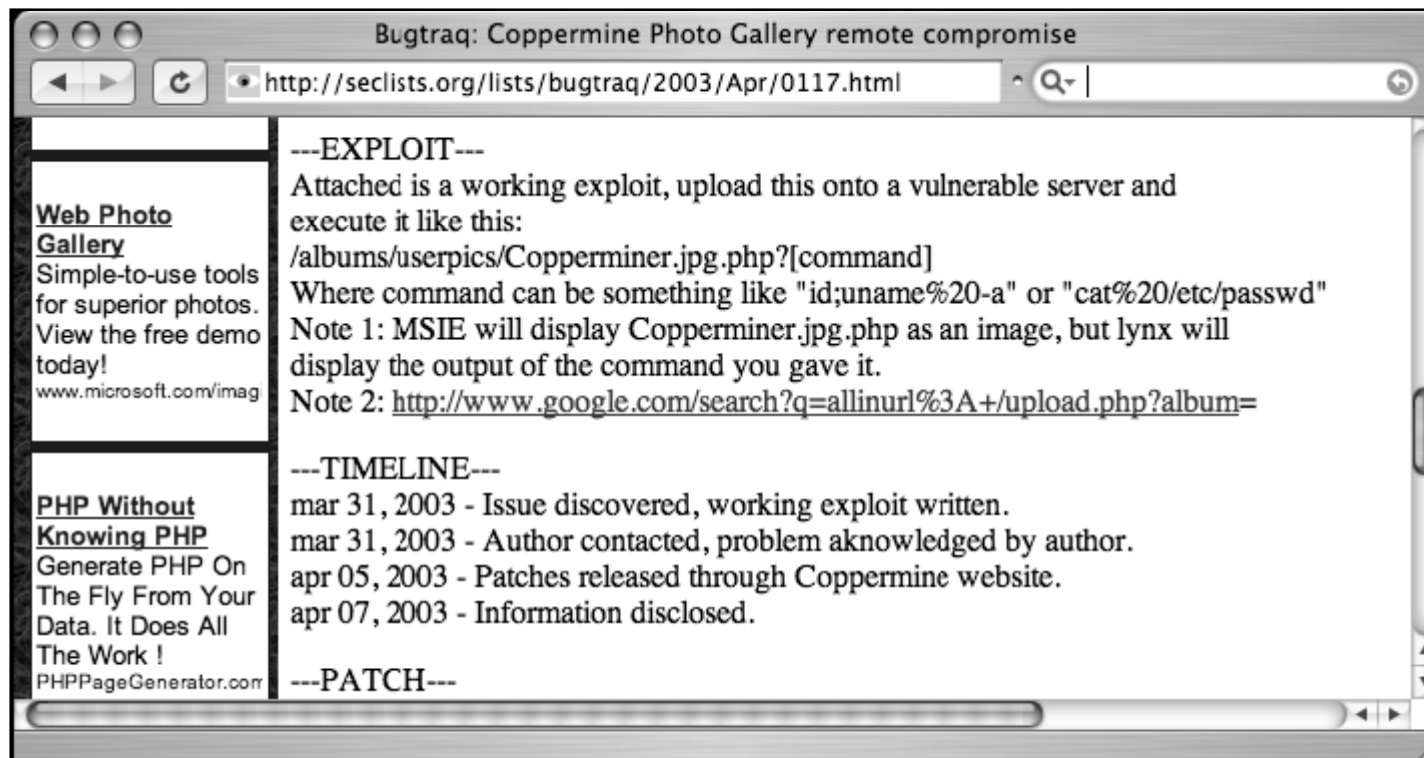
Perl	PERL, PL, PM	<i>"#!/usr/bin/perl"</i>
Python	Py	<i>"#!/usr/bin/env"</i>
VBScript	.vbs	<i>"<%@ language="vbscript" %>"</i>
Visual Basic	Vb	<i>"Private Sub"</i>

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



Locating Targets Via Demonstration Pages

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



“Powered by” Tags Are Common Query Fodder for Finding Web Applications



Locating Targets Via Source Code

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



Vulnerable Web Application Examples

Query	Vulnerability
"Powered by A-CART"	A-CART 2.x vulnerable to cross-site scripting
<i>inurl:"dispatch.php?atknodetype" inurl:class.atkdateattribute.js.php</i>	Achievo .8.x could allow remote code execution
<i>intitle:guestbook "advanced guestbook 2.2 powered"</i>	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
<i>"BlackBoard 1.5.1-f © 2003-4 by Yves Goergen"</i>	BlackBoard 1.5.1 has a remote file inclusion vulnerability
<i>"BosDates Calendar System " "powered by BosDates v3.2 by BosDev"</i>	BosDates 3.2 is vulnerable to SQL injection
<i>inurl:changepassword.cgi -cvs</i>	changepassword.cgi allows for unlimited repeated failed login attempts
<i>"Copyright © 2002 Agustin Dondo Scripts"</i>	CoolPHP 1.0 has multiple vulnerabilities
"Powered by CubeCart 2.0.1"	CubeCart 2.0.1 has an SQL injection vulnerability
<i>"Powered *: newtelligence" ("dasBlog 1.6" "dasBlog 1.5" "dasBlog 1.4" "dasBlog 1.3")</i>	DasBlog versions 1.3–1.6 are susceptible to an HTML injection vulnerability in their request log
<i>"Powered by DCP-Portal v5.5"</i>	DCP-Portal version 5.5 is vulnerable to SQL injection
<i>"2003 DUware All Rights Reserved"</i>	DUForum 3.0 may allow a remote attacker to carry out SQL injection and HTML injection attacks
<i>"inurl:/site/articles.asp?idcategory="</i>	Dwc_Articles 1.6 has multiple input validation problems
<i>inurl:custva.asp</i>	EarlyImpact Productcart v1.5 contains multiple vulnerabilities

<i>inurl:"/becomcommunity/community/index.php?pageurl="</i>	E-market prior to 1.4.0 contains various vulnerabilities
<i>intitle:"EMUMAIL - Login" "Powered by EMU Webmail"</i>	EMU Webmail 5.6 messaging product is susceptible to a cross-site scripting vulnerability
<i>"Powered by FUDforum"</i>	FUDforum 2.0.2 allows manipulation of arbitrary server files
<i>"1999-2004 FuseTalk Inc" -site:fusetalk.com</i>	FuseTalk forums (v4) are susceptible to cross-site scripting attacks
<i>"Powered by My Blog" intext:"FuzzyMonkey.org"</i>	FuzzyMonkey 2.11 has an SQL injection vulnerability
<i>"Powered by Gallery v1.4.4"</i>	Gallery 1.4.4 allows remote code execution
<i>intitle:gallery inurl:setup "Gallery configuration"</i>	Gallery default configuration files allow gallery modification
<i>inurl:"messageboard/Forum.asp?"</i>	GoSmart Message Board (specific versions) are susceptible to SQL injection attack and cross-site scripting attack
<i>intitle:welcome.to.horde</i>	Horde Mail prior to 2.2 has had several reported vulnerabilities
<i>"Powered by IceWarp Software" inurl:mail</i>	IceWarp Web Mail (versions prior to 5.2.8) is reported prone to multiple input validation vulnerabilities
<i>"Ideal BB Version: 0.1" -idealbb.com</i>	Ideal BB 0.1 is susceptible to multiple vulnerabilities
<i>"Powered by Ikonboard 3.1.1"</i>	IkonBoard 3.1.1 allows cross-site scripting
<i>"Powered by Invision Power Board(U) v1.3 Final © *</i>	Invision Power Board v1.3 is vulnerable to SQL injection
<i>inurl:/wiki/MediaWiki</i>	MediaWiki 1.3.5 has a cross-site scripting vulnerability
<i>"Powered by Megabook *" inurl:guestbook.cgi</i>	MegaBook 2.0 is prone to multiple HTML injection vulnerabilities
<i>"Powered by mnoGoSearch - free Web search engine software"</i>	mnGoSearch 3.1.20 and 3.2.10 contain a buffer overflow vulnerability

Vulnerable Web Application Examples (cont'd)

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

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

Locating Targets Via CGI Scanning

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:

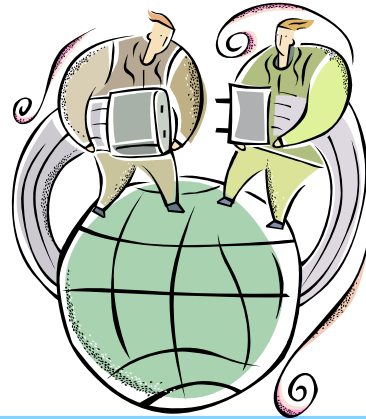
```
/cgi-bin/userreg.cgi  
/cgi-bin/cgiemail/uargg.txt  
/random_banner/index.cgi  
/random_banner/index.cgi  
/cgi-bin/mailview.cgi  
/cgi-bin/maillist.cgi  
/iissamples/ISSamples/SQLQHit.asp  
/iissamples/ISSamples/SQLQHit.asp  
/SiteServer/admin/findvserver.asp  
/scripts/cphost.dll  
/cgi-bin/finger.cgi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



A Single CGI Scan-Style Query

Example: search for *inurl:/cgi-bin/userreg.cgi*

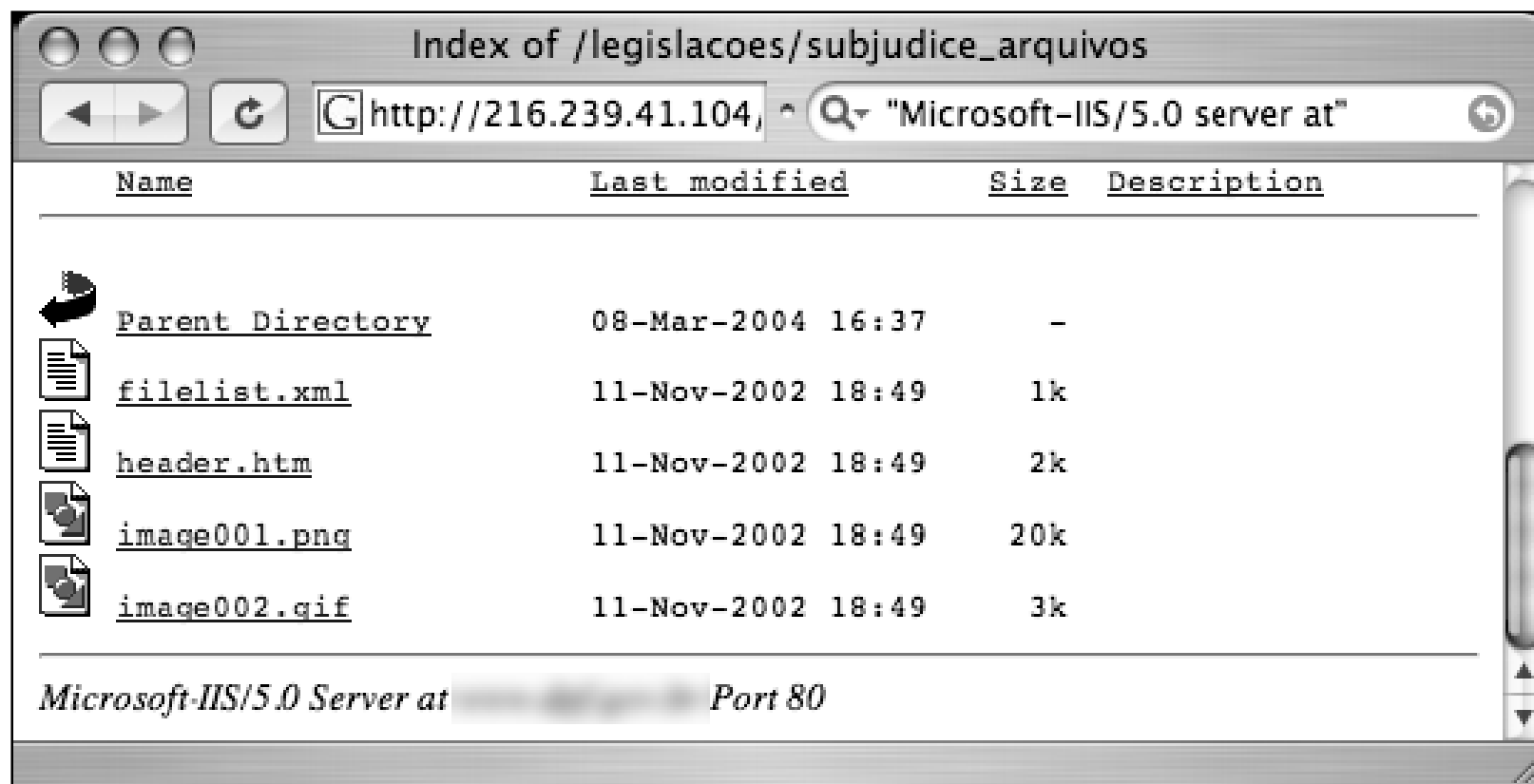




Tracking Down Web Servers, Login Portals, and Network Hardware

Finding IIS 5.0 Servers

Query for *"Microsoft-IIS/5.0 server at"*



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

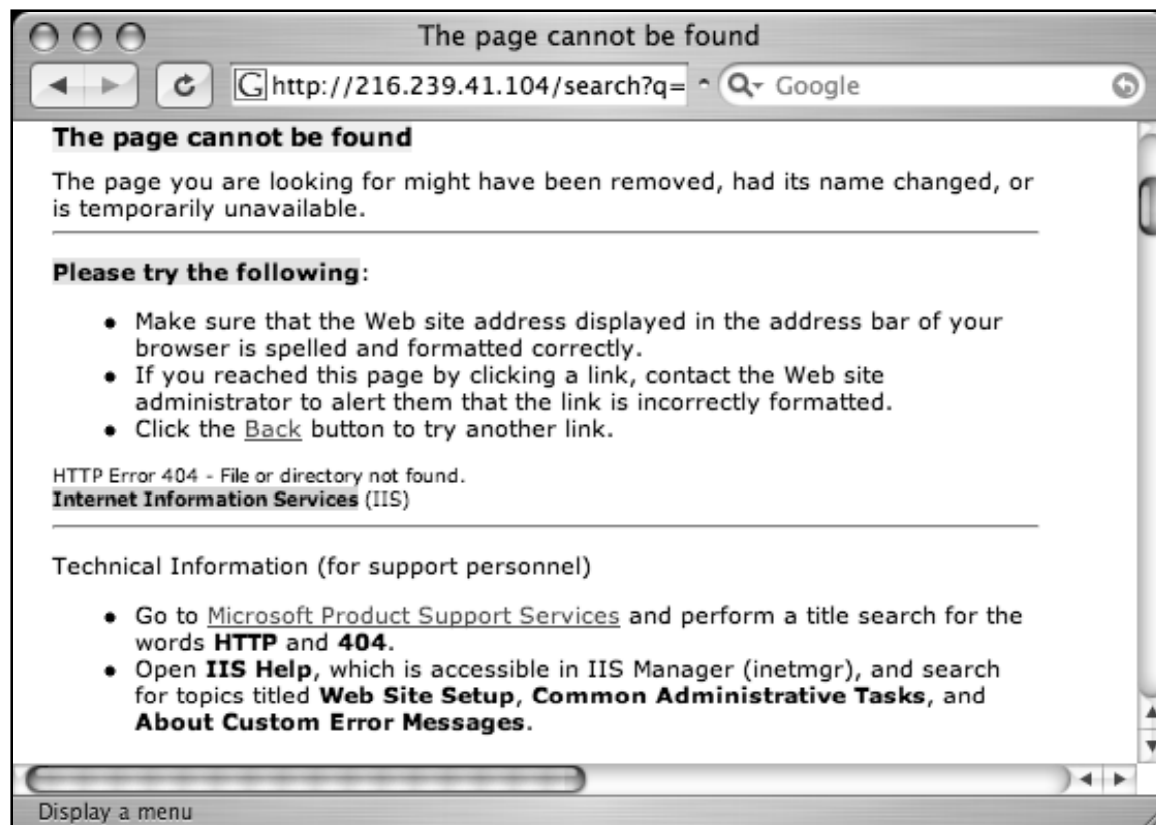
The best way to get this information from IIS is by examining the source code of the error pages themselves

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

Web Server Software Error Messages (cont'd)

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



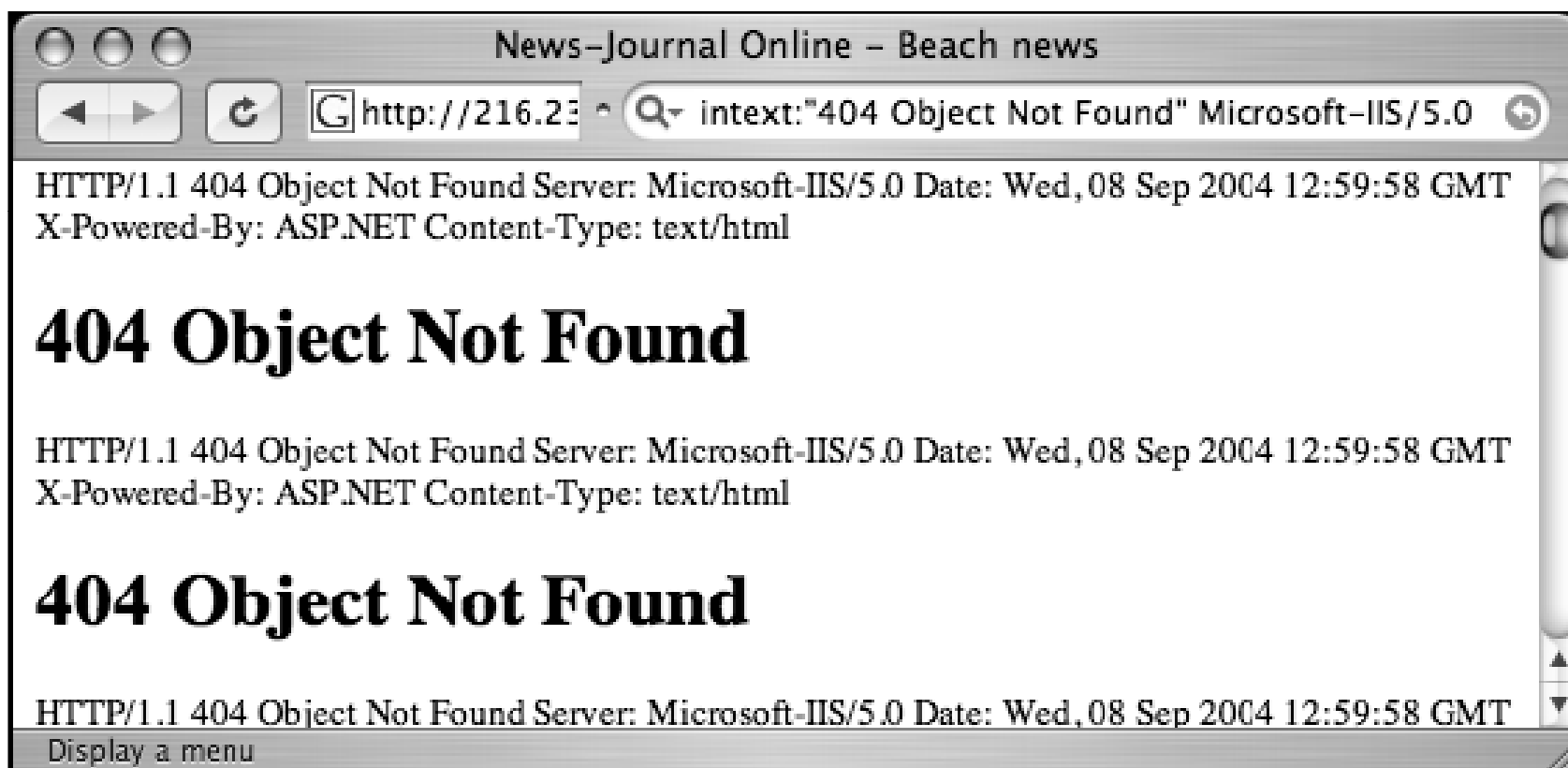
IIS HTTP/1.1 Error Page Titles

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

IIS HTTP/1.1 Error Page Titles (cont'd)

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

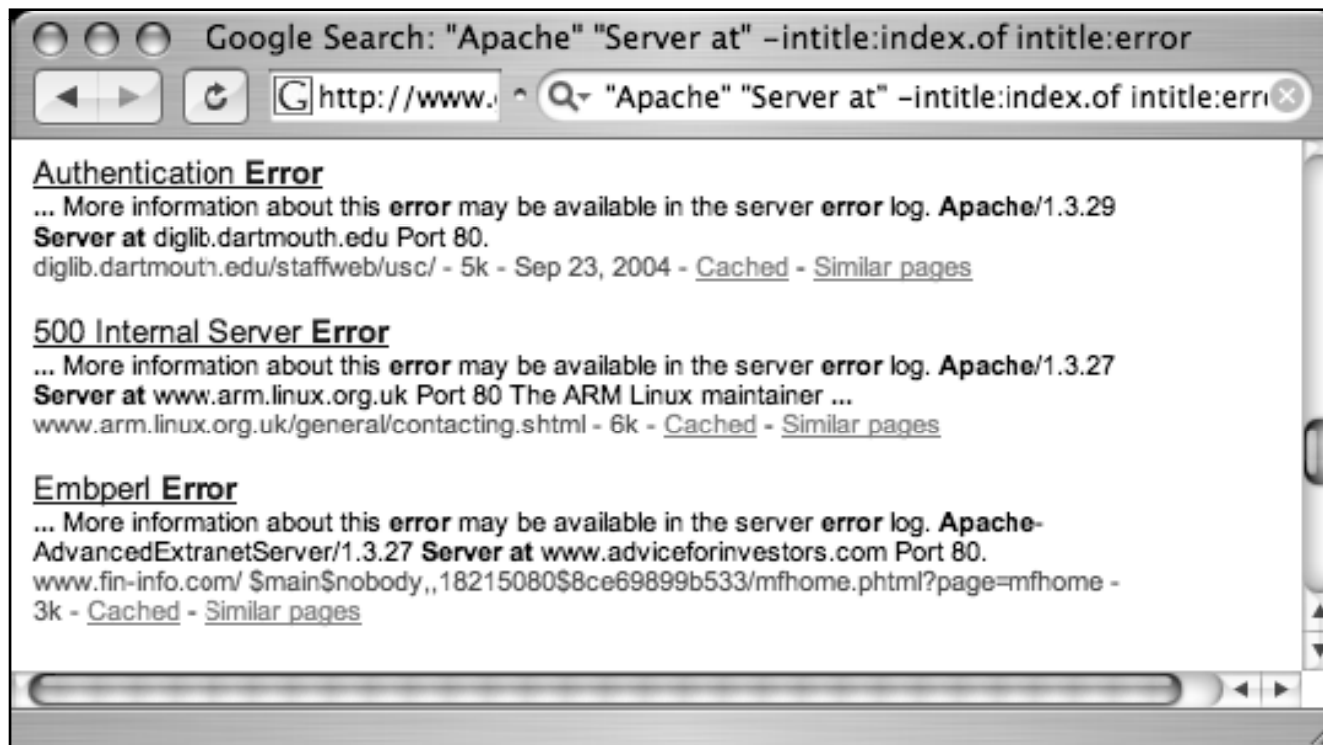
“Object Not Found” Error Message Used to Find IIS 5.0



Apache Web Server

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*



Apache 2.0 Error Pages

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.

Application Software Error Messages



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

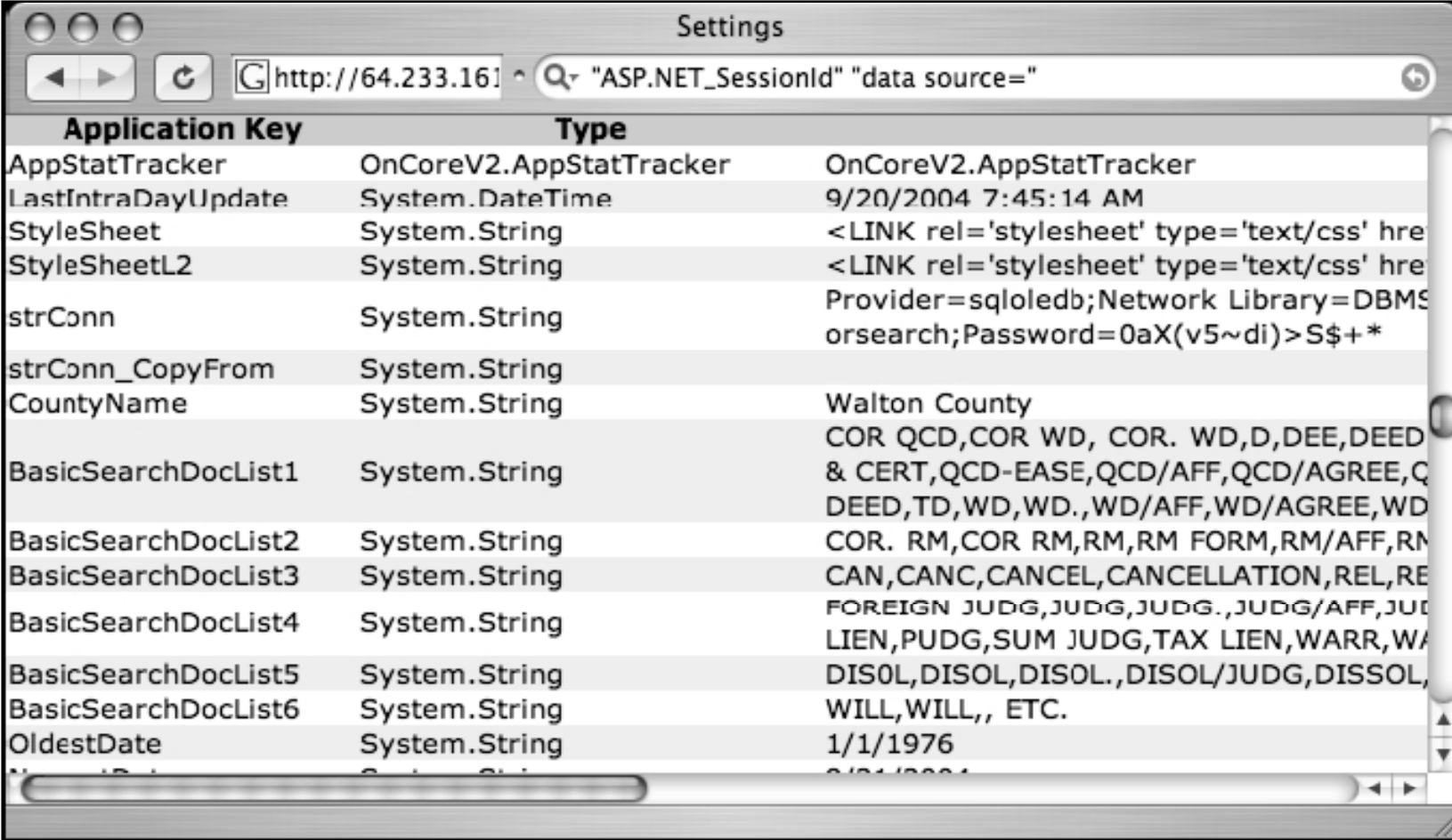
Consider the query *"ASP.NET_SessionId"* *"data source="*, which locates unique strings found in ASP.NET application state dumps



These dumps reveal all sorts of information about the running application and the Web server that hosts that application

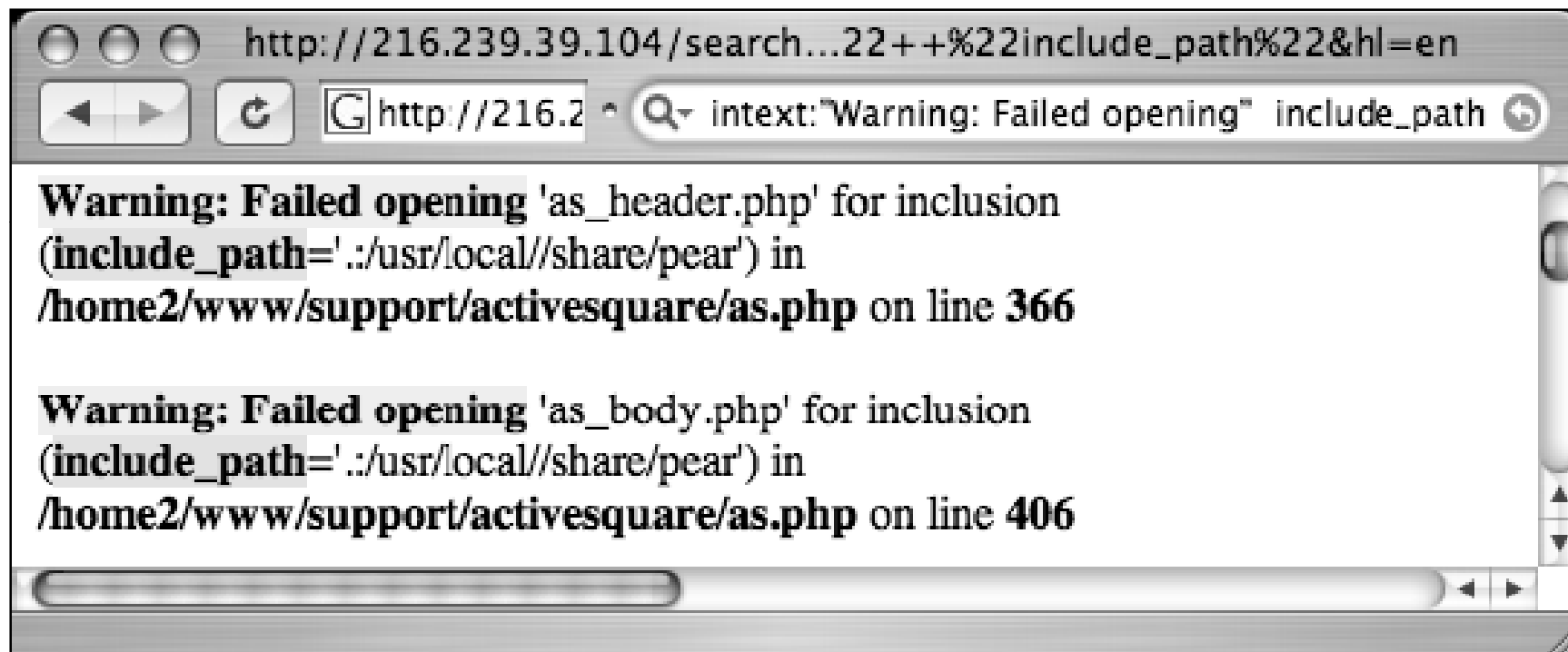
An advanced attacker can use encrypted password data and variable information in these stack traces to subvert the security of the application and perhaps the Web server itself

ASP Dumps Provide Dangerous Details

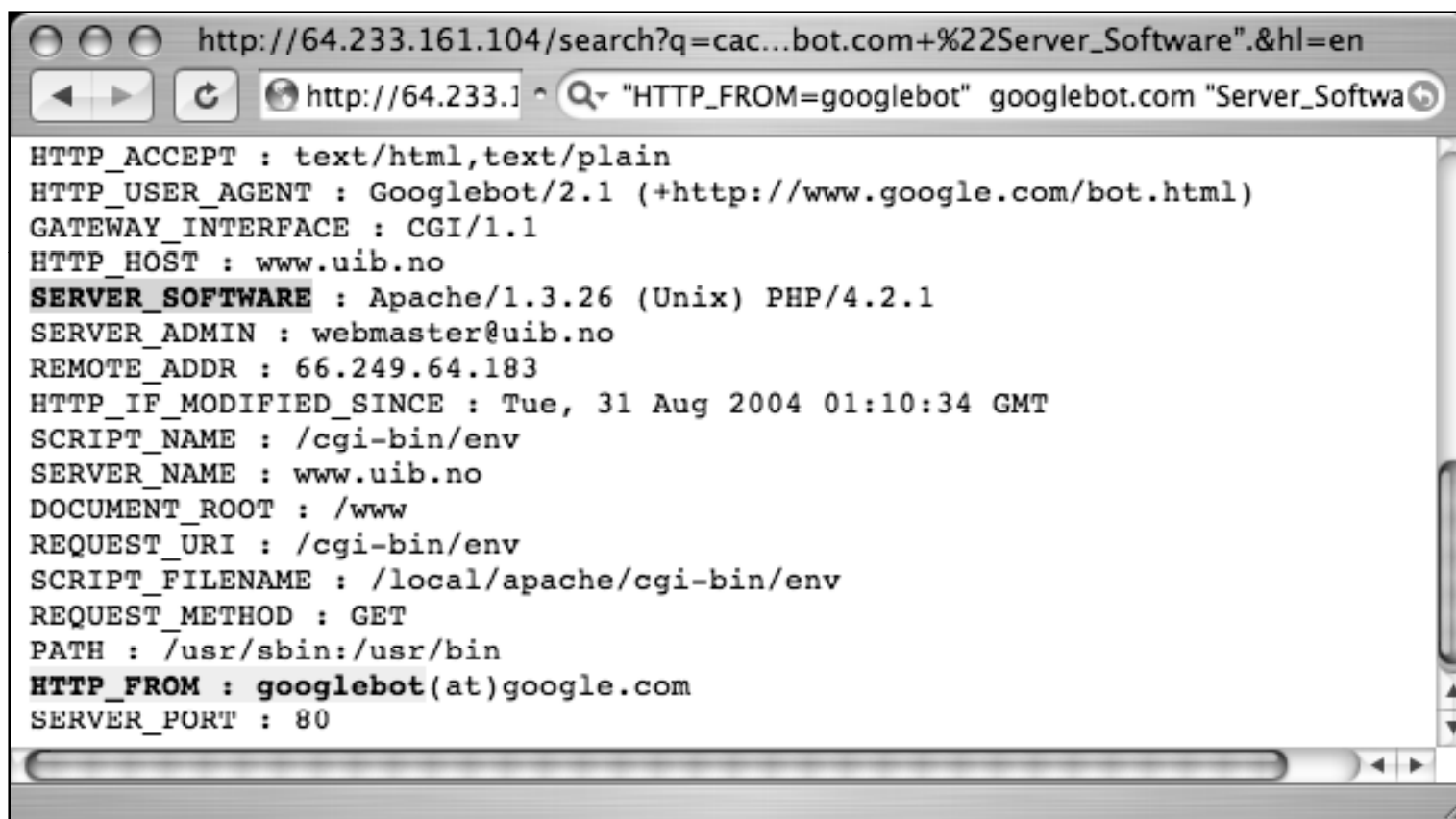


Application Key	Type	Value
AppStatTracker	OnCoreV2.AppStatTracker	OnCoreV2.AppStatTracker
LastIntraDayUpdate	System.DateTime	9/20/2004 7:45:14 AM
StyleSheet	System.String	<LINK rel='stylesheet' type='text/css' href=
StyleSheetL2	System.String	<LINK rel='stylesheet' type='text/css' href=
strConn	System.String	Provider=sqloledb;Network Library=DBMS
strConn_CopyFrom	System.String	orsearch;Password=0aX(v5~di)>S\$+*
CountyName	System.String	Walton County
BasicSearchDocList1	System.String	COR QCD,COR WD, COR. WD,D,DEE,DEED
BasicSearchDocList2	System.String	& CERT,QCD-EASE,QCD/AFF,QCD/AGREE,Q
BasicSearchDocList3	System.String	DEED,TD,WD,WD.,WD/AFF,WD/AGREE,WD
BasicSearchDocList4	System.String	COR. RM,COR RM,RM,RM FORM,RM/AFF,RM
BasicSearchDocList5	System.String	CAN,CANC,CANCEL,CANCELLATION,REL,RE
BasicSearchDocList6	System.String	FOREIGN JUDG,JUDG,JUDG.,JUDG/AFF,JUD
OldestDate	System.String	LIEN,PUDG,SUM JUDG,TAX LIEN,WARR,WA
		DISOL,DISOL,DISOL.,DISOL/JUDG,DISSOL,
		WILL,WILL,, ETC.
		1/1/1976

Many Errors Reveal Pathnames and Filenames



CGI Environment Listings Reveal Lots of Information



The screenshot shows a web browser window with the address bar displaying the URL `http://64.233.161.104/search?q=cac...bot.com+%22Server_Software".&hl=en`. The search bar contains the query `"HTTP_FROM=googlebot" googlebot.com "Server_Softwa`. The main content area displays a list of CGI environment variables in a monospaced font:

```
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
```

Default Pages

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

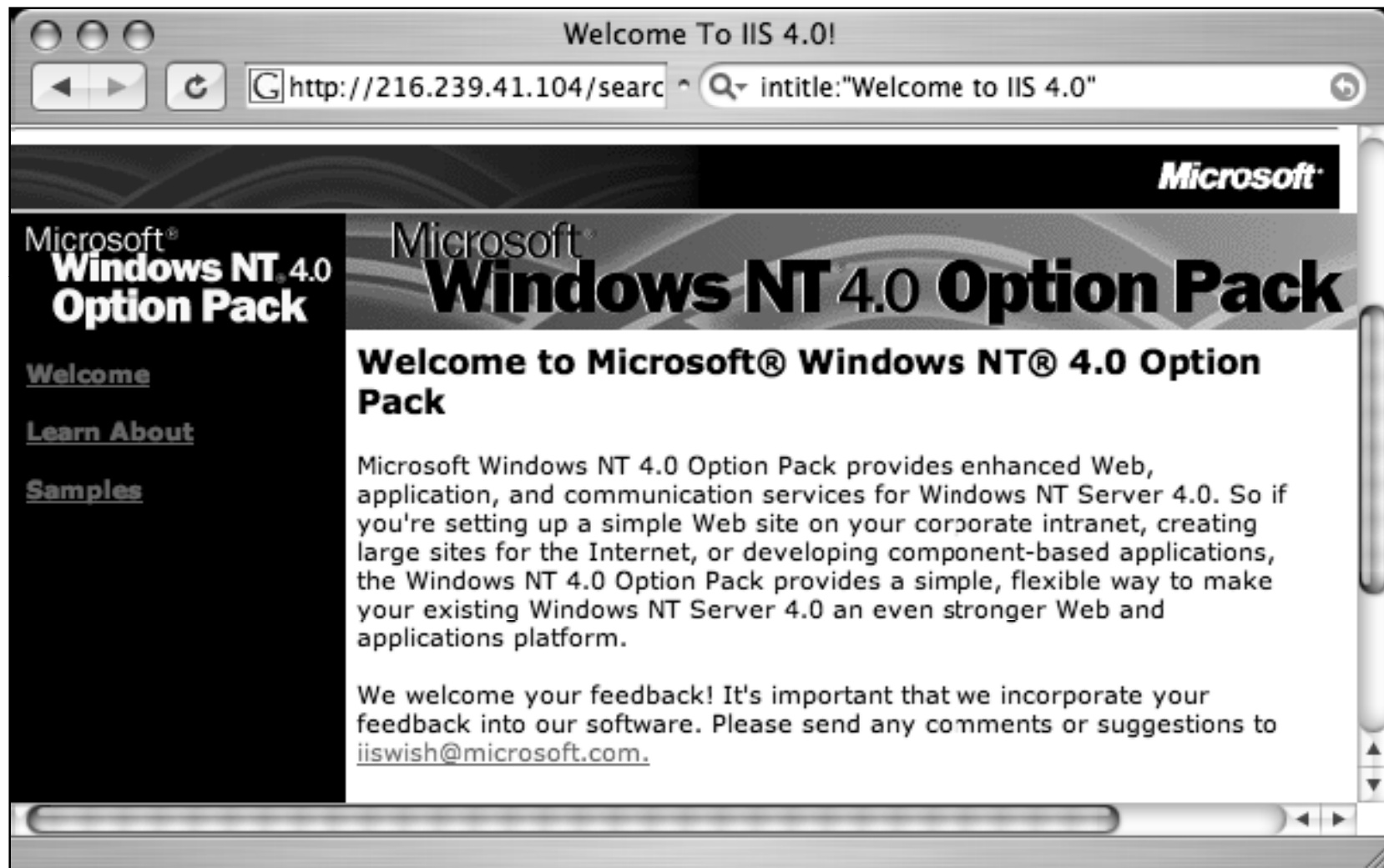
In these cases there is generally a short window of time between the moment when Google crawls the site and when the intended content is actually placed on the server



A Typical Apache Default Web Page



Locating Default Installations of IIS 4.0 on Windows NT 4.0/OP



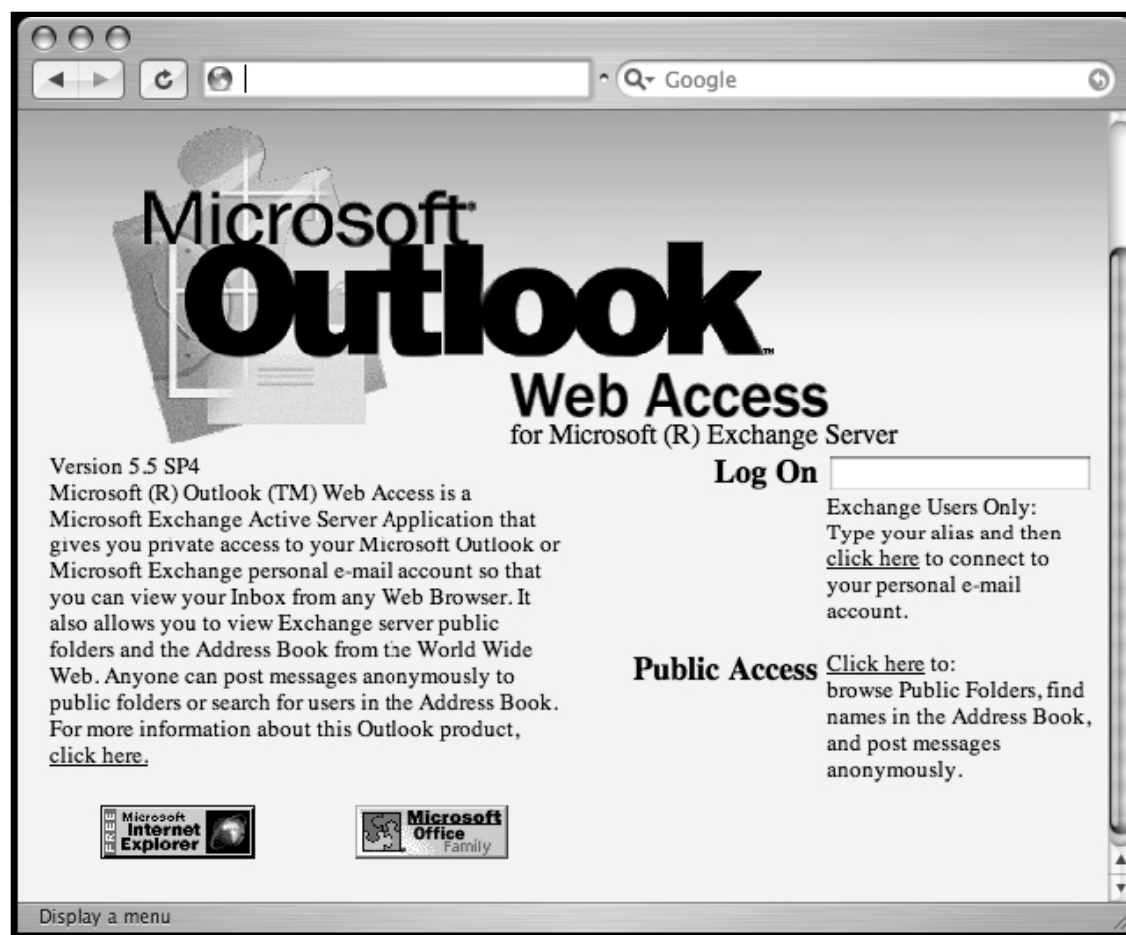
Default Pages Query for Web Server

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

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

Outlook Web Access Default Portal

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

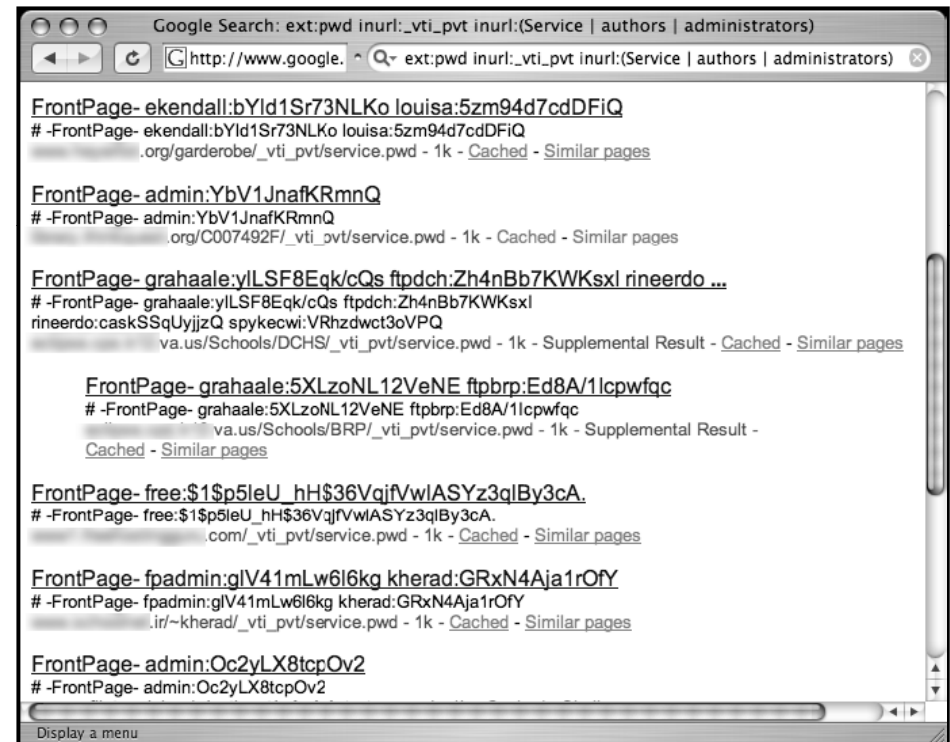


Searching for Passwords

Password data, one of the “Holy Grails” during a penetration test, should be protected

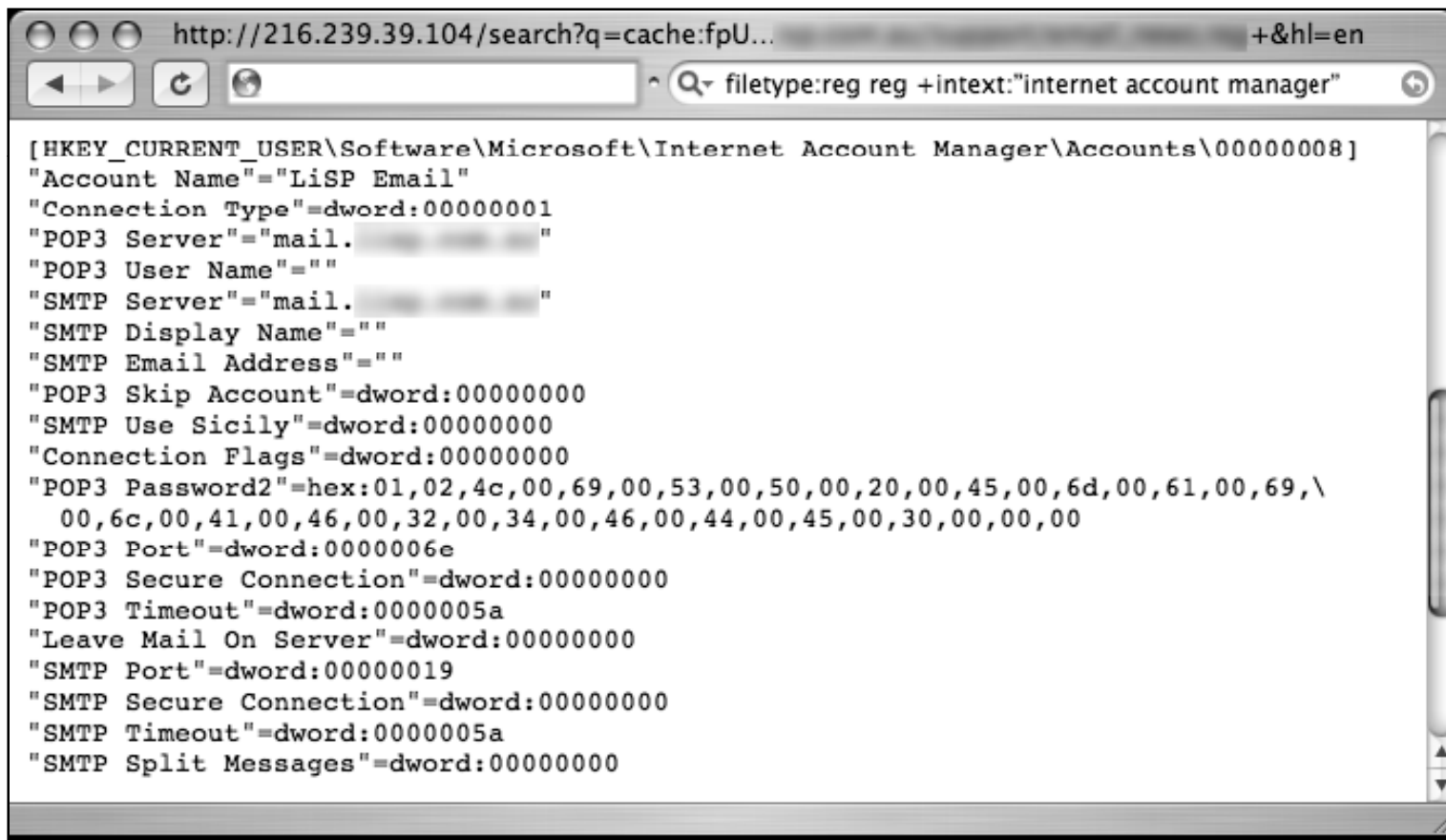


Unfortunately, many examples of Google queries can be used to locate passwords on the Web



Windows Registry Entries Can Reveal Passwords

Query like *filetype:reg intext: "internet account manager"* could reveal interesting keys containing password data



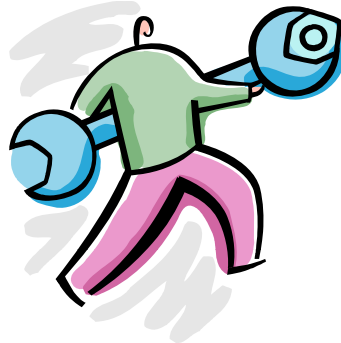
The screenshot shows a web browser window with the address bar displaying `http://216.239.39.104/search?q=cache:fpU...` and a search bar containing `filetype:reg reg +intext:"internet account manager"`. The search results display a registry key path: `[HKEY_CURRENT_USER\Software\Microsoft\Internet Account Manager\Accounts\00000008]`. Below the path, a list of registry values is shown, including:

- "Account Name"="LiSP Email"
- "Connection Type"=dword:00000001
- "POP3 Server"="mail. [REDACTED]"
- "POP3 User Name"=""
- "SMTP Server"="mail. [REDACTED]"
- "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

Username, Cleartext Passwords, and Hostnames!

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
```



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.

SonicWALL® ECLASS
SSL VPN SOLUTIONS

SO MUCH FOR THE STATUS QUO.



SONICWALL®
PROTECTION AT THE
SPEED OF BUSINESS®

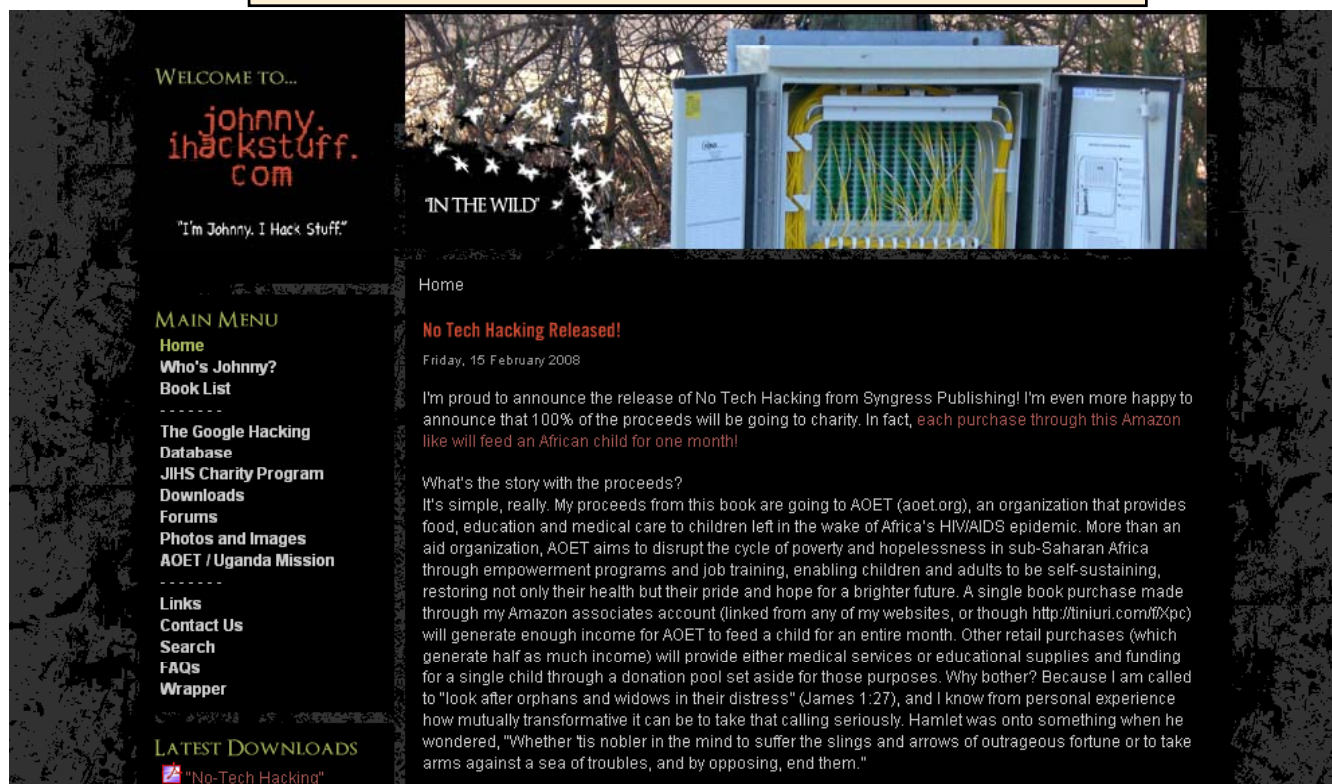
SEE E-CLASS SSL VPN VIDEO NOW.

Source: <http://www.computerworld.com/>

Google Hacking Database (GHDB)

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.

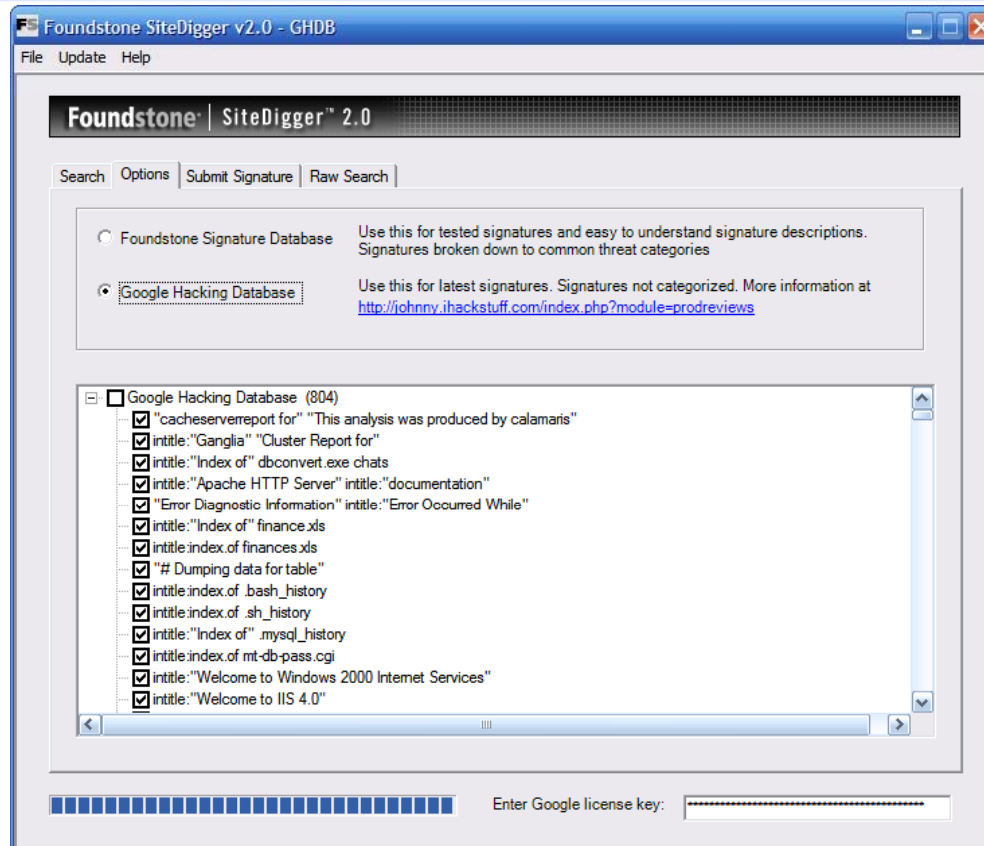
[Pages containing network or vulnerability data](#) (59 entries)

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 is a tool that automates queries against Google search appliances

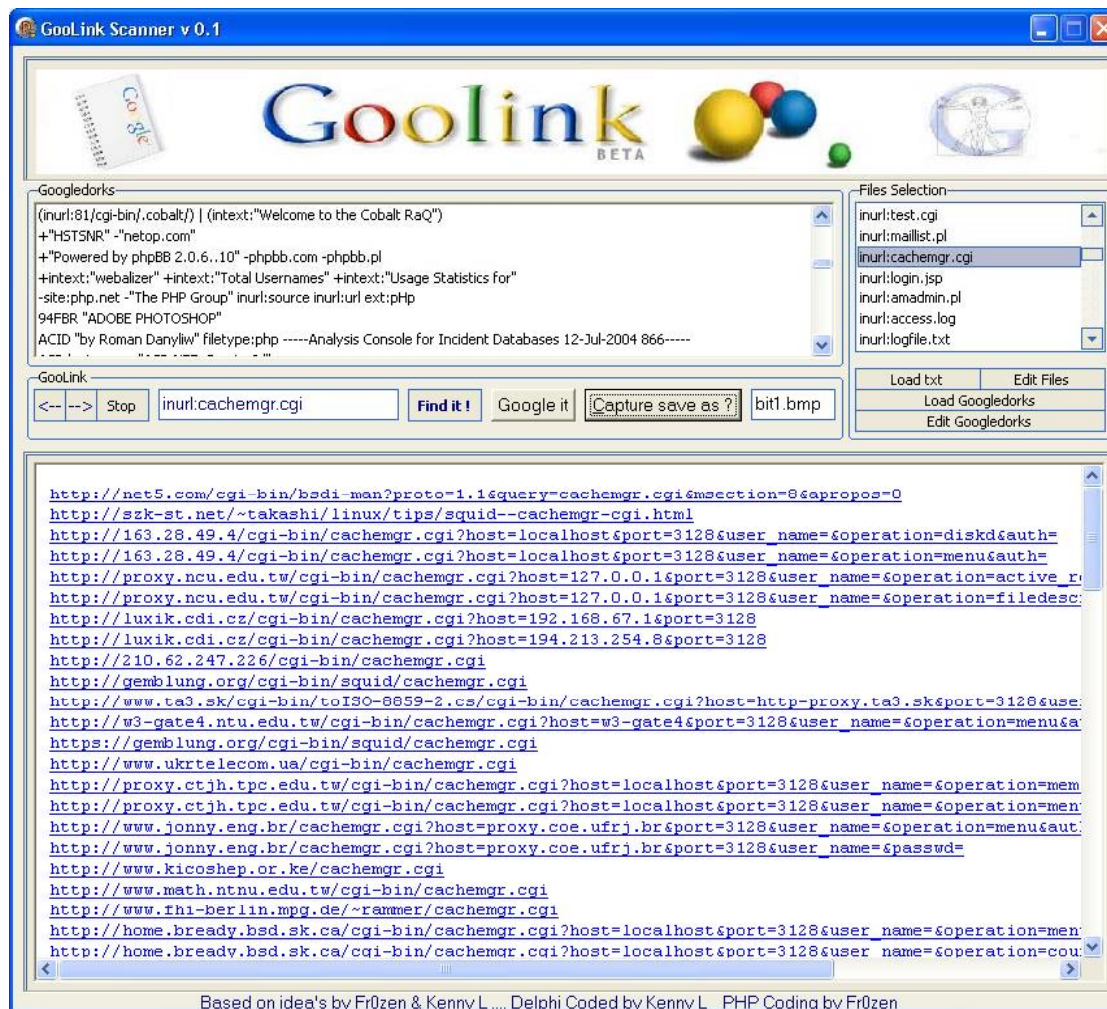
But it can be run against Google itself in direct violation of their Terms of Service

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

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

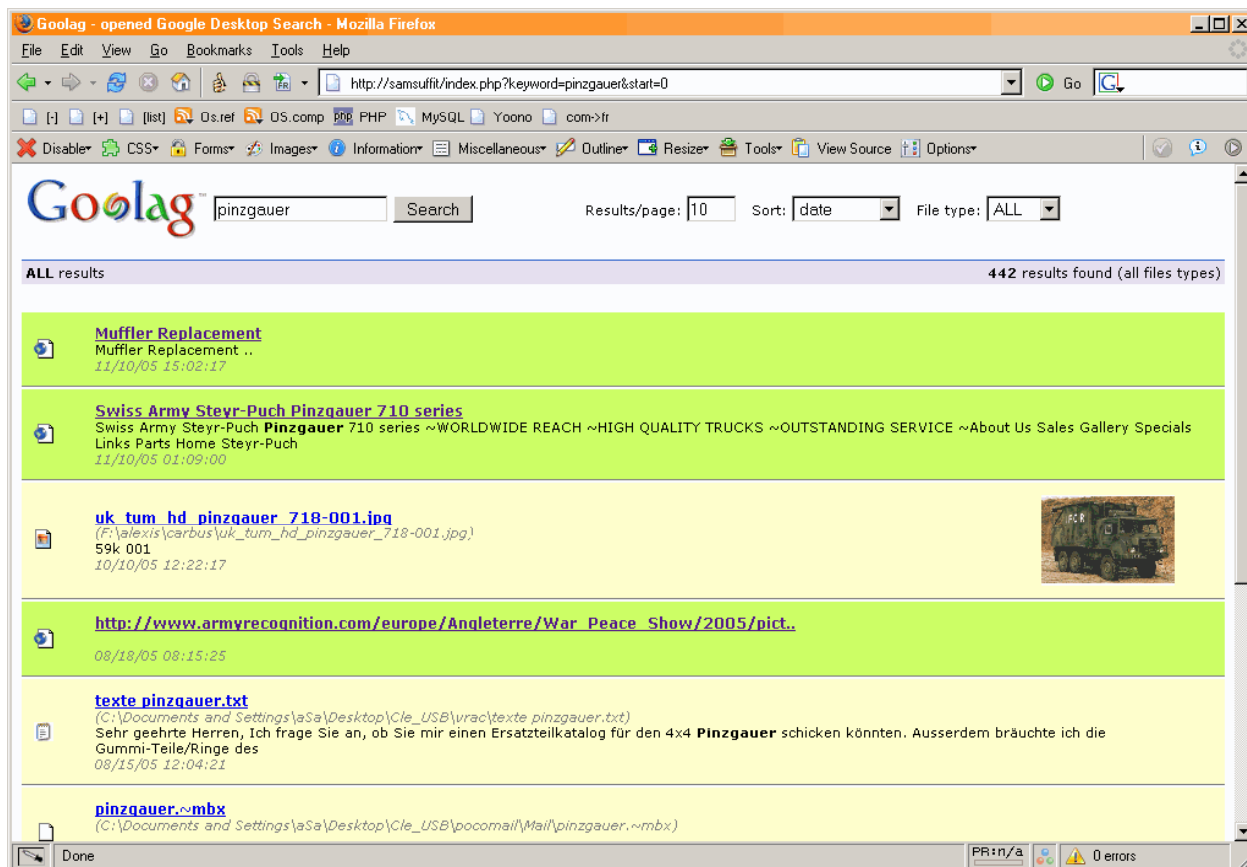
This is very handy for finding vulnerable sites wide open to google and googlebots



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

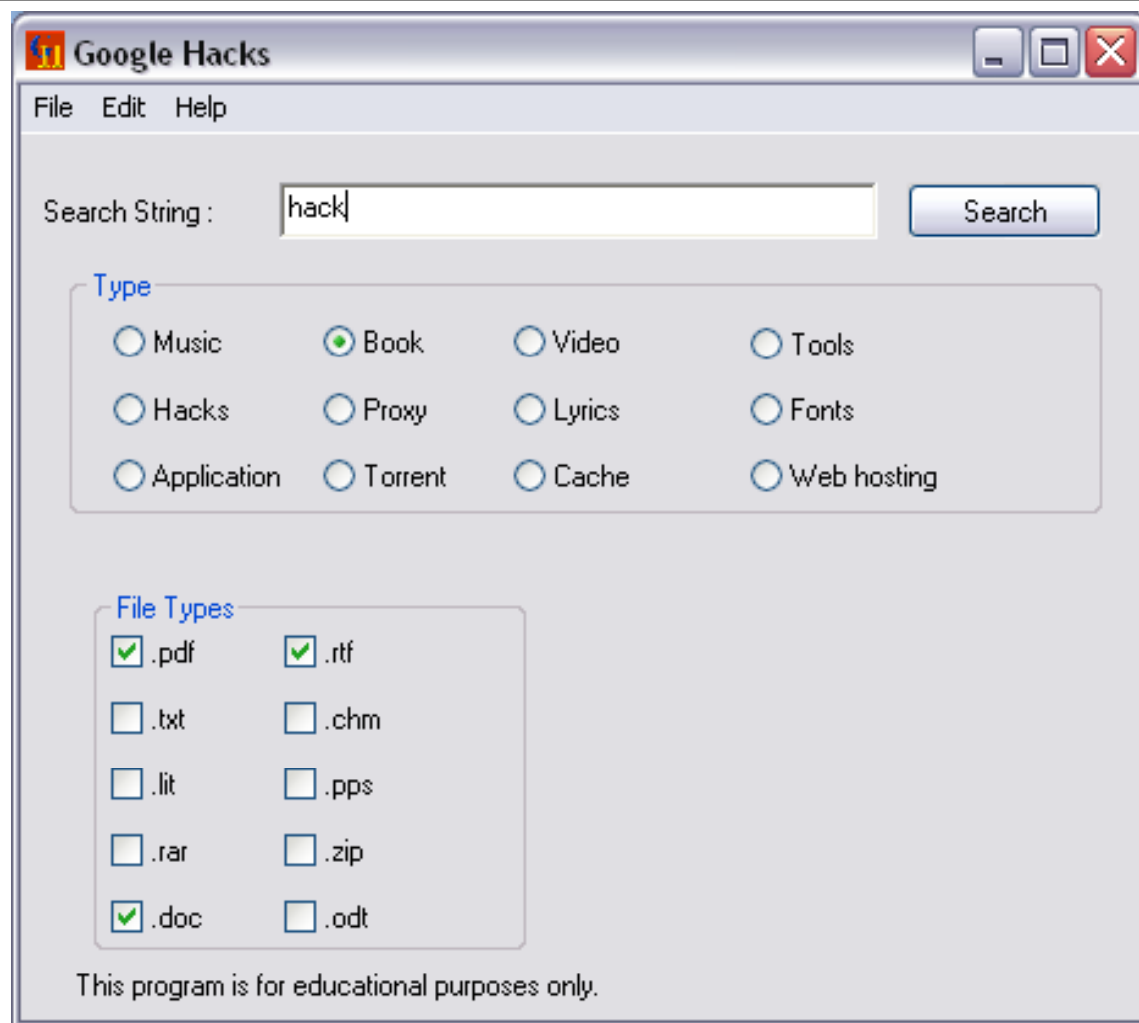
code.google.com/p/googlehacks/

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

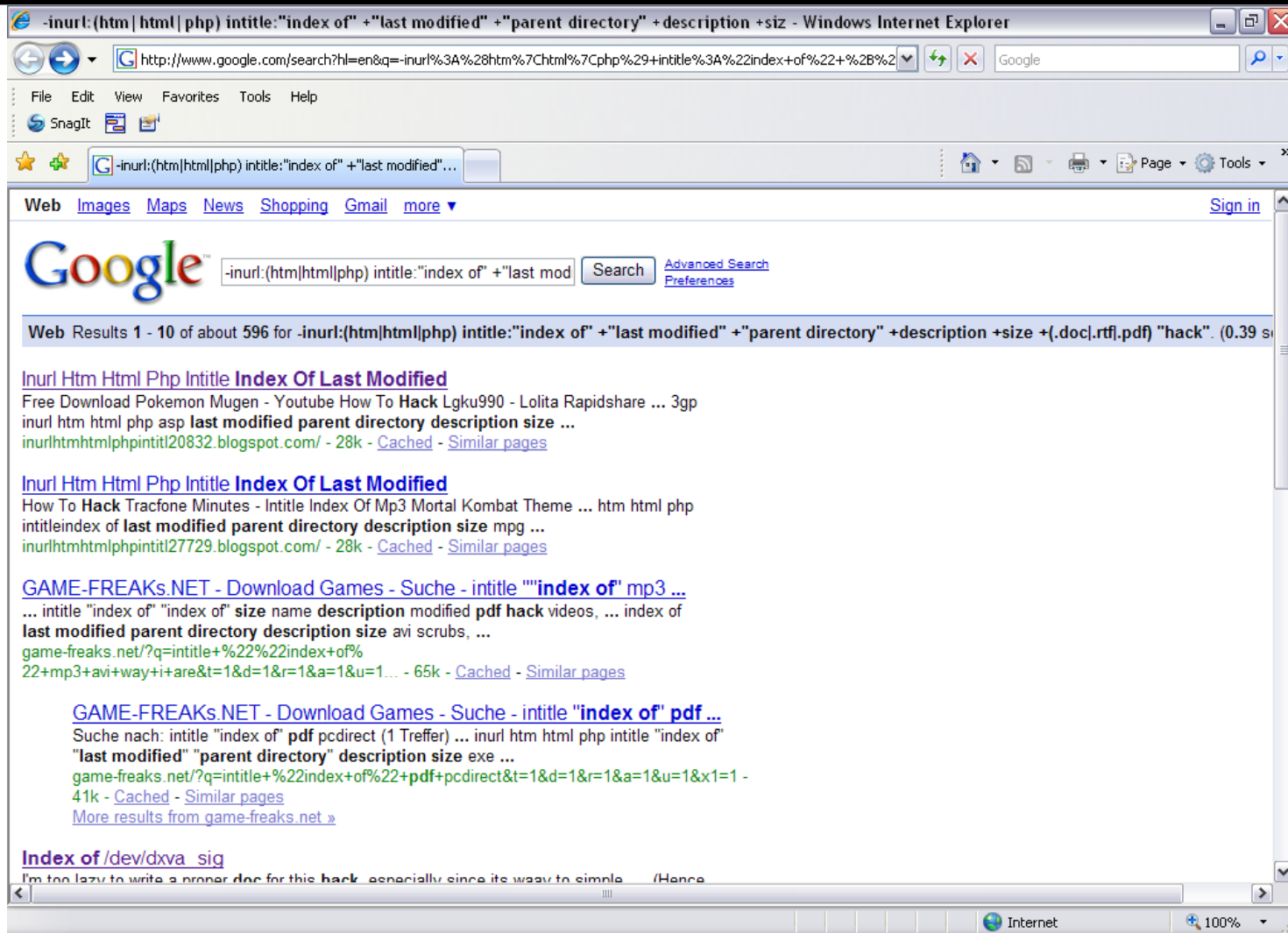
You can also use this program to use google as a proxy

Google Hacks: Screenshot



Google Hacks Input

Google Hacks: Screenshot



Google Hacks Output

Google Hack Honeypot

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



Google Hack Honeypot: Screenshot

THE WEB DECOY TRACK

BRIAN RYAN JOHN JAMIE MICHAEL
ENGERT MCGEEHAN RUERO RIDEN MUETER

GOAL 1

Combine strengths of GHM and PHPHop to create a more advanced, simple, low interaction honeypot tech.



"Google Hack"



PHP
HOP



Web-Deception
Framework

GOAL 2

Create working taxonomy for web application attacks in regards to honeypots, and develop concepts accordingly.

XSS

SQL

XML
RPC

PHP
BB2

SPAM



GOAL 3

Get an outline for a KYE paper, or similar to further web based honeynet development... or make a live cd.



Tool: Google Protocol

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

Certified Ethical Hacker



In this module, Google hacking techniques have been reviewed

The following Google hacking techniques have been discussed:

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



Copyright © Randy Glasbergen. www.glasbergen.com

Copyright 2004 by Randy Glasbergen.
www.glasbergen.com



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